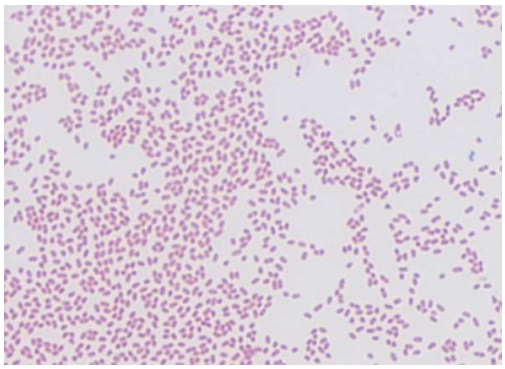
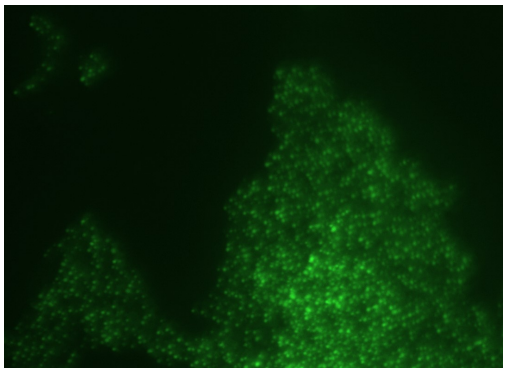




Methylocystis parvus: Genomics, Engineering and Sustainable PHB Production

---

By Benedict Hal Claxton Stevens **Supervised by Ying Zhang**

University of Nottingham
UK | CHINA | MALAYSIA

UKRI Biotechnology and Biological Sciences Research Council

PhD Thesis Submitted to the School of Life Sciences
University of Nottingham 2023

Viva performed by Thomas Smith (Sheffield Halam University)
and Edward Bolt (University of Nottingham)

Abstract

Methylocystis parvus OBBP is a gram-negative aerobic methanotroph of the phylum Alphaproteobacteria. The type species of its genus and type strain of its species. Methanotrophs are industrially relevant organisms defined by their ability to use methane as their sole carbon and energy source. Among many qualities *M. parvus* is capable of the production of poly(3-hydroxybutyrate) (PHB) which is biodegradable and if produced from biogas, is a bioplastic meaning it is produced from renewable biomass. PHB produced by methanotrophs thus has the potential to aid in the current climate crises of plastic waste and greenhouse gas emission fitting into a circular carbon economy.

Across this thesis this goal is developed towards first through genetics assembling a full genome of *M. parvus* OBBP and demonstrating CRISPR genetic editing in it. This forms the first published genetic editing in *M. parvus*. This is followed by chapters investigating PHB fermentation in methanotrophs particularly using real-world biogas comparing multiple strains and scaling production up to 1L fermentation bioreactors. A nanopore based 16S sequencing method for investigation of mixed culture fermentation species abundance is also demonstrated evaluating two PHB producing mixed cultures. Work in this thesis has contributed to three published papers which are included at the end.

Acknowledgements

For funding from the University of Nottingham BBSRC DTP and School of Life Sciences.

David Tooth and James Fothergill for analytics support and advice.

Ruth Cornock assistance on nanopore, keeping the labs show on the road, equipment training at the drop of a hat, passion for the nerdy small things and an uplifting feeling at every conversation.

My secondary supervisor Nigel Minton who has an unmatched talent to cut through the waffle.

My lab mates Hannah, Jake, Ourania and the rest of the SBRC PhDs. We shall all get through it together.

My students Adi Gurung and Asaph Kuria. You both provided a sorely needed refreshing breeze, a push to new heights, helping hands and a reminder of how far I've come.

My supervisor Ying Zhang for letting me spread my wings and patiently letting me learn from my own mistakes.

Huge thanks to my bench supervisor Bashir Rumah for his endless patience, expertise and unending hard work putting the rest of us to shame.

My parents Christopher Claxton Stevens and Diana Halliwell for dragging me as far as the beginning of my academic career. I count myself very lucky.

The forever loyal Odin

Last but never lest, my beloved partner in research, love and fun Maggie Johansen. You helped me both in science and in life more than I can ever say.

Declaration

The following assistance was received in the carrying out of this work with my gratitude:

Ruth Cornock loaded the nanopore flow cell for sequencing in the fermentation chapter.

Biogas analysis performed by Lower Reule Bioenergy Ltd who also produced and supplied the gas.

PHB derivatization and GC/MS operation carried out by James Fothergill and David Tooth.

Bashir Rumah assisted with experimental design and supervision of the biogas experimentation.

Bioreactor fermentations performed with technical support from Alejandro Salinas Vaccaro, Kamran Jawed, Victor Irorere and Swapnika Challa.

Contents

Abbreviations

Acc - Accession

AD - Anaerobic Digestor

bp – base pair

CE - Conjugation Efficiency

CoA – Coenzyme A

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

Ctrl - Control

dCas9 - Endonuclease Dead Cas9

DCW – Dry Cell Weight

dH2O - Distilled Water

DO - Dissolved Oxygen

EdE - Editing Efficiency

g – *g*-force of a Centrifuged Sample

GC - Gas chromatography

GC-MS - Gas chromatography-mass spectrometry

gDNA - Genomic DNA

HA - Homology Arm

HGT – Horizontal Gene Transfer

ICM - Intracytoplasmic membranes

IR - Infrared

Kan - Kanamycin

LBK50 - LB Agar Plates with 50µg/ml Kanamycin

LHA - Left Homology Arm

MeDH - Methanol Dehydrogenase

min - Minutes

MoFe - Molybdenum-Iron Nitrogenase

nCas9 - Cas9 Nickase

NDX - Nalidixic Acid

nfNMS - Nitrate Free NMS

NMR spectroscopy – Nuclear Magnetic Resonance Spectroscopy

NMS Medium – Nitrate Mineral Salt Medium

ntvCRISPR - Native CRISPR

ONT - Oxford Nanopore Technologies

PAM - Protospacer-Adjacent Motif

PDI - Polydispersity Index

PH4B - Poly(4-hydroxybutyrate)

PHA – Polyhydroxyalkanoate

PHB – Poly(3-hydroxybutyrate)

$PHB_{\%DCW}$ - PHB Percent by Dry Cell Weight

$PHB_{m/v}$ - PHB Productivity by Mass per Volume of Culture

pMMO – Particulate Methane Monooxygenase

PQQ - Pyrroloquinoline Quinone

RCF – relative centrifugal force

RHA - Right Homology Arm

RO - Reverse Osmosis

rpm – Rotations Per Minute

RuMP – Ribulose Monophosphate

SCP – Single cell protein

SDM - Site Directed Mutagenesis

sec - Seconds

sMMO – Soluble methane monooxygenase

T/AT - Toxin/Antitoxin

T_a - Annealing Temperature

t_e - Extension Time

Tm - Melting Temperature

TraDIS - Transposon Directed Insertion Sequencing

UHMW-PHB –Ultra-high molecular weight PHB

VFe - vanadium-iron nitrogenase

WT - Wild Type

-ve - Negative

+ve - Positive

Chapter 1: Introduction and Background

1.1 Introduction

In this introduction I will cover the major topics surrounding my project: Methanotrophic bacteria and their core metabolism, PHB (polyhydroxybutyrate) bioplastic production, extraction and usage focusing on methanotroph species and an overview of methanotrophs as industrial production chassis and their genetic engineering prospects.

With this knowledge, data are then presented in later chapters developing knowledge surrounding the industrial application of methanotrophs and my main research organism *Methylocystis parvus* OBBP in PHB production. This is first approached through genetics assembling a full genome of *M. parvus* OBBP and demonstrating CRISPR genetic editing in it. This is followed by chapters investigating PHB fermentation in methanotrophs particularly using real-world biogas comparing multiple strains and scaling this up to 1L fermentation bioreactors. A nanopore based 16S sequencing method for investigation of mixed culture fermentation species abundance is also demonstrated. I have contributed to, or authored three papers as part of this work which are noted in their applicable chapters.

1.2 The Methanotrophs

Methanotrophs are microorganisms defined by their ability to use methane as their sole carbon and energy source, thus they must be capable of all reduction of ATP, NAD(P)H and producing all carbon bonds *de novo* from 1C carbon sources. They can be bacterial or archaeal with bacterial methanotrophs being almost exclusively aerobic and archaeal methanotrophs found in anaerobic or anoxic environments[1,2]. Although bacterial methanotrophs are not generally extremophiles, they are commonly found in many diverse environments of varying temperature (0-70°C), pH (1.5-12) and salinity (0-24%). Species also have diverse oxygen requirements and can grow in aerobic, anaerobic (0.5-60% O2) and rarely anoxic environments but are most prosperous at aerobic/anaerobic interfaces like swamps, soils and sediments along with manmade waste like sewage and landfill[3–6]. Species are also known to form symbiotic relationships with surface plants and marine organisms around hydrocarbon seeps[3,5]. Due to their ability to remove methane, a potent greenhouse gas, from the environment, they are essential in the reduction of the greenhouse gas effects contributed by methane from anthropogenic, biological or geological sources by degrading it in their local environment before it escapes to the atmosphere or by metabolising atmospheric methae[3,5] and as such have been attributed the largest global methane sink[5–7].

Table 1: Characteristics of Proteobacterial methanotroph types separated by family or Type

	Type I	Type X	Type II	Type II
Phyla	γ-proteobacteria	γ-proteobacteria	α-proteobacteria	α-proteobacteria
Family	*Methylococcaceae*	*Methylococcaceae*	*Methylocystaceae*	*Beijerinckiaceae*
Genera	*Methylobacter* *Methylohalobius* *Methylomicrobium* *Methylomonas* *Methylosoma* *Methylosarcina* *Methylosphaera* *Methylothermus* *Crenothrix* *Clonothrix*	*Methylococcus* *Methylocaldum*	*Methylosinus* *Methylocystis*	*Methylocapsa* *Methylocella* *Methyloferula* *Methylorosula,* *Beijerinckia* *Methylovirgula*
Resting stages	*Azotobacter*-type cysts or none	*Azotobacter*-type cysts	Lipoidal cysts or exospores	*Azotobacter*-type cysts or exospores
ICM	Central Disks	Central Disks	Peripheral	Varies
pMMO	+	+	+	Except M*ethylocapsa*
sMMO	Varies	Varies	Varies	Varies
PHB production	-	-	+	+
Carbon assimilation	RuMP	RuMP	Serine	Serine
CBB cycle	-	+	-	-
Major fatty acid chain length	16	16	18	18
Major quinone	Q-8 or MQ-8	MQ-8	Q-10	Q-10
Mol% G+C (Tm)	43–60	56–65	60–67	55–63

+ 90% or more of strains are positive and – 90% or more of strains are negative. CBB - Calvin-Benson-Bessham. ICM - intracytoplasmic membranes. Table adapted from Bowman 2006[5] and Semrau et al. 2010[8] with additions from Alexey et al. 2011[9], Tamas et al. 2014[10], Vorobev et al. 2011[9], Marín, Irma and Ruiz 2014[11] and Dedysh et al. 2002[12]. Not all listed *Beijerinckiaceae* genera are confirmed methanotrophs. sMMO labelled as varies as it has been found in all listed families irregularly.

Most methanotrophs are obligate though a minority are facultative. They are a subset of the methylotrophs; A group of organisms defined by their ability to utilize reduced carbon substrates with no carbon-carbon bonds as their sole carbon and energy source. Methylotrophs utilise a range of C1 compounds including methane, methanol, formate, methylamine, and formamide[3–7]. Some methanotroph species are also capable of methanol metabolism and uptake from their environment, while to others methanotroph species it is highly toxic[5,13].

By some definitions methanotrophs qualify as mixotrophs or heterotrophs depending on the source of methane utilised and considering the

fixing of CO_2 that takes place. The term heterotroph is commonly used use in literature to describe non-methanotrophic heterotrophic bacteria however this is generally incorrect usage. Thus we here use “polytroph” to define non-methanotrophic heterotrophs as suggested by De Marco[14] referring to bacteria that utilise organic substrates with more than one carbon bond, e.g. other than methane and methanol thus excluding methanotrophs and methylotrophs from this group.

The bacterial methanotrophs are a polyphyletic group (a group with multiple independent origins) made up largely of two major classes within the Gram -ve Proteobacteria: γ-proteobacteria, family *Methylococcaceae* (Type I and Type X methanotrophs, Also Called Group I), α-proteobacteria, family *Methylocystaceae* (Group II/Type II methanotrophs)[4,5,15] and family *Beijerinckiaceae*. Other methanotrophs have also been found outside the Proteobacteria including a number of members in the phyla Verrucomicrobia (Group III), and candidate phyla NC10[3,6,16–18] and most recently a mycobacterial species[19]. The details of the Proteobacterial species are summarised in Table 1.

More complete phylogenies have been often disputed between authors and are still emerging but the generally accepted versions are available from Bowman et al. 1993[4] and 2006[5] though these are now somewhat incomplete[3,7–9]. The most up do date phylogenetic studies have been performed by Tamas et al. 2014[10] and Kang et al. 2019[20]. The Type I, II and X definitions are becoming outdated with simplified grouping as α- and γ-proteobacterial methanotrophs becoming more commonplace[21]. Unless specifically differentiating Type X species the α- and γ-proteobacterial methanotroph notation is used in this work. The “group” system seems used in isolated cases in the literature but not commonly[17]. There are likely a large diversity of methanotrophs that have yet to be isolated with some methanotrophic genes isolated from environmental samples that are yet to be aligned with known species and some known but unculturable[3,5] leaving potential for new strain identification to fit industrial niches.

Of the methanotroph species the most studied are within the Proteobacteria and among these species those of specific interest for this work are *Methylococcus capsulatus* Bath and *Methylocystis parvus* OBBP. *M. capsulatus* Bath is a Type X/γ-proteobacterial methanotroph strain and one of the prevalent methanotroph research strains, it has been used in SCP (Single Cell Protein) production at industrial scale and is thermophilic with a preferred temperature of 45°C[4,22,23], notably it is not the type strain of its species which is “Texas” which is less well researched[5]. *M. parvus* is the Methylocystis genus type species[5,13,24] and OBBP is the species type strain and a common organism for Type II/α-proteobacterial methanotroph and PHB bioplastic research[4,5,24]. Both are obligate methanotrophs. Although further strains and species are likely to be of interest any work performed on these two strains is likely tobe generalisable. *Methylosinus trichosporium* OB3b is also a notable research organism for PHB and α-proteobacterial methanotroph research[4,25,26].

1.2.1 *Methylocystis parvus* OBBP

The first methanotroph study was carried out in 1956 by Dworkin and Foster[27] on *Pseudemonas* (now *Methylomonas*) *methanica* a γ-proteobacterial methanotroph. A great number of species, including *M. parvus* were later isolated in a mass collection and description of methanotrophs by Whittenbury et al. 1970[13]. It is almost certain this is the source of the *M. parvus* OBBP strain. This is slightly clouded as the NCIMB documentation for *M. parvus* OBBP (as NCIMB 11129) states isolated by R. Whittenbury, Date of Isolation 01/01/1974, Date of Accession 01/11/1974. 1974 postdates the publication of the 1970 paper. The 1970 paper also does not specifically name the *M. parvus* strain isolate. A later paper in 1979[28] uses the specific title *M. parvus* OBBP and states this was received from R. Whittenbury. Regardless however all usage in the literature cites *M. parvus* OBBP to be from the 1970 paper and this is most likely true, the 1974 date in the NCIMB being in error. Bowman et al. 1993[4] cites Romanovskya et al. 1978[29] for description of *M. parvus* but this has since been considered in error by later publications[30,31] reverting to the Whitunbury et al. 1970[13] originating attribution.

The isolation location of *M. parvus* OBBP described by Whittenbury et al.[13] is vague being "Mud and water (from ponds, rivers, streams and ditches), and soil samples, obtained from the U.K., the European continent (France, Germany and Russia), North America, South America, East and North Africa and Egypt, were used as inocula" and this accounting for a wide variety of isolated species and genera in the paper thus *M. parvus* OBBP's origin cannot be confirmed.

The Latin translation of the species name of *M. parvus*, although not identified in the originating paper[13], means small. Named for the cells very small size as observed in Figure 1D described to be 139 ± 20 nm and a width of 65 ± 12 nm[32]. They are significantly smaller than *M. trichosporium* (Figure 1C ~750nm) and *M. capsulatus* Bath (Figure 1E). *M. parvus* is described and observed to be rod shaped (or vibroid[13]) and Gram -ve often forming rosettes and lipoidal cysts[13,32]. The strain name is short for Oddball – British Petrolium (Personal communication Thomas Smith)

A sequence of *M. parvus* OBBP, broken into 108 contigs without full assembly was released in 2012 (GC 63.40%, 4.47591Mb, 4128 predicted proteins, GenBank GCA_000283235.1)[24]. A fully closed genome sequence has since been produced for *M. parvus* strain BRCS2 isolated from a bog in Mosely, UK (GC 63.35%, 4.52904Mb, 4185 predicted proteins, Genebank GCA_009685195.1)[32]. This consisted of a major chromosome of (4.08Mb CP044331.1) and two unnamed plasmids numbered 1 (0.25Mb, CP044332.1) and 2 of 0.20 Mb (CP044333.1). I assembled a more complete genome of *M. parvus* OBBP presented in this thesis in Chapter 3.

Bashir et al. 2021[32] also predicted restriction modification systems through submission to the REBASE database. Full data available in the supplementary information and from REBASE but notably "CTCGAG" is a predicted methyltransferase target which is the same as the commonly used

XhoI restriction enzyme. This site may need to be avoided in molecular biology applications in *M. parvus*.

1.2.2 Physiology and Methane Metabolism

Proteobacterial methanotrophs are often spore forming during stationary growth phase with Type I and X forming Azobacter-type cysts which allow prolonged survival. Type II may form exospores which are heat resistant to 85°C and can survive drying for over a year or lipoidal cysts, as in *M. parvus,* which are rich in PHB and are desiccation resistant[4,5]. Oxygen starvation experiments in *M. parvus* have indicated PHB can be fermented over a long period of time prolonging cell life until more favourable conditions are available. This state can be maintained for at least three months[33].

A major signature of Proteobacterial methanotrophs are intracytoplasmic membranes (ICM)[5,7] arranged internally as discs in Type I and X, and peripheral membranes in Type II (Figure 1A and B respectively)[4,7]. These function as subcellular compartments housing surface bound particulate methane monooxygenase (pMMO), the electron transport complexes and ATP synthase and are a major location of methane metabolism[3,5]. By varying the amount of ICM the bacteria can increase the surface area available for oxidation in response to methane and oxygen availability[5].

The methane metabolism pathway illustrated in Figure 2A and B is oxygen dependent with methane oxidation to methanol being carried out by MMO (methane monooxygenase). From this point metabolism follows a similar pathway to other methylotrophs[6,7] with methanol further oxidised to formaldehyde by methanol dehydrogenase thus allowing for the subsistence on methanol possible in some facultative methanotrophs[3,5]. Formaldehyde metabolism is rapid as build up is usually highly toxic[6]. Part of the produced formaldehyde is passed to dissimilatory metabolism and oxidised to formate then CO_2 to generate NADH to drive electron transport and ATP synthesis the remainder is assimilated as a carbon source for anabolism[5,7]. Carbon assimilation from formaldehyde differs between methanotroph groups. γ-proteobacterial methanotrophs use the RuMP (ribulose monophosphate) pathway and α-proteobacterial methanotrophs the serine cycle[3,5]. Phylum NC10 and Verrucomicrobia use the Calvin-Benson-Bassham cycle assimilating CO_2 rather than formaldehyde[7]. Of these, the RuMP pathways is more thermodynamically efficient for the incorporation of carbon[3].

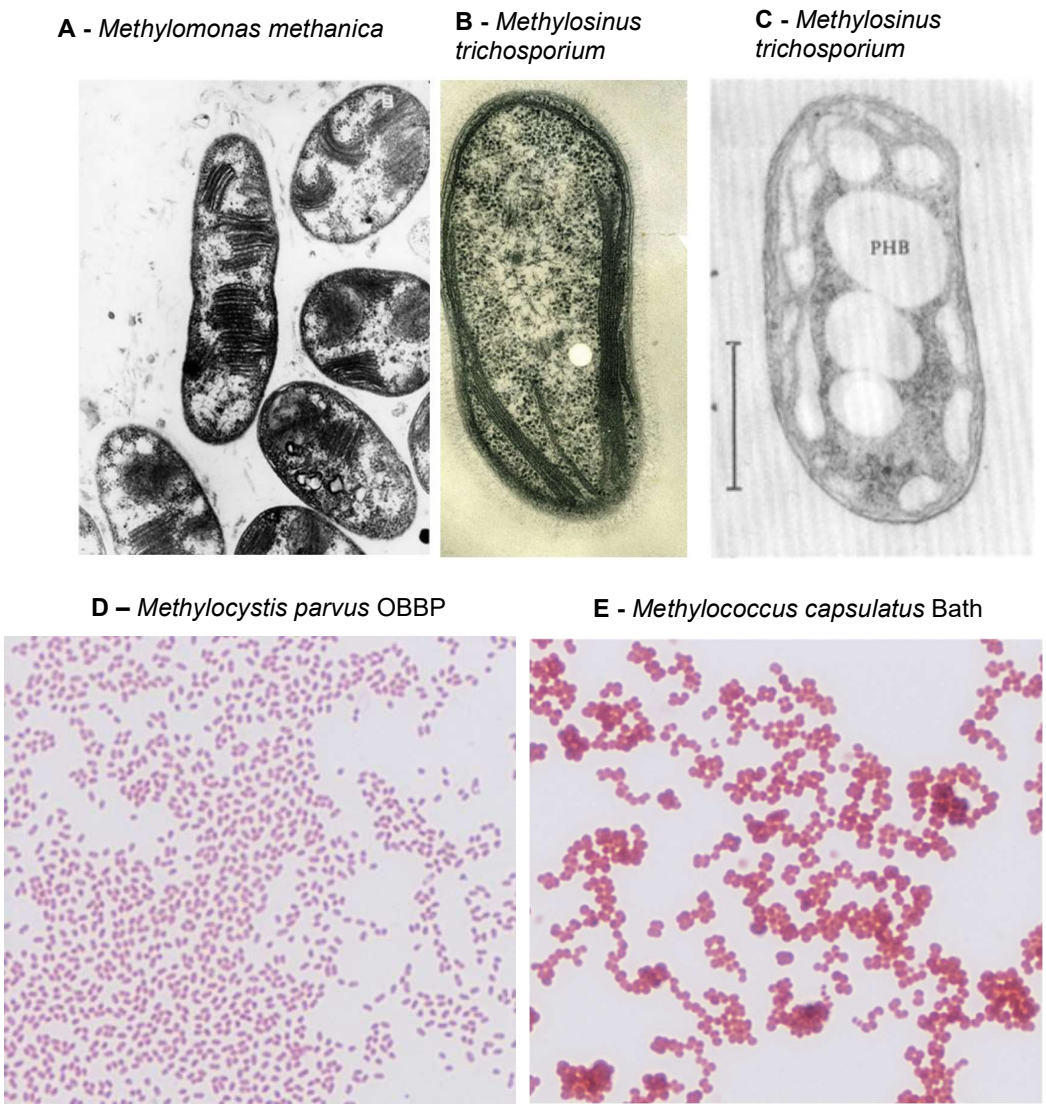
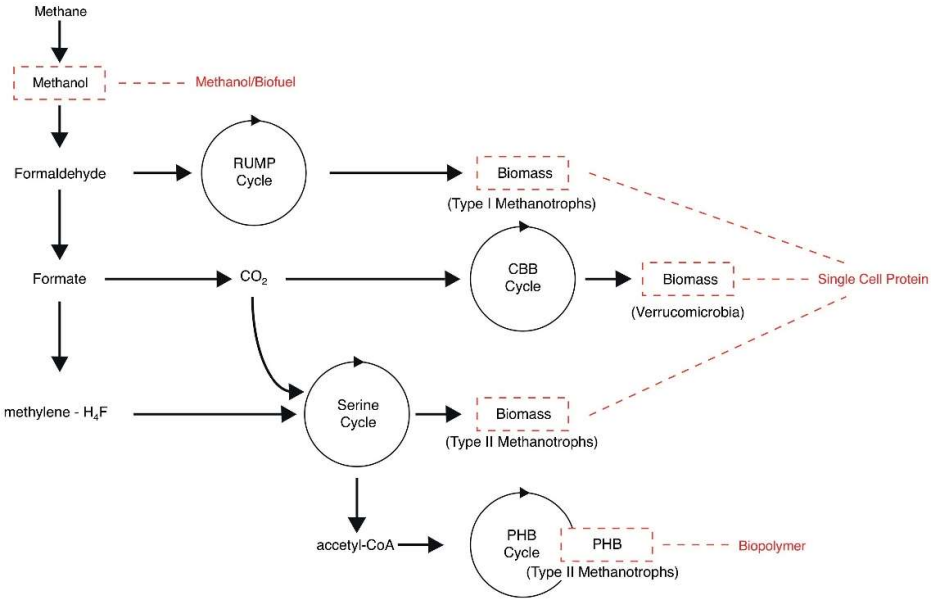


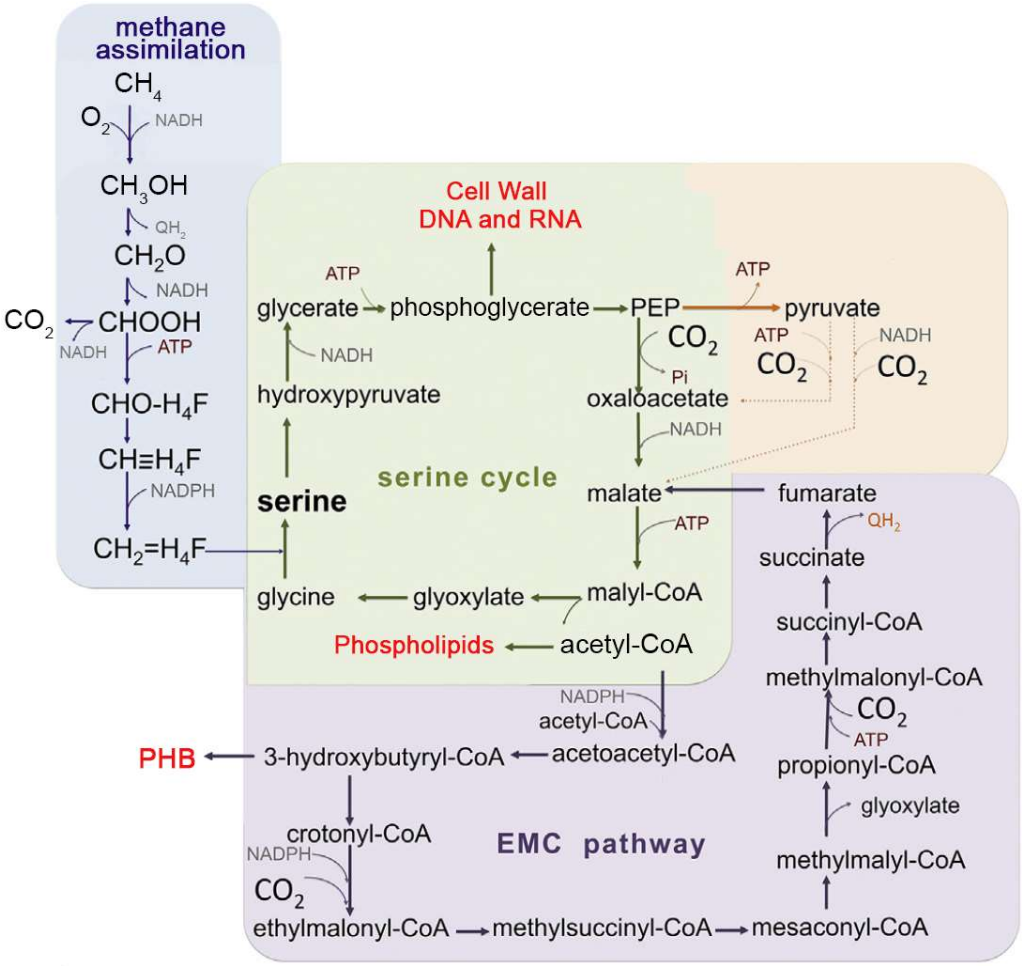
Figure 1A-E: Cross sectional electron microscope images of **A** - *Methylomonas methanica*, γ-proteobacteria, showing internal bundles of ICM. 65,000x magnification. **B** - *Methylosinus trichosporium*, α-proteobacteria, showing ICM along the cell periphery. The white spot is likely a small PHB granule but unconfirmed. 75,000x magnification. Images **A** and **B** by H. Dalton from Kalyuzhnaya et al. 2019[3] **C** - *Methylosinus trichosporium* OB3b showing packed PHB granules. Bar marker represents 0.25 µm. From Best and Higgins 1981[34]. **D** and **E** Show *M. parvus* OBBP and *M. capsulatus* Bath respectively imaged for this thesis at 1,000x magnification and Gram stained with a Remel R400080 Kit following manufacturers protocol. Imaged on a Nikon eclipse Ci with a Nikon CFI Plan Achro DL 100X Oil (MRL21903) objective and DS-Fi2 digital camera in bright field mode.

Figure 2A and **B**: **A** - Diagram of the generalised biochemical assimilatory pathways of three groups of proteobacterial methanotrophs from methane to industrially relevant products shown in red. The tetrahydrofolate (H4F) pathway is shown but a tetrahydromethanopterin (H4MPT) pathway for formaldehyde oxidation that operates in parallel. RuMP - ribulose monophosphate CBB - Calvin Benson-Bassham; PHB - poly-3-hydroxybutyrate. Adapted from Pieja et al. 2017[16] **B** - Pathway of methane assimilation and core metabolism in α-proteobacterial methanotrophs showing the serine and ethylmalonyl-CoA (EMC) pathways. Not all links are shown including amino acid production. Modified from Kalyuzhnaya 2016[35] with *M. trichosporium* OB3b experimental data from Yang et al. 2013[36].

A


Methane
Methanol
Methanol/Biofuel
Formaldehyde
RUMP Cycle
Biomass
(Type I Methanotrophs)
Formate
CO_2
CBB Cycle
Biomass
(Verrucomicrobia)
Single Cell Protein
methylene - H_4F
Serine Cycle
Biomass
(Type II Methanotrophs)
accetyl-CoA
PHB Cycle
PHB
Biopolymer
(Type II Methanotrophs)

B


methane assimilation
CH_4
O_2
NADH
CH_3OH
QH_2
CH_2O
NADH
CO_2
CHOOH
NADH
ATP
$CHO-H_4F$
$CH{\equiv}H_4F$
NADPH
$CH_2{=}H_4F$
Cell Wall
DNA and RNA
ATP
glycerate
phosphoglycerate
PEP
ATP
pyruvate
ATP
CO_2
NADH
CO_2
NADH
hydroxypyruvate
CO_2
Pi
oxaloacetate
NADH
serine cycle
serine
malate
fumarate
ATP
QH_2
succinate
glycine
glyoxylate
malyl-CoA
succinyl-CoA
Phospholipids
acetyl-CoA
methylmalonyl-CoA
NADPH
acetyl-CoA
CO_2
ATP
PHB
3-hydroxybutyryl-CoA
acetoacetyl-CoA
propionyl-CoA
glyoxylate
crotonyl-CoA
NADPH
CO_2
EMC pathway
methylmalyl-CoA
ethylmalonyl-CoA
methylsuccinyl-CoA
mesaconyl-CoA

The RuMP cycle and serine cycle in relation to methanotrophs are fully described in Lidstrom 2006[7] and Anthony 1982[37]. In synopsis, both condense a C_1 compound derived from formaldehyde to another multicarbon acceptor which is cycled up until a C_3 product is cleaved and the acceptor is regenerated. This C_3 product is used in further anabolism. The serine cycle also draws carbon from CO_2 in addition to formaldehyde. This has historically been described as one carbon from CO_2 for every two from formaldehyde[7,37] but more recent authors have found a one formaldehyde to two CO_2 ratio due to new discoveries in the enzymatic pathway[38] among many other considered stochiometric ratios and this is a point of current investigation[35,36].

The methane oxidation to methanol step is performed by MMO which comes in either a soluble form (sMMO- E.C.1.14.13.25) encoded by the genes *mmoXYBZDC*[20] or a particulate (pMMO - EC 1.14.18.3) membrane bound form encoded by *pmoCAB*[5,6]. The particulate and soluble enzymes are biochemically and structurally different. sMMO is a non-haem iron enzyme complex and pMMO is membrane bound and copper reliant with some species also using iron[5,7]. All methanotrophs with the exception of *Methylocella* and *Methyloferula*[9] possess pMMO but sMMO has an irregular distribution across the Proteobacterial species with *M. capsulatus* and *M. trichosporium* being notable possessors of both. *M. parvus* has been found not to possess the sMMO genes or activity[24]. The MMO enzymes are often present in multiple copies. If both types are present, expression is selective depending on copper availability with pMMO favoured in high availability[3,7,25]. sMMO has a notably broad substrate specificity with over 250 known substrates and as such is of major interest in industrial applications and in bioremediation by degrading contaminants including phenol, TCE and cholorofluro-benzines[5,7,17].

Some studies have also found an additional homologue related to pMMO named pXMO that differs in sequence significantly and has not been ascribed a purpose and investigation is pending[39–41]. It has not been identified in *M. parvus* but is present in closely related *Methylocystis* species *rosea* and *hirsuta*[42].

Further detail of methanotroph metabolism pathways has been excellently explained by Kalyuzhnaya 2016[35].

1.2.3 Gene Lineages and Horizontal Gene Transfer

sMMO and *pMMO*s are disparately spread among the polyphyletic of methanotrophs encompassing four separate lineages of members in Verrucomicrobia, NC10, α-proteobacteria and γ-proteobacteria[10,20]. It is believed a horizontal gene transfer (HGT) event passed an ammonia monooxygenase from nitrifying bacteria to a methylotroph where it acted as an MMO forming the first methanotroph within the γ-proteobacteria[20,43]. This then passed by HGT into a methylotroph within the α -proteobacteria which passed through vertical transfer into a monophyletic clade of the *Methylocystaceae* and *Beijerinckiaceae*[10]. A methylotroph is the obvious candidate for this transfer as it already possesses the downstream mechanisms to use the produced methanol and formaldehyde the build-up of which is

otherwise toxic[10]. It is believed the Verrucomocrobial and NC10 methanotrophs received their MMOs in separate transfer events[18,41].

This gain of methanotrophic function within α-proteobacterial and γ-proteobacterial methylotrophs raises the question of why these groups, but not other methylobacterial groups including β-proteobacteria, Actinobacteria, Firmicutes and even eukaryotes have not got examples of gained methanotroph function[20]. This is an open question but Kang et al.[20] suggested presence of a copper uptake systems to supply pMMO may be a deciding factor; they are identified in nearly all methanotroph species except the Verrucomicrobia who are explained to live in already metal rich habitats. It has also been theorised by Tamas et al.[10] that loss of methanotrophy once gained is difficult, at the time of their writing only one species of any of the 4 methanotroph clades, *Beijerinckia indica* had been shown to have subsequently lost the capability. Beijerinckiaceae also consists of nearly all the facultative methanotrophs compared with the other clades being almost exclusively obligate.

Due to its almost universal presence in the methanotrophs, pMMO was almost certainly the original MMO and the work of Kang et al.[20] also strongly supported pMMO HGT. Osborne and Haritos[41] instead suggest *pMMO*, sMMO and pXMO were present in the originating gene arrangement and were inherited together and then individually lost. pMMO and sMMO appear to have developed from different progenitor genes: an ammonia monooxygenase and a "bacterial multicomponent monooxygenase"[41].

In some cases including *M. parvus*, the multiple copies of pMMO genes follow distinct lineages denoted *pmoCAB1* and *pmoCAB2* which possess different kinetic properties[10,44]. *Methylosinus* and *Methylocystis* species all possess two copies of *pmoCAB1* and one of *pmoCAB2*[10,42]. Resolving these lineages will be investigated more closely in section 3.4.3.6.

Horizontal gene transfer of the *nifH* and *nifD* Nitrogen fixation genes has also been shown[45] with more data added by Tamas et al. 2014[10] who suggested that the common ancestor of α-proteobacterial methanotrophs was capable of dinitrogen fixation.

1.2.4 Culturing Methanotrophs

Lab culturing techniques for mesophilic methanotrophs most commonly use purified methane as energy and carbon feedstock mixed with air or pure oxygen. Natural gas, although methane containing, can have a negative effect on growth due to acetylene content which was identified as a suicide substrate for MMO thus is highly inhibitory. In addition, other variable contents within natural gas are also inhibitory so it is not generally used in a regular lab context[5,25]. This is discussed more fully in section 6.2.4.1.

The standard medium is NMS (Nitrate Mineral Salt) described by Whittenbury et al. in 1970[13] providing magnesium sulphate, calcium chloride, monopotassium phosphate, disodium phosphate and potassium nitrate as a nitrogen source, trace vitamins and minerals and NaFeEDTA as an iron source

for sMMO[3,13]. NMS can be used either as a liquid medium or solid with the addition of agar. Modifications of NMS using the same name often include additions of molybdate for nitrogenases in nitrogen fixation and copper to provide the required cofactors for pMMO[3]. Although ammonia can be substituted as a nitrogen source it is a competitive inhibitor of MMO and thus nitrate is preferred in some cases, however stoichiometry shows ammonia assimilation is less energy intensive so has been drip fed into batch cultures at larger scale[5,46,47]. Capability to remove toxic byproducts of ammonia metabolism like hydroxylamine, nitric oxide and nitrite is only possible in some species which possess the required enzymes like hydroxylamine oxidoreductase[10], presence of these genes or engineering them in may therefor make a process organism more desirable. A variety of other media have been published to fit a variety of conditions and specific species, particularly for novel species isolation[3]. Examples that have successfully been used on *M. parvus* include W1 medium[48,49] and JM medium[50].

The common lab technique for liquid culture uses small scale batch growth adding volumes of gas to inoculated sealed media in bottles and shaking resulting in low surface area for transfer of methane and gas into the media, this causes lower growth rates and densities[3]. Gas-liquid mass transfer is a major limiting factor in growth due to low solubility of methane and oxygen[46,51]. In pure water at 1atm and 20°C they are 25 and 44mg/L respectively and at 30°C 21 and 28 mg/L (2sf)[52]. Solubility decreases for both with increase in temperature and is also dependent on partial pressures. Under gas mixing and especially with the diluting effect of using oxygen from air, the solubility will decrease with proportion of other gases utilised as long as the overall pressure remains constant as according to partial pressure and Henry's law[21,53,54].

Conventionally in the lab Methane:Air mixes are supplied in which only the 21% Oxygen constituent of air is of interest excepting N_2 fixation. Higher partial pressures can be achieved using Methane:Oxygen with doubling times of 6.45h being achieved in *M. parvus* OBBP[49]. Industrially oxygen or oxygen enriched air is the standard method however this adds to production cost. Use of stirred bioreactor designs can improve mass transfer beyond bottle scale and is discussed in Chapter 6. Use of high percentage oxygen and methane increase the dissolved gases but this creates an unsafe environment of flammable gas so must be considered carefully[16,21]. Gas flow rates used are limited in our lab setting by the lower and upper explosive limits (LEL and UEL) for safety. Below these limits the mixture is too lean to burn and above this too rich.

It has been demonstrated that many methanotrophs including *M. parvus* OBBP can be grown on methanol vapour in place of methane[28,32]. *M. parvus* has been cultured on up to 2%$_{v/v}$ methanol by direct addition to the medium however these cells grew in an elongated and irregular shape 3-4x longer than when grown on methane[33].

1.3 Polyhydroxybutyrate and Methanotrophs

1.3.1 PHB an Overview

Due to the high accumulation of plastic waste in landfills and sea islands, as well as contamination of aquatic environments with microplastics all having a negative effect on local ecology and also possible human health effects, biodegradable plastics that degrade in the environment are of high industrial and public interest[55,56]. An estimated 10% of all plastics end up in the sea including fishing equipment like nets. Making these from biodegradable plastics would reduce the plastic build up in the oceans drastically[57]. Most biodegradable bioplastics and polymers fit into three groups: bacterial polymers like PHAs (polyhydroxyalkanoates); plant polysaccharides like as cellulose and chitin; and chemically synthesised polymers like PLA (polylactic acid)[58].

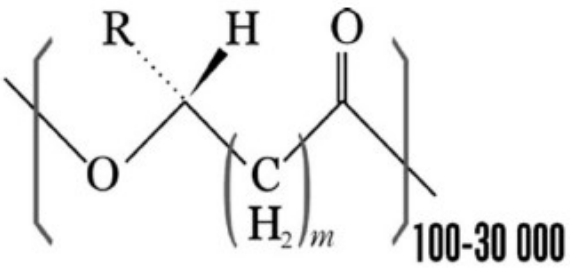
PHB – poly(3-hydroxybutyrate), is a type of PHA: hydroxy fatty acid polymers which have potential as commercial plastics. They are exciting industrially due to their biodegradability, relatively good material properties, biocompatibility and as a bioplastic – a plastic that can be produced from renewable biomass. PHAs are produced inside a broad range of bacteria, where they act as cellular energy and carbon reserves, more specifically PHB which is produced within a range of methanotrophs including *M. parvus* OBBP[15,59].

1.3.2 PHA Material Properties and Co-Polymers

The first discovered and most commonly produced PHA was PHB first described by Maurice Lemoigne, 1926 (in French)[60,61]. Despite their early discovery they received little industrial success due to their inability to compete with low cost petrochemical plastics, a state of affairs that largely continues to this day due to the cost of PHA extraction and feedstock. Increase in environmental concern and avoiding fossil fuel use has invigorated research in the area[59,62].

There is great diversity in the PHA family of thermoplastic biopolymers based around the core hydroxyfatty acid polymer, the most common of which are presented in Figure 3. All biological PHAs are chiral in the R chiral configuration and many enzymes are selective for this configuration[61]. A variety of pendant groups and monomer sizes from C1 to C13 have been observed from biological sources[59]. The properties of PHAs particularly PHB are usually compared closely to those of the fossil fuel derived nonbiodegradable polypropylene[55,59]. PHB itself is described as highly crystalline, stiff but brittle behaving like a hard elastic material. It is superior to many other bioplastics which are often water soluble, moisture sensitive and/or are much more vulnerable to hydrolytic degradation[63]. The properties can be further improved by formation into a co-polymer interleaving multiple PHA types as opposed to a homopolymer that contains only one; poly(3-hydroxybutyrate-co-3-hydroxyvalerate) (PHBV) is a common example[59]. PHBV and other co-polymers are less stiff and brittle and can have a lower

melting temperature but retain other positive properties[50,59], this is evidenced by the comparison of the properties of PHB and some copolymers listed in Table 2. These properties can be further improved by optimised drawing and annealing techniques[58], as high as 1,320 MPa and 96% extension to break for PHB[58]. These material properties allow PHB to be extruded, moulded, spun into fibres and made into films[64].



m=1	R=H	Poly(3-hydroxypropionate)	P(3HP)
	$R=CH_3$	Poly(3-hydroxybutyrate)	P(3HB)
	$R=C_2H_5$	Poly(3-hydroxyvalerate)	P(3HV)
	$R=C_3H_7$	Poly(3-hydroxyhexanoate)	P(3HX)
m=2	R=H	Poly(4-hydroxybutyrate)	P(4HB)
m=3	R=H	Poly(5-hydroxyvalerate)	P(5HV)

Figure 3: Structure of various PHAs generalised to their core structure. Variations in the pendant group R and the length of the number of core carbons "m" define the PHA type. from Jacquel et al. 2008[65]

In addition to the choice of polymer and/or copolymer, the properties of the final plastic are dependent on polymer length which varies greatly with biological polymer sizes, evaluated as molecular weight, with sizes ranging from 100 to 30,000 monomers but may be much larger with 130,000 having been shown[58,59] (see UHMW PBH in Table 2). Previous feeding experiments have shown that co-polymers can be produced *in vivo* by feeding a combination of substrates to certain bacteria, for example glucose and valerate for PHBV[56,59]. PHBV and other co-polymers have also been successfully produced in wild-type α-proteobacterial methanotrophs including *M. parvus* OBBP by feeding with a combination of methane and the fatty acids valerate or propionate [50,66,67]. It may also be possible to produce valerate though installation of exogenous enzyme pathways to produce non-PHB PHAs and a variety of other approaches have been considered by Strong et al.[21].

Table 2: Comparison of PHA copolymers and polypropylene properties

Parameter	PHB	UHMW PHB	PHB-co-3HV	PHB-co-4HB	PHB-co-3Hx	PP
Melting temperature (°C)	177	185	145	150	127	176
Glass transition temperature (°C)	4	4	-1	-7	-1	-10
Crystallinity (%)	60	80	56	45	34	50-70
Tensile strength (MPa)	43	400	20	26	21	38
Extension to break (%)	5	35	50	444	400	400

UHMW – Ultra high molecular weight. PP – polypropylene. For monomer abbreviations see Figure 3. From Tsuge 2002[68]. For specifics of copolymer composition refer to the source.

1.3.3 PHB Biodegradation, Bioplastics and Industrial suitability

Naturally occurring PHAs are fully biodegradable to CO_2 and water by naturally occurring microbes[55,59]. This is logical, as PHAs are an energy storage molecule for many bacteria, it follows that they are capable of its degradation and consumption even if it is from an exogenous source just as if it was released from PHA containing cells. Degradation is achieved by bacteria using PHA hydrolyses and depolymerases the effectiveness of which varies by degrading species and PHA or PHA copolymer type as well as environmental conditions[61]. Degradation can range from years in seawater to a few months in sewage[59] with other estimates suggesting 1-6 months even in an ocean environment[69]. Conversely degradation of petrochemical plastics may rely on the leaching away of plasticisers or solely on UV greatly limiting surface area of degradation resulting in extremely long survival of petrochemical plastic waste[56]. Unlike polypropylene and many other bioplastics, PHB sinks in water, which aids in its removal from surface debris and facilitates anaerobic degradation[57,63].

As PHAs can be made from renewable biomass as a feedstock rather than from petrochemicals they can be bioplastics[56,59]. PHB producing methanotrophs can use methane from biogas, where other PHB bioplastic strategies come from the consumption of refined sugars and fatty acids or even farming waste and sewage[59]. Other gas fermenting bacteria such as *Cupriavidus necator* can also use CO_2 and H_2[64]. Conversely PHB produced using methanotrophs fed natural gas or chemical synthesis from petrochemicals is not a true bioplastic[56].

PHAs have been shown to be biocompatible, i.e. have no harmful or toxic effects in living organisms showing it is both safe and has potential for use in medical implants. When inserted into mammals degradation is slow to non-existent[59,63]. The degradation product of PHB, its monomer 3-hyrdroxybutyric acid, is already present in blood at ~1mM levels and is part of the body's normal ketone body metabolism[56]. PHBV has also been tested as an animal feed directly as a method of recycling, showing no toxic effects by

consumption in sheep and pigs but with variable digestibility possibly depending on gut flora[70]. PHB and its copolymers are already used in a medical setting for stents, bone marrow scaffolds, orthopaedic pins, sutures and in advanced tissue engineering among others[59,71]. Their biocompatibility has the added advantage that PHAs do not have to be removed after recovery[63]. Medical rules about endotoxin contamination mean medical PHAs must be carefully sourced with stringent extraction and purification[59]. Using production chassis not subject to endotoxins, for example methanotrophs, may reduce the cost associated with this paper trail and stringency.

Industrial exploitation of PHAs by fermentation was first attempted by Imperial Chemical Industries in the 1980s producing PHBV under the brand name Biopol, the technology and patents were then transferred to Monsanto in 1995 then Metabolix in 2001 but the Biopol trademark lapsed in 2006[58,59,72]. Other applications and industrial producers have been reviewed by Jacquel et al. 2008[65] and are as of yet limited in industrial success. Current companies in the field include Mango Materials, Newlight Technologies, Genecis, TianAn, Biomer and Tepha among many others. There is also potential for applications new to plastics by chemical modification of the exposed side chains allowing fine tuning of properties or addition of new ones[56]. Chemicals encapsulated within PHAs could be released during PHA breakdown providing a delivery system in medicine and agriculture for example fungicides and fertilizers[56]. There is the potential to replace a wide variety of petrochemical polymers if the full range of PHAs and copolymers is harnessed, particularly for containers and films that can be made disposable and quickly biodegradable[59,73], and microbeads that are normally banned in a non-biodegradable form due to their persistence and risk to marine life and ecology.

Although the focus here shall be on production of PHB within methanotrophs, there are many notable bacterial competitors for PHA production including bacterial producers that can achieve 90% PHAs in dry cell weight. Other hosts such as plants have suffered from low yields (<10%)[59]. Research into PHB production and PHAs in general has often centred around *Cupriavidus necator*, which has undergone a series of renaming previously including *Alcaligenes eutrophus* and *Ralstonia eutropha*[74]. *C. necator* has acted as both a process organism capable of gas fermentation of CO_2 and hydrogen, and as a donor of its PHB production genes to other chassis. PHA production has been attempted by using other species with innate PHA genes or recombinant expression of PHA genes in a variety of chassis including *E. coli*[75] and on a variety of carbon sources, as reviewed by Verlinden et al.[59] .

The maximum reported PHB per dry cell weight (%DCW) in *M. parvus* OBBP is “close to 70%” in Asenjo and Suk 1986[46]. Considering this has not been exceeded since, was acquired using a no longer gold standard PHB analysis technique and the next closest results are in the range of 60% this result is treated with some scepticism and 60% is usually referred to as the maximum attained in *M. parvus*[32].

If handled correctly bacterial PHB production from biogas should have little to no ecological impact during either production or degradation, a great improvement over current non-biodegradable fossil fuel technologies[56]. Cost and process optimisation is the current limiting factor in bioplastic success[56,59,76]. To improve efficiency and thus cost, efforts are being put into improving substrate conversion rate, fermentation efficiency, improved recovery and purification, reduced feedstock cost and improving the PHA material properties[59,76]. With the inevitable increasing price in petroleum and other fossil fuels due to dwindling natural reserves the price of petrochemicals and thus petrochemical based plastics will also increase, thus it may not be necessary to reduce costs of bioplastics like PHB to current levels of fossil fuel plastics to be competitive in the future.

1.3.4 PHB Production in Methanotrophs

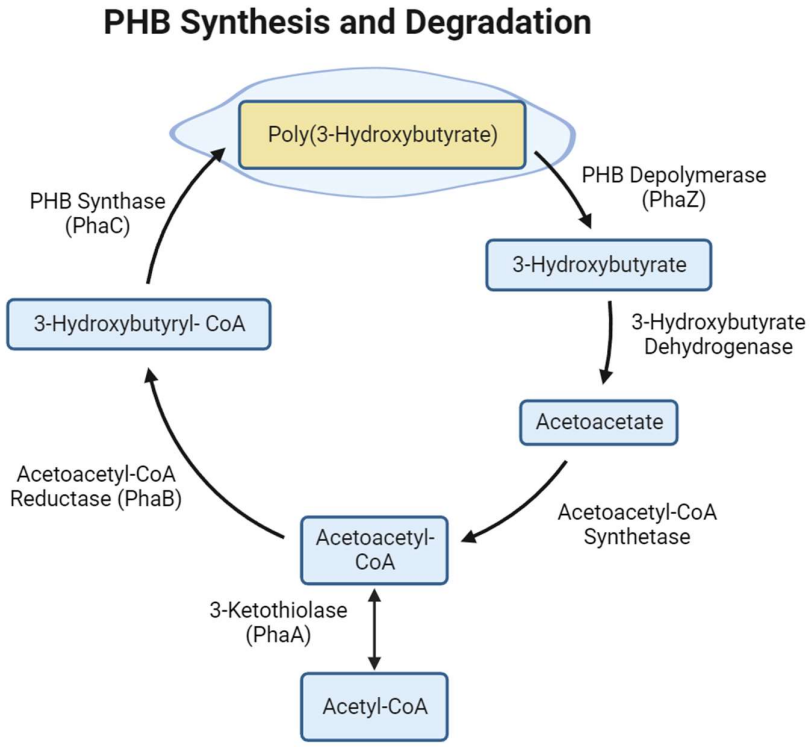


Figure 4: Simplified diagram of the PHB - poly(3-hydroxybutyrate) production and degradation pathway with labelled enzymes for each step. Note this is not stoichiometrically balanced. Data drawn from Liu et al.[66]. Figure created with BioRender.com.

1.3.4.1 The PHB Production Pathway

PHA storage in capable bacteria is induced during stress which can be caused by a variety of factors, the most commonly applied being nitrogen starvation while suitable carbon is still available. Other PHA inducing stress methods are discussed below. PHAs are insoluble and thus it is deposited as inclusion bodies also called granules within the cell[56,59]. These are light transmissive and are clearly visible in transmission electron micrographs (Figure 1C). PHAs are particularly effective as an energy storage molecule as

they are insoluble so have no osmotic effect unlike high concentrations of other molecules. PHA insolubility also prevents seepage from the cell[59]. Granules are coated in protein predominated by phasins, which control the number and size of granules[59,77]. This is a potential target for genetic modification as larger granules are easier to extract from other cell components[62], and might increase the overall yield within the cell. Higher cellular yields are desirable as they result in lower recovery costs as less digestion or extraction needs to take place with lower volumes of solvents or chemicals for the same final purity which also reduces waste water purifying costs[78].

The precise distribution of PHB production capability within the methanotrophs has been conflicted in the literature. However a 2011 study by Pieja et al.[15] found PHB production and the related genes in all tested *Methylocystaceae* α-proteobacterial methanotrophs and none in γ-proteobacterial methanotrophs[15]. PHB has also been found in at least some species of α-proteobacterial *Beijerinckiaceae*[12].

The general assertion that PHB is an energy storage material in *M. parvus* OBBP as in other bacteria has been shown not entirely true. It was found that it was not capable of growth, only maintenance, using PHB reserves in the absence of methane, but when both PHB reserves and methane were present both are utilised simultaneously and PHB conferred a growth advantage. Further experiments indicated PHB was a source of reducing power with the addition of formate (a source of reducing power) delaying PHB consumption where glyoxylate (a source of carbon) did not. PHB can however be used as the only carbon source by at least some non-methanotrophic bacteria[49]. Logically methane fixation through MMO is energy intensive and requires NADH resulting in the greater need for reducing power than carbon alone for methanotrophs. This may have implications on the manipulation of its metabolism to increase PHB production in the future[49]. Accelerated growth through PHB co-utilisation could also be used with repeated stress situations in which PHB production is a selective advantage that might be of use in directed evolution experiments and optimisation.

PHB synthesis occurs over three enzymatic steps through PhaABC and is illustrated in Figure 4. Two acetyl-CoA (Coenzyme A) from the serine cycle are combined into one acetoacetyl-CoA liberating the other CoA. This is reduced with NADH to 3-hydroxybutyryl-CoA and is then polymerised into PHB, one monomer forms an ester bond with the hydroxyl group of the neighbouring monomer, liberating the final CoA. This process is controlled by the availability of CoA; under normal growth conditions PhaA is inhibited by free CoA but under non-carbon nutrient limitation there is an excess of acetyl-CoA which enters PHB synthesis[59]. Production of PHAs in bacteria are stereospecific to the (R)-isomers which is essential for biodegradation and biocompatibility. Complete stereospecificity is not guaranteed from inorganic sources[56,59]. The gene names *phaABC* and *phbABC* are both used in literature to refer to the same genes.

The PHB synthase PhaC comes in two versions often both present in the genome. PhaC1 appears to perform the synthase activity assigned to PhaC while the purpose of PhaC2 is currently not understood. In *C. necator* PhaC2 is present but has been found not expressed in the wild type to detectible levels and by exogenous expression to be catalytically inactive[79,80].

1.3.4.2 PHB Pathway Regulation

The precise structure and control of the PHB synthesis pathway varies between species greatly and relationships are further confused by large amounts of horizontal gene transfer[81,82]. There is more information available about other model species of PHA producers notably *C. necator* H16 but also the very different *Bacillus subtilis*, *Bacillus megaterium, Pseudemonas aeruginosa*, *Pseudomonas oleovorans, Synechocystis sp.* and *Allochromatium vinosum*[81,83]. Parallels drawn between model species and PHB producing α-proteobacterial methanotrophs must be done with care as confirming work is limited and evolutionary distances can be large for example *C. necator* being a β-proteobacterium while *M. parvus* is an α-proteobacterium. The final polymerisation step in PHB synthesis catalysed by PhaC is used to classify PHA producing groups subset into four classes (I-IV) depending on protein structure, substrate specificity and homology. Class I and II PhaC work on short and medium chain length polymers respectively and are made of a single subunit. Classes III and IV consist of two subunits (PhaC and E, or C and R respectively) and both operate on short chain polymers[81,83]. Non-PhaABC pathways for PHA production have also been described in some species[82].

As identified thus far, nothing in the literature explicitly assigns a class to the PhaC of α-proteobacterial methanotroph bacteria. Identifying one would prove indicative in identifying the best model organism to draw inferences from in terms of regulation and expected gene presence. Automated labelling specifies *M. parvus* PhaC as a class I PHA synthase the same type as *C. necator*. Use of the *M. parvus* BRCS2 PhaC amino acid sequence with InterPro[84] for protein functional and familial identification also corroborates this. The utilised hydroxybutyrate monomer in PHB production also fits it within the short chain length monomer groups[83]. This shows convincingly that the PhaC of *M. parvus* and by extension generally α-proteobacterial methanotrophs are in Class I. Class I PhaC forms a homodimer. PHA synthesis genes and regulators are placed in a variety of operon combinations across the evolutionary tree[81]. Generally the *phaB* and *phaC* genes are found together on an operon and surface proteins and transcriptional controllers are expressed separately[81].

There are a number of genes controlling PHB or PHA by transcriptional regulation namely *phaR*, *phaQ* and *phaF* though their presence varies between PHB producers. The literature is somewhat confused with the naming of another regulator AniA with some sources treating it as a different protein to PhaR and others treating them as homologous. They are also ascribed different properties stating *phaR*, *phaQ* and *phaF* are PHA dependent

regulators effecting the expression of synthesis and surface proteins[81] while the PHA independent *aniA* relates to the control of acetyl-CoA flux. tBlastn searching using the *aniA* sequence of *Rhizobium etli* from Encarnación et al. 2002[85] on α-proteobacterial methanotroph genomes (*M. parvus* OBBP and BRCS2) show it to map only to the *phaR* annotated location indicating homology with a 95% coverage and 50% identity. This is significantly higher than the similarity to other PhaR proteins to that location concluding that only one exists in *M. parvus* OBBP. It remains possible that the two activities of phasin control and acetyl-CoA flux are controlled differently in different species by similar proteins, or that the proteins are significantly different. It should be noted that the transcriptional regulator PhaR found in *C. necator* and *M. parvus* is a different protein to the subunit PhaR in Class IV PHA synthesisers[74].

Additional sources of regulation include the activation of PhaC. In *C. necator* the activation and dimerization of PhaC is activated by a control peptide PhaM[83]. PhaM has also been found to be involved in PHB granule location by DNA binding. Its knockout results in one large granule inhibiting passing on of PHB granules to daughter cells post division[86] and its overexpression results in multiple smaller granules[87]. It has also been suggested that PhaC is constitutively expressed awaiting dimerization activation from PhaM[87]. If PhaM is an independent activator of PhaC activity and initiates PHB granule formation it may be effective to induce PHB production detached from nitrogen starvation controls. Other sources have identified PhaM as a phasin[88] which corroborates its described activity with the PhaP phasins described below. *phaM* has not been identified as yet outside of *C. necator* and BLAST searches for this work did not find any matches within *M. parvus*. *phaF* in *Pseudomonas putida*[86,89] (apparently unrelated to *phaF* in *C. necator*) is also expected to serve a similar purpose but BLAST searches also showed no homologues in methanotrophs. Thus any gene performing PhaC activation is yet to be identified in *M. parvus*. Further searching for other regulatory homologs will aid in future annotation and inference from other species.

PHB polymerisation occurs in a polypeptide complex on the surface of the granules of PHB. The PHB is in an amorphous state within the granules but extracted granules become more crystalline[79]. Using data from *C. necator* extracted granules this surface layer was shown to consists of a monolayer of phospholipids and proteins ~4nm thick of which 78% is protein making up ~2.5% of the granule[79]. The presence of the phospholipids *in vivo* was for a long period disputed and was possibly experimental contamination during extraction but nevertheless formed the basis of most granule formation models[79]. Experiments from 2016 with intracellular phospholipid binding tags have shown no phospholipid presence on the granule across α-, β- and γ-proteobacterial examples which appears conclusive that there is no phospholipid layer *in vivo*[77]. Proteomics studies have identified a huge range of associated proteins however it is likely that some identified proteins bound during extraction and are not present *in vivo*. Proteins that have been

identified localised to the granule surface with confidence by immune and fluorescent tagging include PhaC, PhaZ, PhaP, PhaL, PhaM and PhaR[74,81]. Some of these may be species specific among the diversity of PHA producers and specific information on methanotroph species is lacking.

Among the associated proteins a large proportion are phasins (PhaP) which appear to control granule morphology and potentially production. This is achieved by controlling surface to volume ratio and stopping or allowing granules to coalesce by providing a dividing layer of hydrophilic protein. In *C. necator* PhaP loss of function mutation was found to decrease PHB production and result in one large granule, overexpression resulted in many small ones[74]. Phasins have been found in the genomes of all tested PHA producing bacteria including methylotrophs and methanotrophs[24,74,90,91]. The phasin systems exact functionality is not fully understood and its control requires inference from *C. necator* in which the phasin repressor (PhaR) inhibits PhaP production and PhaR mutation causes constitutive PhaP expression. Seven phasins PhaP1-P7, eight including PhaM have been identified in *C. necator* of which only PhaP1 was required for PHB production the function of the remaining phasins are unclear[74,87,88]. PhaP1 and P2 have also been found sufficient to induce PHB production outside of normal stress conditions[88].

PHB depolymerases (PhaZ) are also associated with the granules surface. In vitro PHB depolymerases have been found inactive on non-native PHB granules, native granules that have undergone treatment through chemical means or extraction including freezing and centrifugation[74,79]. This indicates the PHB depolymerase is either required to complex with other proteins or the PHB becomes inaccessible after treatment e.g. structural changes due to crystallisation. A wide range of PhaZ have been identified with seven (depolymerases PhaZa1 to PhaZa5, PhaZd1, and PhaZd2 having been identified in *C. necator*. An additional two proteins PhaZb and PhaZc have been designated oligomer hydrolases by one author but their functional difference is unclear[87].

Recent work in *Methylocystis* sp. B8 has identified two PHB synthase genes, two PHB depolymerases, one acetyl-CoA acetyltransferase, one acetoacetyl-CoA reductase, a polyhydroxyalkanoate synthesis repressor and three phasins[91]. Work in the fragmented *M. parvus* OBBP genome identified two PHB polymerases (*phaCI* and *phaCII*), two PHB depolymerases (*depA* and *depB*), one acetyl-CoA acetyltransferase (*phaA*), one acetoacetyl-CoA reductase (*phaB*), one polyhydroxyalkonate synthesis repressor (*phaR*), and only one phasin mentioned[24]. The fully assembled genome of *M. parvus* BRCS2 noted *phBC1*, *phbC2*, *phbA*, *B* and a phasin along with a *bdhA* D beta hydroxybutyrate dehydrogenase and PHB depolymerase C terminus. This description used the originating authors gene notations. As can be seen annotations differ so future work on the *M. parvus* OBBP genome in this thesis will clarify this, resolving true similarity between annotations. Nevertheless this shows only barebones identification of regulatory genes in the species with two *phaC* variants, one *phaA* and *B* and two depolymerases with at least one phasin across all strains.

The investigation of this work has shown a striking gap in the literature about the specific workings of the PHB system within α-proteobacterial methanotrophs and difficulties in drawing parallels with other PHA/PHB model organism *P. putida*, *C. necator* and others. This raises challenges in engineering the PHB pathway in the future without more intense investigation.

1.3.4.3 PHB Fermentation

Fermentation based PHB and PHA production is usually carried out in a two-stage batch process. The first stage uses complete nutritional media to provide the required biomass, this is replaced in the second stage by nitrogen deficient media to induce PHB production[26,67,92]. Single stage fed batch production have been stated to suffer either from low biomass or low end PHB yield[59]. Typically peak PHB yield is achieved very rapidly with 80% of total achieved in 12h and peak reached in 24h if sufficient oxygen and methane are provided[93]. Stoichiometric calculations for the generalised pathway give a maximum theoretical PHB yields per carbon of 0.67g PHB/g CH_4 by Asenjo and Suk 1986[46] and 0.54g PHB/g CH4 by Yamane 1993[94] the lower value is claimed to be due to an additional NADH regeneration missed in the earlier calculation[94]. More recent work by Rostkowski et al. 2013[95] suggests 0.55 g PHB/g CH_4 by pooling previous experimental data and theoretical work. The authors suggest a molar ratio of O_2 to CH_4 gas of 1.5:1[46] to 2:1[95] which can inform gas concentrations provided in fermentation or compared to headspace usage data. This is corroborated by work directly in *M. parvus* OBBP which found a 1.5:1 ratio of O_2 to CH_4 usage[49]. A patent assigned to Calysta for industrial scale up of *M. capsulatus* Bath for SCP production defines a preferred ratio of 1.8:1 though this potentially included other constituents in mixed culture fermentation and does not account for PHB production[47].

The two-stage process alternation between growth with nitrogen and starvation without it, would imply incompatibility with continuous culturing methods. Previous experiments have used continuous culturing in the biomass growth stage but batch culturing in the PHB productions stage[93]. Mixed-culture and non-sterile cultures of methanotrophs and polytrophs have also been researched for the potential removal of otherwise toxic by-products in the fermentation like methanol and formaldehyde and beneficial nutrient production and potential higher yields which may reduce costs industrially[21].

1.3.5 PHB Induction Methods

As noted previously the dominant stress method for PHB induction in the methanotrophs is nitrogen starvation however other methods have been found worthwhile in methanotrophs or other species. Nitrogen, phosphate, sulphur, oxygen and magnesium starvation have been found effective across the PHA producing bacteria[59,96]. In *M. parvus* OBBP nitrogen, phosphate, potassium, magnesium, and sulphate limitation have been tested. Of these all were found to be capable of inducing PHB production but nitrogen, potassium and phosphate appeared more promising while magnesium and sulphate appeared less so[48].

Wendlandt et. al[97] compared starvation through phosphate, nitrogen (as ammonia) and magnesium starvation in *Methylocystis sp*. GB 25 in mixed culture at the 7L and 70L scale. This found effects from residual phosphorus and magnesium in the medium which were then exhausted. They showed fastest accumulation rates under nitrogen starvation (0.123g/g/h, 51.3% PHB) and roughly equal rates under phosphorous and magnesium starvation but magnesium starvation showed notably lower yield and PHB percentage (~0.06 g/g/h 46.8% and 28.3% PHB).. All stress conditions resulted in almost identical PHB molecular weights of 2.5×10^6 Da.

It appears pertinent that nitrogen fixation, of which many α-proteobacterial methanotrophs are capable, may decrease the impact of nitrogen starvation. Its elimination might lead to increases in PHB yield or the speed of its production. The effect of nitrogen fixation has been shown effective enough that it selects for α-proteobacterial methanotrophs in mixed cultures fed on methane[21,98]. Decoupling PHB production from stress methods by genetic editing would allow PHB production alongside growth or reaching higher PHB concentrations by removing any limiting factors.

1.3.6 PHB Recovery

Extraction methodology is of major concern in the improvement of cost and environmental effects of PHB production[62,65]. Having produced high percentage concentrations within the organism the granules must then be recovered preferably at high purity, with minimal loss and high molecular weight depending on application. Molecular weight is of particular importance due to its corelation with greater mechanical strength of the polymer[58]. Desirable ultra-high molecular weight PHB has been quoted as in the range of $1.3–14\times10^6$ Da[58]. A high starting molecular weight is also important due to thermal degradation during thermoplastic forming that can occur with bioplastics[93]. Molecular weights of 2.5×10^6 Da have been achieved in *Methylocystis* with 51% of dry cell weight being PHB[93,97] but higher sizes and percentages have been achieved in other bacteria though it has been suggested that 0.6×10^6 Da is a common molecular weight for bacterial PHB[58,59]. The polydispersity index (PDI) is also a common measure of polymer quality which measures the variability in polymer length, a PDI of 1 would indicate all chains are the same length but in practice this is always larger. It is calculated as $PDI=M_w/M_n$ where M_w is the weight average molecular weight and M_n is the number average molecular weight. Both PDI and molecular weight will vary with the organism producing it and damage caused by the method of polymer extraction.

The most common lab extraction techniques use solvents like chloroform, methylene chloride or dichloroethane which are both costly and environmentally unpleasant at industrial scale up, often using a reboiler which can reduce molecular weights[55,59,65]. For medical extraction where purity is paramount and cost less essential however solvents are often favoured[59]. Alternatives including ammonia, surfactants, sodium hypochlorite, enzymatic treatments or mechanical disruption with bead mills, sonication, high

pressure homogenisers like a French press or industrial equivalents. More exotic extractions like supercritical fluids, dissolved air flotation and spontaneous cell lysis have also been attempted [59,65]. An even more inventive method of extraction has been demonstrated by Kunasundari et al.[78] who fed PHA containing biomass pellets from *C. necator* to rats and extracted the granules from the faeces without the use of solvents. Although entertaining this could be applied in animal feed SCP situations where faeces could be recovered.

Assessment of PHB yield and purity can be performed by various methods; GC or GC-MS is considered the gold standard and the reference used to calibrate other methods. However GC-MS sample preparation can be intensive especially for monitoring continual growth instead of experimental end points. Alternative options include flow cytometry with Nile red staining[49], NMR spectroscopy, IR spectroscopy or turbidimetry after eliminating non-PHB cellular components with hypochlorite solution. Some authors suggest that calibrated IR spectroscopy, which requires only oven drying, offers a good balance of accuracy and throughput[55,93]. UV-vis spectroscopy of PHB films among other methods has also been used as a measure of quality to assay nitrogen content which should be minimised[93]. Analysis of molecular weight is generally measured by gel permeation chromatography with a refractive index detector which is the standard methodology for polymers[55,93,99]. Other polymer properties of interest include thermal stability assessed by thermogravimetric analysis, crystallinity analysed by X-ray diffraction, melting point, glass transition temperature determined by differential scanning calorimetry and NMR and Fourier transform IR spectroscopy to investigate functional groups and moieties[100].

1.4 Engineering Methanotrophs for Industrial Applications

In addition to SCP and PHB, other products from methanotrophs that have been investigated or suggested include methanol, ethanol, exopolysaccharides, amino acids, carotenoids, biodiesel and lipids[17,101]. Methanotrophs are also of potential use in bioremediation, bioleaching and methane emission control and have already been applied as biofilters removing methane from landfill emissions[17]. To develop these and other engineered applications, use of methanotrophs as chassis organisms would be beneficial. Although PHB production occurs other organisms and has been achieved by recombinant gene expression in model organisms including *E. coli*; methane metabolism for growth has not successfully been transferred into a non-methylotroph chassis organism and a significant breakthrough will be required to achieve this[17,59,75]. This demonstrates a methanotroph chassis is uniquely capable of performing methane metabolising industrial roles.

It has been suggested that the serine cycle of α-proteobacterial methanotrophs is less efficient than the RuMP cycle of γ-proteobacteria supported by lower growth yields on methane in α-proteobacteria. This might suggest that γ-proteobacterial methanotrophs are more suited to industrial

engineering however α-proteobacterial methanotrophs possess other useful qualities including PHB production, a relatively high flux to CoA derivatives, and usage of both CO_2 and CH_4 making them potentially more suited to growth on biogas[17].

Metabolic engineering in methanotrophs is limited so far. Pyruvate in γ-proteobacterial or acetyl-CoA in α-proteobacterial methanotrophs could be readily used for diversion into synthetic pathways to produce useful products and engineering to increase flux to these outputs would improve the strains capabilities. One suggestion is the disabling or removal of the PHB pathway in α-proteobacterial methanotrophs which would otherwise divert the acetyl-CoA to storage providing a usable surplus. Similarly, this could be applied to glycogen in γ-proteobacteria[17].

Formate, lactate and H_2 production that occurs at low O_2 availability can reduce cell yield during fermentation[17,33,35,102]. Elimination of acetate kinase and lactate dehydrogenase could prevent this and elimination of hydrogenase could potentially increase the pool of available NADH. Another contributor to H_2 production include nitrogen fixation in the capable methanotrophs thus elimination of nitrogenases could be beneficial assuming sufficient nitrogen sources are present in the medium[17].

1.5 References

1. Yu H, Susanti D, McGlynn SE, Skennerton CT, Chourey K, Iyer R, et al. Comparative genomics and proteomic analysis of assimilatory sulfate reduction pathways in anaerobic methanotrophic archaea. Front Microbiol. 2018 Dec 1;9:2917.
2. Knittel K, Boetius A. Anaerobic oxidation of methane: Progress with an unknown process. Annu Rev Microbiol. 2009;63:311–34.
3. Kalyuzhnaya MG, Gomez OA, Murrell JC. The Methane-Oxidizing Bacteria (Methanotrophs). In: Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes. Cham: Springer International Publishing; 2019. p. 1–34.
4. Bowman JP, Sly LI, Nichols PD, Hayward AC. Revised taxonomy of the methanotrophs: Description of Methylobacter gen. nov., emendation of Methylococcus, validation of Methylosinus and Methylocystis species, and a proposal that the family Methylococcaceae includes only the group I methanotrophs. Int J Syst Bacteriol. 1993;43(4):735–53.
5. Bowman J. The Methanotrophs — The Families Methylococcaceae and Methylocystaceae. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes. 3rd ed. New York: Springer New York; 2006. p. 266–89.
6. Chistoserdova L, Lidstrom ME. Aerobic Methylotrophic Prokaryotes. In: Rosenberg E, editor. The Prokaryotes. 4th ed. Springer Berlin Heidelberg; 2013. p. 267–85.
7. Lidstrom ME. Aerobic Methylotrophic Prokaryotes. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes. 3rd ed. New York, NY: Springer New York; 2006. p. 618–34.
8. Semrau JD, DiSpirito AA, Yoon S. Methanotrophs and copper. FEMS Microbiol Rev. 2010 Jul;34(4):496–531.
9. Vorobev A V, Baani M, Doronina N V, Brady AL, Liesack W, Dunfield PF, et al. Methyloferula stellata gen. nov., sp. nov., an acidophilic, obligately methanotrophic bacterium that possesses only a soluble methane monooxygenase. Int J Syst Evol Microbiol. 2011;61:2456–63.
10. Tamas I, Smirnova A V, He Z, Dunfield PF. The (d)evolution of methanotrophy in the Beijerinckiaceae—a comparative genomics analysis. ISME J. 2014;8(2):369–82.
11. Marín I, Arahal DR. The Family Beijerinckiaceae. In: The Prokaryotes. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 115–33.
12. Dedysh SN, Khmelenina VN, Suzina NE, Trotsenko YA, Semrau JD, Liesack W, et al. Methylocapsa acidiphila gen. nov., sp. nov., a novel methane-oxidizing and dinitrogen-fixing acidophilic bacterium from Sphagnum bog. Int J Syst Evol Microbiol. 2002;52:251–61.
13. Whittenbury R, Phillips KC, Wilkinson JF. Enrichment, Isolation and

Some Properties of Methane-utilizing Bacteria. J Gen Microbiol. 1970;(61):205–18.

14. De Marco P. Methylotrophy versus heterotrophy: A misconception. Microbiology. 2004;150(6):1606–7.
15. Pieja AJ, Rostkowski KH, Criddle CS. Distribution and Selection of Poly-3-Hydroxybutyrate Production Capacity in Methanotrophic Proteobacteria. Microb Ecol. 2011;62(3):564–73.
16. Pieja AJ, Morse MC, Cal AJ. Methane to bioproducts: the future of the bioeconomy? Curr Opin Chem Biol. 2017;41(1):123–31.
17. Kalyuzhnaya MG, Puri AW, Lidstrom ME. Metabolic engineering in methanotrophic bacteria. Metab Eng. 2015;29:142–52.
18. Kruse T, Ratnadevi CM, Erikstad HA, Birkeland NK. Complete genome sequence analysis of the thermoacidophilic verrucomicrobial methanotroph "candidatus Methylacidiphilum kamchatkense" strain Kam1 and comparison with its closest relatives. BMC Genomics. 2019;20(1):1–15.
19. van Spanning RJM, Guan Q, Melkonian C, Gallant J, Polerecky L, Flot JF, et al. Methanotrophy by a Mycobacterium species that dominates a cave microbial ecosystem. Nat Microbiol. 2022;7(December).
20. Kang CS, Dunfield PF, Semrau JD. The origin of aerobic methanotrophy within the Proteobacteria. FEMS Microbiol Lett. 2019 May 1;366(9):96.
21. Strong P, Laycock B, Mahamud S, Jensen P, Lant P, Tyson G, et al. The Opportunity for High-Performance Biomaterials from Methane. Microorganisms. 2016 Feb 3;4(1):11.
22. Ward N, Larsen Ø, Sakwa J, Bruseth L, Khouri H, Durkin AS, et al. Genomic insights into methanotrophy: The complete genome sequence of Methylococcus capsulatus (Bath). PLoS Biol. 2004;2(10):1616–28.
23. Foster JW, Davis RH. A methane-dependent coccus, with notes on classification and nomenclature of obligate, methane-utilizing bacteria. J Bacteriol. 1966;91(5):1924–31.
24. del Cerro C, García JLJM, Rojas A, Tortajada M, Ramón D, Galán B, et al. Genome sequence of the methanotrophic poly-β-hydroxybutyrate producer Methylocystis parvus OBBP. J Bacteriol. 2012 Oct 15;194(20):5709–10.
25. Prior SD, Dalton H. Acetylene as a suicide substrate and active site probe for methane monooxygenase from Methylococcus capsulatus (Bath) (Inhibitor of methane-oxidising activity). Vol. 29, FEMS Microbiology Letters. 1985.
26. Zhang T, Zhou J, Wang X, Zhang Y. Poly-β-hydroxybutyrate Production by Methylosinus trichosporium OB3b at Different Gas-phase Conditions. Iran J Biotech. 2019;17(1):1866.
27. Dworkin M, Foster JW. STUDIES ON PSEUDOMONAS METHANICA (SÖHNGEN) NOV. COMB. J Bacteriol. 1956 Nov;72(5):646–59.

28. Hou CT, Laskin AI, Patel RN. Growth and polysaccharide production by Methylocystis parvus OBBP on methanol. Appl Environ Microbiol. 1979;37(5):800–4.

29. Romanovskaya VA, Malashenko YR, Bogachenko VN. [Refinement of the diagnosis of the genera and species of methane-using bacteria] - [Article in Russian]. MIKROBIOLOGIYA. 1978;47(1):120–30.

30. LPSN. LPSN - methylocystis-parvus [Internet]. [cited 2023 Feb 27]. Available from: https://lpsn.dsmz.de/species/methylocystis-parvus

31. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. Int J Syst Evol Microbiol. 2020 Nov 1;70(11):5607–12.

32. Rumah BL, Stead CE, Claxton Stevens BH, Minton NP, Grosse-Honebrink A, Zhang Y. Isolation and characterisation of Methylocystis spp. for poly-3-hydroxybutyrate production using waste methane feedstocks. AMB Express. 2021 Dec 6;11(6):1–13.

33. Vecherskaya M, Dijkema C, Saad HR, Stams AJM. Microaerobic and anaerobic metabolism of a Methylocystis parvus strain isolated from a denitrifying bioreactor. Environ Microbiol Rep. 2009;1(5):442–9.

34. Best DJ, Higgins IJ. Methane-oxidizing activity and membrane morphology in a methanol-grown obligate methanotroph, Methylosinus trichosporium OB3b. J Gen Microbiol. 1981;125(1):73–84.

35. Kalyuzhnaya MG. Methane Biocatalysis: Selecting the Right Microbe. Biotechnology for Biofuel Production and Optimization. Elsevier B.V.; 2016. 353–383 p.

36. Yang S, Matsen JB, Konopka M, Green-Saxena A, Clubb J, Sadilek M, et al. Global molecular analyses of methane metabolism in methanotrophic alphaproteobacterium, methylosinus trichosporium OB3b. Part II. metabolomics and 13C-labeling study. Front Microbiol. 2013;4(APR):1–13.

37. Anthony C. The Biochemistry of Methylotrophs. London: Academic Press Inc.; 1982. 342–343 p.

38. Chistoserdova L, Kalyuzhnaya MG, Lidstrom ME. The Expanding World of Methylotrophic Metabolism. Annu Rev Microbiol. 2009;63(1):477–99.

39. Kits KD, Klotz MG, Stein LY. Methane oxidation coupled to nitrate reduction under hypoxia by the Gammaproteobacterium Methylomonas denitrificans, sp. nov. type strain FJG1. Environ Microbiol. 2015;17(9):3219–32.

40. Tavormina PL, Orphan VJ, Kalyuzhnaya MG, Jetten MSM, Klotz MG. A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. Environ Microbiol Rep. 2011;3(1):91–100.

41. Osborne CD, Haritos VS. Horizontal gene transfer of three co-inherited

methane monooxygenase systems gave rise to methanotrophy in the Proteobacteria. Mol Phylogenet Evol. 2018;129:171–81.

42. Oshkin IY, Miroshnikov KK, Grouzdev DS, Dedysh SN. Pan-genome-based analysis as a framework for demarcating two closely related methanotroph genera methylocystis and methylosinus. Microorganisms. 2020;8(5).

43. Ricke P, Erkel C, Kube M, Reinhardt R, Liesack W, Yimga MT, et al. Comparative Analysis of the Conventional and Novel pmo (Particulate Methane Monooxygenase) Operons from Methylocystis Strain SC2 In addition to the conventional pmoA gene (pmoA1) encoding the active site polypeptide of particulate methane monooxygenase, a. Appl Environ Microbiol. 2004;70(5):3055–63.

44. Baani M, Liesack W. Two isozymes of particulate methane monooxygenase with different methane oxidation kinetics are found in Methylocystis sp. strain SC2. Proc Natl Acad Sci U S A. 2008 Jul 22;105(29):10203–8.

45. Dedysh SN, Ricke P, Liesack W. NifH and NifD phylogenies: An evolutionary basis for understanding nitrogen fixation capabilities of methanotrophic bacteria. Microbiology. 2004;150(5):1301–13.

46. Asenjo JA, Suk JS. Microbial Conversion of Methane into poly-β-hydroxybutyrate (PHB): Growth and intracellular product accumulation in a type II methanotroph. J Ferment Technol. 1986 Jan 1;64(4):271–8.

47. Eriksen H, Strand K, Jørgensen L. US7579163B2 - Method of fermentation. United States: United States Patent and Trademark Office; US7579163B2, 2009.

48. Sundstrom ER, Criddle CS. Optimization of methanotrophic growth and production of poly(3-hydroxybutyrate) in a high-throughput microbioreactor system. Appl Environ Microbiol. 2015;81(14):4767–73.

49. Pieja AJ, Sundstrom ER, Criddle CS. Poly-3-hydroxybutyrate metabolism in the type II Methanotroph Methylocystis parvus OBBP. Appl Environ Microbiol. 2011;77(17):6012–9.

50. Myung J, Flanagan JCA, Waymouth RM, Criddle CS. Methane or methanol-oxidation dependent synthesis of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) by obligate type II methanotrophs. Process Biochem. 2016;51(5):561–7.

51. Wendlandt K -D, Jechorek M, Brühl E. The influence of pressure on the growth of methanotrophic bacteria. Acta Biotechnol. 1993;13(2):111–5.

52. Haynes WM. CRC Handbook of Chemistry and Physics. 95th ed. Lide DR, Bruno TJ, Haynes WM, editors. Florida: CRC Press; 2014. 3–488 p.

53. Sander R. Compilation of Henry's law constants (version 4.0) for water as solvent. Atmos Chem Phys. 2015 Apr 30;15(8):4399–981.

54. Henry W. III. Experiments on the quantity of gases absorbed by water, at different temperatures, and under different pressures. Philos Trans R

Soc London. 1803 Dec 31;93(3):29–274.

55. Braunegg G, Lefebvre G, Genser KF. Polyhydroxyalkanoates, biopolyesters from renewable resources: Physiological and engineering aspects. Vol. 65, Journal of Biotechnology. 1998.

56. Zinn M, Hany R. Tailored material properties of polyhydroxyalkanoates through biosynthesis and chemical modification. Adv Eng Mater. 2005;7(5):408–11.

57. Bergmann M, Gutow L, Klages M. Marine anthropogenic litter. Marine Anthropogenic Litter. 2015. 1–447 p.

58. Iwata T. Strong fibers and films of microbial polyesters. Macromol Biosci. 2005;5(8):689–701.

59. Verlinden RAJ, Hill DJ, Kenward MA, Williams CD, Radecka I. Bacterial synthesis of biodegradable polyhydroxyalkanoates. J Appl Microbiol. 2007;102(6):1437–49.

60. Lemoigne M. Produits de dehydration et de polymerisation de l'acide ß-oxobutyrique [Dehydration and polymerization product of β-oxy butyric acid] (French). Bull Soc Chim Biol. 1926;(8):770–782.

61. Lenz RW, Marchessault RH. Reviews Bacterial Polyesters: Biosynthesis, Biodegradable Plastics and Biotechnology.

62. Van Wegen RJ, Ling Y, Middelberg APJ. Industrial production of polyhydroxyalkanoates using Escherichia coli: An economic analysis. Chem Eng Res Des. 1998;76(3):417–26.

63. Kariduraganavar MY, Kittur AA, Kamble RR. Polymer Synthesis and Processing. 1st ed. Natural and Synthetic Biomedical Polymers. Elsevier Inc.; 2014. 1–31 p.

64. Khosravi-Darani K, Mokhtari ZB, Amai T, Tanaka K. Microbial production of poly(hydroxybutyrate) from C1 carbon sources. Appl Microbiol Biotechnol. 2013;97(4):1407–24.

65. Jacquel N, Lo C-W, Wei Y-H, Wu H-S, Wang SS. Isolation and purification of bacterial poly(3-hydroxyalkanoates). Biochem Eng J. 2008;39:15–27.

66. Liu L-Y, Xie G-J, Xing D-F, Liu B-F, Ding J, Ren N-Q. Biological conversion of methane to polyhydroxyalkanoates: Current advances, challenges, and perspectives. Environ Sci Ecotechnology. 2020;2:100029.

67. López JC, Arnáiz E, Merchán L, Lebrero R, Muñoz R. Biogas-based polyhydroxyalkanoates production by Methylocystis hirsuta: A step further in anaerobic digestion biorefineries. Chem Eng J. 2018;(333):529–36.

68. Tsuge T. Metabolic improvements and use of inexpensive carbon sources in microbial production of polyhydroxyalkanoates. J Biosci Bioeng. 2002;94(6):579–84.

69. Rosenboom J, Langer R, Traverso G. Bioplastics for a circular economy. Nat Rev Mater. 2022 Jan 20;7(2):117–37.

70. Forni D, Bee G, Kreuzer M, Wenk C. Novel biodegradable plastics in sheep nutrition 2. Effects of NaOH pretreatment of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) on in vivo digestibility and on in vitro disappearance (Rusitec). J Anim Physiol Anim Nutr (Berl). 1999;81(1):41–50.

71. Marcano A, Haidar NB, Marais SS, Valleton J-M, Duncan AC. Designing Biodegradable PHA-Based 3D Scaffolds with Antibiofilm Properties for Wound Dressings: Optimization of the Microstructure/ Nanostructure. 2017;

72. Metabolix. BIOPOL - Trademark Details. Lisa M. Tittemore; 1375336, 1985.

73. Noda I. Films comprising biodegradable PHA copolymers. United States; US6174990B1, 2001. p. 1–19.

74. Jendrossek D. Polyhydroxyalkanoate granules are complex subcellular organelles (carbonosomes). Vol. 191, Journal of Bacteriology. 2009. p. 3195–202.

75. Balakrishna Pillai A, Jaya Kumar A, Kumarapillai H. Enhanced production of poly(3-hydroxybutyrate) in recombinant Escherichia coli and EDTA–microwave-assisted cell lysis for polymer recovery. AMB Express. 2018;8(1):142.

76. Favaro L, Basaglia M, Casella S. Improving polyhydroxyalkanoate production from inexpensive carbon sources by genetic approaches: a review. Biofuels, Bioprod Biorefining. 2019 Jan 1;13:208–27.

77. Bresan S, Sznajder A, Hauf W, Forchhammer K, Pfeiffer D, Jendrossek D. Polyhydroxyalkanoate (PHA) granules have no phospholipids. Sci Rep. 2016 May 25;6(1):1–13.

78. Kunasundari B, Murugaiyah V, Kaur G, Maurer FHJ, Sudesh K. Revisiting the Single Cell Protein Application of Cupriavidus necator H16 and Recovering Bioplastic Granules Simultaneously. PLoS One. 2013;8(10):78528.

79. Jendrossek D, Pfeiffer D. New insights in the formation of polyhydroxyalkanoate granules (carbonosomes) and novel functions of poly(3-hydroxybutyrate). Environ Microbiol. 2014 Aug 1;16(8):2357–73.

80. Pfeiffer D, Jendrossek D. Localization of poly(3-Hydroxybutyrate) (PHB) granule-associated proteins during PHB granule formation and identification of two new phasins, phap6 and phap7, in Ralstonia eutropha H16. J Bacteriol. 2012 Nov;194(21):5909–21.

81. Kutralam-Muniasamy G, Corona-Hernandez J, Narayanasamy RK, Marsch R, Pérez-Guevara F. Phylogenetic diversification and developmental implications of poly-(R)-3-hydroxyalkanoate gene cluster assembly in prokaryotes. FEMS Microbiol Lett. 2017;364(13):1–9.

82. Kalia VC, Lal S, Cheema S. Insight in to the phylogeny of polyhydroxyalkanoate biosynthesis: Horizontal gene transfer. Gene. 2007;389(1):19–26.

83. Mezzolla V, D'Urso OF, Poltronieri P. Role of PhaC type I and type II enzymes during PHA biosynthesis. Polymers (Basel). 2018;10(8).

84. Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 2021 Jan 8;49(D1):D344–54.

85. Encarnación S, Del Carmen Vargas M, Dunn MF, Dávalos A, Mendoza G, Mora Y, et al. Ania regulates reserve polymer accumulation and global protein expression in Rhizobium etli. J Bacteriol. 2002;184(8):2287–95.

86. Bresan S, Jendrossek D. New insights into PhaM-PhaC-mediated localization of polyhydroxybutyrate granules in Ralstonia eutropha H16. Appl Environ Microbiol. 2017 Jun 1;83(12).

87. Pfeiffer D, Jendrossek D. PhaM Is the Physiological Activator of Poly(3-Hydroxybutyrate) (PHB) Synthase (PhaC1) in Ralstonia eutropha. Appl Environ Microbiol. 2014;80(2):555–63.

88. Tang R, Peng X, Weng C, Han Y. The Overexpression of Phasin and Regulator Genes Promoting the Synthesis of Polyhydroxybutyrate in Cupriavidus necator H16 under Nonstress Conditions. Appl Environ Microbiol. 2022;88(2).

89. Galán B, Dinjaski N, Maestro B, De Eugenio LI, Escapa IF, Sanz JM, et al. Nucleoid-associated PhaF phasin drives intracellular location and segregation of polyhydroxyalkanoate granules in Pseudomonas putida KT2442. Mol Microbiol. 2011;79(2):402–18.

90. Wieczorek R, Steinbbchel A, Schmidt B. Occurrence of polyhydroxyalkanoic acid granule-associated proteins related to the Alcaligenes eutrophus H16 GA24 protein in other bacteria. Vol. 135, FEMS Microbiology Letters. 1996.

