Characterisation of a novel soluble di-iron monooxygenase from the soil organism *Solimonas soli*

Sui Nin Nicholas Yang

Thesis submitted to fulfil the requirements for the degree of

Doctor of Philosophy

School of Life and Environmental Sciences, Faculty of Science

The University of Sydney

February 2025

This research is funded by a Research Training Program scholarship provided by

the Australian Government

Table of Contents

Declaration

I certify that, to the best of my knowledge, the majority of the content of this thesis is my own work. Work performed by other researchers is acknowledged in the Author Attributions section below. This thesis has not been submitted for any degree or other purpose. I certify that the intellectual content of this thesis is the product of my own work and that all assistance received in the preparation of this thesis has been acknowledged.

Sui Nin Nicholas Yang, 4th February 2025

Author attributions

This thesis contains material from the following publications which were altered to better fit the style of a thesis, otherwise the content is largely identical:

1. Yang, S. N. N., Haritos, V., Kertesz, M. A., Coleman, N. V. (2024) A novel soluble di-iron monooxygenase from the soil bacterium *Solimonas soli*. Environmental Microbiology, 26(2), e16567. The publication is represented in Sections 3.1-3.4 in Chapter 3 and Sections 4.2.2, 4.2.3, and 4.3.3, and a paragraph included in Section 4.4. Victoria Haritos (Monash University, Victoria, Australia) undertook the work presented in Figure 3.3.2, and drafted the associated paragraph found in Section 3.3.1. I performed all experimental work under the supervision of Nick Coleman, except for the work noted above. I drafted the manuscript with input and critical revision of the draft by Nick Coleman, Michael Kertesz, and Victoria Haritos.

2. Yang, S. N. N., Kertesz, M. A., Coleman, N. V. (2025) Phylogenetic and functional diversity of soluble di-iron monooxygenases. Environmental Microbiology, 27(2), e70050. This publication is represented in Chapter 1. I designed the scope of the literature review together with Nick Coleman. Nick Coleman prepared Figure 1.1.1 and drafted the associated paragraph found in Section 1.1.1. Michael Kertesz edited the figure legend for Figure 1.1.1 and the associated paragraph in Section 1.1.1. I carried out the main data collection and bioinformatic analysis. I drafted the manuscript, which was then edited and revised by Nick Coleman and Michael Kertesz.

Sui Nin Nicholas Yang, 4th February 2025

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Michael Kertesz, 4th February, 2025

Abstract

Monooxygenase enzymes are