91. Jung GY, Rhee SK, Han YS, Kim SJ. Genomic and physiological properties of a facultative methane-oxidizing bacterial strain of Methylocystis sp. From a wetland. Microorganisms. 2020;8(11):1–20.

92. Koller M, Braunegg G. Potential and prospects of continuous polyhydroxyalkanoate (PHA) production. Bioengineering. 2015;2(2):94–121.

93. Wendlandt KD, Geyer W, Mirschel G, Al-Haj Hemidi F. Possibilities for controlling a PHB accumulation process using various analytical methods. J Biotechnol. 2005;117(1):119–29.

94. Yamane T. Yield of poly-D(-)-3-hydroxybutyrate from various carbon sources: A theoretical study. Biotechnol Bioeng. 1993 Jan;41(1):165–70.

95. Rostkowski KH, Criddle CS, Lepech MD. Cradle-to-gate life cycle assessment for a cradle-to-cradle cycle: Biogas-to-bioplastic (and back).

Environ Sci Technol. 2012;46(18):9822–9.

96. Cantera S, Frutos OD, López JC, Lebrero R, Torre RM. Technologies for the Bio-conversion of GHGs into High Added Value Products: Current State and Future Prospects. In: Fernandez RA, Zubelzu S, Martinez R, editors. Carbon Footprint and the Industrial Life Cycle From Urban Planning to Recycling. Springer International Publishing; 2017. p. 359–88.

97. Wendlandt K-D, Jechorek M, Helm J, Stottmeister U. Producing poly-3-hydroxybutyrate with a high molecular mass from methane. J Biotechnol. 2001 Mar 30;86(2):127–33.

98. Pfluger AR, Wu WM, Pieja AJ, Wan J, Rostkowski KH, Criddle CS. Selection of Type I and Type II methanotrophic proteobacteria in a fluidized bed reactor under non-sterile conditions. Bioresour Technol. 2011;102(21):9919–26.

99. Arikawa H, Sato S, Fujiki T, Matsumoto K. Simple and rapid method for isolation and quantitation of polyhydroxyalkanoate by SDS-sonication treatment. J Biosci Bioeng. 2017 Aug;124(2):250–4.

100. Pradhan S, Dikshit PK, Moholkar VS. Production, ultrasonic extraction, and characterization of poly (3-hydroxybutyrate) (PHB) using Bacillus megaterium and Cupriavidus necator. Polym Adv Technol. 2018;29(8):2392–400.

101. Salehi R, Chaiprapat S. Conversion of biogas from anaerobic digestion to single cell protein and bio-methanol: mechanism, microorganisms and key factors - A review. Environ Eng Res. 2021 Jun 21;27(4):210109–0.

102. Vecherskaya M, Dijkema C, Stams AJM. Intracellular PHB conversion in a Type II methanotroph studied by 13C NMR. J Ind Microbiol Biotechnol. 2001;26(1–2):15–21.

Chapter 2: General Methods, Materials and Growth Dynamics

2.1 Culturing

2.1.1 Stains Used

Methylocystis parvus OBBP was acquired from NCIMB (11129) and was grown at 30°C. *M. parvus* in liquid media generally took 3-4 days to reach OD ~1 and gas exhaustion under the standard conditions without regassing. On solid media growth took 2-3 weeks to show visible colonies. *M. parvus* BRCS2 was provided by Bashir Rumah and treated and grew similarly.

Work by Kevbrina et al.[1] has indicated from extrapolation that the optimal temperature of *M. parvus* OBBP was 34°C, above the 30°C normally quoted and used. This alternative was not explored in this work. The same data also indicated the minimum growth temperature was 10°C allowing cultures to be held at 4°C refrigeration temperatures. Standard deviation of both figures was quoted as 13%.

Methylococcus capsulatus (Bath) was acquired from ATCC (33009) and was grown at 45°C. Its growth times were broadly similar to *M. parvus* OBBP.

General molecular biology was carried out using NEB 5-α Competent *E. coli* (NEB, C2987). *E. coli* S17-1 λpir[2,3], provided by Bashir Rumah and used for conjugation into methanotrophs. Both were grown at 37°C on LB liquid or solid media.

2.1.2 Media

Culture media unless otherwise noted was NMS (Nitrate Mineral Salts). NMS media was originally defined by Whittenbury et al. in 1970[4]. This differs from more modern implementations of NMS which may include a source of molybdate, copper and a vitamin solution and have a higher phosphate concentration and lower $CaCl_2$. A common source is methanotroph.org[5]. This is closer but not identical to the recipe of Dalton and Whittenbury 1976[6]. It is my suspicion that this difference to the original media is not properly elucidated in some publications that use a Whittenbury et al. 1970[4] citation with no additional information.

To ensure clarity, the recipe for NMS media used here is almost the same as that on methanotroph.org[5] differing in the in the addition of FeEDTA after the autoclave step rather than before and is as follows: 1L reverse osmosis (RO) water, 1g $MgSO_4.7H_2O$, 0.2g $CaCl_2.2H_2O$ 1g KNO_3, 0.5ml $Na_2MoO_4.4H_2O$ 0.1% w/v in dH_20 and 1ml trace element stock solution. This solution is autoclaved and cooled to below 60°C, then the following filter sterilised ingredients are added: 100µl FeEDTA 3.8% w/v in dH_2O, 10ml phosphate

buffer stock solution, 10ml vitamin stock solution and 100µl 100mM $CuSO_4$ in dH_2o. The trace element stock solution: 1L RO water, 0.5g $FeSO_4.7H_2O$, 0.4g $ZnSO_4.7H_2O$, 0.02g $MnCl_2.7H_2O$, 0.05g $CoCl_2.6H_2O$, 0.01g $NiCl_2.6H_2O$, 0.015g H_3BO_3 (boric acid), 0.25g EDTA then filter sterilised. The vitamin stock solution at 10x concentration: 1L RO water, 0.02g Biotin, 0.02g Folic Acid, 0.05g Thiamine HCL, 0.05g Capantothenate, 0.001g Vitamin B12, 0.05g Riboflavin, 0.05g Nicotinamide then filter sterilised. 10x vitamin stock solution was diluted 10x with RO water and filter sterilised again to make the final stock. Phosphate buffer stock solution: 1L RO water, 26g KH_2PO_4 and 62g $Na_2HPO_4.7H_2O$ then filter sterilised. Vitamin and trace metal stock solutions were stored at 4°C in the dark. Phosphate buffer stock solution was stored at room temperature to avoid crystallisation.

NMS media appeared stable after autoclaving indefinitely however a precipitate formed upon addition of the post autoclaving additives after a few weeks. Growth in precipitated media appeared successful but for experiments the NMS media was given a shelf life of 1 week after the addition of post-autoclave additives for reproducibility. If the media is not cooled sufficiently before the addition of post-autoclave additives an immediate white precipitate is formed with which growth was also possible. Precipitates have been observed and their avoidance discussed previously[5].

NMS agar plates were prepared by following the above recipe with the addition of 15g/L of agar before autoclaving. This was cooled to 55°C in a water bath before the addition of post autoclave ingredients and pouring.

NMS is a minimal medium with no carbon source and a phosphate buffer. Nitrate is provided in the form of KNO_3. This is discussed more fully in section 1.2.4. An alternative media nfNMS (Nitrate Free NMS) excludes the KNO_3.

LB Broth (Miller) (Sigma-Aldrich) made to instructions and utilised in 5ml aliquots in 50ml falcon tubes incubated at 200rpm and 30°C, or with 15g/L Agar on plates at 30°C both for 48 hours to contaminant check methanotroph cultures at the end of experiments, before freezer storage and before the commencement of major work. Methanotrophs will not grow in these conditions, but many polytrophs will. Due to the prolonged growth times and contaminant risks of regassing methanotroph cultures contaminant checking was necessary to ensure experimental reliability.

Small volume filtering was carried out using 0.22µm PES syringe filters (Sartorious, Minisart Highflow 16532-K). Volumes of larger than 100ml were vacuum sterilised using a 0.2µm Nalgene vacuum filtration system (Thermo Fischer, 566-0020).

2.1.3 Culturing Methods - Methanotrophs

When grown in liquid, methanotrophs were inoculated into NMS media (unless otherwise stated) in serum bottles. This was closed with a butyl (rubberBV, 7395) or bromobutyl (DWK Life Sciences, 224100-331) rubber stopper and closed with an aluminium crimp seal. Methane gas (BOC, Methane N4.5 Research Grade 157682) was added to the bottle measured

using a 50ml syringe then passed into the bottle through a 0.22µm syringe filter and a hypodermic needle through the stopper.

A selection of serum bottles sizes were used: 570, 250, 160, 130 and 60ml. A set volume of media and volume of methane gas was used depending on bottle size to maintain a 20% overpressure and a 25% methane headspace with an approximately 0.15:1 ratio of media to headspace. Specifics are defined in Table 3. For example, in a 160ml bottle 24ml of media were used with 34ml of CH_4. This provides an O_2:CH_4 volume ratio of 0.84:1. This does not match the preferred ratio of 1.5:1 to 2:1 discussed previously (section 1.3.4.3) and would result in oxygen limitation. Pure oxygen could not be used due to health and safety limitations although it has been used in previous literature[7,8]. Experimental conditions were limited by requirements to avoid the flammability limits of methane in air. These are defined by British Standard as a lower flammability limit of 4.4% and upper flammability limit of 17.0% by volume of methane with air[9]. Bottles were then shaken at 200rpm. Samples were taken by withdrawing culture liquid through the stoper with a hypodermic needle and syringe without drawing gas from the headspace.

Table 3: Methane and media fill amounts of each bottle size used and resulting calculations of O_2 levels. Oxygen taken to be 20.9476% of Air by volume[10,11]. Values presented to 2dp where rounded.

Bottle Size (ml)	CH_4 (ml)	Media (ml)	Media %	Air (ml)	O_2 (ml)	CH_4/O2	O_2 ml / Media ml
570	118	100	17.54	470	98.45	0.83	0.98
250	54	35	14.00	215	45.03	0.83	1.29
160	34	24	15.00	136	28.49	0.84	1.19
130	27.5	20	15.38	110	23.04	0.84	1.15
60	13	8	13.33	52	10.89	0.84	1.36

When required the gas in the headspace was replaced (re-gassed) by passing compressed air at 2 bar gauge pressure in through a 0.22µm syringe filter and a hypodermic needle through the stopper and allowing headspace gas to be forced out another hypodermic needle. This was sustained for 2 minutes then compressed air was removed and pressure was allowed to equalise with the atmosphere. Methane was then added as required as when a new bottle is set up.

Multiwell liquid plate growth was performed with 1 ml of media in 48well plates which in an EnzyScreen gas tight plate box (CR1601) and sealed. Methane was introduced at 1.4 bar gauge pressure through a 0.22µm syringe filter and rubber tube through one port allowing displaced air to leave through the other port for 10 seconds before allowing pressure to equalise and closing both ports. This was shaken at 200 rpm.

Methanotrophs on solid media were placed into an airtight box (Sistema Klip 5L it 1850 or LocknLock 5.5L HPL836), Oxoid Anaerojar or Anaerocult jar. Methane was introduced through a 0.22µm syringe filter and Marprene

hose at 1.4 bar gauge pressure for 10 seconds into a partially closed container which was then sealed. Boxes were fully opened to allow an exchange of air and re-gassed in the same way every 3-5 days. In some cases, Vaseline was used around the seals to stop gas escape.

Fungal contamination on plates was a common occurrence and to mitigate this silica beads (SLS, CHE3196) in a petri dish placed in the boxes were used to reduce condensation (Suggestion in person – Colin Murrel) which was found to be partially successful. Following usage by Auman et al.[12], antifungals Nystatin (20mg/ml in DMSO stock, final 10µg/ml) and Cycloheximide (20mg/ml in dH_20 stock, final 20µg/ml) were tested adding to plates. Growth of *M. parvus* OBBP was tested in liquid media supplemented with he antifungals (Figure 5) and not found to not be affected at this concentration, and the antifungals did not appear to inhibit growth on solid plates. However in combination the antifungals were also not found to be successful in inhibiting fungal contamination. The antifungals were also not found effective in curing fugally infected liquid cultures. At least two morphologies of fungus were observed but these were not investigated. Duplicate plates were generally made and incubated separately to allow for fungal losses.

Rubber seals were sterilised with alcohol wipes or 70% industrially methylated spirit solution before penetration with a needle for gassing or taking samples.

It was noted that taking repeat readings from singular culture bottles risked effecting the culture being observed by a) venting some headspace gas b) changing the balance of media to headspace. Thus sampling was limited generally to once a day.

The limiting factor in methanotroph growth rates are considered to be gas mass transfer and gas availability, most likely of methane. At higher ODs however in our culturing setup oxygen limitation becomes a limiting factor. Thus growth rates become linear as oxygen is exhausted between regassing stages. Thus optical density growth charts in this work should generally be observed with the knowledge that exponential growth will not be observed after early growth and will instead be linear.

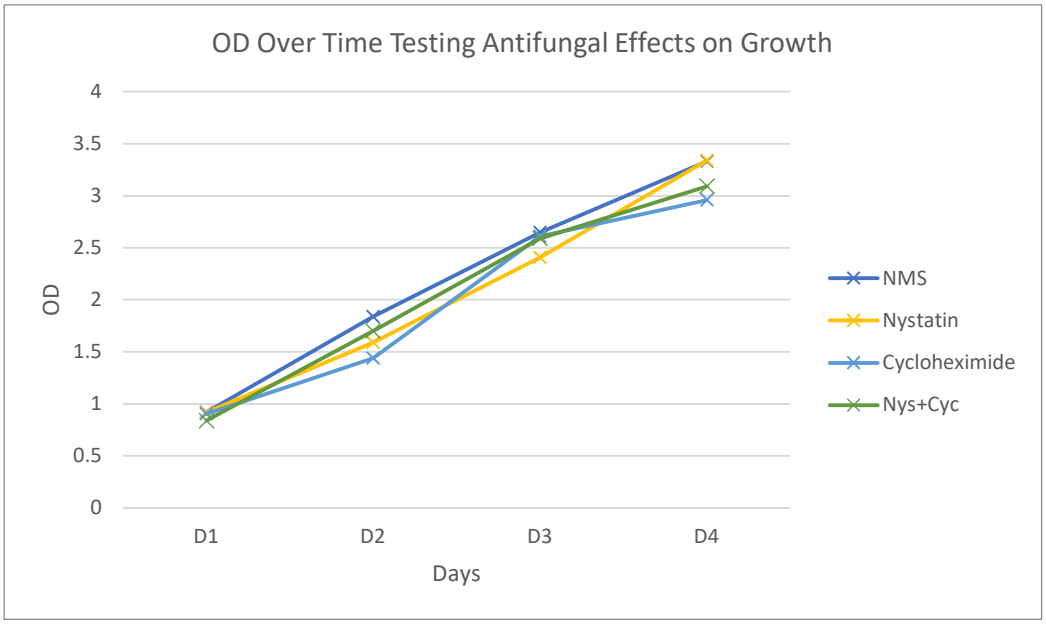


Figure 5: Effect on growth of *M. parvus* OBBP by the two antifungals suggested by Auman et al.[12] were tested individually and in combination added to NMS. This was compared to a control of only NMS. Antifungals were used as follows: Nystatin 10µg/ml and Cycloheximide 20µg/ml. Data points are from single replicates in 250ml bottles with 35ml of media grown over 4 days and regassed and OD taken every day.

2.1.4 Strain Storage

Bacterial strains were stored in 25 bead tubes using the Microbank 2D tube system (Pro-Lab Diagnostics). Due to slow growth on plates this was achieved by removing the storage buffer, applying 1ml of liquid culture to beads, then liquid removal leaving coated beads and stored at -80°C. Single beads were placed in culture media for growth. There are potential concerns about long term revivability of strains using this system. A 0.5ml 50% glycerol in dH_2O solution mixed with 0.5ml culture media was also tested and found effective but no statement of preference could be made without systematic testing. 10% DMSO storage is suggested on methanotroph.org[5] but was not tested.

2.1.5 Centrifugation

Microcentrifugation was carried out on a Micro Star 17R (VWR, 521-1647) or Eppendorf 5424R (5404). Generally at 8,000rpm/6200*g* for 3 minutes for pelleting bacteria to continue growth and 13,000rpm/16,200*g* for 1-5 minutes in other cases. Centrifugation of 15ml and 50ml scale was carried out in an Allegra x-30R (Beckman Coulter, B06321) with C1015 or C0650 conical rotor at the start of the project but due to streaking this was changed. For the remainder of the project an Allegra X-22 (Beckman Coulter, 392185) and SX4250 swing bucket rotor at 4500rpm/3398*g* for 10 minutes or Sorvall Lynx 6000 (Thermo Scientific, 75006590) and BIOFlex HC Swinging-Bucket Rotor (Thermo Scientific, 75003000) at 5,500 rpm/7068*g* for 10 minutes was used. In all these cases maximum ramp speeds were used unless otherwise noted.

Centrifugation of methanotrophs found to behave differently depending on strain. *M. parvus* OBBP was found to streak on the side of falcon tubes if spun in a conical centrifuge rotor (Figure 6). This streaking was time consuming and challenging to resuspend and resulted in sample loss and reduced reliability. A swing bucket rotor was found to avoid this and produce a stable pellet.

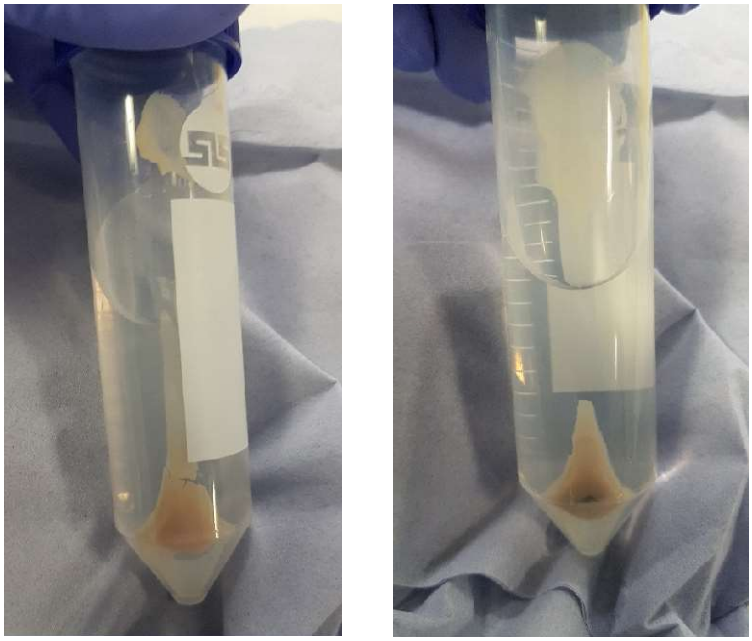


Figure 6: Two images of *M. parvus* OBBP showing streaking down the side of a polypropylene 50ml falcon tube resulting from centrifugation in a conical centrifuge rotor. This did not occur in other bacteria tested in Chapter 5.

2.2 Molecular Biology

2.2.1 General Molecular Biology Techniques

Plasmid extraction (miniprep) from bacteria was carried out using a QIAprep Spin Miniprep Kit (QIAGEN, 27106) using the centrifugation method.

gDNA was extracted with one of two kits. GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich, NA2110) using guidance for minimally sheared DNA, RNAse A treatment and the gram -ve protocol unless otherwise stated. Alternatively Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research, D6005) was used, despite the name this produced gDNA. The bead beating stage was carried out on a Precellys 24 Tissue Homogenizer (Bertin, P000669-PR240-A) using the following setting: 6400rpm for 3x45secs cycles with 60sec breaks between cycles.

Sanger sequencing of PCR products and plasmids was performed by Source Biosciences, Europhins Genomics or Genewiz.

Agarose gel electrophoresis was performed with 0.7%, 1% or 1.5% agarose in 1x TAE (Fischer, BP1332-4) with RO Water, visualised with SYBR Safe

(Invitrogen, S33102). GeneRuler 1kb Plus DNA ladder (Thermo Scientific, SM1331) and TriTrack DNA Loading Dye (Thermo Scientific, R1161) was used in all cases. Where gel extraction was to follow a 0.7% agarose gel was used. Gels were cast and run using a Biometra Compact M and L system (Analytik Jena, 846-025-399 and 846-025-399) and PowerPac Basic Power Supply (Bio-Rad, 1645050). Gels were visualised and imaged on a Gel Doc XR+ (Bio-Rad, 1708195).

DNA quantification and assessment was generally carried out on a SimpliNano spectrophotometer (GE, 29061712). But a Qubit 3.0 (Invitrogen, Q33216) using the broad range (Q32853) or high specificity (Q32854) kits as appropriate for high sensitivity applications.

Routine PCR was carried out at 25µl volume using Q5 High-Fidelity 2X Master Mix (NEB, M0492) with 400nM each of forward and reverse primers. The standard protocol with a gDNA or plasmid template was:

- Initial Denaturation - 98°C 30 secs
- Denaturation - 98°C 10 secs
- Annealing - T_a 30 secs
- Extension - 72°C T_e
- Final Extension - 72°C 2 mins

Where T_a and T_m are specific to the primer pair. T_m was calculated using the NEB Tm Calculator[13]. T_e was calculated as 30secs/kb +30 seconds using the expected length of product. PCR was performed on a Biometra Professional trio (Analytik Jena, 846-2-070-723) or Mastercycler Nexus x2 and x2e (Eppendorf, 6337000043 and 6339000040).

For PCRs directly from live cells initial denaturation was extended to 5 mins to allow cell lysis. From liquid cultures 0.5µl of media was added to the PCR mix. From solid media single colony picks were taken with a pipette tip and touched into the PCR mixture. These were inoculated into liquid media by touching the pipette tip to it prior to adding to PCR mixture for culturing and storage.

PCR products were cleaned up by either QIAquick PCR Purification Kit (QIAGEN, 27206) or using agarose gel electrophoresis followed by band removal under UV transillumination with a scalpel and tweezers and clean up with QIAquick Gel Extraction Kit (QIAGEN, 28076).

2.2.2 General PCR Primers

Primers were acquired from Sigma-Aldrich with desalt purification in water at 100µM. General primers used are listed in Table 4 used for strain identification. Unless noted to be from other publications primers were designed in Benchling[14]. Sequencing primers were designed where possible to have a length of 18-24 bases, a 40-60% G/C content, a G or C clamp in first and last 1-3 bases and to have matched T_ms between primer pairs of +/- 1°C. Min ΔG as calculated by Benchling of homodimers, monomers and

heterodimers between primer pairs was also checked and minimised below -10kcal where possible. Additional primers are listed in their relevant chapters.

Table 4: General Strain Sequencing Primers

Primer	Bases	Source/Purpose
16S_27F	AGAGTTTGATCMTGGCTCAG	[15] as 27f-CM - Strain and contaminant identification
16S_U515F	GTGYCAGCMGCCGCGGTA	[16] – Strain and contaminant identification
16S_U1071R	GARCTGRCGRCRRCCATGCA	[16] - Strain and contaminant identification

2.2.3 Optical Density and Mass

Optical density testing was carried out on 3 instruments. Optical density was always carried out at 600nm giving OD600 values. This measurement does not show optical density in the strictest sense but turbidity, but the OD nomenclature is common in microbiology.

The main spectrophotometer used was a Jenway 7415 (WZ-83056-23), with a minority of experiments using a Biomate 3S (Thermo Scientific, 840208300) and bioreactor fermentation work in Chapter 5 using a Jenway 6300 (WZ-83054-05) for offline readings. For comparability conversions were carried out for data in this work to be equivalent to the Jenway 7415. To create this conversion the same samples of *M. parvus* OBBP were tested on each spectrophotometer over a number of dilutions from 1 to 80x. The data from this is plotted in Figure 7. Conversion factors were calculated as a linear regression (y=mx+c) of one spectrophotometer against another. Zeroing components were minimal ranging from 0.0053 to 0.0079 so regressions were forced through the origin. Resultant conversions were Biomate 3S to Jenway 7415 y= 1.5816x, and from Jenway 6300 to Jenway 7415 y= 1.0774x with Pearson's r values of 0.9999 and 0.9995 respectively. The large difference between the Biomate 3S and the other two spectrophotometers is exemplary of why care must be taken in this regard.

Although it was confirmed using Figure 7 that OD response readings on the three spectrophotometers were linear to at least OD of 0.6 for the Biomate 3S or 0.9 for the Jenway 7415 and 6300, high readings were taken by dilution in an equivalent media (LB or NMS for example). Dilutions were carried out aiming below 0.2 OD and re-diluting for readings above 0.5. The resultant OD readings were multiplied back by the dilution factor to give final results.

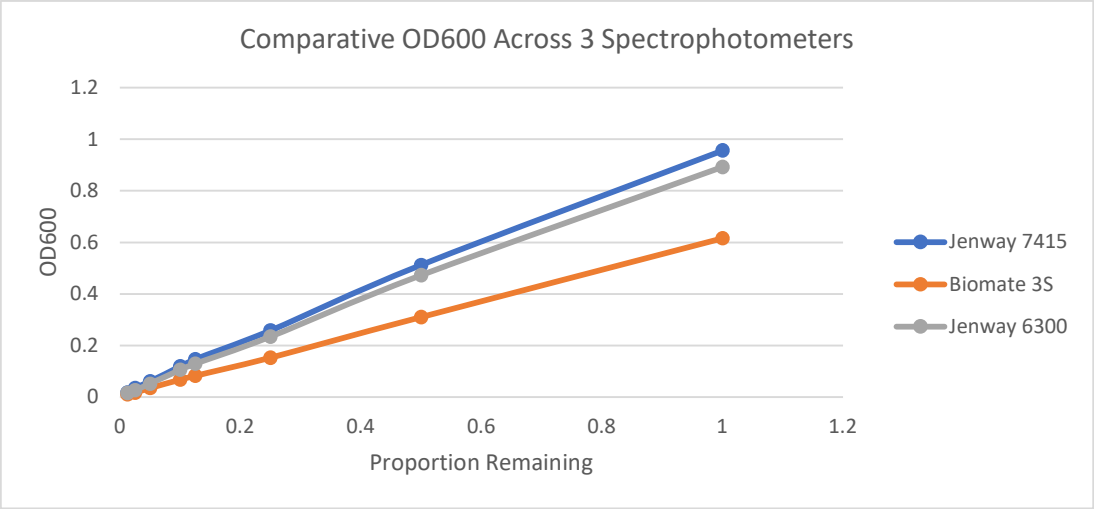


Figure 7: Comparative optical density readings of the same samples on three different spectrophotometers used in this work. Sample was *M. parvus* OBBP grown in NMS and diluted with NMS for dilutions of 1x 2x 4x 8x 10x 20x 40x and 80x.

Equation 1 was generated from pooled growth experiments predicting dry cell weight (DCW) after freeze-drying from optical density. Due to lack of equivalency of spectrophotometers the below formula will need to be adjusted depending on future user instrumentation. This formula allows a freeze-dried pellet of approximate weight to be generated from a volume of culture of known OD, simplifying and increasing reproducibility of downstream work. This conversion was generated utilising both PHB containing and PHB free cultures and appears effective in both cases to +/- 10%.

This demonstrates that OD can be used as an effective analogue of DCW biomass and it is utilised as such in this thesis.

Equation 1

$$DCW(mg) = vol(ml) \times OD \times 0.2545$$

2.3 PHB Production and Analysis

Standard PHB production from methanotrophs had two versions. A two stage and one stage method (Figure 8). In the two-stage method the serum bottles were inoculated with 3-4 day old liquid culture with a target OD600 of 0.02 calculated using Equation 2. These were supplied with methane, sealed and incubated as discussed previously. Bottles were re-gassed daily.

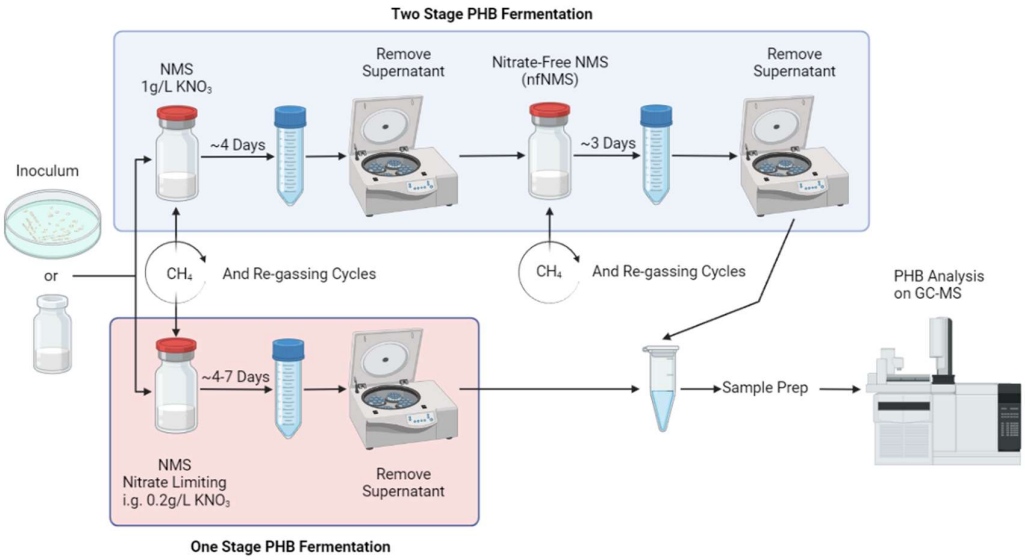


Figure 8: Process Diagram of the One and Two stage PHB fermentation methodologies. Figure created with BioRender.com

Equation 2

$$Volume\ Inocculum\ Added = \frac{OD_{final} \times Vol_{final}}{OD_{inoculum}}$$

After a specified number of days, generally four, seals were broken, and the culture transferred to a 50ml falcon tube and spun down at 8000rpm for 10 minutes. Samples were washed in an equal volume of nfNMS, resuspended in nfNMS again and transferred to a fresh serum bottle sealed and supplied with methane. The bottles were then incubated and regassed daily until experimental termination. Days of experiment are reported as D# where D0 is the day of inoculation or resuspension depending on the experiment.

In the one stage method bottles were set up as above but substituted NMS with a defined nitrate content usually 20% of original (0.2g/L) and incubated with regassing until termination of experiment.

With ether method at experimental termination an OD600 was taken and a defined volume by Equation 1 for a target weight of 10mg was spun down. This was washed once in one volume of dH_2O nfNMS and transferred to a pre-weighed 1.5ml microcentrifuge tube where it was spun down again at 13,000rpm and supernatant removed. The tube was sealed with Parafilm M (Bemis) and the top pierced twice then frozen at -80°C for 1hr+. Tubes were then freeze-dried to completion for ~16hrs either in a FreeZone 2.5L -84C (Labconco, 710201050) with an nXDS6i Vaccume Pump (Edwards, A73501983R) or MicroModulyo (Thermo) and RV5 Vacuum pump (Edwards, A65301903). After freeze-drying tubes were weighed again and the difference before and after (Tube Difference) used to inform PHB concentration in bottles. Pellets were then weighed (Pellet Weight) in aluminium foil lined weighing boats to avoid static and transferred to screw top glass test tubes. If pellets fragmented badly and stuck to the microcentrifuge tube remaining

particles were transferred to the final tube using chloroform and the Tube Difference used for pellet weight. For derivatisation 100µl 1mg/ml benzoic acid in 1-propanol, 2ml 25% conc. HCL in 1-propanol and 2ml chloroform was added to the test tube. This was heated on a dry block 100°C 2hrs then cooled to room temperature. 4ml of deionised water was added and vortexed for 30 seconds then left for phase separation for 5 minutes then the upper aqueous layer was discarded. The addition of water, vortexing, phase separation and discarding was repeated. 1ml of solvent layer was carried on to GC-MS analysis (Agilent Technologies MS 5973N, GC 6890N) using benzoic acid as the internal standard. Separation was carried out on a J & W DB-wax column (20m × 0.18mm, 0.18µm film thickness, Agilent, 121-7022) with a 250°C injection of 1 µL. The program was 5 min hold at 60°C, 20°C/min to 240°C, and hold for 6 min. Split injections were made with 10:1 split ratio. Hydrogen carrier gas was used at 0.6 mL/min. MS analysis was in scanning mode from 40 to 500 m/z produced with EI auto ionization with 3 min MS solvent delay and no additional voltage was to the electron multiplier.

PHB in the pellet was calculated according to Equation 3 as PHB dry cell weight ($PHB_{\%DCW}$) of a freeze-dried pellet. $PHB_{m/v}$ as mass per volume in culture in mg/L was calculated according to Equation 4. Tube difference is used to account for any loss in transfer from the freeze-drying tube and minimise a source of error. The difference was generally observed to be 5-10%. 10-15mg pellets are not essential but are used to maintain an accurate measurable mass that will be fully derivatised.

Equation 3

$$PHB_{\%DCW}\ (Purity\ \%) = \frac{PHB\ from\ GCMS\ (mg)}{Pellet\ Weight\ (mg)}\ \times 100$$

Equation 4

$$PHB_{m/v}\ (mg/L) = \frac{Tube\ Difference\ (mg)\ \times PHB_{\%DCW}}{Culture\ Volume\ Used\ (L)}$$

2.4 Statistical Analysis

Unless otherwise noted error bars are plotted as 95% confidence intervals (95%CI) calculated according to Equation 5. Standard deviations were calculated using Bessel's correction as in Equation 6. Values in text are also specified ±95% CI where appropriate.

Equation 5

$$95\%CI = 1.96\ \times SE = 1.96\ \times \frac{SD}{\sqrt{n}}$$

Where SE is standard error, SD is standard deviation, n is sample size.

Equation 6

$$SD = \sqrt{(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2)}$$

Where n is the sample size, $\bar{x}$ is the sample mean and x_i is each iterated sample in the summation.

All reported Student's t-tests are two-tailed assuming equal variance and a tested p-value of 0.05 is assumed significant. They are reported in the format t-test(df, p=p-value) e.g. Unpaired t-test(13, p=0.012). p-values are reported to 2 significant figures and values of less than 0.001 are presented as p<0.001. Statistical tests performed using jamovi (v2.3)[17] and R[18] with associated packages[19].

2.5 References

1. Kevbrina M V., Okhapkina AA, Akhlynin DS, Kravchenko IK, Nozhevnikova AN, Gal'chenko VF. Growth of mesophilic methanotrophs at low temperatures. Microbiology. 2001;70(4):384–91.
2. Simon R, Priefer U, Pühler A. A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. Nat Biotechnol. 1983 Nov;1(9):784–91.
3. Ferrières L, Hémery G, Nham T, Guérout A-M, Mazel D, Beloin C, et al. Silent Mischief: Bacteriophage Mu Insertions Contaminate Products of Escherichia coli Random Mutagenesis Performed Using Suicidal Transposon Delivery Plasmids Mobilized by Broad-Host-Range RP4 Conjugative Machinery. J Bacteriol. 2010 Dec 15;192(24):6418–27.
4. Whittenbury R, Phillips KC, Wilkinson JF. Enrichment, Isolation and Some Properties of Methane-utilizing Bacteria. J Gen Microbiol. 1970;(61):205–18.
5. Methanotroph Commons. General Culturing Tips | Methanotroph Commons [Internet]. [cited 2023 Jan 23]. Available from: http://www.methanotroph.org/wiki/culturing-tips/
6. Dalton H, Whittenbury R. The acetylene reduction technique as an assay for nitrogenase activity in the methane oxidizing bacterium Methylococcus capsulatus strain bath. Arch Microbiol. 1976;109(1–2):147–51.
7. López JC, Arnáiz E, Merchán L, Lebrero R, Muñoz R. Biogas-based polyhydroxyalkanoates production by Methylocystis hirsuta: A step further in anaerobic digestion biorefineries. Chem Eng J. 2018;(333):529–36.
8. Asenjo JA, Suk JS. Microbial Conversion of Methane into poly-β-hydroxybutyrate (PHB): Growth and intracellular product accumulation in a type II methanotroph. J Ferment Technol. 1986 Jan 1;64(4):271–8.
9. The British Standards Institution. BS EN ISO 80079 - Explosive atmospheres. 2019.
10. Haynes WM. CRC Handbook of Chemistry and Physics. 95th ed. Lide DR, Bruno TJ, Haynes WM, editors. Florida: CRC Press; 2014. 3–488 p.
11. National Oceanic and Atmospheric Administration, National Aeronautics and Space Administration, United States Air Force. U.S. Standard Atmosphere. Washington, D. C.; 1976.
12. Auman AJ, Speake CC, Lidstrom ME. nifH Sequences and Nitrogen Fixation

in Type I and Type II Methanotrophs. Appl Environ Microbiol. 2001;67(9):4009–16.
13. New England BioLabs. NEB Tm Calculator [Internet]. Available from: https://tmcalculator.neb.com/
14. Benchling. Benchling.
15. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. Appl Environ Microbiol. 2008;74(8):2461–70.
16. Wang Y, Qian PY. Conservative Fragments in Bacterial 16S rRNA Genes and Primer Design for 16S Ribosomal DNA Amplicons in Metagenomic Studies. PLoS One. 2009 Oct 9;4(10):e7401.
17. The jamovi project. jamovi.
18. R Core Team. R: A Language and environment for statistical computing. 2021.
19. Fox J, Weisberg S. car: Companion to Applied Regression. 2020.

Chapter 3: Complete genome sequence of the α-proteobacterial methanotroph genera type strain *Methylocystis parvus* OBBP

3.1 Introduction

At time of writing there are a great number of sequencing projects of methanotroph species in various stages of completion notably OMeGA (Organization of Methanotroph Genome Analysis) and the University of Washington in combination with the Joint Genomes Institute, California[1]. The majority of these date from post-2010 allowing much greater data for comparison in recent years. In the following chapter a survey of available genomes was carried out followed by sequencing and assembly of an improved genome of the type II methanotroph type strain *Methylocystis parvus* OBBP and an investigation into its chromosomal annotation and that of its two mega plasmids.

Prior to this work numerous sequencing efforts have been carried out in methanotrophs[2–6]. Within *Methylocystis,* this work's genus of interest, large numbers of assemblies are not assigned at species level (31 of 40 assemblies). Among these 40, 7 are complete genomes (Supplementary Table 1) including the *M. parvus* BRCS2 complete genome in 2021[7]. Until the work presented here however there was no complete genome of the *M. parvus* OBBP, type strain for the genus, and strain used for the majority of this work.

An incomplete *M. parvus* OBBP genome at the contig stage is available released by del Cerro et al. 2012[6]. This assembly (GenBank: GCA_000283235.1) named MetPar_1.0 consists of 108 contigs, N50-95,607, L50-18, 4,475,912bp. A gapped genome could lead to issues in editing including missing genes and unintended edits. This previous genome has been heavily used in research including the production of a genome scale model in 2019[8] so it was aimed to update this to a complete genome for the OBBP strain. As an improved genome, sequencing and error checking of the new genome were exhaustive to ensure a complete and reliable reference was achieved. The new assembly was completed *de novo* using a hybrid PacBio CCS long read, Illumina short read strategy.

The PacBio CCS (circular consensus sequencing) implementation used here forms single DNA fragments into a loop with adapters then replicates the loop repeatedly with DNApolymerase giving a sequence signal by fluorescent nucleotide incorporation. This gives multiple sequencing "subreads" of each single fragment that are aligned and a consensus formed to give a final extremely high accuracy polished read over lengths of 10-30kb. This avoids the low accuracy issues generally associated with long read sequencing from a single pass[9,10].

Part of the work in this chapter was published in "A Complete Genome of the Alphaproteobacterial Methanotroph *Methylocystis parvus* OBBP" Claxton Stevens et al. 2023[11]. A full copy of this paper is presented at the end of this thesis.

3.2 Methods

3.2.1 gDNA and Sequencing

Genomic DNA (gDNA) preparation and submission of sequencing was carried out by Chris Stead and Bashir Rumah. gDNA was extracted using a phenol:chloroform:isoamyl alcohol gDNA extraction method[12]. DNA quantity and purity was analysed by gel electrophoresis and Qubit fluorometric quantification (ThermoFisher Scientific, UK).

For long read PacBio HiFi, sequencing was carried out by Genome Quebec, CA. the DNA library was prepared following the Pacific Biosciences *Procedure & Checklist – Preparing Multiplexed Microbial Libraries Using SMRTbell Express Template Prep Kit 2.0* protocol. 4 µg of high molecular weight genomic DNA (final volume of 150 µl) was sheared using Covaris g-TUBES (Covaris, USA) at 4000 rpm for 60 seconds on each side, on an Eppendorf centrifuge 5424 (Eppendorf, Germany). The DNA Damage repair, End repair and SMRT bell ligation steps were performed as described in the template preparation protocol with the SMRTbell Express Template Prep Kit 2.0 reagents (Pacific Biosciences, USA). The sequencing primer was annealed with sequencing primer v4 at a final concentration of 1 nM and the Sequel II 2.0 polymerase was bound at 0.5 nM. The library went through an AMPure bead cleanup (following the SMRTlink v8 calculator procedure) and no size selection step was carried out. The prepared library was sequenced on a PacBio Sequel II instrument at a loading concentration of 110pM using the diffusion loading protocol, Sequel II Sequencing Kit 2.0, SMRT Cell 8M and 15 hours movies with a 2h pre-extension time. The PacBio SMRTPipe pipeline was followed on the generated raw reads which were collapsed and error corrected resulting in 145,723 ~10kbp CCS reads totalling 1,333,680,793bp and an N50 of 9,860bp (Acc: SRX14359361) from here onwards referred to as PacBio reads. Methylation data was also produced as part of PacBio sequencing.

For short read Illumina, genomic DNA libraries were prepared using a Nextera XT Library Prep Kit (Illumina, San Diego, USA) and processed on an Illumina MiSeq by Deep Seq, University of Nottingham, UK using a 250 bp paired end protocol resulting in 987,357 paired reads totalling 480,751,157bp and an N50 of 250bp. (Acc: SRX14359360).

Quality of read sets was checked using FastQC(v0.11.9)[13] confirming no adapter content and the lack of various biases. Additional sequencing was carried out on PCR products using primers supplied by Sigma-Aldrich, USA and Sanger and listed in Table 5. PCR products were sequenced by Europhins Genomics, DE or Genewiz, Azenta Life Sciences, USA. Additional gDNA for

PCR was extracted using a Monarch Genomic DNA Purification Kit (T3010, NEB) with the gram-negative bacteria protocol with Lysozyme. PCR was performed with Q5 High-Fidelity 2X Master Mix (M0492, NEB) with 400nM primer concentration and gel extracted using a QIAquick Gel Extraction Kit (28706, Qiagen). Sanger reads were deposited in the NCBI SRA (Acc: SRX19212503).

Illumina, PacBio and Sanger reads, and methylation data were deposited as part of NCBI BioProject PRJNA812408. Methylation data can be found in the base modification file linked in the genome GenBank listing.

Table 5: List of Primers used for genome sequencing. Patch, MP_Seq and Dupe primer locations are indicated by the number and L or R showing positioning out from that that end of the numbered region in OBBP-Unicyc in Figure 9A.

MP_Seq_1L	GCCGGTCGCATATCCTATG	**MP_Seq_2R**	TTCACCTTTCCCTCACGG
MP_Seq_3L	CCTTCGATCTCTGTTATTCCGA	**MP_Seq_4R**	AATGGGGAAACCCACCTTC
MP_Seq_5L	CACCTTTCCCTCACGGTAC	**MP_Seq_5R**	TTGACGCAGATGAGCGTC
Patch_5R_Out	AAGGACTGCCAGTAAGCG	**Patch_1L_Out**	GAATACACGTGTCAGCGC
Dupe_2R_Out	CTCTCCAAGCCAAAAGCGT	**Dupe_4R_Out**	CGTCCATCCACCTCGATC
Dupe_5L_Out	GAAGGATCACTTTCTCGGCC	**Dupe_3L_Out**	AACAGCGCGCGAAGATT
Pilon_Er_7_R	CCGCTCGCGTATGTTGA	**Pilon_Er_7_F**	CCATAGGGGTTCCACGTC
Pilon_Er_Plas2_F	CATCCTCGTCTTTCGTGACC	**Pilon_Er_Plas2_R**	CGAATTTTGCAGGAAAGCCG
Pilon_Er_Plas3_1-3_F	ACGCCCCTAATTCGCAAC	**Pilon_Er_Plas3_1-3_R**	CGTAGTAGGCCTCGATCTTG

3.2.2 Genome Assembly

Linux software was operated on Ubuntu 20.04.3 through the Windows Subsystem for Linux in Windows 10. Where necessary, versions of dependent programs are indicated in Supplementary Table 2. Where possible programs were installed and managed using Anaconda 4.11.0[14] with implemented conda environments for each program. Otherwise, they were acquired from respective home websites or Github.

An assembly (*M. parvus* OBBP-Nanuq) was carried out using only PacBio reads by Genome Quebec, CA using SMRTLink 8.0.0.80501 the results of which are summarised in **Table 6** along with comparative data from the *M. parvus* BRCS2 genome. OBBP-Nanuq was supplied tagged as un-circularised across 3 contigs. This could indicate an incomplete gapped region was stopping circularisation or overlap that has not been closed to circularise the contigs[3]. Efforts were made to close the contigs using manual region matching without success. Two additional programs were tested, Simple-Circularise[15] and Circlator[16] both failed to close gaps.

To attempt complete circularisation an assembly was produced using Unicycler[17], an assembler with an integrated circularisation step. Unicycler is a *de novo* assembler pipeline focused on hybrid assembly of bacterial genomes leading with a short read assembly with SPAdes optimising over a variety of k-mer sizes and utilising paired end Illumina data. Long reads are assembled with miniasm and polished with Racon. Short and long read contigs are

combined using its own semi-global aligner and possible joins in the assembly graph are scored on a variety of qualities with the best selected for the final assembly. Final polishing steps with Pilon are carried out until the assembly becomes stable with no changes made on successive rounds. It also carries out contaminant DNA removal among other advantages to nonspecialised assemblers and allows usage of Illumina paired end read data in the assembly stage which are lost in a long-read assemble, short read polish methodology[17].

The resulting assembly (*M. parvus* OBBP-Unicyc) produced 120 initial contigs which were joined and polished into 3 unitigs 4,060,921bp, 248,224bp and 204,886bp totalling 4,514,031bp. The smaller two contigs were complete and circularised the longer was not. This suggests the smaller two are megaplasmids and the larger is the chromosome. By comparison with the *M. parvus* BRCS2 and OBBP-Nanuq assemblies the plasmid sizes are almost identical however OBBP-Unicyc chromosome is 15kb shorter. The previously unnamed plasmids were named pMpar-1 and pMpar-2 respectively in descending order of size.

Conventional hybrid assembly methodologies utilise error prone long reads for structural correctness followed by polishing using a greater read depth of less error prone short reads[17,18]. In this case an exceedingly high coverage of PacBio reads (~300x) was acquired resulting in a greater depth than that acquired by Illumina (~100x). The paired end nature of the Illumina reads however, and reduced error rate still prove beneficial in correction and the identification of structural errors.

Table 6: Genome statistics for the *M. parvus* genomes available or generated.

	BRCS2[7]	MetPar_1.0[6]	OBBP-Nanuq	OBBP-Unicyc	OBBP-Final
Contigs	3	108	3	3	3
Total	4,529,043	4,475,912	4,529,118	4,514,031	4,529,117
Chromosome	4,075,934		4,076,008	4,060,921	4,076,007
Plasmid 1 / pMpar-1	248,223		248,224	248,224	248,224
Plasmid 2/ pMpar-2	204,886		204,886	204,886	204,886
N50	4,075,934	95,607	4,076,008	4,060,921	4,076,007
GC%	63.36				

3.2.3 Comparing and Improving Assemblies

3.2.3.1 Closing the Assembly

An initial comparison was carried out using progressiveMauve in Mauve (v20150226)[19] whole genome aligner and visualiser show in Figure 9A, aligning OBBP-Unicyc, BRCS2 and OBBP-Nanuq. The respective ends of the OBBP chromosome assemblies (5-1 and 3-4) were in different locations indicating coverage of each other's gaps. On inspection the OBBP-Nanuq ends aligned together with no gap to OBBP-Unicyc and BRCS2 suggesting the OBBP-Nanuq chromosome assembly was in fact closed and labelled unclosed in error.

The alignment of the ends in OBBP-Unicyc between segments 1 and 5 indicated a missing region of 15,087bp covered in both other assemblies. PCR

and Sanger sequencing was carried out to confirm the match using Patch_5R_Out+MP_Seq_5R and Patch_1L_Out+MP_Seq_1L. This confirmed the ends of the region. It was concluded the 15kb patch was correct and it was applied to OBBP-Unicyc closing the assembly. This new assembly was named OBBP-UnicycV2

3.2.3.2 Pilon Polish and Reindexing

An additional Pilon[20] polishing step was carried out with the Illumina reads and PacBio reads. To operate Pilon first the Java setting file was updated to increase memory allocation (default_jvm_mem_opts) from 1GB to 8GB. Alignments against OBBP-UnicycV2 to provide to Pilon for this step were produced separately for Illumina and PacBio reads as the optimal aligners differ. Illumina reads were aligned using Bowtie2(v2.4.4)[21] in paired end "end to end" global mode. PacBio read files in .bam format were converted into .fastq.gz format using the bam2fastq command of BAM2fastx(v1.3.1)[22] aligned using the HiFi specific settings in minimap2(v2.24)[23]. Both alignment output SAM files were converted to sorted BAM files and indexed with Samtools(v1.14)[24]. Properties of the alignments were extracted using coverage in Samtools and presented in Table 7. Pilon polishing was run on the Illumina alignment which resulted in no changes. The PacBio alignment was then polished which made 14 SNP changes and a 1bp insertion. Illumina reads were once again aligned and polished and this did not revert any changes made from the PacBio reads. The PacBio changes were thoroughly investigated by visualising the Illumina and PacBio alignments using IGV(v2.11.9)[25] and all changes were mixed reads in Illumina but decisive in PacBio and were confirmed to the PacBio version. These were attributed to multimapping of Illumina reads. The output file constituted the final complete OBBP assembly OBBP-Final which contains 3 closed contigs with no scaffold or ambiguous bases. Pilon also outputs a list of errors which it cannot fix which will be discussed in 3.2.4.2. Results in Table 7 indicate a mean depth of aligned reads of 293x for PacBio data and 104x for Illumina for a total depth of 397x.

With both OBBP-Nanuq and OBBP-Final assemblies closed and circularised they were reindexed for uniformity and ease of comparison. The *dnaA* gene on the chromosome was identified by BLAST search and reindexed to start with *dnaA* in forward orientation on the chromosome in Benchling[26]. The start of the *repABC* operon was identified on each plasmid and each reindexed similarly.

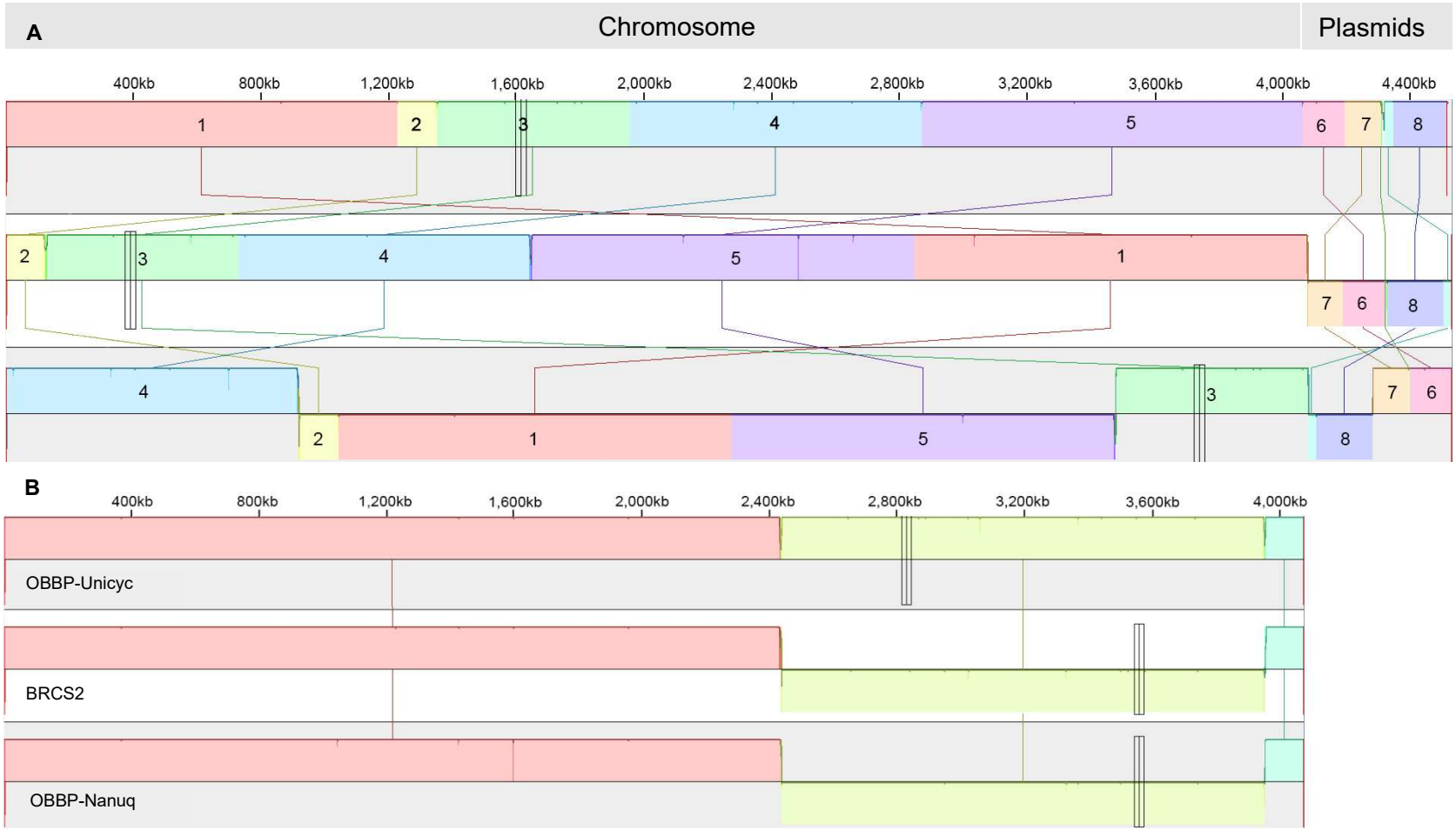


Figure 9A and **B**: Output of progressiveMauve in Mauve full genome aligner **A** - From top to bottom, *M. parvus* OBBP-Unicyc, BRCS2, OBBP-Nanuq. Segment numbers added to aid reference and indicate section alignment. 1-5 are the chromosome, 6 and 7 are pMpar-1, 8 is pMpar-2. More segments are present than might be expected due to the genomes being uncircularised. Sections 2-1-5 are inverted. **B** – The chromosomal region positionally indexed to *dnaA*. From top to bottom, *M. parvus* OBBP-Nanuq, OBBP-Final and BRCS2. Showing the inverted region (yellow) between 2,432,000 – 3,957,000bp. The triple lined marker indicates a similar position on each genome showing the inversion.

Table 7: Properties of produced alignments. Created with *coverage* from SAMtools.

	Length	Reads Mapped	Covered Bases	Coverage %	Mean Depth	Mean Base Quality	Mean Map Quality
OBBP-Final with PacBio							
Chromosome	4,07,6007	138,402	4,076,007	100	307.837	67.8	60
pMpar-1	248,224	4,587	248,224	100	159.288	69.1	59.8
pMpar-2	204,886	3,482	204,886	100	149.685	69.4	60
OBBP-Final with Illumina							
Chromosome	4,076,007	1,771,751	4,074,639	99.9664	105.757	31.3	39.6
pMpar-1	248,224	97,706	248,115	99.9561	95.8108	31.7	38.9
pMpar-2	204,886	75,212	204,630	99.8751	89.3185	31.8	38.1
BRCS2 with BRCS2 Illumina							
Chr	4,075,934	1,813,215	4,075,251	99.9832	108.857	32.7	40
plasmid 1	248,223	88,425	248,223	100	87.0506	33.2	39.2
plasmid 2	204,886	67,000	204,886	100	79.9346	33.4	38.2
OBBP-Final with MetPar_1.0							
Chromosome	4,076,007	87	4,034,731	98.9873	0.99008		55.4
pMpar-1	248,224	9	241,350	97.2307	0.972307		46.7
pMpar-2	204,886	7	195,749	95.5404	0.955404		43

3.2.3.3 1.5Mb Inversion and Other Structural Misassembles

In Figure 9A segments 2-1-5 covering approximately 1.5Mb in OBBP-Nanuq appear inverted in relation to the other two assemblies. This indicates a structural misassembly i.e. one of them is wrong. To investigate more clearly a new whole genome alignment covering only the chromosome was produced (Figure 9B) using BRCS2, OBBP-Nanuq and OBBP-Final. BRCS2 was also reindexed to *dnaA* as Mauve handles circular genomes poorly. This clarified the high similarity between the genomes except this large inverted structural variation. It was identified to occur between 2,432,000 – 3,957,000bp. PCR and Sanger sequencing using primer pairs MP_Seq_2R+3L and 4R+5L and inspection and BLAST search indicated a repeat region of 6048bp present at junction points 2-3 and 4-5 containing 5S, 16S and 23S rRNA genes and the Ile, Ala and Met tRNAs.

To identify the correct structural variant of the inversion, PacBio alignments produced previously were inspected using IGV (Figure 10A and B). This showed a sharp edge in read depth around the 6kb repeat regions (2,431-2,438kb and 3,950-3,957kb, blue arrows Figure 10A) in OBBP-Nanuq. Inspection of reads showed many were soft-clipped by 1000's of bp where reads form a sharp edge as they move from an area they match to into an area they do not shielding the misassembly by allowing incorrectly multi-mapped reads to be placed. Soft-clipping is an intentional and required feature for long read alignment but can produce errors around very long repeats. The same region in OBBP-Final shows no significant drop off in read depth at the edges of the repeat region suggesting this is the correct structural order.

To reveal the error more clearly samclip[27] v0.4.0 was applied which removes clipped reads from an alignment while preserving reads close to the end of an assembly, important in circular genomes (Figure 10 B and C). This was tuned over a range of stringencies allowing for 5, 100, 200, 500 or 1000bp clipped ends or below to remain. 100bp length was chosen removing 0.64% of reads from OBBP-Final and 0.78% from OBBP-Nanuq. Results at the 6kb duplicate region were visualised in IGV and are shown in Figure 10C. This indicated a wide area of reduced read depth in OBBP-Nanuq from ~310x to ~170x with OBBP-Final showing a small dip is from ~285x to ~270x. This decrease is attributable to removal of incorrect multi-mapped reads which is more severe in the incorrect OBBP-Nanuq assembly. Note some multi-mapped reads remain which could be removed by more stringent soft-clip removal, but this could compromise the overall assembly. A complete drop to zero coverage in the area in OBBP-Nanuq was not expected as reads covering from each end but not reading through would still map correctly.

To confirm PCR was carried out using Dupe_2R_Out, Dupe_4R_Out, Dupe_5L_Out and Dupe_3L_Out which were designed to match outside the two 6kb duplicate regions. For OBBP-Final to be correct products with 2R-3L and 4R-5L were expected, for OBBP-Nanuq 3L-5L and 2R-4R was expected. Strong ~7kb expected products were produced by the OBBP-Final pairs and Sanger sequenced confirming the duplicate region ends were placed as

expected. From the above evidence it was concluded OBBP-Final was the correct structural assembly. The same PCR confirmation was carried out on BRCS2 also confirming the structural order present in that assembly is correct around the same regions.

Two other regions (red arrows Figure 10A) were identified at 63kbp-73kb and 2,550-2,560kb the first of which is shown in Figure 10D. These have unexpectedly high read depth of ~440x and a noticeable edge again caused by soft-clipping and multi-mapped with other repeat regions. Application of samclip smoothed the edge and reduces the depth in these regions to ~360x. BLAST search showed these regions were 8,393bp exact repeats containing iron sulphur cluster related genes, *nifSUVW* involved in nitrogen fixation and *cycE* involved in cystine synthesis. As read depth returned to normal it was concluded these were not structural misassembles. No other areas of soft-clipping drop off were observed along the genome.

It is possible that the multi-mapped soft-clipped reads in all four repeat regions (red and blue arrows) could obscure base changes. To confirm, Pilon was run on both OBBP-Nanuq and OBBP-Final samclipped alignments, but no changes were made and no changes in the consensus were observed visually after applying samclip. They appear to be true repeats in all four locations. Due to the short length (250bp) of Illumina reads and of the insert size used (~600bp avg) they would not be helpful in assessing structural variations of this size. The PacBio alignment also covered and confirmed the 15kb patch region applied to OBBP-Unicyc previously.

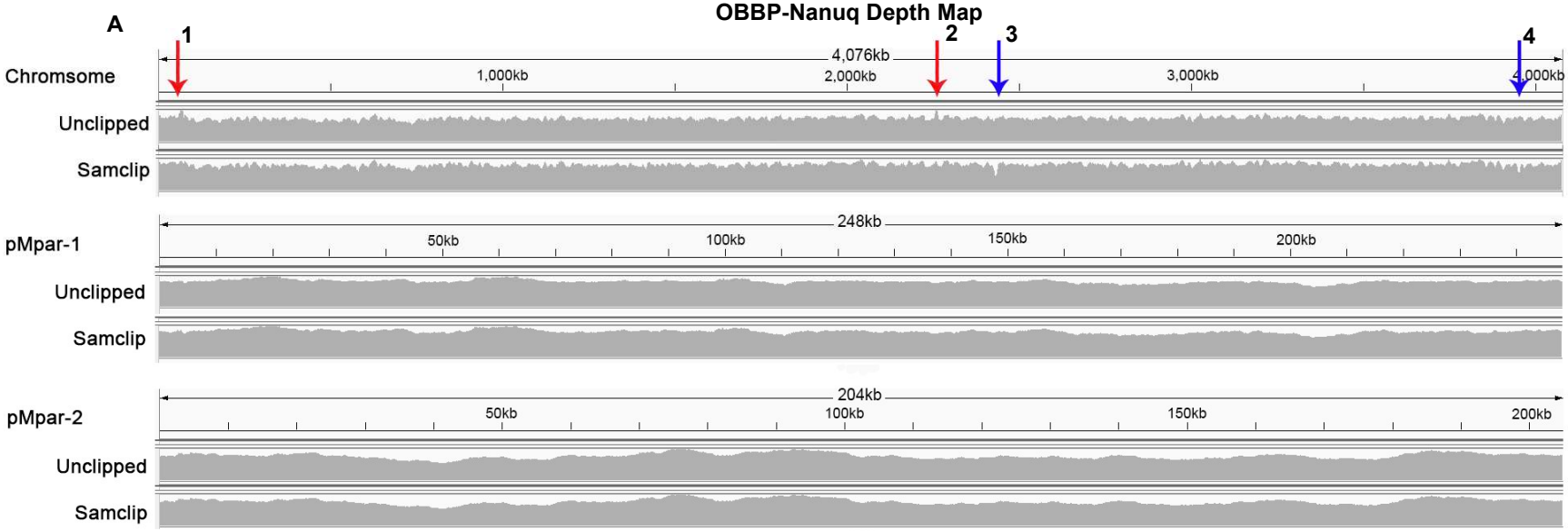
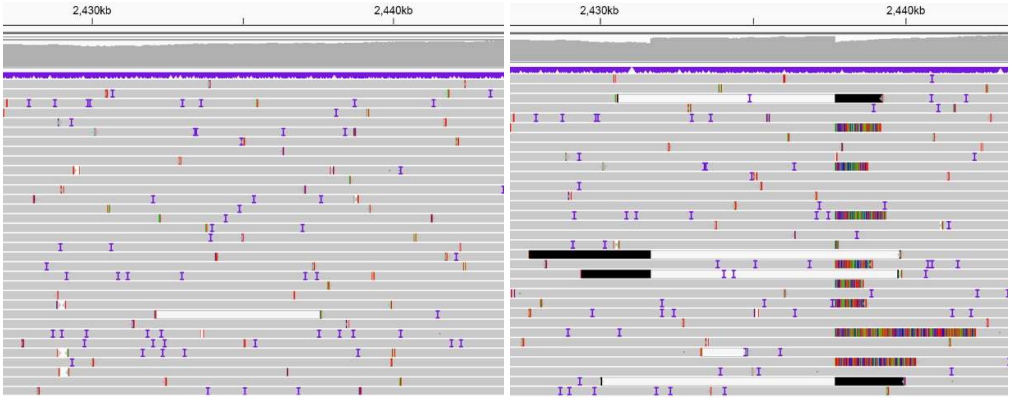


Figure 10 A-D: PacBio alignments showing read depth irregularities and the effect of samclip with a 100bp cutoff removing severely soft-clipped PacBio reads. **A** - Read depth over the whole length of the three OBBP-Nanuq assembly contigs. Red arrows (no. 1 and 2) - two regions with elevated read depth and sharp cut-offs. These peaks decrease in the samclipped alignment. The red arrow regions also occur in OBBP-Final in the same locations. Blue arrows (no. 3 and 4) - the 6kb repeat region ends of the inversion which show a drop in read depth in the samclipped alignment. No change was observed in the plasmid alignments. **B** – Close up of OBBP-Nanuq and OBBP-Final covering the 3' blue arrow (no. 3) location of the 6kb repeat region 2,431-2,438kb including the 6kb duplicate region at the 3' end of the 1.5Mb inverted region. **C** – Close up of the 3' blue arrow (no. 3) location of the 6kb repeat region showing the drop in depth of OBBP-Nanuq and a much smaller drop in OBBP-Final. **D** – Close up of the 5' red arrow (no. 2) in **B** showing the sharp edge smoothed in the samclipped alignment. Results similar for OBBP-Final and OBBP-Nanuq. Grey boxes are PacBio sequencing reads, white boxes are reads that have multi-mapped to more than one location, black and multicoloured regions propagating from them are soft-clipped regions not used in the alignment or polishing but shown for clarity. Read depth is shown in the grey graphs at the top of alignments. Blue Is indicate insertions and coloured bars are base mismatches. Alignments visualised using IGV.

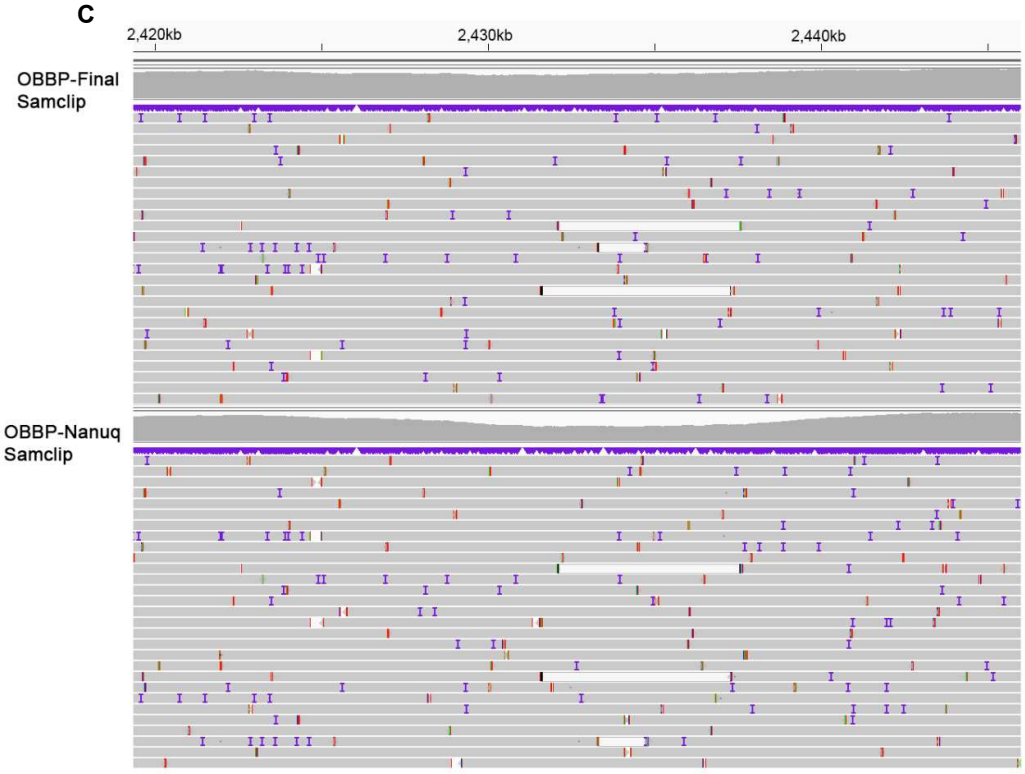
3' 6kb Repeat Read Alignment (Arrow 3)

B OBBP-Final OBBP-Nanuq

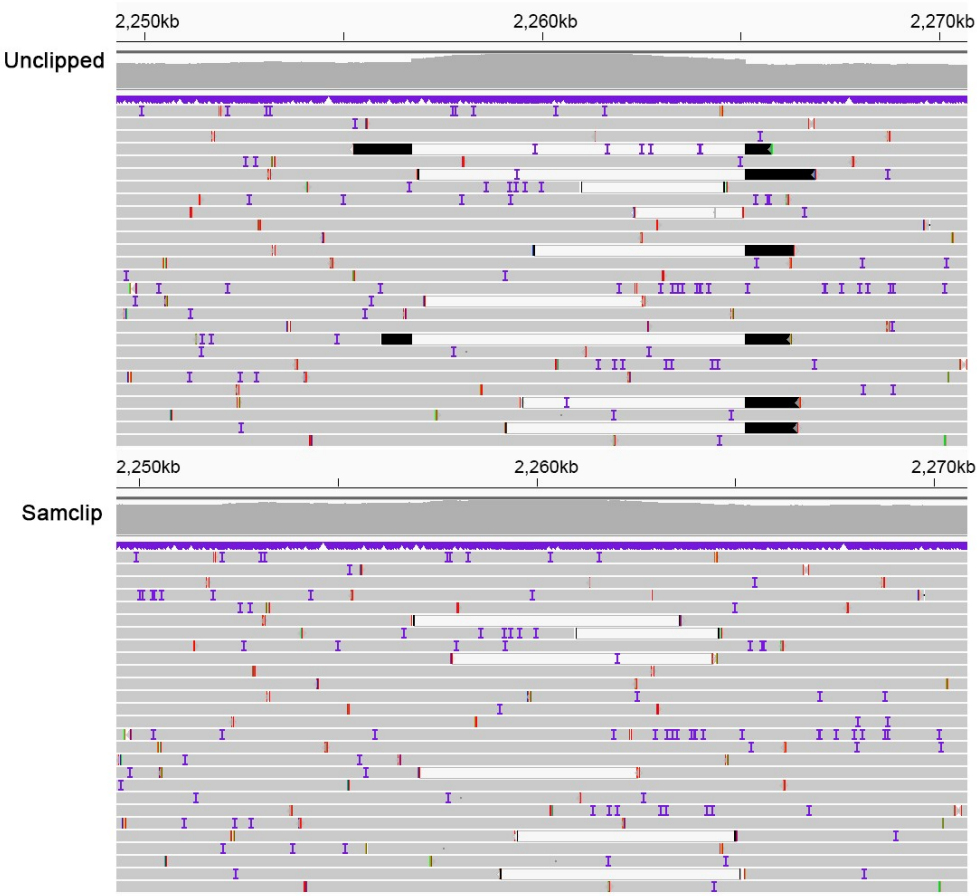


3' 6kb Repeat Read Alignment (Arrow 3)

C



D **5’ High Cover Region Read Alignment (Arrow 2)**



3.2.4 Assessment of Assembly Changes, Errors and Quality

3.2.4.1 Gene Completeness Assessors: QUAST, CheckM and BUSCO

The genome was assessed using a variety of available tools. A QUAST[28] report was generated via the web interface without a reference genome. This incorporates many assessment tools including BUSCO[29] v3.0.2 and HMM based gene prediction. QUAST predicted 4,336 genes of which 4,308 were unique using GlimmerHMM. A BUSCO score is generated by comparison of genes predicted using Augustus (or Prodigal in more recent versions), to a database of near universal single copy ortholog marker genes from OrthoDB[30] that can be tuned to its evolutionary relatives. This BUSCO output is presented in Table 8 showing 97.3% completeness.

The version of BUSCO run as part of QUAST is outdated, an updated version of BUSCO v5.3.0 was run separately with improved libraries for prokaryotes[31]. Two datasets were used: the α-proteobacteria dataset and the Rhizobiales. These datasets showed 99.6% and 98.4% completeness respectively. Expected genes marked as missing or duplicated were identified by lookup in OrthoDB and are listed in Supplementary Table 3.

Table 8: Outcome of running versions and datasets of BUSCO genome completeness analyser. BuscoIDs from the missing genes in the QUAST implementation were not supplied.

	QUAST (BUSCO v3.0.2)	**BUSCO v5.3.0**	**BUSCO v5.3.0**
Dataset	bacteria_odb9	alphaproteobacteria_odb10	rhizobiales_odb10
BUSCOs		**(Number - % of Total)**	
Complete	144 – 97.3%	430 – 99.6%	629 – 98.4%
Complete Single Copy	144 – 97.3%	428 – 99.1%	622 – 97.3%
Complete Duplicated	0	2 – 0.5%	7 – 1.1%
Fragmented	1 – 0.7%	0	0
Missing	3 – 2%	2 – 0.4%	10 – 1.6%
Total Searched	148	432	639

Similarly to BUSCO, CheckM v1.1.3[32] also uses sets of marker genes to assess completeness of a genome however the tool is bacteria specialised and the marker gene set is identified by lineage during processing. This identified 478 marker genes all of which were present in OBBP-Final giving a 100% completeness score, 3 markers appeared in duplicate. 4,340 genes were predicted with a 89.67% coding density. It was given a 0.95% contamination score largely based on the duplication of marker genes. This is similar to results found for *Methylocella tundrae* T4 (completeness 98.35%, contamination 0.73%)[2]. Duplicated marker genes are listed in Supplementary Table 3.

Duplicate genes were inspected for proximity to each other in case of clustering that might indicate a contaminant DNA insertion, but none were identified. If any future issues are found these may assist in identifying contaminant regions. Some changes in essential genes is not unexpected as the metabolism of methanotrophs like *M. parvus* differ greatly from the majority

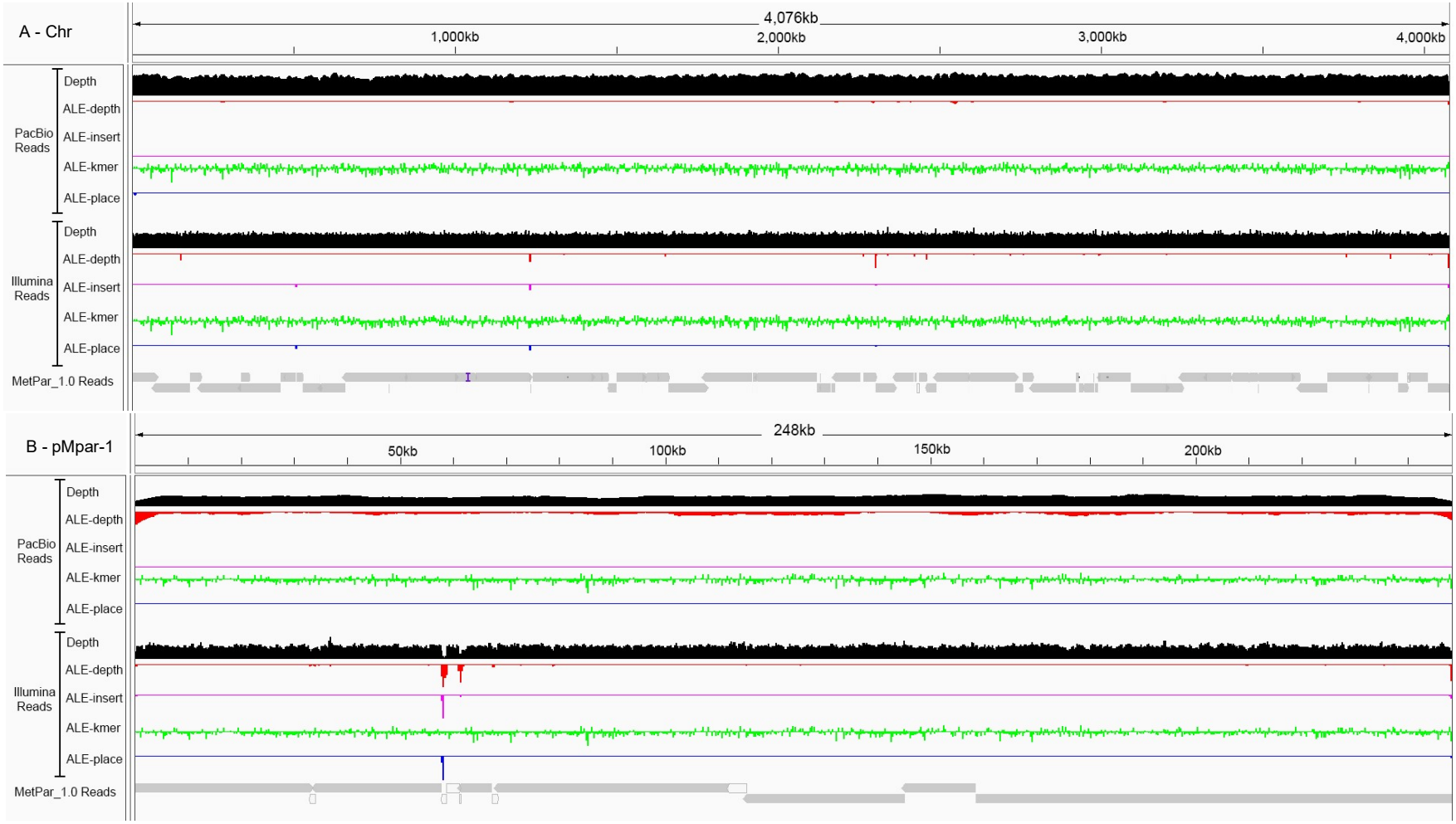
of polytrophs in its order (Rhizobiales) and class (α-proteobacteria) level clade[33].

3.2.4.2 Structural and SNP Assessors: ALE and Pilon

ALE[34] v0.9, a highly diverse assembly assessor was run using both the PacBio reads and Illumina reads. ALE interprets 4 factors: depth, kmer, insert and place. Depth scores for read depth at the location accounting for biases caused by GC rich regions. Kmer uses the assembly to look for over or underrepresented kmers that might indicate contamination. Insert uses Illumina mate pair data to identify areas of misassembly, constriction or expansion. Place identifies disagreements between the aligned reads and the assembled genome. These are combined into a score that can be used to compare genomes and as graphs that can be used in analysis. ALE possesses an arbitrary base quality score cut off of >62 above which it rejects sequencing reads. This does not take into account the high-quality scores awarded to PacBio HiFi reads go up to 93. To allow processing PacBio reads were artificially limited to a base quality of 62 with the reformat program from BBMap[35]. Alignments produced for each read type against the OBBP-Final as described previously were submitted to ALE and visualised with IGV. Resulting ALE graphs are shown in A-C.

Pilon on OBBP-Final using Illumina (section 3.2.3.2), identified 20 breaks in the chromosome, 3 in pMpar-1 and 4 in pMpar-2 which it was unable to fix. None of these issues were identified with Pilon using the PacBio reads. Issues were examined individually and all but one were also identified by ALE, using the Illumina reads, but not by ALE using the PacBio reads. Issues were generally called by ALE-depth but a small number with ALE-insert and place. ALE-kmer showed no issues. Inspection of the Illumina read alignment showed there were gaps with zero coverage at these areas, but they were covered over by the PacBio reads successfully.

Most problem areas indicated by ALE in also line up with breaks in the MetPar_1.0 assembly contigs indicating areas this assembly (produced by short read 454 pyrosequencing[6]) also had difficulties. Pilon was also run using the BRCS2 assembly and the BRCS2 Illumina reads, and this resulted in matching errors in 14 of the 20 chromosomal error regions. This suggests the areas might be generally challenging to short read sequencing. Error areas were extracted using the faidix program from SAMtools, the resulting areas on the chromosome had a length averaged mean GC of 70.47% indicating high GC areas, though this does not account for all regions, four were below 65% GC two ~50% GC. On the plasmids all seven were within normal GC range of 53-64%. It is unclear what property might make these areas hard to sequence for short reads in all cases. Previous studies have suggested causes including active transposons, phage integration, repeats and strong secondary structure[36]. PCR and Sanger sequencing was carried out using Pilon_Er_7_R+F, Pilon_Er_Plas2_F+R, Pilon_Er_Plas3_1-3_F+R which collectively covered 6 of the Pilon identified errors. These all matched OBBP-Final confirming that the assembly was correct.


A - Chr
4,076kb
1,000kb
2,000kb
3,000kb
4,000kb
PacBio Reads
Depth
ALE-depth
ALE-insert
ALE-kmer
ALE-place
Illumina Reads
Depth
ALE-depth
ALE-insert
ALE-kmer
ALE-place
MetPar_1.0 Reads
B - pMpar-1
248kb
50kb
100kb
150kb
200kb
PacBio Reads
Depth
ALE-depth
ALE-insert
ALE-kmer
ALE-place
Illumina Reads
Depth
ALE-depth
ALE-insert
ALE-kmer
ALE-place
MetPar_1.0 Reads

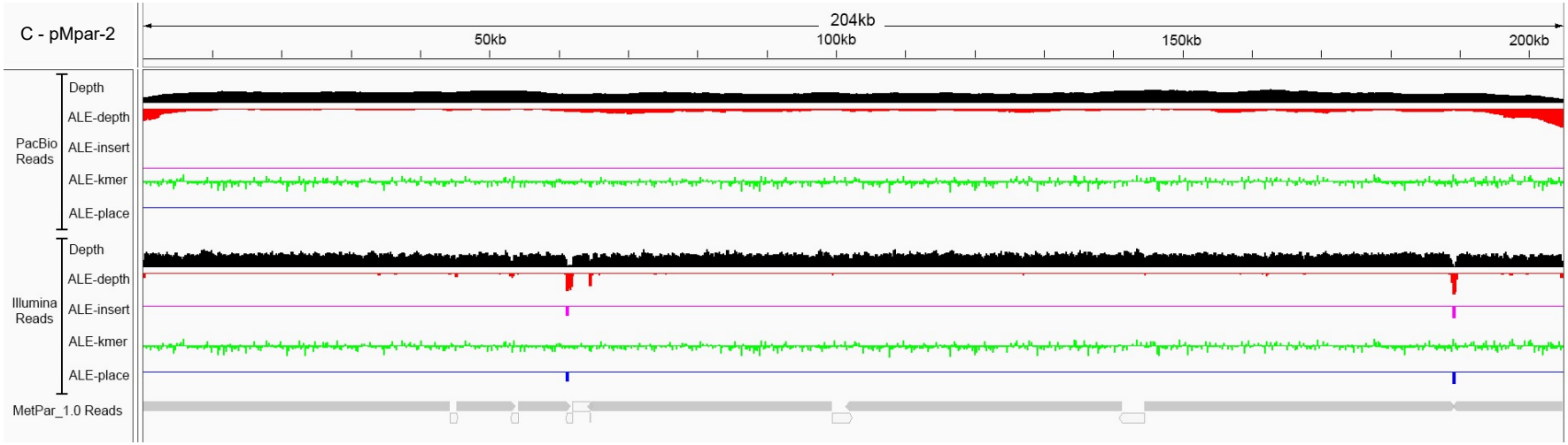


Figure 11 A-C: Read depth of Pac Bio (top) and Illumina (bottom) reads in black, ALE genome assessment outputs of aligned PacBio and Illumina reads and MetPar_1.0 alignments for each component of the OBBP-Final genome **A** - Chromosome **B** - pMpar-1 **C** - pMpar-2. Visualised in IGV. Grey boxes for MetPar_1.0 are individual contigs from the previous genome assembly, white boxes are reads that have multi-mapped to more than one location. Note MetPar_1.0 reads are aligned here for reference but were not used in the OBBP-Final genome assembly.

Alignment and processing with ALE was also carried out on the OBBP-Nanuq assembly. The comparison of the PacBio scores showed OBBP-Final had a 2.37% higher ALE score indicating a better assembly.

3.3 Results and Discussion

3.3.1 Review of NCBI genomes

A survey of available genome data on the NCBI database[37] was made, presented in Supplementary Table 1. Genomes were counted at the genus level by their completion status in accordance with the NCBI definitions into four categories, Complete, Chromosome, Scaffold and Contig. These definitions are made by the highest assembled feature in the assembly for example chromosome level may still include linkage mapped groups, scaffolds and contigs. The level of assembly effects usage of the data in certain circumstances like identification of plasmids, whole genome alignment and confirming the absence of genes. Records were made of the number of represented species, total number of strains and the phylum, family and methanotroph type group.

All notable genera of methanotrophs identified (n=21) in literature were searched for individually, forming the first section, as some of these have no sequenced genomes resulting in empty rows. A search was then made by family to find any additional genera not in the original search and these are presented in the second section, some of which may be of putative nature or outright disputed in some literature[38] . Excluded were additional *Verrucomicrobia* and NC10 genera as their methanotroph capability was unevidenced along with mass sequencing projects with no genera information. This resulted in a total of 64 named species, 224 represented strains with 40 complete, 4 chromosome 70 scaffold and 114 contig level assemblies.

Individually sequenced regions like 16S ribosomal sequences or the signature methane monooxygenases pMMO and sMMO are not included though are existent for many more strains and species.

3.3.2 Assembly Results

The final genome (OBBP-Final in **Table 6**) was assembled at high coverage and checked exhaustively as described in the methods. It consists of a chromosome (4,076,007bp) and two plasmids, pMpar-1 (248,224bp) and pMpar-2 (204,886bp) with no gaps or undecided bases and 63.36% GC. The genome achieved a BUSCO completeness of 99.6% against the v10 Alphaproteobacteria data set and a CheckM score of 100% indicating an expected complete genome. These were deposited in GenBank under the accessions: CP092968 (chromosome), CP092969 (pMpar-1) and CP092970 (pMpar-2).

3.3.2.1 Comparison to MetPar_1.0 assembly and *M. parvus* BRCS2

The *M. parvus* OBBP-Final assembly extends the previous *M. parvus* OBBP MetPar_1.0[6] assembly by 53.2kb in raw length. To get a more accurate estimate and investigate any differences MetPar_1.0 was retrieved and aligned against OBBP-Final using minimap2 with asm5 settings for whole genome alignment (Figure 11). Alignment results are shown in Table 7 and indicate a better estimate of 57.3kb of new coverage in OBBP-Final after eliminating duplicate contig regions in MetPar_1.0. There was also an obvious increase in L50 and N50 by completion of the genome. Investigation of the SAM mapping file showed 9 contigs from MetPar_1.0 did not map, and 15 multi-mapped. 9 unmapped contigs totalling 4.3kb were BLAST searched on the NCBI database and were all identified to be contaminants, six to *Priestia megaterium*, one to *Stenotrophomonas rhizophila*, one to *Mesorhizobium sp* and one to *Pseudomonas putida*. All BLAST matches had >95% identity, generally 99-100% and 95-100% coverage. BUSCO 5.3.0 scores were checked for MetPar_1.0 and showed a 0.6% decrease in completeness for the α-proteobacterial dataset and 0.3% for the Rhizobiales, both with an additional 2 fragmented BUSCO groups compared to the scores for OBBP-Final.

Whole genome alignment of the *M. parvus* OBBP-Final genome against *M. parvus* BRCS2 in Mauve[19] allows SNPs and gaps to be called which resulted in 2 SNPs and 11 gaps, 8 of the gaps were in BRCS2 totalling 220bp and 3 in OBBP totalling 146bp. This indicates very high similarity between the two strains.

3.4 Annotation and Analysis of the Final *M. parvus* OBBP Assembly

Annotation of the genome was performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP)[39] version 6. This resulted in 4,379 annotated genes of which 4,228 were coding, 2 full rRNA gene sets consisting of 5S, 16S and 23S plus48 tRNAs, 4 ncRNAs and 93 pseudo genes. Annotated genes were split across the contigs 3,923 in the chromosome, 226 pMpar-1, 230 in pMpar-2. Of the annotated genes 702 (17.9%), 59 (26.1%) and 62 (27.0%) were hypothetical proteins respectively. Annotated gene counts differ from the MerPar_1.0 and BRCS2 genomes by the addition of 21 genes in both cases and a decrease in pseudogenes by 42 and 31 respectively. However due to changes in the version of PGAP (version 4.9 and 5.1 respectively) overall numbers are not good comparative data. Extensions to the MetPar_1.0 genome[6] include an additional *pmoCAB* cluster and 3 singleton *pmoC* genes, an additional 5S rRNA, completion of partial 16S and 23S rRNA and 3 additional tRNAs. Comparatively the genome scale model based on the 2012 *M. parvus* OBBP genome utilised 2,795 genes forming 1,326 reactions of which 380 reactions were thought to form pathways essential for biomass production[8]. Complementary to this chapter, the genome scale model publication also points out interesting genes or those unique to *M. parvus* within its close relatives[8].

3.4.1 Methylation and Restriction Enzymes

Methylation data from PacBio sequencing was interrogated and also submitted to REBASE[40], which are compared against a database of recognition sites and events for restriction enzymes, methyltransferases and methylation sensitivity. This data might be of use in other studies and in the design of improved molecular biology tools in *M. parvus*. These might avoid the use of recognition sites or ensure correct methylation of inserted DNA by selecting conjugation/transformation species or in vitro methylation. This can be of particular importance in improving plasmid transformation efficiencies[40,41]. Restriction Methylation (RM) systems defend against foreign DNA generally consisting of two parts, a methyltransferase that methylates at or around defined recognition sites, and a restriction endonuclease that cleaves unmethylated sequences of the same site. Type IV methylases however cleave only methylated sites[42]. Some methylases, not always part of an RM system, also play a role in regulating genome replication, repairing mismatched base pairs or small indels that occur during DNA synthesis, and promoting or repressing protein expression[41]. RM systems are split into 4 types with diversity of component makeup and function. Some like the methyltransferase *dcm* are considered outside the typing system and are solitary with no partner restriction enzyme. Solitary restriction endonucleases with no methyltransferase partner are known but rare[42]. Some methylations can block restriction enzymes used in molecular biology, for example *dcm* in *E. coli* is known to block ApaI, BsaI and MscI among others[40].

REBASE (Org No. 56543) identified 13 genes split into 8 clusters with 6 on the chromosome and one each on the plasmids. These were five Type II and one each of Types I, III and IV. Data shown in Supplementary Table 4. These seem, unsurprisingly, parallel to those identified by REBASE in BRCS2[7] with the addition of a Type I cluster on the chromosome and a Type IV cluster on pMpar-1 which was previously found to have no RM genes. The PacBio processing identified 12 potential recognition motifs of which REBASE selected 8. 5 of these were assigned to RM system genes, one (GAYACC) was not assigned directly but was probably linked to MpaBPORFEP. The remaining two motifs RCCGGAGTD and RCCGGAGV were considered likely the same motif but was not assigned[41]. Further research into the assignment of the RM systems might come from future discoveries in other bacteria updating the REBASE database. Alternatively heterologous expression of the potential genes or mutation studies are the accepted methods of assigning further function[41].

One gene was identified as *dcm* by PGAP (M.MpaBPORFGP, MMG94_00260). The recognition sequence identified by REABASE, GGCGCC, does not match the canonical *E. coli* site for *dcm* C^{m}CWGG. Adjacent is a *vsr* (very short repair) endonuclease MMG94_00265 which in *E. coli* cleaves one strand at T/G mismatches in C(T/G)WGG which is repaired preventing mismatches propagating into mutations. These mismatches are caused by the spontaneous deamination of 5-methylcytosineto thymine thus *vsr* offsets the side effect of *dcm* activity[43,44]. The presence of the *dcm* and *vsr*

genes together suggests the activity may be similar though the recognition site might be divergent.

The MpaBPIIP restriction system on pMpar-2 with its conjugate pair was annotated to be of the PaeR7I family and a CTCGAG recognition site was assigned to it by REBASE. PaeR71 is an isoschizomer of the common XhoI restriction enzyme.

3.4.2 Gene Transfer Agents and Mobile Genetic Elements

The OBBP-Final assembly was run through two programs looking for transposable elements and phage. PHASTER[45] looks for prophage and was run in its online implementation identifying 3 regions one placed on the chromosome and two on pMpar-2 (Table 9). PHASTER applies a score establishing completeness of the prophage regions. None of the three found areas register as complete (>90) but highly at 70, 80 and 60 respectively. PHASTER indicated regions 2 and 3 were not expected to contain phage genes only notable transposases. None of the 3 regions picked up by PHASTER overlapped with known hard to sequence regions when inspected with the error data in IGV.

The chromosomal region contained annotation of a variety of phage genes including head, tail, capsid, packaging, and portal proteins, it also contained a gene transfer agent protein. This implies instead of a phage the region codes for a Gene Transfer Agent (GTA).

GTAs are phage like bacterial horizontal gene transfer vectors commonly misidentified as prophages by automated software as they contain many of the same genes. They have been found in a number of bacteria and archaea especially α-proteobacteria[46]. GTAs evolved by repurposing phage genes for the bacteria's gain, functioning similarly to phage but packaging random parts of genomic DNA into the particle for transfer. The majority of work on GTAs has been in the photosynthetic α-proteobacteria *Rhodobacter capsulatus*; in this species there is avoidance of transferring the GTA genes themselves instead packing 4.5kb of genomic dsDNA[46,47]. Experimental methods proving GTA functionality include horizontal gene transfer of functional genes from a media emptied of bacteria, but leaving GTAs present[48]. This could be used to demonstrate GTA activity once a suitable mutant has been generated, for example a specific amino acid auxotroph. The homology annotation is not conclusive, however if corroborated this is to our knowledge the first identification of a GTA within α-proteobacterial methanotrophs. BLAST searches of the region indicate similar clusters in *M. parvus* BRCS2, *Methylocystis* sp. SB2 and SC2, *Methylocystis Rosea* BRCS1 and GW6, *Methylosinus* sp. C49 and *Methylosinus trichosporium* OB3b all with >84% coverage and >69% Identity over the 16kb region. This is followed by a decrease to <60% Cover within other α-proteobacterial species.

Collectively PHASTER regions 2 and 3 (Table 9) placed on pMpar-2 make up 14.6% of its total size. Looking to the PGAP annotation, PHASTER region 2 on pMpar-2 contained 12 annotated transposases and related genes within the

Table 9: Outputs of PHASTER Phage finder applied to the OBBP-Final assembly.

Region Length	**Completeness**	**Keywords**	**Region Position**	**Total Proteins**	**Phage Proteins**	**Hypothetical Proteins**	**Bacterial Proteins**	***att* Site**	**Best Phage Match**	**GC**
15.8Kb	Questionable (70)	lysin, tail, capsid, head, portal	Chromosome: 3,701,765-3,717,619	18	14	4	0	no	33.33%	65.72%
23.5Kb	Questionable (80)	transposase	pMpar-2: 99,010-122,526	16	8	8	0	Yes*	12.50%	58.30%
6.4Kb	Incomplete (60)	transposase	pMpar-2: 150,199-156,604	12	7	5	0	no	16.66%	63.63%
*The Proposed att sites are both TTCGCCATGCGCA										

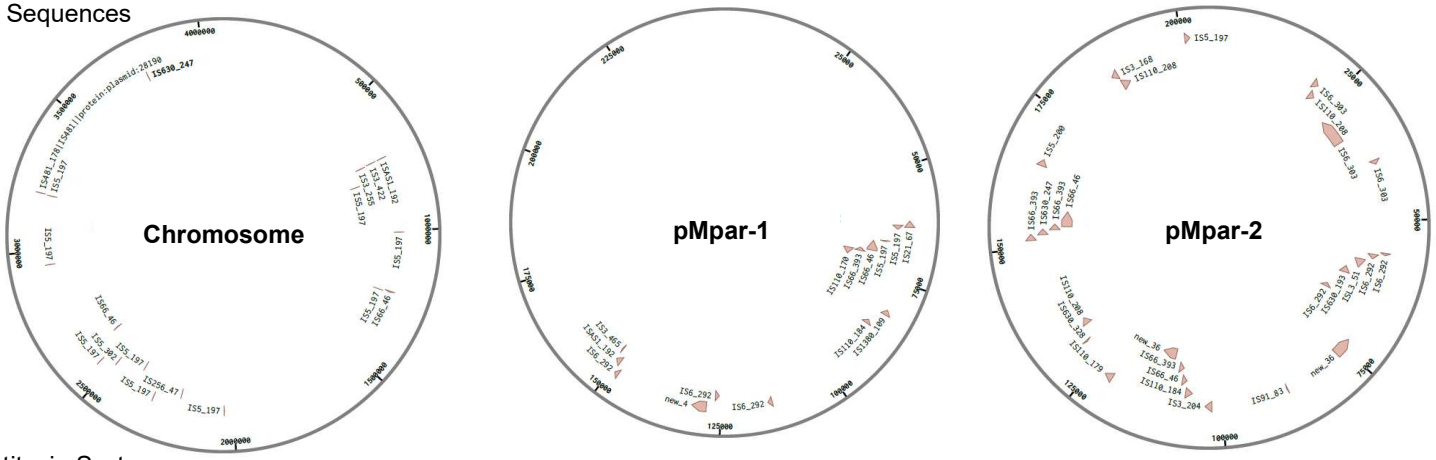
23.5kb including the specifically named *tnpB* and *tniQ* and *tniB*. Of these, 10 occur within 12kb with no intervening annotated genes. PHASTER region 3 on pMpar-2 showed 6 transposases and related genes within the 6.4kb with no other phage genes of note, corroborating findings from PHASTER. PHASTER also identified two regions labelled *attL* and *attR* flanking region 2, implying relation to an integrated tyrosine or serine recombinase phage like ΦC31 or λ-phage[49,50]. The regions are an identical 13bp, smaller than the typical 30-40bp size of recombinase sites[50]. No match to the sites could be identified manually and due to it's content of only transposases it is unlikely to be a prophage but further investigation might reveal more about an integration event. In addition to the copy of *tnpB* found in PHASTER region 2, two more copies were found in PHASTER region 3. Searching other contigs showed 2 additional copies of *tnpB* on the chromosome and 1 on pMpar-1 for a total of 6. One *tnpB* on the chromosome and two on pMpar-1 are within a 3.2kb regions with 99-100% identity to the PHASTER region 3. These three duplicated 3.2kb regions appear to be an IS66 type IS (Insertion Sequence).

tnpB is generally associated with a transposase *tnpA* within a mobile genetic element system of the types IS200/IS605 and IS607. TnpA is considered essential and sufficient for IS mobilisation and insertion and TnpB is neither. Each gene is often found alone as well as in combination, with the *tnpB* only groups being unable to further propagate. The purpose of TnpB is not totally clear. Recent work has found TnpB can act as an RNA directed nuclease in a similar fashion to CRISPR providing a defensive system. It potentially pre-dates CRISPR and may be a future DNA engineering molecular biology tool[51]. No *tnpA* genes were identified anywhere in the genome by gene annotation.

Insertion sequences (IS) are small autonomous transposable elements often flanked by inverted repeats. They generally contain a transposase among other genes but are often poorly covered in databases leading to miss annotation of their containing genes[52]. ISEScan[52] v1.7.2.3 was run to investigate this which performs gene prediction and comparison to known datasets and HMMs returning locations and typing of identified ISs. The outcome showed 59 hits including partial, when run more stringently for complete sites with flanking inversion regions this reduced to 34 hits. Data for all hits including partial were carried forward. These covered 13 of the existent 30 families of IS[53] with an additional 3 regions assigned to new unclassified families. The data show 18 IS on the chromosome, 14 on pMpar-1 and 27 on pMpar-2 making up 0.5%, 7.05% and 19.96% of their total length covered by IS regions. This corroborates and extends the large coverage of pMpar-2 with transposases found by PHASTER. Maps of IS locations are shown in Figure 12A.

Mapping the output of ISEScan identified regions alongside previous error data in IGV corelated with several Illumina zero coverage and Illumina ALE error regions, accounting for ~25% of issues. Having focused on prophage and IS finding, other mobile genetic elements may also be present in the genome including other forms of transposons, and integrons[53].

A – Insertion Sequences



B – Toxin-Antitoxin Systems

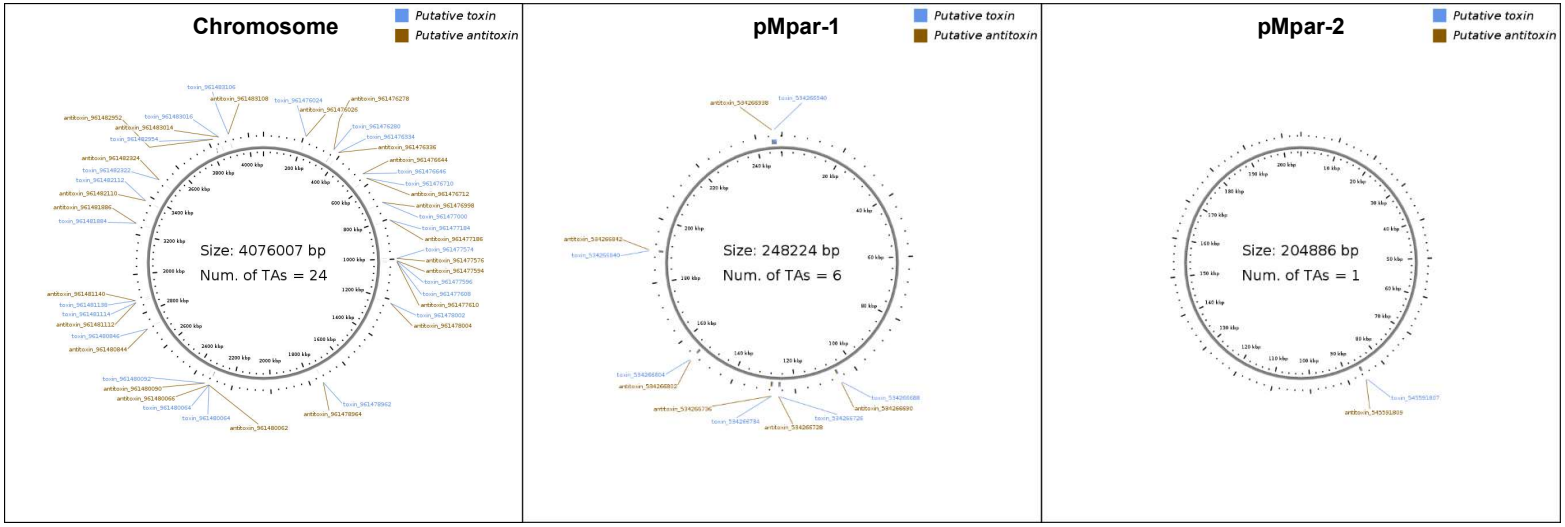


Figure 12 A and **B**: Maps of OBBP-Final showing the **A -** IS elements identified by ISEScan[52] **B** - Toxin-antitoxin system locations identified by TAfinder[54].

3.4.3 Selected Chromosomal Annotation Features

3.4.3.1 Nitrogen Fixation

33 directly annotated *nif* and *nif* associated genes coding for a molybdenum-iron (MoFe) nitrogenase and associated nitrogen fixation pathway genes were identified in two clusters. The first (MMG94_00340-00500) consisting of 25.3kb of 17 *nif* associated genes and 16 others. The second (MMG94_10865- 11020) of 25.1kb consisting of 14 *nif* associated and 19 others. Included were single copies of *nifADKQTZ* and two copies each of *nifBEHNSUVWX* and *nifU* family proteins. The other genes largely consisted of hypothetical proteins and many iron-sulphur cluster related genes and are likely still related to nitrogen fixation function. One copy of *vnfKGD* coding for a vanadium-iron (VFe) nitrogenase was identified in the second cluster[55]. One copy of *cysE* appeared in each cluster. Intervening *cysE* genes in *nif* clusters have been previously reported in the related *Rhizobium leguminosarum* and other more distant species[56].

The *nifKD* genes are located in the first large cluster and *vnfKGD* in the second, both being the subunits expected to form the nitrogenase enzyme itself[55]. This implies the two clusters are independent clusters each producing its own nitrogen fixation system. The MoFe nitrogen fixation system has been found more efficient than the VFe in terms of ATP consumption and many diazotrophs will move from the preferential MoFe to an alternative nitrogen fixation system when low on molybdate. No evidence was found by BLAST search for an *anf* Iron-Iron nitrogenase which sometimes coexists[55]. BLAST searches for other *vnf* genes for example *vnfH* often found matches to their *nifH* annotated counterpart implying interchangeability or poor differentiation in annotation between the two nitrogen fixation systems.

3.4.3.2 CRISPR

CRISPR regions were identified using the online implementation of CRISPRCasFinder[57] (https://crisprcas.i2bc.paris-saclay.fr/) identifying a set of 7 *cas* Type IC genes *cas1C, 2, 3, 4, 5C, 7C* and *8C*[58] (MMG94_07750 – 07780) and 2 CRISPR regions one with 85 spacers (1,590,999 - 1,596,604 bp) adjacent to the *cas* cluster with direct repeat consensus GTCGCGCCTTCACGGGCGCGTGGATCGAAAC and one of 53 spacers 764kb away (2,480,893 - 2,486,628bp) with repeat consensus GTCGTGCCCTTCGCGGGCGCGTGGATCGAAAC. 4 additional CRISPR regions were identified containing 1-3 spacers each with low confidence which were disregarded. An attempt to identify the PAM of this system was carried out in section 4.4.3.

3.4.3.3 Antibiotic Resistance

Genome sequences were passed through CARD[59] (Comprehensive Antibiotic Resistance Database), an online database of antibiotic resistance genes. CARD Identified three hits: a *qacJ* small multidrug resistance (SMR) antibiotic efflux pump MMG94_03960 (40.95% Identity) and two *adeF* resistance-nodulation-cell division (RND) antibiotic efflux pumps MMG94_10530 (53.46%) and MMG94_15225 (42.48%). These three features

fit the CARD "strict hit" criteria. All three annotations were also identified by PGAP. *adeF* is part of a larger complex associated with tetracycline and fluoroquinolone resistance. *qacJ* is associated with resistance to benzalkonium chloride, an antiseptic[59].

PGAP also identified other RND genes in operons alongside the *adeF*s identified by CARD suggesting complete pathways. It is quite possible the suggested targets of these genes from homology are misplaced and do not confirm immunity. *M. parvus* is thought susceptible to tetracycline but information is mixed (section 4.2.3). 219 "loose" criteria hits were also found by CARD and were not considered for brevity.

3.4.3.4 Flagella

A full set of 21 essential flagella genes as laid out by Liu and Ochman[60] was generally well annotated (MMG94_02425-08310, 08735-08820, 16535-16560) excepting a possible *motB* match MMG94_08805. In total this accounted for 32 total genes including *flaF*, *flbT*, *flgABCDEFGHIK*, *flhAB*, *fliCEFGJKLMNOPQRT* and *motAB*. This corroborates recent findings from Oshkin et al.[61] suggesting a full set of flagellar genes are unique to *M. parvus* among the α-proteobacterial methanotrophs, with other species containing some, but not complete sets. This implied motile capability disagrees with previous species definitions suggesting *Methylocystis* is non-motile while *Methylosinus* is highly motile at some stages of its life cycle[62]. Motility in *Methylocystis* has not been observed experimentally[61].

3.4.3.5 Particulate methane monooxygenase

Multiple sets and singletons of copper dependent membrane associated *pmo* particulate methane monooxygenase genes[63] were identified which perform the key enzymatic step of methanotrophy converting methane to methanol. 3 sets of *pmoCAB*, (encoding the λ, β, and α subunit respectively[63]) and 4 singleton *pmoC*s were identified. This shows an additional *pmoCAB* cluster and 3 singleton *pmoC* gene increase on the MetPar_1.0 assembly[6] and an increase of one singleton over *Methaylocystis sp.* SC2[4,64]. As expected, no *mmo* sMMO or *pxm* pXMO genes were identified. The location of the genes on the genome and into clusters is shown in Figure 13.

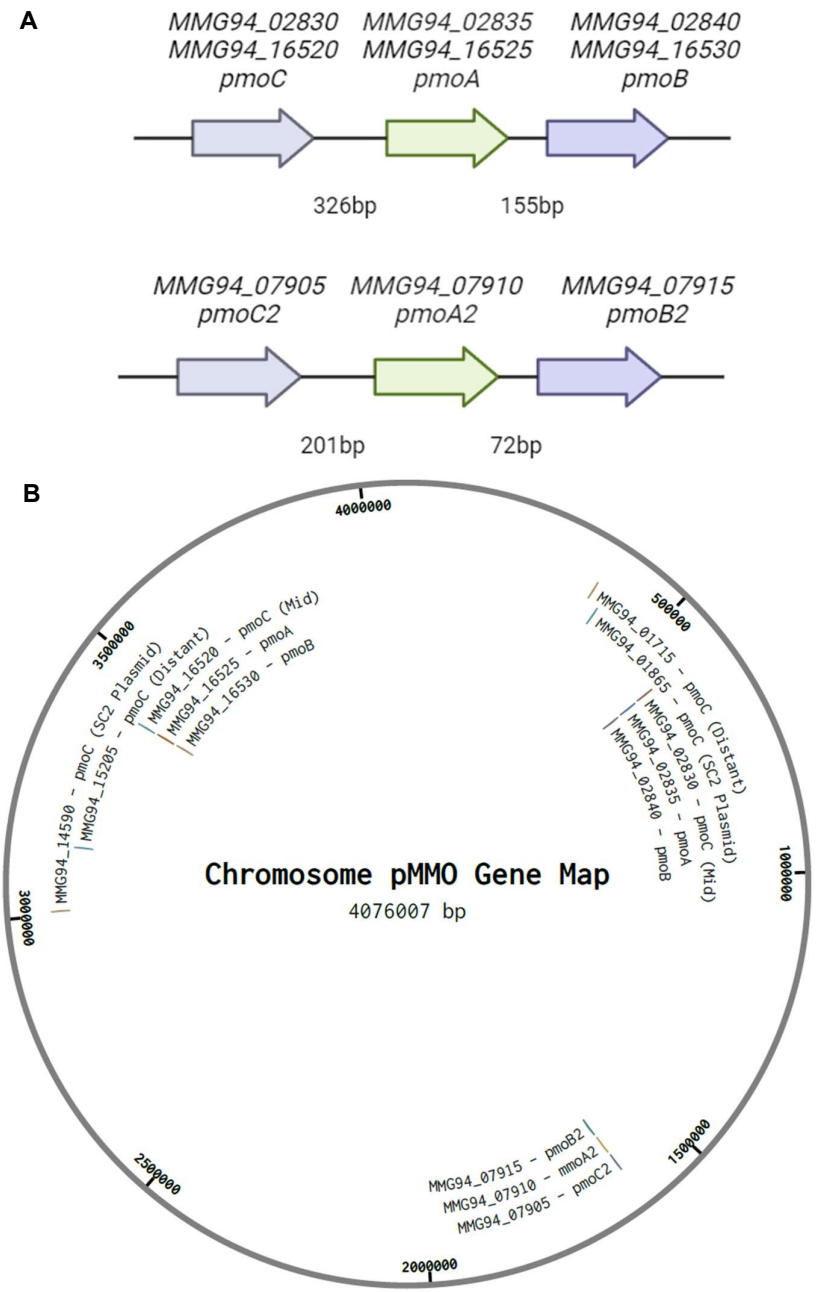


Figure 13: A and **B**: **A** – Gene Maps showing the arrangement of the three *pmoCAB* Clusters as found in OBBP-Final. This accounts for 9 of 13 identified genes. Gaps between the coding sequences are marked. The two *pmoCAB* operons had identical gaps. Figure produced with BioRender.com **B** - Chromosomal map of OBBP-Final showing the location of pMMO genes. Clusters identified in Figure 14 in brackets. Figure produced in Benchling[26].

An attempt was made to type the *pmo* genes present. Prior work in *Methylocystis sp*. SC2 identified two types of operons *pmoCAB1* and *pmoCAB2* and investigated functional differences. This found differential expression with *pmoCAB1* expressed above 600ppmv methane and *pmoCAB2* expressed below this. This is a suggested variability may allow some

methanotrophs capability to use atmospheric methane, while others are limited to higher partial pressure regions like natural gas seeps or the interface of methanogenic environments[65]. Each gene in the *pmoCAB2* operon is expected to differ compared to *pmoCAB1*[64]. The purpose of the singleton subunits has not yet been identified but they have found to be essential[66].

No difference in *pmoCAB* annotation was applied by PGAP. BLAST searching and phylogenetic tree generation for *pmoA* and *pmoB* indicated two similar copies (MMG94_02830-02840 and 16520-16530) and these two clusters were assigned as a *pmoCAB1* operons. One differing version (MMG94_07905-07915) in each case in the same cluster was assigned *pmoCAB2*.

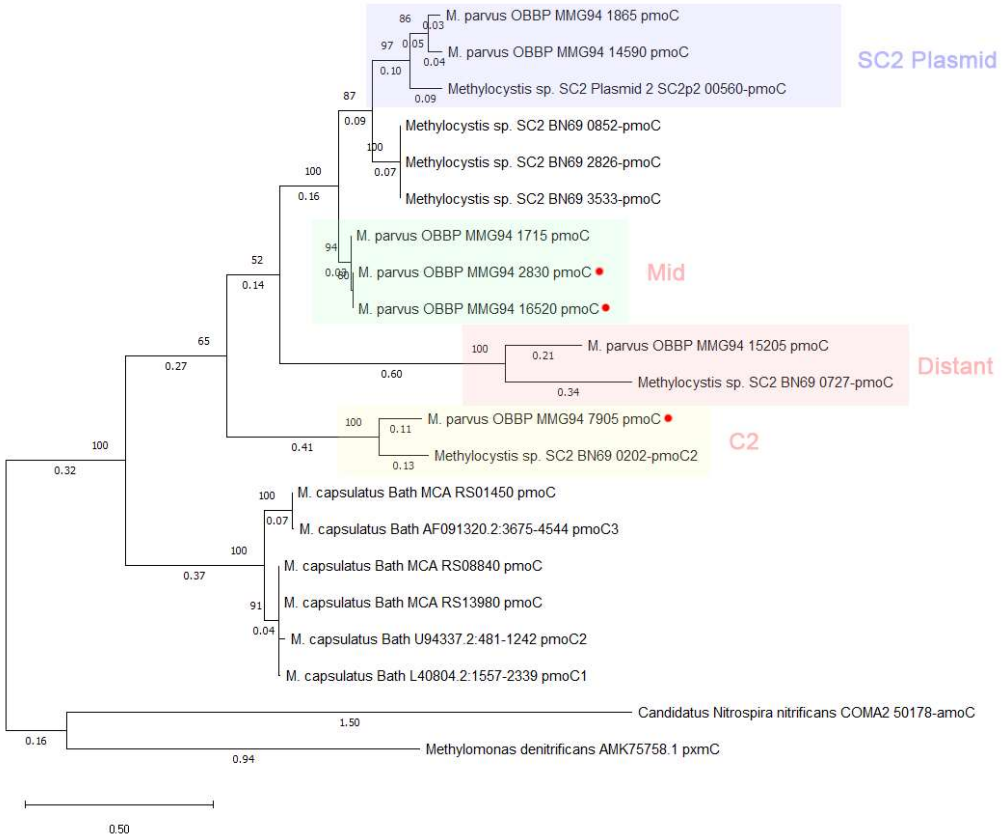
Typing the 7 total copies of *pmoC* was more challenging. A Maximum-Likelihood tree was generated incorporating all annotated *pmoC* gene genes from *Methylocystis sp*. SC2 to investigate shared similarity and aid in type assignment. The SC2 genes included 5 *pmoC1*, one being placed on the SC2 plasmid-2, and one *pmoC2*. An ammonia monooxygenase *amoC* gene from *Nitrospira nitrificans* was selected as the outgroup as *amo* is the suggested lineage from which *mmo* genes evolved[38,64,67]. An additional type of *MMO* gene *pmx*, was also included, until recently only identified in γ-proteobacteria[68] however recent pan-genome work by Oshkin et al 2020 disputes this[61].

Although work in SC2 suggests no *pmoCAB2* equivalents exist in γ-proteobacterial methanotrophs[65] Stolyar et al. 1999[69] annotated three *pmoC* copies 1,2 and 3 in *M. capsulatus* Bath and significant differences in sequence and gene essentiality were observed between types 1/2 and 3. All *pmoC* from *M. capsulatus* Bath were included and also the annotated sequenced regions produced by Stoylar et. al[69] to infer their annotation.

Sequences were gathered in amino acid form as functional rather than evolutionary information was of greatest interest. Trees were produced using MEGA[70] v11.0.11, sequences were aligned using MUSCLE with a gap open penalty of -2.9, no gap extension penalty and a 1.2 hydrophobicity multiplier. MEGAs internal model selection tool was used selecting an LG model with a Gamma distributed rate and invariant sites, the tree was confirmed with 100 bootstraps.

The resulting tree (Figure 14) assigned MMG94_07905 to *pmoC2* as expected forming part of a *pmoCAB2* operon with the type 2 gene from *M. parvus* SC2[65]. There are three other clusters of presumptive *pmoC1* genes: two copies appearing most related to the *pmoC* from the SC2 plasmid, including the additional copy not found in SC2 and three clustering more closely to the genomic SC2 pmoC1 genes two of which are part of the *pmoCAB1* operons. A more distant copy MMG94_15205 also appeared associating with a more distant version of *pmoC* from SC2. This distance might imply another functional difference as the apparent distance is as great as that from *pmoC1* main clusters to *pmoC2*, but this needs confirmation with

other species and future experimental work. The *M. capsulatus* Bath genes all clustered closely confirming the pmoC1/2 and C3 versions are not equivalent to versions found in α-proteobacterial methanotrophs. The *pxmC* associated most closely with *amoC* as expected from previous literature[68]. Tree construction and protein alignments of *pmo* genes on three α-proteobacterial methanotrophs has been performed previously using the 2012 MetPar_1.0 *M. parvus* OBBP genome assembly though this lacks additional copies identified here and incorrectly identifies only a *pmoCAB* and a *pmoAB* operon where here there are three operons of *pmoCAB*[8].



Locus	Gene	Operon	Cluster
MMG94_01715	*pmoC1*	Singleton	Mid
MMG94_01865	*pmoC1*	Singleton	SC2 Plasmid
MMG94_02830	*pmoC1*	*pmoCAB1*	Mid
MMG94_07905	*pmoC2*	*pmoCAB2*	*pmoC2*
MMG94_14590	*pmoC1*	Singleton	SC2 Plasmid
MMG94_15205	*pmoC1*	Singleton	Distant
MMG94_16520	*pmoC1*	*pmoCAB1*	Mid

Figure 14: Maximum Likelihood tree of *pmoC* genes drawn from the genome (*M. parvus* OBBP-Final) and other *pmoC* genes which were annotated with experimental evidence. Red dots indicate OBBP genes associated with a *pmoCAB* operon, others are singletons. Branches are annotated with lengths measured in substitutions per site over 304 positions. Junctions are labelled with their frequency within 100 bootstraps. Branch lengths shorter than 0.01 are not shown. Table shows conclusions drawn for *M. parvus* OBBP genes. Figure produced using MEGA[70].

The tree was also repeated with DNA sequences and no difference to structure or major changes to branch lengths was observed. The process was as above but aligned using MUSCLE in codon sensitive mode and the tree produced with a Hasegawa-Kinshino-Yano model with a Gamma distributed rate.

Annotation by PGAP did not differentiate p*mo* from the highly related possibly interchangeable a*mo* genes (as methane monooxygenase/ammonia monooxygenase subunit C) from which it is believed they evolved[38,67]. Some equivalence of function has been established where insertion of an *amo* into methylotroph *Methylobacterium extorquens* AM1 allowed subsequently grow on methane[71]. No specific *amo* genes have been identified separately in methanotrophs to date and the placement of the outgroup in the above work suggests this still to be the case in *M. parvus* OBBP.

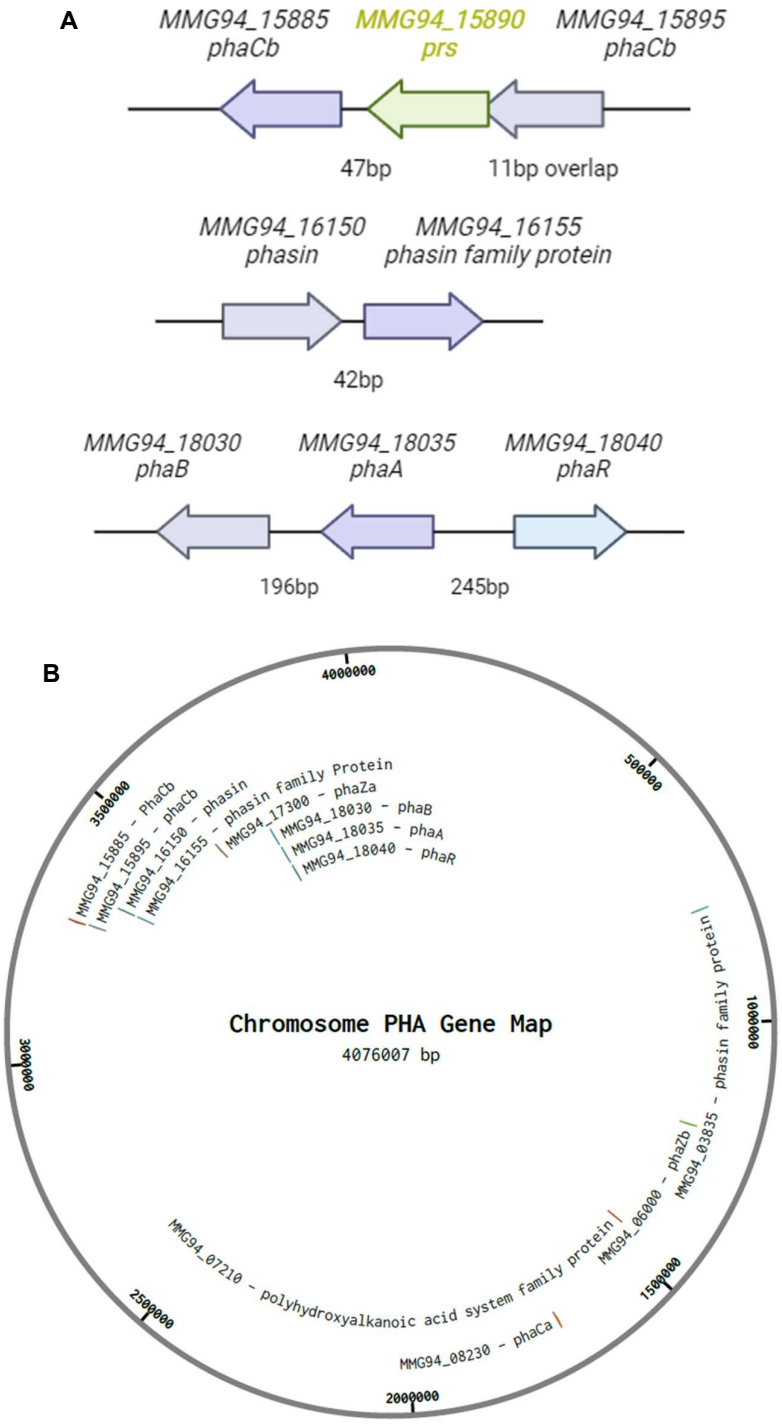
3.4.3.6 PHB Production and Degradation

The PHB production pathway from *phaABC* (alternatively *phbABC*) discussed more thoroughly in the introductory section 1.3.4.1, catalyses the production of PHB or PHAs from Acetyl-CoA. The reverse reaction, following a different pathway is initiated by PHB/PHA depolymerase *phaZ*. The pathway is regulated by the synthesis repressor *phaR* and phasins. The prior M. parvus OBBP genome release paper (MetPar_1.0)[6] identified two PHB synthases *phbCI* and *phbCII*, two depolymerases *depA* and *depB*, one *phaA*, *phaB*, *phaR* and one phasin. More details required to infer the naming scheme against the genome was not supplied so typing was carried out where required with locations supplied to aid future work.

PGAP annotation identified *phaBAR* in a group (MMG94_18030-18040, Figure 15A) with other genes distributed over the chromosome. Two *phaZ* (MMG94_06000 and 17300), one phasin (MMG94_16150) and two phasin family proteins (MMG94_03835 and MMG94_16155). One *phaC* (MMG94_08230) was identified by PGAP and two others (MMG94_15885 and MMG94_15895) were identified as alpha/beta fold hydrolases through InterPro[72] which assigns this to a *phaC* protein family group PTHR36837. BLAST searching indicated some homologues of the latter genes have been annotated as *phaC* others not. A highly divergent PHB depolymerase family esterase was not found in *M. parvus* OBBP but was found in *M. rosea* BRCS1 and GW6, *M. trichosporium* OB3b, *M. bryophillia* S285 and *M. heteri* H2; this was confirmed by BLAST search. *phaM* an activator of *phaC* in *C. necator*[73,74] was not identified by BLAST.

A “polyhydroxyalkanoic acid system family protein” MMG94_07210 was identified by PGAP, this had no matches by BLAST (discontiguous megablast) outside of *Methylocystis* by nucleotide. The annotation appears to draw from protein homology and annotation by Interpro to IPR013433. No literature origin of this group could be found but all members are stated to be within the gammaproteobacterial notably *Pseudemonas*. No further information could be

gained but this could be evidence of an HGT event and/or an unidentified regulatory element. The location of the identified PHA associated genes on the *M. parvus* chromosome are laid out in Figure 15B.



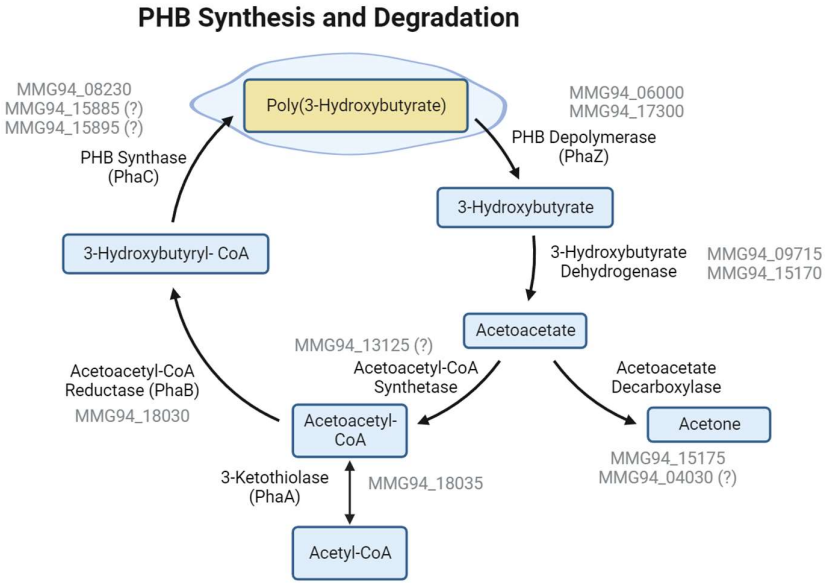


Figure 15A-C: **A** – Gene Maps showing the arrangement of genes into three PHA gene clusters cluster as found in OBBP-Final. This accounts for 7 of 12 identified genes. Gaps between the coding sequences are marked. *prs* - Ribose-phosphate diphosphokinase, not previously considered PHA associated. Figure produced with BioRender.com **B** - Chromosomal map of OBBP-Final showing the location of PHA pathway associated genes. Figure produced in Benchling[26]. **C** – an updated version of Figure 4 illustrating the PHB synthesis and degradation pathway appended with gene locations on the *M. parvus* OBBP genome. Genes with questionable attribution are annotated with "(?)". See text for more details. Figure created with BioRender.com.

To better type gene lineages, gene trees (Figure 16 A-C) were generated for *phaC*, *phaZ* and the phasin and phasin like genes. Sequences were gathered by a combination of BLAST and annotation searching to include all copies within select *Methylocystaceae spp*. And *Cupriavidus necator* the model species for PHB production. α-proteobacterialmethanotrophs appear to be within the PHB production class I group along with *C. necator*[75]. The *phaC2* mentioned by other authors in *C. necator*[76,77] could not be identified by annotation, searching in the NCBI database or from the literature, but BLAST search compared to the *M. parvus* OBBP *phaC2* identified a match in an alpha/beta fold hydrolase E6A55_10265 and this was assumed to be the correct gene. Representative *phaC* genes from class II carrying *Pseudemonas* (*phaC1* and *C2*) and class III *Rhizobium* species (*phaC3*) were selected[75]. PHB depolymerase family esterases were checked in an initial *phaZ* tree version but then excluded due to large divergence (6-7 length branches). A highly divergent PHB depolymerase from *M. rosea* BRCS1 (F7D13_04210), which was also present in *M. rosea* GW6 and *Methylocystis sp*. SB2 and SC2, was also excluded with similar branch lengths.

The *phaC* gene tree (Figure 16A) shows a cluster containing the *phaC* annotated *Methylocystaceae* genes with *Methylobacterium spp*. grouping together. This is considered the true *phaC* cluster. A secondary cluster containing *Methylocystaceae* genes annotated as alpha/beta fold hydrolase or *phaC* grouped separately. The probable *C. necator phaC2* also collocated to

this second group indicating this lineage is shared. This implies literature on PhaC2 in *C. necator* including its catalytic inactivity or unknown separate purpose may also apply here (section 1.3.4.1). As previous annotation[6,7] assumed to be referring to this secondary group as *phbCII* or *phbC2* impinges with annotation of class II or class III *phaC* which can also, subdivide into an unrelated *phaC1* and *phaC2* in *Pseudemonas*[75]. I suggest an alternative scheme naming the found groups *phaCa* and *phaCb*. The known *phaC1, C2* and *phaC3* representatives grouped separately. The *phaC1* in *C. necator* also appeared some distance from the *phaCa* and *phaCb* clusters implying some separation in this part of the lineage. This adds confusion to any inference.

The sequence of identified PhaC genes was translated and aligned to the amino acid sequence for a known crystal structure of *PhaC* in *C. necator* produced by Wittenborn et al.[78] (PDB:5T6O). This allowed identification of highly conserved residues they identified to be of particular importance. All specified residues were conserved in all three OBBP *phaC* genes covering types *phaCa* and *phaCb* suggesting both are correctly identified as *phaC* genes. Specifically, they were Cys319, Asp480 and His508 which define the active site, Arg398 which might bind CoA, His481 which is proposed to stabilise CoA and Asp421 thought to be involved with PHB chain termination. Note that PDB:5T6O has a C319A mutation made for crystallisation. Identifications used here referring to the position in the *C. necator* crystal structure.

phaZ depolymerase genes (Figure 16B) formed two clusters I named *phaZa* and *phaZb*. All *Methylocystaceae* species had one copy of *phaZa* and one of *phaZb*. It is not clear how these correlate with the *depA* and *depB* stated by del Cerro et al.[6]. Genes manually annotated in facultative methylotroph *Methylobacterium extorquens* AM1[79] as *depA* and *depB* both appear to cluster together near the *phaZa* group. *C. necator phaZ* clustered separately except for a single gene within *phaZb*.

Phasin and phasin family proteins (Figure 16C) from the methanotrophs all group separately from the 6 identified in *C. necator* H16 making inferred function from individual homologous phasins challenging. Four of the five annotated true phasins cluster together with one *M. trichosporium* OB3b CQW49_11760 grouping with the other phasin family proteins.

Some additional PHB degradation pathway genes were identified, two 3-hydroxybutyrate (3HB) dehydrogenases (MMG94_09715 and 15170) one of which is closely adjacent to an acetoacetate decarboxylase (MMG94_15175). The 3HB dehydrogenase converts the monomer into acetoacetate with release of reduced NAD[80]. The acetoacetate decarboxylase converts acetoacetate into acetone and CO_2[81] providing an alternative to the Acetoacetyl-CoA synthetase route of degradation illustrated in Figure 4. Another acetoacetate decarboxylase family protein was identified at MMG94_04030. No Acetoacetyl-CoA synthetase as illustrated by Liu et al.[82] could be readily identified but an "acyl CoA:acetate/3-ketoacid CoA transferase" named YdiF by PGAP (MMG94_13125) is a potential candidate[83]. These may be important

enzymes to downregulate in redirecting flux to greater PHB accumulation and stopping its degradation. An updated version of Figure 4 illustrating the PHB synthesis and degradation pathway with appended genome locations of genes is presented in Figure 15C.

A – *phaC* PHA Synthase

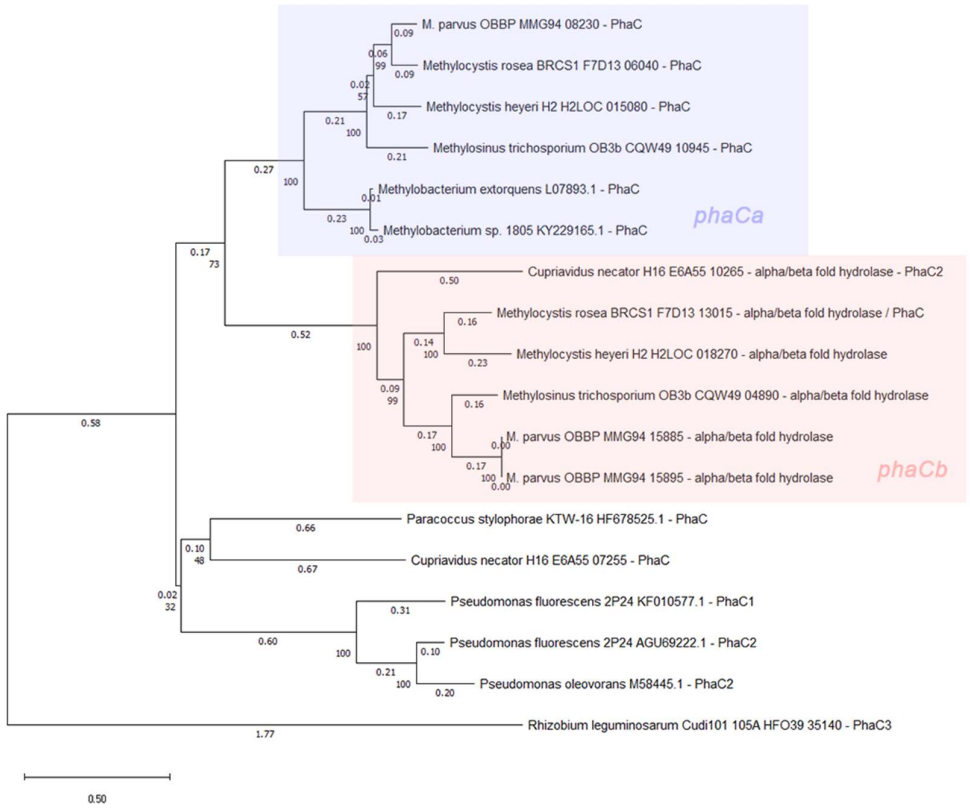
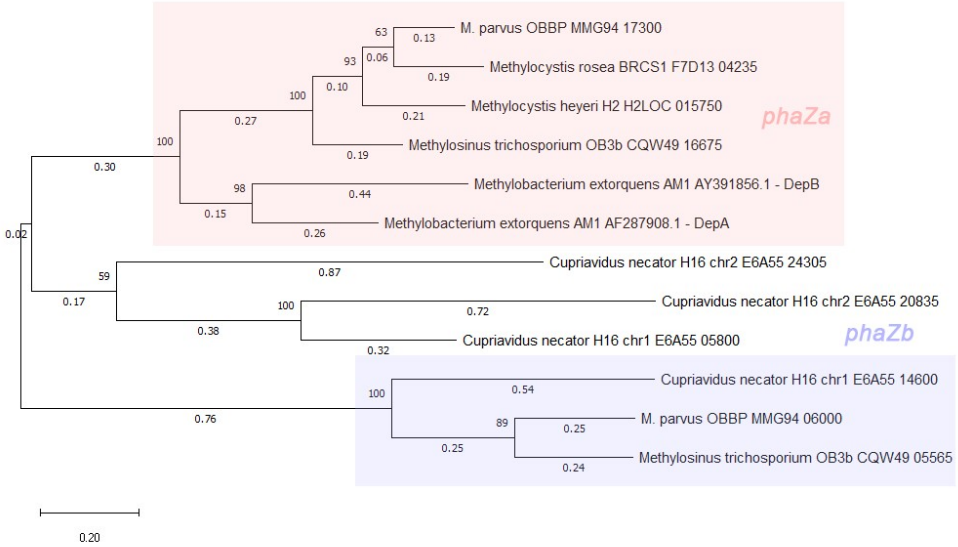
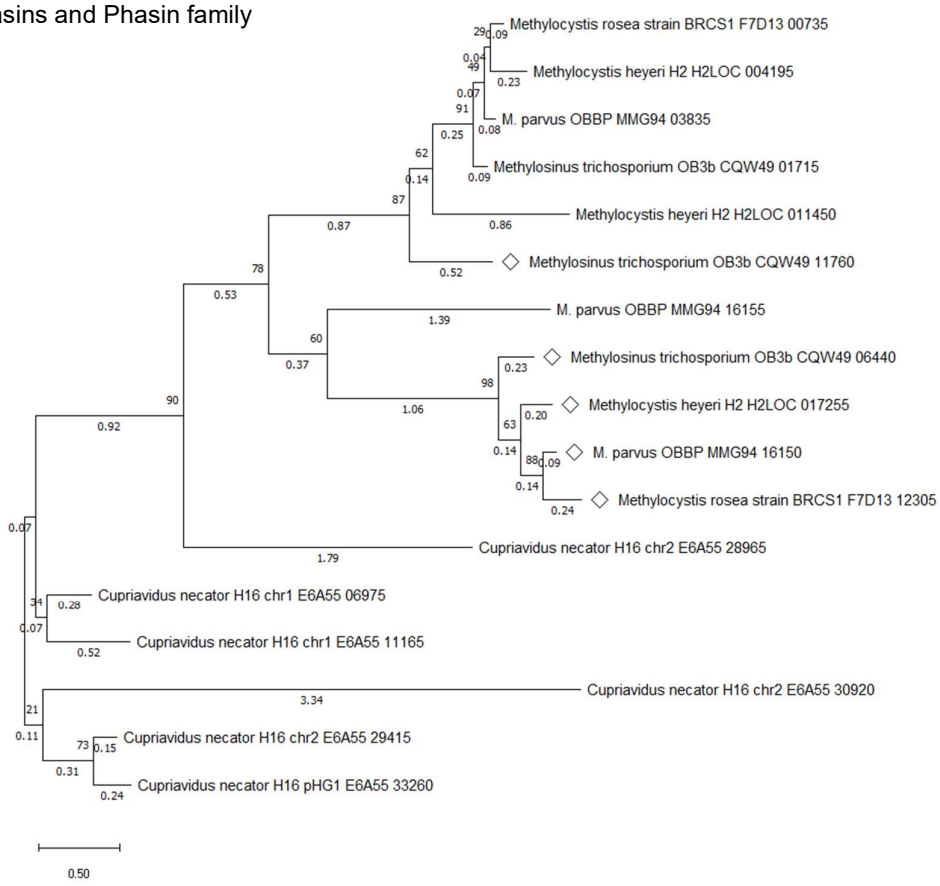


Figure 16 A-C: Maximum Likelihood trees of A – *phaC,* B – *phaZ* and C – phasin genes drawn from the genome (*M. parvus* OBBP-Final). Models used are LG+G+F, LG+G and LG+G+F respectively. In C diamonds indicate genes annotated as phasins, others were phasin family proteins. Branches are annotated with lengths measured in substitutions per site. Junctions are labelled with their frequency within 100 bootstraps. Figures produced using MEGA[70]

B – *phaZ* PHA Depolymerase



C – Phasins and Phasin family



3.4.3.7 .Toxin Antitoxin Systems

A total of 36 toxin-antitoxin (T/AT) system genes or groups, which generally code for a long lived toxin and short lived antitoxin, were identified by PGAP: 25 on the chromosome, 10 on pMpar-1 and 1 on pMpar-2. This is not an unusual number sitting between the 12 on *E. coli* K-12 and 76 in *Mycobacterium tuberculosis* H37Rv[84]. Significantly represented families include 7 RelE/ParE and 8 VapC. All of the genes belong to the most well understood and common Type II class where both components are peptides[84].

The selection found is unlikely to be complete as many toxins or antitoxins are missing their conjugate pair which would be expected to collocate in an operon. TAfinder[54] a specialist HMM and BLAST based Type II TA finder using a curated database was applied. OBBP-Final sequences were uploaded to TAfinder using their auto-annotation method. This identifies pairs and individual genes, finding 31 pairs in total with 24 on the chromosome 6 on pMpar-1 and 1 on pMpar-2. Positions are indicated in Figure 12B. Inspection indicated annotation of some genes had been misassigned by PGAP likely due to the method of action of the toxin appearing similar to a native protein e.g. transcriptional regulation. This data could be used to improve overall annotation.

Many of the plasmid borne T/AT systems are duplicated to the chromosome, at least at the family level though confirmation of equivalence using a gene tree could illuminate more thoroughly within families. Using the PGAP annotation pMpar-1 has a unique YhaV family (MMG94_19885) and PemK/MazF family (MMG94_20225). The only T/AT identified on pMpar-2 (RelE/ParE family MMG94_21215) appears to have many duplicates on the chromosome and pMpar-1.

The original discovery of T/AT systems was as plasmid addiction modules carrying out post-segregational killing upon plasmid loss during cell division where remaining toxins kill daughter cells now lacking a plasmid borne antitoxin to inactivate them. T/AT systems are often present in many copies on chromosomes[84] as seen in *M. parvus* OBBP. They are often seen in concert with mobile genetic elements. Proposed purposes for chromosomal T/AT systems include protection against phage infection and anti-addiction repelling the addictive properties of the TAs of new plasmids[84]. It has been suggested on many occasions that T/AT systems might provide regulatory or survival functions in stressful conditions however this has often been hard to reproduce and recently disputed[84,85]. Elimination of some T/AT systems may allow curing of the megaplasmids for streamlining of the genome, investigation of the megaplasmids purpose more thoroughly or reducing transcriptional and translational load[86]. Other chemical means have also been used to cure megaplasmids including those possessing T/AT system addiction for example in *Pseudemonas putida* using mitomycin C[85,87].

3.4.4 Plasmids pMpar-1 and pMpar-2 and their Annotation

In addition to the previously mentioned high IS content of the plasmids, T/AT systems and possible GTA on pMpar-2 the purpose of the megaplasmids is not fully clear and has not been elucidated in papers on other *Methylocystis* megaplasmids[66].

3.4.4.1 Plasmid Copy Number

Using alignment data shown in Table 7 plasmid copy numbers can be inferred assuming reads are drawn from the genetic material uniformly, depth of reads should indicate the ratio of each plasmids DNA to chromosomal DNA. Illumina data showed a ratio of 0.91 and 0.85 and PacBio 0.52 and 0.48 respectively for pMpar-1 and pMpar-2. These results were checked using the Illumina data for BRCS2 mapped to its assembly which showed a ratio of 0.80 and 0.73 respectively to its highly similar plasmids. This analysis was repeated using the same data with CLC Genomics[88] assembly algorithms with outcomes within 1% similarity. The Illumina data shows a roughly 1:1 ratio indicating a plasmid copy number of 1. It is not clear why the indicated copy numbers deviate below 1 especially in the PacBio data, the presence of previously identified hard to sequence areas or plasmid loss during culturing may provide part of the explanation though this is considered unlikely. This copy number of 1 corroborates work in *Methylocystis sp*. SC2 which showed its two megaplasmids were essential or at least could not be easily cured, with ratios of 1.02 and 0.95. It should be noted the SC2 plasmids are not equivalent and differ in size from *M. parvus* but share the low copy *repABC* system[66].

3.4.4.2 Plasmid Annotation

A more thorough investigation of genes present on the megaplasmids was undertaken to understand their relationship to the chromosome particularly to identify unique genes where the plasmid might provide essential or accessory processes unavailable in the chromosome.

A single *repABC* cluster was identified on each plasmid. The *repABC* plasmid cluster results in very low copy number plasmids with multiple possible incompatibility groups. The first two genes are involved in plasmid segregation and the final gene contains the origin of replication and encodes a protein required for replication[89]. Two *parS* centromere like regions were identified by the GTTNNCNGCNGNNAAC 16bp palindrome motif[89] 277-300bp 3' of *repA* on pMpar-1. No correct *parS* regions were identified on pMpar-2, with an additional mismatch 8 possible sites emerged including two 10,523bp-11,878bp 3' of *repA*, but none found were closer or correct palindromes however not all *parS* sites fit the above consensus[89].

A BLAST search of the *repABC* regions was carried out visualised using the NCBI tree view. pMpar-1 had a close cluster containing *M. parvus* BRCS1 plasmids 1, *Methylocystis sp*. SC2 plasmid 1 and 2, *M. rosea* BRCS1 plasmid 2, *M. rosea* GW6 pGW6_2. pMpar-2 had a close cluster containing *M. parvus* BRCS2 plasmid 2, *M. rosea* BRCS1 plasmid 1 and *M. rosea* GW6 pGW6_1. These associations might equate to plasmid lineage and could form the groundwork to compare evolutionary similarity of the plasmid as a whole.

Inspections for any horizontal plasmid transfers and inspection of conserved genes within a lineage might give a greater understanding of the megaplasmid function.

On pMpar-1 a particularly large 8.5kb CDS making up 3.4% of the plasmid codes for a putative glycosyl transferase (MMG94_19975) unique to the plasmid. Protein BLAST searching produced strong homology (76% Ident 99% cover) to cyclic β -1,2-glucan synthetase. Cyclic β-1,2-glucans are localised to the periplasm and are generally related to plant pathogenesis and symbiotic host interactions also providing osmoregulation and have been found essential in some flagella assembly[90,91].

Other identified loci of note on pMpar-1 included 2 cold-shock proteins (MMG94_20055 and 20090), another being present on the chromosome (MMG94_03005). A unique *nhaA* Na+/H+ antiporter (MMG94_20050) related to pH homeostasis in alkaline conditions which in *E. coli* possesses its own sensing function[92]. A *queC* locus coding for 7-cyano-7-deazaguanine synthase (MMG94_20165), in addition to another already present on the genome, which is part of queuosine production, a modified nucleoside present in some tRNAs which can improve anticodon recogntion[93]. A nitrile hydratase (MMG94_20355) alpha subunit was identified with no adjacent other subunits which might have diverse roles converting nitrile to amide groups in post translational modification or the utilisation of nitriles as a nitrogen source. The alpha subunit alone has been found to still be sufficient for catalytic activity and this may be a monomeric form[94].

No evidence of a plasmid located *pmoC* was identified like the single copy found in *Methylocystis sp.* SC2[66].

3.4.4.3 Hydrogenase

Several *hyp* hydrogenase genes were identified in a cluster at MMG94_20980- 21045 on pMpar-2. These code for a nickel iron hydrogenase which reversibly breaks diatomic hydrogen[95,96]. This feature on pMpar-2 does not appear unique with many *hyp* genes also appearing in a larger 17 gene MMG94_15085-15165 region on the chromosome. A putative full set of *hypABCDEF* chaperon and maturation proteins are present in both clusters with additional annotated related genes though the order of the cluster and additional constituents appears divergent between the two. The actual hydrogenase enzyme consists of a large and small subunit[95,96]. One large subunit was found on pMpar-2 and two copies on the chromosome while the small subunit had one copy on the chromosome and could not be found on pMpar-2. This suggests poor annotation, or the plasmid cluster is incomplete. The chromosomal copy is adjacent to the 3HB dehydrogenase and acetoacetate decarboxylase mentioned previously.

3.4.4.4 Methanobactin

Methanobactin is a fluorescent secreted copper chelating (chalkophore) peptide identified in various methanotrophs used for copper uptake. Likely common among them to service the high requirements of copper by pMMO

for methane oxidation. Methanobactin appear to have no equivalents outside the methanotrophs but is related to iron binding siderophores[97,98]. Methanobactins also promiscuously bind other transition metals including some that might otherwise be toxic like mercury which could act to sequester them reducing toxicity in bioremediation. A 6 gene methanobactin containing cluster was identified on pMpar-2 (MMG94_21255- MMG94_21280) with a set of *mbnABC* (MMG94_21260-21270) where *mbnA* codes for the methanobactin precursor itself and *mbnBC* are biosynthesis proteins. Adjacent to this was an identified di-heme enzyme and "metallo-mystery pair system four-Cys motif protein". A *matE* multi antimicrobial extrusion protein was identified just preceding the *mbn* cluster which may tentatively be involved in methanobactin export. Five of the six genes appear in the same order on the chromosome MMG94_05915- 05935 sharing 59 % overall identity (Needleman-Wunsch alignment). However, no methanobactin *mbnA* precursor was identified by PGAP on the chromosome leaving a blank region. An open reading frame in this region (1196951-1197034b) was matched manually with 87% amino acid identity to a methanobactin identified by DiSpirito et al.[98] (WP_157212416.1) filling the space. This corroborates work by DiSpirito et al. [98] which identified two groups of methanobactins, most methanotrophs only carrying one, but *M. parvus OBBP* and *Methylosinus sp.* LW3 and LW5 carry one from each group. In the case of *M. parvus* OBBP the group I (OB3b like) is on the pMpar-2 and the group II (SB2 like) is chromosomal. The gene order of the two methanobactin regions identified in the OBBP-Final genome match those described for the two groups by DiSpirito et al. and the annotated surrounding genes have been assigned *mbn* gene annotations improving on the annotation provided by PGAP[98].

3.4.4.5 TonB-dependent transporters

pMpar-1 contains a cluster of 6 genes (MMG94_19690-19715) related to TonB-dependent transport. TonB dependent transporters bind and transport vitamin B12, nickel complexes, carbohydrates and most commonly siderophores (ferric iron chelating compounds)[99,100]. Siderophores are secreted to allow iron scavenging from the environment and are non-ribosomal peptides produced independently of mRNA and are not transcribed directly from the genome[99]. The pMpar-1 cluster is annotated with a siderophore specific TonB dependent receptor indicating its function from the previous mentioned options. The TonB dependent region on pMpar-1 is not unique in its annotation with multiple being identified on the chromosome including 2 siderophore specific. Therefore, unique function of this cluster cannot be established.

Two TonB-dependent transporters are also found on pMpar-2 one of which (MMG94_21285) is adjacent to the methanobactin cluster.

3.4.4.6 Type IV Secretion Systems

Four Type IV Secretion Systems (T4SS) were identified. One on the chromosome, two on pMpar-1 and one on pMpar-2. All tentatively classed as Type IVA. Presence of at least one Type IV secretion system implies capability

of conjugation, extracellular DNA deposition/uptake or protein transfer though the last is generally only found in pathogenic bacteria[101]. The presence of a *virD4* or *traG* coupling protein near each system suggest they are all DNA transfer related[102]. These may equate to the T4SS identified on both pBSC2-1 and pBSC2-2 plasmids in *Methylocystis sp*. SC2 previously[66].

The most easily defined putative T4SS region of 12kb was identified on pMpar-2 containing *virB1-11* and *virD4* (MMG94_ 21725- 21785). It possesses the same order and gene set as the classical *Agrobacterium tumefaciens* Ti plasmid[101]. A putative *virD2* relaxase and *mobC* mobilisation protein (MMG94_21705 and 21710) were identified 2kbp away. *virB1*, *B6*, *B7*, *D2* and *D4* were identified by protein BLAST of hypothetical or ambiguously annotated proteins to manually improve the annotation. The first pMpar-1 T4SS region is a 12.7kb operon of putative *vir* genes consisting of *virB1-4, virB6*, *virB8-11 and virD4*. (MMG94_20575- 20630). These were found in order with no gap for *virB7* and in the location expected for *virB5* were two genes, one an unassignable hypothetical protein, and a second gene annotated to belong to a Type VI secretion system. It cannot yet be concluded this is a complete operon however T4SS often differ from the classical order and gene set[101]. *virB1*, *B6* and *D4* were identified by protein BLAST of hypothetical or ambiguously annotated proteins.

The second putative pMpar-1 T4SS (MMG94_19730-19825) of 25kb contained 11 genes generally annotated to the *trb* lineage (trbBCEFGJK-altLI among others) which is related to the RP4 family of conjugation systems. However, the structural proteins in the two families are often homologous with many equivalents identified resulting in mixed annotation[103]. A *traG* conjugal transfer coupling protein (MMG94_19825) was identified 14kb away.

To investigate plasmid transfer capability more thoroughly oriTfinder[104] (https://bioinfo-mml.sjtu.edu.cn/oriTfinder/) was applied to look for an *oriT* and associated conjugation genes from a database of over 1000 known *oriT*. The OBBP-Final annotated PGAP file was submitted with CDSs preidentified. In pMpar-1 no *oriT* was predicted. A 990aa relaxase was predicted from 192928- 195900bp not correlating to a PGAP predicted CDS. It is adjacent to MMG94_20520: a *mobC* family mobilization relaxosome protein, which is not sufficient for relaxosome formation itself[105], and the predicted relaxosome covers two predicted but inconclusively annotated CDS (MMG94_ 20525 and 20530). The 990aa relaxase is also annotated in the BRCS2 genome however the feature oriTfinder used (NP_387457) for inference has been depreciated potentially indicating it is unreliable and suggesting why it was not annotated in in the genome by PGAP v6.

oriTfinder confidently identified the *trb* lineage genes in pMpar-1 (MMG94_19730-19825) to be trbBCDEJKLFGI and the *traG* identified previously. *trbK* was substituted for the PGAP annotated *trbK-alt* which are not considered homologous. This covers 9 of the 10 essential *trb* genes from the Tra2 core region of the RP4 mating pair formation system, missing *trbH* but including *trbK* which was not found to be necessary. No *traF*, the other

essential part of the mating pair formation system was identified on pMpar-1 but two potential *traF* genes were identified on the chromosome noted below. None of the genes (*traHIJK*) of the RP4 Tra1 core region coding for the relaxosome, or other Tra1 genes related to DNA processing were identified in the genome[106]. oriTfinder also annotated the genes in the pMpar-1 T4SS region MMG94_20575- 20630 finding an almost complete set of *virB1-B11* and *virD4* excepting no *virB7* as identified previously and the Type VI secretory system protein (MMG94_20605) was instead annotated as *virB5*.

oriT finder found a further 14.6kb RP4 like T4SS system region on the chromosome (MMG94_12390-12450) containing *trbBCDEJLFGI* generally corroborated by PGAP annotation with the addition of *trbK-alt* at MMG94_12410 which was unannotated by oriTfinder. In addition, oriTfinder annotated a VirD4 at MMG94_12450, two nearby *traF* at MMG94_12465 and 12505 and *traG* at MMG94_02065 further away on the chromosome. None of these additions were supported by PGAP directly which annotated the putative *virD4* at MMG94_12450 as *traG*.

OriT finder identified no *oriT* in pMpar-2 either but did identify the full set at MMG94_21710-21785 noted previously and no additional regions.

The presence and extent of the four T4SS including three across the two megaplasmids suggest plasmid mobility that has been considered but not shown experimentally[66]. Multiple conjugation and replication machineries on a single plasmid theorised to broaden their host range has been observed previously[107]. Mobilisation of larger *repABC* plasmids of 400kb+ has been shown[107,108], if mobilisation is possible there might be evidence of plasmid lineage through HGT that would not align with the vertically inherited phylogenetic tree of established methanotroph (or wider) bacteria.

3.4.4.7 Gene essentiality in the chromosome, pMpar-1 and pMpar-2 plasmids

TraDIS (Transposon Directed Insertion Sequencing) data produced by Bashir Rumah (generally unpublished but preliminarily mentioned in[109]) in the highly related *M. parvus* BRCS2 used randomly inserted disrupting transposons to knock out gene function. Only surviving mutants reproduced and were sequenced giving a total of 1,648,896 unique insertion sites giving an average insert every 2.75bp proving results that should be exhaustive. The data were analysed showing 596 of 3307 (15.3%) genes on the chromosome were essential. On pMpar-1 6 genes of 224 (2.7%) were essential and one ambiguous. In pMpar-2 only 3 of 228 (1.3%) were essential. In both plasmids the *repABC* genes were all essential this making up half of the genes in pMpar-1 and all of pMpar-2. Assuming that knock out of the *repABC* genes result in elimination of the plasmid this implies the plasmids are essential despite no/few genes on them being classed as such. The additional essential genes in pMpar-1 were MMG94_20735 helix-turn-helix domain-containing protein, MMG94_20245 helix-turn-helix domain-containing protein, MMG94_20255 DNA-binding transcriptional regulator and an ambiguously essential MMG94_20005 hypothetical protein. MMG94_20735 and MMG94_20255 each correspond to the antitoxin section of two antitoxin systems implying

these systems are active. This is 2 of the 6 pMpar-1 identified T/AT systems. The one T/AT system on pMpar-2 (MMG94_21215) was not identified as essential.

Looking beyond the two megaplasmids, TraDIS data for the PHB pathway genes showed *phaZa, phaZb, phaR* and the alternate *phaCb* were not essential, while *phaA*, *B*, and *Ca* were essential. This essentiality of the core *phaABC* pathway confirms published findings that *phaCa* could not be knocked out in *M. parvus* BRCS2[109]. This contrasts to findings in other PHA producers for example *C. necator* where *phaC* was not essential[77]. None of the 3 identified phasin or phasin family proteins were found to be essential. The essentiality of *phaA* and *B* seem logical due to their usage in the first steps of the EMC regeneration pathway in addition to the PHB pathway[110].

3.5 Conclusion

A complete and ungapped genome with no ambiguous bases has been produced with high confidence for the α-proteobacterial methanotroph *Methylocystis* type strain *Methylocystis parvus* OBBP improving on the previous contig level assembly. Disappointingly no PHB production control genes beyond *phaR* were identified. This limits goals of uncoupling PHB production from nitrate or other nutrient starvation and suggests more work is needed to understand the PHB regulatory systems within *M. parvus* OBBP. A potential new lineage of *pmoC* PHA synthases (the pmoC "distant" group Figure 14) also require further investigation. Investigation of duplicate and missing genes identified by BUSCO and CheckM provide potential leads on metabolic specialisation within *Methylocystis* and α-proteobacterial methanotrophs overall.

The two megaplasmids in *M. parvus* OBBP make up 10% of the total genome but large portions consist of IS elements, T/AT systems and plasmid transfer elements and a surprising diversity of T4SS. The specific annotations of the four T4SS systems must be taken with care however the existence of systems in those locations appears well evidenced.

The established essentiality of the two single copy megaplasmids through failed attempts at curing[66] in *Methylocystis sp.* SC2 could be ascribed to essential genes present upon them or to T/AT systems causing plasmid addiction mentioned previously. A limited selection of apparently plasmid unique genes including the *nhaA* Na+/H+ antiporter and cyclic beta-1,2-glucan synthetase might indicated improved survivability in challenging pH and osmotic conditions provided by the megaplasmids but this has not been exhaustive. Future work disabling T/AT systems to improve plasmid curing will elucidate this essentiality more thoroughly.

3.6 Data Availability

The final genome has been deposited in GenBank under the accessions: CP092968 (chromosome), CP092969 (pMpar-1) and CP092970 (pMpar-2). PacBio, Illumina and Sanger sequencing data and PacBio methylation data has been deposited in the NCBI BioProject PRJNA812408.

3.7 References

1. Methanotroph Commons. Genomes | Methanotroph Commons [Internet]. 2013. Available from: http://www.methanotroph.org/wiki/genomes/
2. Kox MAR, Farhan M, Haque U, Van Alen TA, Crombie AT, Jetten MSM, et al. Complete Genome Sequence of the Aerobic Facultative Methanotroph Methylocella tundrae Strain T4. Microbiol Resour Announc. 2019 May 16;8(20).
3. Heil JR, Lynch MDJ, Cheng J, Matysiakiewicz O, D'Alessio M, Charles TC. The Completed PacBio Single-Molecule Real-Time Sequence of Methylosinus trichosporium Strain OB3b Reveals the Presence of a Third Large Plasmid. Genome Announc. 2017 Dec 1;5(49).
4. Dam B, Dam S, Kube M, Reinhardt R, Liesack W. Complete genome sequence of Methylocystis sp. strain SC2, an aerobic methanotroph with high-affinity methane oxidation potential. J Bacteriol. 2012;194(21):6008–9.
5. Stein LY, Yoon S, Semrau JD, DiSpirito AA, Crombie A, Murrell JC, et al. Genome sequence of the obligate methanotroph Methylosinus trichosporium strain OB3b. J Bacteriol. 2010 Dec;192(24):6497–8.
6. del Cerro C, García JLJM, Rojas A, Tortajada M, Ramón D, Galán B, et al. Genome sequence of the methanotrophic poly-β-hydroxybutyrate producer Methylocystis parvus OBBP. J Bacteriol. 2012 Oct 15;194(20):5709–10.
7. Rumah BL, Stead CE, Claxton Stevens BH, Minton NP, Grosse-Honebrink A, Zhang Y. Isolation and characterisation of Methylocystis spp. for poly-3-hydroxybutyrate production using waste methane feedstocks. AMB Express. 2021 Dec 6;11(6):1–13.
8. Bordel S, Rojas A, Muñoz R. Reconstruction of a Genome Scale Metabolic Model of the polyhydroxybutyrate producing methanotroph Methylocystis parvus OBBP. Microb Cell Fact. 2019;18(104):1–11.
9. Kanwar N, Blanco C, Chen IA, Seelig B. PacBio sequencing output increased through uniform and directional fivefold concatenation. Sci Rep. 2021;11(1):1–13.
10. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 2020;21(1):1–16.
11. Claxton Stevens BH, Rumah BL, Stead CE, Zhang Y. A Complete Genome of the Alphaproteobacterial Methanotroph Methylocystis parvus OBBP. Klepac-Ceraj V, editor. Microbiol Resour Announc. 2023 Apr 18;12(4).
12. Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. 2nd ed. NY: Cold Spring Harbor Laboratory Press; 1989.
13. Andrews S. FastQC. 2019.
14. Anon. Anaconda Software Distribution. Anaconda Inc.; 2021.
15. Kitson E. Simple-Circularise. 2018.

16. Hunt M, Silva N De, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: Automated circularization of genome assemblies using long sequencing reads. Genome Biol. 2015 Dec 29;16(1):1–10.
17. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Comput Biol. 2017 Jun 1;13(6):e1005595.
18. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotechnol 2012 307. 2012 Jul 1;30(7):693–700.
19. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Res. 2004 Jul 1;14(7):1394–403.
20. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS One. 2014 Nov 19;9(11):e112963.
21. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012 94. 2012 Mar 4;9(4):357–9.
22. PacificBiosciences. BAM2fastx. 2021.
23. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018 Sep 15;34(18):3094–100.
24. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021 Jan 29;10(2):1–4.
25. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011 291. 2011 Jan 1;29(1):24–6.
26. Benchling. Benchling.
27. Seemann T. samclip. 2020.
28. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013 Apr 15;29(8):1072–5.
29. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: Assessing Genomic Data Quality and Beyond. Curr Protoc. 2021 Dec 1;1(12):e323.
30. Kriventseva E V., Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, et al. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 2019 Jan 8;47(D1):D807–11.
31. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol. 2021 Sep 27;38(10):4647–54.
32. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015 Jul 1;25(7):1043–55.
33. Bowman J. The Methanotrophs — The Families Methylococcaceae and Methylocystaceae. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes. 3rd ed. New York: Springer New York; 2006. p. 266–89.
34. Clark SC, Egan R, Frazier PI, Wang Z. ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. Bioinformatics. 2013 Feb 15;29(4):435–43.
35. Bushnell B. BBMap. 2017.

36. Utturkar SM, Klingeman DM, Hurt RA, Brown SD. A case study into microbial genome assembly gap sequences and finishing strategies. Front Microbiol. 2017 Jul 18;8(JUL):1272.
37. National Center for Biotechnology Information (NCBI). NCBI Genome Browser. Bethesda (MD): National Library of Medicine (US); 2022.
38. Tamas I, Smirnova A V, He Z, Dunfield PF. The (d)evolution of methanotrophy in the Beijerinckiaceae—a comparative genomics analysis. ISME J. 2014;8(2):369–82.
39. Tatusova T, Dicuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016 Aug 19;44(14):6614–24.
40. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. 2015 Jan 28;43(D1):D298–9.
41. Beaulaurier J, Schadt EE, Fang G. Deciphering bacterial epigenomes using modern sequencing technologies. Nat Rev Genet. 2019;20(3):157–72.
42. Vasu K, Nagaraja V. Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. Microbiol Mol Biol Rev. 2013 Mar;77(1):53–72.
43. Macintyre G, Doiron KMJ, Cupples CG. The Vsr endonuclease of Escherichia coli: An efficient DNA repair enzyme and a potent mutagen. J Bacteriol. 1997;179(19):6048–52.
44. Bhagwat AS, Lieb M. Cooperation and competition in mismatch repair: Very short-patch repair and methyl-directed mismatch repair in Escherichia coli. Mol Microbiol. 2002;44(6):1421–8.
45. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016 Jul 8;44(W1):W16–21.
46. Esterman ES, Wolf YI, Kogay R, Koonin E V., Zhaxybayeva O. Evolution of DNA packaging in gene transfer agents. Virus Evol. 2021 Jan 20;7(1).
47. Kogay R, Neely TB, Birnbaum DP, Hankel CR, Shakya M, Zhaxybayeva O, et al. Machine-Learning Classification Suggests That Many Alphaproteobacterial Prophages May Instead Be Gene Transfer Agents. Genome Biol Evol. 2019 Oct 1;11(10):2941–53.
48. Stanton TB. Prophage-like gene transfer agents—Novel mechanisms of gene exchange for Methanococcus, Desulfovibrio, Brachyspira, and Rhodobacter species. Anaerobe. 2007 Apr 1;13(2):43–9.
49. Suzuki S, Yoshikawa M, Imamura D, Abe K, Eichenberger P, Sato T. Compatibility of Site-Specific Recombination Units between Mobile Genetic Elements. iScience. 2020;23(1):100805.
50. Groth AC, Calos MP. Phage integrases: Biology and applications. J Mol Biol. 2004;335(3):667–78.
51. Karvelis T, Druteika G, Bigelyte G, Budre K, Zedaveinyte R, Silanskas A, et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. Nat 2021 5997886. 2021 Oct 7;599(7886):692–6.
52. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics. 2017 Nov 1;33(21):3340–7.
53. Makałowski W, Gotea V, Pande A, Makałowska I. Transposable Elements: Classification, Identification, and Their Use As a Tool For Comparative Genomics. Methods Mol Biol. 2019;1910:177–207.
54. Xie Y, Wei Y, Shen Y, Li X, Zhou H, Tai C, et al. TADB 2.0: an updated

database of bacterial type II toxin–antitoxin loci. Nucleic Acids Res. 2018 Jan 4;46(D1):D749–53.
55. Demtröder L, Narberhaus F, Masepohl B. Coordinated regulation of nitrogen fixation and molybdate transport by molybdenum. Mol Microbiol. 2019 Jan 1;111(1):17–30.
56. Parker G, Walshaw D, O'Rourke K, Broad S, Tingey A, Poole PS, et al. Evidence for redundancy in cysteine biosynthesis in Rhizobium leguminosarum RL3841: Analysis of a cysE gene encoding serine acetyltransferase. Microbiology. 2001 Sep 1;147(9):2553–60.
57. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Eron BN′, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018 Jul 2;46(W1):W246–51.
58. Makarova KS, Koonin E V. Annotation and Classification of CRISPR-Cas Systems. Methods Mol Biol. 2015;1311:47–75.
59. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020;48(D1):D517–25.
60. Liu R, Ochman H. Stepwise formation of the bacterial flagellar system. Proc Natl Acad Sci U S A. 2007;104(17):7116–21.
61. Oshkin IY, Miroshnikov KK, Grouzdev DS, Dedysh SN. Pan-genome-based analysis as a framework for demarcating two closely related methanotroph genera methylocystis and methylosinus. Microorganisms. 2020;8(5).
62. Whittenbury R, Phillips KC, Wilkinson JF. Enrichment, Isolation and Some Properties of Methane-utilizing Bacteria. J Gen Microbiol. 1970;(61):205–18.
63. Murrell JC, Gilbert B, McDonald IR. Molecular biology and regulation of methane monooxygenase. Arch Microbiol. 2000 Apr 4;173(5):325–32.
64. Ricke P, Erkel C, Kube M, Reinhardt R, Liesack W, Yimga MT, et al. Comparative Analysis of the Conventional and Novel pmo (Particulate Methane Monooxygenase) Operons from Methylocystis Strain SC2 In addition to the conventional pmoA gene (pmoA1) encoding the active site polypeptide of particulate methane monooxygenase, a. Appl Environ Microbiol. 2004;70(5):3055–63.
65. Baani M, Liesack W. Two isozymes of particulate methane monooxygenase with different methane oxidation kinetics are found in Methylocystis sp. strain SC2. Proc Natl Acad Sci U S A. 2008 Jul 22;105(29):10203–8.
66. Dam B, Kube M, Dam S, Reinhardt R, Liesack W. Complete sequence analysis of two methanotroph-specific repabc containing plasmids from methylocystis sp. Strain SC2. Appl Environ Microbiol. 2012 Jun;78(12):4373–9.
67. Kang CS, Dunfield PF, Semrau JD. The origin of aerobic methanotrophy within the Proteobacteria. FEMS Microbiol Lett. 2019 May 1;366(9):96.
68. Tavormina PL, Orphan VJ, Kalyuzhnaya MG, Jetten MSM, Klotz MG. A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. Environ Microbiol Rep. 2011;3(1):91–100.
69. Stolyar S, Costello AM, Peeples TL, Lidstrom ME. Role of multiple gene copies in particulate methane monooxygenase activity in the methane-oxidizing bacterium Methylococcus capsulatus Bath. Microbiology. 1999 May 1;145(5):1235–44.
70. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics

Analysis Version 11. Mol Biol Evol. 2021 Jun 25;38(7):3022–7.
71. Crossman LC, Moir JWB, Enticknap JJ, Richardson DJ, Spiro S. Heterologous expression of heterotrophic nitrification genes. Microbiology. 1997 Dec 1;143(12):3775–83.
72. Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 2021 Jan 8;49(D1):D344–54.
73. Mezzolla V, D'Urso OF, Poltronieri P. Role of PhaC type I and type II enzymes during PHA biosynthesis. Polymers (Basel). 2018;10(8).
74. Bresan S, Jendrossek D. New insights into PhaM-PhaC-mediated localization of polyhydroxybutyrate granules in Ralstonia eutropha H16. Appl Environ Microbiol. 2017 Jun 1;83(12).
75. Rehm BHA. Polyester synthases: Natural catalysts for plastics. Biochem J. 2003;376(1):15–33.
76. Jendrossek D, Pfeiffer D. New insights in the formation of polyhydroxyalkanoate granules (carbonosomes) and novel functions of poly(3-hydroxybutyrate). Environ Microbiol. 2014 Aug 1;16(8):2357–73.
77. Pfeiffer D, Jendrossek D. Localization of poly(3-Hydroxybutyrate) (PHB) granule-associated proteins during PHB granule formation and identification of two new phasins, phap6 and phap7, in Ralstonia eutropha H16. J Bacteriol. 2012 Nov;194(21):5909–21.
78. Wittenborn EC, Jost M, Wei Y, Stubbe JA, Drennan CL. Structure of the catalytic domain of the class I polyhydroxybutyrate synthase from Cupriavidus necator. J Biol Chem. 2016;291(48):25264–77.
79. Korotkova N, Lidstrom ME. Connection between poly-β-hydroxybutyrate biosynthesis and growth on C1 and C2 compounds in the methylotroph Methylobacterium extorquens AM1. J Bacteriol. 2001;183(3):1038–46.
80. Takanashi M, Shiraki M, Saito T. Characterization of a novel 3-hydroxybutyrate dehydrogenase from Ralstonia pickettii T1. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol. 2009 Feb 15;95(3):249–62.
81. Highbarger LA, Gerlt JA, Kenyon GL. Mechanism of the reaction catalyzed by acetoacetate decarboxylase. Importance of lysine 116 in determining the pKa of active-site lysine 115. Biochemistry. 1996;35(1):41–6.
82. Liu L-Y, Xie G-J, Xing D-F, Liu B-F, Ding J, Ren N-Q. Biological conversion of methane to polyhydroxyalkanoates: Current advances, challenges, and perspectives. Environ Sci Ecotechnology. 2020;2:100029.
83. Gu Y, Lu H, Shao Y, Fu D, Wu J, Hu J, et al. Acetoacetyl-CoA transferase ydiF regulates the biofilm formation of avian pathogenic Escherichia coli. Res Vet Sci. 2022;153(September):144–52.
84. Fraikin N, Goormaghtigh F, Van Melderen L. Type II Toxin-Antitoxin Systems: Evolution and Revolutions. J Bacteriol. 2020 Mar 11;202(7).
85. Kusumawardhani H, van Dijk D, Hosseini R, de Winde JH. Novel toxin-antitoxin module SlvT-SlvA regulates megaplasmid stability and incites solvent tolerance in Pseudomonas putida S12. Appl Environ Microbiol. 2020 Jul 1;86(13).
86. Lau MSH, Sheng L, Zhang Y, Minton NP. Development of a Suite of Tools for Genome Editing in Parageobacillus thermoglucosidasius and Their Use to Identify the Potential of a Native Plasmid in the Generation of Stable Engineered Strains. ACS Synth Biol. 2021;10(7):1739–49.
87. Trevors JT. Plasmid curing in bacteria. FEMS Microbiol Lett. 1986;32(3–

4):149–57.
88. QIAGEN Digital Insights. CLC Genomics Workbench 20.0.4. 2020.
89. Cevallos MA, Cervantes-Rivera R, Gutiérrez-Ríos RM. The repABC plasmid family. Plasmid. 2008 Jul 1;60(1):19–37.
90. Guidolin LS, Ciocchini AE, De Iannino NI, Ugalde RA. Functional mapping of Brucella abortus cyclic β-1,2-glucan synthase: Identification of the protein domain required for cyclization. J Bacteriol. 2009;191(4):1230–8.
91. Guidolin LS, Arce-Gorvel V, Ciocchini AE, Comerci DJ, Gorvel JP. Cyclic β-glucans at the bacteria–host cells interphase: One sugar ring to rule them all. Cell Microbiol. 2018 Jun 1;20(6):e12850.
92. Krulwich TA, Sachs G, Padan E. Molecular aspects of bacterial pH sensing and homeostasis. Nat Rev Microbiol 2011 95. 2011 Apr 5;9(5):330–43.
93. Tuorto F, Lyko F. Genome recoding by tRNA modifications. Open Biol. 2016;6(12):160287.
94. Nelp MT, Astashkin A V., Breci LA, McCarty RM, Bandarian V. The Alpha Subunit of Nitrile Hydratase Is Sufficientfor Catalytic Activity and Post-Translational Modification. Biochemistry. 2014 Jun 24;53(24):3990.
95. Chan KH, Lee KM, Wong KB. Interaction between hydrogenase maturation factors HypA and HypB is required for [NiFe]-hydrogenase maturation. PLoS One. 2012;7(2).
96. Vignais PM, Billoud B. Occurrence, Classification, and Biological Function of Hydrogenases: An Overview. Chem Rev. 2007 Oct;107(10):4206–72.
97. Matsen JB, Yang S, Stein LY, Beck D, Kalyuzhnaya MG. Global molecular analyses of methane metabolism in Methanotrophic alphaproteobacterium, methylosinus trichosporium OB3b. Part I: Transcriptomic study. Front Microbiol. 2013;4(APR):1–16.
98. DiSpirito AA, Semrau JD, Murrell JC, Gallagher WH, Dennison C, Vuilleumier S. Methanobactin and the Link between Copper and Bacterial Methane Oxidation. Microbiol Mol Biol Rev. 2016;80(2):387–409.
99. Carroll CS, Moore MM. Ironing out siderophore biosynthesis: a review of non-ribosomal peptide synthetase (NRPS)-independent siderophore synthetases. Crit Rev Biochem Mol Biol. 2018 Jul 4;53(4):356–81.
100. Noinaj N, Guillier M, Barnard TJ, Buchanan SK. TonB-Dependent Transporters: Regulation, Structure, and Function. http://dx.doi.org.nottingham.idm.oclc.org/101146/annurev.micro112408134247 . 2010 Sep 29;64:43–60.
101. Wallden K, Rivera-Calzada A, Waksman G. Type IV secretion systems: versatility and diversity in function. Cell Microbiol. 2010 Sep 1;12(9):1203–12.
102. Llosa M, Gomis-Rüth FX, Coll M, De la Cruz F. Bacterial conjugation: A two-step mechanism for DNA transport. Mol Microbiol. 2002;45(1):1–8.
103. Virolle C, Goldlust K, Djermoun S, Bigot S, Lesterlin C. Plasmid transfer by conjugation in gram-negative bacteria: From the cellular to the community level. Genes (Basel). 2020;11(11):1–33.
104. Li X, Xie Y, Liu M, Tai C, Sun J, Deng Z, et al. oriTfinder: a web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. Nucleic Acids Res. 2018 Jul 2;46(W1):W229–34.
105. Zhang S, Meyer R. The relaxosome protein MobC promotes conjugal plasmid mobilization by extending DNA strand separation to the nick site at the origin of transfer. Mol Microbiol. 1997 Aug 1;25(3):509–16.
106. Grahn AM, Haase J, Bamford DH, Lanka E. Components of the RP4

conjugative transfer apparatus form an envelope structure bridging inner and outer membranes of donor cells: Implications for related macromolecule transport systems. J Bacteriol. 2000;182(6):1564–74.
107. Hall JPJ, Botelho J, Cazares A, Baltrus DA. What makes a megaplasmid? Philos Trans R Soc B. 2022;377(1842).
108. Cazares A, Moore MP, Hall JPJ, Wright LL, Grimes M, Emond-Rhéault JG, et al. A megaplasmid family driving dissemination of multidrug resistance in Pseudomonas. Nat Commun. 2020;11(1):1–13.
109. Rumah BL, Claxton Stevens BH, Yeboah JE, Stead CE, Harding EL, Minton NP, et al. In Vivo Genome Editing in Type I and II Methanotrophs Using a CRISPR/Cas9 System. ACS Synth Biol. 2023 Feb 17;12(2):544–54.
110. Kalia VC, Lal S, Cheema S. Insight in to the phylogeny of polyhydroxyalkanoate biosynthesis: Horizontal gene transfer. Gene. 2007;389(1):19–26.

Supplementary Table 1: Methanotroph Genomes Present in the NCBI Database and Their Levels of Completion

Searches - Genus	Species Represented	Strains	Complete	Chromosome	Scaffold	Contig	Phylum	Family	Type
Methylocystis	6 + Sp	40	7		17	16	α-proteobacteria	*Methylocystaceae*	II
Methylosinus	2 + Sp	14	2		3	9	α-proteobacteria	*Methylocystaceae*	II
Methylocapsa	3 + Sp	6			5	1	α-proteobacteria	*Beijerinckiaceae*	II
Methylocella	2	4	2			2	α-proteobacteria	*Beijerinckiaceae*	II
Methyloferula	1	1				1	α-proteobacteria	*Beijerinckiaceae*	II
Beijerinckia	2 + Sp	4	1		2	1	α-proteobacteria	*Beijerinckiaceae*	II
Methylovirgula	1+ Sp	3	1		2		α-proteobacteria	*Beijerinckiaceae*	II
Methylococcus	1 + Sp	15	10		4	5	γ-proteobacteria	*Methylococcaceae*	X
Methylocaldum	2 + Sp	6	1		2	3	γ-proteobacteria	*Methylococcaceae*	X
Methylobacter	3 + Sp + 1 candidate	25			12	13	γ-proteobacteria	*Methylococcaceae*	I
Methylohalobius	1	1				1	γ-proteobacteria	*Methylococcaceae*	I
Methylomicrobium	5 + Sp	8	3	1	1	3	γ-proteobacteria	*Methylococcaceae*	I
Methylomonas	4 + Sp	27	6		3	18	γ-proteobacteria	*Methylococcaceae*	I
Methylosoma	0	0					γ-proteobacteria	*Methylococcaceae*	I
Methylosarcina	2	2			1	1	γ-proteobacteria	*Methylococcaceae*	I
Methylosphaera	0	0					γ-proteobacteria	*Methylococcaceae*	I
Methylothermus	0	0					γ-proteobacteria	*Methylococcaceae*	I
Crenothrix	1 + Sp	3				3	γ-proteobacteria	*Methylococcaceae/ Crenotrichaceae*	I
Clonothrix	0	0					γ-proteobacteria	*Methylococcaceae*	I
Methylacidiphilum	3 + Sp	11	3		6	2	Verrucomicrobia	*Methylacidiphilaceae*	
Verrucomicrobium	1 + Sp	6		2	1	3	Verrucomicrobia	*Verrucomicrobiaceae*	
Methylomirabilis	3 + sp	10	1		2	7	NC10		
Additional Genera from searching NCBI genomes by Family									
Hansschlegelia	1 + Sp	2			1	1	α-proteobacteria	*Methylocystaceae*	II
Methylopila	Sp	4				4	α-proteobacteria	*Methylocystaceae*	II
Oharaeibacter	1	2				2	α-proteobacteria	*Methylocystaceae*	II
Pleomorphomonas	4 + Sp	7		1	3	3	α-proteobacteria	*Methylocystaceae*	II
Terasakiella	1 + Sp + 1 candidate	3	1			2	α-proteobacteria	*Methylocystaceae*	II
Rhodoblastus	2	5			2	3	α-proteobacteria	*Beijerinckiaceae*	II

Searches - Genus	Species Represented	Strains	Complete	Chromosome	Scaffold	Contig	Phylum	Family	Type
Methylomagnum	1	1				1	γ-proteobacteria	*Methylococcaceae*	I or X
Candidate Methylospira	1	1	1				γ-proteobacteria	*Methylococcaceae*	I or X
Methylicorpusculum	1	1				1	γ-proteobacteria	*Methylococcaceae*	I or X
Methylocucumis (oryzae)	1	1				1	γ-proteobacteria	*Methylococcaceae*	I or X
Methylogaea (oryzae)	1	1				1	γ-proteobacteria	*Methylococcaceae*	I or X
Methyloglobulus	1 + Sp	2				2	γ-proteobacteria	*Methylococcaceae*	I or X
Methylomarinum	1	1				1	γ-proteobacteria	*Methylococcaceae*	I or X
Methyloprofundus	1	1				1	γ-proteobacteria	*Methylococcaceae*	I or X
Methyloterricola (oryzae)	1	1			1		γ-proteobacteria	*Methylococcaceae*	I or X
Methylotetracoccus (oryzae)	1	1				1	γ-proteobacteria	*Methylococcaceae*	I or X
Methylovulum	2 + Sp	4	1		2	1	γ-proteobacteria	*Methylococcaceae*	I or X
Total		224	40	4	70	114			

Sp - members of undefined species. *Methylocystis* and *Methylococcus* are in bold as the type Genus for Alpha and Gammaproteobacterial methanotrophs respectively. Grey bands separate Changes in phylum or type for clarity.Data from NCBI Genome Browser[37].

Supplementary Table 2: List of software and version numbers used through Linux with notable dependencies allowing those to be listed alongside.

Main Program											
Circlator	**1.5.5**	**Unicycler**	**0.4.9**	**Pilon**	**1.24**	**BUSCO**	**5.3.0**	**CheckM**	**1.1.3**	**BBMap**	**38.93**
Major Dependencies											
BWA	0.7.17	SPAdes	3.13.0	OpenJDK	10.0.2	hmmsearch	3.1	HMMer	3.3.2	OpenJDK	11.0.9.1
Canu	2.1	Racon	1.4.20			Prodigal	2.6.3	Prodigal	2.6.3		
Nucmer	3.1	BLAST	2.5.0					pplacer	1.1.alpha19		
Prodigal	2.6.3	Bowtie2	2.4.2					Numpy	1.22.2		
SAMtools	1.11	SAMtools	1.11					Pysam	0.16.0.1	**Other Programs**	
SPAdes	3.13.0	OpenJDK	10.0.2					Matplolib	3.5.1	**Samclip**	0.4.0
Python	3.8.5	Pilon	1.24							**ALE**	0.9
Openpyxl	3.0.5										
Pyfastaq	3.17.0										
Pymummer	0.11.0										
Pysam	0.16.0.1										

Supplementary Table 3: Missing and Duplicated Genes Identified by BUSCO and CheckM Listed are BUSCO/OrthoDB ID or TIGRFAM that can be looked up in their respective databases for more information

BUSCO v5.3.0 alphaproteobacteria_odb10		**BUSCO v5.3.0** rhizobiales_odb10		**BUSCO v5.3.0** rhizobiales_odb10	
Missing		**Missing**		**Duplicated**	
182731at28211	DNA recombination RmuC	124366at356	holo-[acyl-carrier-protein] synthase	8940at356	ATP-dependent zinc metalloprotease FtsH
276656at28211	holo-[acyl-carrier-protein] synthase	141504at356	Ubiquinone biosynthesis accessory factor UbiK	21695at356	NADH-quinone oxidoreductase subunit D
BUSCO v5.3.0 alphaproteobacteria_odb10		144882at356	DNA gyrase inhibitor YacG	30115at356	Citrate synthase
Duplicated		19628at356	Folate-dependent ribothymidyl synthase	68602at356	RNA polymerase factor sigma-32
20198at28211	ATP-dependent zinc metalloprotease FtsH	35941at356	peptidase M16	86198at356	DNA-3-methyladenine glycosylase
69911at28211	NADH-quinone oxidoreductase subunit D	42874at356	LysR family transcriptional regulator	126093at356	Cytochrome c-type biogenesis protein
CheckM Duplicated		67663at356	Permease LptG/LptF-related export ATP transporter	134388at356	integration host factor subunit alpha
TIGR01798		Citrate synthase	Biotin--acetyl-CoA-carboxylase ligase		
TIGR00624	DNA-3-methyladenine glycosylase I	78242at356	Signal transduction response regulator, C-terminal effector CheY family		
TIGR00558	pyridoxamine 5'-phosphate oxidase	87379at356	Ubiquinone biosynthesis hydroxylase UbiH/COQ6		

Supplementary Table 4: Compiled data from REBASE showing identified DNA methyltransferases and restriction enzymes.

Type	Gene	Name	Predicted Recognition Seq	Methylation Type	Type (detailed)	Coordinates	Locus
Chromosome							
II	M	M.MpaBPORFGP	GGCGCC	m5C	II	50317-51582	MMG94_00260
II	V	V.MpaBPORFGP			II	51587-52000	MMG94_00265
II	M	M.MpaBPIV	GTAC	m4C	II alpha	709117-710340	MMG94_03480
II	RM	MpaBPORFEP	GAYACC *	m6A	II G,alpha	2052616-2055414 c	MMG94_09980
II	M	M.MpaBPIII	GANTC	m6A	II beta	2761142-2762431 c	MMG94_13370
III	R	MpaBPORFCP			III	3802438-3805467 c	MMG94_18360
III	M	M.MpaBPORFCP		m6A	III beta	3805478-3807481 c	MMG94_18365
IV	R	MpaBPMrrP			IV Methyl-directed (fragment)	3881763-3882299 c	MMG94_18730
pMpar-1							
I	M	M.MpaBPI	GATCNNNNNCTC	m6A	I gamma	233487-235055 c	MMG94_20695
I	S	S.MpaBPI	GATCNNNNNCTC		I	231402-232544 c	MMG94_20685
I	R	MpaBPIP	GATCNNNNNCTC		I	226668-229871 c	MMG94_20670
pMpar-2							
II	R	MpaBPIIP	CTCGAG		II P	195660-196424 c	MMG94_21855
II	M	M.MpaBPII	CTCGAG	m6A	II gamma	196414-198165 c	MMG94_21860

*GAYACC was assigned tentatively. Methylation sites of recognition sequence underlined where available. A RCCGGAGTD and RCCGGAGV site was considered likely indicative of the same site but could not be assigned to an enzyme. Source data deposited on REBASE under Org No. 56543 Reference 34031.

Chapter 4: Developing CRISPR and ntvCRISPR for genome editing in *M. parvus* OBBP

4.1 Introduction

With the establishment of a genome scale model for *M. parvus* OBBP[1] (based on the 2012 fragmented version of the genome[2]) rational editing for strain development becomes an enticing target. Genome scale models are also available for α-proteobacterial methanotrophs *M. hirsuta*, *Methylocystis sp.* SC2 and SB2[3] and the γ-proteobacterial *M. capsulatus* Bath[4], *Methylomicrobium buryatense*[5] and *Methylomicrobium alcaliphilum*[6]. Development of a directed CRISPR(Clustered Regularly Interspaced Short Palindromic Repeats)-Cas(CRISPR-associated) protein mediated cleavage and homology repair system to make genetic edits within methanotrophs will improving efficiency and decrease the time taken compared to current techniques. This will improve the capability of methanotroph species as industrial platform organisms.

The targeting of CRISPR-Cas systems, and gene editing as a whole in *M. parvus* OBBP builds on the completed genome laid out in the previous Chapter 3. This raised a number of potential editing targets. This included the presence of T/AT systems on the plasmids alongside little to no essential genes (by unpublished TraDIS) suggesting the potential for plasmid KO. As collectively the plasmids make up 10% of the genome reducing the genome size by their removal will reduce metabolic load and nutrient requirements. If the resulting bacteria were less stable or resilient this would also prove an interesting point of study.

The overexpression or duplication of PHB genes and phasins to increase PHB production, or elimination of *phaZ* to stop PHB degradation is an enticing target for the present work. Optimising metabolic flux and co-factor generation like NADH and ATP may also be required to stop them becoming limiting factors in a fully engineered PHB production system but selective knockout will build towards that future goal. Conversely elimination of PHB production has been suggested to free up metabolic load especially acetyl-CoA to be redirected into engineered pathways[7]. However, the core PHB pathway has so far been found essential (section 3.4.3.6)[8]. Nitrogenase and hydrogenase elimination has been suggested to stop needless nitrogen fixation and wasted H_2 production increasing NADH availability. Acetate kinase and lactate dehydrogenase have also been suggested to eliminate organic acid production[7]. In addition to improving upon pre-existing methanotroph production targets like ectoine, PHB and methanol, the possible engineered pathway products utilised so far and of potential interest in the future are diverse and better reviewed elsewhere[7,9,10].

The relevance of discussing tools for α-proteobacterial methanotrophs like *M. parvus* and γ-proteobacterial methanotrophs like *M. capsulatus* Bath

collectively is in my view questionable. Methanotrophs form a polyphyletic group and the similarity of methanotroph metabolism would be expected to have little impact on genetic tool design. Thus, focus here is on tools used in α-proteobacterial methanotrophs. Nevertheless, more flexible tools that work across a range of methanotrophs from both classes would be a boon to research in the area allowing editing to be performed in parallel for comparison.

Within α-proteobacterial methanotrophs *Methylosinus trichosporium* OB3b has generally been the focus of genetic tool testing and development with some work in *Methylocystis sp*. SC2[11] with no editing performed in *M. parvus* OBBP so tentative parallels are drawn where necessary. That makes the work described here and in the accompanying paper[8] the first genetic editing carried out in *M. parvus* OBBP.

Part of the work in this chapter was published as part of "In Vivo Genome Editing in Type I and II Methanotrophs Using a CRISPR/Cas9 System" Rumah et al. 2023[8]. A full copy of this paper is presented at the end of this thesis. Only work completed by me is described here except where noted. The system utilised in this chapter was developed by co-author Bashir Rumah[8] and more details of its development are available in the published paper. As part of this publication, I assisted with conjugation, assays and colony counting for the main system development and eYFP fluorescent assays not described in this chapter.

Additionally for the publication I designed, assembled and tested 15 new seeds which is described in Part 1 of this chapter (4.4.1) examining the influence of seed design on editing success and demonstrating the first genetic edits published in *M. parvus* OBBP.

In Part 2 (4.4.2) an array of further CRISPR knockouts were attempted targeting amino acid production genes to develop an auxotrophic strain, disrupt the PHB production pathway and eliminate one or both of the megaplasmids by targeting T/AT and plasmid replication genes. In Part 3 (4.4.3) a plasmid invasion assay was performed trying to identify the PAM of the endogenous CRISPR system in *M. parvus*. Part 2 and Part 3 are not part of the publication.

4.2 Background

4.2.1 Introduction of DNA into Methanotroph Cells

A challenge universal to gene editing techniques is the method DNA, either as linear fragments or plasmids, are introduced to the cell. For *E. coli* this can be achieved by conjugation, electroporation or chemical transformation and heat shock. In methanotrophs conjugation and electroporation may be available depending on strain. Of these electroporation is preferable due to the decrease in steps and labour involved[7,9].

Although electroporation has been achieved previously in γ-proteobacterial methanotrophs, steps in achieving electroporation have generally utilised the copper sMMO/pMMO switch to move to sMMO production and reduce internal membrane complexity[9,12]. In *M. parvus* which does not possess sMMO this is not an option. *Methylocella silvestris* has been successfully electroporated with plasmids and linear DNA and although an α-proteobacterial methanotroph it does not possess pMMO[13]. Electroporation has successfully been carried out in the α-proteobacteria *Methylocystis sp.* SC2 indicating pMMO is not an entirely limiting factor[11]. Work in a PhD thesis has attempted electroporation of three α-proteobacterial methanotrophs: *Methylocystis rosea* BRCS1 (as Isolate 6), *M. parvus* BRCS2 (Isolate 12) and another probably *M. rosea* (Isolate 3*) based on the SC2[11] and γ-proteobacterial[13] methods and had no success[14]. Other reviews have noted electroporation implementation reliability challenges in many labs even in strains where it has been previously demonstrated[10].

Due to these challenges conjugation has been the most popular method of introducing vectors[7,13]. Most cited methods refer back to Martin and Murrell 1995[15] using RP4 based mobilisation and *E. coli* S17-1 λpir as the donor cell. A method based on this was used throughout this thesis. Triparental mating using an additional strain with a helper plasmid like pRK2013 or *E. coli* HB101 is also possible but adds complexity compared to the biparental approach[7,16].

While some species have DNA uptake issues due to restriction modification system activity which may require methylation or gene knockout, this has not been a noted problem in α-proteobacterial methanotrophs with conjugation[9]. It has been attributed a potential factor in the failure of electroporation however where *E. coli* grown plasmids may not electroporate but methanotroph sourced plasmids will[10].

4.2.2 Plasmid Vectors and Expression Systems in Methanotrophs

RP4 *oriT* conjugation vectors with the *Tra* initiation proteins and a suitable *E. coli oriV* work with the S17-1 λpir system and this is sufficient for suicide vector conjugation. For example the pHM32m suicide vector for marker exchange mutagenesis was demonstrated effective in 1995, replicating in *E. coli* and conjugating into methanotrophs where they are then unable to replicate[15]. Similar plasmid series include pBR329*mob*, pK18*mob*[17] and pCM184[9] the latter of which contains *loxP* sequences flanking the insert site[7]. Where electroporation is possible vectors without RP4 *oriT* can be used for example pUC18[11].

For gene expression and CRISPR a more stable replicating vector is required with the capability to replicate within the methanotroph as well as the editing and conjugation donor *E. coli*. Along with a requirement of RP4 *oriT* for conjugation, an *oriV* effective in methanotrophs must be included. The following *oriV* have been shown effective in all tested α- and γ-proteobacterial methanotrophs: pBBR, IncP group RK2[18], and IncQ group RSF1010[7,18]. These *oriV* also all serve for replication in *E. coli*.

An inducible promotor has been achieved using the tetracycline promotor/operator in *M. buryatense*[19] and *M. capsulatus* Bath[20]. Phenol and benzoate inducible promotors have also been demonstrated in *M. capsulatus* Bath[16]. Although constitutive α-proteobacterial methanotroph expression systems have been available for some time[18,21] the lack of inducible systems limited their application, relying on tuning promotors and ribosomal binding sites to control expression[22]. In 2023 the phenol inducible system has now been demonstrated in the α-proteobacterial *M. trichosporium* OB3b[16].

A survey of nine constitutive promotors in *M. parvus* OBBP have been carried out using eYFP[8]. Various other reporters including LacZ, XyleE, GFP and dTomato have also been utilised in methanotrophs though concerns have been raised over the impact of internal membranes on the performance of fluorescent promotors and it has been suggested cell-free extracts may be required for reliability[7,9,20].

More complete lists and examples of utilised plasmids are available at Methanotroph Commons[23], Khmelenina et al.[9] and Kalyuzhnaya et al.[7].

4.2.3 Antibiotic Selection in Methanotrophs

Nalidixic acid (NDX) is commonly resisted by methanotrophs at 25μg/ml[7] and was utilised in the original implementation of our conjugation method in *M. trichosporium* OB3b to select against donor *E. coli*[15]. Rifamycin resistant methanotrophs strains have also been selected for and used for counterselection[7,24]. Ampicillin/AmpR has been noted unsuitable as a selective marker for methanotroph use due to its instability over the 2-4 weeks required for methanotroph plate incubation[15]. Streptomycin has also been used as a selectable marker at least once in *M. trichosporium* OB3b[18] but it was found resistant to chloramphenicol[15]. Methanotroph general susceptibility to kanamycin and successful expression of its resistance gene *kanR* has made this a commonly utilised selectable marker in various methanotrophs[15] and was also utilised in this work.

Comparative antibiotic resistance data for *M. parvus* OBBP is not available in paper form to the best of our knowledge. A PhD thesis[14] investigated this by antibiotic minimum inhibitory concentration testing on the closely related *M. parvus* BRCS2 (as Isolate 12) and showed susceptibility of kanamycin at 0.5, tetracycline at 0.025, gentamicin at 6, streptomycin at 1 and erythromycin at 0.0016 μg/ml. It was resistant to NDX at >256 μg/ml. It was noted the tetracycline result may be unreliable due to tetracycline's instability over extended periods[14,18]. Previous authors have also found tetracycline selection challenging[18]. The thesis also showed NDX resistance was universal across 5 tested strains covering α- and γ-proteobacterial methanotrophs[14].

4.2.4 Prior Gene Editing Techniques in Methanotrophs

Except for slow growth periods of 2-4 weeks on plates[9,13] and the presence of internal membranes in pMMO expressing species, to my

knowledge there is little unique to methanotrophs that makes editing tools methanotroph specific compared to non-methanotrophic bacteria.

In early work non-rational mutation through application of UV and chemical mutagenesis or transposon insertion was utilised to induce adaptation to conditions, however this was found challenging, attributed to effective DNA repair machinery[7]. The current standard form of modification is marker exchange mutagenesis[13] and its developments the Cre-*loxP*[13], *FRT*-Flp[25] and *sacB* system[26,27].

In marker exchange mutagenesis two homologous recombination events up and downstream of the region of interest are required, inserting a selectable marker by homology recombination. This can be performed in a single step inserting a selectable marker and cargo leaving the selectable marker in the strain[15]. Alternatively for marker removal two steps are performed where a selectable marker is inserted alone then it is exchanged again with the cargo via a second round of homology recombination and counterselected. The homology recombination DNA can be provided on a suicide plasmid or linear DNA[9,13]. In Cre-*loxP*[13] the process is as the single step marker exchange but the selectable marker is then removed by site-specific recombination by Cre between two *loxP* sites leaving a scar in the form of a minimal *loxP* site with or without an adjacent inserted sequence[13]. Cre expression comes from a secondary expression plasmid which must subsequently be cured[13]. The *FRT*-Flp system has been used in a similar fashion to Cre-*loxP* with similar drawbacks[10,25]. Marker exchange mutagenesis runs the risk of full plasmid integration from single rather than double recombination so edited strains must be checked[15]. Marker exchange mutagenesis by electroporation has been reported in the α-proteobacterial *Methylocella silvestris* from the *Beijerinckiaceae* family[13] and *M. trichosporium* OB3b[15,17].

sacB editing utilises the tendency for full plasmid integration if a single homology recombination event occurs rather than two. s*acB* is present on the incorporated plasmid along with homology markers and provides a counterselection system when provided with sucrose which forces the elimination of the plasmid resulting in unmarked deletions and this has been applied in *Methylocystis sp*. SC2 and *M. trichosporium* OB3b[26–28].

4.2.5 CRISPR/Cas Editing Background

CRISPR editing in its common engineered implementation uses the Cas9 endonuclease to instigate double strand breaks in DNA, guided in a sequence specific manner by single guide RNA (sgRNA) to a target site. These breaks are lethal unless repaired. Repair is enabled by the homology-directed repair mechanism integrating a provided DNA template and replacing DNA between homologous sites. Alternatively, the cell may self-repair *via* the non-homologous end joining mechanism which is mutation prone. If the target is reformed fully it is recut by the Cas9[29,30]. Escapees are usually due to mutation in the target region after non-homologous end joining[29]. The sgRNA consists of two sections, one that binds to the Cas9, and the other often referred to as

the “seed”, is 20bp complementary to the target sequence which is cut. Targets have to be located immediately 3’ of a protospacer-adjacent motif (PAM), of 2-5bp, which is specific to the CRISPR system[31] and varies between species. *Streptococcus pyogenes* Cas9, the most commonly utilised, including in this section, has a PAM of 5’-NGG-3’[30,32]. Seeds can tolerate some mismatches which decreases escapees by target mutation, but also risks off-target mutation[30]. To minimise this risk automated systems for seed design have been implemented with formula for scoring the uniqueness and probable success of targets[33,34]. In some cases, the Cas9 and sgRNA are separated from the repair cassette on different plasmids to aid transformation or conjugation. CRISPR can also be used with multiple targets (multiplexed) from a single plasmid and Cas9 allowing simultaneous mutations to be achieved reducing the requirement for selection testing and regrowth if multiple edits are intended. This added complexity would however be expected to reduce editing efficiency (EdE)[20,29]. Standard CRISPR-Cas9 editing is illustrated in Figure 17.

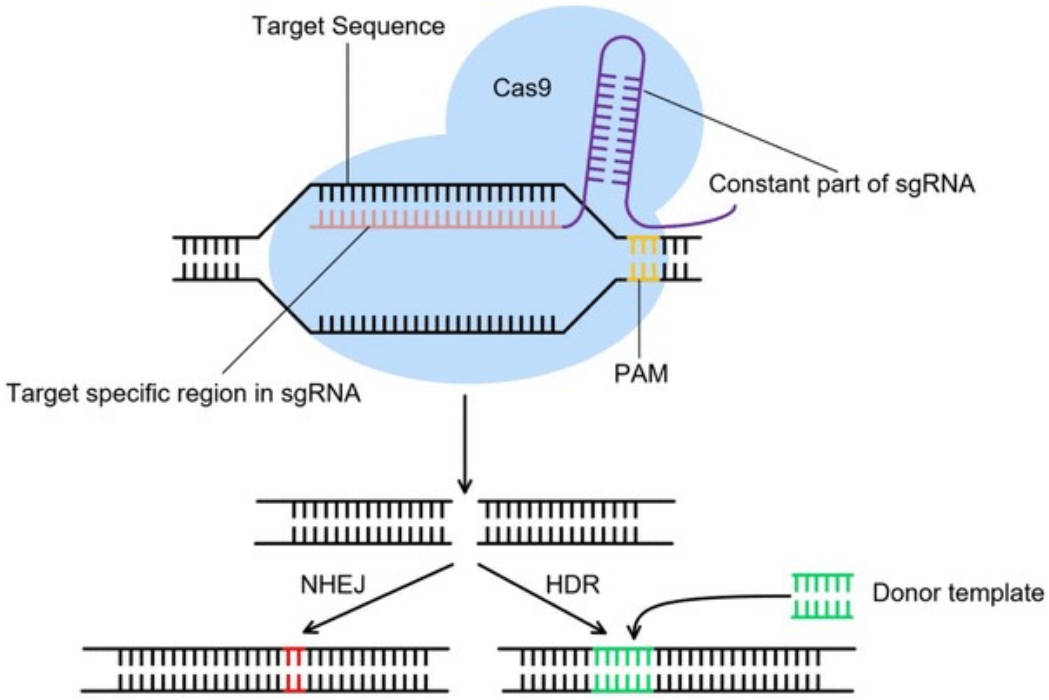


Figure 17: A diagram illustrating CRISPR-Cas9 editing showing how the Cas9 in blue is directed by the sgRNA to bind and cut at a specified region adjacent to the PAM. This results in a break that can be repaired by non-homologous end joining (NHEJ) or homology directed repair (HDR) utilising a provided donor sequence to enact insertion, replacement, or deletion. In this work the Cas9, sgRNA and donor template are all provided on a single plasmid. Figure from Cui et al. [35].

nCas9 (nicking) and dCas9 (endonuclease dead) are modified versions of the Cas9 which cause single strand breaks or no breaks only binding to the target DNA respectively. nCas9 can also be used to initiate single-nick-assisted homology directed repair from a template[20], or can be targeted twice to produce a staggered double strand break with a sticky end theoretically decreasing off-target effects by requiring two events for the break[30]. dCas9 allows blocking of specific regions as in CRISPRi (interference) or localisation

of cargos bound to the Cas9 and has other uses in more sophisticated gene regulation control and editing techniques[30].

CRISPR has several advantages over marker exchange mutagenesis and other homology recombination-based methods:

1) Cas9 cutting is a common event unlike homology recombination which requires two rare recombinations to function. 2) It is not effected by plasmid incorporation which is a danger in other methods (except in *sacB* where it is a feature) 3) No selectable marker is installed on the genome, antibiotic selection is for plasmid conjugation and the Cas9 provides its own selection against non-mutants[30]. 4) This internal selection leads to high EdE reducing the number of colonies screened to find successful mutants[29,30]. This is reflected by efficiencies of 70+% generally being observed[29]. 5) Edits are scarless and exactly as designed. 6) Editing takes place in a single step of conjugation and selection unlike other methods which require further steps for site specific recombination or *sacB* exposure. Reducing selection steps is of particular importance in with extended growth times in methanotrophs on plates.

The major downside compared to marker exchange mutagenesis based methods which can utilise linear DNA or suicide vectors is the requirement to cure the CRISPR plasmid[30], however in our system Bashir Rumah found it possible to counterselect spontaneous curing with one cycle into liquid then solid media[36] . CRISPR can also be challenging to implement in some species due to cytotoxicity of the Cas9[30,31], however this appears not to be the case in at least *M. capsulatus* Bath and *M. parvus*[8].

4.2.6 Previous CRISPR Development in Methanotrophs

Three papers on CRISPR in methanotrophs have been published, two focusing on *M. capsulatus* Bath[16,20] described in this heading, and ours published alongside this chapter which approaches *M. capsulatus* Bath and *M. parvus* OBBP[8].

Work by Tapscott et al. 2019[20], showed progress towards production of a CRISPR system in *M. capsulatus* Bath but limited success. Modification of a plasmid borne GFP reporter was achieved with both Cas9 and nCas9 (71%EdE). When attempting genomic editing they found Cas9 caused excessive cell death and no successful edits, instead using an nCas9 to make single stranded breaks to achieving knock out of *mmoX* at a low EdE (2%). Editing by homology repair making four point-mutations was demonstrated but no insertion or deletion. This used a two-plasmid system separating the Cas9 from the sgRNA and template using an inducible tetracycline promotor to control Cas9 expression.

Jeong et al. 2023[16] published a month after our CRISPR paper, provided a heterologous pathway for the production of mevalonate by expression on plasmid of *phaA*, *fxpk*, *pta*, *mvaE* and *mvaS* while knocking out *ackA* on the *M. capsulatus* Bath genome, this blocked acetate production and achieved impressive mevalonate titres (2,090 mg/L). Their CRISPR system did not

utilise homology repair, instead using cytosine base editing, their Cas9 being described as an "nCas9-CD (cytidine deaminase)-UGI" effector protein under the control of a phenol inducible promotor. This is a specialised system that makes single base edits of C -> T generally applied to introduce stop codons[30,37]. It is thus unable to make insertions and deletions or other substitutions. Jeong et al. [16] inserted premature stop codons in 4 genes, *mmoX*, *acs*, *ackA* and *acyP*. Up to three passages were applied to allow the base editing to take place with EdE increasing each time ranging from 0 to 87% which they quote to be >20%.

Comparatively our work uses constitutive promotors and a single plasmid with WT Cas9 demonstrating large deletions of 1000+ bps using homologous recombination. Insertion of eYFP in place of *ligD* is also demonstrated in the accompanying paper[8].

4.2.7 Utilising Endogenous/Native CRISPR

The engineered CRISPR/Cas9 system is based on CRISPR systems found in 50% of bacteria and 90% of archaea utilised for adaptive immunity by targeting invading DNA e.g. from bacteriophage, mobile genetic elements and unwanted plasmids and incorporating it into its genome for future targeting against that sequence[29–31]. These can also be utilised in engineered native CRISPR (ntvCRISPR) genome editing.

CRISPR systems come in types I to VI differing by the *cas* genes present and the method of function. The Cas9 type generally used for editing is Type II[30]. In their wild state Cas9 requires two RNA molecules a tracrRNA which binds to the Cas9 and a crRNA which provides the targeting by complementary base pairing and binds to the tracRNA. In modern engineered systems these have been combined into a single sgRNA. Other CRISPR types may only require a crRNA and no tracRNA[38]. A wild type CRISPR array consists of alternating regions of direct repeats and variable spacers which code for the crRNA. The whole array is expressed as a long pre-crRNA transcript and the direct repeats form cutting targets releasing the individual crRNAs. The DNA targeted by the crRNA are referred to as protospacers[30,31]. Self-targeting against the spacers is prevented by the requirement of the PAM[31]. As noted in the previous chapter (section 3.4.3.2) *M. parvus* has an endogenous CRISPR system of the 1C Type. In this type Cas1, 2 and 4 perform spacer acquisition, Cas5 controls crRNA release and Cas3 is the effector endonuclease which binds to target DNA in a "cascade" complex along with 5, 7 and 8[31]. Comparatively in Type II systems the Cas9 operates in place of the Cas3 endonuclease without need for the cascade genes making it a simpler and more tractable system for heterologous expression[31].

It has been noted that Cas3 produces single not double strand breaks and initiates exonuclease activity, degrading in a 3' -> 5' direction from the cut site[31,38] logically this might be more challenging to repair and lead to more killing than the point double strand break endonuclease activity of Cas9, however Cas3 has previously been proved effective in ntvCRISPR editing. By

elimination of a greater section of DNA it may avoid non-homologous end joining putting preference on homology directed repair increasing EdE.

To utilise endogenous CRISPR systems or avoid their attack on introduced DNA, the PAM must be identified. In early CRISPR editing work this was achieved by combinatorial testing, for example for a 5bp PAM 4^5 possibilities exist making 1,024 variants, doubling that if unsure of the PAM placement up or downstream of the protospacer. These can be pooled, tested and deep sequenced to indicate those that are underrepresented implying they were subject to CRISPR activity[29]. A rational design approach was tested and is discussed later in this chapter.

Type I ntvCRISPR systems and examples of their use in genetic editing have been further reviewed by Zheng et al. 2020[31] including applications of a disabled dCas3 making a highly effective strain for CRISPRi silencing using only an artificial CRISPR array.

4.3 Methods

4.3.1 The CRISPR-Cas Editing System Utilised in this Chapter

The system utilised in this chapter has a fully functioning WT Cas9 nuclease from *Streptococcus pyogenes* on a single plasmid active in *M. parvus* OBBP or *M. capsulatus* Bath. The base plasmid pMTL9BR2-Cas9 is based on the pMTL90882 plasmid from the SBRC Nottingham[39], itself produced from the pMTL80451 clostridium shuttle plasmid[40]. It's assembly was carried out by Bashir Rumah and is described in the paper[8]. I resequenced the plasmid using nanopore and expanded upon previous annotation to better understand its constituent parts using previous publications on the pMTL8 series plasmids and a variety of tools described in Figure 18[40,41]. The backbone contains a ColE1 *oriV*, RK2 *oriV*, *trfA* needed for RK2 *oriV* replication, RP4 *oriT* and *traJR* for conjugation into the methanotroph host, *kanR* for kanamycin resistance selection and a multiple cloning site. The CRISPR system elements consist of the Cas9 nuclease, an sgRNA consisting off a 150bp handle and 20bp seed region and a homology repair cassette of two 500 or 1000bp left and right arms which may flank an insertion/replacement cargo or with nothing between for a deletion (Figure 18). For use in *M. parvus* OBBP Cas9 is under the methanol dehydrogenase promoter P_{mdh} and the sgRNA is under the acetolactate synthase promoter P_{als}. Both promotors are native to *M. parvus*, constitutive and have low expression in *E. coli*. In *M. capsulatus* Bath the system is identical except for the replacement of the two promotors[8]. The sgRNA is replaced and the homology arms, and cargo are inserted to customise the plasmid for its target.

As RK2 *oriV* is functional in both *M. parvus* and E. coli[42,43] the inclusion of ColE1 *oriV* for *E. coli*[7] is redundant and could be disposed of to reduce plasmid size and metabolic load as has been carried out by other authors[16]. On the other hand ColE1 is expected to provide a higher copy number in *E. col* by ~6x [43] improving yields of plasmid extraction for molecular biology. It is unknown but possible copy number impacts conjugation efficiency (CE).

4.3.2 Molecular Biology Techniques

General Molecular Biology Techniques are outlined in section 2.2.1 including, miniprep, gDNA extraction, PCR, DNA quantification and agarose gel electrophoresis.

All restriction enzymes were sourced from New England Biolabs (NEB) and digests were carried out using the most effective recommended restriction buffer, generally CutSmart and the enzymes specific protocol and heat inactivated. Double digests were performed using the protocol recommended by NEBCloner RE Digest[44] for the enzyme combination. Antarctic Phosphatase (M0289, NEB) was added to restriction digests alongside restriction enzymes where specified. Ligation was performed with T4 DNA Ligase (M0202, NEB) following the protocol.

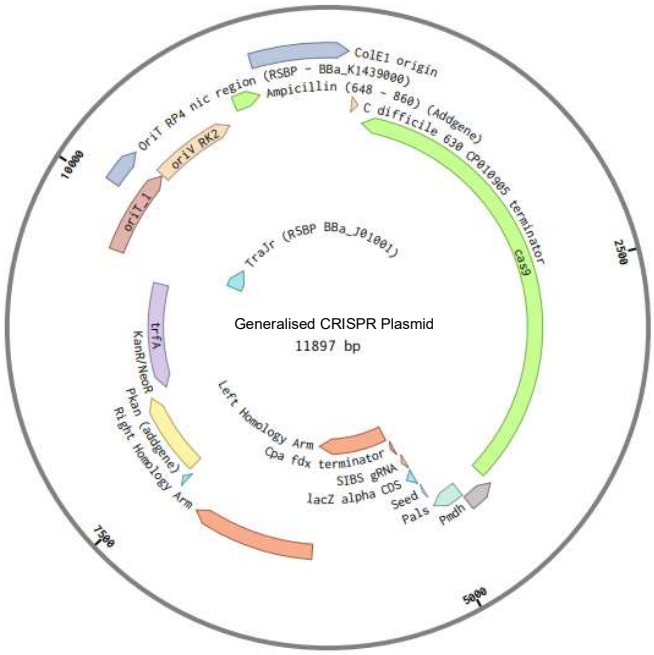


Figure 18: A map of the CRISPR-Cas Plasmid pMTL9BR2-Cas9 with generalised 1kb homology arms and no insertion cargo. The left homology arm, right homology arm, intervening region and seed are customised to the target. Additional work was done to improve the plasmid map identifying regions using a combination of the Benchling Auto annotate tool, addgene sequence analyser, and parts from the Standard Registry of Biological Parts (SRBP). The relevant sources are labelled where this took place. Cpa Fdx/terminator is the terminator from the ferredoxin gene of *Clostridia pasteurianum*. An identified incomplete Ampicillin resistance gene should be inactive and was not previously identified. Plasmid map generated using Benchling[45].

4.3.3 Chemically Competent Cells and Transformation

NEB 5-α Competent *E. coli* (NEB, C2987) were generally purchased ready to use. For all uses of *E. coli* S17-1 λpir and in some instances of NEB 5-α, competent cells were produced in house with a protocol adapted from NEB[46] (Available on Wayback Machine). Bacteria were subcultured overnight 0.5 ml in 50ml LB in a shaker flask. This was grown to 0.5 – 0.7 OD600 transferred to a 50ml falcon tube then centrifuged 6200*g*, 5 min at 4°C. Supernatant was discarded, and the pellet resuspend in 5ml ice-cold 30 mM $CaCl_2$. This was distributed 1.6ml each to three chilled 2ml tubes and centrifuged at max speed for 30 seconds in a microcentrifuge. Supernatant was discarded and each pellet was resuspended in 0.5ml ice-cold 30 mM $CaCl_2$ without vortexing. This was aliquoted in 50µl portions into chilled 1.5ml centrifuge tubes. This results in 30 tubes of competent cells.

Plasmids were transformed into chemically competent cells using the NEB High Efficiency Transformation Protocol[47] and spread on the relevant antibiotic plate. Unless specified this was LB agar with 50µg/ml kanamycin (LBK50). Successful conjugants were identified by colony PCR or miniprep followed by sequencing. Plasmids in S17-1 λpir were all miniprepped and sequenced before strain storage for later conjugation.

4.3.4 Conjugation into *M. parvus*

Conjugation of plasmids into methanotrophs were performed as follows after the method of Martin and Murrell[15]. First plasmids were transformed into *E. coli* S17-1 λpir. This was grown overnight and an equivalent volume of an OD600 of 1 at 1ml was added (Equation 2) to a 1.5ml centrifuge tube. e.g. for a starting OD600 of 2, 0.5ml was used. This was washed 3 times with 600µl NMS to remove residual antibiotics by centrifugation at 6200*g* for three minutes. An equivalent volume of an OD600 of 1 at 1ml of methanotroph set up for growth 3-4 days before was added. This was centrifuged as above, resuspended in 50µl NMS and placed in the middle of a pre-dried mating plate (NMS agar+0.5% yeast extract) and allowed to dry. Plates were incubated with methane at 30°C for 24hrs. After this time the mating pellet was picked in its entirety using a plastic loop and resuspend in 1ml of NMS. This was vortexed for 15 seconds to terminate mating. Dilutions were then made and 100µl spread onto NMS selective plates (NMS agar + nalidixic acid 25µg/ml + plasmid resistant antibiotic) unless otherwise noted the antibiotic is kanamycin at 50µg/ml. The plasmid resistant antibiotic selects against non-conjugated methanotrophs, nalidixic acid selects against donor *E. coli*. Spread plates were incubated with methane and successful transconjugants identified by PCR or miniprep and sequencing.

For data on conjugation efficiency (CE) calculated as transconjugants/donor cell, two conjugations were performed in parallel as replicates. The mating mixture was serial diluted to 10^{-8} in NMS. Three 10µl spots were placed for each conjugation and dilution onto LBK50 agar (showing *E. coli* S17-1 λpir donor colonies) and appropriate NMS selective plates (showing methanotroph transconjugants) and allowed to dry. The general range required on NMS selective plates was Neat to 10^{-3} and LBK50 10^{-4} to 10^{-7}. This allowed plates to be sectioned into eight and one plate of each type used per plasmid being conjugated. Example plates are shown in Figure 22. LBK50 plates were counted after overnight growth at 37°C. NMS selective plates were incubated with methane until sufficient growth had occurred. Colonies were counted where possible and an average of the three spots were formed. Appropriate dilution counts were aimed for between 4 and 40 to avoid merged cells biasing results. Transconjugants were divided by the donor count for that replicate to give a CE. The CE of the two replicates was averaged to give a final value.

4.3.5 Picking Seeds

CRISPR seeds were selected using Benchling's[45] CRISPR design tool which provides a list of possible targets in a select region that are correctly placed upstream of a PAM. Choices were made optimising for high on-target scores, high off-target scores, GC of 40-80% and a central position on the gene unless otherwise noted. Chosen seeds for seed specificity testing are listed in Table 10 and other seeds in Table 11 picked out in red text in the sgRNA F primers.

Table 10: Details and outcome of Seeds used in Alt-KO experiments.

Name	Seed Sequence	GC %	On Target %	Off Target %	Strand	PAM	Distance along gene /gene length (bp) (%)	Distance from end (bp)	KO %
LigD_Seed*	GGAAGCGGGCTGTCCAATCG	65	64.3	100	+ve	agg	1616/2488 (65)	872	58%
AltKO_LigDSeed1	CGAGAGGATGGTCTTCCGTG	60	73.4	100	+ve	cgg	2183/2488 (88)	305	50%
AltKO_LigDSeed2	ATGGTCGCGAATTTCCCCCG	60	71.8	100	+ve	cgg	438/2488 (18)	438	50%
AltKO_LigDSeed3	CATCACCCATGCAAGCCGGG	65	68.1	100	-ve	tgg	827/2488 (33)	827	90%
AltKO_LigDSeed4	GCGCCATATAAAGTTCGTCG	50	71	100	+ve	tgg	614/2488 (25)	614	80%
AltKO_LigDSeed5	TTTCAGCTCGAAGAGCCAAT	45	64.2	100	+ve	cgg	1704/2488 (68)	784	60%
AltKO_LigDSeed6	GATCAAGGGCGACTTTCGAG	55	66.5	100	-ve	agg	1211/2488 (49)	1211	60%
PntA_Seed*	GGACGCGGCGACGACGCTCG	80	47.2	98.6	+ve	cgg	176/1134 (16)	176	30%
AltKO_PntA _Seed_1	TTTGAAACGTTCGTCCGTAA	40	55.2	100	-ve	cgg	1105/1134 (97)	29	22%
AltKO_PntA _Seed_2	TCATCTCGAAGGAGACGAAG	50	69.3	98.2	+ve	cgg	1021/1134 (90)	113	60%
AltKO_PntA _Seed_3	CTTCGCGTAGAGGCTCGAAG	60	72.5	99.8	-ve	cgg	976/1134 (86)	158	60%
MPA_0518_Seed*	GGAGCCGGAATCGCGTCGCG	75	65.2	100	+ve	agg	50/855 (6)	50	30%
AltKO_MPA_0518_Seed_1	TTCGACACCGAGACGCCGCG	70	74.7	98.2	+ve	ggg	228/855 (27)	228	70%
AltKO_MPA_0518_Seed_2	AAGCGGTAGAGCGTGCCCCG	70	70.1	99.8	-ve	cgg	246/855 (29)	246	0%
AltKO_MPA_0518_Seed_3	AAGGCCGACCAGAACCACGA	60	67.9	99.6	+ve	agg	9/855 (1)	9	0%
BcsB_Seed*	CACGCCGGTGAATACGACGC	65	66.4	99.8	+ve	cgg	167/2712 (6)	167	30%
AltKO_BcsB _Seed_1	TCGTCAAGAATGTCGGCAGG	55	63.7	99.2	+ve	agg	2680/2712 (99)	32	50%
AltKO_BcsB _Seed_2	GTATGTGAACGAGCGCACGG	60	79.6	100	+ve	cgg	1382/2712 (51)	1330	50%

Percent knockouts achieved from 10 sampled colony PCRs except LigD_Seed which was from 12. *Provided by Bashir Rumah.

Scoring was calculated using Benchling with *S. pyogenes* Cas9 specific calculations for on-target scoring[34] and weighted off-target scoring[33] calculated against the *M. parvus* OBBP MetPar_1.0 genome. The on-target score shows the likelihood of cutting at the chosen location with that seed and the off-target score shows the likelihood of off target cutting based on the occurrence of similar sequences in the genome that could also be targeted.

4.3.6 HiFi Assembly

HiFi assemblies were designed using the Benchling[45] tool or manually. The resulting assembly primers are shown in Table 11. HiFi overlap sizes of 20-50bp were used during design. HiFi Assembly was performed using NEBuilder HiFi DNA Assembly Master Mix (E2621, NEB).

4.3.7 Correct Construct Identification

Overall plasmid sizing was confirmed by gel electrophoresis of single and double diagnostic digests, however due to the multiple variants of similar sizes this was not sufficient for confirmation of individual plasmid constituents. Plasmids were confirmed by miniprep and Sanger sequencing using the following primers for the CRISPR KO Plasmids: sgRNA and the near end of the LHA by Pals_Seq_F or Pals_Xba_F. If homology arms were in situ KanR and KanR_1 sequenced the RHA, if they were not it also sequenced the sgRNA and upstream area. One 1000-1400 bp Sanger read from each end was sufficient to sequence the entire sgRNA and Homology cassette depending on size or cargo. For the ntvCRISPR plasmids sequencing with M13R was sufficient to show the entire insert.

4.3.8 Plasmid Nanopore

Selected constructs were full plasmid sequenced by nanopore by Ruth Cornock with a Rapid Barcoding Kit (SQK-RBK004) on an R9.4.1 MinION Flow Cell (FLO-MIN106) and a 5 hour run time. I then basecalled with Guppy v6.4.2+97a7f06 to SUP accuracy and assemble with the Epi2Me Agent “Fastq Clone Validation” protocol. This achieved coverage of ~x2000 with ~30% full plasmid length reads (5000-10,000 bp depending on plasmid).

The source plasmids (pMTL90882 and pMTL9BR2-Cas9) and at least one representative from each series (pMTL9BR2-Cas9ΔligD_Alt6, pMTL9BR2-Cas9ΔpntA_Alt2, pHisD-A, pPhaR-A, pAra3-36-3-CGG) were sequenced to ensure no assembly method errors had been incorporated and that the design was as expected. A disagreement was identified between the plasmid map received and sequenced for pMTL9BR2-Cas9 substituting a 33bp region for a 52bp region between the Cas9 and ColE1. This is expected to be downstream of both genes and have no effect. The success of the pMTL9BR2-Cas9 has also already been demonstrated. This difference was also identified in all resulting plasmids in its lineage sourcing from the original. This difference was also identified by Sanger sequencing with the primers Cas9_JoinSeq_R and ColE1_F1.

Table 11: Primers used for plasmid sequencing, genomic edit sequencing and plasmid assembly.

Primer	Bases	Notes
Plasmid Sequencing		
Kan_R	CACTGCAAGCTACCTGCTTTC	RHA or sgRNA
Kan_R1	TCCAGATAGCCCAGTAGCTGAC	RHA or sgRNA
PaIs_Seq_F	GGTTCTCCACTTTTCACTTGCG	sgRNA and LHA
M13R	CAGGAAACAGCTATGAC	Cas9 in from *oriV* side
Cas9_JoinSeq_R	AGATGCCACTCTTATCCATC	Cas9 out towards *oriV*
ColE1_F1	ACGGTTCCTGGCCTTTTGCT	Cas9 in from *oriV* side
PaIs_Xba_F	TAATCTTCTAGACACCCAGGGATGGTCCGGC	
Part 1: CRISPR and Alt-KO Assembly		
AsiSI_sgRNA-R	ATTACAGTGCGATCGCATAAAAATAAGAAGCCTGCAAATGC	LigD sgRNA primer
AltKO_LigD_Seed_1	TAAAGAGGGGCCGAAGCTTAATTAACGAGAGGATGGTCTTCCGTGGTTTTAGAGCTAGAAATAGCAAGTT	"
AltKO_LigD_Seed_2	TAAAGAGGGGCCGAAGCTTAATTAAATGGTCGCGAATTTCCCCCGGTTTTAGAGCTAGAAATAGCAAGTT	"
AltKO_LigD_Seed_3	TAAAGAGGGGCCGAAGCTTAATTAACATCACCCATGCAAGCCGGGGTTTTAGAGCTAGAAATAGCAAGTT	"
AltKO_LigD_Seed_4	TAAAGAGGGGCCGAAGCTTAATTAAGCGCCATATAAAGTTCGTCGGTTTTAGAGCTAGAAATAGCAAGTT	"
AltKO_LigD_Seed_5	TAAAGAGGGGCCGAAGCTTAATTAATTTCAGCTCGAAGAGCCAATGTTTTAGAGCTAGAAATAGCAAGTT	"
AltKO_LigD_Seed_6	TAAAGAGGGGCCGAAGCTTAATTAAGATCAAGGGCGACTTTCGAGGTTTTAGAGCTAGAAATAGCAAGTT	"
LigD_LHA-AsciSI-F	AGGCTTCTTATTTTTATGCGATCGCTTTCTTTTTTGCGTATCTCCGGC	
LigD_LHA-R	GCCGCGCAACGCCTCGCAATGTGGATCAAGCCCGGCGGCGCGTT	
LigD_RHA-F	AACGCGCCGCCGGGCTTGATCCACATTGCGAGGCGTTGCGCGGC	
LigD_RHA-AscI-R	CATGGCCGGCCCAGTCGGCGCGCCACCTCGCGCTCGCCCGCGCGCC	
PntA_sgRNA_R	TCGACGCGGTCCTTCTTCTCGGCGCGCCGCGATCG	pntA sgRNA primer
AltKO_PntA_Seed_1	AAGAGGGGCCGAAGCTTAATTTTGAAACGTTCGTCCGTAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	"
AltKO_PntA_Seed_2	AAGAGGGGCCGAAGCTTAATTCATCTCGAAGGAGACGAAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	"
AltKO_PntA_Seed_3	AAGAGGGGCCGAAGCTTAATCTTCGCGTAGAGGCTCGAAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	"
PntA_LHA_F	TTATGCGATCGCGGCGCGCCGAGAAGAAGGACCGCGTCGAGAGAAGAAGGACCGCGTCGA	
PntA_LHA_R	TTTGAGGGATGCTATGCGGCGACGAACTCCGGTTTCGAGAGACGAACTCCGGTTTCGAGA	
PntA_RHA_F	TCTCGAAACCGGAGTTCGTCGCCGCATAGCATCCCTCAAACGCCGCATAGCATCCCTCAAAC	
PntA_RHA_R	TGGCCGGCCCAGTCGGCGCGTCGGTCGAGACGAAAAGCAGTCGGTCGAGACGAAAAGCAG	
MPA_0518_sgRNA_R	CGACCACGACCTATCGCTACGGCGCGCCGCGATC	MPA_0518 sgRNA primer
AltKO_MPA_0518_Seed_1	AAGAGGGGCCGAAGCTTAATTTCGACACCGAGACGCCGCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	"
AltKO_MPA_0518_Seed_2	AAGAGGGGCCGAAGCTTAATAAGCGGTAGAGCGTGCCCCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	"
AltKO_MPA_0518_Seed_3	AAGAGGGGCCGAAGCTTAATAAGGCCGACCAGAACCACGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	"
MPA_0518_LHA_F	TTATGCGATCGCGGCGCGCCGTAGCGATAGGTCGTGGTCCGTAGCGATAGGTCGTGGTC	
MPA_0518_LHA_R	GCGGCGACTCATTCAATCCTGCGGTCTCCTATGGCGTGCGGTCTCCTATGGCG	
MPA_0518_RHA_F	CAAGCGCCATAGGAGACCGCAGGATTGAATGAGTCGCCGCACAGGATTGAATGAGTCGCCGCAC	
MPA_0518_RHA_R	TGGCCGGCCCAGTCGGCGCGCTTCGCCGCGTCCGGAGGCTTCGCCGCGTCCGGAG	
Part 2: Further CRISPR Edits Assembly		
BcsB_sgRNA_R	TGAAGGCCACCGCTTCGTCGGGCGCGCCGCGATCG	bcsB sgRNA primer
AltKO_BcsB _Seed_1	AAGAGGGGCCGAAGCTTAATTCGTCAAGAATGTCGGCAGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	"
AltKO_BcsB _Seed_2	AAGAGGGGCCGAAGCTTAATGTATGTGAACGAGCGCACGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	"
BcsB _LHA_F	TTATGCGATCGCGGCGCGCCCGACGAAGCGGTGGCCTTCCGACGAAGCGGTGGCCTTC	

Primer	Bases	Notes
BcsB _LHA_R	GGCGACGCGGAGCGCGAAGCTCATGCCGATTTGCGCGCTCATGCCGATTTGCGCGC	
BcsB _RHA_F	CGGCGCGCAAATCGGCATGAGCTTCGCGCTCCGCGCATGAGCTTCGCGCTCCGCG	
BcsB _RHA_R	TGGCCGGCCCAGTCGGCGCGTCGACGCCGTTCGTCACGTCGACGCCGTTCGTCACG	
LeuB_sgRNA_F1	AGGGGCCGAAGCTTAATTAAATATTCGCGCTTGTGCAGCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	
LeuB_sgRNA_F2	AGGGGCCGAAGCTTAATTAAGCGCAAGCTCGAAAGCGACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	
LeuB_sgRNA_R	TAGGAGCGCTTGAAGACGACGCGATCGCATAAAAATAAGAAGCCT	
LeuB_LHA_F	TCTTATTTTTATGCGATCGCGTCGTCTTCAAGCGCTCCTA	
LeuB_LHA_R	AACGCGCAAACGAGGCTTTCGCCGCCTCGCCCAT	
LeuB_RHA_F	AGCGCGATGGGCGAGGCGGCGAAAGCCTCGTTTGCGCGT	
LeuB_RHA_R	GCCGGCCCAGTCGGCGCGCCCGCGTCACCTCGAATTTCTCG	
HisD_sgRNA_F1	AGGGGCCGAAGCTTAATTAACGAGGATCAGCACTTCCGAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	
HisD_sgRNA_F2	AGGGGCCGAAGCTTAATTAAGATGATTTCGAGATGTTCGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	
HisD_sgRNA_R	GACGGCGATGGGCGAGCGCCGCGATCGCATAAAAATAAGAAGCCTG	
HisD_LHA_F	TCTTATTTTTATGCGATCGCGGCGCTCGCCCATCG	
HisD_LHA_R	AAAACCGGTTGGAGATGGAAGCGTGGAGGCAGGCG	
HisD_RHA_F	CCGCGCGCCTGCCTCCACGCTTCCATCTCCAACCGGTTTTCG	
HisD_RHA_R	GCCGGCCCAGTCGGCGCGCCCTTCATGCCGGGCAAGC	
ThrC_sgRNA_F1	AGGGGCCGAAGCTTAATTAAGTTCGCGGAAAGTTTGATTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	
ThrC_sgRNA_F2	AGGGGCCGAAGCTTAATTAAGCATCCCGCGAAATTCCCCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	
ThrC_sgRNA_R	GAACGAAGCCGCGTTCGACCGCGATCGCATAAAAATAAGAAGCCTG	
ThrC_LHA_F	TCTTATTTTTATGCGATCGCGGTCGAACGCGGCTTCG	
ThrC_LHA_R	AGCCATGCCGAAACTCGCCGATTTTCAATCCGTCAGGCAGGG	
ThrC_RHA_F	CTGCCTGACGGATTGAAAATCGGCGAGTTTCGGCATGG	
ThrC_RHA_R	GCCGGCCCAGTCGGCGCGCCGAAGCCGACGCGCTCC	
P2AT-T_sgRNA_F1	AGGGGCCGAAGCTTAATTAAGGAAGGCGCCCGTTTCATCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	
P2AT-T_sgRNA_F2	AGGGGCCGAAGCTTAATTAAGAAGCCGATGACGGCCGTGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	
P2AT-T_sgRNA_R	GGGGCTCGTGCGGAAGCTGAGCGATCGCATAAAAATAAGAAGCCTG	
P2AT-T_LHA_F	TCTTATTTTTATGCGATCGCTCAGCTTCCGCACGAGC	
P2AT-T_LHA_R	GTTCCATCGTCGCGCCTCCGGTCAGCCATGATAGTTTATGTCAGTTTTG	
P2AT-T_RHA_F	CATAAACTATCATGGCTGACCGGAGGCGCGACGATG	
P2AT-T_RHA_R	GCCGGCCCAGTCGGCGCGCCAAGGCGAGTTTGTGAACGTTCG	
PhaZa_sgRNA_F1	AGGGGCCGAAGCTTAATTAATTTCTACGACGAATATCTCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	
PhaZa_sgRNA_F2	AGGGGCCGAAGCTTAATTAATACGCGCGTCAATCCGACAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	
PhaZa_sgRNA_R	TTCGTAATAGCCGAAGTCCCGCGATCGCATAAAAATAAGAAGCCTG	
PhaZa_LHA_F	TCTTATTTTTATGCGATCGCGGGACTTCGGCTATTACGAAATTC	
PhaZa_LHA_R	GAAGCGTAATTATGGACCGTTATGGTTTCCGTCTAATCGCTTCG	
PhaZa_RHA_F	GCGATTAGACGGAAACCATAACGGTCCATAATTACGCTTCTCC	
PhaZa_RHA_R	GCCGGCCCAGTCGGCGCGCCCGTCGTCAACCTGATTCCCG	
PhaZb_sgRNA_F1	AGGGGCCGAAGCTTAATTAATCAGATCAGGACCTTCTACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	
PhaZb_sgRNA_F2	AGGGGCCGAAGCTTAATTAACGACCACGAAGTCTATGTCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	
PhaZb_sgRNA_R	CCCGCATTTCAAGGTGTGGTGCGATCGCATAAAAATAAGAAGCCT	
PhaZb_LHA_F	TCTTATTTTTATGCGATCGCACCACACCTTGAAATGCGGG	
PhaZb_LHA_R	GCAGGCTCCCGGGAAAGCGCGCGCTGGAGAGGAGCC	
PhaZb_RHA_F	CAAGGGCTCCTCTCCAGCGCGCGCTTTCCCGGGAGC	
PhaZb_RHA_R	GCCGGCCCAGTCGGCGCGCCGACGCGATCTTGGTGACGC	
PhaR_sgRNA_F1	AGGGGCCGAAGCTTAATTAACTCGATCGACAAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	

Primer	Bases	Notes
PhaR_sgRNA_F2	AGGGGCCGAAGCTTAATTAACAGATCATCTTCGAGCAGGAGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
PhaR_sgRNA_R	GCGCCACCGCTTTCGGCGTCGCGATCGCATAAAAATAAGAAGCCTG	
PhaR_LHA_F	TCTTATTTTTATGCGATCGCGACGCCGAAAGCGGTGG	
PhaR_LHA_R	AAGCGATGCCCGCGCCTTCCACGCCTGACAGTGCTCCC	
PhaR_RHA_F	AGGGGAGCACTGTCAGGCGTGGAAGGCGCGGGCATC	
PhaR_RHA_R	GCCGGCCCAGTCGGCGCGCCTCCAACTGCCCGACGAGAC	
Phas_sgRNA-F1	AGGGGCCGAAGCTTAATTAACCTCGCGCAGAAGCTCGTAAGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
Phas_sgRNA-F2	AGGGGCCGAAGCTTAATTAACGCCGAAGCCAATGTCAATGGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
Phas_sgRNA_R	CGCCGCCGCCCGCGCGGTTTGCGATCGCATAAAAATAAGAAGCCTG	
Phas_LHA_F	TCTTATTTTTATGCGATCGCAAACCGCGCGGGCG	
Phas_LHA_R	TGATCGGGAATCGCGGACGAGGATGTCTCCTGTTGCGGAC	
Phas_RHA_F	GTCCGCAACAGGAGACATCCTCGTCCGCGATTCCCGAT	
Phas_RHA_R	GCCGGCCCAGTCGGCGCGCCTTCGCGAAATCCATCAACTCCTT	
Phaf1_sgRNA_F1	AGGGGCCGAAGCTTAATTAAGCTGGCGGAAAAGCTCATCGGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
Phaf1_sgRNA_R	CGTTTCTACGCATCGACCTCGCGATCGCATAAAAATAAGAAGCCT	
Phaf1_LHA_F	TCTTATTTTTATGCGATCGCGAGGTCGATGCGTAGAAACGA	
Phaf1_LHA_R	GCGCATCGGCGTAAGATTTTGGGGTTTGTCCTCATAAAATGGC	
Phaf1_RHA_F	ATTTTATGAGGACAAACCCCAAAATCTTACGCCGATGCGC	
Phaf1_RHA_R	GCCGGCCCAGTCGGCGCGCCTCGGCGCGAGCAGC	
Phaf2_sgRNA_F1	AGGGGCCGAAGCTTAATTAAGAAATCGCTCGATCTGTGGGGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
Phaf2_sgRNA_F2	AGGGGCCGAAGCTTAATTAAGGCGAGATCGAAAAACGCCGGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
Phaf2_sgRNA_R	GCGAGGCGGACTTTGCCCGAGCGATCGCATAAAAATAAGAAGCCTG	
Phaf2_LHA_F	TCTTATTTTTATGCGATCGCTCGGGCAAAGTCCGCC	
Phaf2_LHA_R	CAATCGAAACGTCGGGAGCCGCTACCGCTCCTCGGTTG	
Phaf2_RHA_F	CGCAACCGAGGAGCGGTAGCGGCTCCCGACGTTTCGATT	
Phaf2_RHA_R	GCCGGCCCAGTCGGCGCGCCGCGCTTGCGTGATCATCG	
P1repA_sgRNA_F1	AGGGGCCGAAGCTTAATTAACGAAAAACATTAAATCCCCGGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
P1repA_sgRNA_F2	AGGGGCCGAAGCTTAATTAAACTGTGATGAACTTCAAAGGGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
P1repB_sgRNA_F1	AGGGGCCGAAGCTTAATTAATCGGTTTGAAAACTGGACGAGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
P1repB_sgRNA_F2	AGGGGCCGAAGCTTAATTAAGTCGGAAAGAGCCTTGACGAGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
P2repA_sgRNA_F1	AGGGGCCGAAGCTTAATTAAATGGATGGTGATGAGAACAGGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
P2repA_sgRNA_F2	AGGGGCCGAAGCTTAATTAAGGAGCGTCTTCAGGTCATCGGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
P2repB_sgRNA_F1	AGGGGCCGAAGCTTAATTAAAGATTACGTTCATCTCCGGGGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
P2repB_sgRNA_F2	AGGGGCCGAAGCTTAATTAAATCCTCAAGATTCTTGACGAGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	
p1n2gen_sgRNA_R	GCCGGCCCAGTCGGCGCGCCGCGATCGCATAAAAATAAGAAGCCT	Matches with P1 and P2 repA and repB sgRNA F primers
Part 3: ntvCRISPR Site Directed Mutagenesis Primers		
Ntv_SDM-F	atcccagcgacgcccgcgtcgtGATTACGAATTCGGTACCC	
Ntv_SDM-R	cgccgcgatcgtcgcgcggtttcATGGTCATATGCCGCCTC	
TTC_SDM-_F	cgatcccagcgacgcccgcGATTACGAATTCGGTACCC	Also ATT_SDM-F and ACT_SDM-F
TTC_SDM-_R	ccgcgatcgtcgcgcggaaATGGTCATATGCCGCCTC	
ATT_SDM-R	ccgcgatcgtcgcgcgaatATGGTCATATGCCGCCTC	
ACT_SDM-R	ccgcgatcgtcgcgcgagtATGGTCATATGCCGCCTC	
3-CGC_SDM-F	tcccagcgacgcccgccgcGATTACGAATTCGGTACCC	
3-CGC_SDM-R	tcgccgcgatcgtcgcgcgATGGTCATATGCCGCCTC	Also 3-CGG_SDM-R
3-CGG_SDM-F	tcccagcgacgcccgccggGATTACGAATTCGGTACCC	
Genomic Edit Screening Primers		
LigD_LHA_OHA-F	CCCGTCAACCTTCGATCCG	
LigD_LHA_OHA-R	CGCCTCCTTTGACCTCATCTAC	

Primer	Bases	Notes
pntA-KO_OHA_F	CGACCGTGAGAAGCTCCAGG	†
pntA-KO_OHA_R	GATCGGGATGATCAGCGTCAC	†
bcsB_seq-F	GAAGCTCGGCTATCTCTCGAG	†
bcsB_seq_R	TGGCGGGATAATAATTGTCG	†
HYPROT_seq-F	CAGATAGAGCCCGTTCATCAG	
HYPROT_seq-R	TGGTAGAGATCGACATAATCGG	
LeuB_OSeq_F	TTTGGGAGCGCCTGGAGAA	
LeuB_OSeq_R	GCTCGCCTTGAAATAGAACTCGTA	
LeuB_OSeq2_F	CTATGAAATCCTCACGCAGCAC	
LeuB_OSeq2_R	GTTCTGGATGGTGGTCGAGTAATAG	
HisD_OSeq_F	CGCACCGTGTTCAGCGT	
HisD_OSeq_R	ACAGCGAAGATCTCGCGC	
ThrC_OSeq_F	GGTCTCGAGCCGGGCTATAT	
ThrC_OSeq_R	AAATGGTCGCGCCACTCG	
P2AT-T_OSeq_F	GTGCGTCGCTATGTCGGTC	
P2AT-T_OSeq_R	CTCGTCGGTCCCGTTCGAT	
PhaZa_OSeq_F	CAAAGAGCTTCGTGCTCTTCGAC	
PhaZa_OSeq_R	CGATCAACACGGCCGACAG	
PhaZb_OSeq_F	CGCTCGTTTCGAGGCATTCG	
PhaZb_OSeq_R	GGCGATAGAGGGCAATCTGAGA	
PhaR_OSeq_F	GACGACGATCTGGGCGTC	
PhaR_OSeq_R	CGCGGGCGAATATTGCTCC	
Phas_OSeq_F	TCGACGAGTTCCGCGAGAC	
Phas_OSeq_R	GAAACGTCGGGAGCCTCAG	
Phaf1_OSeq_F	GCTATTGGCGTCCATGACGATG	
Phaf1_OSeq_R	GTTGCGGGCGAAACAGCT	
Phaf2_OSeq_F	ATGACGCGCCGCACAATTTA	
Phaf2_OSeq_R	GTCGCGTCGCTTCACGAA	
P1repA_Seq_F	ACTCCGCAGCTTTGATTTGGT	
P1repA_Seq_R	CAGCGACGGACTTGTCGTC	
P1repB_Seq_F	GTCAACGGAGAGATCGAAGGC	
P1repB_Seq_R	TACTTTTTGACGCCACGCGAAA	
P2repA_Seq_F	GCGTGAAAAAGCGTTATTCCGG	
P2repA_Seq_R	GGCGGTTTTACCATTCTGCAC	
P2repB_Seq_F	GCGATTGGCGCAAGAGGA	
P2repB_Seq_R	CAGAGAAAGGCTTCCGGGC	

Primer names have been matched to those presented in the publication where applicable[8].
†Primer and sequences for *pntA* and *bcsB* sequencing primers are swapped in error in the publication. The ones presented here are correct.

4.3.9 Testing Genome Edits

Unless otherwise noted 10 colonies from selection plate were tested by colony PCR using their respective primer pairs under the "Genomic Edit Screening Primers" heading of Table 11. These primer pairs were designed to amplify the edited region and only genomic DNA not the plasmid HAs by positioning primers external to the homology regions. When run on a gel the PCR indicated deletion by a smaller product size which was counted as a positive edit. A selection of PCR products from successful and failed edits were validated by Sanger sequencing. Editing efficiency (EdE) was calculated as the number of successfully edited colonies divided by the total number of successful and WT colonies according to the PCR results.

All pairs were tested against *M. parvus* OBBP gDNA to confirm expected size WT bands were produced. All were as expected except LeuB_OSeq_F/R which failed to amplify. This was redesigned as LeuB_OSeq2_F/R which was successful.

4.3.10 Methods for Part 1 – Seed Specificity Testing with Alt-KO Plasmids

4.3.10.1 Picking seeds

In addition to four seeds designed by Bashir Rumah, six additional *ligD*, three *pntA*, three MPA_0518 and two *bcsB* targeting CRISPR Seeds were selected using Benchling's[45] CRISPR design tool for a spread of factors including targeting the template and non-template strand of the gene, 3 different PAMs within the NGG requirements of our Cas9 and a range of lengths along the gene, then prioritising by on-target score and GC of 40-80%. Listed in Table 10.

4.3.10.2 Assembly of Part 1 Plasmids

Assembly of Alt-KO (Alternative Knock Out) *ligD* plasmids pMTL9BR2-Cas9Δ*ligD*_Alt1-6 was achieved in a three-part HiFi assembly followed by sgRNA exchange by restriction digest as follows. Homology arms were produced by PCR of LigD_LHA-AsciSI-F and LigD_LHA-R or LigD_RHA-F and LigD_RHA-AscI-R against *M. parvus OBBP* gDNA. The sgRNAs were produced by PCR combining AsiSI_sgRNA-R with one of the six AltKO_LigDSeed_1-6 primers using pMTL9BR2-Cas9Δ*ligD* as a template. pMTL9BR2-Cas9 which contained no homology arms, was digested with AscI and AsiSI, and joined with the LHA and RHA by NEBuilder HiFi DNA Assembly Master Mix (NEB) and transformed into *E. coli* NEB 5-α. The resulting plasmid with homology arms was verified then had its sgRNA exchanged to the six possible versions as follows. The plasmid and the sgRNA PCRs were digested with PacI and AsiSI and the plasmid was treated with antarctic phosphatase, then the digested plasmids and sgRNAs were ligated and transformed into *E. coli* NEB 5-α. The choice of restriction enzyme pair PacI and AsiSI was problematic as they produce complementary AT-- sticky ends making ligation not orientation specific requiring additional colony picks. This was corrected in future designs.

The Alt-KO *pntA*, MPA_0518 and *bcsB* plasmids were assembled in one pot HiFi reactions. The homology arms were produced by PCR of the LHA-F and LHA-R or RHA-F and RHA-R of each target against *M. parvus* OBBP gDNA. The sgRNAs containing the alternative seeds were produced by PCR combining AltKO_PntA_Seed_1-3 with PntA_sgRNA_R, AltKO_MPA_0518_Seed_1-3 with MPA_0518_sgRNA_R and AltKO_BcsB_Seed_1-2 with BcsB_sgRNA_R all using pMTL9BR2-Cas9 as the template. The pMTL9BR2-Cas9 backbone was digested with PacI and AscI and antarctic phosphatase treated then the sgRNA, digested backbone, homology arms and sgRNA were assembled in a one pot HiFi reaction and transformed into *E. coli* NEB 5-α.

4.3.11 Methods for Part 2 – Further CRISPR Edits

Two seeds were designed against each of 14 targets following the method and priorities described previously. Each target had one pair of 500bp homology arms designed to produce an in-frame deletion of the coding region. These shorter homology arms than the 1000bp used in Part 1 were demonstrated effective by Bashir Rumah in the accompanying paper[8].

The assembly of the 28 plasmids listed in Table 12 and Table 13 was achieved by one pot HiFi reaction. Homology arms were produced by PCR of the LHA-F and LHA-R or RHA-F and RHA-R of each target against *M. parvus* OBBP gDNA. sgRNAs were produced by PCR combining the relevant sgRNA_F1 or F2 with sgRNA_R using pMTL9BR2-Cas9 as the template. The pMTL9BR2-Cas9 backbone was digested with PacI and AscI and antarctic phosphatase treated then the sgRNA, digested backbone, homology arms and sgRNA were assembled in a one pot HiFi reaction and transformed into *E. coli* NEB 5-α. Plasmid pPhaf1-B had an error in assembly and was not included in testing.

P1repA, P1rebB, P2repA and P2repB _sgRNA F1 and F2 were all paired with p1n2gen_sgRNA_R and assembled without homology arms in the HiFi mix.

4.3.12 Methods for Part 3 - ntvCRISPR Plasmid Production

Six ntvCRISPR PAM testing plasmids listed in Table 12 and Table 14 were generated by site directed mutagenesis (SDM) using the NEB Q5 Site-Directed Mutagenesis Kit (E0554). Primers were designed with the NEBaseChanger[48] designer tool in insertion mode, this was also used to define the Ta. For each final plasmid pMTL90882 was subject to SDM using the manufacturers protocol except with a 20µl reaction size with an appropriate pair of SDM primers named in the format [Seed]- SDM-F/R. Designed inserts were placed 5bp from the upstream NdeI site and 6bp from the downstream EcoRI site. SDM product were transformed into *E. coli* NEB 5-α and correct transformants were identified. Final plasmids were 5569bp for the PAM testing plasmids and 5576 for the Native control plasmid as shown in Figure 19 varying only at the "ntvCRISPR Test Insert" region.

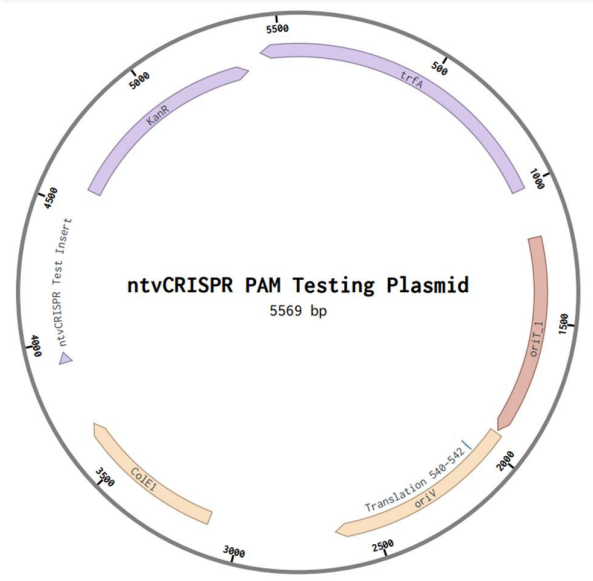


Figure 19: A map of the ntvCRISPR PAM testing plasmid. This is based on pMTL90882 and shares backbone features with the Knockout Plasmids. Map generated using Benchling[45].

Table 12: Plasmids used in this chapter.

Plasmids	Description	Source
pMTL9BR2-Cas9	*phaC sgRNA*	Bashir Rumah[8]
Alt-KO Plasmids (Part 1)		
pMTL9BR2-Cas9Δ*ligD*	*ligD* sgRNA, *ligD* LHA, *ligD* RHA	Bashir Rumah[8]
pMTL9BR2-Cas9Δ*ligD*_Alt1	*ligD* sgRNA Alt1, *ligD* LHA, *ligD* RHA	This study
pMTL9BR2-Cas9Δ*ligD*_Alt2	*ligD* sgRNA Alt2, *ligD* LHA, *ligD* RHA	This study
pMTL9BR2-Cas9Δ*ligD*_Alt3	*ligD* sgRNA Alt3, *ligD* LHA, *ligD* RHA	This study
pMTL9BR2-Cas9Δ*ligD*_Alt4	*ligD* sgRNA Alt4, *ligD* LHA, *ligD* RHA	This study
pMTL9BR2-Cas9Δ*ligD*_Alt5	*ligD* sgRNA Alt5, *ligD* LHA, *ligD* RHA	This study
pMTL9BR2-Cas9Δ*ligD*_Alt6	*ligD* sgRNA Alt6, *ligD* LHA, *ligD* RHA	This study
pMTL9BR2-Cas9Δ*pntA*	*pntA* sgRNA, *pntA* LHA, *pntA* RHA	Bashir Rumah[8]
pMTL9BR2-Cas9Δ*pntA*_Alt1	*pntA* sgRNA Alt1, *pntA* LHA, *pntA* RHA	This study
pMTL9BR2-Cas9Δ*pntA*_Alt2	*pntA* sgRNA Alt2, *pntA* LHA, *pntA* RHA	This study
pMTL9BR2-Cas9Δ*pntA*_Alt3	*pntA* sgRNA Alt3, *pntA* LHA, *pntA* RHA	This study
pMTL9BR2-Cas9ΔMPA_0518	MPA_0518 sgRNA, MPA_0518 LHA, MPA_0518 RHA	Bashir Rumah[8]
pMTL9BR2-Cas9ΔMPA_0518_Alt1	MPA_0518 sgRNA Alt1, MPA_0518 LHA, MPA_0518 RHA	This study
pMTL9BR2-Cas9ΔMPA_0518_Alt2	MPA_0518 sgRNA Alt2, MPA_0518 LHA, MPA_0518 RHA	This study
pMTL9BR2-Cas9ΔMPA_0518_Alt3	MPA_0518 sgRNA Alt3, MPA_0518 LHA, MPA_0518 RHA	This study
pMTL9BR2-Cas9ΔBcsB	*bcsB* sgRNA, *bcsB* LHA, *bcsB* RHA	Bashir Rumah[8]
pMTL9BR2-Cas9ΔBcsB_Alt1	*bcsB* sgRNA Alt1, *bcsB* LHA, *bcsB* RHA	This study
pMTL9BR2-Cas9ΔBcsB_Alt2	*bcsB* sgRNA Alt2, *bcsB* LHA, *bcsB* RHA	This study
Part 1 Plasmids also all contained RK2 *oriV*, RP4 *oriT*, *trfA*, ColE1 *oriV*, kan^R, Cas9, P_{mdh}, P_{als} and *tfdx*. Homology Arms were 1000bp each. These are based on the pMTL9BR2-Cas9 plasmid.		
Additional CRISPR Plasmids (Part 2)		
pLeuB-A	*leuB*-A sgRNA, *leuB* LHA, *leuB* RHA	This study
pLeuB-B	*leuB*-B sgRNA, *leuB* LHA, *leuB* RHA	This study
pHisD-A	*hisD*-A sgRNA, *hisD* LHA, *hisD* RHA	This study
pHisD-B	*hisD*-B sgRNA, *hisD* LHA, *hisD* RHA	This study
pThrC-A	*thrC*-A sgRNA, *thrC* LHA, *thrC* RHA	This study
pThrC-B	*thrC*-B sgRNA, *thrC* LHA, *thrC* RHA	This study
pP1repA-A	pMpar-1 *repA*-A sgRNA	This study
pP1repA-B	pMpar-1 *repA*-B sgRNA	This study
pP1repB-A	pMpar-1 *repB*-A sgRNA	This study
pP1repB-B	pMpar-1 *repB*-B sgRNA	This study
pP2repA-A	pMpar-2 *repA*-A sgRNA	This study
pP2repA-B	pMpar-2 *repA*-B sgRNA	This study
pP2repB-A	pMpar-2 *repB*-A sgRNA	This study
pP2repB-B	pMpar-2 *repB*-B sgRNA	This study
pP2AT-T-A	pMpar-2 *AT-T*-A sgRNA, pMpar-2 *AT-T* LHA, pMpar-2 *AT-T* RHA	This study
pP2AT-T-B	pMpar-2 *AT-T*-B sgRNA, pMpar-2 *AT-T* LHA, pMpar-2 *AT-T* RHA	This study
pPhaZa-A	*phaZa*-A sgRNA, *phaZa* LHA, *phaZa* RHA	This study
pPhaZa-B	*phaZa*-B sgRNA, *phaZa* LHA, *phaZa* RHA	This study
pPhaZb-A	*phaZb*-A sgRNA, *phaZb* LHA, *phaZb* RHA	This study
pPhaZb-B	*phaZb*-B sgRNA, *phaZb* LHA, *phaZb* RHA	This study
pPhaR-A	*phaR*-A sgRNA, *phaR* LHA, *phaR* RHA	This study
pPhaR-B	*phaR*-B sgRNA, *phaR* LHA, *phaR* RHA	This study
pPhas-A	*phas*-A sgRNA, *phas* LHA, *phas* RHA	This study
pPhas-B	*phas*-B sgRNA, *phas* LHA, *phas* RHA	This study
pPhaf1-A	*phaf1*-A sgRNA, *phaf1* LHA, *phaf1* RHA	This study
pPhaf1-B	*phaf1*-B sgRNA, *phaf1* LHA, *phaf1* RHA	This study
pPhaf2-A	*phaf2*-A sgRNA, *phaf2* LHA, *phaf2* RHA	This study

Plasmids	Description	Source
pPhaf2-B	*phaf2*-B sgRNA, *phaf2* LHA, *phaf2* RHA	This study
pPhasin1-KO	*phaf1*-C sgRNA, *phaf1-B* LHA, *phaf1-B* RHA	Bashir Rumah
pPhasin2-KO	*phas*-C sgRNA, *phas* LHA, *phas* RHA	Bashir Rumah
pPHBC2-KO	*phbC2*-A sgRNA, *phbC2* LHA, *phbC2* RHA	Bashir Rumah
pPHBC3-KO	*phbC3*-A sgRNA, *phbC3* LHA, phbC3 RHA	Bashir Rumah
pPHBdpol-KO	*phaZb*-C sgRNA, *phaZb* LHA, *phaZb* RHA	Bashir Rumah
pPHBinDepol-KO	*phaZa*-C sgRNA, *phaZa* LHA, *phaZa* RHA	Bashir Rumah
Part 2 Plasmids also all contained RK2 *oriV*, RP4 *oriT*, *trfA*, ColE1 *oriV*, kan^R, Cas9, P_{mdh}, P_{als} and *tfdx*. Homology arms were 500bp. These are based on the pMTL9BR2-Cas9 plasmid.		
ntvCRISPR Source Plasmid		
pMTL90882	RK2 *oriV*, RP4 *oriT*, *trfA*, ColE1 *oriV* and kan^R	Bashir Rumah[49]
ntvCRISPR PAM Testing Plasmids (Part 3)		
pAra3-36-Ntv	5' GAAAC, protospacer, 3' GTCGT	This study
pAra3-36-TTC	5' TTC, protospacer	This study
pAra3-36-ATT	5' ATT, protospacer	This study
pAra3-36-ACT	5' ACT, protospacer	This study
pAra3-36-3-CGC	Protospacer, 3' CGC	This study
pAra3-36-3-CGG	Protospacer, 3' CGG	This study
ntvCRISPR PAM Testing Plasmids also all contained RK2 *oriV*, RP4 *oriT*, *trfA*, ColE1 *oriV* and kan^R. These are based on the pMTL90882 plasmid.		

4.4 Results and Discussion

4.4.1 CRISPR Part 1 – Seed Specificity

Work by Bashir Rumah, previously confirmed CRISPR knockouts using a single seed in two target genes: *ligD* - Multifunctional non-homologous end joining protein and *bcsB* - Cellulose biosynthesis cyclic di-GMP-binding regulatory protein. Two additional genes were identified to be non-essential (unpublished TraDIS) *pntA* - NAD(P)transhydrogenase subunit alpha, and MPA_0518, a hypothetical protein (CP044331.1 F7D14_02490) which could not be identified by nucleotide or amino acid sequence or structural homology except to similarly identified hypothetical proteins. Seeds were designed for these later two by Bashir Rumah but had not been successful. Further plasmids were designed and produced with alternative seeds for the same target genes to explore reliability of the CRISPR system. 6 for *ligD*, 3 for *pntA*, 3 for MPA_0518 and 2 for *bcsB* all designed to perform scarless in frame deletions. Prior challenges experienced knocking out *phaC*[8] caused suspicion that gene essentiality was the reason for failure. Alternative seed designs were intended to show the non-essential nature of the gene was the important feature in the knockout, not seed design, and explore the finer points of seed design in the CRISPR system.

The Alt-KO plasmids were produced as described in methods and conjugated into *M. parvus* OBBP on selective media. Plasmids pMTL9BR2-Cas9Δ*ligD*, Δ*pntA,* ΔMPA_0518 and ΔBcsB from Bashir Rumah were treated similarly for a total of 18 plasmids. They were grown to colonies after 3 weeks under a methane/air atmosphere and picked and tested by colony PCR with the appropriate screening primers. Two example gels are presented in Figure 20.

The data of knock outs was plotted against the variables in seed design in Figure 21 A-D. Visually a trend appears to show lower knock out success at lower distances from the end of the gene (Figure 21C). Spearman's Rho (ρ) calculation was carried out as a statistical test of monotonic relationship as it was judged from the figure that any relationship was unlikely to be linear. This results in a test value $\rho=0.512$ which passes $\rho_{crit(0.05,18)}=0.475$ at $p=0.030$ therefore, there is a significant positive correlation. The non-linearity of this correlation is likely due to the sensitivity of the positioning to the first few bases with less sensitivity to positing within the central section of the gene. This shows that seed design should be targeted within the central portion of the gene, and by inspection of the Figure 21C more specifically outside the end 250bp, to maximise EdE.

A Spearman's Rho test was also performed on the other variables for completeness, all of which failed to show correlation as predicted visually. The lack of trend does not indicate these variables are unimportant or uncorrelated, the values used particularly for GC, and on and off-target score do not fully explore the variable space as broad variation was not intended during seed design. The means for positive and negative strands showed no

difference (50 and 51% KO) and the PAM choice showed equal averages of agg and cgg (40 and 41% KO), and with only 1 and 2 samples for tgg and ggg not being enough data to be conclusive.

Conjugation efficiency for a selection of plasmids (Figure 21E) indicated broadly similar efficiencies for all plasmids ($1.00\text{-}4.10\text{x}10^{-7}$) except for pMTL9BR2-Cas9ΔMPA_0518_Alt2 which had a higher CE ($1.71\text{x}10^{-3}$); as a lower efficiency is expected due to CRISPRs effect on survival, this implies pMTL9BR2-Cas9ΔMPA_0518_Alt2 did not cut successfully reflected in its failure to cause any knock outs. Conversely pMTL9BR2-Cas9ΔMPA_0518_Alt3 caused no detected knockouts but did have a CRISPR range CE of $1\text{x}10^{-7}$ suggesting it had cut but possibly at an off-target location or failed to undergo homologous recombination repair.

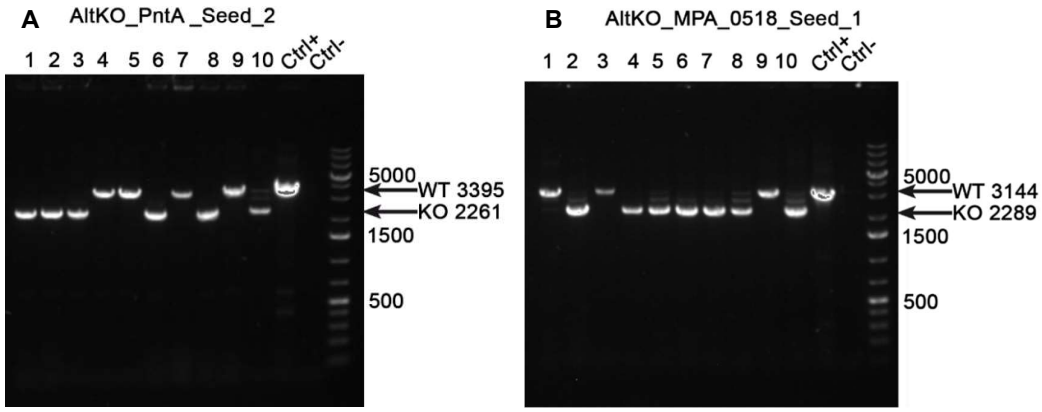


Figure 20 A and **B**: Example of two KO detection PCRs run on a 1% Agarose gel showing evidence of knockouts in *M. parvus* OBBP colonies on selection plates. Each image is produced by a different plasmid and seed. Each lane number represents an independent picked colony. PCR Bands at the WT (Wild Type) position have no deletion, bands at the KO (Knockout) position have had the gene removed. Numbers shown indicate the predicted sizes. **A -** from pMTL9BR2-Cas9Δ*pntA*_Alt2 shows 6 KOs out of 10 PCRs giving a 60% EdE and used screening primers pntA-KO_OHA_F and R. **B -** from pMTL9BR2-Cas9ΔMPA_0518_Alt1 shows 7 KOs of 10, a 70% EdE and used screening primers HYPROT_seq-F and R. Ctrl+ used unedited gDNA in the PCR and showed bands at the WT position as expected. Ctrl- used molecular biology grade water and showed no band as expected. Ladder is GeneRuler 1 kb Plus DNA Ladder.

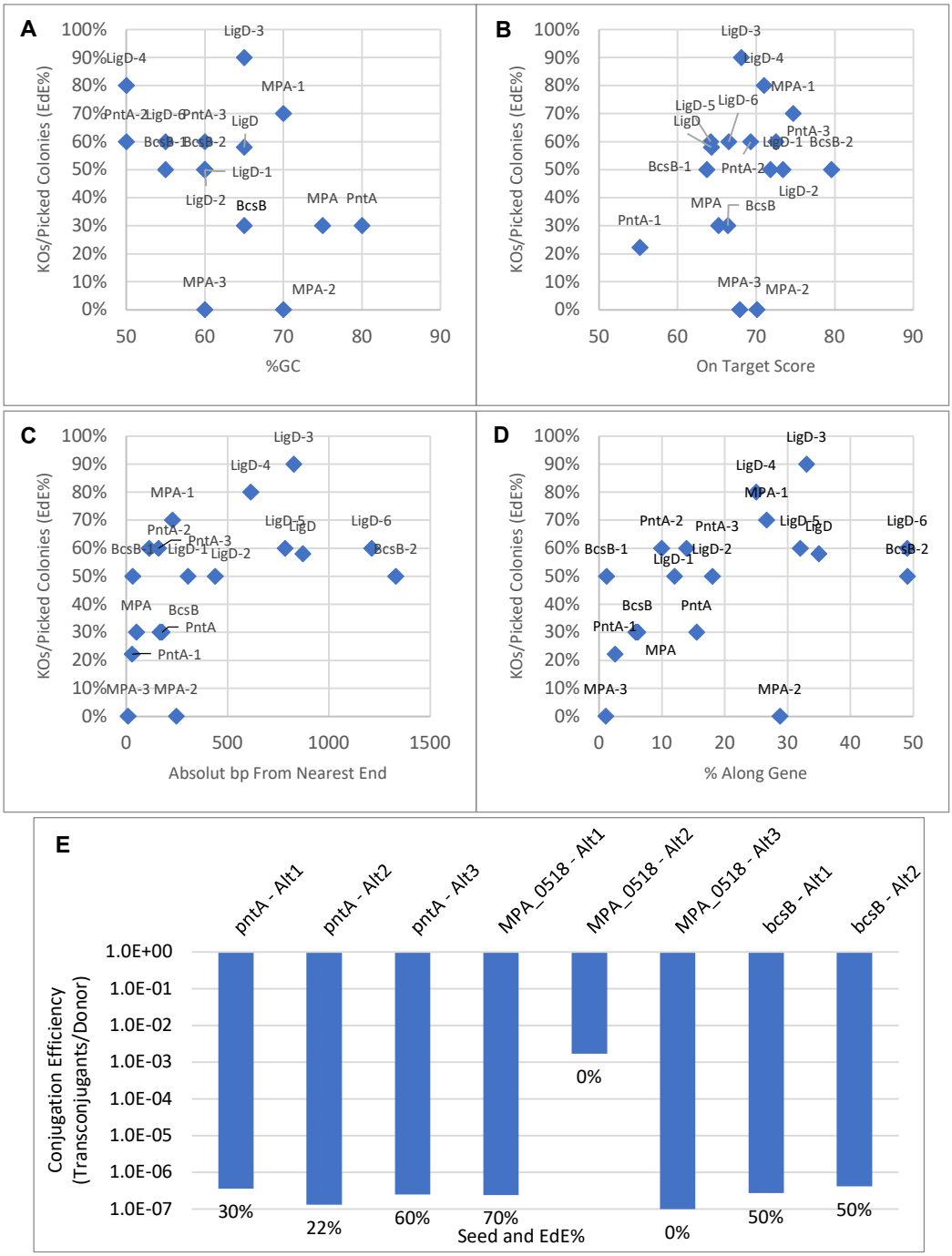


Figure 21 A-E: A-D Graphs of the knock-out success percentages of the 18 tested CRISPR Knock out seed designs including data across 4 different genes. Plotted against 4 potential influences on seed effectiveness. Individual seed identities are labelled. MPA_0518 has been shortened to MPA. **A** - GC Percentage **B** - On Target Score **C** - Absolute base pair distance from the nearest gene end **D** - Percent distance along the gene, rescaling the data in **C** to account for varied gene lengths. **E** – Conjugation efficiencies observed for 8 seeds with the KO% efficiency heading the columns. Values are the mean of two biological replicates of three 10µl spot counts. The y-axis is log plotted.

4.4.1.1 CRISPR Part 1 – Discussion

The knockouts were successful in 4 genes of *M. parvus* OBBP attempted using 18 seeds, of which only 2 seeds failed, both against a single gene (MPA_0518), conclusively showing the robustness and effectiveness of this implementation of the CRISPR/Cas technique within *M. parvus* OBBP. An average of all successful plasmids showed a 53% EdE. This improves upon previous published genomic editing CRISPR work in methanotrophs which could only function with nCas9 and shown in a single gene at very low efficiencies (2% EdE), or utilised base editing rather than deletion. Successful knock out does appear gene dependent as published work showed multiple other genes (*phaC*, *ligA*, *copD* and *glg*[8]) could not be knocked out and were likely essential, though multiple seeds were not tested thoroughly against those targets.

4.4.2 CRISPR Part 2 – Further CRISPR Edits

A further array of CRISPR edits was attempted targeting the production of a specific amino acid auxotrophic *M. parvus* OBBP strain, disruption of the PHB production machinery and curing of the two native megaplasmids.

Targets are listed in Table 13 and described below. All targets were in-frame scarless deletions with 500bp homology arms except for *repA* and *repB* which had no homology arms included.

4.4.2.1 Amino Acids Auxotrophs

Three amino acid auxotrophs were attempted against leucine, histidine and threonine, these have hydrophobic, positively charged and polar uncharged side chains respectively selected for their appearance in separate synthesis pathways and different properties increasing chance of a success. They are also intermediately low abundance amino acids in *M. parvus* so were thought to cause minimal disruption while maintaining essentiality[50]. Gene targets to accomplish these auxotrophs were chosen from tested strains in the literature including the *E. coli* K-12 Keio collection[51,52].

Essentiality of the chosen genes, and thus their likelihood of producing an auxotroph was empirically determined from TraDIS data in *M. parvus* produced by Bashir Rumah (unpublished). This found that *leuB* and *thrC* were essential and that *hisD* had mixed results. A BioCyc[53] model based on the MetPar_1.0 *M. parvus* genome and a model from my own implementation of Pathway Tools[54] based on CP092968 were also interrogated to find missing or additional predicted amino acid pathways. This was to identify if any bypass pathways are present.

Capability of amino acid uptake by *M. parvus* was unknown as they are not capable polytrophs. As the standard NMS medium is minimal containing no additional amino acids, *M. parvus* is capable of producing the full range of required amino acids natively. Two amino acid transporters were identified in *M. parvus* by BioCyc WP_016920876.1 and WP_016919425.1. I could not ascertain the applicability of these specific transporters to individual amino acids. This is significantly less than the 7 amino acid transporters identified in

E. coli according to BioCyc. Past amino acid auxotroph work in methylotrophs and methanotrophs has had mixed results. In the methylotroph *Methylophilus methylotrophus* surviving auxotrophs could not be generated due to an inability ot uptake proline. This was fixed by the expression of a heterologous permease[55]. Conversely a Calysta patent describes production of an *M. capsulatus* Bath proline auxotroph strain without requirement for any additional transporters[56].

Table 13: Targets of the further CRISPR edits.

Target Gene	Gene Locus	Aim	Plasmid 1	Plasmid 2	Plasmid 3	TraDIS essentiality
leuB	MMG94_16110	Leucine Auxotroph	pLeuB-A	pLeuB-B		No
hisD	MMG94_05520	Histidine Auxotroph	pHisD-A	pHisD-B		Yes
thrC	MMG94_17585	Threonine Auxotroph	pThrC-A	pThrC-B		Yes
phaZ	MMG94_17300	Disrupt PHB degradation	pPhaZa-A	pPhaZa-B	pPHBinDepol-KO	No
phaZ	MMG94_06000	Disrupt PHB degradation	pPhaZb-A	pPhaZb-B	pPHBdpol-KO	No
phaR	MMG94_18040	Disrupt PHB production	pPhaR-A	pPhaR-B		No
phas	MMG94_16150	Disrupt PHB production	pPhas-A	pPhas-B	pPhasin2-KO	No
phaf1	MMG94_03835	Disrupt PHB production	pPhaf1-A		pPhasin1-KO	No
phaf2	MMG94_16155	Disrupt PHB production	pPhaf2-A	pPhaf2-B		No
phaCb	MMG94_15895	Disrupt PHB production			pPHBC2-KO	No
phaCb	MMG94_15885	Disrupt PHB production			pPHBC3-KO	No
pMPar-1 *repA*	MMG94_19620	pMPar1 curing	pP1repA-A	pP1repA-B		Yes
pMPar-1 *repB*	MMG94_19625	pMPar1 curing	pP1repB-A	pP1repB-B		Yes
pMPar-2 *repA*	MMG94_20750	pMPar2 curing	pP2repA-A	pP2repA-B		Yes
pMPar-2 *repB*	MMG94_20755	pMPar2 curing	pP2repB-A	pP2repB-B		Yes
pMPar-2 Type II T/AT RelE/ParE family toxin	MMG94_21215	pMPar2 curing	pP2AT-T-A	pP2AT-T-B		No

pMPar-1 and pMPar-2 are the M. parvus OBBP megaplasmids[36]. The following notations are based on the PGAP annotation of the genome: *phaf*, phasin family protein, *phas*, phasin.

Six CRISPR plasmids targeting the three amino acid genes were produced and validated by Sanger sequencing. Testing of their editing capability however was limited by an inability to grow *M. parvus* on supplemental amino acids.

Selective plates produced as described were supplemented with 3mM of L-leucine, L-histadine or L-threonine this value being drawn from *E. coli* auxotroph experimental literature[51]. Each conjugation mix was plated onto selective media with its respective supplemented amino acid and a control on selective media. Two additional broad range amino acid supplemented selective plates were tested with either 0.5%$_{w/v}$ yeast extract or 2g/L casamino acids and 3mM tryptophan.

Growth appeared completely inhibited on all amino acid or yeast extract supplemented plates (Figure 22). Failure of growth on yeast extract at 0.5%$_{w/v}$ was noted in the original Whittenbury et al. 1970[57] paper but does not appear to have been explored since. A full test of minimum inhibitory concentration of yeast extract and various supplemental ammino acids would be illuminating. With this knowledge the inclusion of yeast extract in the mating plates of the conjugation method is curious and might bear improvement. This inability to grow in yeast extract was also tested in NMS liquid medium with addition at 0.5%w/v concentration and confirmed failure to grow for 2 weeks.

This limitation until solved, most likely by using a lower concentration of amino acid, halted progress towards engineering an amino acid auxotroph in *M. parvus*.

Figure 22 also illustrates a high number of escapees on the selective media. Correctly engineered colonies appear larger and outstanding and can be picked. This was ascribed to degradation of antibiotic selective kanamycin over the 3-4 week incubation period. This could also be due to a high number of CRISPR escapees. Colonies on the selective media were PCR amplified and run on a gel using the relevant PCR primers and showed no successful edits. This is as expected as successful auxotrophic edits would be unable to grow without their respective supplemental amino acid.

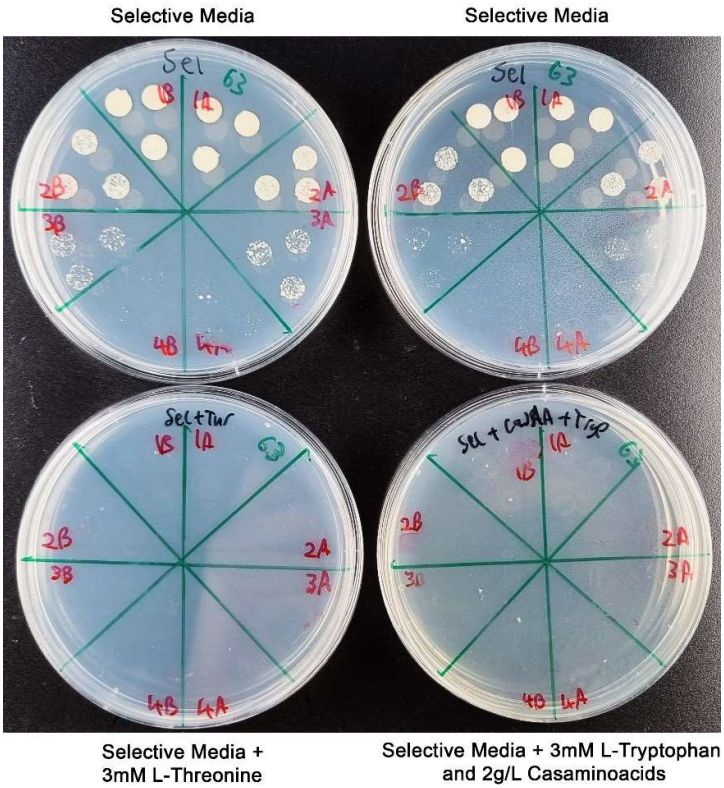


Figure 22: Four example selective media plates illustrating failure of growth with supplemental amino acids. Growth is observed up to 1000x dilution on the selective plates but the amino acid supplemented plates show no growth. The base selective medium is NMS agar with 50µg/ml Kanamycin and 25 µg Nalidixic acid. Supplements to this are listed. Plates have been inoculated with three 10µl spots of conjugation mixture per segment. Segments numbers 1-4 indicate dilution factors of 10^x i.e. 1A is the conjugation mixture dilute 10x. The left four segments and right four segments (A and B) of each plate are independent biological replicates. *M. parvus OBBP* on Yeast extract $0.5_{w/v}$ supplemented plates behaved similarly to the amino acid supplemented plates.

4.4.2.2 PHB Pathway Disruption

As established experimentally by Bashir Rumah in our publication[8], for *phaC* and by TraDIS for *phaABC*, the PHB synthesis genes are essential in *M. parvus*. Though this was not identified by the *M. parvus* genome scale model[1,58]. This is unusual compared to other strains like *C. necator* where PHB production has been successfully eliminated[59]. Other genes thought to regulate PHB production were found non-essential and so were targeted. This included the PHA repressor *phaR*, the annotated phasin *phas* and two "phasin family proteins" *phaf1* and *phaf2*. The two *phaZ* PHA depolymerase genes were found non-essential and were also targeted for KO. Their elimination could act to increase PHB accumulation by stopping its degradation. I theorise the non-essentiality of the *phaZ* genes identified in the TraDIS may be due to gene duplication-based redundancy. The PHB production pathway is discussed more completely in section 1.3.4. Plasmids provided by Bashir Rumah targeted two subsidiary *phaC* genes both identified by me as *phaCb* (section 3.4.3.6 and Figure 16A) both of which were identified as non-essential.

11 plasmids targeting six genes were produced and tested and produced colonies but no successful edits according to PCR and gel electrophoresis testing. This was repeated with the same results.

Six plasmids designed by Bashir Rumah against similar targets in *M. parvus* BRCS2 were supplied for use in *M. parvus* OBBP and are listed in Table 12 and the "plasmid 3" column of Table 13. These were supplied already transformed into S17-1 λpir and were conjugated and tested by me. pPhasin1-KO and pPHBinDepol-KO produced no edits. pPHBdpol-KO (EdE - 40%) pPhasin2-KO (32%) pPHBC2-KO (20%) pPHBC3-KO (76%) all produced successful and verified in frame deletions.

As described all 11 preliminary PHB gene targeted plasmids failed to produce edits, however, 4 of the 6 plasmids provided by Bashir Rumah succeeded. The plasmids supplied by Bashir Rumah were almost identical to the plasmids designed for the additional edits section of this work (except the two *phaCb* genes which I did not target) in terms of backbone, homology arms and arrangement but differing in seed. As the successful plasmids from Bashir Rumah were provided in the conjugation *E. coli* strain S17-1 λpir it seems likely the failure of my plasmids was due to this strain in my application. It is possible an additional plasmid was already present in the S17-1 λpir strain which was conjugated over providing kanamycin resistance, though this is unlikely as all S17-1 λpir transformed strains were miniprepped and the plasmids sequenced confirming they contained the correct plasmid before conjugation. The alternative is that the strain I used was not S17-1 λpir but another *E. coli* strain i.e. NEB 5-α or DH5α.

It is also possible this effected the amino acid auxotroph knockouts as they were carried out in parallel, but this could not be ascertained as the supplemental amino acid toxicity was more prevalent.

4.4.2.3 Plasmid Curing

As the native pMpar-2 megaplasmid contained only one identified T/AT system (section 3.4.3.7) this target was selected for removal. Prior work by other authors has shown in some cases removal of a T/AT system was sufficient for a megaplasmid to be cured from the strain[60]. As discussed previously the lack of other essential genes on pMPar-2 suggest this should be possible. The adjacent toxin gene MMG94_21215 was identified as non-essential by TraDIS but MMG94_21220, its theorised conjugate partner, was found mixed evidencing its potential functionality and activity in retaining the plasmid. As pMpar-1 consisted of 6 T/AT pairs multiple of which might be required for plasmid curing, editing of these genes was not attempted.

The *repA* and *repB* origins of replication and replication machinery were also targeted in both megaplasmids in case their targeted attack was sufficient for curing. For these genes no homology arms were included providing only CRISPR targeted attack. Success is not expected as TraDIS data indicated the plasmids essentiality, however if the T/AT elimination proved successful a second stage of targeted plasmid removal could be attempted in series.

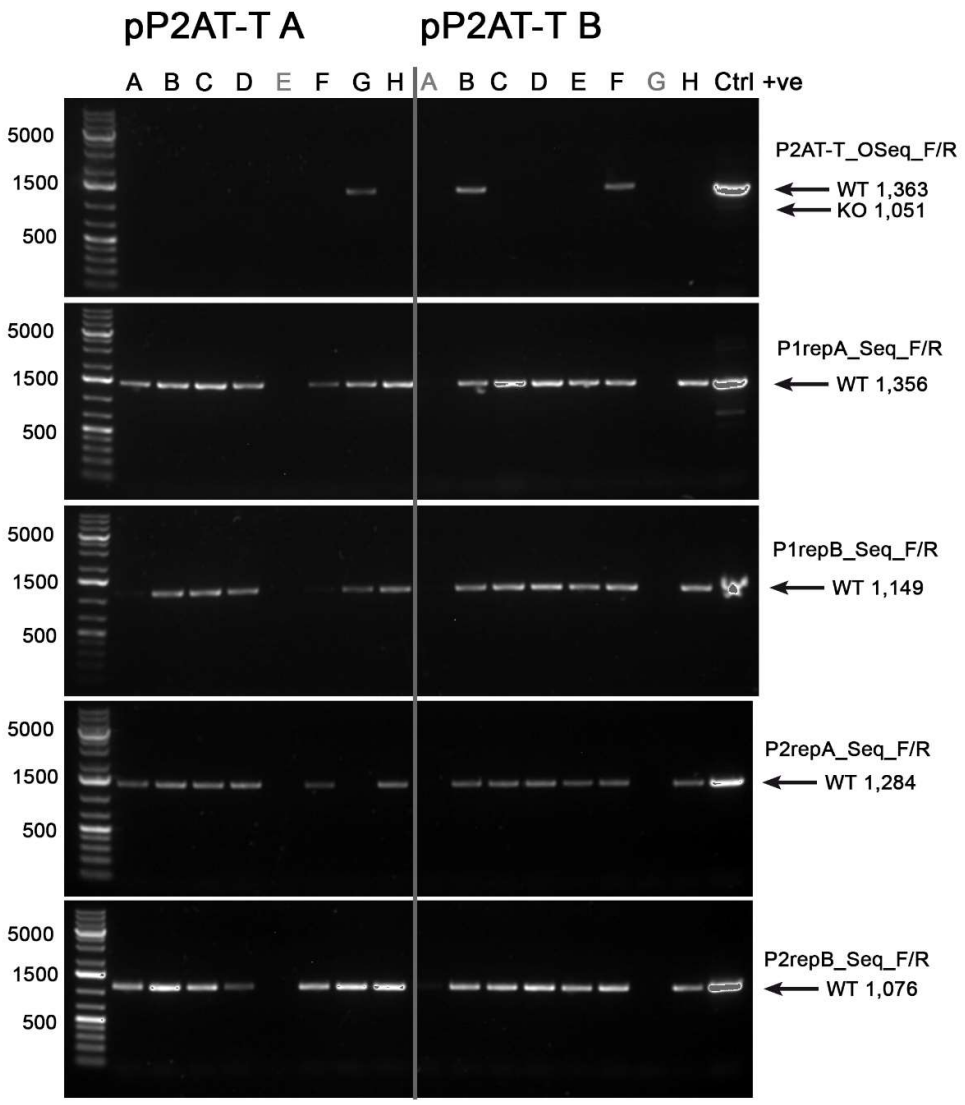


Figure 23: Gel showing PCR bands from attempted KO of pMPar-2 RelE/ParE Toxin using pP2AT-T A and B run on a 1% Agarose gel. The first row uses the correct primer pair for detecting Toxin knockouts. The remaining 4 rows utilise *repA* and *B* primer pairs against each plasmid. *repAB* on pMpar-1 were used to confirm if plasmid knockout had been achieved. pMpar-2 were used to confirm successful PCRs overall. Each column represented by a letter uses gDNA from one picked and grown colony. Eight colonies were picked for each KO plasmid. Three columns pP2AT-T: E and p2AT-T B: A and G are shown in gray as these are believed to be failed PCRs. Ctrl+ used unedited gDNA in the PCR and showed bands at the WT position as expected. Ladder is GeneRuler 1 kb Plus DNA Ladder.

The 12 plasmids were assembled successfully and tested by conjugation. Testing for curing by pP1repA, pP1repB, pP2repA and pP2repB showed the megaplasmids had not been cured. Prior work in curing the somewhat related plasmids in *Methylocystis sp.* SC2 by chemical exposure has also been unsuccessful leading to theories of their essentiality[26].Upon PCR testing pP2AT-T-A and B produced no bands in some PCRs. This could be indicative of complete elimination of the plasmid. To test these 8 resultant colonies for each plasmid were picked, grown, gDNA extracted and PCR checked again using the primers for *repA* and *B* in pMpar-1 and -2 for a total of five primer

pairs with a repeated test for P2AT-T induced KO (Figure 23). Bands appeared for the *repA and repB* primers in all lanes indicating no plasmid curing occurred (three lanes with both *repA* and *repB* missing were classed as PCR failures). The control showed positive bands, as did other lanes but p2AT-T bands were absent in some lanes . This may be indicative of deletion which the PCR did not capture if so p2AT-T was deleted at an EdE of 85.7% by pP2AT-T-A and 66.6% by pP2AT-T-B. Without a redesigned primer pair to repeat this test and indicate smaller deletion bands this cannot be confirmed.

4.4.3 CRISPR Part 3 – Native CRISPR

A native CRISPR (ntvCRISPR) implementation was explored which would allow editing of the genome without providing an external Cas9 protein. As the Cas9 is a large gene (4107bp, 35% of the plasmid - Figure 18) its elimination would shrink the plasmid considerably increasing CE and counts of recombinant cells[30]. As the ntvCRISPR is regulated by the bacteria it should avoid cytotoxicity and thus increase surviving cells that could go on to be successful edits[31].

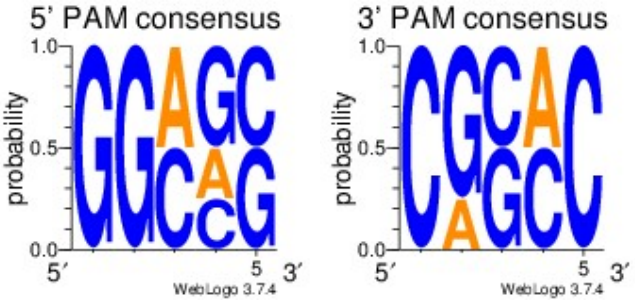
4.4.3.1 Identifying potential PAMs

Margaux Delavelle developed a tool to identify native CRISPR (ntvCRISPR) PAMs (unpublished). It matches spacers in the arrays against sequences in a database, those likely to be of interest here limited to phage either identified free living or as predicted prophage. These are selected as protospacers and the region upstream of the target was collected to form a pool of possible PAMs. The most likely bases in the PAM are identified by repeat presence at that position.

This was applied to the *M. parvus* OBBP on two CRISPR clusters utilising 85 spacers on one array and 53 spacers on the other array. The resulting bases in the PAM regions are illustrated in Figure 24. The array locations are indicated in the figure. Collectively the arrays produced 13 matches to known phage. Of these 7 were associated with *Rhizobium* genera hosts, and 2 to *Bosea,* both of these genera being related to *M. parvus* at the order level (Hyphomicrobiales). The closeness and prevalence of these relations suggest these are real hits to phage that may interact with *M. parvus*.

The predicted PAM regions of Array 1 5’ and 3’ and Array 2 5’ are GC rich. The PAM for the two arrays would be expected to be similar as the PAM is defined by the Cas effector however the two arrays did not form convincing similarity. As it consisted of the most phage datapoints (n=9) the PAMs indicated by array 2 was selected for further testing.

CRISPR Array 1: 1,590,999-1,596,735bp n=4



CRISPR Array 2: 2,360,214-2,363,776bp n=9

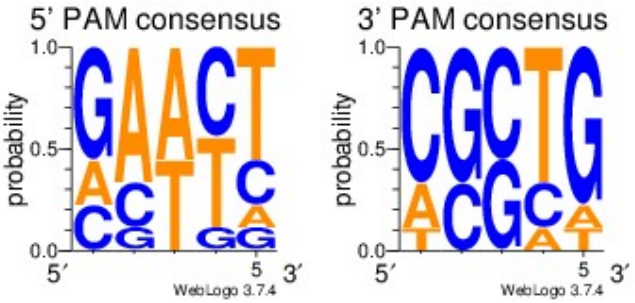


Figure 24: Four consensus graphs showing cumulative probability of bases within the likely PAM regions at the 3' and 5' ends of genomic regions matched to spacers in *M. parvus* OBBP CRISPR arrays. These are used to indicate likely ntvCRISPR PAM sequences. The two CRISPR arrays were analysed separately. Array 1 consisted of 85 spacers producing 68 matches of which 4 were phage and used in this analysis. Array 2 consisted of 53 spacers producing 106 matches of which 9 were phage and used in the analysis. The figures were produced by Margaux Delavelle. Positions refer to the CRISPR Array locations in *M. parvus* OBBP Genbank Accesion: CP092968.

PAM testing followed prior work in *Clostridium* spp[38,61] a "plasmid invader assay". A sequence was assembled containing an *M. parvus* spacer selected from those identified, with an upstream and downstream leader. The leader could be a prospective PAM or a native control sequence as shown in Table 14. Controls were formed by inserting known non-PAM sequences from the start and end of the repeat regions in its place 5'- GAAAC – spacer – GTCG(T/C) – 3' i.e. if this was the PAM the ntvCRISPR would self-target its own genome which would be lethal. If the correct PAM is identified the plasmid will be unable to survive in *M. parvus* OBBP as the plasmid will be cut by the ntvCRISPR and thus will not conjugate. If the PAM is incorrect the plasmid will conjugate. The control should always conjugate. The bases adjacent to the tested PAMs were 5' AT-insert region-GA 3' and is listed in case of effect on the overall PAM recognition if it proves to be 4 or 5bp.

It has been specified in the literature that PAM position, 3' or 5' is dependent on CRISPR type with Type I as 5′-PAM-protospacer-3′ and Type II being 5′-protospacer-PAM-3′[38,61]. This would suggest an expected 5' PAM in the *M. parvus* system which has a Type IC (section 3.4.3.2). Prior work has tested both PAM positions in Type I systems[38] and due to this and strong

signal at the 5' end (Figure 24) it was concluded both should be tested. Three 5' possibilities were tested and two 3' PAMs. These are identified in Table 14.

Table 14: Tested PAMs used in this work.

Plasmid	PAM	Insert Region
pAra3-36-Ntv	Native Control	**GAAAC***CGCGCGACGATCGCGGCGATCCCAGCGACGCCCGC***GTCGT**
pAra3-36-TTC	5-TTC	**TTC***CGCGCGACGATCGCGGCGATCCCAGCGACGCCCGC*
pAra3-36-ATT	5-ATT	**ATT***CGCGCGACGATCGCGGCGATCCCAGCGACGCCCGC*
pAra3-36-ACT	5-ACT	**ACT***CGCGCGACGATCGCGGCGATCCCAGCGACGCCCGC*
pAra3-36-3-CGC	3-CGC	*CGCGCGACGATCGCGGCGATCCCAGCGACGCCCGC***CGC**
pAra3-36-3-CGG	3-CGG	*CGCGCGACGATCGCGGCGATCCCAGCGACGCCCGC***CGG**

PAMs are identified in bold and underlined. The protospacer is in italics and is identical in each insert region and was chosen from the existing repeats in Array 2 2,362,579 – 2,362,613bp in *M. parvus* OBBP Genbank Accession: CP092968.

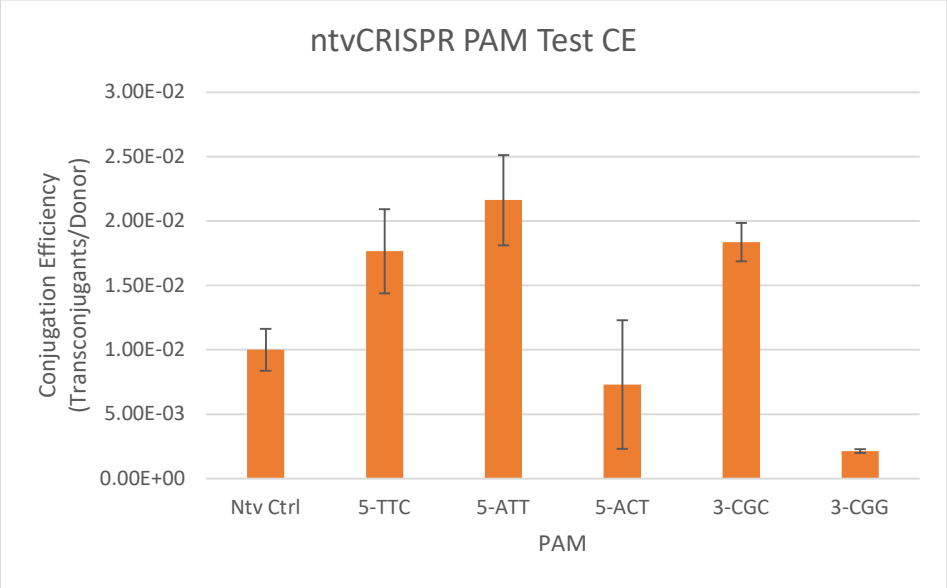


Figure 25: Conjugation efficiency results of the conjugation of six ntvCRISPR PAM testing plasmids. Results are the mean of three biological replicates of three 10µl spot counts except 5-ACT which is drawn from two. Error bars are 95%CL.

4.4.3.2 Conjugation of ntvCRISPR PAMs

The six ntvCRISPR PAM testing plasmids were conjugated into *M. parvus OBBP* and colony counts of donors and recipients were evaluated to give the CE (Figure 25). This indicated no significant deviation from the native control. A correctly identified PAM would be expected to not conjugate successfully as the native CRIPSR machinery would cut the plasmid in the exconjugant removing antibiotic resistance and result in no colonies as was observed by other authors in *Clostridia*[38]. The data indicate most tested PAMs were not correct for the *M. parvus* OBBP ntvCRISPR system actually showing higher CE than the native control, 5' ACT had variable data. The 3' CGG PAM (CE = $2.13x10^{-3}$) was significantly lower than the native control of CE=$1x10^{-2}$ (Tukey's

test p=0.005) suggesting a possible success however the value is still higher than the total lack of conjugation seen by other authors in correct PAMs, or the 10^{-7} observed with CRISPR KO plasmids in Figure 24E. The 3' position is also unexpected for Type I CRISPR systems. From this I conclude this is likely not the correct PAM.

4.4.3.3 Discussion ntvCRISPR

It has not been confirmed that the native CRISPR system in *M. parvus* is functional and it remains possible it would not react to any PAM explaining the above results. It is considered more likely however the correct PAM was not tested. The testing here was limited in scope and non-exhaustive, testing only the most obvious signals. Further testing may yet identify the correct PAM based on the current data including expanding to 5bp PAMs. As prediction of PAMs is dependent on identification of phage to either match in the database directly or matching prophage this is a limiting factor. This form of analysis will prove more effective as databases of phage and prophage identification improves.

A better plasmid design would have utilised the native sequence at the 3' or 5' end of the spacer if a PAM was not tested in that location in case the adjacent region influenced the results and incidentally contained an active PAM. This was however of no concern as all plasmids conjugated. Only one protospacer was tested and it remains possible this protospacer was not correctly recognised. Prior work in *Clostridium*[38] used three different protospacers for PAM testing and found the same success and failures in all three indicating this is at least not a prevalent issue.

Had the ntvCRISPR PAM identification experiment succeeded the next step would have been to insert an artificially expressed CRISPR array targeting a genomic region which could be used in concert with a homology repair cassette to make genomic edits. Alternatively a confirmational test could be performed with a plasmid borne artificial CRISPR array targeting a protospacer also on the plasmid, providing self-elimination with the correct PAM choice[61].

4.5 General Conclusions

The CRISPR system was proved effective against nine genes (*ligD*, *pntA*, *bcsB*, MPA_0518, *phbC2*, *phaCb*, *phas*, *phaZb* and potentially pMPar-2 RelE/ParE Toxin) using 22 seeds and an overall EdE of 52.76%. These form the first genetic editing carried out in *M. parvus* OBBP to date. The high reliability found against the first four genes in Part 1 was not repeated in Part 2 due to a range of factors likely unrelated to the underlying CRISPR system. Developing the individual testing methods and rectifying issues in conjugation will likely lead to better results in auxotroph production and PHB pathway editing in Part 2 if approached again.

M. parvus incapability to grow on a range of amino acids, casaminoacids or with yeast extract was strongly evidenced, however a full minimum inhibitory concentration test would be an obvious step to find the lower bounds of this. Until that has been identified work on developing amino acid auxotroph's remains impossible.

Due to time constraints phenotypic testing on the edited strains could not be carried out but testing particularly for cell morphology and PHB accumulation would be vital to further understanding of the PHB pathway in *M. parvus* OBBP especially with a potential to increase in PHB accumulation from *phaZb* deletion.

Reliability of CE testing was challenging. Successful CRSIPR edited colonies appeared larger than unsuccessful edits but there was often a lawn of smaller colonies on the selection plates, and these may be mixed in lowering EdE. I suggest this is due to a satellite colony effect where the correctly transformed colonies degrade the selective kanamycin in surrounding areas as wild type *M. parvus* OBBP would not grow naturally on the selective media. This effect is likely exacerbated by the required extended growing time for methanotroph cultivation on solid media which allows time for antibiotic degradation. Alternative antibiotics may prove more effective in *M. parvus* editing.

The PAM of the endogenous *M. parvus* CRISPR system was not successfully identified. However, with the success of the Cas9 system demonstrated in this chapter with high EdE, abating concerns of cytotoxicity found by the previous two methanotroph CRISPR papers, the ntvCRISPR system is not essential for the continuation of editing work in *M. parvus* OBBP. It could however still be of use in increasing efficiency of multiplexed edits.

4.6 References

1. Bordel S, Rojas A, Muñoz R. Reconstruction of a Genome Scale Metabolic Model of the polyhydroxybutyrate producing methanotroph Methylocystis parvus OBBP. Microb Cell Fact. 2019;18(104):1–11.
2. del Cerro C, García JLJM, Rojas A, Tortajada M, Ramón D, Galán B, et al. Genome sequence of the methanotrophic poly-β-hydroxybutyrate producer Methylocystis parvus OBBP. J Bacteriol. 2012 Oct 15;194(20):5709–10.
3. Bordel S, Rodríguez Y, Hakobyan A, Rodríguez E, Lebrero R, Muñoz R. Genome scale metabolic modeling reveals the metabolic potential of three Type II methanotrophs of the genus Methylocystis. Metab Eng. 2019;54(December 2018):191–9.
4. Lieven C, Petersen LAH, Jørgensen SB, Gernaey K V., Herrgard MJ, Sonnenschein N. A Genome-Scale Metabolic Model for Methylococcus capsulatus (Bath) Suggests Reduced Efficiency Electron Transfer to the Particulate Methane Monooxygenase. Front Microbiol. 2018;9(December):1–15.
5. Torre A, Metivier A, Chu F, Laurens LML, Beck DAC, Pienkos PT, et al. Genome-scale metabolic reconstructions and theoretical investigation of methane conversion in Methylomicrobium buryatense strain 5G(B1). Microb Cell Fact. 2015;14(1):1–15.
6. Akberdin IR, Thompson M, Hamilton R, Desai N, Alexander D, Henard CA, et al. Methane utilization in Methylomicrobium alcaliphilum 20ZR: A systems approach. Sci Rep. 2018;8(1):1–13.
7. Kalyuzhnaya MG, Puri AW, Lidstrom ME. Metabolic engineering in methanotrophic bacteria. Metab Eng. 2015;29:142–52.
8. Rumah BL, Claxton Stevens BH, Yeboah JE, Stead CE, Harding EL, Minton NP, et al. In Vivo Genome Editing in Type I and II Methanotrophs Using a CRISPR/Cas9 System. ACS Synth Biol. 2023 Feb 17;12(2):544–54.
9. Khmelenina VN, But SY, Rozova ON, Oshkin IY, Pimenov N V., Dedysh SN. Genome Editing in Methanotrophic Bacteria: Potential Targets and Available Tools. Microbiol (Russian Fed. 2022;91(6):613–30.
10. Henard CA, Guarnieri MT. Metabolic Engineering of Methanotrophic Bacteria for Industrial Biomanufacturing. In: Kalyuzhnaya MG, Xing X-H, editors. Methane Biocatalysis: Paving the Way to Sustainability. Springer International Publishing; 2018. p. 117–32.
11. Baani M, Liesack W. Two isozymes of particulate methane monooxygenase with different methane oxidation kinetics are found in Methylocystis sp. strain SC2. Proc Natl Acad Sci U S A. 2008 Jul 22;105(29):10203–8.
12. Nguyen DTN, Lee OK, Hadiyati S, Affifah AN, Kim MS, Lee EY. Metabolic engineering of the type I methanotroph Methylomonas sp. DH-1 for production of succinate from methane. Metab Eng. 2019;54(March):170–9.
13. Crombie A, Murrell JC. Development of a system for genetic manipulation of the facultative methanotroph Methylocella silvestris BL2. 1st ed. Vol. 495, Methods in Enzymology. Elsevier Inc.; 2011. 119–133 p.
14. Rumah BL. Isolation, characterisation and development of genetic tools for methane-oxidising bacteria. PhD Thesis. 2018;
15. Martin H, Murrell J. Methane monooxygenase mutants of Methylosinus trichosporium constructed by marker-exchange mutagenesis. FEMS Microbiol Lett. 1995 Apr;127(3):243–8.

16. Jeong J, Kim TH, Jang N, Ko M, Kim SK, Emelianov G, et al. A highly efficient and versatile genetic engineering toolkit for a methanotroph-based biorefinery. Chem Eng J. 2023;453(P2):139911.
17. Stafford GP, Scanlan J, McDonald IR, Murell JC. rpoN, mmoR and mmoG, genes involved in regulating the expression of soluble methane monooxygenase in Methylosinus trichosporium OB3b. Microbiology. 2003;149(7):1771–84.
18. Lloyd JS, Finch R, Dalton H, Murrell JC. Homologous expression of soluble methane monooxygenase genes in Methylosinus trichosporium OB3b. Microbiology. 1999;145(2):461–70.
19. Henard CA, Smith H, Dowe N, Kalyuzhnaya MG, Pienkos PT, Guarnieri MT. Bioconversion of methane to lactate by an obligate methanotrophic bacterium. Sci Rep. 2016 Feb 23;6(1):21585.
20. Tapscott T, Guarnieri MT, Henard CA. Development of a CRISPR/Cas9 system for Methylococcus capsulatus in vivo gene editing. Appl Environ Microbiol. 2019;85(11):e00340-19.
21. Lloyd JS, De Marco P, Dalton H, Murrell JC. Heterologous expression of soluble methane monooxygenase genes in methanotrophs containing only particulate methane monooxygenase. Arch Microbiol. 1999;171:364–70.
22. Lee HM, Ren J, Yu MS, Kim H, Kim WY, Shen J, et al. Construction of a tunable promoter library to optimize gene expression in Methylomonas sp. DH-1, a methanotroph, and its application to cadaverine production. Biotechnol Biofuels. 2021;14(1):1–12.
23. Methanotroph Commons. Genetics | Methanotroph Commons [Internet]. Available from: http://www.methanotroph.org/wiki/genetics/
24. Lidstrom ME, Korotkova ; N, Chistoserdova L, Kuksa V. Poly-Hydroxybutyrate Biosynthesis in the Facultative Methylotroph Methylobacterium extorquens AM1: Identification and Mutation of gap11, gap20, and phaR. J Bacteriol. 2002;184:1750–8.
25. Yan X, Chu F, Puri AW, Fu Y, Lidstrom ME. Electroporation-based genetic manipulation in type I methanotrophs. Appl Environ Microbiol. 2016;82(7):2062–9.
26. Dam B, Kube M, Dam S, Reinhardt R, Liesack W. Complete sequence analysis of two methanotroph-specific repabc containing plasmids from methylocystis sp. Strain SC2. Appl Environ Microbiol. 2012 Jun;78(12):4373–9.
27. Puri AW, Owen S, Chu F, Chavkin T, Beck DAC, Kalyuzhnaya MG, et al. Genetic tools for the industrially promising methanotroph Methylomicrobium buryatense. Appl Environ Microbiol. 2015;81(5):1775–81.
28. Borodina E, Nichol T, Dumont MG, Smith TJ, Murreil JC. Mutagenesis of the “leucine gate” to explore the basis of catalytic versatility in soluble methane monooxygenase. Appl Environ Microbiol. 2007;73(20):6460–7.
29. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol. 2013;31(3):233–9.
30. Arroyo-Olarte RD, Bravo Rodríguez R, Morales-Ríos E. Genome editing in bacteria: Crispr-cas and beyond. Microorganisms. 2021;9(4).
31. Zheng Y, Li J, Wang B, Han J, Hao Y, Wang S, et al. Endogenous Type I CRISPR-Cas: From Foreign DNA Defense to Prokaryotic Engineering. Front Bioeng Biotechnol. 2020 Mar 4;8(March):1–17.
32. Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature.

2018;556(7699):57–63.
33. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013;31(9):827–32.
34. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. 2016;34(2):184–91.
35. Cui Y, Xu J, Cheng M, Liao X, Peng S. Review of CRISPR/Cas9 sgRNA Design Tools. Interdiscip Sci – Comput Life Sci. 2018;10(2):455–65.
36. Claxton Stevens BH, Rumah BL, Stead CE, Zhang Y. A Complete Genome of the Alphaproteobacterial Methanotroph Methylocystis parvus OBBP. Klepac-Ceraj V, editor. Microbiol Resour Announc. 2023 Apr 18;12(4).
37. Haopeng Yu, Wu Z, Chen X, Ji Q, Tao S. CRISPR-CBEI : a Designing and Analyzing Tool Kit for Cytosine. mSystems. 2020;5:e00350-2(September):1–9.
38. Pyne ME, Bruder MR, Moo-Young M, Chung DA, Chou CP. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. Sci Rep. 2016 May 9;6(1):25666.
39. Burbidge A, Minton NP. SBRC-Nottingham: sustainable routes to platform chemicals from C1 waste gases Background to the CRG. Biochem Soc Trans. 2016;44(3).
40. Heap JT, Pennington OJ, Cartman ST, Minton NP. A modular system for Clostridium shuttle plasmids. J Microbiol Methods. 2009 Jul 1;78(1):79–85.
41. Zhang Y, Grosse-Honebrink A, Minton NP. A Universal Mariner Transposon System for Forward Genetic Studies in the Genus Clostridium. Paredes-Sabja D, editor. PLoS One. 2015 Apr 2;10(4):e0122411 1-21.
42. Durland RH, Toukdarian A, Fang F, Helinski DR. Mutations in the trfA replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy numbers. J Bacteriol. 1990;172(7):3859–67.
43. Jahn M, Vorpahl C, Hübschmann T, Harms H, Müller S. Copy number variability of expression plasmids determined by cell sorting and droplet digital PCR. Microb Cell Fact. 2016;15(1):1–12.
44. New England BioLabs. NEBCloner - RE Digest [Internet]. Available from: https://nebcloner.neb.com/#!/redigest
45. Benchling. Benchling.
46. New England Biolabs. Making your own chemically competent cells [Internet]. 2020 [cited 2020 Sep 1]. Available from: https://international.neb.com/protocols/2012/06/21/making-your-own-chemically-competent-cells
47. New England BioLabs. High Efficiency Transformation Protocol (C2987H) V.1. protocols.io. 2014;
48. New England BioLabs. NEBaseChanger. 2021.
49. Rumah BL, Stead CE, Claxton Stevens BH, Minton NP, Grosse-Honebrink A, Zhang Y. Isolation and characterisation of Methylocystis spp. for poly-3-hydroxybutyrate production using waste methane feedstocks. AMB Express. 2021 Dec 6;11(6):1–13.
50. Han D, Link H, Liesack W. Response of Methylocystis sp. strain SC2 to salt stress: Physiology, global transcriptome, and amino acid profiles. Appl Environ Microbiol. 2017;83(20):1–14.
51. Bertels F, Merker H, Kost C. Design and characterization of auxotrophy-based amino acid biosensors. PLoS One. 2012;7(7).

52. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006 Jan 21;2(1).
53. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, et al. The BioCyc collection of microbial genomes and metabolic pathways. Brief Bioinform. 2018;20(4):1085–93.
54. Karp PD, Paley SM, Midford PE, Krummenacker M, Billington R, Kothari A, et al. Pathway Tools version 23.0: Integrated Software for Pathway/Genome Informatics and Systems Biology. Brief Bioinform. 2015 Oct 14;11(1):40–79.
55. Yomantas YAV, Tokmakova IL, Gorshkova N V., Abalakina EG, Kazakova SM, Gak ER, et al. Aromatic amino acid auxotrophs constructed by recombinant marker exchange in methylophilus methylotrophus AS1 cells expressing the aroP-Encoded transporter of escherichia coli. Appl Environ Microbiol. 2010;76(1):75–83.
56. Stumpe J. US10995316B2: Proline auxotrophs. United States; US10995316B2, 2016.
57. Whittenbury R, Phillips KC, Wilkinson JF. Enrichment, Isolation and Some Properties of Methane-utilizing Bacteria. J Gen Microbiol. 1970;(61):205–18.
58. Bordel S. ModelsMethanotrophs GitHub repository. 2020.
59. Pfeiffer D, Jendrossek D. Localization of poly(3-Hydroxybutyrate) (PHB) granule-associated proteins during PHB granule formation and identification of two new phasins, phap6 and phap7, in Ralstonia eutropha H16. J Bacteriol. 2012 Nov;194(21):5909–21.
60. Lau MSH, Sheng L, Zhang Y, Minton NP. Development of a Suite of Tools for Genome Editing in Parageobacillus thermoglucosidasius and Their Use to Identify the Potential of a Native Plasmid in the Generation of Stable Engineered Strains. ACS Synth Biol. 2021;10(7):1739–49.
61. Maikova A, Kreis V, Boutserin A, Severinov K, Soutourina O. Using an Endogenous CRISPR-Cas System for Genome Editing in the Human Pathogen Clostridium difficile. Atomi H, editor. Appl Environ Microbiol. 2019 Oct 15;85(20):1–15.

Chapter 5: Use of Real World Biogas and Comparison of PHB Producing Strains

5.1 Introduction

Methanotroph based biodegradable bioplastic production provides progress against two important anthropogenic pollutants: non-degrading plastic waste and waste methane. These are both current major environmental issues. A circular carbon economy has been envisaged in which biogas is used as the feedstock for industrial methanotroph fermentation, producing biodegradable bioplastics, which at the end of their lifespan can be degraded by anaerobic digestion into more biogas and re-enter the cycle[1] (Figure 26). This provides a net-zero carbon solution. Though it must be noted recycling and reuse are more energy efficient methods of reducing plastic waste, many angles of approach are needed to tackle the issue[2].

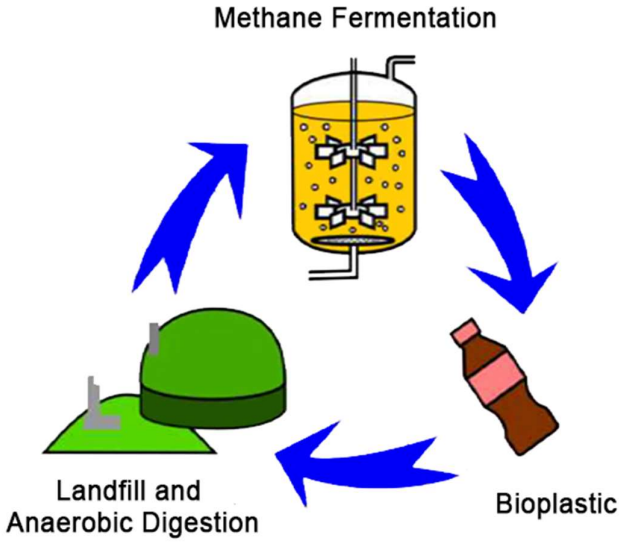


Figure 26: Illustration of the Circular economy of methanotroph based bioplastic production. Adapted from a figure by Alexander Grosse-Honebrink.

Biogas from anaerobic digesters and landfills consist of a mixture of methane (CH_4), carbon dioxide (CO_2) and nitrogen (N_2), with traces of other compounds such as hydrogen sulphide (H_2S), siloxanes and aromatic and halogenated compounds[3,4]. Biogas composition is highly dependent on waste composition, temperature and moisture among other factors, and can thus vary between different anaerobic digestor (AD) facilities and landfill sites[3,4]. Concerns have been raised over the use of biogas due to the possibility of H_2S toxicity and CO_2 acidity along with many others[5,6]. Other compounds especially from landfill biogas may source from the waste itself e.g. pesticides in agricultural waste[3]. This, along with economic viability have led to hesitancy in progression of the technology. Though one industrial implementation by

the Technical Research Centre of Finland (VTT) has been reported[7,8] there has been no news since 2016.

Most studies reported to date on methanotrophic PHB production have focused on the use of pure methane or natural gas with a small amount using artificial biogas as substrate leaving the renewable sources of methane open to investigation [6,9–12]. Only one study has properly investigated biogas growth in pure culture methanotrophs[13] despite many reviews and life cycle assessments on the subject[1,14–17].

Table 15: Constituents of Biogas and Natural Gas

	General gas makeup		Gases used in this study	
	Biogas from anaerobic fermentation[18]	**Natural Gas**[18]	**Anaerobic Digestor Direct (AD-Dir)**[a]	**Anaerobic Digestor Carbon Filtered (AD-Filt)** [a]
Methane	50-85%	83-98%	62.4%	58.0%
Carbon Dioxide	15-50%	0-1.4%	40.5%	38.0%
Nitrogen	0-1%	0.6-2.7%	0.0	0.0
Oxygen	0.01-1%	-	0.5%	0.4%
Hydrogen (ppmv)	trace	-	23.0	16.0
Hydrogen Sulphide (ppmv)	Up to 4000 ppmv	-	682.0	34.0
Ammonia	trace	-	0	0
Ethane	-	Up to 11%	-	
Propane	-	Up to 3%	-	-
Siloxane	0-5 mg/m^3	-	-	-

Values percent by volume unless otherwise noted. ppmv-parts per million by volume. [a]Data for samples for this work. Data taken by Lower Reule Bioenergy Ltd. Values for AD-Dir total 103.4% and AD-Filt total 96.4%. In publication[19] AD-Dir is AD3 and AD-Filt is AD4. Other From Lambert 2017[18]

5.1.1 Methane and Biogas as a Feedstock

After carbon dioxide, methane is the most important greenhouse gas in terms of damage[20] and up to 60% of world methane output is anthropogenic[6] with major efforts being carried out to reduce that including control under the Kyoto protocol. The major production sources in the UK being landfill waste (35.6%) and agriculture (43.6%) in 2010[21]. Both these sources have great potential for reduction in the UK and abroad improving humanities environmental footprint[22]. Captured methane containing gases from landfill and wastewater treatment along with gases from AD using agricultural and food waste, are collectively called biogas. The methane available here is an efficient use of the degraded carbon; up to 50%, is converted into methane during anaerobic degradation[23] giving end methane percentages of 50-85% for anaerobic fermenters and ~45% for landfill gas[18](Table 15). Biogas below an economically valuable methane composition is considered “waste gas” which would normally be flared[24] (burned on site as disposal) or released to atmosphere resulting in greenhouse gas emission either as CO_2 or methane[25].

Compared to other fuel sources methane is very cheap and widely available as both natural gas and from biogas[26,27]. Biogas can be used for energy generation and heating directly providing a renewable energy source

however it still results in CO_2 emissions and is considered low quality and inefficient with a lower methane percentage than natural gas (Table 15)[18,28]. To rectify this specialised equipment can enrich methane content into a form equivalent to natural gas called biomethane for use in transport fuel and energy generation directly however this adds cost[18,28]. Biomethane production prices ranging from 0.5 to 1.5 US$/m^3 with an average of 0.65 US$/m^3 depending on source, struggles to compete in cost with the natural gas prices[18,28]. Recent natural gas pricing history for the last 10 years fluctuations range 0.30-0.06, Mean 0.12, Median 0.10 US$/m^3 putting biomethane currently out of economic feasibility without additional incentives (To end 2022, Henry Hub Spot Price[29] - Monthly Average, Conversion 1 Metric Million British thermal units -MMBTU = 29.31 m^3 [30]). Outside the US market natural gas pricing is significantly higher for example Europe (2018) is closer to 0.24US$/m^3[28]. It is estimated in 2018 only 5% of potential biomethane production was being achieved compared to available fermentable waste, without any competition for food or agricultural land, thus the potential for economy of scale in production is large but unexploited[28].

Unenriched biogas as a feedstock for industrial methanotrophs making bioplastics, single cell protein, methanol or other higher value products has the potential to add value to biogas without the need for enrichment and provides carbon capture that can enter the carbon cycle without venting CO_2 to the atmosphere or entering long term artificial storage[31]. "Waste gas" that is too low in methane for economically viable enrichment can still be harnessed by methanotrophs[6]. Despite this potential, at an industrial scale natural gas is used as a methanotroph feedstock rather than biogas (e.g. Calysta[32]), failing to deliver on the potential of the technology. This is partly due to the previously mentioned cost difference and due to the variability of biogas content which can add uncertainty to industrial outcomes[3,27]. The situation is expected to change in the future with natural gas demand forecast to stabilise or increase marginally out to 2050 and beyond and with increasing renewables usage which might see a fall in price[33]. This would lower the feedstock costs of biogas for methanotrophic growth further improving economic viability. Ensuring the viability of PHB production and methanotroph growth in general on biogas at lab scale will prepare for this future. Thus a fermentation experiment was carried out examining growth and PHB production on real world AD biogas sources using *M. parvus* OBBP, three novel isolated *Methylocystis* strains and *M. capsulatus* Bath as a PHB negative control. A preliminary experiment confirming time taken for PHB production was also carried out.

Work in this chapter was published as part of "Isolation and characterisation of *Methylocystis spp.* for poly-3-hydroxybutyrate production using waste methane feedstocks" Rumah et al. 2021[19]. A full copy of this paper is presented at the end of this thesis.

5.2 Methods

5.2.1 Strains Used

In addition to *M. parvus* OBBP four additional strains were utilised described in Table 16. BRCS1, BRCS2 and Isolate 3* (Iso 3*) were isolated and provided by Bashir Rumah and Chris Stead.

Table 16: Additional methanotroph strains utilised in Chapter 5

Species	Strain	Properties	Source
Methylococcus capsulatus	Bath	γ-proteobacterial PHB -ve Control	ATCC - 33009
Methylocystis rosea	BRCS1	α-proteobacterial PHB +ve	Bashir Rumah and Chris Stead[19] (SBRC5822)[b] – Isolated from a Recreational Lake Nottingham University, UK
Methylocystis parvus	BRCS2	α-proteobacterial PHB +ve	Bashir Rumah and Chris Stead[19] (SBRC5824)[b] - – Isolated from a Bog in Mosley, UK
Methylocystis rosea	"Isolate 3*"	Expected α-proteobacterial PHB +ve	Bashir Rumah and Chris Stead (Thesis)[34]
[b] – Available from the SBRC Nottingham culture collection (https://store.nottingham.ac.uk/product-catalogue/schools-and-departments/synthetic-biology-research-centre) **or from NCIMB as numbers 15262 and 15263**			

Genomes of BRCS1, BRCS2 and Isolate 3* were supplied by Bashir Rumah and were used to identify the closest extant strains. The first two were closed and described in literature[19] the latter was in 100 contigs. A BLAST search was performed on a 1400bp section following on from primer 16S_27F which contained the 16S rRNA. This resulted in BRCS1 and Isolate 3* exactly matching (100% coverage and identity) *Methylocystis rosea* (Strain GW6 - CP034086.1) and BRCS2 matching *M. parvus* (OBBP - CP092968.1). Additional close matches were found within *Methylocystis* all with some lack in coverage or identity. From this, species identities were concluded presented in Table 16. Full genome alignments carried out using Mauve with their closest published relatives showed BRCS1 to share 94.96% identity with *M. rosea* GW6 and BRCS2 had 99.99% identity with *M. parvus* OBBP[19].

5.2.2 Anaerobic Digestor Gas

Gas was acquired in from Lower Reule Bioenergy Ltd, UK from their Staffordshire plant 26/2/2019 in Supelco 2 L Tedlar bags (Sigma-Aldrich, UK). Gas measurement carried out by Lower Reule Bioenergy Ltd are shown in Table 15. Lower Reule Bioenergy generally uses agricultural waste and household and industrial food waste[35].

AD-Dir was taken directly from the AD biogas collector. AD-Filt was taken post treatment as follows: a condensate pit to remove moisture and biological scrubbers to remove further H_2S, through a chiller and finally a carbon filter to remove any excess H_2S. This did result in a decreased H_2S level in AD-Filt

(34.0ppmv) compared to AD-Dir (682ppmv). However, the gases are also from different periods of fermentation so other compositions may also differ.

5.2.3 Fermentation

In both experiments fermentation was carried out according to the two-stage method (section 2.1.3) unless otherwise noted with 35ml media in 250ml bottles. They each utilised 5 strains, *M. parvus* OBBP and the 4 listed in Table 16.

Exp. 1 compared strains and PHB over time, 10 bottles for each strain were set up for a total of 50 samples. Bottles were grown for 3 days without regassing before resuspension in nfNMS, OD readings taken and regassed. There was no regassing on subsequent days. After resuspension (D0) each day, one pair (making two replicates) of each strain was sacrificed for OD and PHB testing giving a timecourse of PHB accumulation over 4 days (D0-D3). Two replicates (BRCS1, D2-B and BRCS2 D2-B) were discarded due to lack of growth thus these datapoints use single samples.

In Exp. 1 OD increase from before resuspension in nfNMS to the point of bottle sacrifice was calculated rather than raw OD. This was used as each data point was an independent bottle to remove noise caused by variation in starting OD. These data appeared considerably less disordered than raw OD values. In later experiments in this thesis a pooled resuspension inoculum was used making initial conditions uniform.

In Exp. 2 three replicates of each gas/strain combination were carried out using the 4 α-proteobacterial species for a total of 36 samples. 54ml of either pure CH_4, AD-Dir or AD-Filt was added to the headspace at each regassing step from inoculation. This resulted in less CH_4 in the biogas bottle headspace compared to the pure CH_4 but as oxygen limitation is the major factor (section 6.4.1) influence should be minimal. This did allow for maximal effect of any biogas to be observed. Calculated O_2:CH_4 ratios: Pure CH4 - 0.83, AD-Dir - 1.34, AD-Filt - 1.44. Bottles were regassed two days after inoculation and were resuspended in nfNMS two days after that (D0). Bottles were regassed on the following two days (D1 and D2) and the experiment was terminated on D3. OD was taken D0-D3.

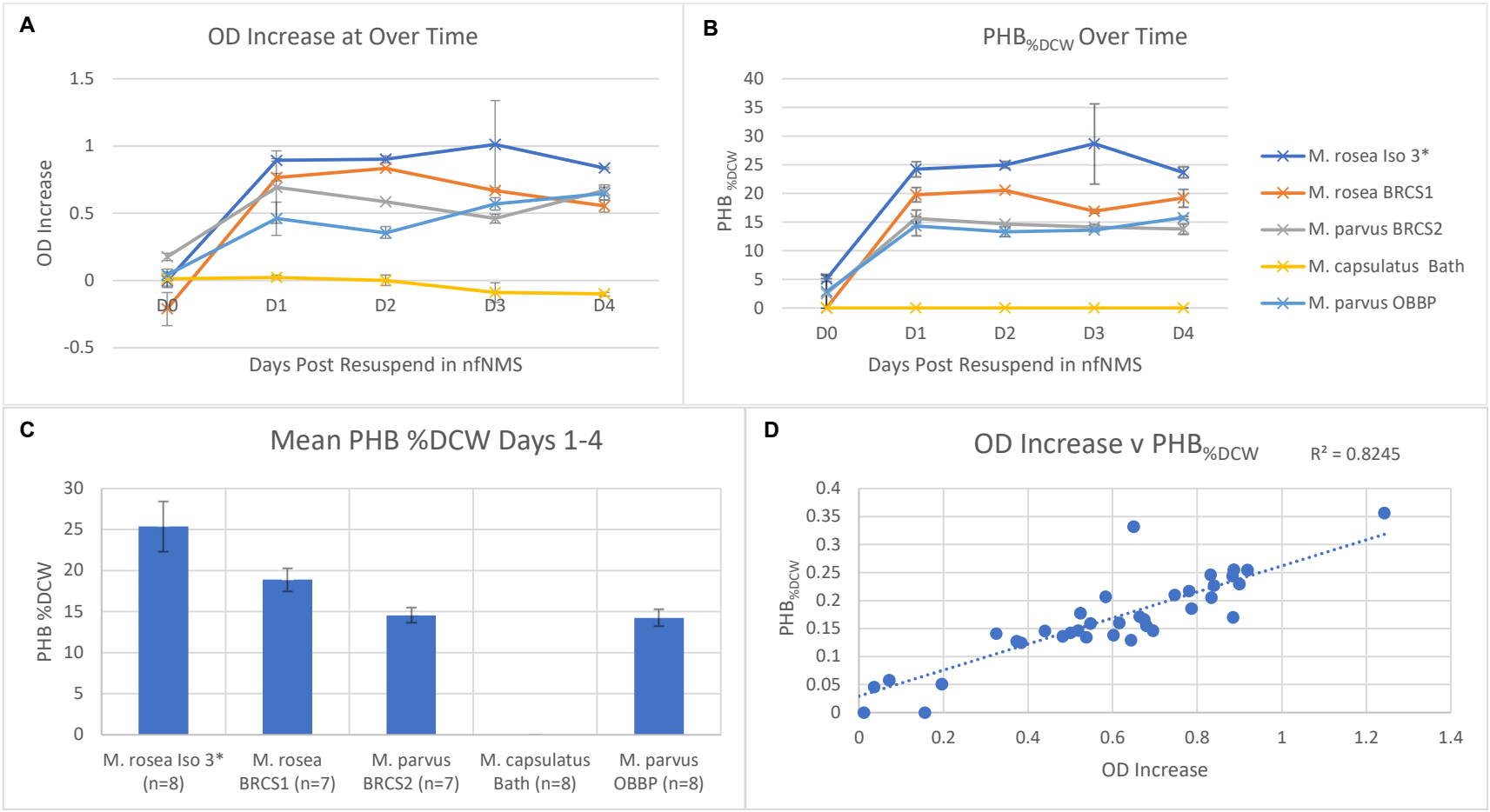


Figure 27 A-D: Comparing five strains of methanotroph for the 4 days after resuspension in nfNMS. **A** – OD increase over 4 days resuspension to sacrifice. **B** – $PHB_{\%DCW}$ over 4 days **C** – Mean of $PHB_{\%DCW}$ over days 1-4. **D** – OD % increase against $PHB_{\%DCW}$ showing corelation these two factors. r^2=0.82. Data for *M. capsulatus* Bath excluded. Data drawn from 40 independent sacrificial bottles. Each datapoint in **A** and **B** are a mean of two (except see text). **C** is a mean of 8 samples or 7 for BRCS1. **D** each point is a single sample (n=40). Error bars are 95%CL.

5.3 Results

5.3.1 Experiment 1 – Comparison of PHB Production in Strains Over Time

In the strain comparison experiment (Figure 27) OD for the four α-proteobacterial methanotrophs increased 0.46-0.90 in one day then remained approximately steady in this range. Of these Iso 3* showed a higher OD at all points. *M. capsulatus* Bath remained steady then decreased progressively to -0.1 OD. $PHB_{\%DCW}$ for the 4 α-proteobacterial methanotrophs increased in 1 day to 13-30% and remained within this range. $PHB_{\%DCW}$ of each species remained level so a mean of the 3 days after inoculation (D1-4) was made for a final comparison in Figure 27C. A two-way ANOVA of $PHB_{\%DCW}$ separated by Strain and Day validated this decision indicating strain was responsible for 78.1% of variation (as η^2) and day only 0.1%.This indicated *M. parvus* BRCS2 and OBBP had similar $PHB_{\%DCW}$ of 14.5±0.87 and 14.3±0.97%, while *M. rosea* Iso 3* was significantly higher than all others at 25.4±3.0% (Tukey's post-hoc p=0.006 or less). *M. rosea* BRCS1 at 18.9±1.35% was significantly above BRCS2 and OBBP (p=0.015 and p=0.008). *M. capsulatus* Bath reported no PHB presence at any point. As no regassing took place, it is believed gas exhaustion of oxygen occurred within 1-2 days.

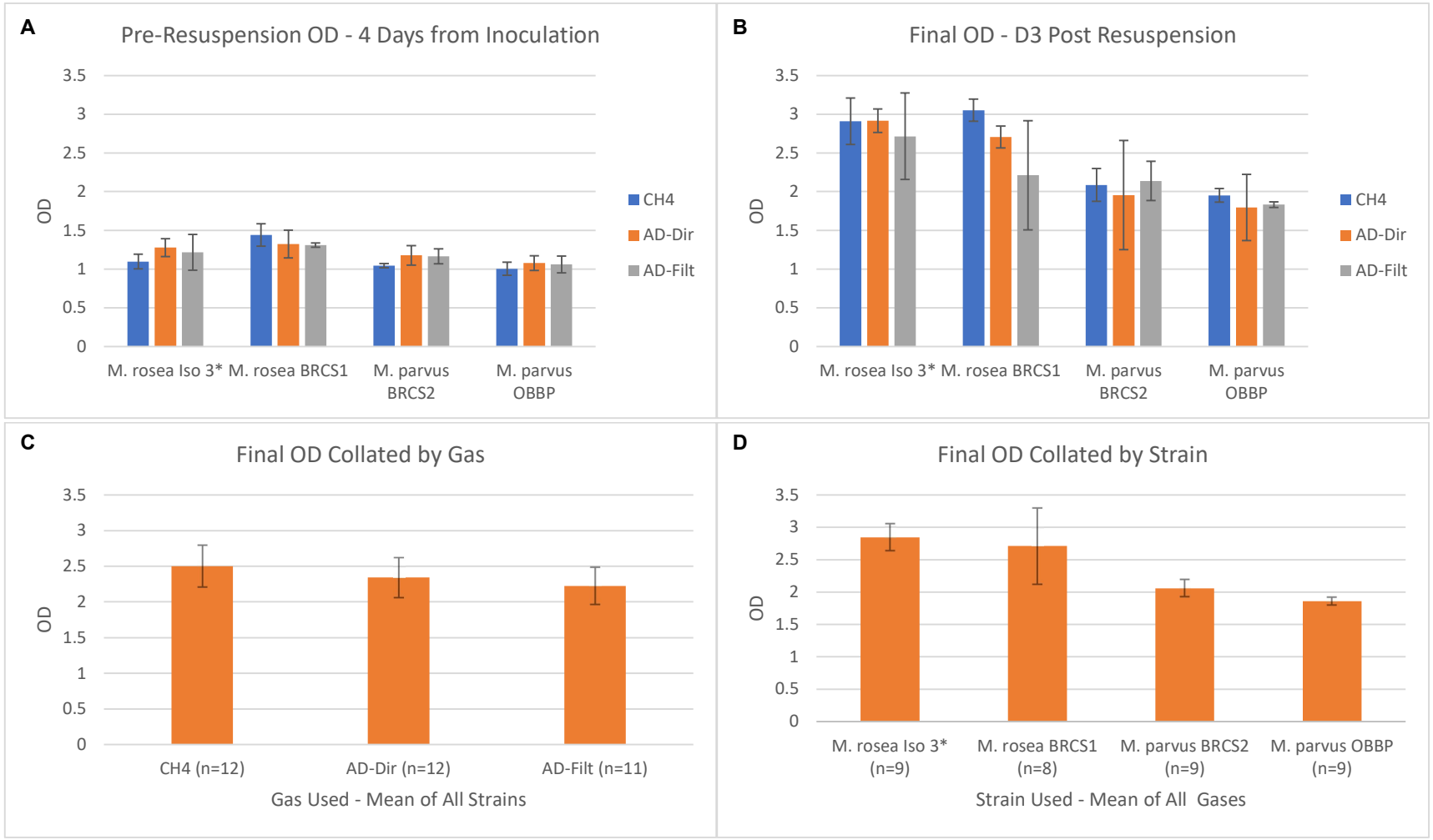
The $PHB_{\%DCW}$ from producing strains (n=40) were plotted against increase in OD (Figure 27D) and a linear regression performed. This indicated an r^2 of 0.82 indicating 82% of the variation in OD was explained by PHB accumulation. The r^2 value of pellet weight of all samples (n=50) in all strains against final OD was 0.96 (not illustrated) indicating a very strong corelation regardless of species or PHB content.

5.3.2 Experiment 2 – Comparison of PHB Production in Strains with Methane or Biogas

Growth data from Exp. 2 showed visually similar growth across strains regardless of gas during the initial phase in NMS for 4 days with a single regassing after 2 days (Figure 28A). A two-way ANOVA showed significant difference by strain but not gas or interaction (p= <0.001, 0.35 and 0.33 respectively) and that via η^2 strain was responsible for 52% of variation. Tukey's post-hoc test indicated BRCS1 was significantly higher than all others and 3* being higher than OBBP with no difference among others. After resuspension in nfNMS and daily regassing for 3 days the endpoint OD (Figure 28B) appeared to further favour Iso 3* and BRCS1. A two-way ANOVA confirmed a difference by strain and gas but not interaction (p= <0.001, 0.031 and 0.097 respectively). Within gas Tukey's test indicated difference was between CH_4 and AD-Filt (p=0.025) with no other differences. Examination of η^2 values showed strain was still responsible for 72% of variation while gas only for 5% meaning gas input was marginal. There was also no difference in the final OD of the samples for each gas taken regardless of strain (Figure 28C). Tukey's identified no difference between the two *M. rosea* strains Iso 3* and BRCS1 (p=0.359) but both were higher than the *M. parvus* strains BRCS2

and OBBP (p=<0.001). Figure 28D illustrates final OD collated by strain regardless of gas.

Two-way ANOVA of $PHB_{\%DCW}$ and $PHB_{m/v}$ indicated no differences by gas type or interaction ($PHB_{\%DCW}$ Gas p=0.83, Interaction p=0.54 $PHB_{m/v}$ Gas p=0.48, Interaction p =0.24) but a significant difference by strain in both cases (p<0.001) making up 51% and 69% of variation respectively (as η^2). Tukey's tests showed for $PHB_{\%DCW}$ Iso 3* was greater than all others (p all <0.013) the rest not being significantly different. For $PHB_{m/v}$ Iso 3* was higher than all others and BRCS1 was higher than BRCS2 and OBBP.


A
Pre-Resuspension OD - 4 Days from Inoculation
OD
3.5
3
2.5
2
1.5
1
0.5
0
M. rosea Iso 3*
M. rosea BRCS1
M. parvus BRCS2
M. parvus OBBP
CH4
AD-Dir
AD-Filt
B
Final OD - D3 Post Resuspension
OD
3.5
3
2.5
2
1.5
1
0.5
0
M. rosea Iso 3*
M. rosea BRCS1
M. parvus BRCS2
M. parvus OBBP
CH4
AD-Dir
AD-Filt
C
Final OD Collated by Gas
OD
3.5
3
2.5
2
1.5
1
0.5
0
CH4 (n=12)
AD-Dir (n=12)
AD-Filt (n=11)
Gas Used - Mean of All Strains
D
Final OD Collated by Strain
OD
3.5
3
2.5
2
1.5
1
0.5
0
M. rosea Iso 3* (n=9)
M. rosea BRCS1 (n=8)
M. parvus BRCS2 (n=9)
M. parvus OBBP (n=9)
Strain Used - Mean of All Gases

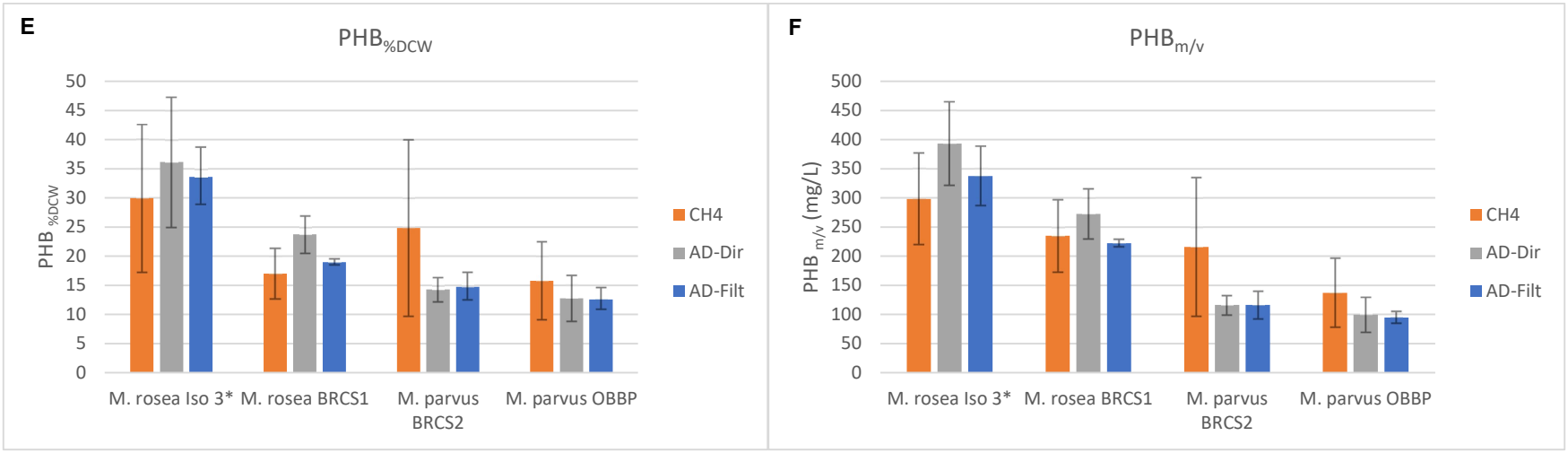


Figure 28 A-F: Comparing four strains of methanotroph grown on three different gas mixtures **A** –OD after 4 days growing in NMS from OD 0.02 with a single regassing after 2 days and before resuspension **B** – OD at end of experiment 3 days after resuspension in nfNMS with a regassing each day. **C –** Data from **B** collated by gas **D –** Data from **B** collated by strain **E -** Final $PHB_{\%DCW}$ **F** – Final Total PHB production as $PHB_{m/v}$. Each datapoint in **A**, **B**, **E** and **F** are a mean of three replicates (except see text). Error bars are 95%CL.

5.4 Discussion

5.4.1 PHB Production

In Exp. 1 (Figure 27) most if not all PHB accumulation occurred in the first 24 hours. However, as this experiment ran without regassing and likely exhausted oxygen within this time it remains possible PHB production could continue further. This was supported by comparing the $PHB_{\%DCW}$ from the PHB producing strains in Exp. 1 (mean 18.3%) to Exp. 2 (mean 21.2%) showing 11% greater levels where daily regassing took place during PHB accumulation. This combined with the lack of PHB decrease over the 4 days of observation resulted in the decision that future PHB bottles would be sacrificed after 3 days. Alternatively due to the strong corelation of OD and PHB accumulation, sacrifice would occur once OD was found to no longer be increasing. *M. capsulatus* Bath reported no PHB presence at any point. Lack of PHB production in *M. capsulatus* Bath was as expected for a γ-proteobacterial species not previously found to PHB producing and included as a -VE control[25]. With the evidence of a lack of PHB production in *M. capsulatus* Bath and, only PHB was found by this analytics methodology in expected strains, this negative control was dispensed with for further PHB production experiments in this thesis.

Results showed *M. rosea* Isolate 3* was a superior PHB producing strain across both experiments achieving a mean of $PHB_{\%DCW}$ 33.2±6.1% and $PHB_{m/v}$ 343±46 mg/L in Exp. 2. *M. rosea* BRCS1 also showed a better $PHB_{m/v}$ 246±46 mg PHB/L than the *M. parvus* strains BRCS2 and OBBP (149±51 and 110±25 mg/L). Both *M. rosea* strains showed better OD and thus biomass but this is mostly due to increase PHB production as the two measures are highly correlated (Figure 27D). Due to extreme similarity as discussed in Chapter 5, no difference would be expected between *M. parvus* OBBP and BRCS2 and none were observed.

PHB levels attained for OBBP were generally lower than expected from other experiments within this thesis so data cannot be used to compare peak production with great authority. Other authors achieved highs of 60% $PHB_{\%DCW}$ in *M. parvus* OBBP[36] and in the accompanying paper to this chapter[19] co-authors achieved higher values for OBBP, BRCS1 and BRCS2. However there is a long published history of widely varying PHB production levels for example other cross strain comparisons achieved *M. rosea* SV99 as 9% and OBBP to be 36% $PHB_{\%DCW}$[23,25,27]. Even under supposedly exhaustive media optimised conditions authors have only reached 50% $PHB_{\%DCW}$ in OBBP[37]. It seems likely OBBP underperformed here. Reliable conclusions drawn on individual strain PHB production is best achieved with a review of literature alongside for example Karthikeyan et al.[16]. Later experiments using the one stage production methodology achieved PHB concentrations closer to those found at the mid to upper end of literature for OBBP. Reasons behind this have not been successfully explained in the literature though various media constituent sensitivities have been examined[37]. Though media used

between researchers is generally a quite consistent recipe for NMS. Age of inoculum has also been shown important in 1986[23] but not examined thoroughly since to the best of my knowledge.

This work is not a full comparison of PHB production across methanotroph strains, rather a study of newly isolated strains with particular absence of the popular PHB producing methanotrophs *Methylosinus trichosporium* OB3b, *Methylocystis sp.* SC2 or a more published *M. rosea* strain like SV99 and other α-proteobacterial methanotroph representatives.

5.4.2 Real world biogas testing

Examination of biogas choice on growth (as OD) showed no significant difference in the preliminary steps with a nitrogen source in the medium. ANOVA showed a difference in OD caused by gas after culturing in nitrogen free medium, but the effect size was small (5%) and was only significant between CH_4 and AD-Filt with strain choice being a much greater component (72%). No significant difference caused by gas choice was found in either of the measures of PHB. This, combined with the marginal p value (0.025) lends us to the suspicion that the CH_4 to AD-Filt difference is in error. The minor differences in biogas OD did not follow any expected pattern, being slightly lower in AD-Filt, which had reduced H_2S by 20x compared to AD-Dir and with assumed none in the pure CH_4 used showing no corelation with H_2S as a factor. The large CO_2 component of both biogases also appear to have had no impact.

Work by López et al.[6] in *Methylocystis hirsuta* utilising artificial biogas (70% CH_4, 29.5% CO_2, 0.5% H_2S – equiv 5000ppmv) achieved growth and a 45% $PHA_{\%DCW}$ purity. No difference was identified in growth with this artificial mix with or without the H_2S. CO_2 has even been considered a potential benefit for α-proteobacterial methanotrophs due to its high use in the serine cycle[38]. This is discussed further in section 6.6.1.2. The lack of CO_2 impact on growth has been corroborated in a second study[5].

Another study by Henard et al.[5] utilising biogas in *M. capsulatus* Bath, *M. trichosporium* OB3b, and *Methylomicrobium alcaliphilum* 20Z^R focused on growth and engineered lactate production. The biogas was generated by the authors from six different feedstocks which differed in resultant biogas. Two had exceedingly high H_2S content >10,000 ppmv. This showed detrimental effects in two of the AD gases used in a strain dependent manner but did not corelate with any monitored biogas constituent including H_2S. Metabolic analysis of *M. alcaliphilum* 20Z^R showed increases in metabolites suggesting oxidative stress and a stress response when grown under biogas compared to pure methane feedstock. Other work in *Methylocystis sp.* M6 found overall growth rates and methane oxidation rates reduced by ~45% with 200ppmv H_2S[39], 3x lower than those found in our AD-Dir gas which showed no such limitation.

As H_2S can also be highly corrosive to metal equipment in biogas to heat and energy plants its potential usage with methanotrophs would avoid this and costly steps in its removal with carbon filters that is currently undertaken[40]. Other reviews concerned with methanotrophic methanol production have suggested H_2S removal is still essential but suggested H_2S tolerant strains could be an alternative[40,41]. It is unclear how the methanol producing studies which require manipulation of methanotroph metabolism, for example H_2 as an additional reducing equivalent, would reflect in a more natural methanotroph metabolism as seen here[27]. Overall, the literature appears conflicted on H_2S but data collected here lends credence to the side of minimal effect.

One contaminant mentioned previously but not evaluated here are siloxanes (a group of organosilicates). Although mentioned in some scientific literature nothing of actual concern about siloxanes and methanotroph fermentation has been illustrated to the best of our knowledge. They are diverse as a group with 350+ different siloxanes being registered for use[42]. Siloxanes were not directly measured here so no specific statement can be made on their effect. Measurements of the same site used here taken 1 year prior (supplementary information in[19]) assessing 10 different siloxanes found levels of 2.4mg/m^3 or below and only finding measurable quantities of four over two measurement events. These values were considerably lower than the upper limits that might be found in biogas shown in literature (50mg/m^3 maximum but highly sample and siloxane type dependent)[4]. The biogas used here were also produced from agricultural waste which would be expected to be lower in siloxanes compared to waste water or landfill biogas with might contain siloxane sources directly, so these lower levels are not surprising. As siloxanes are an established concern in biogas to energy facilities due to their build up on machinery, their lack of impact might be another point in favour of methanotroph biogas fermentation if illustrated satisfactoraly[4].

Additional concerns have been raised about other non-monitored constituents including CH_3SH, COS or CS_2 which can be present in biogas[6,43]. Experiments using soil samples found CH_3SH and CS_2 were also inhibitory to methane oxidation but indicated effects were more severe in Type I compared to Type II methanotrophs[43]. The *Methylocystis sp.* M6 study mentioned previously also found reduced growth and oxidation rates with addition of ethylbenzene, m-xylene, p-xylene, methanthiol (CH_3SH) and dimethylsulfide ($(CH_3)_2S$) suggested to be other potential volatile products found in landfill gas biogas[39].

It is possible other biogas contaminants are toxic to methanotrophs during fermentation, but they are not present in our tested gases in large quantities. Due the huge array of potential low-level content on biogas from the varied inputs into landfill and anaerobic digestion facilities it is hard to exclude everything. Running fermentations with intentional spikes of suspected contaminants like H_2S, a selection of siloxanes, CH_3SH, COS and CS_2 would establish minimum inhibitory concentrations in these contaminants if any exist. Previous work has not established these responses

in a dose-response fashion. This would be vital in running a scaled-up plant indicating what monitoring is required for reliability and quality.

Combined industrial setups of methanotrophs and algae biogas plants either in series or a consortia have been suggested and appear promising both on a scientific[11,40] and technoecanomic[16] level with CO_2 photosynthetically fixed by the algae generating O_2. This can be used to feed methanotrophs which use the remaining CH_4 component for PHB/PHA. This collectively results in lower CO_2 emissions than either microbe individually[16].

5.5 Conclusions

Of the three novel strains mentioned here *M. rosea* Isolate 3* and to a lesser extend BRCS1 both appear highly effective PHB production strains and are worthy of further examination. From the data here we conclude there is little if any detriment from the use of real work biogas to growth or PHB production in any of the strains tested from both *M. parvus* and *M. rosea*. The broad application of real-world biogas rather than artificial biogas used in some previous studies captures any unevaluated contaminant compounds adding surety to their usage in industrial fermentation applications. I hypothesise accumulated toxins in media over time may result in issues unseen here when utilising real world biogas sources and as such a sparged fermenter with high gas throughput to test this is a recommended next step.

Experimentation with real world biogas here is hoped to reassure industry of the concept. The demonstration of the reliability of growth on biogas are applicable to the production of other products from methanotrophs in addition to PHB including the establishing single cell protein as animal feed production methods from Unibio and Calysta among others. Waste methane consumption as an integrated process also allows other higher value products from new or engineered methanotroph strains to be pivoted to with minimal change to the fermentation set up further adding value.

The technology still has to deal with certain obstacles. Largely relating to economic viability. Although the cost of biogas in comparison to fossil fuel methane or oil for plastic production is currently unfavourable, it appears inevitable that fossil fuel prices will rise as resources are exhausted. At that time biogas utilising technologies will become very attractive and in the interim subsidy schemes and green credentials motivate usage. Ensuring scientific barriers to the adoption of the technology are bypassed by that time will maximise its uptake and benefit to the climate and economy. Continually improving the fermentation methods at scale including induction of PHB production, optimised reactor designs and new or optimised strains to push yields higher will help it be cost competitive and achieve its goals for carbon recycling and capture, and reducing plastic waste.

5.6 References

1. Rostkowski KH, Criddle CS, Lepech MD. Cradle-to-gate life cycle assessment for a cradle-to-cradle cycle: Biogas-to-bioplastic (and back). Environ Sci Technol. 2012;46(18):9822–9.
2. Rosenboom J, Langer R, Traverso G. Bioplastics for a circular economy. Nat Rev Mater. 2022 Jan 20;7(2):117–37.
3. Rasi S, Veijanen A, Rintala J. Trace compounds of biogas from different biogas production plants. Energy. 2007 Aug;32(8):1375–80.
4. Tansel B, Surita SC. Managing siloxanes in biogas-to-energy facilities: Economic comparison of pre- vs post-combustion practices. Waste Manag. 2019;96:121–7.
5. Henard CA, Smith H, Dowe N, Kalyuzhnaya MG, Pienkos PT, Guarnieri MT. Bioconversion of methane to lactate by an obligate methanotrophic bacterium. Sci Rep. 2016 Feb 23;6(1):21585.
6. López JC, Arnáiz E, Merchán L, Lebrero R, Muñoz R. Biogas-based polyhydroxyalkanoates production by Methylocystis hirsuta: A step further in anaerobic digestion biorefineries. Chem Eng J. 2018;(333):529–36.
7. VTT Group. Protein feed and bioplastic from farm biogas [Internet]. 2016 [cited 2019 Aug 26]. Available from: https://www.vttresearch.com/media/news/protein-feed-and-bioplastic-from-farm-biogas
8. Technical Research Centre of Finland (VTT). Protein feed and bioplastic from farm biogas. ScienceDaily. 2016 Nov 17;
9. Pieja AJ, Sundstrom ER, Criddle CS. Poly-3-hydroxybutyrate metabolism in the type II Methanotroph Methylocystis parvus OBBP. Appl Environ Microbiol. 2011;77(17):6012–9.
10. Listewnik H-F, Wendlandt K-D, Jechorek M, Mirschel G. Process Design for the Microbial Synthesis of Poly-β-hydroxybutyrate (PHB) from Natural Gas. Eng Life Sci. 2007 Jun;7(3):278–82.
11. van der Ha D, Nachtergaele L, Kerckhof F, Rameiyanti D, Bossier P, Verstraete W, et al. Conversion of Biogas to Bioproducts by Algae and Methane Oxidizing Bacteria. Environ Sci Technol. 2012 Dec 18;46(24):13425–31.
12. Rahnama F, Vasheghani-farahani E, Yazdian F, Abbas S. PHB production by Methylocystis hirsuta from natural gas in a bubble column and a vertical loop bioreactor. Biochem Eng J. 2012;65:51–6.
13. Henard CA, Franklin TG, Youhenna B, But S, Alexander D, Kalyuzhnaya MG, et al. Biogas Biocatalysis: Methanotrophic Bacterial Cultivation, Metabolite Profiling, and Bioconversion to Lactic Acid. Front Microbiol. 2018 Oct 31;9(October):1–8.
14. Chen X, Rodríguez Y, López JC, Muñoz R, Ni B, Sin G. Modeling of

Polyhydroxyalkanoate Synthesis from Biogas by Methylocystis hirsuta. ACS Sustain Chem Eng. 2020 Mar 9;8(9):3906–12.

15. Pérez V, Mota CR, Muñoz R, Lebrero R. Polyhydroxyalkanoates (PHA) production from biogas in waste treatment facilities: Assessing the potential impacts on economy, environment and society. Chemosphere. 2020 Sep;255:126929.

16. Karthikeyan OP, Chidambarampadmavathy K, Cirés S, Heimann K. Review of Sustainable Methane Mitigation and Biopolymer Production. Crit Rev Environ Sci Technol. 2015 Aug 3;45(15):1579–610.

17. Strong PJ, Kalyuzhnaya M, Silverman J, Clarke WP. A methanotroph-based biorefinery: Potential scenarios for generating multiple products from a single fermentation. Bioresour Technol. 2016;215:314–23.

18. Lambert M. Biogas: A significant contribution to decarbonising gas markets? Oxford; 2017.

19. Rumah BL, Stead CE, Claxton Stevens BH, Minton NP, Grosse-Honebrink A, Zhang Y. Isolation and characterisation of Methylocystis spp. for poly-3-hydroxybutyrate production using waste methane feedstocks. AMB Express. 2021 Dec 6;11(6):1–13.

20. Bowman J. The Methanotrophs — The Families Methylococcaceae and Methylocystaceae. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes. 3rd ed. New York: Springer New York; 2006. p. 266–89.

21. Department of Energy and Climate Change. Agriculture GHG Inventory Summary Factsheet. Environment. 2013.

22. Jardine CN, Boardman B, Osman A, Vowles J, Palmer J. Methane UK. Environ Chang Inst. 2003;44:1–96.

23. Asenjo JA, Suk JS. Microbial Conversion of Methane into poly-β-hydroxybutyrate (PHB): Growth and intracellular product accumulation in a type II methanotroph. J Ferment Technol. 1986 Jan 1;64(4):271–8.

24. Golder Associates Ireland/EPA. Management of Low Levels of Landfill Gas. 2013.

25. Pieja AJ, Rostkowski KH, Criddle CS. Distribution and Selection of Poly-3-Hydroxybutyrate Production Capacity in Methanotrophic Proteobacteria. Microb Ecol. 2011;62(3):564–73.

26. Wendlandt KD, Geyer W, Mirschel G, Al-Haj Hemidi F. Possibilities for controlling a PHB accumulation process using various analytical methods. J Biotechnol. 2005;117(1):119–29.

27. Pieja AJ, Morse MC, Cal AJ. Methane to bioproducts: the future of the bioeconomy? Curr Opin Chem Biol. 2017;41(1):123–31.

28. International Energy Agency. Outlook for biogas and biomethane - World Energy Outlook Special Report. 2020.

29. U.S. Energy Information Administration. Henry Hub Natural Gas Spot

Price 1997-2022 [Internet]. [cited 2023 Feb 1]. Available from: https://www.eia.gov/dnav/ng/hist/rngwhhdM.htm

30. BP. Approximate conversion factors. Stat Rev World Energy. 2021;July.

31. Ward N, Larsen Ø, Sakwa J, Bruseth L, Khouri H, Durkin AS, et al. Genomic insights into methanotrophy: The complete genome sequence of Methylococcus capsulatus (Bath). PLoS Biol. 2004;2(10):1616–28.

32. Calysta. Calysta Acquires BioProtein A/S: Proven Methane to Feed Technology – Enters $370 Billion Nutritional Market with Approved Product. 2014.

33. U.S. Energy Information Administration. Annual Energy Outlook 2022. U.S. Energy Information Administration. 2022.

34. Rumah BL. Isolation, characterisation and development of genetic tools for methane-oxidising bacteria. PhD Thesis. 2018;

35. Lower Reule Bioenergy Ltd. Lower Reule Bioenergy Ltd - Environmental Management System Manual. 2020.

36. Rostkowski KH, Pfluger AR, Criddle CS. Stoichiometry and kinetics of the PHB-producing Type II methanotrophs Methylosinus trichosporium OB3b and Methylocystis parvus OBBP. Bioresour Technol. 2013;132:71–7.

37. Sundstrom ER, Criddle CS. Optimization of methanotrophic growth and production of poly(3-hydroxybutyrate) in a high-throughput microbioreactor system. Appl Environ Microbiol. 2015;81(14):4767–73.

38. Strong P, Laycock B, Mahamud S, Jensen P, Lant P, Tyson G, et al. The Opportunity for High-Performance Biomaterials from Methane. Microorganisms. 2016 Feb 3;4(1):11.

39. Lee E-H, Yi T, Moon K-E, Park H, Ryu HW, Cho K-S. Characterization of Methane Oxidation by a Methanotroph Isolated from a Landfill Cover Soil, South Korea. J Microbiol Biotechnol. 2011 Jul;21(7):753–6.

40. Salehi R, Chaiprapat S. Conversion of biogas from anaerobic digestion to single cell protein and bio-methanol: mechanism, microorganisms and key factors - A review. Environ Eng Res. 2021 Jun 21;27(4):210109–0.

41. Patel SKS, Mardina P, Kim D, Kim S-Y, Kalia VC, Kim I, et al. Improvement in methanol production by regulating the composition of synthetic gas mixture and raw biogas. Bioresour Technol. 2016 Oct;218:202–8.

42. Lassen C, Hansen CL, Mikkelsen SH, Maag J. Siloxanes - Consumption , Toxicity and Alternatives. Danish Ministry of the Environment. 2005.

43. Börjesson G. Inhibition of methane oxidation by volatile sulfur compounds (CH3SH and CS2) in landfill cover soils. Waste Manag Res J a Sustain Circ Econ. 2001 Aug 2;19(4):314–9.

Chapter 6: Scale Up of Production to 750ml with Analytics, Stirring and Sparging with Mixed Culture Nanopore Analysis

6.1 Introduction

Scale up of methanotroph growth to sparged fermenters is not a new development by any means with work on advanced bioreactors dating back to at least 1992[1]. However, it provided a major challenge in proceeding with this work where growth would not progress after inoculation. The original intention was to produce PHB containing biomass via sparged fermenter and utilise this biomass for later PHB extraction experiments. Experiments in this chapter are split into three sections. First, eight failed fermentations in bioreactors at 750ml scale are presented along with two successful fermentations. Of the two successful runs one was contaminated only after successful growth and one contaminated from inoculation. In the second section, four small experiments were run to investigate the parameters eliminating or informing on possible causal factors of failure in section one. In the third section the contaminated runs were considered more favourably as mixed cultures. An in house 16S nanopore sequencing analysis was developed to evaluate the species mix present and its variation over time in a fast and efficient fashion that may have broader applications in industry.

6.2 Background

6.2.1 Bioreactors and Fermenters

A fermenter or bioreactor (equivalent terms) is a controlled environment for cell growth with greater capabilities than bottled growth. These come in the form of:

1. Improved monitoring by use of online probes for example OD, pH, DO (Dissolved Oxygen) and temperature at high polling rates without impacting volume.
2. Additions can be made continually during fermentation e.g. acid and base for pH balancing, nitrate and carbon sources, air or oxygen and feed gases for gas fermentation. This allows faster growth rates and increased peak cell density.
3. Stirring by impeller and gas introduction from a sparger providing greater gas mass flow than incubator shaking.
4. Increased liquid volume capacity.

These factors allow higher biomasses and growth rates to be attained without limitations being reached and a greater understanding of the

fermentation. Scaled up fermentations in bioreactors also provide a closer environment to industrial fermentation indicating the effectiveness of a process for commercialisation in a more applicable fashion. For gas fermentation the major advantage of a bioreactor over bottled experiment is the increase in mass transfer of methane and oxygen into the media and continual feeding of gases[2]. The setup used in this chapter is shown in Figure 29A.

The use of a sparger with continual gas flow of air/oxygen and methane allows higher cell densities and growth rates to be achieved, with *M. capsulatus* Bath and *M. trichosporium* reaching up to 5g/L DCW and specific growth rates of up to 0.2/h[3]. Optimised and specialised reactors have attained densities of 65g/L and growth rates of 0.17/h in *Methylocystis sp.*[4] with other species reporting up to 0.4/h[3]. This is a reasonable rate when compared to *E. coli* which have been reported from 0.4 to 0.7/h at batch culture despite methanotrophs generally being considered slow growing[2]. Peak biomass concentrations are likely further optimisable. This compares to the maximum cell densities personally achieved in serum bottles of 1.1g/L with *M. parvus* OBBP with repeated regassing. Greater cell density reduces the amounts of water required in fermentation and allows a greater biomass to be achieved on smaller equipment.

As discussed, in section 1.2.4 p21 gas-liquid mass transfer of oxygen and methane is the major limiting factor in growth due to low solubility of methane and oxygen[2,5,6]. Oxygen mass flow is of general concern in aerobic heterotrophs to maximise respiration. In methanotrophs the distribution of methane is an added challenge. This challenge is great enough that even sparged stirred tank fermenters do not scale well in industry for gas fermentation[7]. Several novel reactor designs and methods have been utilised to combat gas mass flow challenges, developing on conventional bioreactors. Examples include the U-loop reactor in use by Unibio for SCP production[8,9], high pressure reactors[4,5] and bubble column reactors[10–12]. Calysta also has its own horizontal loop reactor design with fluid moving along a tube with gases added through ports along it[13]. Two-phase partitioning bioreactors also have potential to aid in methane and oxygen transfer using a non-aqueous vector like paraffin, propanol or silicone oil[14–16]. Two-phase reactors have seen some success at 5L scale with *M. trichosporium* where the use of paraffin oil increased growth rates drastically[17]. Unless designed with a recirculation or retrieval system, gas not utilised by methanotrophs is wasted. This reduces efficiency and increases cost per output, thus efficient mass transfer is of great importance at industrial scale[5,13].

6.2.2 pH Control Requirements and Byproducts

pH control is a requirement in most if not all bioreactor systems. CO_2 dissolution is often a cause of fermenter acidification[18]. However the overall direction of pH change is generally caused by fermentation by-products and is dependent on the bacteria and feedstock consumed. A well understood microbial system often only utilises acid or base for control in one direction

e.g. a fermentation of *Methylomicrobium buryatense* only utilised base control due to the production of formate, acetate and lactate causing acidification[19]. Generally NaOH is the most common base and HCl is the most common acid but H_2SO_4 has also been utilised as an acid[13,20].

γ-proteobacterial methanotrophs produce organic acids in the form of formate, acetate, succinate and lactate, especially in oxygen limited conditions[21]. These are excreted and their build up may become a limiting factor[19,21]. α-proteobacterial *M. parvus* has also been found experimentally to excrete low levels of acetate and succinate along with acetone, 2,3-butanediol and isopropanol under oxygen limiting conditions[22–24], though this has been disputed in the genome scale model[25]. The presence of a significant proportion of methylotrophs in the *Methylocystis sp*. GB25 mixed culture fermentation mentioned below (6.2.4.2 p252) also implies methanol may also be excreted which could become a limiting factor[26] and methanol and formaldehyde production has been confirmed to occur in *M. trichosporium* OB3b and IMV 3011[27,28].

The major methanotroph isolation paper by Whittenbury et al.[29] states all strains (including *M. parvus* OBBP) grew over a pH range of 5.8 to 7.4 and that 6.6 to 6.8 was optimal. Pieja et al.[30] carried out enrichment experiments from mixed cultures using a carbonate buffer adjusted with HCl and achieved pHs of 4, 5, 6, 7, and 8 and showed α-proteobacterial methanotroph growth at pH 4, 5 and 7. Collectively this data indicates *M. parvus* should be capable of growth at a range of pHs and the standard pH of 6.8 used in NMS medium under standard conditions was chosen. No detailed studies examining the effect of pH on PHB accumulation in methanotrophs have been carried out[31].

6.2.3 PHB and Continual Fermentation

Scale-up was originally envisaged to use a process similar to the two-stage fermentation process described in section 2.1.3, with growth in the bioreactor, culture removal, resuspension in nfNMS to limit nitrate and return to the reactor to accumulate PHB. This removal, spin down and resuspension step would pose increased workload, cell loss and a risks contamination. Adoption of a one-stage methodology where nitrate becomes limiting as cells consume the original media during growth eliminates this step. The next goal after successful growth would be continual fermentation.

In continual fermentation like that of a chemostat fresh medium is continually added and biomass containing culture is continuously removed to maintain culture volume and biomass. In this way maximum productivity can be maintained over an extended period. This is far more efficient compared to running a single fermentation to completion, emptying, cleaning, refilling, and sterilisation between runs. Continual fermentation also serves to eliminate batch to batch variation, down time and growth lag time.

Continual fermentation of methanotrophs has previously been achieved on multiple occasions[5,26,32]. Continual PHB production from methanotrophs however is more challenging due to the requirements of nutrient limitation. A

semi-continuous fermentation in which the biomass growth phase is continuous but culture is removed to complete nutrient exhaustion and allow PHB accumulation separately appears the most effective solution and has been achieved over long periods in mixed culture by previous authors[20,26,33]. Alternatively feast/famine cycling where PHB is allowed to accumulate, a portion of culture removed then more nitrate is added has also been atempted[34].

6.2.4 Mixed Culture Fermentation

A mixed culture is one that contains more than one microbial species as compared to a pure culture which only contains one. Methanotrophs are often more successful in mixed than in pure culture and have improved growth and stability characateristics[35,36]. Mixed culturing is essential in some cases e.g. when using natural gas as a feedstock over a prolonged period[32]. Mixed cultures are based on a mutual relationship where methanotrophs perform carbon fixation and energy input and other constituent polytrophs consume by-products and may produce nutrients utilised by the methanotroph[26,35,36]. By-products may be toxic to methanotrophs if left to accumulate. It has been suggested partner polytrophs may also provide vitamins[2] but from experimentation within our lab *M. parvus* is capable of growth without additional vitamins. As only by-products are utilised by the polytrophs this theoretically would not impact carbon conversion efficiency but might increase oxygen utilisation.

A successful mixed culture is kept in balance dominated by the methanotroph as it provides the only route to the introduction of carbon for utilisation by the other constituent species stopping them overtaking the culture[35]. These balances can however be disturbed by contamination with unintended species or exhaustion of nutrients, oxygen or methane stalling methanotroph growth[26]. A mixed culture may be more resistant to contamination from outside sources as a totally sterile process at industrial scale may be unfeasiable[26,32] and non-sterile cultures avoid many costly processes and procedures[2].

Research into bacterial relationships in mixed culture methanotroph fermentation are lacking. Environmental microbiology and biotechnology have long been interested in methanotroph/polytroph interactions and the influences on their balance. This field may hold information on effective relationships not identified in studies on mixed culture industrial fermentation[36,37]. For example, mixed cultures of *M. parvus* with denitrifying bacteria in denitrification bioreactors (generally woodchip filled open pits) has been explored[22]. Cobalamin (vitamin B12) produced by *Rhizobium spp.,* in an environmentally sourced co-culture has been identified as a stimulator of growth and methane oxidation in accompanying methanotrophs[38]. Five polytrophs have been shown to improve *M. parvus* OBBP growth believed to provide volatile organic compounds[36]. Mixed culturing can also unlock access to combined feedstocks like H_2 which can work to capture the CO_2 developed by methane oxidation[39].

6.2.4.1 Calysta Process Case Study

A mixed culture bacterial set developed by Norferm, now used by Calysta is particularly well studied and constituents are listed in patents for their process and associated literature[13,32]. It is used for the production of SCP at industrial scale using *M. capsulatus* Bath and 3 other polytrophic strains: DB3, DB4 and DB5 (Table 17). These species were found as contaminants in their semi-sterile fermentations (as occurred in Ferm 9 and 10 of this chapter). The contaminant bacteria were found to be beneficial and were adopted as part of a defined mixed culture. As the best studied methanotroph mixed culture process this is used as a case study. A study of mixed cultures arising alongside industrial cultivation of *M. capsulatus* Bath for SCP by Gasprin is also notable[40].

M. capsulatus Bath could not be grown in continuous fermentation on natural gas for more than 2-3 days in pure culture but the presence of the polytrophs allowed this to continue indefinitely. The relative abundance of each species was stable in continuous culture but differed between methane and natural gas feedstock. All three polytrophs were found capable of growing in cell-free medium in which *M. capsulatus* Bath had previously grown. The polytrophs reduced acetate, total organic carbon and free amino acid levels. This indicates they were capable of growth only on methanotrophy by-products and dead cells. C-C bond natural gas components like ethane and propane are not fully metabolised by methanotrophic bacteria and are oxidised to carboxylic acids by cross reactivity by MMO. These products can be inhibitory to methanotrophs. Polytrophicic bacteria in the mixed culture, mostly DB3, metabolises and remove these components[13,32]. In this way the mixed culture deals with toxic by-products arising from the use of natural gas feedstock. A similar feature may also prove important for stability from accumulated biogas components as discussed in Chapter 5. The Calysta bacterial set also metabolise lysed cell content which would otherwise contribute to foaming[13,32].

6.2.4.2 PHB Mixed Methanotroph Cultures

The mixed culture used by Calysta is a defined culture with 4 total components, but most studied mixed methanotroph cultures are more diverse being sourced from environmental samples, waste decomposition sludge or prolonged exposure to external contamination in non-aseptic conditions[26,41–43]. These cultures are challenging to understand, particularly in terms of individual constituents' contribution to the process forming more of a "black box". Often little to no investigation is carried out into the microbial constituents of the culture making comparisons of studies challenging.

In a well-studied example mixed culture, a *Methylocystis sp.* GB25 dominated (~86% prevalence) open culture has run in multiple experiments and up to 29 months[4,20,26,44]. This isolated two methylotrophs: *Acidovorax sp.* and *Methylophilus methylotrophus* along with three heterotrophs: *Leifsonia sp.*, *Bosea thiooxidans* and *Chryseobacterium sp.* At least 2 other species were present but not identified. Prevalence of *Methylocystis sp.* GB25 remained

Table 17: Defined polytrophic species constituents of the Norferm/Calysta mixed culture industrial process alongside *M. capsulatus* Bath.

Strain Identifier	NCIMB Number	Published Species[32]	BLAST Species	Notes	Purpose	Relative Abundance
DB3	13287	*Ralstonia sp.* (formerly *Alcaligenes acidovorans*)	*Cupriavidus sp.* *Cupriavidus gilardii* *Cupriavidus cauae*	Can metabolise ethanol, acetate, propionate and butyrate.	Utilize acetate, propionate and other carboxylic acids produced by *M. capsulatus* Bath from ethane and propane in the natural gas.	6-13%
DB4	13288	*Aneurinibacillus sp.*	*Aneurinibacillus danicus*	*Aneurinibacillus danicus* in NCIMB. Can metabolise acetate, D-fructose, D-mannose, ribose and D-tagatose. Noted to be less essential than DB3 and DB5.	Utilize acetate and lysis products and metabolites in the medium.	<1%
DB5	13289	*Brevibacillus agri* (formerly *Bacillus firmus*)	No Sequence	*Brevibacillus sp.* in NCIMB. Can metabolise acetate, N-acetyl-glucosamine, citrate, gluconate, D-glucose, glycerol and mannitol.	Utilize lysis products and metabolites in the medium including DNA.	<1%

As these species definitions source from 2002, it was attempted to re-BLAST the original sequences however the accessions listed were incorrect covering DB3 and DB4 but not DB5 as stated[32]. AF369868 and AF369869 DB3 sequences were replaced with updated accession AH010744.2. AF369870 was DB4. Table content sourced from [13,32].

constant +/- 2.7% over the length of the experiment indicating a stable mixed culture.

More mixed cultures have been utilised in PHB production from methane and are listed by Singh et al[35], there are however to the best of our knowledge no defined mixed cultures published for this purpose.

Section 1: Scaled Up 750ml Fermentation

6.2.5 Methods

These methods use the terminology "online" indicating an instrument monitoring the fermentation directly as part of the fermenter setup, and "offline" meaning data taken from aliquots of the culture.

6.2.5.1 Fermenter Details

Scale up fermentations were performed on a DASGIP Parallel Bioreactor System with DASware control software (Eppendorf, Germany). Fermentations were all performed on a single fermentation vessel of an 8 unit system. The system consisted of the following DASGIP control modules: TC4SC4 Temperature and Agitation Controller, MX4/4 Gas Mixing Module, OD4 OD Monitor, PH4PO4L for pH monitoring and PH4O4RD4 for OD monitoring, 2x MP8 Multipump Modules for acid and base flow control. Attached to these were the following online probes: 405-DPAS-SC pH Electrode (Mettler Toledo, USA), InPro6800 Ampomeric O_2 Sensor (Mettler Toledo, USA), VisiFerm DO ECS 225 DO Sensor (Hamilton, Switzerland), Platinum RTD Temperature Sensor (Eppendorf, USA), DASGIP OD Sensor (78103409, Eppindorf, Germany). 2.1L Glass Fermentation vessels were placed within a DASGIP Bioblock CWD4 Cooling Water Distribution Unit fed from an MC 1200 Microcool Circulation Chiller (Lauda Brinkmann, USA). Air, CH_4 and N_2 gas flow rates were controlled by red-y GSC smart controllers (Vögtlin Instruments, Switzerland). This set up is shown in Figure 29 B and C.

Each system had a shadow reactor with an O_2 probe to monitor for explosive gas mixtures fed from the off-gases from the main fermentation vessel. DO is expressed as $\%_{DO}$, the percentage of media saturation by oxygen compared to the predicted maximum content at 30°C under air at 1atm.

pH probes were 2-step calibrated at pH 4 and pH 7 before each run and verified with offline readings before and during fermentation. Fermenter vessels were autoclaved in its entirety with DO, OD and pH probes in place. Sterile medium was then added by displacement with nitrogen through a filter.

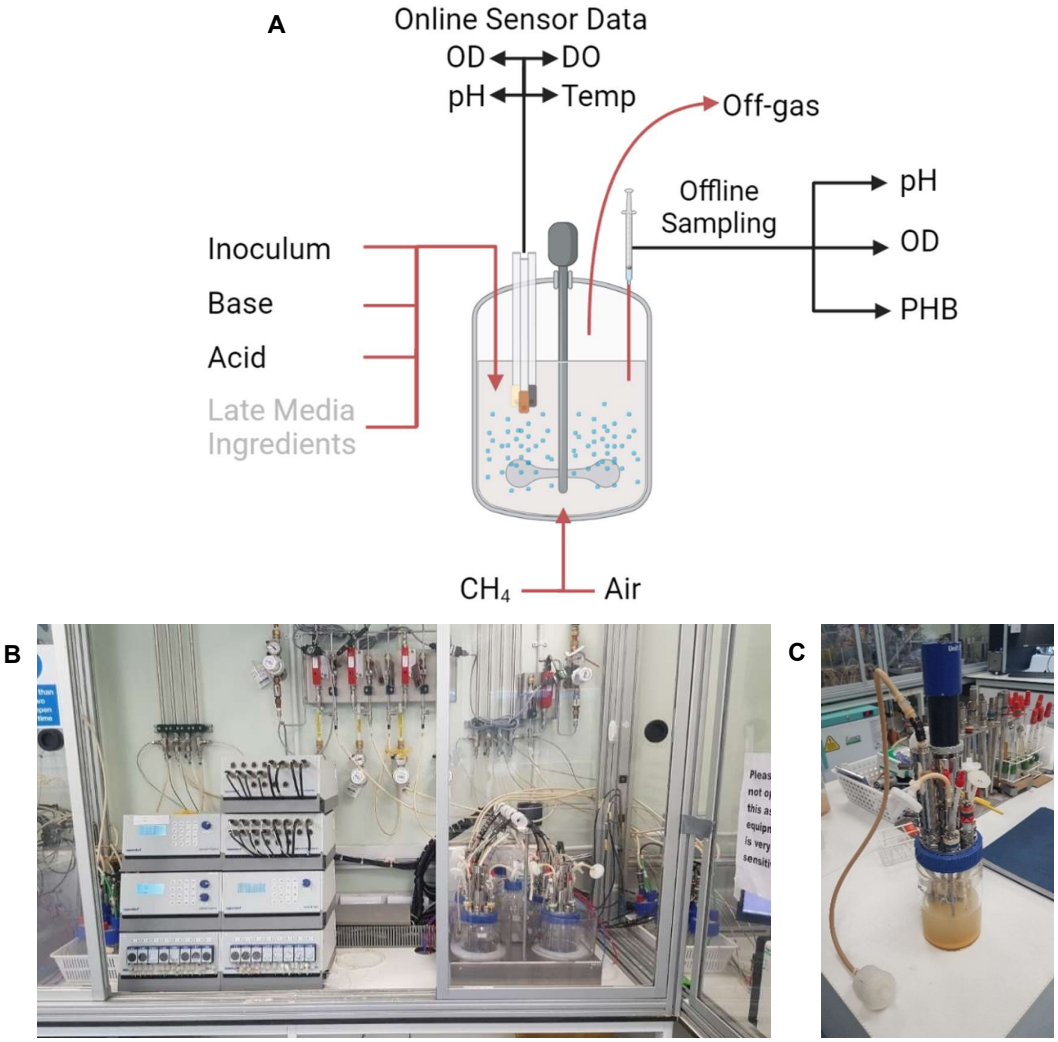


Figure 29 A - C: **A –** Diagram of inputs and outputs of the fermenter as used in this process. Physical additions and removals are indicated with red arrows, data only with black. Late medium ingredients sometimes added. **B** and **C** Photographs showing **B** - the DASGIP 750ml scale fermentation setup for 4 units **C** – A fermenter out of its housing. Specifically, Ferm 9 after its end point. **A** was created with BioRender.com

6.2.5.2 Fermenter Run Parameters

DASGIP Fermentations were run with the following parameters unless indicated otherwise in Table 18: Total initial volume 750 ml of NMS, 2-3 drops (~90µl) antifoam 204 (A8311, Sigma-Aldrich), temperature 30°C, pH control by addition of 1M NaOH and 1M HCl to a set point of 6.8. Impeller speed and Acid/Base flow rate were set by PID (proportional–integral–derivative) function dependent on $\%_{DO}$ and pH respectively.

CH_4 flow rate was a constant 27 sL/h (standard litres per hour) and air flow rate 18 sL/h giving 12.9 and 8.6 reactor volumes per hour respectively. This is equivalent to 60% methane 40% Air of which 8.4% is oxygen well above the UEL (section 1.2.4 p21). This equates to an O_2 to CH_4 ratio of 0.14:1 compared to an optimum of 1.5:1[6] to 2:1[45]. To maximise mass flow in industry and some published research a 50% oxygen 50% methane mixture is common which is closer to the stochiometric usage predictions[46] but this was not possible due to safety limitations.

Methanotroph cultures bound for fermenter inoculum were freshly grown in glass bottles. Culture from multiple bottles were combined in a calculated volume to give a final OD in the fermenter, centrifuged and resuspended in 30ml of NMS. This was injected into the fermenter using a syringe port which was sealed with a clip when not in use. This combined with 720 ml of medium in the fermenter for the total volume of 750 ml.

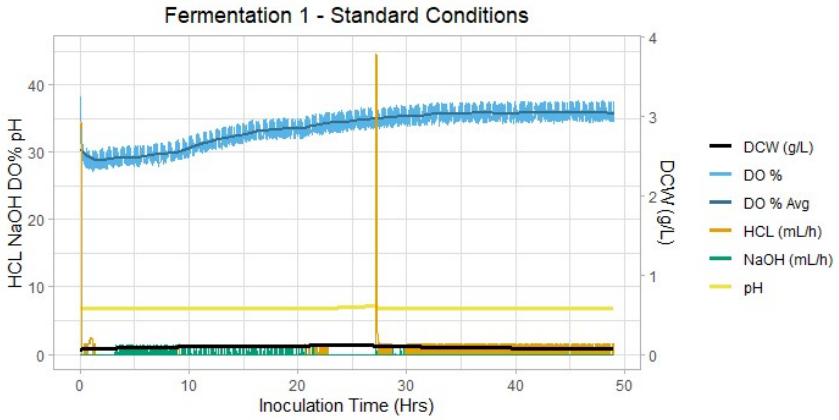
Aliquots for offline analysis were drawn using a one-way flow syringe port. A 3 ml sample was drawn to clear residual culture in the tube line then samples of necessary size were drawn. For pH, OD and contaminant testing a 2 ml sample was taken and tested in duplicate. Sample volumes for PHB analysis were calculated to give 10-15 mg using Equation 1 limited at 40ml maximum. At fermentation termination, vessels were flushed with N_2 50sL/h to remove residual CH_4 and O_2 and return safe conditions.

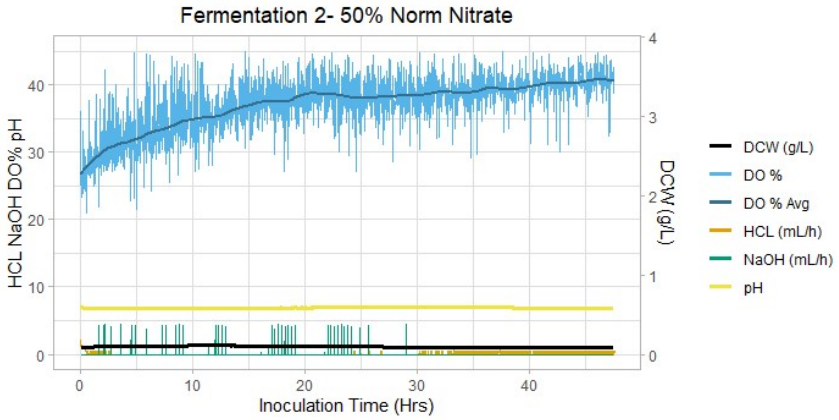
Table 18: Fermentation Parameters and Outcomes

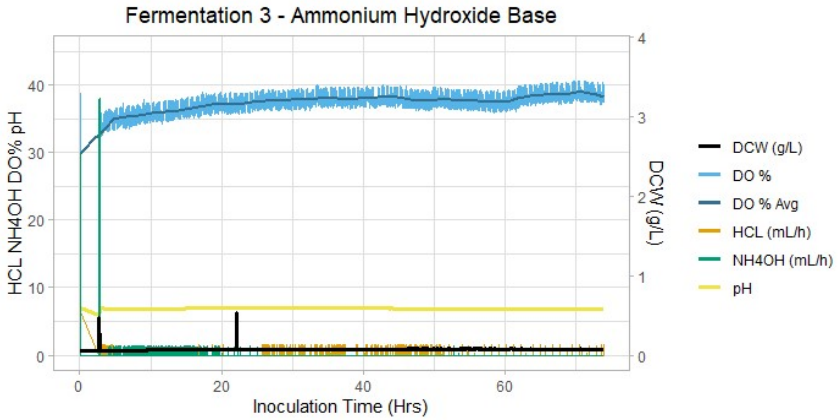
Fermentation No.	Purpose	Notes/Changes in Detail	Starting OD	Run Length (Hrs)	Increase in DCW (mg)	Peak $PHB_{\%DCW}$	pH Setpoint	Acid Added (ml)	Base Added (ml)	Acid:Base Ratio	Contaminated
1	Standard Conditions		0.23	49	53.85	18.22[a]	6.8	4.2	1.9	2.24	No
2	Nitrate Limitation	Used NMS with 50% standard (0.5g/L) KNO3	0.34	48	19.83		6.9	3.3	1.3	2.51	No
3	Change Base	Used NH_4OH in place of NaOH	0.25	74	32.10		6.9	2.5	6.2	0.40	No
4	No Antifoam No pH Control		0.18	72	17.24		None	0.0	0.0	N/A	No
5	Late Addition of NMS. Force Increase rpm.	Set up with 330ml NMS. 300ml NMS Added along with inoculum. Sparger Blockage Starts after 13H. rpm set to 800	0.29	29	249.34		6.8	9.0	5.1	1.76	No
6	High OD Start. Medium Ingredients Late. Force Increase rpm	Phosphates, Vitamins, Fe-EDTA and CuSO4 Added with Inoculum. rpm set to 800. Sparger Blockage 53 Hrs	1.01	58	113.68	8.69	6.8	7.1	4.9	1.45	No
7	Medium OD Start. Medium Ingredients Late. Force High rpm	Phosphates, Vitamins, Fe-EDTA and CuSO4 Added with Inoculum. rpm set to 1200. Online OD Failed, Used Offline Data	0.58	38	21.94	14.21	6.8	0	2.8	-----	No

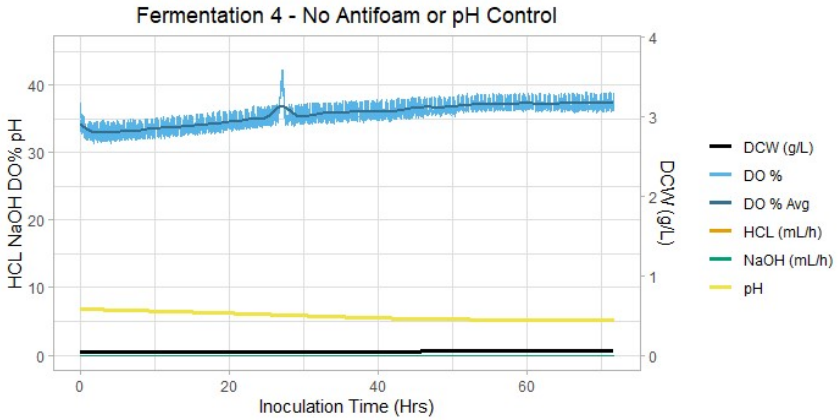
Fermentation No.	Purpose	Notes/Changes in Detail	Starting OD	Run Length (Hrs)	Increase in DCW (mg)	Peak $PHB_{\%DCW}$	pH Setpoint	Acid Added (ml)	Base Added (ml)	Acid:Base Ratio	Contaminated
8	*M. parvus* BRCS2		0.22	191	75.27	10.17	6.8	11.7	3.0	3.96	No
9	High OD Start. Medium Ingredients Late	Phosphates, Vitamins, Fe-EDTA and CuSO4 Added with Inoculum	1.64	140	2,402.89	39.04	6.8	27.6	5.6	4.94	After 43hrs, not before
10	High OD Start		1.31	118	2,868.29	35.73	6.8	32.9	6.5	5.05	Throughout

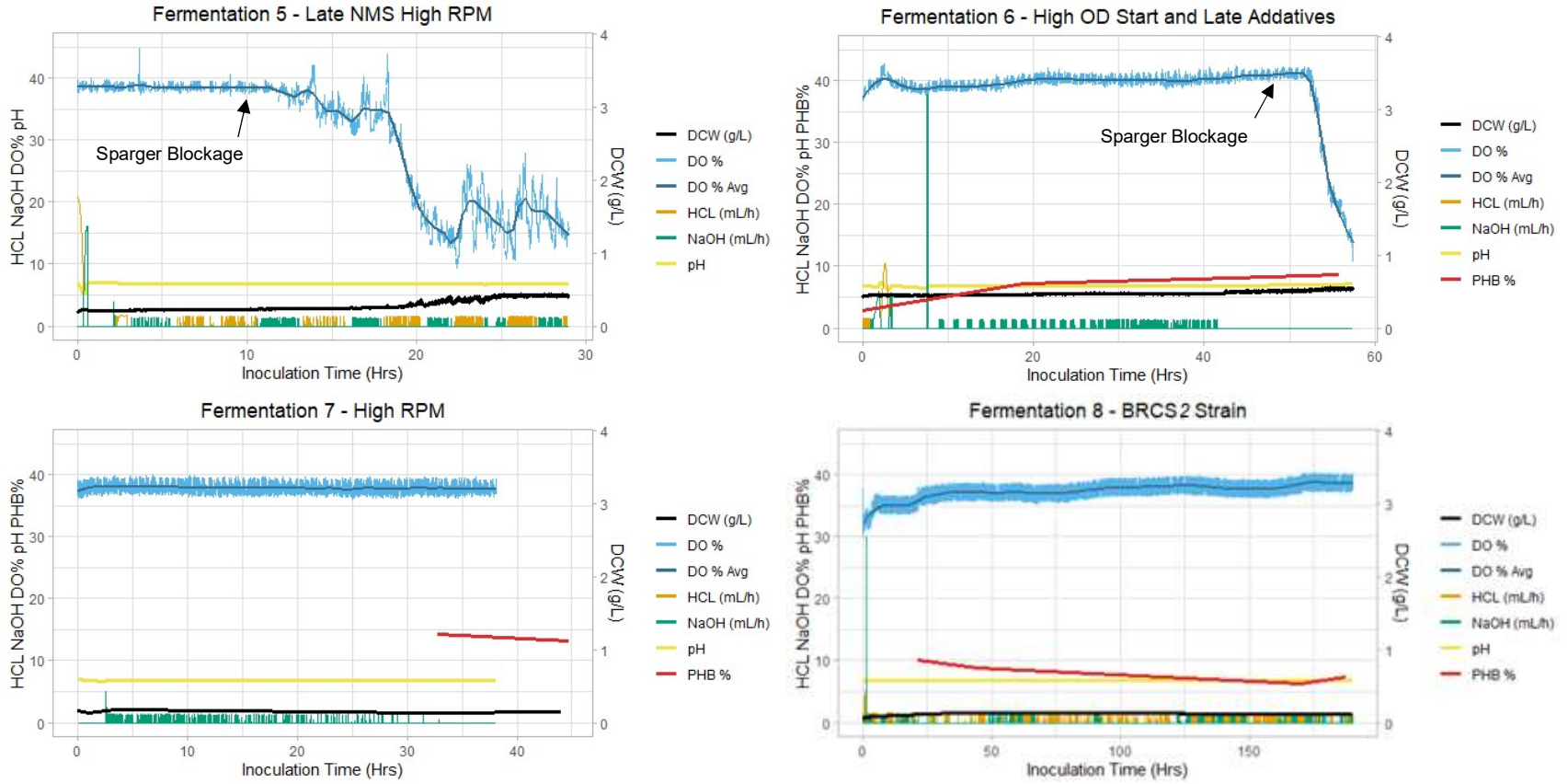
Run length rounded up to whole hours. Other values to 2dp where rounded. Standard Conditions are those listed in 6.3.1.2. All fermentations run under those conditions unless otherwise noted. Acid in all cases is HCl at 1M. Base was NaOH 1M in all cases except Ferm 3 which was NH_4OH 1M. Values for Ferm. 9 and 10 DCW taken directly from freeze-dried samples taken at the start and end of fermentation, others converted from starting and peak fermenter OD using Equation 1 and Equation 7. Peak $PHB_{\%DCW}$ calculated directly against weight of freeze-dried pellet. [a]PHB for fermentation one was taken after flushing with nitrogen gas.


Fermentation 1 - Standard Conditions
HCL NaOH DO% pH
DCW (g/L)
Inoculation Time (Hrs)
DCW (g/L)
DO %
DO % Avg
HCL (mL/h)
NaOH (mL/h)
pH


Fermentation 2- 50% Norm Nitrate
HCL NaOH DO% pH
DCW (g/L)
Inoculation Time (Hrs)
DCW (g/L)
DO %
DO % Avg
HCL (mL/h)
NaOH (mL/h)
pH


Fermentation 3 - Ammonium Hydroxide Base
HCL NH4OH DO% pH
DCW (g/L)
Inoculation Time (Hrs)
DCW (g/L)
DO %
DO % Avg
HCL (mL/h)
NH4OH (mL/h)
pH


Fermentation 4 - No Antifoam or pH Control
HCL NaOH DO% pH
DCW (g/L)
Inoculation Time (Hrs)
DCW (g/L)
DO %
DO % Avg
HCL (mL/h)
NaOH (mL/h)
pH


Fermentation 5 - Late NMS High RPM
Sparger Blockage
HCL NaOH DO% pH
DCW (g/L)
Inoculation Time (Hrs)
DCW (g/L)
DO %
DO % Avg
HCL (mL/h)
NaOH (mL/h)
pH
Fermentation 6 - High OD Start and Late Addatives
Sparger Blockage
HCL NaOH DO% pH PHB%
DCW (g/L)
Inoculation Time (Hrs)
DCW (g/L)
DO %
DO % Avg
HCL (mL/h)
NaOH (mL/h)
pH
PHB %
Fermentation 7 - High RPM
HCL NaOH DO% pH PHB%
DCW (g/L)
Inoculation Time (Hrs)
DCW (g/L)
DO %
DO % Avg
HCL (mL/h)
NaOH (mL/h)
pH
PHB %
Fermentation 8 - BRCS2 Strain
HCL NaOH DO% pH PHB%
DCW (g/L)
Inoculation Time (Hrs)
DCW (g/L)
DO %
DO % Avg
HCL (mL/h)
NaOH (mL/h)
pH
PHB %

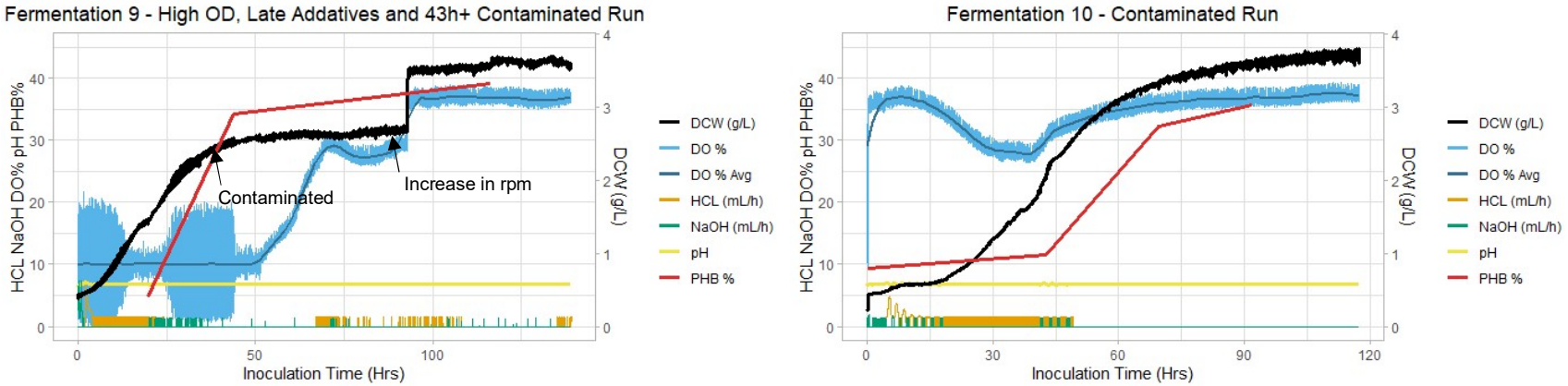


Figure 30: Plots of 10 Fermentations illustrating various changes over time. More details provided in Table 18. Fermentation 9 and 10 here show calculated DCW using on-line data which is informative in comparison, however due to the content of non-*M. parvus* bacteria this is imprecise. The actual values used in assessing outcomes were from offline data. DO% Avg was calculated from DO% using geom_smooth in ggplot2[47] using a LOESS function with a span of 0.1. Figures produced using ggplot2[47] and R[48].

6.2.5.3 Conversion of Online Fermenter Data

6.2.5.3.1 Online OD to Offline OD

OD data on the DASGIP fermenters were acquired by online instrumentation, here referred to as OD_{online}. This data was highly numerous being polled every 30 seconds allowing granular analysis of fermentation. This instrumentation had a range of 0-2.57OD. Comparison of this data with corresponding $OD_{offline}$ datapoints taken during fermentation (Figure 31) indicated the online data was informative but transformed compared to the offline data which had a range 0-15.64OD. Regression was performed to produce a data transformation allowing the OD_{online} data to be compared with other data in this thesis and infer biomass production.

53 datapoints from 8 experiments Ferm 1-5 and 8-10 taken offline in duplicate and averaged were compared to the OD_{online} readings at the same time points (Figure 31). As the relationship was clearly not linear alternatives were tested. An exponential ($0.1915e^{2.0354x}$), 2nd order polynomial ($4.1394x^2 - 3.2854x + 0.9699$) and 3rd order polynomial ($2.0772x^3 - 3.2632x^2 + 3.6735x - 0.46$) were optimised. A sum of square residuals (SSR) was calculated (sum of the squared differences between offline and online values) resulting in values as follows: linear 180.20, exponential 88.63, 2nd order polynomial 17.65, 3rd order polynomial 14.19. A lower SSR indicated a better fit. Although the 2nd order polynomial would be a better representation of the theoretical relationship in data, plotting transformed OD_{online} data against the $OD_{offline}$ data indicated the 2nd order relationship had a curve in the central region where the 3rd order was linear. The 3rd order polynomial was selected (Equation 7) for the transformation as minimal data distortion was preferable. The transformed data (Figure 31B) had an R^2 of 0.991.

Equation 7

$$OD_{offline}\ Equivalent = 2.0772\ OD_{online}{}^3 - 3.2632\ OD_{online}{}^2 + 3.6735\ OD_{online} - 0.46$$

A figure in the DASGIP software reported an ODCX value indicating biomass however linear regression of this data (not featured) indicated this was calculated as 27x OD_{online} so this output was dismissed.

OD_{online} appears affected by impeller rpm, likely due to an increase in fine bubbles. This is thought responsible for the non-zero intercept in the transformation from OD_{online} to $OD_{offline}$ in Equation 7. This also means in fermentations which had increased impeller speeds introduce a source of error. However, the transformation should account for this as it utilises data from those experiments and the rpm is expected to corelate with OD via $\%_{DO}$ reduction which in turn would cause impeller rpm to increase. Construction of a general model alternative to Equation 7 utilising rpm as a factor might be able to account for this more fully.

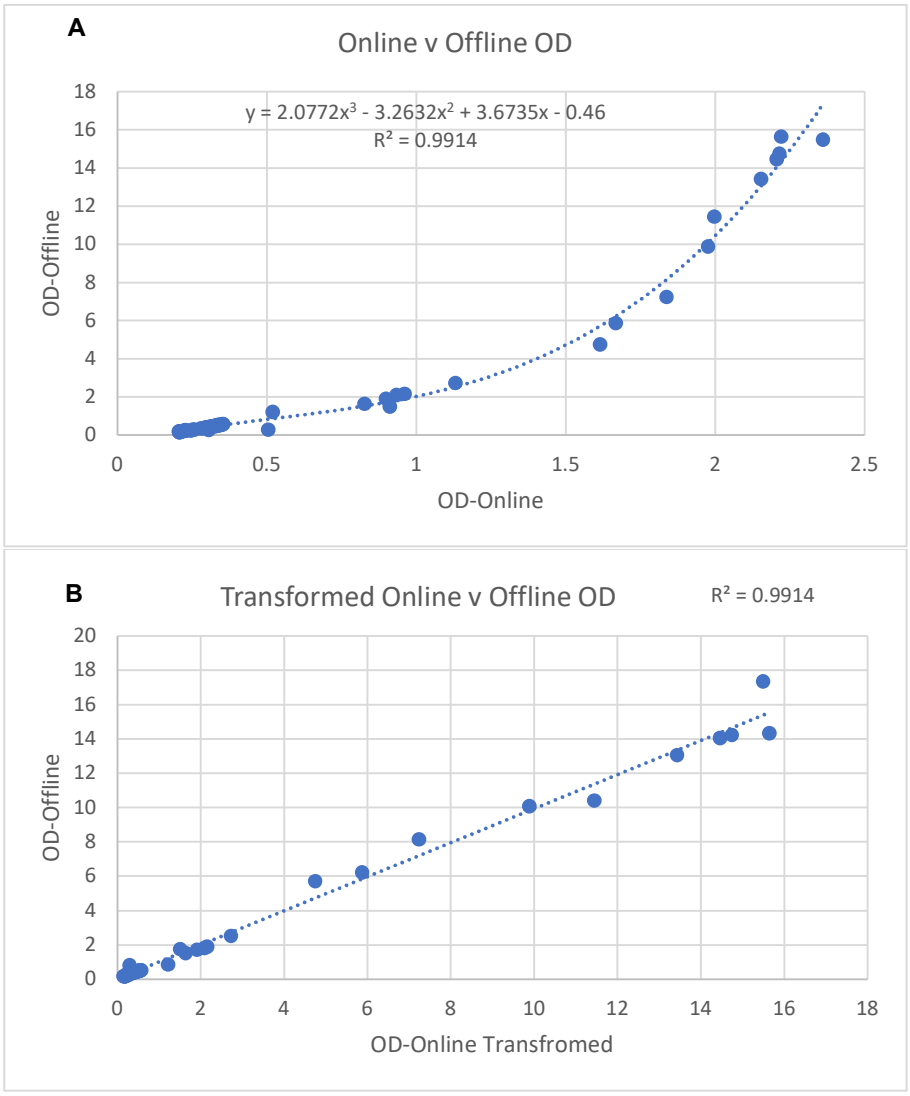


Figure 31 A and **B:** Scatter plot comparing OD_{online} and $OD_{offline}$ data gathered in 8 experiments producing a conversion between the two. n=53. **A** - Raw data plotted with the 3rd order polynomial regression. **B** – The same date with OD_{online} transformed with Equation 7.

As the mixed cultures Ferm 9 and 10 contain bacteria other than *M. parvus* OBBP the conversion by Equation 1 is not confirmed reliable and biomass drawn directly from taken samples for PHB analysis and Biomass values is utilised instead where possible.

6.2.5.3.2 Time Stamps and Dry Cell Weight

Time stamps in the DASGIP online data were converted to decimal hours using lubridate[49] implemented in R[48]. OD_{online} data was converted to DCW using Equation 1 and Equation 7. In Ferm 7 OD_{online} was not available due to probe failure and $OD_{offline}$ was used instead.

Data was trimmed on the time axis discarding anything after the application of nitrogen gas and termination of a run. In Ferm 5 and 6 sparger blockage stopped addition of new methane and air after 13 and 53 Hrs respectively and this data was retained and is discussed in section 6.3.2.3 and

the time points marked on their respective figures. In Ferm 3 a power cut occurred 8min after setup. Monitoring was resumed 2.5 hours later.

Difference in DCW from inoculation to end point was used rather than raw DCW to eliminate the effect of variation in starting DCWs/ODs.

6.2.6 Scaled Up Fermentation Results

Data from Fermentations 1-10 are shown in Figure 30 and further data in Table 18.

6.2.6.1 Biomass and General Growth Observations in Fermentations 1-10

Fermentations 1-8 (Figure 30) all appeared to show failed growth increasing in biomass from 1.2-2.3X their starting DCW. There was a tendency for biomass to increase slightly in the first 10-30 hours then decrease again over the length of fermentation, but effects were minor.

Attempting to rectify this failed growth a range of conditions and changes were attempted. These included:

- Limiting nitrate provision to 50% (Ferm 2) to allow PHB production.
- A pH set point of 6.9 rather than 6.8 (Ferm 2 and 3)
- Changing the provided base from NaOH to NH_4OH (Ferm 3)
- Elimination of any pH control (Ferm 4)
- Eliminating antifoam (Ferm 4)
- Increasing rpm to 800, 800 or 1200 rpm (Ferm 5, 6 and 7 respectively) to increase oxygen availability.
- Later addition of post autoclave NMS ingredients: Phosphate Buffer, Vitamin Stock, Fe-EDTA and $CuSO_4$ (Ferm 6, 7 and 9). This was in case of additive ingredient degradation making them unavailable for growth.
- Addition of 300ml of fresh NMS with additives at inoculation (Ferm 5). This was in case of NMS main ingredient degradation making them unavailable for growth.
- Use of a higher starting OD with an OD of 1.01 (Ferm 6) 0.58.(Ferm 7) and 1.64 (Ferm 9) rather than 0.1-0.3 (all Others)
- Use of *M. parvus* BRCS2 as an alternative strain (Ferm 8).

These changes in treatment did not produce desired growth in Ferm 1-8. Ferm 9 and 10 increased 6.8x and 9.6x respectively representing DCW increases of 2402 and 2868 mg/L based on offline samples and were considered successful fermentations on this basis. This increase can be observed in their charts (Figure 30). An increase in OD in Ferm 9 at 93hrs coincides with a manual increase in impeller speed to increase $\%_{DO}$ and push growth further, but this OD increase is believed resultant of increased turbidity not real increase in growth, and this was supported by offline data. The end of growth in Ferm 9 and 10 are believed to be due to nitrate exhaustion.

Both fermentation 9 and 10 were shown to be contaminated at some point by plating on LB plates at 30°C for 48 hours. Testing plates on LB medium indicated non-methanotrophic polytroph growth at 4 tested points in Ferm 10 indicating contamination throughout and was visibly yellow in colour by its endpoint which is not expected for pure *M. parvus*. Ferm 9 tested negative for non-methanotrophic growth at two time points 18 and 43 hrs, but positive at 116 and 165 hrs. This finding was also supported by nanopore S16 sequencing data discussed in Section 3 of this chapter (Figure 39).

From this it was concluded that under the conditions in Ferm 9 growth with only *M. parvus* OBBP was achieved to at least 43 hrs by which point growth was leaving the exponential phase and entering stationary phase and thus almost complete. Conditions in Ferm 9 were identical to those in Ferm 6 and 7 except an even higher starting OD (1.64 rather than 1.01 and 0.58) and no forced increase in rpm starting at 400 rather than 800 or 1200. In common they had late addition of post autoclave medium ingredients.

As Ferm 9 was successful with a lower starting rpm than unsuccessful fermentations with similar parameters this adds credence to over exposure to a gas component in the air or methane stream causing a negative effect. A much higher starting OD may have allowed the *M. parvus* to become metabolically active before such an effect took hold. It was suspected that some component essential to *M. parvus* is reacted with or stripped out by the fast gas flow. Options include CO_2 being essential for growth, or toxicity caused by overexposure to oxygen. As high gas flow and pH balancing are the major differences from the reliable growth in serum bottles this link to gas flow seems likely.

6.2.6.2 pH and Neutralisation Salts in Fermentations 1-10

A pH set point of 6.8 was utilised generally following the known successful growth achieved in the NMS medium set to pH 6.8[29]. This was controlled by addition of acid and base and in standard conditions this was HCl and NaOH. pH set point preference and resilience were tested more thoroughly in bottles in section 6.4.3.

More acid than base was added in all but one fermentation (Ferm 3, using NH_4OH) by an average of 2.78x (excluding the un-pH controlled Ferm 4) and the two contaminated runs that ran to complete growth had higher ratios of ~5x. This suggests an alkaline by-product is being produced during the fermentation.

With no pH control in Ferm 4, pH decreased over time from an initial 6.78 to 5.14 and continued decreasing until experimental termination at 72hrs. This was not accompanied by growth, indicating the acid base control was not responsible for fermentation failure. This is in line with suggestions that dissolved CO2 should result in a decreased pH[18]. This decrease in pH does not correlate with Acid/Base additions observed previously where additional acid was added thought to imply the pH would otherwise become more alkaline. With little to no growth from 34mg/L of bacteria I hypothesise that the change

in pH was not caused by the bacteria but by interaction of air or methane with medium ingredients. There may be an interaction of input gas in the presence of HCl and NaOH causing alkalisation in other fermentations where these are not present here. A run with no inoculum, only medium and no pH control and a similar run with pH control would illuminate this further.

It is of potential concern that excess HCl and NaOH neutralise to form NaCl salt (and H_2O) which could impede growth. This has been studied in *Methylocystis sp*. SC2[50] at ranges from 0 to 1.5% NaCl (0-256.7 mM) and was found to have a major impact on maximum OD attained at 128.34 mM and above. *M. parvus* strain MTS has been found to tolerate at least 85.6mM NaCl[22]. Using the acid and base addition data in Ferm 1-2 and 5-8 (Excluding the NH_4OH based Ferm 3 and non-pH controlled Ferm 4) had a calculated range of NaCl from 3.75 to 9.56mM. The two contaminated fermentations Ferm 9 and 10 which had more pH balancing had 28.80 and 34.88mM respectively. These are all below the effective levels outlined in the *Methylocystis sp*. SC2 study thus it was concluded that NaCl build-up was unlikely to be a contributing factor to the lack of growth.

To test the effect of NaCl production ammonium hydroxide (NH_4OH) was used in place of NaOH in Ferm 3 resulting in NH_4Cl and H_2O upon neutralisation rather than NaCl. This also proved ineffective at producing growth. It is noted that high levels of ammonia are known to be toxic and also provide an additional nitrogen source[6,51]. The levels of ammonium added here (9.47mM) were comparable to the 13.5mM NH_4Cl used as a nitrogen source in JM medium which has been used with *M. parvus* OBBP previously[52]. Whittenbury's Ammonia Mineral Salt medium[29] also contains 9.89mM NH_4Cl. Conversely other work in *M. parvus* has found no growth on 2.5mM NH_4^+ and above but this has not been evidenced by our own work. From this it was concluded that the NH_4Cl produced here should not have impeded growth.

6.2.6.3 Dissolved Oxygen in Fermentations 1-10

DO can be used as an indicator of oxidative cellular activity and growth, where oxygen is consumed this results in a decrease in $\%_{DO}$. DO is monitored by online sensors and impeller speed is increased to increase DO. The maintained set point for DO was 10% in all cases, below this value the impeller would increase above the baseline 400rpm to increase the DO. In all fermentations that did not successfully grow (excepting Ferm 5 and 6 which incurred a sparger blockage) $\%_{DO}$ did not decrease below 30% indicating oxygen availability was not a limiting factor in growth and that little cellular activity was taking place. DO generally trended upwards from inoculation towards 40% in failed fermentations.

With no growth under the standard control algorithm rpm remained at a minimal 400. In experiments Ferm 5, 6 and 7 rpm was forced higher in case starting OD capacity was limiting and caused failed culture growth. Lack of response to this, and lack of any decreases in $\%_{DO}$ in failed cultures suggest this was not a limiting factor.

In both experiments where a sparger blockage occurred (Ferm 5 and 6) stopping introduction of new methane or air, $\%_{DO}$ rapidly decreased. Ferm 5 appeared to show gradual increase in DCW over the length of experiment with a major increase in rate from 233mg/L at 12hrs to 433mg/L after 28hrs. The major increase started after sparger blockage at 13hrs which caused a gradual decrease in methane and air flow rates until the end of the fermentation (Figure 32). The increase in DCW was not corelated with an increase in rpm which remained level at 800 for the entire fermentation. This might indicate the reduced gas flow rates were a positive effect. The effect of oxygen availability is discussed further in section 6.6.1.1.

6.2.6.4 PHB Productions in Fermentations 1-10

Offline samples tested for PHB content in Ferm 1, 6, 7 and 8 indicated elevated levels of 6-14 $PHB_{\%DCW}$ despite not reaching a stationary phase OD indicative of nitrate exhaustion. This elevated PHB may be indicative of a stress response. This is supported by the PHB data point taken at inoculation in Ferm 6 of 2.88% which progressively increase to 8.69 $PHB_{\%DCW}$ over the length of fermentation. All other PHB datapoints were taken after 19+hrs of fermentation so do not illuminate this period.

The contaminated Ferms 9 and 10 reached 39% and 36% $PHB_{\%DCW}$ respectively. This is indicative of nitrate exhaustion and the increased levels correlate roughly with plateaus in growth (Figure 30). This is comparable but lower than the average 46.2% achieved by Helm et al.[26] and superior to others like Cattaneo et al.[42] (32%) and Helm et al.[53] (33.6%) in mixed culture fermentation.

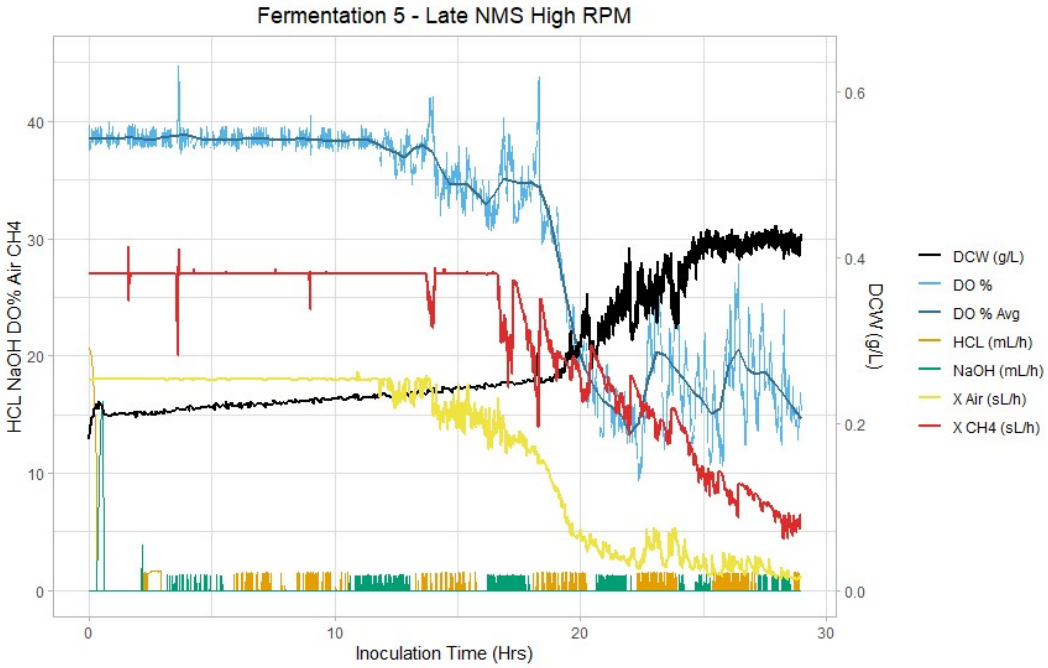


Figure 32: Fermentation 5 shown with Air and Methane flow rates decreasing after sparger blockage about 13Hrs. DCW increased after this time. This is an alternate version of that shown in Figure 30 with additional details and a rescaled DCW axis.

6.3 Section 2: Investigation of Fermentation Parameters

Due to difficulties incurred in scaled up *M. parvus* OBBP fermentation a deeper inspection of underlying causes was carried out over four experiments.

6.3.1 Oxygen Limitation

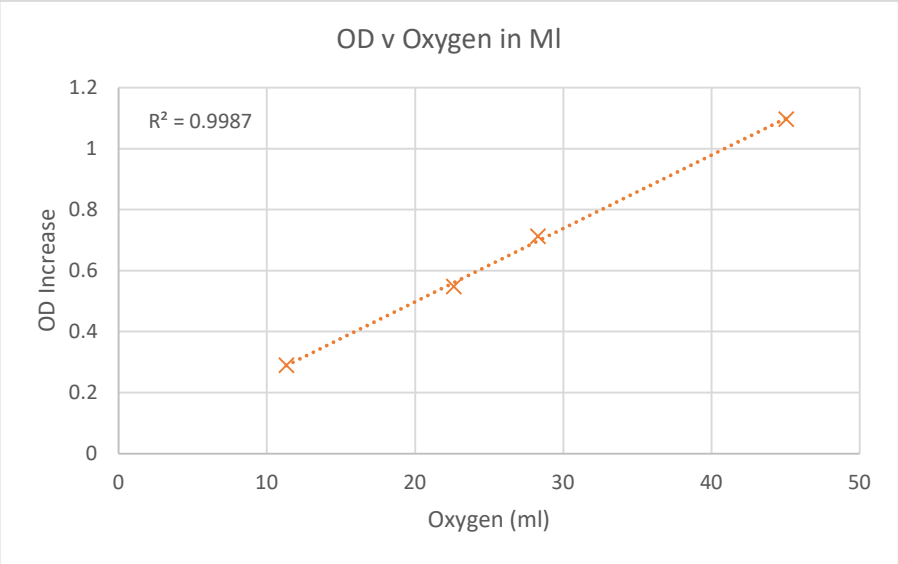
To test if the gas mix present in the fermenter feed was an issue (6.3.1.2) 5x250ml bottles each with 35ml of NMS medium but differing levels of air and methane were set up. These were inoculated to 0.02OD with *M. parvus* OBBP and grown for 3 days then bottles were sacrificed, and OD taken (Figure 33). Gas ratios were achieved by removal of a calculated volume of air or methane with a syringe allowing for gas expansion and supplying either air or methane to result in the 20% overpressure common to bottled fermentation experiments in this thesis (section 2.1.3).

As the ratios used were far below the 1.5-2 O_2/CH_4 optimal ratio, it was inferred oxygen was the limiting factor. Linear regression between OD increase from inoculation and O_2 volume supplied indicated strong correlation with complete predictive power ($R^2 = 0.9987$).

To allow these results to be applied more broadly a mean was taken establishing a ratio of 0.870±0.002 OD/ml O_2/ml medium. Using the established OD/mg conversion (Equation 1) a value for total DCW per ml of oxygen was calculated to be 0.221±0.006g DCW/L O_2 or 4.95±0.1344g DCW/mol O_2. This should be applicable to all fermentations using *M. parvus* OBBP though O_2 consumption may be affected by background maintenance and therefore deviate with faster or slower growth rates.

Using an inert replacement to methane such as argon might have been preferable but results here appear reliable. With pure gassed oxygen, higher oxygen to methane ratios and total oxygen could have been achieved capturing the oxygen to methane balance point expected to be 1.5 – 2 O_2/CH_4.

This experiment encompassed the 0.133 O_2/CH_4 ratio present in the DASGIP fermenter setup and showed no effect on growth beyond the expected oxygen limitation. In sparged fermentation there is a very high flow of gas compared to the static gas mixture here, and as indicated by $\%_{DO}$ in the failed experiments oxygen was far from a limiting factor. Thus this does not eliminate all possibility of gas flow causing issues with growth but shows the specific ratio was not the causing factor.



CH4 %	Air %	Oxygen %	Oxygen ml	Ratio O_2/CH_4
20	80	16.76	45.08	0.84
50	50	10.47	33.81	0.21
60	40	8.38	28.17	0.14
80	20	4.19	22.54	0.05

Figure 33: OD against ml of O_2 supplied to 250ml bottles of *M. parvus* OBBP in 35ml NMS after 3 days of growth. Plotted linear regression has an R^2 value of 0.9987. Accompanying table values rounded to 2dp. Oxygen taken to be 20.9476% of Air by volume[54,55]. A 40% CH4 60% Air bottle failed proper growth and was eliminated as an outlier. 20% CH4 80% Air is the standard condition for bottle fermentations used in this thesis.

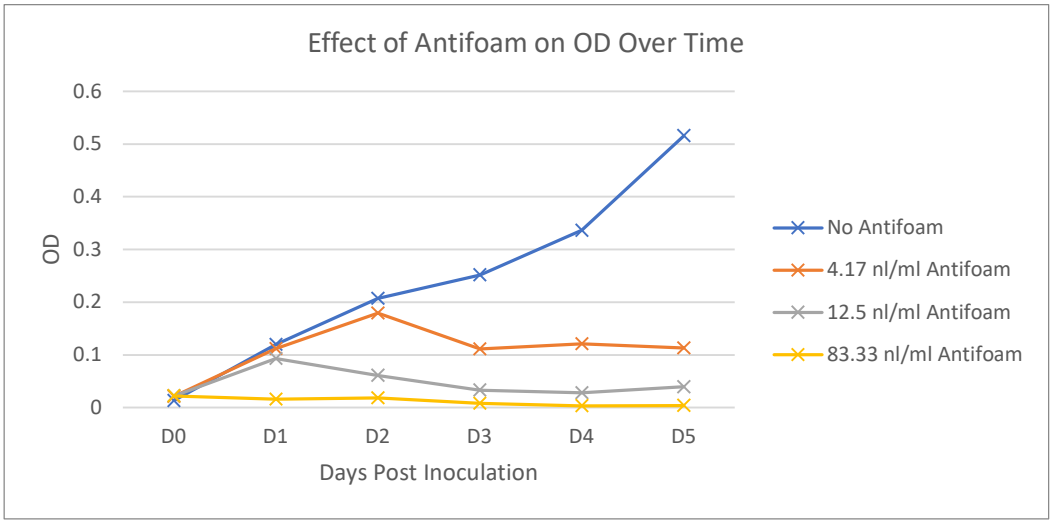


Figure 34: Effect of addition of antifoam on the growth of *M. parvus* OBBP over time. Each point is a single datapoint.

6.3.2 Antifoam

Antifoam 204 used in this work according to the manufacturer consists of a mixture of organic non-silicone polypropylene-based polyether dispersions. The possibility of antifoam causing fermentation failure was tested in a small experiment. Four 160ml bottles were setup with 24ml NMS and inoculated to 0.02OD with *M. parvus* OBBP. To each was added an amount of antifoam giving final concentrations ranging from 0-83.3nl/ml. This was incubated at 200rpm and 30°C for 5 days with OD taken on each day and no regassing took place. Results in Figure 34 showed a strong effect with any addition of antifoam resulting in suppression of growth increasing with dose to complete suppression at 83.33nl/ml. Growth in this experiment of even the control was surprisingly slow compared to other experiments, the effect of antifoam addition is strong, nonetheless.

I hypothesise this growth inhibition was due to antifoam forming a layer on the media surface inhibiting already challenging gas mass flow from the headspace into the media. In the highly perturbed state of a fermenter with an impeller it is suggested this effect would be diminished. Antifoam during fermenter setup was not tightly controlled but the standard 2-3 drops was measured to give ~90µl in 750ml or 14nl/ml so within the experimental range.

Antifoam elimination was tested in Ferm 4 and did not improve the state of fermentation. Other gas fermentation users of the same equipment with different microbes had successful fermentation with antifoam, and antifoam was used in the successful Ferm 9 and contaminated Ferm 10. From this it was concluded that antifoam was not the key inhibitory issue in scaled up fermentation, though it remains possible it could be an issue for *M. parvus* OBBP alone and the effect of any possible toxicity should be tested on a scaled up system.

6.3.3 Starting pH Set Points

The discrepancy between bottled growth success and bioreactor failure due to pH control was considered. It was hypothesised the standard starting point of pH 6.8 in bottles under standard setup may be modified to be higher or lower during growth resulting in the bacteria's actual preferred pH. Meanwhile in the fermenter the pH of 6.8 was forced to continue causing failure.

To investigate pH effects an experiment was run taking pH readings of the culture over time. Starting pH was achieved by varying buffer constituents. pH was taken on a SevenCompact S210 pH Meter with an InLab Flex-Micro pH electrode (Mettler Toledo, 30130862 and 51343164). This electrode allows readings of samples as low as 500 µl allowing repeat sampling from single bottles without affecting total volume unduly, this not possible with larger electrodes.

Table 19: Amounts of phosphate used and resultant pH values in pH-controlled experiment. pH values taken from inoculated bottles at initiation of experiment. All Values to 2dp.

Amount (mmol)		Starting pH	Molar Ratio
KH_2PO_4	$Na_2HPO_4.7H_2O$	Mean	KH_2PO_4 / $Na_2HPO_4.7H_2O$
1.48	0.15	5.96	10.14
1.17	0.30	6.39	3.88
0.75	0.51	6.81	1.47
0.56	0.61	7.02	0.92
0.39	0.70	7.19	0.56
0.26	0.76	7.34	0.34
0.11	0.84	7.52	0.13
0.04	0.87	7.58	0.05

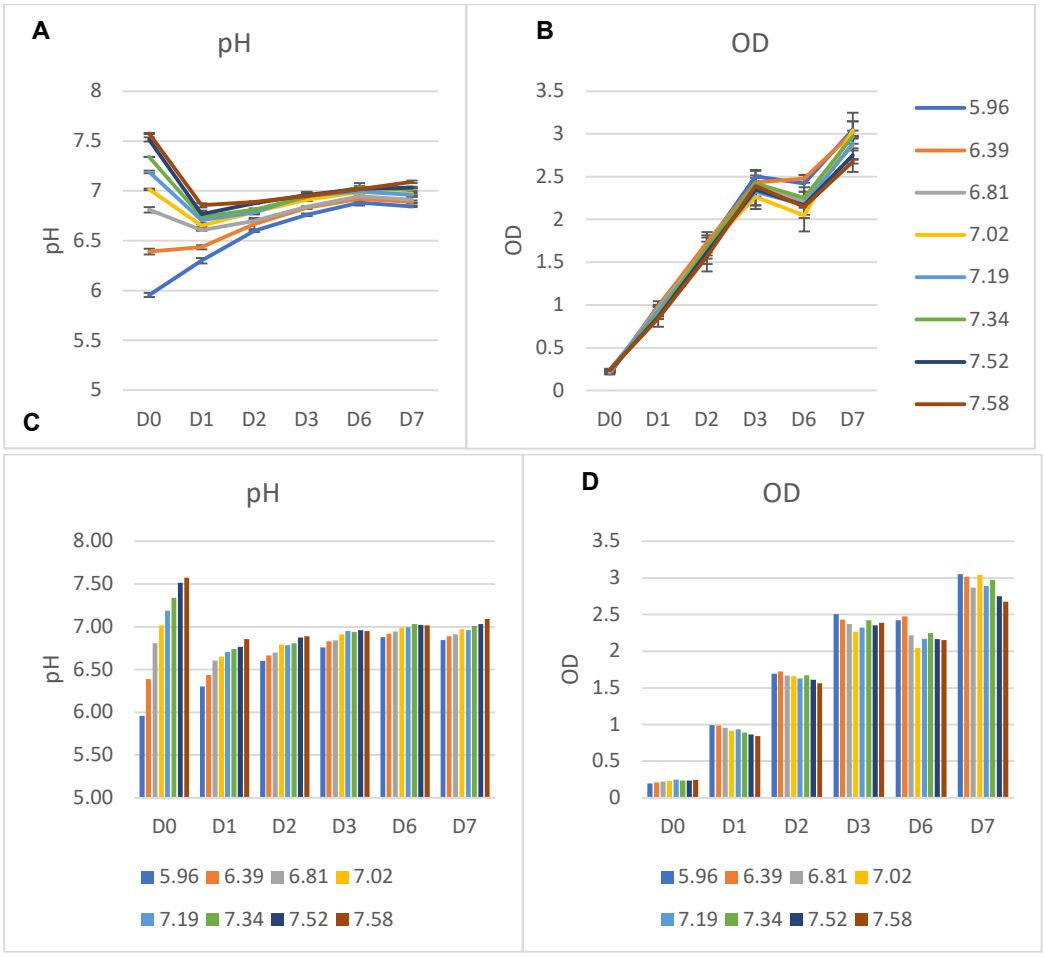


Figure 35 A-D: Changes in pH and OD over 7 days in 8 different starting pH conditions controlled by phosphate buffer balance. Each condition was run in duplicate, and averages are presented. Error bars are 1.96xSE. Upper and lower charts show the same data in different formats for clarity. Data in legends indicates the starting pH.

The pH-controlled experiment Figure 35 used varied ratios of KH_2PO_4 and $Na_2HPO_4.7H_2O$ which set the starting pH. These were added in place of the standard phosphate buffer to make a final volume of 24 ml NMS medium. *M. parvus* OBBP was centrifuged and resuspended in fresh phosphate free medium to eliminate carryover and inoculated into 160 ml bottles to give a starting OD of ~0.2 in in duplicate. Amounts of phosphate added and resultant starting pHs are listed in Table 19. Bottles were sealed, gassed and incubated at 200rpm and 30°C. The experiment ran for 7 days with 1ml samples taken for OD and pH analysis and bottles regassed daily except D4 and D5 where bottles were not regassed or data taken.

Starting pHs converge across multiple days and slowly rise over time towards a value of pH 7. This is comparable to the expected pH preference of *M. parvus* which grows effectively in NMS medium (pH 6.8)[29] equivalent to a $KH_2PO_4/Na_2HPO_4.7H_2O$ molar ratio of 0.83. pH values separate somewhat on day 7 but this may be confounded by the lack of regassing that occurred due to COVID regulations on D4 and D5. The values order was conserved relative to each other. The starting pH had no significant effect on OD up to D3 with a potential lower final OD corelating with lower starting pH by D7. This showed resilience to pH variation and some level of regulation within the culture over prolonged growth. The increase in pH over time is contradictory to expectations of acidification[19]. A similar trend of pH tending towards 6.9 was noted by other authors in the α-proteobacterial *M. trichosporium* OB3b[1] with an extended lag time and decreased growth rate appearing above a starting pH of 7.5. This data allays worries over pH sensitivity contributing to fermentation failure at least within the presented ranges.

This experiment only tests the range possible from the phosphate buffer, careful design might achieve the maximum range of the buffer expected to be 5.7 to 8.0[56]. Other buffers might provide pH ranges beyond this that would be inhibitory to growth and of interest in future investigation. A buffer free datapoint would also be illuminating however a phosphate source is necessary for growth thus some must still be provided in an alternate form.

6.3.4 Transfer of Failed Fermentation to New Media

100ml of culture from the endpoint of failed Ferm 6 was transferred either directly to 500ml bottles ("re-bottled") and gassed, or centrifuged and re-suspended in 100ml NMS, bottled and gassed. Bottles were grown over 10 days with OD and regassing on D3, D7 and D10 (Figure 36). This showed growth in the resuspended culture after a lag phase and gradual decline in the rebottled culture.

From this it was concluded at least part of the fermentation failure was an irreversible change to the medium that occurred during fermentation. Possible causes include degradation of the medium components or the build-up of a toxic by-product which were cleaned away during resuspension.

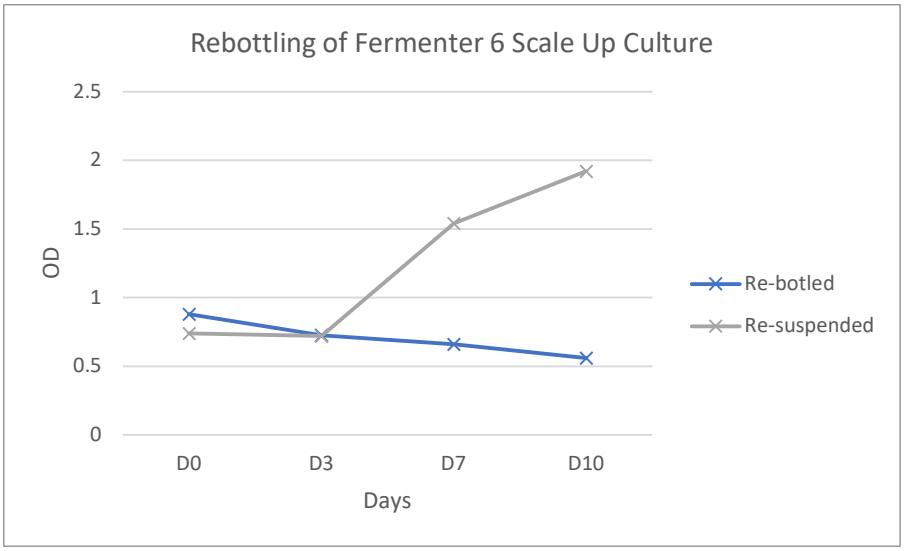


Figure 36: OD over 10 Days of 100ml Ferm 6 endpoint culture either untreated or resuspended in an equal volume of NMS medium in 500ml bottles. Datapoints are from single replicates.

6.4 Section 3: Mixed Culture Species Prevalence

The contaminated fermentations 9 and 10 showed growth until apparent nitrate exhaustion and high PHB production resulting in an overall successful fermentation. 16S nanopore sequencing was carried out on stored fermentation broth samples taken at four time points in each fermentation to investigate the species content.

6.4.1 Nanopore Sequencing Method

gDNA extraction was carried out using GenElute Bacterial Genomic DNA Kit (NA2110, Sigma-Aldrich) using guidance for minimally sheared DNA, RNAse A treatment and the Gram +ve protocol for final results. The Gram -ve protocol was also tested for comparison. Extracted gDNA was quantified by nano spectrophotometer and Qubit with high sensitivity kit (Q33230, Invitrogen).

Samples (Table 20) were processed with a 16S Barcoding Kit 1-24 (SQK-16S024, Oxford Nanopore Technologies - ONT). This kit uses supplied 27F (5' - ATCGCCTACCGTGAC - barcode - AGAGTTTGATCMTGGCTCAG - 3') and 1492R (5' - ATCGCCTACCGTGAC - barcode - CGGTTACCTTGTTACGACTT - 3') 16S primers[57] providing an expected read length of 1451bp in *M. parvus* OBBP. Primers also include additional regions for barcoding. The primer region binding to the genome is underlined. The forward primer was typed 27F-CM by comparison to literature[58].

Table 20: Nanopore Samples Discussed

Sample	Hours Post Inoculation	gDNA Extraction Gram +ve	gDNA Extraction Gram -ve
Ferm 9	18	Yes	
Ferm 9	43	Yes	Yes
Ferm 9	116	Yes	
Ferm 9	165	Yes	Yes
Ferm 10	19	Yes	Yes
Ferm 10	53	Yes	
Ferm 10	70	Yes	
Ferm 10	92	Yes	Yes
C. acetobutylicum		n/a	n/a
M. parvus OBBP		Yes	Yes
gDNA -ve Ctrl.		n/a	n/a

* *Clostridium acetobutylicum* gDNA provided by Ruth Cornock. 16 samples in total.

Sequencing of samples in Table 20 were run on a MinION Mk1C with a FLO-MIN106 R9.4.1 cell (ONT). The run length was 18hrs and 48min generating a total 97.5 Gigabytes of fast5 format raw nanopore data and an estimate 8.59 Gbp.

Basecalling was carried out using GUPPY v6.3.7 (ONT) in GPU accelerated mode with CUDA 12.0 on a GeForce Super 2070 Maq-Q with 8GB dedicated RAM (N18E-G2 MAX-Q TU106, NVIDIA) using the appropriate basecalling model for the kit and flow cell combination. Basecalling was tested at the three accuracy levels FAST, HAC (High Accuracy) and SUP (Super Accuracy) for comparison but SUP was used for final results. All base calling parameters were at default except *gpu_runners_per_device* to 3 from 12 in SUP. Changed parameters of basecaller beyond accuracy level only affects processing time not outcome. SUP basecalling generated 3,832,505 reads across all samples and controls. This resulted in ~130,00-180,000 reads per sample, in excess of that required for interpretation but thus giving precise results.

Initial analysis of basecalled sequences used Epi2Me Agent v3.5.7 (ONT) "Fastq 16S v2022.01.07" protocol with BLAST e-value filter of <0.01, min coverage 30% and min identity of 77%. This pipeline uses BLAST search to assign species data to each read the results of which can be used quantitively. Resulting outputs from Epi2Me Agent was missing data due to an issue in the analysis pipeline, most likely Kraken2 giving a *genus_taxid* of -1 instead of a number, and/or an *lca* of 0 instead of 1 resulted in species and genus not being assigned in their respective columns on some lines, despite it being assigned to strain level on the same line in another column. This is believed due to multiple genera being present in the top blast hits due to limited sequencing data in the database. This behaviour is applied to avoid erroneous lineage assignment, but the issue appeared greater when more data was provided to the pipeline so was concluded to be in error. The missing species, genus and sometimes family affected 25-86% of the data. As *M. parvus* OBBP was the major affected species the correctness of its identification in these reads was accepted.

To rectify this first a legacy deprecated script from ONT "*Analysis_of_EPI2ME_16S_CSV_Output.ipynb*" available at the end of this chapter and implemented in Epi2Me Labs (ONT), was used to improve taxonomy data present in the output from the Fastq 16S analysis. A script "*R_CSV_Fixing_Script.R"* was then written and implemented in R[48] to fix the missing strain and genus data. The .ipynb was then used to summarise reads by taxonomic level for final analysis. The R script corrected missing species and genus data but not family data which affected 2.4-4% of sequences depending on basecaller accuracy after correction.

From this summary relative abundance of each species, genus or family in a sample was calculated as a percentage of the total reads in that sample.

6.4.2 Comparison of gDNA Extraction Methods

As the bacterial makeup of the contaminated fermentations was unknown it was deemed most correct to use the Gram +ve gDNA extraction protocol to not underrepresent those species. To identify if this produced any biases select Gram -ve controls were also performed. Quantified gDNA extracted on paired samples with both methods are compared in Figure 37.

DNA quantities extracted showed for *M. parvus* OBBP gDNA yields were significantly higher (paired t-test, 3, p=0.037) with the Gram +ve methodology despite being a Gram -ve bacteria. This was true for each of the 4 tested paired samples. Conversely for the contaminated fermenter samples Gram -ve yields were greater than Gram +ve yields (paired t-test, 7, p=0.007). The surprisingly low yield of *Methylocystis* DNA has also been noted by previous authors attributed to the complexity of methanotroph internal membranes[26].

The actual quantity run on the nanopore is standardised to 10ng input to the 16S kit, but extracting differing quantities of DNA from the same sample may indicate some bacterial are extracted more completely than others, changing the observed relative abundance.

Comparison of the processed nanopore reads indicated some variation in results when the Gram -ve gDNA isolation method was used (Table 21). In Ferm 9 - 165hrs the Gram +ve *Humibacillus xanthopallidus* made up 12.82% of the Gram +ve extracted gDNA but 0.17% of the Gram -ve extracted gDNA. Underrepresented samples result in a relative increase in the content of other species particularly *M. parvus*. The Ferm 10 samples show considerable variability despite any discernible consistent Gram +ve or -ve effects. This indicates a sizable source of error. This illustrates the necessity of the Gram +ve protocol and that the sequencing method is only as good as the samples supplied to it.

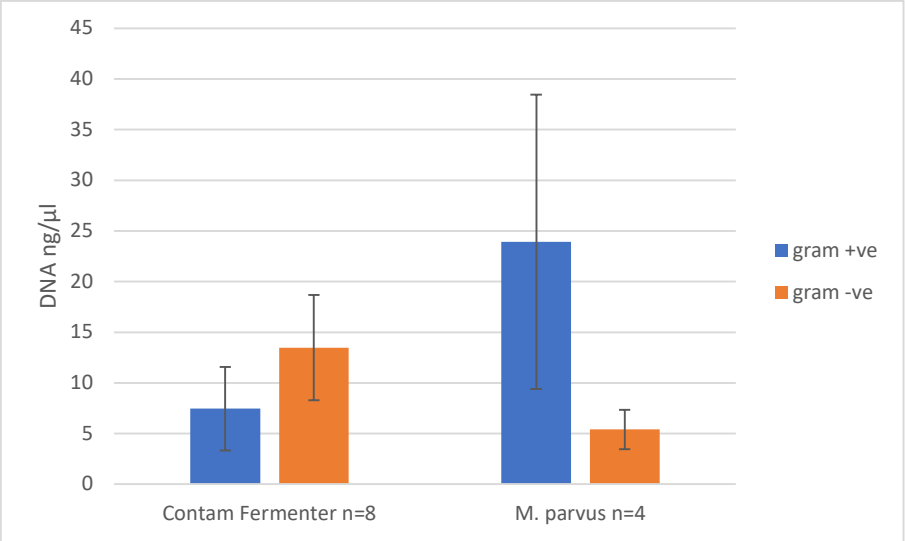


Figure 37: Comparison of gDNA quantities extracted from 8 Fermenter samples and 4 *M. parvus* samples using either a Gram +ve or -ve methodology. Paired samples were used for the treatments. Error bars are 95% Cl.

Table 21: Comparison of species make-up of samples depending on gDNA extraction method.

gDNA Extraction Method - % of Sample Total		
	gram +ve	**gram -ve**
Ferm 9 – 165 Hrs		
Methylocystis parvus	63.06	80.72
***Humibacillus xanthopallidus* (gram +ve)**	12.82	0.17
Paracoccus yeei	12.11	9.95
Paracoccus sanguinis	2.68	0.76
Roseomonas mucosa	1.77	0.36
Ferm 10 – 19 Hrs		
Methylocystis parvus	78.48	86.45
Sphingomonas desiccabilis	8.86	5.52
Sphingomonas molluscorum	4.02	2.59
Massilia suwonensis	3.72	1.58
Massilia brevitalea	1.12	0.48
Ferm 10 - 92hrs		
Methylocystis parvus	72.23	70.84
Sphingomonas desiccabilis	15.54	14.39
Sphingomonas molluscorum	7.38	6.27
Massilia suwonensis	1.11	3.60
Massilia brevitalea	0.34	1.08

Variations of greater than 0.5% between the two extraction methods. All species are Gram -ve unless otherwise noted. Data taken with basecalling at SUP accuracy.

6.4.3 Comparison of Basecaller Accuracy Levels

Raw data from nanopore sequencing needs to be basecalled to produce sequences before they can be analysed. A range of software is available for this, but the standard is GUPPY. GUPPY can be utilised at three accuracy levels with a trade-off in computational time. To inform future experimental design the effect on the results of lower accuracy was investigated. If this technique were to be applied industrially, faster data response times by using lower accuracy might be preferred.

Some amount of misclassification of barcodes is expected but with SUP accuracy 3 reads of the total ~4 million were misclassified to the -ve control indicative of the underlying level. Compared to FAST - 11 and HAC – 8. This indicates higher accuracy calling decreases barcode misallocation, but the level of misallocations was already at an acceptable level. Various other outcomes are shown in Table 22.

Table 22: Outcomes of three levels of Basecaller model used in GUPPY

Accuracy Level	FAST	HAC	SUP
Run Time (hrs)	2.15	4.4	30.1
Total Good Reads	4,096,541	4,087,603	3,833,536
Misclassified To -ve Ctrl	11	8	3
Misclassified to Unused Barcodes (n=8)	39	17	13
Misclassified % of Total*	$3.35x10^{-3}$%	$1.65x10^{-3}$%	$1.13x10^{-3}$%
Unclassified barcode (% of Total)	116,193 (2.83%)	58,235 (1.42%)	47,185 (1.23%)
Mode Quality	11.85	13.75	14.65
Mode Length (bp)	1,513	1,421	1,485
% Accuracy†	92	95	96

Basecalling the same 97.5 Gigabytes of fast5 format raw nanopore data with an estimate 8.59Gbp. †Average accuracy to the top BLAST result for each read classified[59]. *Calculated as average reads misclassified to unused and -ve barcodes (n=9) multiplied by the total number of barcodes possible (24) assuming equal misallocation rate.

Note the FAST, HAC and SUP have differing automatic quality score cutoff parameters (8.0, 9.0 and 10.0 respectively) in addition to improved base calling capability. This will result in more discarded “bad” reads and increase in final quality beyond higher quality base calling alone. Reads unclassified to barcode provide a source of error if they are not evenly distributed across strains. Their reduction by 2X in HAC and 2.46X in SUP significantly reduces this error.

Nanopore suggest that HAC takes 5x to 8x longer than FAST and SUP takes 3x longer than HAC. Results using the computational setup here showed FAST 2.05x-> HAC 6.84x-> SUP. Note speeds here could be more optimised but are indicative, SUP particularly was limited by available RAM.

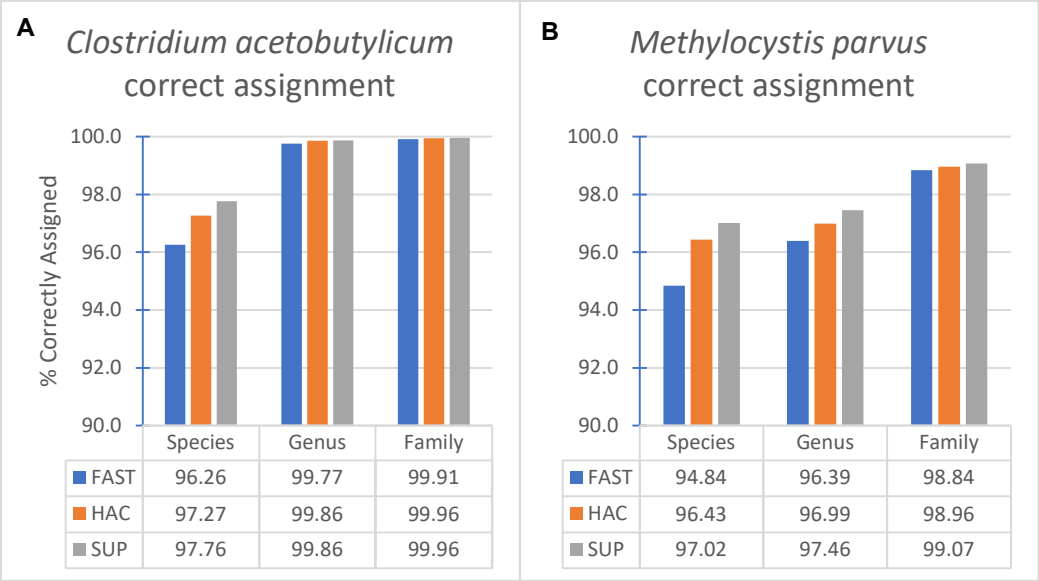


Figure 38: Correct lineage assignment rates in single species control samples for **A** - *C. acetobutylicum* and **B** - *M. parvus*. Correct assignment calculated as by number of correctly assigned reads in that sample divided by the total reads in that sample . Correct family assignments of *C. acetobutylicum* was to Clostridiaceae and *M. parvus* was to Methylocystaceae.

16S sequencing of mixed cultures suffers from background noise of reads with misassigned lineage data. These are due to sequencing errors and are generally assigned to closely related species, thus viewing data at genus level will eliminate much of this noise and at family level even more. This is particularly applicable here where a small number of species are being monitored compared to environmental, gut samples, or undefined mixed cultures containing a high diversity of species. This noise and its reduction are illustrated by the high number of identified lineages despite most containing little data e.g. FAST across all samples produced 1947 species, 643 genera and 107 families where 98.61% of species contain less than 0.1% of the data and so are considered spurious. These spurious lineages were reduced by higher accuracy basecalling but were still significant e.g. SUP 1561 species, 518 genera, 87 families.

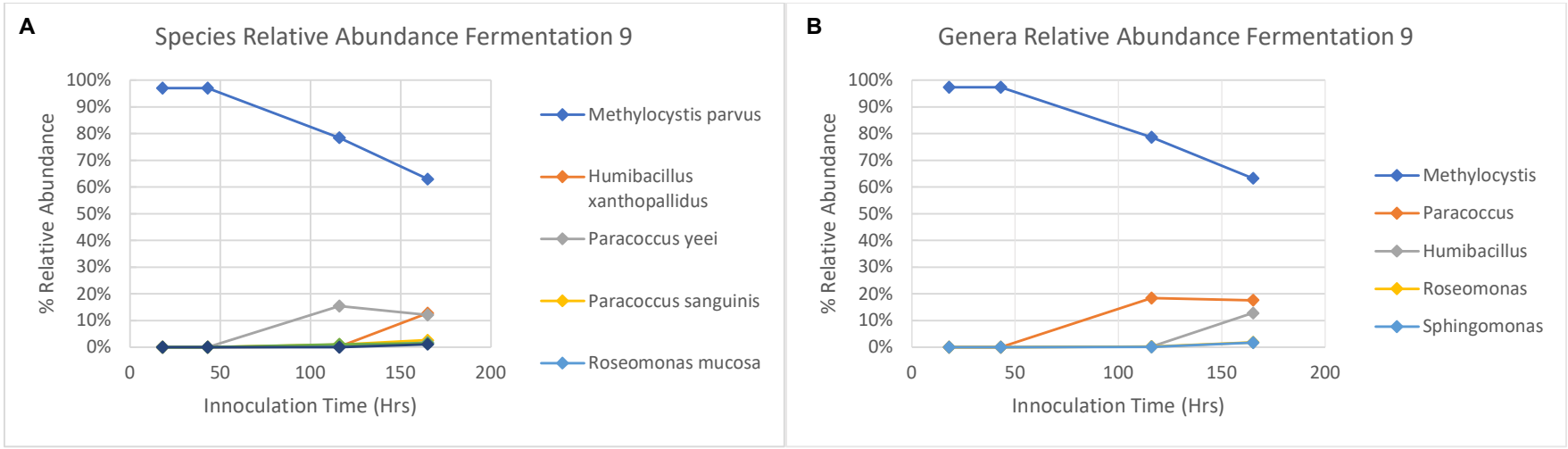
This type of error will vary for each species depending on the number of closely related species sequenced in the comparison database. Two examples from control data are laid out in Figure 38. All data points indicate improved assignment accuracy with higher accuracy basecalling. There is almost perfect assignment in *C. acetobutylicum* at genus level and above. *M. parvus* is less strongly assigned at each combined accuracy/lineage level. This would suggest a preference for family level data but due to the previously mentioned issue with lineage assignment, data was more complete at genus and species level and provides more granularity of information.

Species that makeup the misassigned reads in Figure 38 for *M. parvus* were all at 0.24% abbundance or lower, and were largely other closely related methanotrophs and methylotrophs e.g. *Methylosinus sporium*, *Methylocella*

silvestris, Methylocystis heyeri and *Methylosinus trichosporium* with less related species making up less of the overall count. At genera level the highest count of non-*Methylocystis* was 0.35% being the closely related *Methylosinus*. As *Methylocystis* was the majority in all samples this provides an effective noise floor for identifying real species and genera as above this prevalence. A level of 0.5% relative abundance was chosen.

As the *C. acetobutylicum* control is the only one to contain that species this gives the ability to quantify misclassified barcodes from one barcode to another. This could be a source of error if prevalent. Calculations indicated at FAST accuracy 0.1% of sequences were assigned to the wrong barcode, at HAC 0.05% and at SUP 0.06%.

It is notable that along with basecaller accuracy, the chemistry of the flow cell is also influential on accuracy. This experiment was run on a R9.4.1 flow cell with an expected raw read accuracy of 98.3% when called with SUP. The new R10.4.1 flow cell should be more accurate achieving 99.6% accuracy in ONT's published data[60] significantly improving results, but the 16S Barcoding Kit is not available for that chemistry at time of writing.



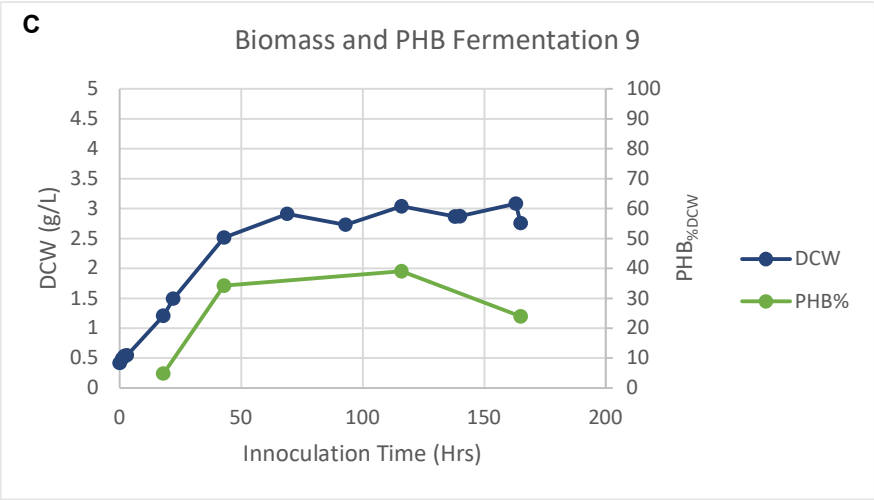


Figure 39 A-C: Showing the percent makeup of fermenter samples from Ferm 9 at four time points. Data acquired by nanopore of gDNA extracted samples processed using the 16S Barcoding Kit 1-24. Lineage levels making up >1% of total relative abundance are shown. **A** - Shows break down by species and **B** - by genera. **C** - Biomass and $PHB_{\%DCW}$. Biomass data as DCW here is taken from offline samples rather than the online readings in Figure 30.

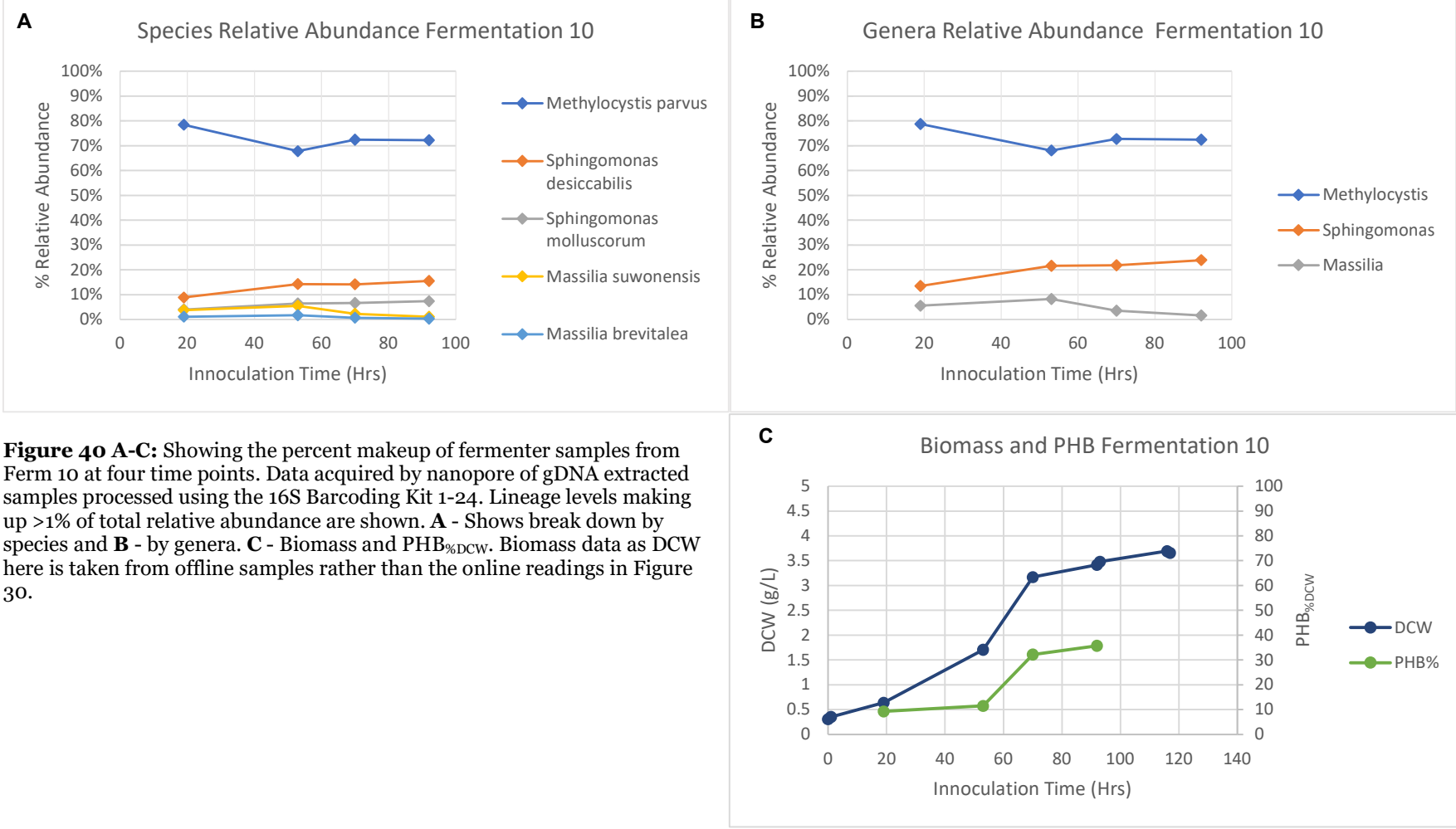


Figure 40 A-C: Showing the percent makeup of fermenter samples from Ferm 10 at four time points. Data acquired by nanopore of gDNA extracted samples processed using the 16S Barcoding Kit 1-24. Lineage levels making up >1% of total relative abundance are shown. **A** - Shows break down by species and **B** - by genera. **C** - Biomass and $PHB_{\%DCW}$. Biomass data as DCW here is taken from offline samples rather than the online readings in Figure 30.

6.4.4 Nanopore Data Results for Fermentation 9 and 10

6.4.4.1 Genera Makeup and Relation with Fermentation Data

The final data for Fermentations 9 and 10 are shown in Figure 39 and Figure 40 respectively. These were processed as discussed previously with SUP level basecaller accuracy. In text constituents are referred to at genera level. No *Methylocystis* species except *M. parvus* was identified to a meaningful level (>0.5%) at any point and so this is used throughout. The species of these genera are considered in the following section (6.5.4.2).

Ferm 9 (Figure 39B) found 5 meaningful genera present above 0.5%. No genera except *Methylocystis* was present for the two timepoints in the first 43 hours showing successful growth was initially in pure culture. *Paracoccus* was present by 116 hrs at 18.41% reducing the relative abundance of *M. parvus* to 78.68%. By 165hrs *M. parvus* had reduced further to 63.23%, *Paracoccus* remained similar and the Gram +ve *Humibacillus* had emerged to 12.81%. *Roseomonas* and *Sphingomonas* were present in the 116hrs sample at low levels (0.23% and 0.10% respectively) but became more prevalent in the final sample both to ~1.75%.

When compared with the PHB data, 34.19% $PHB_{\%DCW}$ was produced with only *M. parvus* presence in the first 43 hrs and *it* exhausted most available nitrate. Overall DCW remained generally stable with a slight increase for the remainder of the experiment. The growth of the other genera in apparently nitrate exhausted conditions suggest they were consuming fermentation byproducts, dead cell matter or predating on the *M. parvus*. As PHB increased marginally to 39.04%$_{DCW}$ with the appearance of *Paracoccus* this suggests this might have also produced PHB. PHB decreased to 29.32 %$_{DCW}$ by 165hrs likely consumed by another genera appearing at this time.

Ferm 10 (Figure 40B) found three meaningful genera present above 0.5%. All three were present from initial sampling. *M. parvus* initially decreased from 78.74% then remained consistent between 72.86% and 68.05% , *Sphingomonas* initially increased from 13.49% to remain between 21.65% and 23.91%, *Massilia* was more variable initially being 5.55% then increasing to 8.27% before decreasing across the next two sample to 1.66%.

The relative abundance of the genera remained approximately similar throughout as DCW increased to the point of levelling off and assumed nitrate exhaustion after 70hrs and beyond. $PHB_{\%DCW}$ increased to 32.27% to the stationary growth phase and increase further to 35.73 $PHB_{\%DCW}$ by the last sample. Though it is noted this was at 92hrs and the decrease in Ferm 9 was recorded after 116hrs.

6.4.4.2 Species Makeup

All species mentioned in this section are Gram -ve unless otherwise noted. *M. parvus* was the only identified *Methylocystis* and it was concluded that in all cases this was the intended *M. parvus* OBBP inoculum. Species not found at above 0.5% relative abundance in any samples are considered spurious and not discussed here.

Humibacillus xanthopallidus[61] is the only species in its genus thus its species identity in Ferm 9 was assigned confidently.

Paracoccus in Ferm 9 appears to be made up in majority by *Paracoccus yeei* (>70.6% of *Paracoccus*) with a minority of *Paracoccus sanguinis* (<15.6%), *Paracoccus chinensis* (<8.7%) and *Paracoccus marinus* (<4.8%). These high percentages cause tentative suggestion that all 4 species were present but it's possible they share highly similar 16S sequences resulting in high levels of misallocation. This exemplifies the advantage of genera level for the overview of the entire experiment with current sequence accuracy. *P. yeei* and *P. sanguinis* are generally associated with human clinical samples[62]. As noted previously *Paracoccus* addition did not result in a decrease in PHB. *Paracoccus denitrificans* though not identified here, is known to produce PHB, as are other less well studied species of the genus[63]. I was unable to establish if this trait is shared in the species identified here.

Roseomonas mucosa[64] identified in Ferm 9 was the only species identified of is genus and is noted for its mucoid characteristics and is commonly associated with human clinical samples.

Sphingomonas desiccabilis[65] and *Sphingomonas molluscorum*[66] were identified in both Ferm 9 and 10 and always in a 3:1 to 2:1 ratio. I suspect this ratio is due to incorrect lineage assignment as the introduction of both species to both fermenters is unlikely. *S. molluscorum* has been noted to have remarkable antimicrobial effect on some Gram +ve bacteria[66]. Species of the genus are generally mucoid due to polysaccharide secretion and yellow or orange coloured. The yellow colour was clearly visible in the colour of the Ferm 10 and not as strongly visible in Ferm 9 (Figure 29B) where they were less prevalent. Some *Sphingomonas* species have been noted to be effective PHB producers[67].

Massilia suwonensis[68] and *Massilia brevitalea*[69] both appeared at a constant 3:1 ratio in Ferm 10 throughout its increase and decline again potentially indicating the separation in species sources from lineage assignment as noted for *Sphingomonas*. Some *Massilia* species have also been noted to be effective PHB producers[70].

Though not interrogated here, lineage data is also produced down to the level of individual strain by selecting the best 16S BLAST match but does not subject this to the same level of stringency as higher levels of lineage identification. Where strains are almost identical in their 16S sequence this may also be misleading. Species level lineage data may also be misleading in its specificity where 16S data is not available for new or otherwise described species.

16S sequencing using 16S_U515F and 16S_U1071R primers of colonies grown on LB plates from the final fermenter cultures were colony PCR'd and sent for Sanger sequencing. These were identified using NCBI BLAST[71] and showed *P. yeei*, *P. sanguinis* and *S. molluscorum* were isolated pure from Ferm 9 and *S. molluscorum* only was isolated from Ferm 10.

6.5 General Discussion

6.5.1 Failure of Ferm 1-8

As shown by the number of publications in the field of methanotrophic fermentation including *M. parvus*, the challenges found here with successful fermenter growth have either not been experienced by other authors in the literature or due to the paucity of negative result publication they have never been reported. Solving the issues found here is important in informing failure states that must be avoided if applied industrially.

As discussed in the background and illustrated by the bottled fermentation experiment (6.4.3), *M. parvus* should be capable of growth at a broader range of pHs than those found in any of the fermentations. Ferm 4 showed that pH control was not responsible for lack of growth. It has been conclusively shown that pH itself is not a cause of fermentation failure but the difference in acidifying pH change from the non-pH balanced Ferm 4 compared to alkylation in other fermentations and 6.4.3 suggest some medium interaction including HCl and NaOH could be a major factor. Antifoam (6.4.2) also showed strong negative effects in bottles was considered an unlikely factor in stirred and sparged reactors.

6.5.1.1 O_2 Toxicity

Although the fermentation gas mix as tested in 6.4.1 showed no effect on growth in bottles, sparged gas remains the major difference between successful bottled growth and the failed bioreactor fermentations and so warrants further consideration.

It has been shown in pressurised reactors which can attain higher gas dissolution there was an optimal growth of *Methylocystis sp*. GB25 at O_2 of 1 mg/L[5]. This equates to 13.24%$_{DO}$ at 30°C. There was however no major drop off and in the range 0-25mg/L (0-330.91%$_{DO}$) growth rates only decreased by 50% from optimum. This optimum is in line with the %$_{DO}$ set point of our fermenter runs at 10%$_{DO}$ however without competing growth the high gas flow rate trended towards 40%$_{DO}$ as predicted by the 60% CH_4 40% Air input ratio. This should not have been sufficient for complete failure, however in the α-proteobacterial methanotroph *M. trichosporium* OB3b formaldehyde formation was found at high oxygen levels leading to cell toxicity and death[27].

A patent by Calysta for a reactor design for SCP production from *M. capsulatus* Bath in mixed culture notes excess oxygen can be an issue for growth and defines a maximum of 10ppm O_2 and a target of 3ppm equivalent to 10mg/L and 3mg/L[13]. This indicates higher optimal oxygen levels than found in the pressurised experiment, but this requirement may be modified by the presence of the other bacteria present and expectation of very fast activity in high growth and optimised conditions. A failure state caused by high DO was also specified in which growth completely stopped and could sometimes not be revived. Observation showed internal bacterial structures had become disrupted. This was potentially attributed to activation of prophage[13]. As noted in section 3.4.2, no prophage have been identified in *M. parvus* OBBP. A

recent 1L scale fermentation[72] specified a reduced gas flow rate for the first 15h of 2.4 reactor volumes per hour, 9x lower than the 21.5 used at all points in this work. This was specified to reduce growth inhibition due to excessive oxygen transfer. Another failure state in the Calysta process has been described in which pH is lowered and becomes unretrievable. This was attributed to the build-up of formic acid for which the lowering of ammonia and/or oxygen inputs was a suggested response[13].

From the above it is concluded that high $\%_{DO}$ may have contributed to the failure of Ferm 1-8 especially via build-up of a toxic metabolite like formaldehyde.

6.5.1.2 CO_2 Essentiality

The previously mentioned pressurised reactor experiment found an optimal dissolved CO_2 occurs at 150 mg/L[5]. This does indicate some benefit of CO_2 presence in the media but did not elucidate observations as levels approach zero.

From our fermentation set up CO_2 made up 0.0165% of input gas. Using Henry's Law[54,73,74] this was calculated to give an expected CO_2 content of 0.24mg/L far below that found to be optimal in the pressurised experiment. Degassing of CO_2 from culture media is utilised in the Calysta process by passing a "driving gas" like nitrogen through culture media to remove excess CO_2 illustrating this effect[13].

It has been discussed in literature and illustrated in section 1.2.2 that CO_2 is essential to metabolism by *M. parvus* as its assimilated as part of the serine cycle in the production of oxaloacetate from phosphoenolpyruvate among other reactions. Experimentally 50-60% of biomass is derived from CO_2 in Type II methanotrophs like *M. parvus*[75,76]. Stoichiometrically one CO_2 is expected to be utilised for every two formaldehydes entering the serine cycle[1]. The inclusion of the ethylmalonyl-CoA (EMC) pathway for the regeneration of glyoxylate which also takes in CO_2 makes the balance of methane to CO_2 carbon 1:1 though the exact balance varies by literature source as high as 3:5[23,76].

Methane metabolism for biomass and PHB production are net producers of CO_2 and under bottled conditions CO_2 is generated and used, however it is hypothesised the high gas flow rates with very low partial pressures seen here stripped out the CO_2 stalling metabolism and thus growth. Where a critical mass of methanotrophs are present in the fermentation (e.g. the high starting OD in Ferm 9) enough CO_2 is produced that it is no longer effectively stripped out faster than it can be utilised. This could explain the failure of a number of fermentations. The forced high starting impeller speed in Ferm 5, 6 and 7 would enhance this effect by increasing gas mass transfer and may outbalance the higher starting ODs in Ferm 5 and 6.

CO_2 addition has been shown to improve methanotroph growth at 5% of headspace gas[36] and a few scaled up fermentation experiments have added CO_2 to the gas mix[77,78]. Park et al.[1] notably found greatly reduced lag times and

growth rate with the supplementation of CO_2 or as $NaHCO_3$. This was attributed to internal CO_2 levels becoming limiting otherwise due to the sparged gases striping CO_2, as suggested in the above hypothesis.

Low CO_2 as a causal factor in the failure of Ferm 1-8 could be tested by the addition of CO_2 to the gas mix or the addition of $NaHCO_3$ to the media. The gas striping effect can be tested by reduced gas flow rates. Testing for build-up of stalled pyruvate, phosphoenolpyruvate or formaldehyde that cannot be assimilated may also illustrate this.

6.5.2 Success in Ferm 9 and 10

As Ferm 9 grew successfully with only *M. parvus* prior to contamination this indicates the other final species constituents were not essential for its success. This casts some doubt over the definitive key nature of the additional species present in Ferm 10 to its success. However, if toxic metabolite build up was responsible for the failure of fermentation the contaminated Ferm 10 may have been successful due to removal of these by-products. Toxic byproduct build up may also explain why resuspension but not rebottling of fermenter culture (Figure 36) allowed the revival of growth.

6.5.2.1 Suitability of Ferm 9 and 10 culture for future PHB fermentation

Ferm 9 was contaminated after significant growth with only *M. parvus* and although no detriment was observed by the introduction of the *Paracoccus* species, a decline in PHB at the end and the relative abundance of *M. parvus* declined throughout, indicating this is not a stable mixed culture. Though by the point of contamination growth was almost complete so behaviour may be more favourable under non-nitrate limiting conditions. The presence of two genera with likely species being associated with human infections also pose a risk to health if scaled up industrially. *Paracoccus* only as a mixed culture partner may still prove promising.

Ferm 10 provided general stability of species constituents throughout and no negative effect on PHB content giving it potential as a defined mixed culture for future use. Innate antimicrobial activity as mentioned in *S. molluscorum* could be a positive feature in a mixed culture to stop contamination from other sources. The *Sphingomonas* polysacheride secretions might affect behaviour of any final product but also have industrial applications and have been found easily extracted. *Sphingomonas* polysaccharide and PHB co-production have previously been explored[67].

In industrial production the extended phase after PHB production would not occur and processing would occur at an economically viable trade off of time and PHB accumulation. Thus, the long period of fermentation after nitrate exhaustion tested here is not highly applicable in industrial applications. If a continuous or semi-continuous fermentation was established at lab pilot scale this might illustrate the capability and stability of these mixed culture inoculua more thoroughly.

None of the polytrophic species identified here have previously been identified in methanotroph mixed cultures to the best of our knowledge

including those in the Calysta process mentioned previously[32,40]. In the Calysta process this is of no great surprise as they are generally thermotolerant for a 45°C fermentation compared to the 30°C here. Broadly *Sphingomonadaceae*, the family containing *Sphingomonas* identified in Ferm 10 was a constituent by Zhang et al. [79] but with differing genera.

6.5.3 Nanopore 16S Method

6.5.3.1 16S Accuracy and Read Depth

16S sequencing was chosen for nanopore as it should provide a result unbiased by genome size unlike other methods like a WIMP (Whats In My Pot) shotgun sequencing approach. 16S sequencing does have its own biases discussed more fully elsewhere[80,81]. Particular factors include DNA copy number, deviation in the primer binding areas and the copy number of 16S genes[80]. For example the 27-CM primer in the nanopore kit differs from the binding site in *Sphingomonadales* (which includes the *Sphingomonas* species found in Ferm 9 and 10) by two consecutive base pairs[58] requiring two consecutive mispairings which may reduce primer binding and negatively influence the perceived relative abundance of these species. Comparative fold changes in relative abundance however should remain unbiased. 16S gene copy number could be accounted for with a correction from a databases and tools have been developed to carry this out[81]. Any biases can also be corrected empirically against other microbe quantification techniques like dilution plate counting. As whole genome sequences become more common larger sequenced regions like the 16S-ITS-23S or larger may become preferred as they provide greater variation between species and thus resolution[82].

What cutoff should be used for eliminating species or genera as spurious is a challenging question. Previous values of cutoff have often used partial 16S genes and read lengths as low as 250 or 100bp, this will have a much lower read assignment accuracy compared to 1500bp full length reads achieved with the 16S nanopore method[83,84]. The methodology here also has less focus on rare species than in environmental and gut microbiome testing. Johnson et al. [85] found that with full 16S sequencing >200 reads depth there was no additional benefit in avoiding error, however this was in 2019 and increasing sequencing accuracy achieved since reduces these errors and thus the acceptable depth. Statistical analysis could prove the actual critical value for required read count considerably lower with recent accuracy capabilities.

During experimental design a decision should be made as to the minimum relative abundance of interest and set nanopore run time and thus data collection volume to fit that requirement. For example, using the >200 read depth requirement for a 1% relative abundance cutoff 20,000 reads per sample are required, for 0.5% 40,000 and 0.2% 100,000 reads. Samples in this chapter had a depth of ~130,00-180,000 reads per sample. Thus, this experiment, which disregarded species of lower than 0.5% abundance would be expected to achieve similar outcomes with ~30% of the data acquired. The high depth of reads in this study did add additional information; for example, in our experiment *Roseomonas* and *Sphingomonas* in Ferm 9 could be

identified at 116hrs with 339 and 153 reads respectively, making 0.23% and 0.10% relative abundance respectively and as these went on to increase in abundance to ~1.75% each shows these were real datapoints. However below 0.40% there was a large amount of noise caused by misidentified *M. parvus* reads that conceal these reads. Work on the Calysta mixed fermentation set found two of three polytrophs generally had <1% relative abundance[13,32] indicating resolution below this level should still be targeted as species can be active. A non-barcoding 16S kit is available from ONT placing all reads on a single sample for very high read depth but this removes the advantage of splitting cost between multiple barcoded samples.

6.5.3.2 Accuracy and Alternative Methodologies

Although correct barcode and species assignment improved with better accuracy, FAST accuracy basecalling for this application appears sufficient with little effect on conclusions. Higher accuracy basecalling would be more essential in genome assembly applications. Nevertheless, if the computational capability is available increasing the accuracy of the final data would still be beneficial. Sources of error in the analysis have been quantified and indicated high accuracy of results with the likely greatest contributor being gDNA extraction variability. Where sequencing accuracy will improve with basecaller algorithm refinement and flow cell chemistry, gDNA extraction must be standardised to improve its error.

This nanopore 16S methodology, completable in 2 days, is favourable compared to dilution and selective plate counting that is generally used to monitor mixed culture fermentations including at industry level[32]. This method will indicate both the presence of known species and contaminants. It also allows identification of contaminants immediately where dilution and selection plate methods require a combination of microbial knowledge from visual identity and further analysis like 16S PCR+sequencing or Maldi-TOF to identify species adding additional cost and time. Illumina sequencing of 16S libraries in mixed culture has also been carried out but results in much shorter reads with lower lineage assignment accuracy[40,85].

Other methods used in the literature for monitoring mixed culture bacterial prevalence and identification include fatty acid content, immunofluorescent staining, light microscopy cell counting and using an automated cell counter with size distribution[26].

6.5.4 Conclusions

A range of reasons for fermentation failure in Ferm 1-8 have been discussed. In my view the most probable being CO_2 striping by sparged gas staling metabolism, followed by high levels of oxygen causing toxic byproduct build-up. There also remains a possibility of media instability caused by interaction with sparged gas, but this is considered least likely. Although fermentation has been successfully scaled up without reported issues in the past, solving the issues found here is important in informing failure states that must be avoided if applied industrially.

Retesting of the successful fermentation method (Fermentation 9) to confirm its reliability would add confidence to these conclusions. The application of analytical techniques to monitor by-products like methanol, acetone, acetate, 2,3-butanediol and isopropanol would illuminate the metabolic background to the failure states observed. A broader spectrum metabolomic study would also be of use especially in finding the unknown alkalising chemical responsible for the acid addition required in all fermentations to maintain pH parameters. Such byproducts will indicate inefficiencies that could be corrected by improved fermentation methods, strain engineering or selection of co-culturing partner microbes.

It is unclear if issues found here are applicable to *M. parvus* OBBP only, α-proteobacterial methanotrophs or methanotrophs as a whole including γ-proteobacterial. This depends on the true source of the issues found and if they effect metabolic pathways in other parts of the polyphyletic methanotroph group.

Two mixed cultures have been presented, one of which appears stable and produces equivalent PHB percentage to a methanotroph pure culture. This is to the best of our knowledge the first defined mixed culture for the production of PHB from methane described.

6.6 Code Availability

"Analysis_of_EPI2ME_16S_CSV_Output.ipynb" script was used in the analysis of nanopore data. This has been removed from Epi2Me Labs but was found in the digital change log of Epi2Me Labs tutorials GitHub[86]. This has been archived along with the CSV repair script at the following GitHub for reference and future use with more in depth instructions.
github.com/StarburstCLA/Epi2Me-16S-CSV.git

6.7 References

1. Park S, Shah NN, Taylor RT, Droege MW. Batch cultivation of Methylosinus trichosporium OB3b: II. Production of particulate methane monooxygenase. Biotechnol Bioeng. 1992;40(1):151–7.
2. Pieja AJ, Morse MC, Cal AJ. Methane to bioproducts: the future of the bioeconomy? Curr Opin Chem Biol. 2017;41(1):123–31.
3. Kalyuzhnaya MG, Gomez OA, Murrell JC. The Methane-Oxidizing Bacteria (Methanotrophs). In: Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes. Cham: Springer International Publishing; 2019. p. 1–34.
4. Wendlandt KD, Geyer W, Mirschel G, Al-Haj Hemidi F. Possibilities for controlling a PHB accumulation process using various analytical methods. J Biotechnol. 2005;117(1):119–29.
5. Wendlandt K -D, Jechorek M, Brühl E. The influence of pressure on the growth of methanotrophic bacteria. Acta Biotechnol. 1993;13(2):111–5.
6. Asenjo JA, Suk JS. Microbial Conversion of Methane into poly-β-hydroxybutyrate (PHB): Growth and intracellular product accumulation in a type II methanotroph. J Ferment Technol. 1986 Jan 1;64(4):271–8.
7. Strong P, Laycock B, Mahamud S, Jensen P, Lant P, Tyson G, et al. The Opportunity for High-Performance Biomaterials from Methane. Microorganisms. 2016 Feb 3;4(1):11.
8. Ritala A, Häkkinen ST, Toivari M, Wiebe MG. Single cell protein-state-of-the-art, industrial landscape and patents 2001-2016. Vol. 8, Frontiers in Microbiology. 2017. p. 1–18.
9. Petersen LAH, Villadsen J, Jørgensen SB, Gernaey K V. Mixing and mass transfer in a pilot scale U-loop bioreactor. Biotechnol Bioeng. 2017;114(2):344–54.
10. Kantarci N, Borak F, Ulgen KO. Bubble column reactors. Process Biochem. 2005;40(7):2263–83.
11. Lee J, Yasin M, Park S, Chang IS, Ha KS, Lee EY, et al. Gas-liquid mass transfer coefficient of methane in bubble column reactor. Korean J Chem Eng. 2015;32(6):1060–3.
12. Rahnama F, Vasheghani-farahani E, Yazdian F, Abbas S. PHB production by Methylocystis hirsuta from natural gas in a bubble column and a vertical loop bioreactor. Biochem Eng J. 2012;65:51–6.
13. Eriksen H, Strand K, Jørgensen L. US7579163B2 - Method of fermentation. United States: United States Patent and Trademark Office; US7579163B2, 2009.
14. Liu L-Y, Xie G-J, Xing D-F, Liu B-F, Ding J, Ren N-Q. Biological conversion

of methane to polyhydroxyalkanoates: Current advances, challenges, and perspectives. Environ Sci Ecotechnology. 2020;2:100029.
15. Kim K, Kim Y, Yang J, Ha KS, Usta H, Lee J, et al. Enhanced mass transfer rate and solubility of methane via addition of alcohols for Methylosinus trichosporium OB3b fermentation. J Ind Eng Chem. 2017;46:350–5.
16. Zúñiga C, Morales M, Le Borgne S, Revah S. Production of poly-β-hydroxybutyrate (PHB) by Methylobacterium organophilum isolated from a methanotrophic consortium in a two-phase partition bioreactor. J Hazard Mater. 2011;190(1–3):876–82.
17. Han B, Su T, Wu H, Gou Z, Xing XH, Jiang H, et al. Paraffin oil as a "methane vector" for rapid and high cell density cultivation of methylosinus trichosporium OB3b. Appl Microbiol Biotechnol. 2009;83(4):669–77.
18. López JC, Arnáiz E, Merchán L, Lebrero R, Muñoz R. Biogas-based polyhydroxyalkanoates production by Methylocystis hirsuta: A step further in anaerobic digestion biorefineries. Chem Eng J. 2018;(333):529–36.
19. Gilman A, Laurens LM, Puri AW, Chu F, Pienkos PT, Lidstrom ME. Bioreactor performance parameters for an industrially-promising methanotroph Methylomicrobium buryatense 5GB1. Microb Cell Fact. 2015;14(1):1–8.
20. Wendlandt K-D, Jechorek M, Helm J, Stottmeister U. Producing poly-3-hydroxybutyrate with a high molecular mass from methane. J Biotechnol. 2001 Mar 30;86(2):127–33.
21. Kalyuzhnaya MG, Yang S, Rozova ON, Smalley NE, Clubb J, Lamb A, et al. Highly efficient methane biocatalysis revealed in a methanotrophic bacterium. Nat Commun. 2013;4(May):1–7.
22. Vecherskaya M, Dijkema C, Saad HR, Stams AJM. Microaerobic and anaerobic metabolism of a Methylocystis parvus strain isolated from a denitrifying bioreactor. Environ Microbiol Rep. 2009;1(5):442–9.
23. Kalyuzhnaya MG. Methane Biocatalysis: Selecting the Right Microbe. Biotechnology for Biofuel Production and Optimization. Elsevier B.V.; 2016. 353–383 p.
24. Vecherskaya M, Dijkema C, Stams AJM. Intracellular PHB conversion in a Type II methanotroph studied by 13C NMR. J Ind Microbiol Biotechnol. 2001;26(1–2):15–21.
25. Bordel S, Rojas A, Muñoz R. Reconstruction of a Genome Scale Metabolic Model of the polyhydroxybutyrate producing methanotroph Methylocystis parvus OBBP. Microb Cell Fact. 2019;18(104):1–11.
26. Helm J, Wendlandt KD, Rogge G, Kappelmeyer U. Characterizing a stable methane-utilizing mixed culture used in the synthesis of a high-quality biopolymer in an open system. J Appl Microbiol. 2006;101(2):387–95.
27. Costa C, Vecherskaya M, Dijkema C, Stams AJM. The effect of oxygen on methanol oxidation by an obligate methanotrophic bacterium studied by in vivo 13C nuclear magnetic resonance spectroscopy. J Ind Microbiol Biotechnol. 2001;26(1–2):9–14.
28. Xin J, Cui J, Niu J, Hua S, Xia C, Li S, et al. Production of methanol from methane by methanotrophic bacteria. Biocatal Biotransformation. 2004 May 11;22(3):225–9.
29. Whittenbury R, Phillips KC, Wilkinson JF. Enrichment, Isolation and Some Properties of Methane-utilizing Bacteria. J Gen Microbiol. 1970;(61):205–18.
30. Pieja AJ, Rostkowski KH, Criddle CS. Distribution and Selection of Poly-3-Hydroxybutyrate Production Capacity in Methanotrophic Proteobacteria.

Microb Ecol. 2011;62(3):564–73.
31. Karthikeyan OP, Chidambarampadmavathy K, Cirés S, Heimann K. Review of Sustainable Methane Mitigation and Biopolymer Production. Crit Rev Environ Sci Technol. 2015 Aug 3;45(15):1579–610.
32. Bothe H, Møller Jensen K, Mergel A, Larsen J, Jørgensen C, Bothe H, et al. Heterotrophic bacteria growing in association with Methylococcus capsulatus (Bath) in a single cell protein production process. Appl Microbiol Biotechnol. 2002;59(1):33–9.
33. Listewnik H-F, Wendlandt K-D, Jechorek M, Mirschel G. Process Design for the Microbial Synthesis of Poly-β-hydroxybutyrate (PHB) from Natural Gas. Eng Life Sci. 2007 Jun;7(3):278–82.
34. Rodríguez Y, Firmino PIM, Pérez V, Lebrero R, Muñoz R. Biogas valorization via continuous polyhydroxybutyrate production by Methylocystis hirsuta in a bubble column bioreactor. Waste Manag. 2020;113(2020):395–403.
35. Singh R, Ryu J, Kim SW. Microbial consortia including methanotrophs: some benefits of living together. J Microbiol. 2019;57(11):939–52.
36. Veraart AJ, Garbeva P, Van Beersum F, Ho A, Hordijk CA, Meima-Franke M, et al. Living apart together - Bacterial volatiles influence methanotrophic growth and activity. ISME J. 2018;12(4):1163–6.
37. Semrau JD, DiSpirito AA, Yoon S. Methanotrophs and copper. FEMS Microbiol Rev. 2010 Jul;34(4):496–531.
38. Iguchi H, Yurimoto H, Sakai Y. Stimulation of methanotrophic growth in cocultures by cobalamin excreted by rhizobia. Appl Environ Microbiol. 2011;77(24):8509–15.
39. Kerckhof FM, Sakarika M, Van Giel M, Muys M, Vermeir P, De Vrieze J, et al. From Biogas and Hydrogen to Microbial Protein Through Co-Cultivation of Methane and Hydrogen Oxidizing Bacteria. Front Bioeng Biotechnol. 2021;9(August):1–17.
40. Oshkin IY, Belova SE, Khokhlachev NS, Semenova VA, Chervyakova OP, Chernushkin D V., et al. Molecular Analysis of the Microbial Community Developing in Continuous Culture of Methylococcus sp. Concept-8 on Natural Gas. Microbiol (Russian Fed. 2020;89(5):551–9.
41. Fergala A, AlSayed A, Eldyasti A. Factors affecting the selection of PHB accumulating methanotrophs from waste activated sludge while utilizing ammonium as their nitrogen source. J Chem Technol Biotechnol. 2018;93(5):1359–69.
42. Cattaneo CR, Rodríguez Y, Rene ER, García-Depraect O, Muñoz R. Biogas bioconversion into poly(3-hydroxybutyrate) by a mixed microbial culture in a novel Taylor flow bioreactor. Waste Manag. 2022;150(April):364–72.
43. Jiang H, Duan C, Jiang P, Liu M, Luo M, Xing XH. Characteristics of scale-up fermentation of mixed methane-oxidizing bacteria. Biochem Eng J. 2016;109:112–7.
44. Wendlandt KD, Jechorek M, Helm J, Stottmeister U. Production of PHB with a high molecular mass from methane. Polym Degrad Stab. 1998;59(1–3):191–4.
45. Rostkowski KH, Criddle CS, Lepech MD. Cradle-to-gate life cycle assessment for a cradle-to-cradle cycle: Biogas-to-bioplastic (and back). Environ Sci Technol. 2012;46(18):9822–9.
46. Pieja AJ, Sundstrom ER, Criddle CS. Poly-3-hydroxybutyrate metabolism in the type II Methanotroph Methylocystis parvus OBBP. Appl Environ Microbiol. 2011;77(17):6012–9.

47. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
48. R Core Team. R: A Language and environment for statistical computing. 2021.
49. Grolemund G, Wickham H. Dates and Times Made Easy with lubridate. Vol. 40, Journal of Statistical Software. 2011. p. 1–25.
50. Han D, Link H, Liesack W. Response of Methylocystis sp. strain SC2 to salt stress: Physiology, global transcriptome, and amino acid profiles. Appl Environ Microbiol. 2017;83(20):1–14.
51. Bowman J. The Methanotrophs — The Families Methylococcaceae and Methylocystaceae. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes. 3rd ed. New York: Springer New York; 2006. p. 266–89.
52. Myung J, Flanagan JCA, Waymouth RM, Criddle CS. Methane or methanol-oxidation dependent synthesis of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) by obligate type II methanotrophs. Process Biochem. 2016;51(5):561–7.
53. Helm J, Wendlandt KD, Jechorek M, Stottmeister U. Potassium deficiency results in accumulation of ultra-high molecular weight poly-β-hydroxybutyrate in a methane-utilizing mixed culture. J Appl Microbiol. 2008;105(4):1054–61.
54. Haynes WM. CRC Handbook of Chemistry and Physics. 95th ed. Lide DR, Bruno TJ, Haynes WM, editors. Florida: CRC Press; 2014. 3–488 p.
55. National Oceanic and Atmospheric Administration, National Aeronautics and Space Administration, United States Air Force. U.S. Standard Atmosphere. Washington, D. C.; 1976.
56. Sigma Aldrich. Buffer Reference Center [Internet]. [cited 2021 Jan 27]. Available from: https://www.sigmaaldrich.com/GB/en/technical-documents/protocol/protein-biology/protein-concentration-and-buffer-exchange/buffer-reference-center#phosphate
57. Oxford Nanopore. Nanopore Methodology - CHEMISTRY TECHNICAL DOCUMENT V1 07jul2016 [Internet]. 2016 [cited 2023 Mar 13]. Available from: https://community.nanoporetech.com/technical_documents/chemistry-technical-document/v/chtd_500_v1_revai_07jul2016/barcoding-kits
58. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. Appl Environ Microbiol. 2008;74(8):2461–70.
59. Oxford Nanopore. 16S v2022.01.07 Workflow Document [Internet]. 2022. Available from: https://epi2me.nanoporetech.com/workflow-docs-7?version=2022.01.07
60. Oxford Nanopore. Accuracy [Internet]. 2023 [cited 2023 Mar 13]. Available from: https://nanoporetech.com/accuracy
61. Kageyama A, Matsumoto A, Ömura S, Takahashi Y. Humibacillus xanthopallidus gen. nov., sp. nov. Int J Syst Evol Microbiol. 2008;58(7):1547–51.
62. Puri A, Bajaj A, Singh Y, Lal R. Harnessing taxonomically diverse and metabolically versatile genus Paracoccus for bioplastic synthesis and xenobiotic biodegradation. J Appl Microbiol. 2022;132(6):4208–24.
63. Bordel S, van Spanning RJM, Santos-Beneit F. Imaging and modelling of poly(3-hydroxybutyrate) synthesis in Paracoccus denitrificans. AMB Express. 2021;11(1).
64. Han XY, Pham AS, Tarrand JJ, Rolston K V., Helsel LO, Levett PN.

Bacteriologic characterization of 36 strains of Roseomonas species and proposal of Roseomonas mucosa sp nov and Roseomonas gilardii subsp rosea subsp nov. Am J Clin Pathol. 2003;120(2):256–64.
65. Reddy GSN, Garcia-Pichel F. Sphingomonas mucosissima sp. nov. and Sphingomonas desiccabilis sp. nov., from biological soil crust in the Colorado Plateau, USA. Int J Syst Evol Microbiol. 2007;57(5):1028–34.
66. Romanenko LA, Uchino M, Frolova GM, Tanaka N, Kalinovskaya NI, Latyshev N, et al. Sphingomonas molluscorum sp. nov., a novel marine isolate with antimicrobial activity. Int J Syst Evol Microbiol. 2007 Feb 1;57(2):358–63.
67. Wu M, Li G, Huang H, Chen S, Luo Y, Zhang W, et al. The simultaneous production of sphingan Ss and poly(R-3-hydroxybutyrate) in Sphingomonas sanxanigenens NX02. Int J Biol Macromol. 2016 Jan 1;82:361–8.
68. Weon HY, Yoo SH, Kim SJ, Kim YS, Anandham R, Kwon SW. Massilia jejuensis sp. nov. and Naxibacter suwonensis sp. nov., isolated from air samples. Int J Syst Evol Microbiol. 2010 Aug 1;60(8):1938–43.
69. Zul D, Wanner G, Overmann J. Massilia brevitalea sp. nov., a novel betaproteobacterium isolated from lysimeter soil. Int J Syst Evol Microbiol. 2008 May 1;58(5):1245–51.
70. Han X, Satoh Y, Kuriki Y, Seino T, Fujita S, Suda T, et al. Polyhydroxyalkanoate production by a novel bacterium Massilia sp. UMI-21 isolated from seaweed, and molecular cloning of its polyhydroxyalkanoate synthase gene. J Biosci Bioeng. 2014;118(5):514–9.
71. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, et al. Database resources of the national center for biotechnology. Vol. 31, Nucleic Acids Research. Oxford University Press; 2003. p. 28–33.
72. Jeong J, Kim TH, Jang N, Ko M, Kim SK, Emelianov G, et al. A highly efficient and versatile genetic engineering toolkit for a methanotroph-based biorefinery. Chem Eng J. 2023;453(P2):139911.
73. Sander R. Compilation of Henry's law constants (version 4.0) for water as solvent. Atmos Chem Phys. 2015 Apr 30;15(8):4399–981.
74. Henry W. III. Experiments on the quantity of gases absorbed by water, at different temperatures, and under different pressures. Philos Trans R Soc London. 1803 Dec 31;93(3):29–274.
75. Trotsenko YA, Murrell JC. Metabolic Aspects of Aerobic Obligate Methanotrophy. In: Advances in Applied Microbiology. 2008. p. 183–229.
76. Yang S, Matsen JB, Konopka M, Green-Saxena A, Clubb J, Sadilek M, et al. Global molecular analyses of methane metabolism in methanotrophic alphaproteobacterium, methylosinus trichosporium OB3b. Part II. metabolomics and 13C-labeling study. Front Microbiol. 2013;4(APR):1–13.
77. Jiang H, Chen Y, Jiang P, Zhang C, Smith TJ, Murrell JC, et al. Methanotrophs: Multifunctional bacteria with promising applications in environmental bioengineering. Biochem Eng J. 2010;49(3):277–88.
78. Choi SY, Rhie MN, Kim HT, Chan Joo J, Cho IJ, Son J, et al. Metabolic engineering for the synthesis of polyesters: A 100-year journey from polyhydroxyalkanoates to non-natural microbial polyesters. Metab Eng. 2019;(May):1–35.
79. Zhang T, Wang X, Zhou J, Zhang Y. Enrichments of methanotrophic–heterotrophic cultures with high poly-β-hydroxybutyrate (PHB) accumulation capacities. J Environ Sci (China). 2018;65:133–43.

80. Brooks JP, Edwards DJ, Harwich MD, Rivera MC, Fettweis JM, Serrano MG, et al. The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies Ecological and evolutionary microbiology. BMC Microbiol. 2015;15(1):1–14.
81. Kembel SW, Wu M, Eisen JA, Green JL. Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. PLoS Comput Biol. 2012;8(10):16–8.
82. Milani C, Alessandri G, Mangifesta M, Mancabelli L, Lugli GA, Fontana F, et al. Untangling Species-Level Composition of Complex Bacterial Communities through a Novel Metagenomic Approach. mSystems. 2020;5(4).
83. Schloss PD. The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. PLoS Comput Biol. 2010;6(7):19.
84. Ladin ZS, Ferrell B, Dums JT, Moore RM, Levia DF, Shriver WG, et al. Assessing the efficacy of eDNA metabarcoding for measuring microbial biodiversity within forest ecosystems. Sci Rep. 2021;11(1):1–14.
85. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun. 2019;10(1):1–11.
86. Oxford Nanopore. Analysis_of_EPI2ME_16S_CSV_Output.ipynb for Epi2Me Labs.

Chapter 7: General Remarks and Conclusions

Due to the polyphyletic nature of the methanotroph group authors should take care when discussing similarities and inferring between the groups. While methanotrophy genes themselves are more closely related due to horizontal gene transfer, they sit within chassis that have evolved separately. Some differences between α- and γ-proteobacterial methanotrophs are documented but there is space in the literature for a comparison in behaviour, possession of genes and features that might show more in common with their own α-proteobacterial or γ-proteobacterial relatives than each other and so provided sources of information in the literature from overlooked close relatives. The ease of falling into a silo of methanotroph publications and not venturing into broader relatives was a hazard in the progression of this work limiting access to useful data. For example, Type I γ-proteobacterial methanotrophs like *M. capsulatus* Bath are more related to *E. coli*, also a γ-proteobacteria, than they are to Type II α-proteobacterial methanotrophs like *M. parvus*. How the 4-5 methanotroph groups adapted to methanotrophy separately, and any convergent evolution between them, may shed light on interesting new parts of methanotroph metabolism. This is especially true when incorporating the more diverse and less well understood mycobacterial, NC 10 and Verrucomicrobia.

With the creation of a complete genome for the *Methylocystis* genus type species *M. parvus* OBBP I have attempted to elucidate these similarities and differences in the phylogenetic gene trees for MMOs and PHB pathway genes. Knowing what model organisms are most similar and thus from which inferences can most safely be drawn, will accelerate work in the species and methanotrophs as a whole. This is especially true where gene notation is not preserved between publications or, as with phaC1 and 2, causing false appearance of similarity between groups.

Along with my collaborators a fully functional CRISPR-Cas9 editing system has been developed and demonstrated. This is the first fully functional system in methanotrophs and proved flexible and reliable over a range of seed designs more limited by gene essentiality. Discussions and investigation of target pathways and genes of interest for engineering lay the ground for future work in the area, utilising this CRISPR system and other previous engineering systems.

Difficulties in scale up from bottle to 750 ml bioreactors proved challenging but solving those issues will be essential in the reliability of any future industrial applications. Certain essential information about fermentation byproducts particularly those responsible for unexpected alkalisation over time are missing and would be straightforward to assess with the application of metabolomic analytical techniques. Successes with two new contamination sourced mixed cultures and their PHB production provide two further avenues of investigation and potential industrial application. The

nanopore based relative abundance method developed here helped understand these mixed cultures and I commend its use to other researchers and industrial users in the mixed culture field. Often this type of data is surprisingly absent and this could provide an improvement in time and accurate classification of species compared to plate count methods currently in use.

The use of biogas, although a key selling point of methanotroph produced SCP and PHB green credentials, remains underutilised, with industrial applications tending towards natural gas. The work demonstrating comparable growth and production of PHB on biogas and pure methane adds another piece of evidence that biogas should still be the final goal of the technology and is usable once it becomes price competitive to do so.

In addition to the directions of further work listed above, a full comparison of nanopore based and Illumina based metagenomic abundance monitoring of mixed cultures compared to selective media spread plate methods would have the greatest immediate industrial impact. Especially if carried out with an industrial partner in parallel to their standard workflows. With the advent of our effective CRISPR-Cas9 editing system more work in molecular biology and genetic knockout to illuminate the origins, functionality, and essentiality of the two megaplasmids would improve the understanding of *M. parvus* and its relatives as a platform organism. Similarly, a better understanding of the purpose of its four secretion systems and a full range of flagella motility genes would be of great interest.

The regulatory system of PHB production in *M. parvus* and related methanotrophs remains unknown and different to that of other PHB model organisms. If this could be understood and the induction machinery detached from nutrient starvation the ease of industrial use would be greatly increased. Additionally, the major challenge of PHB based bioplastics is the cost of extraction, and engineered improvements to this by weakening cell walls or instigating an autolysis system could prove immediately helpful.

Methanotrophs remain under investigated, perhaps due to their slow growth, and safety challenges in working with flammable gases. Many questions still present especially those that lead to variation in culture growth and PHB yields. Despite this methanotrophs remain a fascinating group of organisms in terms of their genetic origins and metabolic capability particularly considering its place in environmental microbiology as an atmospheric methane sink. I commend their research and the promise of their application in the green technology of the future.

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

A Complete Genome of the Alphaproteobacterial Methanotroph *Methylocystis parvus* OBBP

Benedict H. Claxton Stevens,[a] Bashir L. Rumah,[a] Christopher E. Stead,[a] Ying Zhang[a]

[a]BBSRC/EPSRC Synthetic Biology Research Centre, School of Life Sciences, Biodiscovery Institute, University of Nottingham, Nottingham, United Kingdom

ABSTRACT A complete genome is presented for *Methylocystis parvus* OBBP, a Gram-negative aerobic methanotroph of the phylum *Alphaproteobacteria*. *M. parvus* OBBP is the genus type strain and of interest in the production of polyhydroxybutyrate and environmental microbiology. The genome consists of two plasmids (248 kbp and 205 kbp) and a chromosome (4.076 Mbp).

Methylocystis parvus OBBP is an obligate methanotroph of interest in the production of the biodegradable bioplastic polyhydroxybutyrate (PHB) and is the genus type strain of *Methylocystis* (1–3). An unclosed *M. parvus* OBBP genome (MetPar_1.0-AJTV00000000.1) of 108 contigs has been previously released (1), which has been widely used, including a genome-scale model in 2019 (4). As an improvement upon this, sequencing and error checking of the new genome were exhaustive to ensure a complete and reliable reference was achieved.

Strain OBBP, isolated in 1970 from soil and water samples (2), was acquired from NCIMB 11129 (National Collection of Industrial, Food and Marine Bacteria). OBBP was grown in pure culture in nitrate mineral salt (NMS) medium (2) at 30°C, and genomic DNA (gDNA) was extracted using a phenol-chloroform-isoamyl alcohol method (5). A library of sheared large insert gDNA was prepared using g-TUBES (Covaris, Woburn, MA, USA) at 4,000 rpm once per end for 60 s and sequenced without size selection using HiFi PacBio on an RS II SMRTcell by Genome Quebec, Canada. The PacBio SMRTPipe pipeline collapsed and error corrected the reads, giving 145,723 ~10-kbp circular consensus sequencing reads totaling 1,333,680,793 bp, an N_{50} of 9,860 bp, and 293× depth. Short-read gDNA libraries were prepared using the Nextera XT library prep kit (Illumina, San Diego, CA, USA) and processed on an Illumina MiSeq (Deep Seq, Nottingham, UK) using a 250-bp paired-end protocol, resulting in 987,357 paired reads totaling 480,751,157 bp and 104× depth. Quality of reads was checked using FastQC (v0.11.9) (6).

A hybrid assembly strategy was followed. PacBio reads were assembled with SMRTLink (v8.0.0.80501), and PacBio and Illumina reads were assembled together in Unicycler (v0.4.9) (7), providing two assemblies. Realignment was carried out with Bowtie2 (v2.4.4) (8) for short reads and minimap2 (v2.24) (9) for long reads with iterated polishing with Pilon (v1.24) (10). Assemblies were compared with Mauve (v20150226) (11) and ALE (v0.9) (12) and visualized with IGV (v2.11.9) (13). Both assemblies informed the final genome. Discrepancy decisions were confirmed by Sanger sequencing of PCR products by Genewiz, USA.

The final genome (Table 1) consists of a chromosome and two plasmids (pMpar-1 and pMpar-2) with no gaps or undecided bases. The genome achieved a BUSCO (v5.3.0) (14) completeness of 99.6% against the v10 *Alphaproteobacteria* data set and a CheckM (v1.1.3) (15) score of 100%.

Annotation was carried out using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP, v6.0) (16) and summarized in Table 1. Of these annotations, 93 were pseudogenes. Annotations included two full sets of rRNAs, 48 tRNAs, and two CRISPR arrays. The new genome adds 57.3 kbp to the MetPar_1.0 assembly including six additional *pmo* genes and

Editor Vanja Klepac-Ceraj, Wellesley College

Address correspondence to Ying Zhang, ying.zhang@nottingham.ac.uk.

The authors declare no conflict of interest.

Received 2 February 2023
Accepted 9 March 2023
Published 23 March 2023

TABLE 1 Assembly and annotation summary

Parameter	Total	Chromosome	pMpar-1	pMpar-2
GenBank accession no.		CP092968	CP092969	CP092970
Size, bp	4,529,117	4,076,007	248,224	204,886
GC, %	63.39	63.7	60.8	60.3
Genes, no.	4,379	3,923	226	230
Hypothetical genes, no. (%)	823 (18.8)	702 (17.9)	59 (26.1)	62 (27.0)

removes nine unplaced contaminant contigs: six *Priestia megaterium*, one *Stenotrophomonas rhizophila*, one *Mesorhizobium* sp., and one *Pseudomonas putida*.

Data availability. This genome has been deposited in GenBank as CP092968 (chromosome), CP092969 (pMpar-1), and CP092970 (pMpar-2). The version described in this paper is the first version. Sequencing data are under the NCBI BioProject accession no. PRJNA812408 containing Illumina, PacBio, and Sanger sequencing data with additional methylation data.

ACKNOWLEDGMENTS

This work was supported by the UK Biotechnology and Biological Sciences Research Council (grant numbers BB/L013940/1, BB/L013800/1, BB/M008770/1, and BB/S009833/1), Petroleum Technology Development Fund Nigeria, the Nottingham BBSRC DTP (grant number BB/J014508/1), and "Proof of Concept" funding from the BBSRC NIBBs C1Net and the Carbon Recycling Network.

REFERENCES

1. del Cerro C, García JLJM, Rojas A, Tortajada M, Ramón D, Galán B, Prieto MA, García JLJM. 2012. Genome sequence of the methanotrophic poly-β-hydroxybutyrate producer *Methylocystis parvus* OBBP. J Bacteriol 194:5709–5710. https://doi.org/10.1128/JB.01346-12.
2. Whittenbury R, Phillips KC, Wilkinson JF. 1970. Enrichment, isolation and some properties of methane-utilizing bacteria. J Gen Microbiol 61:205–218. https://doi.org/10.1099/00221287-61-2-205.
3. Bowman J. 2006. The methanotrophs—the families Methylococcaceae and Methylocystaceae, p 266–289. *In* Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (ed), The prokaryotes, 3rd ed. Springer, New York, NY.
4. Bordel S, Rojas A, Muñoz R. 2019. Reconstruction of a genome scale metabolic model of the polyhydroxybutyrate producing methanotroph Methylocystis parvus OBBP. Microb Cell Fact 18:104. https://doi.org/10.1186/s12934-019-1154-5.
5. Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
6. Andrews S. 2019. FastQC. 0.11.9.
7. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
8. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
9. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
10. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
11. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704.
12. Clark SC, Egan R, Frazier PI, Wang Z. 2013. ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. Bioinformatics 29:435–443. https://doi.org/10.1093/bioinformatics/bts723.
13. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nat Biotechnol 29:24–26. https://doi.org/10.1038/nbt.1754.
14. Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021. BUSCO: assessing genomic data quality and beyond. Curr Protoc 1:e323. https://doi.org/10.1002/cpz1.323.
15. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
16. Tatusova T, Dicuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.

Rumah *et al. AMB Expr* (2021) 11:6
https://doi.org/10.1186/s13568-020-01159-4

AMB Express

ORIGINAL ARTICLE

Open Access

Isolation and characterisation of *Methylocystis* spp. for poly-3-hydroxybutyrate production using waste methane feedstocks

Bashir L. Rumah[†], Christopher E. Stead[†], Benedict H. Claxton Stevens, Nigel P. Minton, Alexander Grosse-Honebrink and Ying Zhang[*]

Abstract

Waste plastic and methane emissions are two anthropogenic by-products exacerbating environmental pollution. Methane-oxidizing bacteria (methanotrophs) hold the key to solving these problems simultaneously by utilising otherwise wasted methane gas as carbon source and accumulating the carbon as poly-3-hydroxybutyrate, a biodegradable plastic polymer. Here we present the isolation and characterisation of two novel *Methylocystis* strains with the ability to produce up to 55.7 ± 1.9% poly-3-hydroxybutyrate of cell dry weight when grown on methane from different waste sources such as landfill and anaerobic digester gas. *Methylocystis rosea* BRCS1 isolated from a recreational lake and *Methylocystis parvus* BRCS2 isolated from a bog were whole genome sequenced using PacBio and Illumina genome sequencing technologies. In addition to potassium nitrate, these strains were also shown to grow on ammonium chloride, glutamine and ornithine as nitrogen source. Growth of *Methylocystis parvus* BRCS2 on Nitrate Mineral Salt (NMS) media with 0.1% methanol vapor as carbon source was demonstrated. The genetic tractability by conjugation was also determined with conjugation efficiencies up to 2.8×10^{-2} and 1.8×10^{-2} for *Methylocystis rosea* BRCS1 and *Methylocystis parvus* BRCS2 respectively using a plasmid with ColE1 origin of replication. Finally, we show that *Methylocystis* species can produce considerable amounts of poly-3-hydroxybutyrate on waste methane sources without impaired growth, a proof of concept which opens doors to their use in integrated bio-facilities like landfills and anaerobic digesters.

Keywords: Methanotrophy, *Methylocystis* species, Poly-3-hydroxybutyrate, Bioplastic, Biogas

Keypoints

- *Methylocystis rosea* BRCS1 was isolated from a lake while *Methylocystis parvus* BRCS2 was isolated from a bog, both in England.
- Both species showed normal growth and PHB accumulation on landfill and anaerobic digester gas which contain trace contaminants speculated to be inhibitory to growth.
- *Methylocystis parvus* BRCS2 showed the highest PHB accumulation 55.7 ± 1.9% PHB of cell dry weight when grown using landfill gas as methane source.

Introduction

Methane (CH_4) is the second most abundant greenhouse gas (GHG) produced by human activity with a global warming potential up to 105 times higher than CO_2 over a 20-year period (Rodhe 1990; Shindell et al. 2009). Methane is emitted from a variety of anthropogenic and non-anthropogenic sources including wetlands, natural

*Correspondence: Ying.Zhang@nottingham.ac.uk
[†]Bashir L. Rumah and Christopher E. Stead contributed equally to this work
BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK

gas exploration sites and landfill sites (Boeckx et al. 1996; Allen et al. 2013; Zhang et al. 2017). High quality biogas from Anaerobic Digester (AD) and landfill sites is currently economically used for energy production (Allen et al. 2013). However, biogas with low methane content is often flared with the aforementioned environmental impact (EPA 2011). To improve incentive for biogas capture, new technologies for utilisation of the gas need to be explored.

As the only known biological sink for atmospheric methane, methane-oxidizing bacteria are largely responsible for balancing methane flux in the environment through oxidation of methane for a source of carbon and energy (Anthony 1982). The use of methanotrophs to produce platform chemicals, single cell protein or biopolymers has high economic potential (Strong et al. 2016). Biopolymer production in particular has received renewed societal and industrial interest with reports of petrochemical, non-biodegradable plastics polluting the environment and earth's oceans (Derraik 2002; Eriksen et al. 2014). Typically, plastic compounds cannot be degraded by microorganisms. Rather they disintegrate into ever smaller fragments called microplastics. Microplastics have been found throughout the marine ecosystem and pose possible adverse effects on ecological and human health (Cole et al. 2013; Wright and Kelly 2017). Poly-3-hydroxybutyrate (PHB) is a short-chained polyhydroxyalkanoate (PHA) with mechanical properties comparable to isotactic polypropylene (PP) and polyethylene (PE), with the advantage that it is biodegradable (Tokiwa et al. 2009; Yeo et al. 2018). PHB is produced by type II methanotrophs during nutrient limitation and it serves as a source of reducing equivalents (Asenjo and Suk 1986; Wendlandt et al. 2001; Listewnik et al. 2007; Pieja et al. 2011). Therefore, the utilisation of PHB producing methanotrophic organisms grown on comparably cheap or waste sources of methane such as AD or landfills could represent a consolidated solution to two major environmental problems from anthropogenic activity.

Factors such as ability to utilise methane feedstock and PHB accumulation capability of the chosen methanotrophic chassis need to be taken into consideration when selecting a bacterial strain. Most studies reported to date on methanotrophic PHB production have mainly focused on the use of pure methane, natural gas or artificial biogas as substrate, leaving the renewable sources of CH_4 (biogases) open to investigation (Pieja et al. 2011; Listewnik et al. 2007; López et al. 2018). Biogas from anaerobic digesters and landfills consist primarily of a mixture of methane, carbon dioxide (CO_2) and nitrogen, with traces of toxic compounds such as hydrogen sulphide (H_2S), siloxanes and aromatic and halogenated compounds (Rasi et al. 2007). Also, biogas composition is highly dependent on waste composition, temperature and moisture among other factors, and can thus vary between different AD facilities and landfill sites (Rasi et al. 2007). Here we investigate the effect of biogas from three landfill sites and four different AD sources on growth and PHB production of two newly isolated strains of *Methylocystis* species, and compare their performance against the type strain *Methylocystis parvus* OBBP.

Materials and methods

Bacterial strains and culture conditions

All strains used in this study are listed in Additional file 1: Table S1. Methanotrophic strains were cultured in liquid Nitrate Mineral Salt (NMS), in dNMS medium (5 times diluted NMS medium with H_2O) or on solid plates of NMS or dNMS supplemented with 1.5% Agar Bacteriological (Thermo Scientific, UK) (Whittenbury et al. 1970). Unless otherwise stated, liquid cultures were grown at 30 °C in serum bottles capped with rubber stoppers with a 5:1 headspace to culture ratio and headspace was adjusted to a 2:1 molar oxygen to methane ratio with 0.5 bar overpressure. Cultures on solid medium were grown at 30 °C in anaerobic Oxoid jars (Thermo Scientific, UK) by addition of methane to the headspace. Methanotrophs were stored at − 80 °C on microbeads (Microbank™ Bacterial and Fungal Preservation System, Pro-Lab Diagnostics, UK) according to the supplier's instructions and revived on solid medium before inoculation into liquid medium.

Isolation of methanotrophs

Environmental samples leading to isolation of *Methylocystis rosea* BRCS1 were collected from a recreational lake at the University of Nottingham campus (52° 56′ 13.9″ N 1° 11′ 29.4″ W) on 16th of March 2015. Enrichment started within 24 h of sampling. The sample was vortexed and centrifuged at 1000 rpm for 1 min. 10 µL of the supernatant was added to 60 mL serum bottles containing 10 mL dNMS media. Serum bottles were incubated at 30 °C and 150 rpm for 25 days with a ratio of air:CH_4:CO_2 of 76:20:4. All samples were processed in duplicates. Samples exhibiting visible growth were subcultured by adding 10 µL of the enrichment culture to fresh 60 mL serum bottles containing 10 mL of dNMS media and incubated as above. After five days of growth, the samples were serially diluted up to 10^{-5} and 100 µL of each dilution was spread on dNMS agar plates. Plates were incubated in Oxoid jars as described above.

On day 5 and 21, growth on plates was analysed. Colonies growing on agar were resuspended in 15 µL Nuclease Free Water (NFW) and re-spread on dNMS agar plates. Once colonies formed, they were analysed for Methane Monooxygenase (MMO) gene presence by PCR

using specific primers (*pmoA* and *mmoX*) and PCR products were Sanger sequenced (Eurofins Scientific, UK) (Bourne et al. 2001). Colonies testing positive for MMO genes (*pmoA* and/or *mmoX*) were purified through multiple rounds of growth in liquid culture starting from serial dilutions followed by growing to single colony on dNMS agar. After several rounds of such purification, vitamins in the medium were omitted to inhibit growth of non-methanotrophic bacteria. A pure culture of BRCS1 was obtained after further rounds of purification.

Samples leading to isolation of *Methylocystis parvus* BRCS2 were obtained from a bog in Moseley UK (52° 26′ 10.5" N 1° 51′ 55.0" E) and stored at room temperature overnight. 3 g of solid bog samples (gravel sediment and bog sediment) were homogenised in 27 mL dNMS media supplemented with 10 µM $CuSO_4{\cdot}7H_2O$ using a vortex. Sediments in the samples were settled and supernatant was serially diluted with supplemented dNMS medium to 10^{-7}. 11 mL of each dilution was transferred to a 60 mL serum bottle. Serum bottle headspace was adjusted to 20:80 CH_4:air ratio and samples were incubated at 30 °C for five weeks shaking at 150 rpm.

Samples were visually analysed for growth after five weeks and highest dilutions per sample showing growth were plated on dNMS agar plates. Single colonies were further purified after two rounds of liquid culturing and plating, as explained above. At this stage, colonies were analysed by PCR and Sanger sequencing as described above and MMO positive colonies were further purified by extinction-dilution as follows: colonies were resuspended in dNMS media omitting vitamins, diluted to 10^{-7} in 96-well plates and incubated at 30 °C and 200 rpm in the gas-tight box CR1601 (EnzyScreen, NL) for two weeks. From the highest dilutions showing visible growth, 5 µL were streaked on dNMS agar plates and grown for 10 days. This process was repeated until pure isolates were obtained.

Purity of isolates were tested by observation of cells under Phase Contrast Microscope (PCM) and by PCR of 16S rRNA with primers U515f: GTGYCAGCMGCC GCGGTA and U1071r: GARCTGRCGRCRRCCATG CA (Wang and Qian 2009). Growing liquid cultures of isolates were spread on LB agar, 10% LB agar, Tripticase Soy Agar (TSA) (Sigma-Aldrich, UK) and 10% TSA. The plates were incubated at 30 °C under normal atmospheric pressure. Absence of growth on rich media suggested pure methanotroph cultures.

Conjugation of methanotrophs

Conjugation was carried out based on modifications of the method used by Martin and Murrell (1995). 5 mL of *E. coli* S17-1 λ pir harbouring plasmids pMTL90882 or pMTL71401 (Dr Muhammad Ehsaan, University of Nottingham, unpublished) was grown overnight in LB media containing 50 µg/mL kanamycin. Absorbance (OD_{600}) of the grown culture was measured and used to calculate the volume required to get 1 mL of *E. coli* donor at OD_{600} of 1. The calculated volume of *E. coli* was pipetted in 2 mL Eppendorf tubes and washed three times with NMS media to remove the antibiotics by spinning at 8000 rpm for three minutes. After the third wash, the *E. coli* donor pellet was mixed 1:1 with recipient methanotrophic isolates. The mixture was spun down at 8000 rpm for three minutes and resuspended in 50 µL of NMS which was spotted on NMS agar supplemented with 0.5% yeast extract. The conjugation was incubated at 30 °C with methane for 48 h. The conjugation spot was scraped with a plastic loop and resuspended in 1 mL NMS which was serially diluted to 10^{-7}. Each dilution was spotted in triplicates on NMS agar with 50 µg/mL kanamycin for plasmid retention and 25 µg/mL nalidixic acid as selection against *E. coli*. Dilutions were also spotted on LB media containing 50 µg/mL kanamycin to calculate number of donor cells. NMS media Plates were incubated with methane at 30 °C for two weeks after which transconjugants were enumerated to calculate conjugation efficiency (transconjugants/donor cell) (Phornphisutthimas et al. 2007). LB media plates were incubated at 37 °C overnight to calculate number of donor cells.

Growth of methanotrophs on anaerobic digester and landfill gas

Biogas samples from anaerobic digesters 1–4 (AD1–AD4) and landfill gas samples 1- 3 (LG1–LG3) were collected in 2 L Teddlar bags (Sigma-Aldrich, UK) in different locations around the UK on different dates (Table 1). The AD gases were collected from Staffordshire at different dates while the LG were collected from different sites around the East Midlands of England. Bulk gas composition was measured using Trace GC (see below) while trace gas composition for AD1 and AD2 was measured by Lucideon Ltd in Staffordshire to understand potential trace contaminants.

The first set of experiments involved growth on AD1 and AD2 to determine if methanotrophs could grow on AD biogas. Methanotrophs were grown in duplicates on AD1 and AD2 gases as follows: 9 mL of methanotroph culture at an OD_{600} of 0.02 was added to 165 mL serum bottle together with 75% air and 25% of either AD1, AD2 or CH_4 as carbon source. Samples were incubated at 30 °C for 4 days.

PHB analysis

PHB accumulation in methanotrophs was achieved using a method similar to preliminary assays described in Additional file 1. In short, one 250 mL serum bottle

Table 1 Gas composition measured by GC for CH_4, CO_2, O_2 and by Lucideon Ltd for trace contaminants H_2S, H_2 and NH_3

Sample	Source location	Date of sampling	CH_4 (%)	CO_2 (%)	O_2 (%)	N_2 (%)	H_2S (ppm)	H_2 (ppm)	NH_3 (mg/m^3)
AD1	Staffordshire	17/12/17	49.0	40.3	2.1	7.7	94.0	0	2.9
AD2	Staffordshire	16/01/18	58.6	33.4	2.1	7.8	462.0	0	>0.7
AD3	Staffordshire	26/02/19	62.4	40.5	0.5	0.0	682.0	23.0	0
AD4	Staffordshire	26/02/19	58.0	38.0	0.4	0.0	34.0	16.0	0
LG1	East Midlands	07/01/2019	52.2	37.3	0.4	10.1	n/a	n/a	n/a
LG2	East Midlands	08/01/2019	45.6	31.5	0.4	22.5	n/a	n/a	n/a
LG3	East Midlands	08/01/2019	51.9	38.7	0.4	9.0	n/a	n/a	n/a

n/a not applicable (not measured)

with 35 mL NMS medium was inoculated from colonies growing on plates and grown for several days with pure methane. This culture was used to inoculate main cultures (35 ml NMS in 250 ml serum bottles) to OD_{600} 0.05 and gassed with the respective biogas and air. Biogas was added to result in 0.65 mM methane and air to make up at least a 2:1 molar ratio and to result in 1.5 bar pressure in the serum bottle. This culture was grown for three days, re-gassing every day, before the grown cells were resuspended in NMS media without nitrogen source (potassium nitrate) to trigger PHB accumulation. The culture was then re-gassed daily for another 3 days and then harvested for PHB analysis. Daily re-gassing during PHB accumulation phase was not incorporated in experiments involving anaerobic digester gases.

After three days of PHB accumulation, the cell cultures were pelleted and freeze-dried overnight using the Thermo Micromodulyo Freeze Dryer (Thermo Scientific, UK). The pellets were transferred to a pre-weighed 2 mL Eppendorf and weighed again. These pellets were then transferred to 16 mm diameter round-bottom screw cap centrifuge tube. In a fume cupboard, 100 µL of 1 mg/mL benzoic acid solution in 1-propanol was added. This was followed by 4 mL of 25% concentrated hydrochloric acid in 1-propanol. The glass tubes were then heated at 100 °C for two hours. After cooling, 4 mL of deionised water was added, and each tube was vortexed for 30 s to cause phase separation. The top layer was discarded, 4 mL of deionised water was added, and each tube was vortexed again for 30 s. The top layer was discarded for the second time and 1 mL of the bottom layer was transferred into a GC snap vial cap which was analysed on an Agilent 6890 N Series gas chromatograph, equipped with an Agilent 7983 autosampler, an Agilent 5973 MS detector and a J & W DB-wax column (20 m × 0.18 mm, 0.18 mm film thickness). Injection temperature of 250 °C was applied with standard single split insert with a glass wool packing. The oven program following 1 µL injection was as follows: 5 min hold at 60 °C, 20 °C/min to 240 °C, and hold for 6 min. Split injections were made with 10:1 split ratio. Hydrogen carrier gas with constant flow control was used at 0.6 mL/min. MS analysis was in scanning mode from 40 to 500 m/z produced with EI auto ionization, the MS solvent delay was 3 min and no additional voltage was applied to the electron multiplier.

Results

Isolation of methanotrophs

Samples for methanotroph isolation were taken from two UK locations, a bog in Moseley and a recreational lake at the campus of The University of Nottingham. Isolation and purification procedures for samples of each location varied as outlined in "Materials and methods", and both resulted in the isolation of pure methanotrophic cultures. Purity of both isolates was confirmed by absence of non-methanotrophic bacterial growth on rich media without methane addition. Phase Contrast Microscopy (PCM) also confirmed single bacterial morphology. The 16S rRNA of the isolate from the lake sample showed 100% similarity to *Methylocystis rosea* GW6 and was subsequently designated as *Methylocystis rosea* BRCS1. The isolate from the bog sample shared 100% 16S rRNA sequence similarity to *Methylocystis parvus* OBBP and was designated *Methylocystis parvus* BRCS2.

Morphology of isolated methanotrophs

Morphology of the isolated methanotrophs was studied based on colony formation on agar plates, Phase Contrast Microscopy (PCM) and Transmission Electron Microscopy [TEM (method in Additional file 1)]. *Methylocystis rosea* BRCS1 colonies appeared cream-coloured after 10 days of growth on NMS agar plates. After 2–4 weeks, the cream colour gradually converted to and remained pink (Fig. 1a). Colonies were concave and grew up to 3 mm in diameter. Single cells appeared oval according to PCM imaging (Fig. 1b). However, TEM revealed polymorphic cells with a prominent head-like structure and a thin to thick tail-like structure (Fig. 1c, d). The tail-like structure resembles a rod-like appendage in some cells and a stalk in others. It is speculated that the cells use the

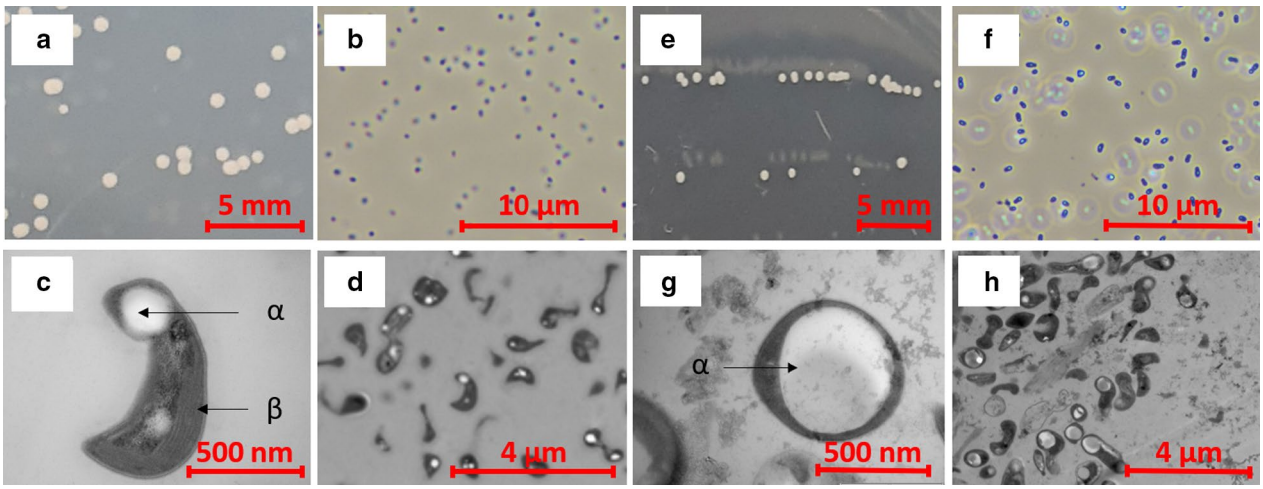


Fig. 1 Morphological studies of BRCS1 on the left (**a**–**d**) and BRCS2 on the right (**e**–**h**). **a**, **e** Respective colony morphology on NMS agar plates. **b**, **f** Phase-contrast photomicrograph at ×100 magnification. **c**, **g** Transmission electron photomicrograph of single cells at ×60,000 magnification. **d**, **h** Transmission electron photomicrograph at ×4200 and ×8200 magnification, respectively. **c** The dumbbell-like shape with a head- and tail-like part. α) The head part includes a light-transmissive granule thought to consist of the storage compound PHB. β) The tail part show striata which are believed to be the ICM. **g** A single coccus-shaped cell with a big light-transmissive granule (α) also thought to consist of the storage compound PHB

tail for adherence on surfaces or to one another as it is the case in other stalked species (Curtis 2017). Measuring cell length and width from TEM images using ImageJ revealed the whole length of a cell including stalk to be 116 ± 19 nm and the width of the head-like structure to be 53 ± 10 nm (mean ± SD, n = 20) (Pérez and Pascau 2013). The head-like structure comprised of a prominent white granule which is suspected to be PHB, used for redox balancing (Fig. 1c). Striations visibly circumscribing the cell periphery are thought to be the Intracytoplasmic Membrane (ICM) (Fig. 1c).

Methylocystis parvus BRCS2 colonies were concave and had a cream appearance (Fig. 1e). Prolonged incubation (more than a month) was observed to lead to drying out and solidifying of the colonies. Single cells had a short, thick, dumbbell-like morphology with a length of 139 ± 20 nm and a width of 65 ± 12 nm (measured from TEM images using ImageJ, mean ± SD, n = 14). BRCS2 presented a prominent white storage granule suspected to be PHB and can be observed in most of the cells (Fig. 1g).

Conjugation efficiency of *Methylocystis rosea* BRCS1 and Methylocystis parvus BRCS2

Efficiency of DNA transfer is an important characteristic of a newly isolated strain with biotechnological potential. Therefore, conjugation efficiency of the newly isolated strains compared to the established laboratory strain *M. parvus* OBBP was determined using two plasmids with different origins of replication (ORI), ColE1 and pBBR1 (Lovett et al. 1974; Antoine and Locht 1992). Efficiency of conjugation was measured as number of transconjugants per donor cell (TC/DC). Plasmid pMTL90882 (ColE1) conjugation efficiency was significantly higher for strain BRCS1 ($2.8 \times 10^{-2} \pm 1.5 \times 10^{-3}$) compared to OBBP ($8.9 \times 10^{-3} \pm 3.3 \times 10^{-3}$) tested by Unpaired t-test ($p = 0.035$) (Fig. 2a). Conjugation Efficiency of pMTL71401 featuring the pBBR1 ORI was $4.3 \times 10^{-3} \pm 1.2 \times 10^{-3}$, $9 \times 10^{-4} \pm 3 \times 10^{-4}$ and $2.6 \times 10^{-4} \pm 3.8 \times 10^{-5}$ for BRCS1, BRCS2 and OBBP respectively (mean ± SEM, n = 2). pMTL71401 (pBBR1) conjugation efficiency does not differ significantly from the new isolates compared to OBBP tested by Unpaired t-test (Fig. 2b).

Genomic DNA sequencing and analyses

Chromosomal DNA of *M. rosea* BRCS1 and *M. parvus* BRCS2 was extracted and sequenced as described in Additional file 1. Both species were found to carry two autonomous replicating plasmids. Annotation was carried out by NCBI with accession number of CP044328, CP044329 and CP044330 for BRCS1 chromosome and plasmids, corresponding to 3,386,331 bp; 195,485 bp and 213,640 bp respectively. GC content was calculated as 62.67%. BRCS2 chromosome and plasmids accession numbers are CP044331, CP044332 and CP044333, corresponding to sizes of 4,075,934 bp; 248,223 bp and 204,886 bp respectively. GC content was calculated as 63.35%.

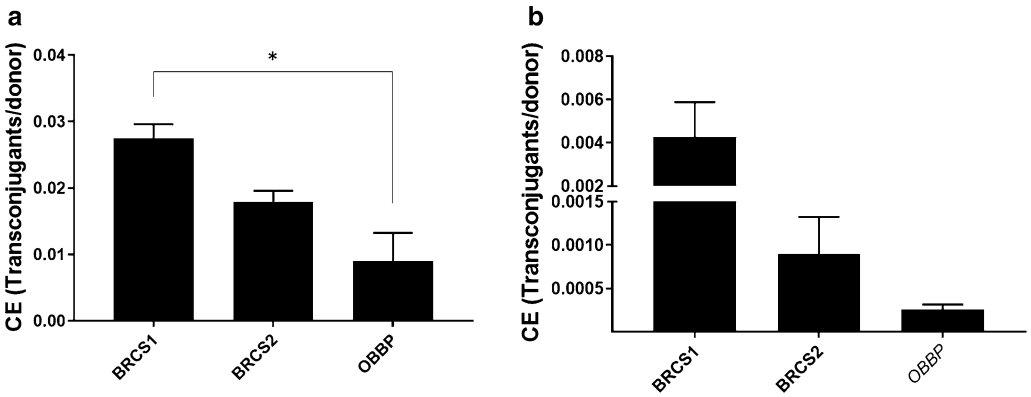


Fig. 2 Conjugation efficiencies of BRCS1 and BRCS2 compared to established strain *M. parvus* OBBP (OBBP) with two differing plasmids. **a** Plasmid pMTL90882 harbouring the ColE1 replicon and **b** pMTL70401 harbouring the pBBR1 replicon. Statistically significant difference is indicated with an asterisk compared to the OBBP control and calculated by t-test as indicated in the text. All samples n = 2 with error-bars representing SEM. Graphs made using GraphPad Prism

The analysis file received from PacBio was sent to Rebase to reveal the Restriction Modification (RM) systems of BRCS2 which was shown to have both type II and III systems (Roberts et al. 2015). BRCS2 was shown to possess RM systems on its genome and mega plasmid that recognise the sequence GANTC, GATC and CTCGAG (Additional file 1: Figure S1 and Table S2). The information obtained from Rebase can be used for increasing genetic transformation efficiencies of isolates. One area where high transformation efficiencies can be desirable is when making transposon mutant libraries in which hundreds of thousands of transconjugants are required which is easier to achieve with high transformation efficiencies.

The completed genome sequences were analysed for genes responsible for carrying out cellular tasks such as DNA repair, homologous recombination and PHB metabolism. DNA repair genes play important roles during genome editing, an area planned for investigation in future studies. Genes involved in Non-Homologous End Joining (NHEJ) during DNA repair were found in isolated strains (Additional file 1: Table S3). *ykoV* and *ykoU* were present in BRCS1 while in BRCS2, the genes involved in NHEJ during DNA repair were *ligD* and the Ku protein genes. In *M. parvus* (OBBP), *ligD* and Ku protein genes were also present. PHB metabolism genes found in *M. rosea* BRCS1 were also found in *M. parvus* BRCS2 (similar to *M. parvus* OBBP) including *phbA*, *phbB* and *phbC*. In addition, an esterase family of PHB depolymerase was found in BRCS1 (Additional file 1: Table S3). Understanding genes involved in PHB metabolism can enable genetic engineering of strains with better PHB accumulation properties.

Phylogenetic tree and genomic alignment of novel isolates

Phylogenetic analysis was carried out to determine the relationship of isolated strains to known strains of methanotrophs and distantly related species (Fig. 3a). *M. rosea* BRCS1 was closely related to the already published *M. rosea* GW6, both of which fell under the same clade. The same relationship was observed between *M. parvus* BRCS2 and the type strain *M. parvus* OBBP. The close relationship was expected due to 100% similarity of their 16S rRNA sequence.

Genomic DNA alignment provided more insight into the similarity of isolates to closely related strains in terms of genomic structure and nucleotide similarity. The comparison of *M. rosea* BRCS1 to *M. rosea* GW6 genome revealed relatively high level of structural dissimilarity (Fig. 3b). Additionally, there was high level of nucleotide dissimilarity when genes of both strains were compared with Average Nucleotide Identity of 94.96%. This was unexpected considering both species had 100% 16S rRNA sequences, suggesting the 16S rRNA does not necessarily imply whole genome resemblance. On the other hand, *M. parvus* BRCS2 and *M. parvus* OBBP showed high similarity in terms of genomic structure and nucleotide comparison which was expected as a result of the 100% 16S rRNA similarity (Fig. 3c). The calculated Average Nucleotide Identity was 99.99%.

Growth characteristics of *Methylocystis rosea* BRCS1 and *Methylocystis parvus* BRCS2

Two crucial media components are important for growth of methanotrophs–nitrogen and carbon source. As such, we set out to test various sources of both growth components.

Growth on different nitrogen sources

Growth was tested on various potential nitrogen sources (potassium nitrate, ammonium chloride, asparagine, glutamine, ornithine, aspartate, lysine, and putrescine) by measuring final OD_{600} after 14 days of growth. Both strains BSRC1 and BSRC2 grew best with potassium nitrate as nitrogen source. While BRCS1 can grow on ammonium chloride, glutamine and ornithine, final OD_{600} is reduced to about 50% whereas final OD_{600} of BRCS2 is reduced to roughly 20% on these nitrogen sources compared to growth on potassium nitrate (Additional file 1: Figure S2).

Growth on methanol

After determining the suitability of potassium nitrate as nitrogen source in the media, both strains were tested on the ability to grow on methanol instead of methane as sole carbon source. After 10 days of shaking and incubation at 30 °C in 65 mL serum bottles, *M. parvus* BRCS2 like its closest evolutionary relative *M. parvus* OBBP was able to grow with 0.1% methanol vapour, while BRCS1 was not able to grow with the methanol concentrations tested in the range 0.01–1% (Fig. 4a).

Methanotrophic growth on biogas from anaerobic digesters (AD)

Anaerobic digester gas composition

Bulk composition of biogases from anaerobic digesters (AD1–AD4) and gases from landfill sites (LG1–LG3) were measured by trace GC (Table 1). Methane content ranged from 49% in AD1 to 62.4% in AD3. Carbon dioxide content was measured at 32% in LG2 and up to 40.5% in AD3. All samples contained traces of oxygen which are assumed to originate from gas exchange through the Teddlar gas collection bags and not from the original sample as those environments are expected to be anaerobic. The rest of the gas composition is made up of nitrogen, hydrogen sulphide and other trace gases. Ammonia, hydrogen sulphide and siloxane composition of gases AD1 and AD2 were assessed and found to be minimal (Table 1 and Additional file 1: Table S4).

The ability of the isolates to grow on renewable forms of biogas from anthropogenic sources such as anaerobic digesters (AD) and landfills is crucial if they are to be utilised for industrial biotechnology. Therefore, an initial experiment was conducted to investigate growth of the novel strains on un-purified gas from AD which carries potentially toxic contaminants such as ammonia, siloxanes, hydrogen sulphide and aromatics as well as halogenated compounds (Rasi et al. 2007). Growth of BRCS1, BRCS2 and OBBP using two biogas samples (AD1 and AD2) was compared to growth on CH_4 as control (Fig. 4b). No significant growth difference was observed of strains growing on pure CH_4 compared to growth on AD gas sources, tested by two-way ANOVA with Dunnett's post hoc test ($p > 0.05$). These findings suggest that for these species, contamination of up to 1.1 mg/m^3 siloxane (Additional file 1: Table S4) and 682 ppm H_2S as well as 2.9 mg/m^3 ammonia are non-problematic in methanotrophic culturing as speculated, opening the possibility of integrated methanotrophic facilities at AD sites.

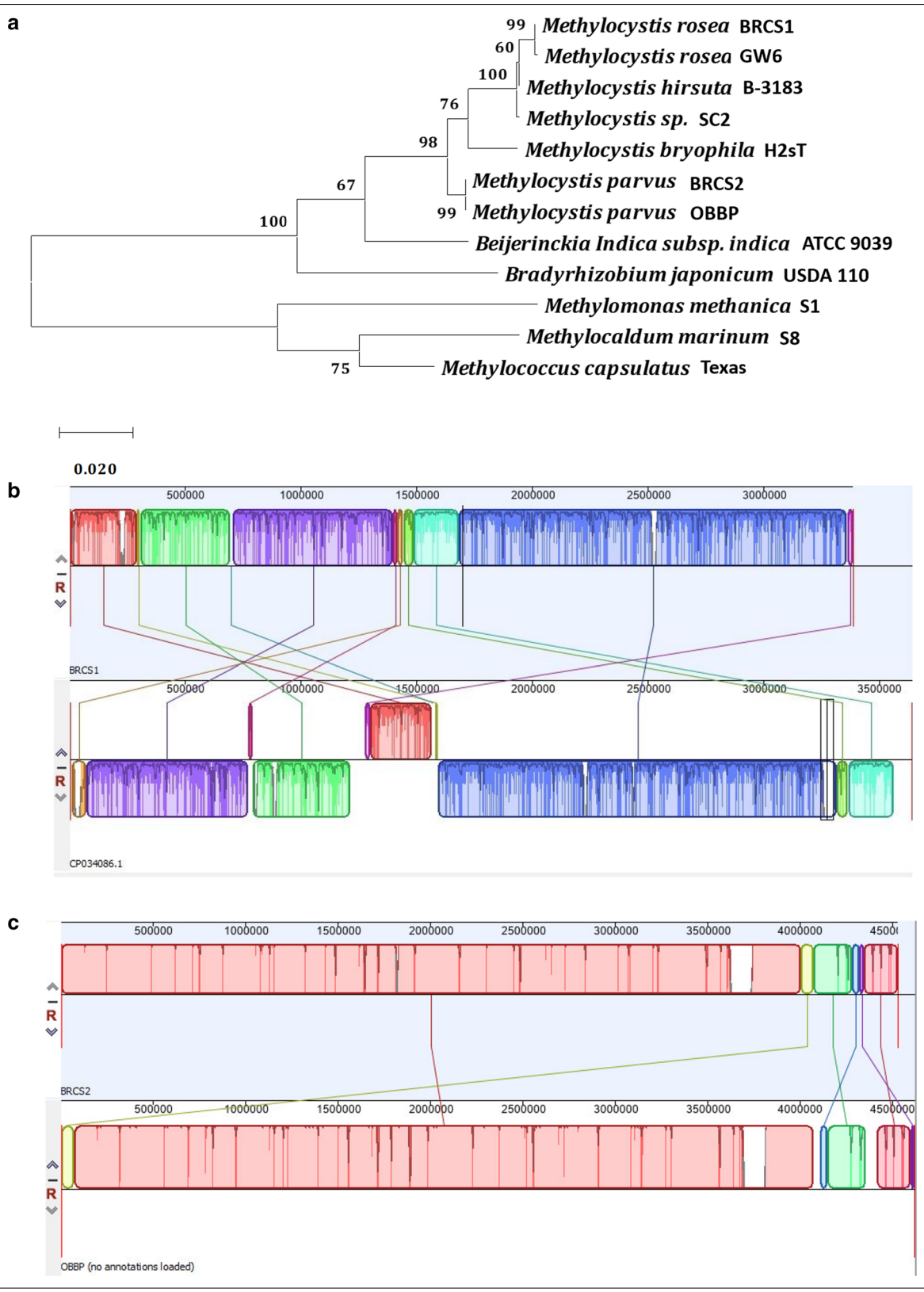
PHB production in methanotrophs

Having shown that contaminants in biogas from anaerobic digesters are non-toxic to the new isolates and the type strain *M. parvus* OBBP, it was further tested if the isolates produce the biopolymer poly-3-hydroxybutyrate (PHB) when biogases from landfill sites (LG1/2/3) and anaerobic digesters (AD3/4) are used as CH_4 source. It has been shown that PHB accumulation can be triggered by nitrogen limitation in the medium (Wendlandt et al. 2001; Listewnik et al. 2007; Pieja et al. 2011). Hence, we adopted a two-stage growth and production method to maximise PHB accumulation. Peak PHB accumulation was observed on day 3 of the second growth stage (nitrate-limiting conditions). Incubation lasting longer than 3 days under nitrate-limiting conditions led to decreased PHB yield as experiments showed (Fig. 5a, b). For this initial experiment, the type strain *Methylocystis parvus* OBBP was used with pure CH_4 serving as the source of carbon.

Once it was established that PHB accumulation under nitrate-limiting conditions peaked on day 3 in the preliminary experiment, subsequent experiments harvested cell cultures for PHB assays on day 3 of incubation on nitrate-free media. Four strains (*M. rosea* BRCS1, *M. parvus*

(See figure on next page.)

Fig. 3 Phylogenetic tree comparing isolated strains of *Methylocystis* species with related species and whole genome alignment of isolated strains with closest relatives based on 16S rRNA similarity. **a** The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 12 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. Bootstrap method was used as test of phylogeny with 1000 number of Bootstrap Replications. There were a total of 1547 positions in the final dataset. Evolutionary analyses were conducted in MEGA X9. b/c) The alignment shows structural and single nucleotide similarity between isolated strains and the closely related strains they were compared to. With Average Nucleotide Identity of 94.96%, *M. rosea* BRCS1 had significant levels of structural and single nucleotide differences when compared to *M. rosea* GW6 with accession number CP034086.1. **a** There was significantly less structural and single nucleotide difference between *M.* parvus BRCS2 and *M. parvus* OBBP with Average Nucleotide Identity of 99.99% (**b**)


a
99
60
100
76
98
67
99
100
75
Methylocystis rosea BRCS1
Methylocystis rosea GW6
Methylocystis hirsuta B-3183
Methylocystis sp. SC2
Methylocystis bryophila H2sT
Methylocystis parvus BRCS2
Methylocystis parvus OBBP
Beijerinckia Indica subsp. indica ATCC 9039
Bradyrhizobium japonicum USDA 110
Methylomonas methanica S1
Methylocaldum marinum S8
Methylococcus capsulatus Texas
0.020
b
500000
1000000
1500000
2000000
2500000
3000000
3500000
BRCS1
CP034086.1
c
4000000
4500000
BRCS2
OBBP (no annotations loaded)

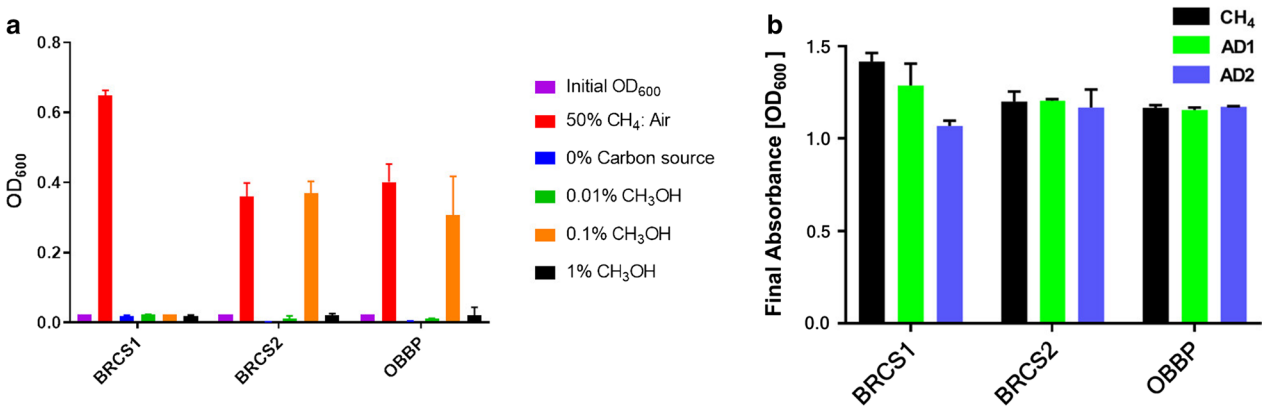


Fig. 4 Growth of methanotrophs on different carbon sources. **a** Growth of methanotrophs with different CH_3OH concentrations. Growth in CH_4 was used as control. There was significant difference across all absorbance values measured with $p < 0.0001$ using one-way ANOVA. **b** Growth characteristics of new isolates BRCS1 and BRCS2 were compared with control strain OBBP on biogases from anaerobic digesters (AD1/2). Final absorbance at OD_{600} was compared after 4 days of growth (n = 2). Graphs made using GraphPad Prism

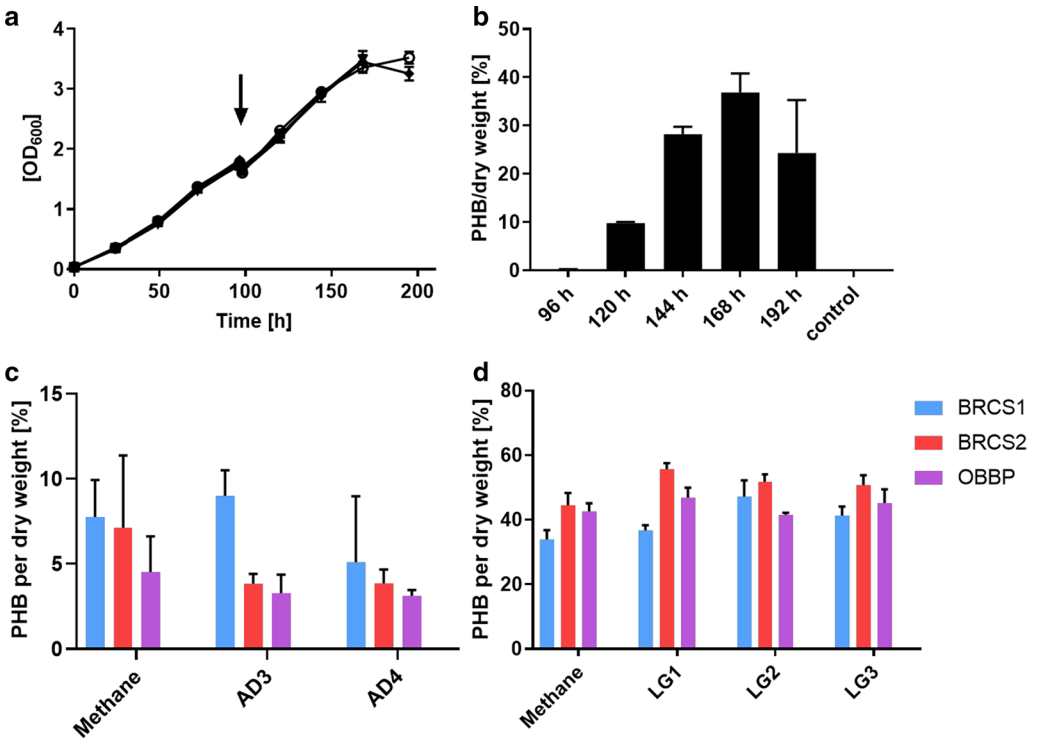


Fig. 5 Poly-3-hydroxybutyrate (PHB) production of new methanotroph isolates. **a** Preliminary experiment to test optimal PHB accumulation with type strain OBBP. Growth was measured during the pre-culture which grew supplemented with nitrate (NO^-_3) which was harvested after 96 h (indicated by and arrow) and resuspended in medium without nitrate. Samples with black symbols and control (grown with nitrate for 192 h) with empty symbol. **b** PHB accumulation over time during nitrate limited phase (n = 2). PHB production of new methanotroph isolates compared to PHB producing type strain *M. parvus* OBBP and PHB negative strain *Methylococcus capsulatus* Bath from two individual experiments (**c** and **d**). **c** Test of methanotrophic PHB production on biogas from AD unfiltered (AD3) and filtered (AD4). **d** Test of methanotrophic PHB production on biogas from three different landfills. OBBP served as PHB producing positive control of the type II methanotrophs and Bath as negative control. The bar chart shows PHB per cell dry weight accumulated in a 35 mL culture after 3 days growth followed by three days growth without nitrate. All samples n = 3 with error-bars representing SEM. Graphs made using GraphPad Prism

BRCS2, *M. parvus* OBBP and *M. capsulatus* Bath which served as control) were then grown using the respective landfill and AD gases as CH_4 source. The results showed that all type II methanotrophs tested were able to grow and accumulate PHB using pure CH_4 as well as landfill and AD gas as source of CH_4. It was observed that PHB accumulation of all strains when grown on landfill gas was higher than when grown on AD gas. Although there is a possibility that landfill gas components can trigger higher PHB accumulation, it is more likely as a result of the method used during PHB accumulation of the separate experiments. The results further show that *M. parvus* BRCS2 produced the most PHB per dry cell weight ($55.7 \pm 1.9\%$) compared to the other strains on all gases tested. When grown on AD gases, *M. parvus* BRCS2 also performed better in producing PHB per dry cell weight ($24.8 \pm 2.0\%$) than other strains. The control strains performed as expected. *M. capsulatus* Bath as a type I methanotroph acting as negative control did not produce any PHB, while *M. parvus* OBBP acting as positive control produced up to $46.8 \pm 3.2\%$ dry cell weight of PHB when grown on landfill gas (Fig. 5c, d). This is close to the highest value of $50.3 \pm 3.3\%$ obtained in the study carried out by Pieja et al. (2011). When PHB accumulation was compared between filtered (AD3) and unfiltered (AD4) AD gas, no significant difference ($P = 0.253$) was observed for all strains using one-way Anova. The unfiltered AD gas has a H_2S content about 20 times higher than the filtered gas.

Discussion

Isolation of methanotrophs was carried out in this study which resulted in two pure *Methylocystis* species. The characteristics of these isolates including morphology, physiology, genetics, genomics and the ability to accumulate PHB on biogas were investigated.

Isolation of methanotrophs from bog and lake samples was achieved without immediate addition of methane when the environmental samples were collected. Further processing of the samples took place after 24 h suggesting ability of methanotrophs to survive short term in the absence of methane. In fact, separate experiments leading to the isolation of a *Methylomonas* species and a novel strain of *Methylococcus capsulatus* were carried out after storage of the initial samples for four months at 4 °C (unpublished data). Previous studies reported survival of *Methylosinus trichosporium* OB3b after 10 weeks of CH_4 starvation (Roslev and King 1994). These observations suggest that it is not essential to collect samples planned for methanotroph isolation under methane-enriched conditions. This potentially eases the process of sample collection for future isolations.

Isolation of methanotrophs was done according to protocols established previously with some adjustments aiding the purification of the strains presented here (Hoefman et al. 2012). Firstly, samples were cyclically alternated between growth in liquid culture and on agar plates. This alternation on the one hand allowed quick purification of desired methanotrophs from non-methanotrophic bacteria which can feed on methanotrophic metabolites such as acetate, formate and lactate (Whittenbury et al. 1970). On the other hand, some non-methanotrophs can grow on polysaccharides which make up agar, therefore switching from agar to liquid media potentially eliminated these contaminants (Payton and Roberts 1976; Imran et al. 2016). Furthermore, extinction dilution culturing was employed when transferring to liquid medium which further benefitted purification (Hoefman et al. 2012). The other important adjustment made to previous studies was the omission of vitamins in the defined NMS medium, potentially reducing the number of contaminating organisms auxotrophic for these vitamins. A disadvantage of this approach is the simultaneous loss of auxotrophic methanotrophs. However, isolation of methanotrophic species not requiring expensive vitamins in the medium (such as the isolates presented here) can be advantageous for subsequent industrial applications.

The newly isolated strains showed similar colony morphology at the outset of culturing on agar plates which manifests in round, cream coloured colonies. However, after about two weeks of growth, *M. rosea* BRCS1 transitioned to a pink colour potentially due to biosynthesis of carotenoids as was observed in other methanotrophic isolates (Leadbetter and Foster 1958). This hypothesis is further substantiated by the identification of the operon *crtBCDL* responsible for carotenoid biosynthesis in the genome of BRCS1. A subset of these genes, not including *crtD* were identified in BRCS2 as well, however the pathway does not seem active or is incomplete without *crtD*, suggested by the lack of pigment synthesis.

All strains could be transformed with plasmids carrying ColE1E and pBBR1 replicons via bacterial conjugation, with ColE1 replicon having a higher conjugation efficiency. The superiority of pMTL90882 over pMTL71401 was not surprising as ColE1 is a high copy number replicon (approximately 40 copies per chromosome in *E.coli*) compared to pBBR1 which has a low copy number (approximately 5 copies per chromosome in *E.coli*) (Jahn et al. 2016). The ability to accept heterologous plasmid DNA is advantageous for novel strains with biotechnological applicability. This enables designed genetic manipulation that can increase flux along desired biosynthetic pathways for products such as PHB. Furthermore, the complete genome sequence of the isolated

strains offers more insight and explanation of some of the characteristics observed and suspected. Already, key genes have been mentioned involved in PHB metabolism and DNA repair pathways. The presence of mega plasmids was revealed in *M. parvus* BRCS2 which is likely present in *M. parvus* OBBP sequenced in 2012 but was not observed probably due to the sequencing technology at the time (del Cerro et al. 2012). This finding is crucial as it points towards greater understanding of the industrially relevant and widely studied *M. parvus* species. Mega plasmids can be used for plasmid addiction systems which enable industrial biotechnology applications of genetically engineered microorganisms.

With full 16S rRNA sequence from the genome of all strains, a more accurate phylogenetic analysis was possible. The phylogenetic tree constructed placed all type II methanotrophs such as *Methylocystis* species closer to one another compared to type I *methanotrophs.* This is not unreasonable, especially when taking into consideration that type II methanotrophs such as *Methylocystis* species which are *alphaproteobacteria* have a different process of carbon assimilation compared to type I methanotrophs like *Methylococcus capsulatus* which are methanotrophs *gammaproteobacteria* (Hanson and Hanson 1996). Further insight into the distant evolutionary ancestry between *alphaproteobacteria* and *gammaproteobacteria* is provided by evidence hinting that members of the family *Methylocystaceae* which are *alphaproteobacteria* have not always had the ability to oxidise methane. This ability is likely the result of lateral gene transfer from a methanotrophic *gammaproteobacteria* (Tamas et al. 2014).

Two sources of macronutrients are essential for normal growth of methanotrophs, nitrogen and carbon source. As such the effect they have on growth was investigated. The importance of investigating various nitrogen sources cannot be overemphasized because nitrogen starvation is directly linked to PHB accumulation. Furthermore, the choice of nitrogen source during methanotrophic growth phase was shown to influence PHB accumulation (Rostkowski et al. 2013). Potassium nitrate used throughout this study and in the commonly used NMS media was shown to be the best at supporting growth of the strains tested (Whittenbury et al. 1970).

Both strains of *M. parvus* were able to grow on methanol. The growth of *M. parvus* OBBP in methanol was already demonstrated in a previous study (Hou et al. 1979). However, it was important to verify this characteristic in the isolated strains *M. rosea* BRCS1 and *M. parvus* BRCS2. The ability of *M. parvus* BRCS2 to grow on methanol is noteworthy since this allows handling of the organism in places where methane atmosphere is not an option such as liquid handling robots. It can furthermore seamlessly slot into a methanol-based economy as proposed by Olah et al. (2006).

Investigating biogas sources is crucial because the ability of methanotrophs to grow can be influenced by the chemical composition of the biogas sources. For example, acetylene commonly found in natural gas was shown to inactivate the soluble methane monoxygenase used for methane oxidation in *Methylococcus capsulatus* (Prior and Dalton 1985). The successful growth of *Methylocystis* species on two different sources of AD gas (AD1 and AD2) provided the foundation and supporting evidence for further investigation of methanotrophic growth and PHB accumulation using other AD and landfill biogas sources. Additionally, the findings suggest that trace contaminants are non-problematic in methanotrophic culturing. Biogas from AD and landfill sites are currently mostly used for electricity generation but are flared or vented especially when CH_4 concentration is low (Cashdollar et al. 2000; Tollefson 2016). In such circumstance, methanotrophs offer the possibility to utilise methane from low quality biogas.

In type II methanotrophs such as *M. parvus* OBBP, PHB is suspected to play a role in redox balancing and its accumulation usually manifests when nitrogen becomes limiting (Pieja et al. 2011). In this study, PHB accumulation was stimulated by nitrate starvation where nitrate-free media was used to incubate already grown cultures of methanotrophs. PHB accumulation peaked on day 3 of incubation in nitrate-free media as shown by (Asenjo and Suk 1986). PHB accumulation was higher when landfill gas was used as source of CH_4. Although there is a possibility that landfill gas components can trigger higher PHB accumulation, it is more likely as a result of the method used during PHB accumulation of the separate experiments. Whereas daily air replenishment was carried out during PHB accumulation phase when methanotrophic strains were grown on landfill gas, that was not the case when methanotrophs were grown on AD gas. This suggests a positive influence of oxygen on PHB accumulation which was replenished daily when landfill gas was used as a source of CH_4. The negative effect of oxygen limitation on PHB accumulation has been previously reported (Zhang et al. 2019).

The unfiltered AD gas has H_2S content about 20 times higher than the filtered gas which was the only significant difference observed with the limited analysis conducted. The absence of significant H_2S effect on PHB accumulation was also reported by López et al. (2018). However, the study conducted by López et al. (2018) significantly differs from ours considering synthetic biogas was used. Additionally, the study used *Methylocystis hirsuta,* a different species in the *Methylocystis* genus.

The highest PHB content per cell dry weight was measured to be 55.7 ± 1.9% with *M. parvus* BRCS2 on LG1. Similar yields were found in optimised bioreactor cultures which suggests a good performance of the isolate (Listewnik et al. 2007). However, as high as 60% PHB accumulation in *M. parvus* OBBP was reported by Rostkowski et al. (2013) although ammonium chloride was used as nitrogen source during the growth phase.

PHB can be harvested and used to make bioplastic which is not only from biological source but also biodegradable, a differentiation not always made in sustainability research (Mekonnen et al. 2013). The CO_2 produced by the methanotrophs can be captured easily as fermenter off-gas and recycled in greenhouses for the growth of plants or in secondary fermentation by algae or syngas fermenting bacteria, further aiding mitigation of greenhouse gases (Mortensen 1987; Sayre 2010; Bengelsdorf 2013).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13568-020-01159-4.

Additional file 1. Supplementary information providing additional methods and results. Methods and results include: Bacterial strains and plasmid used in this study (Table S1); Method and result for next generation sequencing analysis; Rebase analysis of restriction patterns (Figure S1 and Table S2); Comparison of genes shared by isolates (Table S3); Method of Phase Contrast and Transmission Electron Microscopy; Method of phylogenetic tree analysis and whole genome alignment; Growth on different nitrogen sources (Figure S2); Siloxane composition of AD gases (Table S4); Method of preliminary PHB accumulation assay.

Acknowledgements
BBSRC, EPSRC, The University of Nottingham and PTDF (Nigeria) played no role in the design of the study and collection, analysis, and interpretation of data, or in writing the manuscript. The authors thank BioG-UK (http://www.biog-uk.co.uk/) for AD biogas and Freeland Horticulture (http://www.freelandhorticulture.co.uk/) for landfill gases. The authors acknowledge the help of Matthew Abbott and James Fothergill in the analysis of methanotrophic metabolites, Denise Mclean for TEM imaging, Dr Muhammad Ehsaan for constructive discussions and Dr Gareth Little for the in-silico analysis of Illumina and PacBio data.

Authors' contributions
BLR and CES contributed equally to the work. BLR, CES and AGH carried out the laboratory work, data analysis and drafted the manuscript. NPM helped design the study and edited the manuscript. BCS participated in the laboratory work and coordination of the final part of the study. The manuscript was written through contributions of all authors. YZ conceived the study, oversaw its design and coordination, helped with the data analysis and revised the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC; Grant numbers BB/L013940/1, BB/N010701/1 and BB/L013800/1) and the Engineering and Physical Sciences Research Council (EPSRC; Grant number BB/L013940/1). Genome sequencing was provided by MicrobesNG (http://www.microbesng.uk), which is supported by the BBSRC (grant number BB/L024209/1). We thank The University of Nottingham supporting the PhD studentships of CES. BLR acknowledges the financial support of the Petroleum Technology Development Fund (PTDF) Nigeria.

Availability of data and materials
The data that support findings in this study are openly available in NCBI as referenced in Additional file 1 of this study. All strains are deposited in NCIMB (National Collection of Industrial, Food and Marine Bacteria https://www.ncimb.com/) with the following accession numbers: NCIMB 15262 *Methylocystis parvus* BRCS2; NCIMB 15263 *Methylocystis rosea* BRCS1.

Ethics approval and consent to participate
This article does not contain any studies with human participants or animals performed by any of the authors.

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

Received: 28 August 2020 Accepted: 7 December 2020
Published online: 06 January 2021

References

Allen DT, Torres VM, Thomas J, Sullivan DW, Harrison M, Hendler A, Herndon SC, Kolb CE, Fraser MP, Hill AD, Lamb BK (2013) Measurements of methane emissions at natural gas production sites in the United States. Proc Natl Acad Sci 110(44):17768–17773

Anthony C (1982) The biochemistry of methylotrophs, vol 439. Academic press, London

Antoine R, Locht C (1992) Isolation and molecular characterization of a novel broad-host-range plasmid from *Bordetella Bronchiseptica* with sequence similarities to plasmids from gram-positive organisms. Mol Microbiol 6(13):1785–1799. https://doi.org/10.1111/j.1365-2958.1992.tb01351.x

Asenjo JA, Suk JS (1986) Microbial conversion of methane into poly-β-hydroxybutyrate (PHB): growth and intracellular product accumulation in a type II methanotroph. J Ferment Technol 64(4):271–278. https://doi.org/10.1016/0385-6380(86)90118-4

Bengelsdorf FR, Straub M, Dürre P (2013) Bacterial synthesis gas (syngas) fermentation. Environ Technol 34(13–14):1639–1651. https://doi.org/10.1080/09593330.2013.827747

Boeckx P, Van Cleemput O, Villaralvo I (1996) Methane emission from a landfill and the methane oxidising capacity of its covering soil. Soil Biol Biochem 28(10–11):1397–1405. https://doi.org/10.1016/S0038-0717(96)00147-2

Bourne DG, McDonald IR, Murrell JC (2001) Comparison of PmoA PCR primer sets as tools for investigating methanotroph diversity in three danish soils. Appl Environ Microbiol 67(9):3802–3809. https://doi.org/10.1128/AEM.67.9.3802-3809.2001

Cashdollar KL, Zlochower IA, Green GM, Thomas RA, Hertzberg M (2000) Flammability of methane, propane, and hydrogen gases. J Loss Pre Process Ind 13(3–5):327–340. https://doi.org/10.1016/S0950-4230(99)00037-6

Cole M, Lindeque P, Fileman E, Halsband C, Goodhead R, Moger J, Galloway TS (2013) Microplastic ingestion by zooplankton. Environ Sci Technol 47(12):6646–6655. https://doi.org/10.1021/es400663f

Curtis PD (2017) Stalk formation of *Brevundimonas* and how it compares to caulobacter crescentus. PLoS ONE. https://doi.org/10.1371/journal.pone.0184063

Del Cerro C, García JM, Rojas A, Tortajada M, Ramón D, Galán B, Prieto MA, García JL (2012) Genome sequence of the methanotrophic poly-β-hydroxybutyrate producer *Methylocystis parvus* OBBP. J Bacteriol 194:5709–5710

Derraik JGB (2002) The pollution of the marine environment by plastic debris: a review. Mar Pollut Bull. https://doi.org/10.1016/S0025-326X(02)00220-5

Environmental Protection Agency (2011) Management of low levels of landfill gas. https://www.epa.ie/pubs/advice/waste/waste/EPA%20_Management_Of_Low_Levels_Of_Landfill_Gas.pdf. Accessed 31 Mar 2020

Eriksen M, Lebreton LCM, Carson HS, Thiel M, Moore CJ, Borerro JC, Galgani F, Ryan PG, Reisser J (2014) Plastic pollution in the world's oceans: more than 5 trillion plastic pieces weighing over 250,000 tons afloat at sea. PLoS ONE. https://doi.org/10.1371/journal.pone.0111913

Hanson RS, Hanson TE (1996) Methanotrophic bacteria. Microbiol Mol Biol Rev 60(2):439–471

Hoefman S, van der Ha D, De Vos P, Boon N, Heylen K (2012) Miniaturized extinction culturing is the preferred strategy for rapid isolation of fast-growing methane-oxidizing bacteria. Microb Biotechnol 5(3):368–378. https://doi.org/10.1111/j.1751-7915.2011.00314.x
Imran M, Poduval PB, Ghadi SC (2016) Bacterial degradation of algal polysaccharides in marine ecosystem. Marine pollution and microbial remediation. Springer, Singapore, pp 189–203. https://doi.org/10.1007/978-981-10-1044-6_12
Jahn M, Vorpahl C, Hübschmann T, Harms H, Müller S (2016) Copy number variability of expression plasmids determined by cell sorting and Droplet Digital PCR. Microb cell fact 15(1):211
Leadbetter E, Foster J (1958) Studies on some methane-utilizing bacteria. Arch Microbiol 30(1):91–118
Listewnik HF, Wendlandt KD, Jechorek M, Mirschel G (2007) Process design for the microbial synthesis of poly-β-hydroxybutyrate (PHB) from natural gas. Eng Life Sci 7(3):278–282. https://doi.org/10.1002/elsc.200620193
López JC, Arnáiz E, Merchán L, Lebrero R, Muñoz R (2018) Biogas-based polyhydroxyalkanoates production by *Methylocystis Hirsuta*: a step further in anaerobic digestion biorefineries. Chem Eng J 333:529–536. https://doi.org/10.1016/j.cej.2017.09.185
Lovett MA, Katz L, Helinski DR (1974) Unidirectional replication of plasmid ColE1 DNA. Nature 251(5473):337–340. https://doi.org/10.1038/251337a0
Martin H, Murrell J (1995) Methane Monooxygenase mutants of *Methylosinus Trichosporium* constructed by marker-exchange mutagenesis. FEMS Microbiol Lett 127(3):243–248. https://doi.org/10.1111/j.1574-6968.1995.tb07480.x
Mekonnen T, Mussone P, Khalil H, Bressler D (2013) Progress in bio-based plastics and plasticizing modifications. J Mater Chem A 1(43):13379–13398. https://doi.org/10.1039/c3ta12555f
Mortensen LM (1987) Review: CO2 enrichment in greenhouses. Sci Horticult, Crop Responses. https://doi.org/10.1016/0304-4238(87)90028-8
Olah GA, Goeppert A, Prakash GKS (2006) Beyond oil and gas: the methanol economy. Ang Chem Int Ed 44(18):2636–2639
Payton M, Roberts CF (1976) Agar as a carbon source and its effect on the utilization of other carbon sources by acetate non-utilizing (Acu) mutants of aspergillus nidulans. J Gen Microbiol 94(r976):228–233
Pérez J, María M, Javier P (2013) Image processing with ImageJ. Packt Publishing Ltd, Birmingham
Phornphisutthimas S, Thamchaipenet A, Panijpan B (2007) Conjugation in *Escherichia Coli*: A laboratory exercise. Biochem Mol Biol Educ 35(6):440–445. https://doi.org/10.1002/bmb.113
Pieja AJ, Sundstrom ER, Criddle CS (2011) Poly-3-Hydroxybutyrate metabolism in the type II methanotroph *Methylocystis Parvus* OBBP. Appl Environ Microbiol 77(17):6012–6019. https://doi.org/10.1128/AEM.00509-11
Prior SD, Dalton H (1985) Acetylene as a suicide substrate and active site probe for methane monooxygenase from *Methylococcus capsulatus* (Bath). FEMS Microbiol Lett 29(1–2):105–109
Rasi S, Veijanen A, Rintala J (2007) Trace compounds of biogas from different biogas production plants. Energy 32(8):1375–1380. https://doi.org/10.1016/j.energy.2006.10.018
Roberts RJ, Vincze T, Posfai J, Macelis P (2015) REBASE—a database for DNA restriction and modification: enzymes genes genomes. Nucleic Acid Res 43(D1):D298–D299
Rodhe HA (1990) Comparison of the Contribution of Various Gases to the Greenhouse Effect. Science (80-) 248(4960):1217–1219. https://doi.org/10.1126/science.248.4960.1217
Roslev P, King GM (1994) Survival and recovery of methanotrophic bacteria starved under oxic and anoxic conditions. Appl Environ Microbiol 60(7):2602–2608
Rostkowski KH, Pfluger AR, Criddle CS (2013) Stoichiometry and kinetics of the PHB-producing Type II methanotrophs *Methylosinus trichosporium* OB3b and *Methylocystis parvus* OBBP. Biores technol 132:71–77. https://doi.org/10.1016/j.biortech.2012.12.129
Sayre R (2010) Microalgae: the potential for carbon capture. Bioscience 60(9):722–727. https://doi.org/10.1525/bio.2010.60.9.9
Shindell DT, Faluvegi G, Koch DM, Schmidt GA, Unger N, Bauer SE (2009) Improved attribution of climate forcing to emissions. Science 326(5953):716–718
Strong PJ, Kalyuzhnaya M, Silverman J, Clarke WP (2016) A Methanotroph-based biorefinery: potential scenarios for generating multiple products from a single fermentation. Biores Technol. https://doi.org/10.1016/j.biortech.2016.04.099
Tamas I, Smirnova AV, He Z, Dunfield PF (2014) The (d) evolution of methanotrophy in the *Beijerinckiaceae*—a comparative genomics analysis. ISME j 8(2):369–382
Tokiwa Y, Calabia BP, Ugwu CU, Aiba S (2009) Biodegradability of plastics. Int J Mol Sci. https://doi.org/10.3390/ijms10093722
Tollefson J (2016) "Flaring" wastes 3.5% of world's natural gas. Nature. https://doi.org/10.1038/nature.2016.19141
Wang Y, Qian PY (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. PLoS ONE. https://doi.org/10.1371/journal.pone.0007401
Wendlandt KD, Jechorek M, Helm J, Stottmeister U (2001) Producing poly-3-hydroxybutyrate with a high molecular mass from methane. J Biotechnol 86(2):127–133. https://doi.org/10.1016/S0168-1656(00)00408-9
Whittenbury R, Phillips KC, Wilkinson JF (1970) Enrichment, isolation and some properties of methane-utilizing bacteria. J Gen Microbiol 61(2):205–218. https://doi.org/10.1099/00221287-61-2-205
Wright SL, Kelly FJ (2017) Plastic and human health: a micro issue? Environ Sci Technol 51(12):6634–6647. https://doi.org/10.1021/acs.est.7b00423
Yeo JCC, Muiruri JK, Thitsartarn W, Li Z, He C (2018) Recent advances in the development of biodegradable phb-based toughening materials: approaches, advantages and applications. Mat Sci Eng. https://doi.org/10.1016/j.msec.2017.11.006
Zhang Z, Zimmermann NE, Stenke A, Li X, Hodson EL, Zhu G, Huang C, Poulter B (2017) Emerging role of wetland methane emissions in driving 21st century climate change. Proc Natl Acad Sci 114(36):9647–9652
Zhang T, Zhou J, Wang X, Zhang Y (2019) Poly-β-hydroxybutyrate production by *Methylosinus trichosporium* OB3b at different gas-phase conditions. Iran j biotechnol. https://doi.org/10.21859/ijb.1866

Publisher's Note

ACS Synthetic Biology

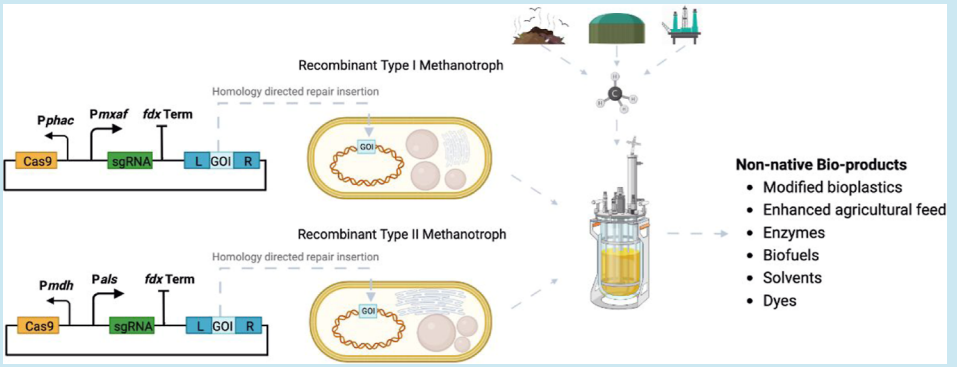
pubs.acs.org/synthbio Research Article

In Vivo Genome Editing in Type I and II Methanotrophs Using a CRISPR/Cas9 System

Bashir L. Rumah,[†] Benedict H. Claxton Stevens, Jake E. Yeboah, Christopher E. Stead, Emily L. Harding, Nigel P. Minton, and Ying Zhang*

Cite This: *ACS Synth. Biol.* 2023, 12, 544–554 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Methanotrophic bacteria are Gram-negative, aerobic organisms that use methane as their sole source of carbon and energy. In this study, we constructed and exemplified a CRISPR/Cas9 genome editing system and used it to successfully make gene deletions and insertions in the type I methanotroph *Methylococcus capsulatus* Bath and the type II methanotroph *Methylocystis parvus* OBBP. High frequencies of gene deletions and insertions were achieved in combination with homology-directed repair. In *M. parvus* OBBP, we also investigated the impact of several parameters on the CRISPR/Cas9 genome editing, where the *ligD* gene was targeted with various PAM sequences and guide RNA spacer sequences, homology arms of variable length, differences in the duration of mating during conjugation, and exploiting promoters of different strengths to control the expression of cas9 and sgRNA. Although not the first attempt to develop a CRISPR/Cas system in methanotrophs, this work demonstrated for the first time an efficient CRISPR/Cas9 system generating scarless clean gene deletions and insertions in methanotroph genomes.

KEYWORDS: *methanotrophs, CRISPR, genome editing, methane monooxygenase, gene deletion, gene insertion, DNA ligase, promoter library, homology-directed repair*

INTRODUCTION

Methane (CH_4) ranks as the second most abundant anthropogenic greenhouse gas (GHG) next to carbon dioxide.[1] It can serve as the sole source of carbon and energy source for a group of microorganisms called methanotrophs, which are either aerobic bacteria or anaerobic archaea.[2,3] As methanotrophs represent the primary biological sink of methane in the atmosphere and soil, they are extremely important in helping to control the levels of methane in the environment[4] and therefore play an important role in global climate change. In recent times, methanotrophs are also utilized in industrial biotechnology, most notably in the manufacture of animal feed in the form of single cell protein.[5]

Methanotrophs are ubiquitous and can be found growing in most places where methane is emitted, such as landfill sites, lake sediments, wetlands, and marine environments.[6−9] The distinguishing characteristic that enables methanotrophs to oxidize CH_4 and survive in different environments is their exclusive possession of a broad-spectrum methane monooxygenase (MMO) enzyme.[10] This enzyme is present in two forms, each found in different parts of the cell. The particulate form (pMMO) is membrane-associated, while the soluble form (sMMO) is found in the cytoplasm.[11] Based on the morphology of their internal structures, particularly the intracytoplasmic membrane (ICM) where pMMO is found, and other characteristics such as the method of carbon assimilation, methanotrophs are divided into type I and type II. The ICM of type I organisms resembles disc-shaped vesicles and can be present all over the cell, whereas in type II

Received: October 18, 2022
Published: January 23, 2023

https://doi.org/10.1021/acssynbio.2c00554
ACS Synth. Biol. 2023, 12, 544−554

methanotrophs, the ICM appears as paired membranes assembled at the cell periphery usually running around the whole cell. Type I strains utilize the ribulose monophosphate pathway for carbon assimilation as opposed to the serine cycle of type II methanotrophs.[2]

Some genetic tools have been developed for both type I and II methanotrophs. However, few are available for rational genome editing. Examples include marker-exchange mutagenesis,[12] the *cre-lox* system,[13] and the *sacB*-based deletion system.[14] Marker-exchange mutagenesis has been successfully demonstrated in *Methylocella silvestris*[12] and reported in *Methylosinus trichosporium*[15] for gene deletion (*mmoX*). This method, however, has a number of disadvantages, including (1) a relatively high risk of introducing unwanted insertions or deletions of genetic elements due to the large genetic changes that are made[16] and (2) a low frequency of gene replacement;[12] and (3) once inserted, markers cannot be reused. The *cre-lox* system used in *M. silvestris* lacks consistency in terms of the rate at which double-cross over events are achieved (ranging from 5 to 50%)[13,16] and leaves a recombinase recognition site scar,[17] which can potentially affect phenotype and can be responsible for recombination events where multiple genes are deleted. The *sacB*-based deletion system, as with the *cre-lox* system, makes marker-less genomic edits[18] and is an easily applicable system designed for use in a wide range of organisms. As with all counter-selections systems, however, there is a high rate of false positives due to spontaneous sucrose-resistant colonies. Moreover, as the method requires the initial isolation of single cross-over integrants, followed by the subsequent isolation of double cross-over mutants, the two-step process is time-consuming when manipulating slow growing methanotrophs.

More recently, genome editing based on Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) has been developed as a quick and easy-to-design tool for a wide variety of living organisms.[19−21] Its deployment in methanotrophs would potentially increase mutagenesis efficiency, leave no markers or scars, and decrease the time needed to generate the required edits, an important consideration in slow growing methanotrophs. Genome editing is achieved in CRISPR/Cas9 systems when DNA strand breaks induced by Cas proteins are repaired using repair templates (homology-directed repair—HDR) or using error-prone template-free cellular machinery (non-homologous end joining—NHEJ).[22] Although genome editing using CRISPR has been demonstrated once in *Methylococcus capsulatus* Bath, the CRISPR system used was a $Cas9^{D10A}$ nickase, the efficiency achieved was extremely low (2%), and no clean deletion of the targeted gene was demonstrated.[23]

Here, we demonstrate the deployment of a highly effective, *Streptococcus pyogenes*-based CRISPR/Cas9 system that generates scarless gene deletions and/or insertions in both a type I (*M. capsulatus* Bath) and type II (*M. parvus* OBBP) methanotroph. Two methods of gene insertion were developed and exemplified by genomic insertion of the eYFP reporter gene into the *ligD* of *M. parvus* OBBP, and one method involving eYFP reporter gene insertion (gene replacement) in MCA_0145 gene of *M. capsulatus* Bath. This work also provides a Tn*5* transposon-based strategy for identifying the non-essential genes to be targeted when first implementing genome editing tools, such as CRISPR/Cas9 in a microbe. The importance of identifying essential genes, though often overlooked, can be a deciding factor in the successful implementation of any new genome editing system as demonstrated in this study.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions. Details of all bacterial strains used in this study are in Table SI. *Escherichia coli* XL-1 blue was used for routine cloning, *E. coli* DH5α was used for HiFi assembly, and *E. coli* S17-1 λ*pir* was used for biparental conjugation. Unless when making chemically competent cells, *E. coli* strains were grown at 37 °C in Lysogeny broth (LB) media supplemented with kanamycin 50 μg/mL, while shaking at 200 rpm. Methanotrophs were grown in nitrate mineral salt (NMS) media supplemented with 10 μM $CuSO_4{\cdot}7H_2O$.[24] Shaking was at 200 rpm for liquid cultures. *M. parvus* OBBP was incubated at 30 °C, while *M. capsulatus* Bath was incubated at 37 °C or 45 °C both with a 1:5 CH_4/air mixture when grown in serum bottles. When methanotrophs were grown with solid media on plates, CH_4 was gassed into anaerobic jars. During conjugation, kanamycin was used at 15 μg/mL for *M. capsulatus* Bath and 50 μg/mL for *M. parvus* OBBP. Nalidixic acid at 25 μg/mL was used for counter-selection against *E. coli* S17-1 λ*pir* during conjugation.

Promoter Strength Assays. The promoters in Table S2 were tested for activity in *E. coli* S17-1 λ*pir*, *M. parvus* OBBP, and *M. capsulatus* Bath. Most of the promoters were derived by cloning DNA regions upstream of the start codon of genes from *M. parvus* OBBP and *M. capsulatus* Bath. DNA regions approximately 300 bp from the start codon of selected genes were analyzed with three promoter prediction algorithms (Promoter Prediction by Neural Network, BPROM and PePPER)[25] and the DNA sequences predicted to have promoters by at least two of the software packages were sent to Twist Bioscience for synthesis. After synthesis, each promoter was cloned upstream of a reporter gene encoding enhanced yellow fluorescent protein (eYFP) and transformed into *E. coli* S17-1 λ*pir*. Details of the cloning and plasmid design can be found in Supporting Information.

The various plasmids made were conjugated into *M. parvus* OBBP and *M. capsulatus* Bath as described below; the level of eYFP expression was estimated using fluorescence assays. Actively growing cultures were diluted to approximately OD_{600} 0.5. Quadruplicates of each were pipetted (100 μL) into Black Greiner 96 Well Flat Bottom (Chimney well) plates. Culture harbouring the control plasmid (plasmid with a promoter-less eYFP gene) was pipetted in wells (100 μL) in 12 replicates. The media was also pipetted (100 μL) into 12 different wells. Fluorescence was measured with a Tecan M1000 using the following parameters: excitation wavelength—495 nm; emission wavelength—530 nm; optical density wavelength—600 nm; mode—read from the bottom; gain—70; shaking frequency—408 rpm; kinetic cycles—5. Promoter activity was measured as fluorescence/OD, normalizing for fluorescence from media and the promoter-less eYFP reporter gene.

Plasmid Design and Construction. Oligonucleotide primers used for cloning and sgRNA were synthesized by Merck UK and Eurofins in Germany and are listed in Table S3. Detailed plasmid construction is described in Supporting Information, and the plasmids used in this study are listed in Table S4. Sanger sequencing was carried out by Source Bioscience Nottingham.

Conjugation of Methanotrophs and Mutant Screening. Conjugation was carried out based on modifications of the method used by Martin and Murrell.[26] *E. coli* S17-1 λ*pir*

https://doi.org/10.1021/acssynbio.2c00554
ACS Synth. Biol. 2023, 12, 544−554

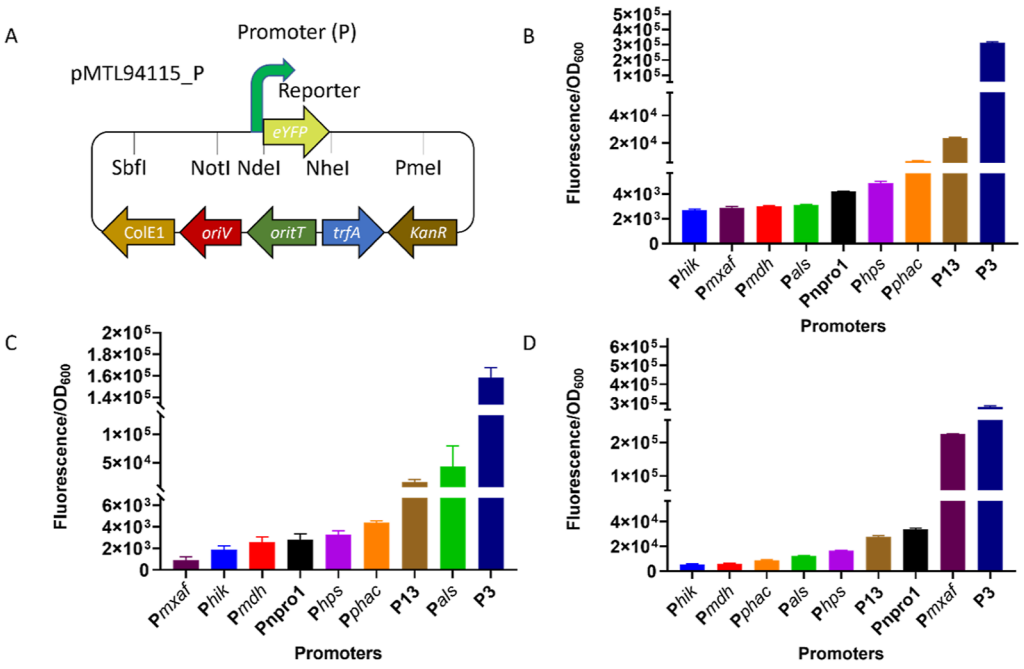


Figure 1. Fluorescence assay of promoters. (A) pMTL94115_P plasmid used for promoter assay. Fluorescence assay of promoters in (B) *E. coli* S17-1 *λpir*; (C) *M. parvus* OBBP; (D) *M. capsulatus* Bath. All promoters are expressed constitutively using eYFP as the reporter. Fluorescence values for promoterless eYFP was used to normalize all promoter-eYFP data.

harboring the plasmid intended for conjugation was grown overnight in LB media containing kanamycin (50 μg/mL). The absorbance (OD_{600}) of the grown overnight culture was measured, and the volume required to give 1 mL of *E. coli* S17-1 *λpir* at OD_{600} of 1 was calculated and pipetted into an Eppendorf tube. For example, 0.4 mL of OD_{600} 2.5 gave 1 mL of OD_{600} 1. The *E. coli* S17-1 *λpir* culture was then washed three times with NMS media to remove the antibiotics by centrifuging at 8000 rpm for 3 min. After the third wash, a calculated volume of methanotroph culture was mixed to have 1:1 donor/recipient ratio by OD. The mixture of *E. coli* S17-1 *λpir* and methanotroph was spun at 8000 rpm for 3 min and resuspended in 50 μL of NMS, which was spotted on a dry plate of NMS Bacto Agar containing 0.5% yeast extract. The spot was allowed to dry, and the plates were placed in air-tight anaerobic jars. The jars were sparged for 3−5 s with CH_4 and incubated at 30 °C for *M. parvus* OBBP and 37 °C for *M. capsulatus* Bath typically for 48 h unless otherwise stated. After 48 h of mating, the spot was scraped with a sterile loop and resuspended in 1 mL of NMS. The resuspension was diluted to 10^{-7} in NMS media. In triplicate, 10 μL of each dilution including neat was plated in a sector of LB agar and NMS Bacto agar plate divided into eight sectors with the required antibiotics. NMS Bacto agar plates were incubated for 2 weeks while LB agar plates were incubated overnight. After incubation, colonies that appeared on plates were counted. Conjugation efficiency (CE) was calculated as the number of transconjugants per donor cell. In other words, the number of methanotrophs that grew in a particular dilution in NMS Bacto agar plates with antibiotics per number of *E. coli* S17-1 *λpir* that grew in the same dilution on LB media plate with antibiotics.[27]

After conjugating CRISPR/Cas9 plasmids into methanotrophs, to identify mutants generated by our CRISPR/Cas9 system, transconjugant colonies growing on NMS antibiotic plates were picked, patch-plated on NMS antibiotic plates, and dipped into a 25 μL PCR mix according to NEB UK Q5 high-fidelity polymerase instructions, using specific flanking primers (Table S3). In all cases, mutations were further confirmed with Sanger sequencing.

Transposon Mutagenesis. To determine non-essential genes, transposon mutagenesis was used to generate mutant cells with transposon insertion in non-essential genes and gene regions. This was carried out by conjugation of *M. parvus* BRCS2 and *M. capsulatus* Bath with pMTL90531_Tn5 plasmid using *E. coli* S17-1 *λpir*. The location of transposon insertion and subsequently designated non-essential gene was determined by inverse PCR and Sanger sequencing. Cloning pMTL90531_Tn5, conjugation of the plasmid, and inverse PCR are described in detail in Supporting Information.

Plasmid Curing. During screening of colonies for mutants, each colony was patch-plated on NMS antibiotic plates before screening with PCR. After confirming genome editing in colonies using PCR, the patch plates were revisited and re-streaked on antibiotic-free NMS plates. After growth, a colony is transferred to NMS liquid media and allowed to grow to exponential phase. The culture is then diluted up to 10^{-6}. Dilutions are then spotted on NMS plates in replicates. After growth on NMS plates, 25 colonies are scooped and dipped in PCR tubes containing 50 μL of NMS media. From this mixture, 5 μL is spotted on both NMS and NMS antibiotic plates. The presence of growth on NMS plates and absence of growth on NMS antibiotic plates suggest that the plasmid has been cured. The colony confirmed with plasmid cured cells is then grown in liquid NMS media and cryo-stocked. Example plates are shown in Figure S2.

RESULTS AND DISCUSSION

Promoter Strength Assays. Although previous studies have evaluated the performance of the *mxaF* promoter,[23,28] the focus here was on testing uncharacterized promoters in type I (*M. capsulatus* Bath) and type II (*M. parvus* OBBP)

https://doi.org/10.1021/acssynbio.2c00554
ACS Synth. Biol. 2023, 12, 544−554

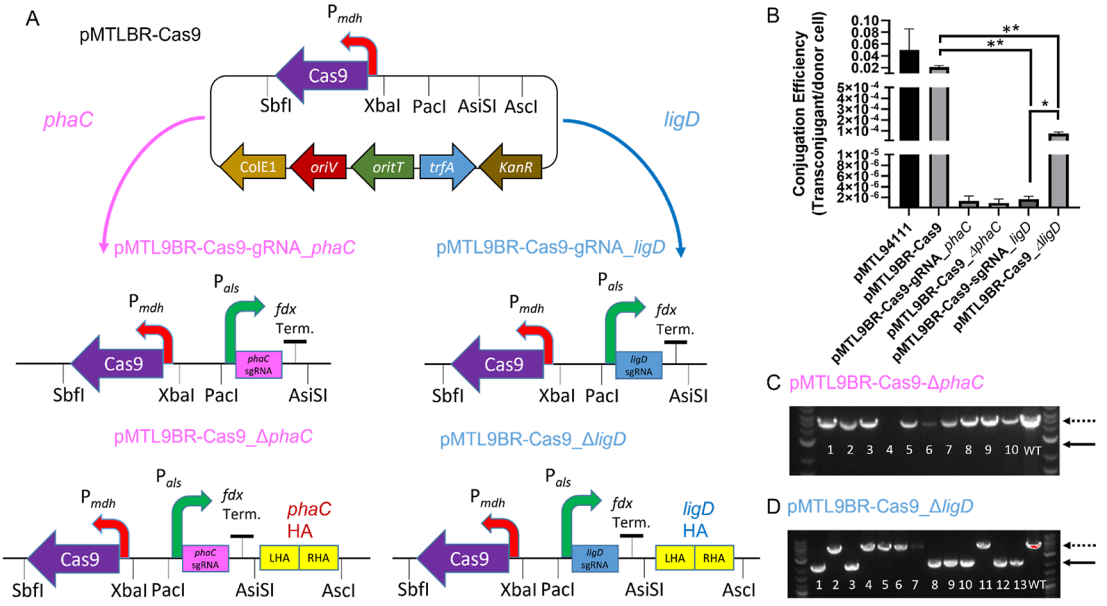


Figure 2. Plasmids, CE, and genome editing screens of *M. parvus* OBBP. (A) Plasmid pMTL9BR2-Cas9 carrying *S. pyogenes cas9* gene cloned into pMTL94111 backbone and subsequent plasmids designed from pMTL9BR2-Cas9. (B) CE graph showing significant difference between pMTL9BR2-Cas9 when compared to pMTL9BR2-Cas9-gRNA_*phaC* or pMTL9BR2-Cas9_Δ*ligD* ($P = 0.0088$). pMTL9BR2-Cas9-gRNA_*ligD* CE was significantly lower than pMTL9BR2-Cas9_Δ*ligD* ($P = 0.02$). There was no significant difference between pMTL94111 and pMTL9BR2-Cas9 ($P = 0.3681$). Unpaired *t*-test used, $n = 2$, error bars represent the standard error of the mean. (C) Bands of PCR screens of unsuccessful *phaC* gene deletion. *phaC* is a 2070 bp gene, and edited colonies were expected to have bands approximately 2070 bp smaller than the control (WT), which is in the last lane. (D) Bands of PCR screen of successful *ligD* gene deletion. *ligD* is 2448 bp in size, and edited colonies were expected to have bands approximately 2448 bp smaller than the control (*M. parvus* OBBP WT gDNA), which is in the last lane. The numbers of colonies screened were represented by numbers. NEB 1 kb Plus DNA ladder was used. Dotted lines represent the size of WT PCR amplicon, while solid lines represent mutants.

methanotrophs. These comprised synthetic and constitutive promoters native to organisms listed in Table S2. Their activity was tested in *E. coli* S17-1 *λpir*, *M. parvus* OBBP, and *M. capsulatus* Bath using a reporter gene encoding eYFP. pMTL94115 modular plasmid which has eYFP inserted in the multiple cloning site was used as the backbone to give pMTL94115_P where P stands for the specific promoter used as shown in Figure 1A. Relatively speaking, the promoters P_{als} and P3 led to the moderate and high expression, respectively, of eYFP in *M. parvus* OBBP. Only a low level of eYFP expression occurred from the remaining promoters. In *M. capsulatus* Bath, the use of P_{hps} and P_{mxaf} led to moderate expression of eYFP compared to the high level of expression observed from P3. In *E. coli* S17-1 *λpir*, the P3 promoter also led to the highest level of eYFP expression (Figure 1B–D). The relative strength of the promoters in the three hosts was subsequently taken into account in selecting the promoter to be used to express *cas9* and sgRNA in the developed CRISPR-based gene editing system.

It was important to assess the expression levels of promoters in *E. coli* S17-1 *λpir* to avoid overexpression of *cas9* during cloning, as this could lead to reduced transformation efficiency as reported in *E. coli* ER1821.[29] Hence, promoters that weakly express *cas9* and express guide RNAs at relatively stronger levels in methanotrophs were selected for use, and those such as P3, shown to direct very high level of expression in both *E. coli* and methanotrophs, were avoided. The weakest promoters were also not selected for use to avoid expression of *cas9* below levels required for efficient double strand break. Promoter expression levels proved to be different for the two methanotroph species under investigation. The selected promoters were expressed constitutively, which contrasts with many other systems that place *cas9* under inducible control.[30−32] However, our study clearly demonstrates that constitutive expression of *cas9* can be just as effective for successful genome editing so long as the level of *Cas9* expression is low in the cloning host to avoid mutations and DNA rearrangement of the plasmid during cloning. Nonetheless, inducible promoter-controlled *cas9* is advantageous in organisms with relatively low DNA transfer efficiencies, weak recombinases, and organisms that are susceptible to Cas9 toxicity.[33] For example, in a study involving *Clostridium acetobutylicum*, no colonies were observed after transformation when *cas9* was under the transcriptional regulation of constitutive promoters, whereas efficient editing was observed when *cas9* was regulated with inducible promoters.[30] This supports the importance of CRISPR system designs that incorporate inducible promoters such as those seen in the methanotroph study referred to earlier.[23] Further experiments in this study were designed to investigate the toxicity of Cas9 in methanotrophs and CE of *Cas9* plasmids. It was decided to use promoter P_{phaC} to express *cas9* gene and promoter P_{mxaF} to express guide RNA in type I methanotroph *M. capsulatus* Bath. In type II methanotroph *M. parvus* OBBP, promoter P_{mdh} was used to express *cas9* gene and promoter P_{als} to express sgRNA.

Establishment of a CRISPR/Cas9 Genome Editing Plasmid System in *M. parvus* OBBP. CRISPR/Cas9 plasmid systems were designed in stages that will enable the functional study of various components such as *cas9*, guide RNAs, and repair templates in the form of homology arms. To investigate the specific effect of Cas9, for instance its toxicity in *M. parvus* OBBP, pMTL9BR2-Cas9 was constructed. In this plasmid, *cas9* was under the transcriptional regulation of a methanol dehydrogenase promoter P_{mdh}. Second, the addi-

https://doi.org/10.1021/acssynbio.2c00554
ACS Synth. Biol. 2023, 12, 544−554

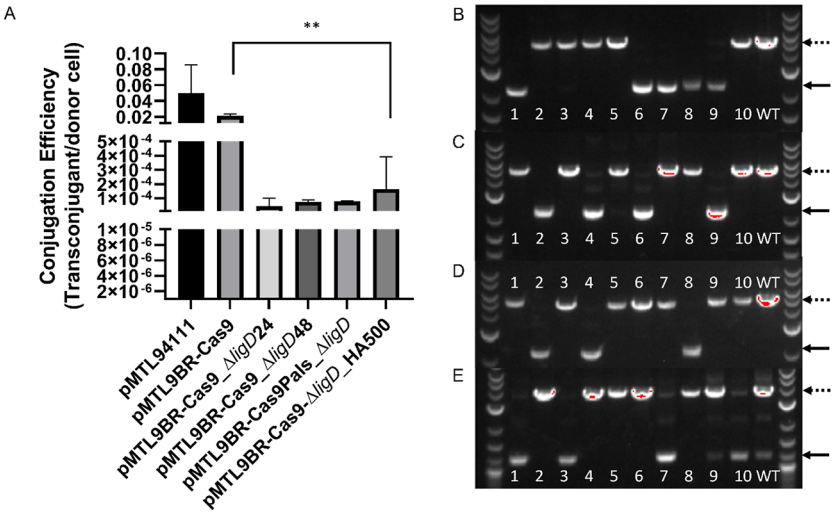


Figure 3. Δ*ligD* gene deletion under different conditions (A) CE of additional *ligD* deletions in *M. parvus* OBBP. (B) *M. parvus* OBBP *ligD* gene deletion using pMTL9BR2-Cas9_Δ*ligD*24 plasmid conjugated with *E. coli* (24 h mating time). (C) *M. parvus* OBBP *ligD* gene deletion using pMTL9BR2-Cas9_Δ*ligD*48 plasmid conjugated with *E. coli* (48 h mating time). (D) *M. parvus* OBBP *ligD* gene deletion using pMTL9BR2-Cas9Pals_Δ*ligD* plasmid with a relatively stronger Cas9 promoter (P_{als}) compared to sgRNA promoter (P_{mdh}) conjugated with *E. coli*. (E) *M. parvus* OBBP *ligD* gene deletion using pMTL9BR2-Cas9-Δ*ligD*_HA500 plasmid with 500 bp HA conjugated with *E. coli*. NEB 1 kb Plus DNA ladder was used. Dotted lines represent the size of WT PCR amplicon, while solid lines represent mutants. The number of colonies screened were represented by numbers.

tional effect of guide RNA was investigated by cloning into pMTL9BR2-Cas9 a guide RNA with spacer targeting either *phaC* or *ligD* genes, resulting in plasmids pMTL9BR2-Cas9-sgRNA-*phac* and pMTL9BR2-Cas9-sgRNA-*ligD*, respectively. In both cases, the sgRNAs were driven by acetolactate synthase promoter P_{als}. Finally, complete CRISPR/Cas9 plasmids pMTL9BR2-Cas9_Δ*phaC* and pMTL9BR2-Cas9_Δ*ligD*, which consist of *cas9*, sgRNA, and homology arms, were designed for Cas9-induced double strand break with the targeting effect of sgRNA and the repair function of homology arms (Figure 2A). The combined action of these three components was expected to result in genome editing. Details of cloning can be found in Supporting Information.

Initial *M. parvus* OBBP Deletion Target, *phaC*. The first *M. parvus* OBBP gene to be targeted, *phaC*, encoded poly(*R*)hydroxybutyrate (PHB) synthase. Its selection was based on the fact that it has been deleted in other PHB-producing species such as *Cupriavidus necator*,[34] and it represents an important biotechnological target since its deletion would potentially allow carbon flux to be re-directed from PHB into alternative biosynthetic products. Using the promoters selected above and the modular plasmid pMTL94111, an appropriate CRISPR/Cas9 plasmid targeting *phaC* was constructed expressing *cas9* and sgRNA against *phaC* from the P_{mdh} and P_{als} promoters, respectively, together with a repair template comprising 1000 bp left and right homology arms flanking the intended 2070 bp *phaC* gene (pMTL9BR2-Cas9_Δ*phaC*) (Figure 2A). Control plasmids were also constructed lacking the repair template (pMTL9BR2-Cas9-sgRNA-*phaC*) or repair template and sgRNA (pMTL9BR2-Cas9). All three plasmids, together with the pMTL94111 progenitor, were conjugated into *M. parvus* OBBP and their CE estimated from the number of colonies obtained (Figure 2B). Plasmid pMTL9BR2-Cas9 transformed at the same high frequency as its progenitor, pMTL94111 (Figure 2B), suggesting that expression of Cas9 in the absence of a sgRNA is not toxic in *M. parvus* OBBP. This is in contrast to other microorganisms where expression of Cas9 can reduce transformation frequencies by two orders of magnitude.[29] Although the non-toxic effect observed could indicate that Cas9 was not being produced, nevertheless, the plasmid carrying a combination of *cas9* and sgRNA targeting *phaC* (pMTL9BR2-Cas9-sgRNA-*phaC*) exhibited a significant reduction in CE (Figure 2B), consistent with the production of a functional Cas9-sgRNA complex and a reduction in cell numbers through double strand cleavage of the chromosome.[30] An equivalent low frequency was seen with pMTL9BR2-Cas9_Δ*phaC*, suggesting that the addition of homology arms to the pMTL9BR2-Cas9-sgRNA-*phaC* did not increase the chances of transconjugants surviving. Colonies were screened to establish whether *phaC* deletion mutants had been generated using genome-specific primers flanking the repair template homology arms. No mutants were detected in three independent experimental attempts to delete *phaC* (Figure 2C). Failure to isolate *phaC* deletion mutants was suspected to be due to low homologous recombination (HR) efficiency in *M. parvus* OBBP. Past studies have reported that HR efficiency was increased by inhibiting or knocking out NHEJ genes such as the ATP-dependent DNA ligase *ligD*.[35,36] A recent study reported 22% of bacteria carry this pathway.[37] In common with 22% of bacteria, *M. parvus* OBBP appears to carry this pathway.[27] Attempts were, therefore, made to delete *ligD*.

***ligD* Gene Deletion in *M. parvus* OBBP.** To target *ligD*, equivalent plasmids to those made for *phaC* were constructed (Figure 2A) and conjugated in *M. parvus* OBBP from an *E. coli* S17-1 *λpir* donor. After 2 weeks of incubation at 30 °C, colonies were counted, and the CE was estimated. Similar to the observation made with the *phaC* specific constructs, the combined presence of Cas9 and sgRNA directed against *ligD* significantly reduced the transformation frequency of the plasmid concerned, pMTL9BR2-Cas9-sgRNA-*ligD*, compared to pMTL9BR2-Cas9. This observation further supports our contention that the designed system is able to produce a functional Cas9/sgRNA complex. However, in the case of

https://doi.org/10.1021/acssynbio.2c00554
ACS Synth. Biol. 2023, 12, 544−554

Table 1. Effect of gRNA Spacer Sequence and PAM on Success of CRISPR/Cas9 Gene Deletion in *M. parvus* OBBP[a]

name	seed code	GC (%)	on target score	off target score	strand	PAM	dist. along gene/gene length in bp (%)	actual dist. from end (bp)	genome editing eff. (%)
ligD Seed	GGAAGCGGGCTGTCCAATCG	65	64.3	100	+ve	agg	1616/2488 (65)	872	58
AltKO_*ligD*Seed1	CGAGAGGATGGTCTTCCGTG	60	73.4	100	+ve	cgg	2183/2488 (88)	305	50
AltKO_*ligD*Seed2	ATGGTCGCGAATTTCCCCCG	60	71.8	100	+ve	cgg	438/2488 (18)	438	50
AltKO_*ligD*Seed3	CATCACCCATGCAAGCCGGG	65	68.1	100	−ve	tgg	827/2488 (33)	827	90
AltKO_*ligD*Seed4	GCGCCATATAAAGTTCGTCG	50	71	100	+ve	tgg	614/2488 (25)	614	80
AltKO_*ligD*Seed5	TTTCAGCTCGAAGAGCCAAT	45	64.2	100	+ve	cgg	1704/2488 (68)	784	60
AltKO_*ligD*Seed6	GATCAAGGGCGACTTTCGAG	55	66.5	100	−ve	agg	1211/2488 (49)	1211	60

[a]Percent knockouts achieved from 10 sampled colony PCR's except *ligD*_Seed which was from 12.

targeting *ligD*, the presence of a repair template for *ligD* on plasmid pMTL9BR2-Cas9_Δ*ligD*, together with *cas9* and sgRNA targeting *ligD*, significantly increased the CE, suggesting that the addition of homology arms to the plasmid increased the chances of transconjugants surviving by enabling removal of the PAM target through allelic exchange (Figure 2B,C). Evidence that this had occurred was provided by PCR screening of the colonies and the demonstration of DNA bands on agarose gels of a size consistent with the intended deletion in the *ligD* gene (Figure 2D). Gel extraction of these DNA fragments and their analysis by Sanger sequencing confirm that they were *ligD* gene deletions. This represents the first reported example of wild type *S. pyogenes* CRISPR/Cas9-mediated scarless genome editing in methanotrophs. The editing efficiency was 58%. Having demonstrated that the system was working, the effect of various parameters was investigated with a view to optimizing the process. These included mating times during conjugation, sgRNA spacer target score, promoter strength, and the length of homology arms used in the repair template. A reduction in the size of the homology arms used from 1000 to 500 bp in plasmid pMTL9BR2-Cas9-Δ*ligD*_HA500 appeared to have little effect on editing frequencies (Figure 3E) and gave the highest CE ($P = 0.089$) among the CRISPR/Cas9 plasmids tested (Figure 3A). Similar, if not slightly better, editing efficiency was demonstrated when mating times were reduced from the routinely used 48 to 24 h (Figure 3B,C). This reduction could represent a significant saving in the time needed to generate mutants. This contrasts with studies in *Bacillus subtilis* where additional incubation of the transformation mixture led to an increase in editing efficiency from 16 to 80%.[38] Genome editing was also successful when a promoter of moderate strength was used for *cas9* (P_{als}) expression and a relatively weaker promoter for gRNA (P_{mdh}) (Figure 3D). This provides more flexibility when cloning CRISPR/Cas9 plasmids for genome editing. The ability of *M. parvus* OBBP to withstand moderate expression of Cas9 further supports its tolerance to Cas9 toxicity unlike in other organisms such as *Clostridium* species.[29] In these experiments, genome editing efficiency varied from 30 to 50%, as shown in Figure 3.

Further experiments to understand the effect of different sgRNAs on genome editing showed that six additional sgRNAs successfully mediated the deletion of *ligD* regardless of the on-target score, genomic DNA strand, PAM, and position of spacer sequence along *ligD* gene (Table 1). This allows for a high level of flexibility in selecting the sgRNA to be used. With the sgRNA AltKO_ligDSeed3, up to 90% genome editing efficiency was achieved. Overall, of the parameters investigated, changing the sgRNA gave the highest variation in genome editing efficiency (10−90%) and likely represents the most important starting point when trying to optimize genome editing efficiency.

Other Gene Targets for CRISPR/Cas9 Genome Editing in *M. parvus* OBBP. Having generated a *ligD* mutant, additional genes were targeted, including *glg* (a gene involved in glycogen metabolism), *copD* (a copper tolerance gene), and *ligA* (an ADP-dependent DNA ligase). No evidence of gene deletion was obtained. In addition to low HR efficiency earlier suspected when *phaC* failed to be deleted, another potential reason to explain the failure to delete a particular target was that the gene was essential. Non-essential genes may be potentially identified through their inactivation using a transposon. Accordingly, a Tn*5*-based transposon plasmid pMTL90531_Tn5 was assembled and conjugated in *M. parvus* BRCS2 to create random transposon mutants (Figure S1). This particular methanotroph was chosen, because unlike *M. parvus* OBBP whose genome is only in draft form, the *M. parvus* BRCS2 genome is complete.[27,39] The genomes of the two strains in any case share 99.99% homology. After screening a number of colonies with inverse PCR, 37 transposon insertion sites within *M. parvus* BRCS2 genome were identified, and the genes affected were identified by nucleotide sequencing. This led to the identification of three non-essential genes, which were identical in the DNA sequence to the type strain *M. parvus* OBBP, namely, *pntA* (an NAD(P) trans-hydrogenase subunit), MPA_0518 (a hypothetical protein), and *bcsB* (a cyclic di-GMP binding protein precursor). Equivalent CRISPR/Cas9 plasmids to those used to knock out *ligD* were constructed and shown to generate mutants in the three genes of *M. parvus* OBBP with 30−40% genome editing efficiency (Figure 4A−C). The colonies with edited genomes were re-streaked on kanamycin plates and subsequently grown in kanamycin-free media to start a plasmid curing process. Up to 100% plasmid curing rate was obtained based on 25 colonies for *bcsB* gene deletion. As with *ligD*, further experiments using eight additional sgRNA spacers were carried out to investigate gene deletion success in these new targets. The use of six of the eight alternative sgRNAs led to successful gene deletion (Table S5).

This finding emphasizes the need to choose gene targets for the exemplification of new genome editing systems that are proven to be non-essential as opposed to hypothesized to be based on indirect findings, for example, *phaC* is shown to be dispensable in other organisms such as *C. necator*.[34] Hence, our strategy of using transposons to identify non-essential genes provided ideal targets to validate our genome editing system and would make a significant difference whenever such systems are being developed in new and understudied organisms. Once

https://doi.org/10.1021/acssynbio.2c00554
ACS Synth. Biol. 2023, 12, 544−554

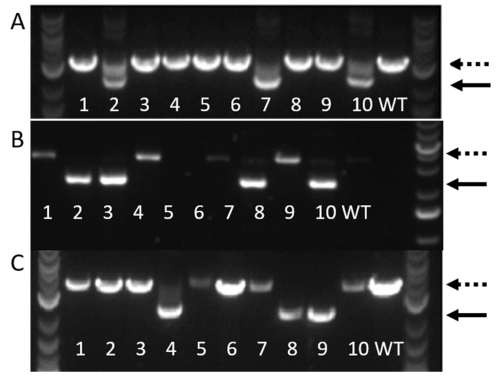


Figure 4. Deletion of additional genes in *M. parvus* OBBP. (A) *M. parvus* OBBP gene deletion using pMTL9BR2-Cas9_Δ*pntA* plasmid. (B) *M. parvus* OBBP gene deletion using pMTL9BR2-Cas9_Δ*bcsB* plasmid. (C) *M. parvus* OBBP gene deletion using pMTL9BR2-Cas9_ΔMPA_0518 plasmid. NEB 1 kb Plus DNA ladder was used. Dotted lines represent size of WT PCR amplicon, while solid lines represent mutants. Lanes 1−10 represent the 10 colonies screened.

non-essential genes were identified, the deletion of 4 genes using 16 different sgRNAs was easily achieved.

Effect of *M. parvus* OBBP ATP-Dependent DNA Ligase (*ligD*) on Genome Editing Efficiency. The successful knock-out of *ligD* may have been a consequence of the ablation of NHEJ, which compelled the cell to repair its genome via HDR. *ligD* inhibition has been shown to increase the efficiency of HR in rice[35] and to increase genome editing efficiency in mammalian cells and mice.[36] Therefore, it was hypothesized that deletion of *phaC* could succeed if the function of *ligD* was compromised, leading to an increased HR and genome editing efficiency. To test this, *phaC* gene deletion gene was pursued in a Δ*ligD*$^{-}$ mutant strain. In a second experiment, *phaC* was targeted in the presence of the *ligD* inhibitor SCR7 added to the media during conjugation. No gene deletion was observed in either case. This outcome strongly suggests the essentiality of *phaC* and casts doubts to any significant role played by *ligD* with regard to HR and gene editing efficiency.

The effect on the editing efficiency of a gene that could be knocked out when performed in a Δ*ligD* background was further explored. Accordingly, the efficiency of gene editing of MPA_0518 in a wild type and Δ*ligD* background was compared. MPA_0518, a gene coding for a hypothetical protein, was already knocked out in this study. The deletion efficiency was higher in the wild type (61%) compared to the Δ*ligD* mutant strain (50%), suggesting that the *ligD* deletion does not offer genome editing advantages (Figure S4).

Gene Insertion in *M. parvus* OBBP. After the successful deletion of multiple gene targets in *M. parvus* OBBP, CRISPR/Cas9 plasmids were designed to insert DNA cargo into *M. parvus* OBBP genome. A gene encoding eYFP under the transcriptional regulation of a strong promoter (P3) was chosen for the proof-of-concept experiments. Two approaches were investigated. In the first, gene replacement approach, the eYFP gene was introduced concomitantly with the deletion of *ligD* (Figure 5A), through its insertion between two 1000 bp homology arms flanking the *ligD* gene. This method was demonstrated in *Corynebacterium glutamicum,* where *ldhA* was replaced with a *rfp* cassette.[40] In the second approach, the region between *ligD* sgRNA spacer and PAM was targeted for insertion. In this design, the left homology arm extended 1000 bp all the way to the end of the gRNA spacer, while the right homology arm started from the PAM extending 1000 bp in the 5′−3′ direction. The inserted eYFP gene was positioned between the two homology arms (Figure 5B). This is a slight modification of the method used in *Clostridioides difficile* where

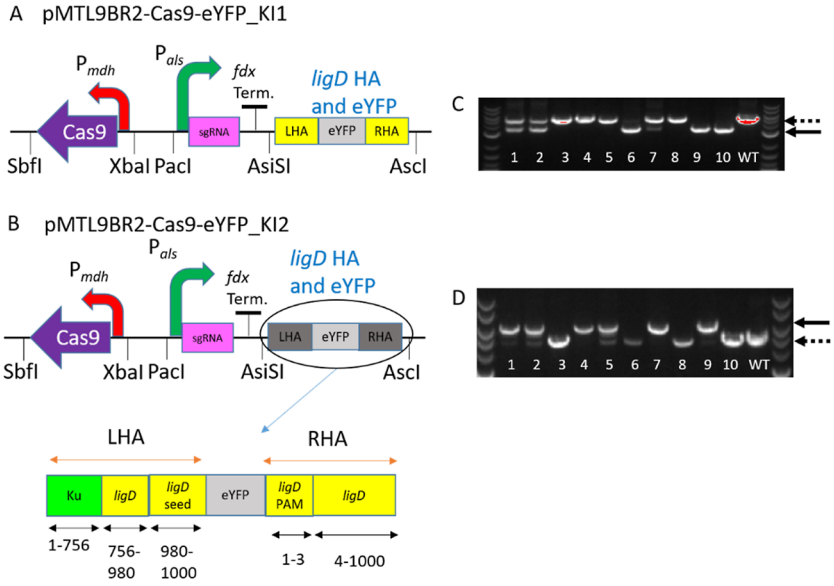


Figure 5. Gene insertion in *M. parvus* OBBP. (A) plasmid pMTL9BR2-Cas9-eYFP_KI1 used for inserting eYFP gene and simultaneous replacement of *ligD* gene. (B) Plasmid pMTL9BR2-Cas9-eYFP_KI2 used for insertion of eYFP gene without gene replacement. (C) eYFP and its P3 promoter are both 1091 bp long. Using pMTL9BR2-Cas9-eYFP_KI1 to Replace *ligD* with eYFP will give a band around 1357 bp smaller than the control which is *M. parvus* OBBP WT gDNA. (D) pMTL9BR2-Cas9-eYFP_KI2 colonies with successful gene insertions have bands approximately 1091 bp higher than the control, which is *M. parvus* OBBP WT gDNA. The same primer set was used for screening in (C,D). NEB 1 kb Plus DNA ladder was used. Dotted lines represent size of WT PCR amplicon, while solid lines represent mutants. Lanes 1−10 represent the 10 colonies screened.

https://doi.org/10.1021/acssynbio.2c00554
ACS Synth. Biol. 2023, 12, 544−554

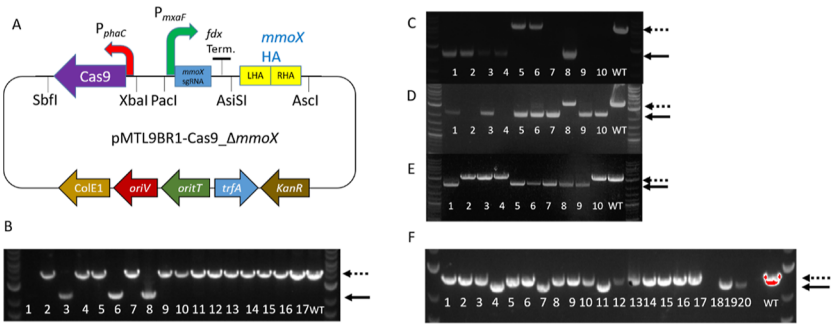


Figure 6. Genome editing in *M. capsulatus* Bath. (A) *M. capsulatus* Bath *mmoX* gene deletion plasmid pMTL9BR1-Cas9_Δ*mmoX*. (B) *M. capsulatus* Bath *mmoX* gene deletion PCR screen. (C) *M. capsulatus* Bath gene deletion using pMTL9BR1-Cas9_Δ*czcA* plasmid. (D) *M. capsulatus* Bath gene deletion using pMTL9BR1-Cas9_ΔMCA_0145 plasmid. (E) *M. capsulatus* Bath gene deletion using pMTL9BR1-Cas9_ΔMCA_2158 plasmid. (F) Gene insertion of eYFP with simultaneous gene deletion of MCA_0145 using plasmid pMTL9BR1-Cas9_ΔMceYFPKI1. NEB 1 kb Plus DNA ladder was used. Dotted lines represent the size of WT PCR amplicon, while solid lines represent mutants.

an anaerobic active green fluorescent protein was inserted between the nucleotide 101 and 102 downstream from a gene terminator.[41] Both methods led to the successful insertion of the eYFP gene with 60% efficiency, as confirmed by PCR and Sanger sequencing of the amplified product.

Gene Deletion in *M. capsulatus* Bath. To test the versatility of the developed method to other methanotrophs, *M. capsulatus* Bath was targeted. The two targets selected were *mmoX* and *ligA*. The former was considered non-essential as it has been disrupted previously,[23] while the latter was suggested to be involved in a possible NHEJ system referred to as alternative end-joining. Inactivation of *ligA* could potentially increase the efficiency of genome editing.[42] Though *ligA* is an NAD-dependent DNA ligase, unlike *ligD* which is ATP-dependent and proven to increase genome editing efficiency when inhibited in mammalian cell lines and plants, as earlier reported.[35,36] For *M. capsulatus* Bath the CRISPR/Cas9 plasmids constructed used the P_{phac} and P_{mxaf} promoters to control the expression of *cas9* and sgRNA, respectively (Supporting Information). The *mmoX*-specific CRISPR/Cas9 plasmid (pMTL9BR1-Cas9_Δ*mmoX*) produced gene deletions with 19% efficiency (Figure 6B). However, the use of the *ligA*-specific CRISPR/Cas9 plasmid did not result in the isolation of deletion mutants. This is likely due to the essentiality of *ligA*, as is the case in *B. subtilis*.[43] To identify further non-essential gene targets, the transposon approach developed in *M. parvus* OBBP was implemented (Supporting Information). Among the non-essential genes identified in *M. capsulatus* Bath, three were targeted using our CRISPR/Cas9 system. These were genes encoding a porin family membrane protein (MCA_0145), a heavy metal efflux pump (*czcA*), and a transcriptional regulator from the *AraC* family (MCA_2158). Using appropriately constructed CRISPR/Cas9 vectors, deletion mutants of the three genes were isolated with an efficiency of 50, 70, and 60% for MCA_0145, *czcA,* and MCA_2158, respectively (Figure 6C−E). Plasmids of all successful knockouts in *M. capsulatus* Bath were cured using the same method for *M. parvus* OBBP with up to 84% plasmid curing rate based on 25 colonies for *mmoX* gene.

Gene Insertion in *M. capsulatus* Bath. Having demonstrated that the CRISPR/Cas9 system was able to generate deletions, the ability of the system to mediate knock-in was tested essentially as exemplified in *M. parvus* OBBP using the eYFP gene as cargo via the gene replacement approach. In the initial experiments, *mmoX* was targeted, but very few colonies were obtained following conjugation with the requisite CRISPR/Cas9 insertion vector and none carried the desired eYFP insertion. Accordingly, the gene replacement target was changed to the MCA_0145 gene as its deletion efficiency was higher than *mmoX*. Initially, the number of colonies obtained after conjugation remained low (less than 3 colonies per plate) and no eYFP insertion mutants were isolated. To improve the number of colonies obtained after conjugation, a few modifications were made to the conjugation protocol including mating at 30 °C instead of 37 °C, doubling the number of mating and donor cells, increasing the concentration of $CaCl_2$ in the media 10-fold and, finally, halving the antibiotic concentration used to 7.5 and 12.5 μg/mL of kanamycin and nalidixic acid, respectively. Of these modifications, only reducing the antibiotic concentration resulted in a significant increase in the number of colonies, which led to the successful gene insertion of the eYFP gene under the transcriptional control of the P3 promoter. A 25% gene insertion efficiency was obtained (Figure 6F).

So far, our study has demonstrated, for the first time, two methods of gene insertion in *M. parvus* OBBP and one method in *M. capsulatus* Bath. They pave the way for the rapid introduction of exogenous genetic pathways into the genome of methanotrophs, promoting a desirable chassis for biotechnological applications. The genome editing efficiency was also monitored and was usually around 30−50%, although up to 90% genome editing efficiency was attained in some instances. The non-edited colonies in genome editing experiments often referred to as escapers survive for several reasons.[33] Among them are possible *cas9* gene mutation in plasmids found in escaper cells and/or cellular modification of sgRNA target site in CRISPR/Cas9 plasmids or DNA of such cells. We found that CRISPR/Cas9 genome editing in *M. parvus* OBBP was successful regardless of homology arm length, duration of mating in conjugation, and promoter expression levels as long as the promoter controlling *cas9* gene was not highly expressed in *E. coli*. However, the choice of sgRNA is crucial, as demonstrated in two experiments targeting the hypothetical protein MPA_0518 gene; two sgRNAs did not result in gene deletions suggesting that more than one sgRNA should be tested in parallel.

To provide more understanding around why this study led to successful wild type Cas9-initiated HDR-assisted scarless genome editing compared to a previous study,[23] it is important to highlight differences between both studies. The first

https://doi.org/10.1021/acssynbio.2c00554
ACS Synth. Biol. 2023, 12, 544−554

consideration is the methanotrophic bacteria used. In our study, type I and II methanotroph genomes were edited using wild type *S. pyogenes cas9* and HDR was used for DNA repair after Cas9-induced double strand break. In the previous study,[23] genome editing of only the type I *M. capsulatus* Bath was attempted. Wild type *S. pyogenes cas9* did not lead to successful genome editing, and NHEJ was used instead of HDR when success was achieved using the mutated $Cas9^{D10A}$ nickase to achieve genome editing via nonsense mutation.

With regards to plasmid design, different promoter sets were used for genome editing in type II *M. parvus* OBBP involved in our study. For type I *M. capsulatus* Bath plasmid design where there are more similarities in both studies, the first difference observed was the *cas9* promoter used. In our study, the *phaC* promoter from *C. necator* H16 was constitutively expressed. P_{phaC} is a relatively weak promoter in both *E. coli* that was used as a cloning and conjugative host, as well as in *M. capsulatus* Bath which possesses the genome target. In the previous study,[23] the inducible tetracycline promoter (P_{tetA}) was used. P_{tetA} can lead to strong transcription of *cas9* depending on the concentration of the inducer (anhydrotetracycline) added. A high level of *cas9* expressions can lead to plasmid DNA mutation in regions essential for the functioning of the CRISPR-Cas9 systems. Another difference is in the guide RNA spacer used. Whereas our selection of guide RNA spacers was guided by tools provided by Benchling and CRISPy, it is unclear what guided spacer selection in the previous study. As noted in our spacer investigation, guide RNA spacer may determine efficiency and ultimately success of genome editing. Additionally, the guide RNA terminators used in both studies were likely different. Finally, we used a one-plasmid system, whereas a two-plasmid system seemed to have been used in the previous study.[23] One or a combination of the factors highlighted may have been crucial in efficient wild type Cas9-initiated HDR-assisted scarless genome editing.

The demands to extend advanced synthetic biology techniques to more exotic chassis are high, especially for the development of industrial biocatalysts in an array of non-model microbes. In this study, we developed a broad-host-range CRISPR/Cas9 system and demonstrated the efficient genetic manipulation of type I methanotroph *M. capsulatus* Bath as well as type II *M. parvus* OBBP. Furthermore, our challenges of identifying non-essential gene knockout targets highlight the dire need for basic research on bacterial molecular biology in this class of ubiquitous yet understudied organisms. We anticipate that novel molecular mechanisms underlying methanotroph biology will be probed and accelerated with the addition of CRISPR/Cas9 to the limited methanotroph genetic toolbox.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.2c00554.

Plasmid design and cloning; alternative CRISPR knock-outs spacers; strains used in this study; promoters and origin; primers used for designing plasmids and screening gene deletions; plasmids used in this study; transposon mutagenesis plasmid; plasmid curing on NMS agar plates; knock-out success percentages of the 18 tested CRISPR knock out spacer; and gene knock out in *M. parvus* OBBP wild type strain and *M. parvus* OBBP Δ*ligD* strain (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Ying Zhang − *BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Biodiscovery Institute, University of Nottingham, Nottingham NG7 2RD, U.K.*; Phone: +44(0)115 7486119; Email: Ying.Zhang@nottingham.ac.uk

Authors

Bashir L. Rumah − *BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Biodiscovery Institute, University of Nottingham, Nottingham NG7 2RD, U.K.*; orcid.org/0000-0002-5186-7654

Benedict H. Claxton Stevens − *BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Biodiscovery Institute, University of Nottingham, Nottingham NG7 2RD, U.K.*

Jake E. Yeboah − *BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Biodiscovery Institute, University of Nottingham, Nottingham NG7 2RD, U.K.*

Christopher E. Stead − *BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Biodiscovery Institute, University of Nottingham, Nottingham NG7 2RD, U.K.*

Emily L. Harding − *BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Biodiscovery Institute, University of Nottingham, Nottingham NG7 2RD, U.K.*

Nigel P. Minton − *BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Biodiscovery Institute, University of Nottingham, Nottingham NG7 2RD, U.K.*; orcid.org/0000-0002-9277-1261

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.2c00554

Author Contributions

†First author.

Author Contributions

Conceptualization: Y.Z.; Methodology: B.L.R., B.H.C.S., C.E.S.; Investigation: B.L.R., B.H.C.S., C.E.S., J.E.Y., E.L.H., N.P.M.; Visualization: B.L.R., B.H.C.S.; Supervision: Y.Z., N.P.M.; Writing—original draft: B.L.R., Y.Z.; Writing—review & editing: B.L.R., B.H.C.S., Y.Z., N.P.M.

Funding

This work was supported by the UK Biotechnology and Biological Sciences Research Council [grant numbers BB/L013940/1, BB/L013800/1 & BB/S009833/1]—NPM. Petroleum Technology Development Fund Nigeria—BLR. BBSRC iCASE Studentship supported by Calysta (https://www.calysta.com/)—JEY. Nottingham BBSRC DTP [grant number BB/J014508/1]—BCS. "Proof of Concept" funding from the BBSRC NIBBs C1Net and the Carbon Recycling Network—Y.Z.

Notes

The authors declare no competing financial interest.
All data generated or analyzed during this study are included in this published article [and Supporting Information].

■ ACKNOWLEDGMENTS

We thank Drs. Lori Giver and Renee Saville from Calysta for helpful discussions, and all members of SBRC who helped to carry out this research. We would finally like to thank Dr Andrew Crombie (University of East Anglia) for gifting us *Methylococcus capsulatus* Bath.

■ REFERENCES

(1) Rodhe, H. A Comparison of the Contribution of Various Gases to the Greenhouse Effect. *Science* **1990**, *248*, 1217−1219.
(2) Hanson, R. S.; Hanson, T. E. Methanotrophic bacteria. *Microbiol Rev.* **1996**, *60*, 439−471.
(3) Knittel, K.; Boetius, A. Anaerobic oxidation of methane: progress with an unknown process. *Annu. Rev. Microbiol.* **2009**, *63*, 311−334.
(4) Dunfield, P. F. The Soil Methane Sink. *Greenhouse Gas Sinks*; CABI, 2007; p 152.
(5) Jones, S. W.; Karpol, A.; Friedman, S.; Maru, B. T.; Tracy, B. P. Recent advances in single cell protein use as a feed ingredient in aquaculture. *Curr. Opin. Biotechnol.* **2020**, *61*, 189−197.
(6) Lee, E. H.; Yi, T. W.; Moon, K. E.; Park, H. J.; Ryu, H. W.; Cho, K. S. Characterization of methane oxidation by a methanotroph isolated from a landfill cover soil, South Korea. *J. Microbiol. Biotechnol.* **2011**, *21*, 753−756.
(7) Khmelenina, V. N.; Kalyuzhnaya, M. G.; Starostina, N. G.; Suzina, N. E.; Trotsenko, Y. A. Isolation and characterization of halotolerant alkaliphilic methanotrophic bacteria from Tuva soda lakes. *Curr. Microbiol.* **1997**, *35*, 257−261.
(8) Dedysh, S. N.; Panikov, N. S.; Liesack, W.; Großkopf, R.; Zhou, J.; Tiedje, J. M. Isolation of acidophilic methane-oxidizing bacteria from northern peat wetlands. *Science* **1998**, *282*, 281−284.
(9) Takeuchi, M.; Kamagata, Y.; Oshima, K.; Hanada, S.; Tamaki, H.; Marumo, K.; Maeda, H.; Nedachi, M.; Hattori, M.; Iwasaki, W.; Sakata, S. Methylocaldum marinum sp. nov., a thermotolerant, methane-oxidizing bacterium isolated from marine sediments, and emended description of the genus Methylocaldum. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 3240−3246.
(10) Pandey, V. C.; Singh, J. S.; Singh, D. P.; Singh, R. P. Methanotrophs: promising bacteria for environmental remediation. *Int. J. Environ. Sci. Technol.* **2014**, *11*, 241−250.
(11) Semrau, J. D.; DiSpirito, A. A.; Yoon, S. Methanotrophs and copper. *FEMS Microbiol. Rev.* **2010**, *34*, 496−531.
(12) Crombie, A.; Murrell, J. C. Development of a system for genetic manipulation of the facultative methanotroph Methylocella silvestris BL2. *Methods Enzymol.* **2011**, *495*, 119−133.
(13) Ishikawa, M.; Yokoe, S.; Kato, S.; Hori, K. Efficient counterselection for Methylococcus capsulatus (Bath) by using a mutated pheS gene. *Appl. Environ. Microbiol.* **2018**, *84*, No. e01875.
(14) Yan, X.; Chu, F.; Puri, A. W.; Fu, Y.; Lidstrom, M. E. Electroporation-based genetic manipulation in type I methanotrophs. *Appl. Environ. Microbiol.* **2016**, *82*, 2062−2069.
(15) Smith, T. J.; Murrell, J. C. Mutagenesis of soluble methane monooxygenase. *Methods Enzymol.* **2011**, *495*, 135−147.
(16) Singh, U. S.; Singh, R. P. *Molecular Methods in Plant Pathology*. CRC Press; 2017.
(17) Marx, C. J.; Lidstrom, M. E. Broad-host-range cre-lox system for antibiotic marker recycling in gram-negative bacteria. *Biotechniques* **2002**, *33*, 1062−1067.
(18) Marx, C. J. Development of a broad-host-range sacB-based vector for unmarked allelic exchange. *BMC Res. Notes* **2008**, *1*, 1−8.
(19) Ran, F. A.; Hsu, P. D.; Wright, J.; Agarwala, V.; Scott, D. A.; Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **2013**, *8*, 2281−2308.
(20) Ma, H.; Marti-Gutierrez, N.; Park, S. W.; Wu, J.; Lee, Y.; Suzuki, K.; Koski, A.; Ji, D.; Hayama, T.; Ahmed, R.; Darby, H.; Van Dyken, C.; Li, Y.; Kang, E.; Park, A.-R.; Kim, D.; Kim, S.-T.; Gong, J.; Gu, Y.; Xu, X.; Battaglia, D.; Krieg, S. A.; Lee, D. M.; Wu, D. H.; Wolf, D. P.; Heitner, S. B.; Belmonte, J. C. I.; Amato, P.; Kim, J.-S.; Kaul, S.; Mitalipov, S. Correction of a pathogenic gene mutation in human embryos. *Nature* **2017**, *548*, 413−419.
(21) Gratz, S. J.; Rubinstein, C. D.; Harrison, M. M.; Wildonger, J.; O'Connor-Giles, K. M. CRISPR-Cas9 genome editing in Drosophila. *Curr. Protoc. Mol. Biol.* **2015**, *111*, 31.2.1−31.2.20.
(22) Jiang, W.; Bikard, D.; Cox, D.; Zhang, F.; Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **2013**, *31*, 233−239.
(23) Tapscott, T.; Guarnieri, M. T.; Henard, C. A. Development of a CRISPR/Cas9 system for Methylococcus capsulatus in vivo gene editing. *Appl. Environ. Microbiol.* **2019**, *85*, e00340−e003419.
(24) Whittenbury, R.; Phillips, K. C.; Wilkinson, J. F. Enrichment, isolation and some properties of methane-utilizing bacteria. *Microbiology* **1970**, *61*, 205−218.
(25) Promoter and Terminators, Online Analysis Tools Promoters. https://molbiol-tools.ca/Promoters.htm (accessed October 2017).
(26) Martin, H.; Murrell, J. C. Methane monooxygenase mutants of Methylosinus trichosporium constructed by marker-exchange mutagenesis. *FEMS Microbiol. Lett.* **1995**, *127*, 243−248.
(27) Rumah, B. L.; Stead, C. E.; Claxton Stevens, B. H.; Minton, N. P.; Grosse-Honebrink, A.; Zhang, Y. Isolation and characterisation of Methylocystis spp. for poly-3-hydroxybutyrate production using waste methane feedstocks. *AMB Express* **2021**, *11*, 6.
(28) Puri, A. W.; Owen, S.; Chu, F.; Chavkin, T.; Beck, D. A.; Kalyuzhnaya, M. G.; Lidstrom, M. E. Genetic tools for the industrially promising methanotroph Methylomicrobium buryatense. *Appl. Environ. Microbiol.* **2015**, *81*, 1775−1781.
(29) Pyne, M. E.; Bruder, M. R.; Moo-Young, M.; Chung, D. A.; Chou, C. P. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Sci. Rep.* **2016**, *6*, 25666.
(30) Wasels, F.; Jean-Marie, J.; Collas, F.; López-Contreras, A. M.; Lopes Ferreira, N. L. A two-plasmid inducible CRISPR/Cas9 genome editing tool for Clostridium acetobutylicum. *J. Microbiol. Methods* **2017**, *140*, 5−11.
(31) Cañadas, I. C.; Groothuis, D.; Zygouropoulou, M.; Rodrigues, R.; Minton, N. P. RiboCas: a universal CRISPR-based editing tool for Clostridium. *ACS Synth. Biol.* **2019**, *8*, 1379−1390.
(32) Pyne, M. E.; Bruder, M.; Moo-Young, M.; Chung, D. A.; Chou, C. P. Technical guide for genetic advancement of underdeveloped and intractable Clostridium. *Biotechnol. Adv.* **2014**, *32*, 623−641.
(33) Vento, J. M.; Crook, N.; Beisel, C. L. Barriers to genome editing with CRISPR in bacteria. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 1327−1341.
(34) Pfeiffer, D.; Jendrossek, D. Localization of Poly(3-Hydroxybutyrate) (PHB) Granule-Associated Proteins during PHB Granule Formation and Identification of Two New Phasins, PhaP6 and PhaP7, in Ralstonia eutropha H16. *J. Bacteriol.* **2012**, *194*, 5909−5921.
(35) Endo, M.; Mikami, M.; Toki, S. Biallelic gene targeting in rice. *Plant Physiol.* **2016**, *170*, 667−677.
(36) Maruyama, T.; Dougan, S. K.; Truttmann, M. C.; Bilate, A. M.; Ingram, J. R.; Ploegh, H. L. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **2015**, *33*, 538−542.
(37) Sharda, M.; Badrinarayanan, A.; Seshasayee, A. S. Evolutionary and comparative analysis of bacterial nonhomologous end joining repair. *Genome Biol. Evol.* **2020**, *12*, 2450.
(38) So, Y.; Park, S. Y.; Park, E. H.; Park, S. H.; Kim, E. J.; Pan, J. G.; Choi, S. K. A highly efficient CRISPR-Cas9-mediated large genomic deletion in Bacillus subtilis. *Front. Microbiol.* **2017**, *8*, 1167.
(39) Del Cerro, C.; García, J. M.; Rojas, A.; Tortajada, M.; Ramón, D.; Galán, B.; Prieto, M. A.; García, J. L. Genome Sequence of the Methanotrophic Poly-β-hydroxybutyrate Producer Methylocystis Parvus OBBP. *J. Bacteriol.* **2012**, *194*, 5709.
(40) Liu, J.; Wang, Y.; Lu, Y.; Zheng, P.; Sun, J.; Ma, Y. Development of a CRISPR/Cas9 genome editing toolbox for Corynebacterium glutamicum. *Microb. Cell Fact.* **2017**, *16*, 205.

(41) Wang, S.; Hong, W.; Dong, S.; Zhang, Z. T.; Zhang, J.; Wang, L.; Wang, Y. Genome engineering of Clostridium difficile using the CRISPR-Cas9 system. *Clin. Microbiol. Infect.* **2018**, *24*, 1095−1099.
(42) Chayot, R.; Montagne, B.; Mazel, D.; Ricchetti, M. An end-joining repair mechanism in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 2141−2146.
(43) Petit, M. A.; Ehrlich, S. D. The NAD-dependent ligase encoded by yerG is an essential gene of Bacillus subtilis. *Nucleic Acids Res.* **2000**, *28*, 4642−4648.

https://doi.org/10.1021/acssynbio.2c00554
ACS Synth. Biol. 2023, 12, 544−554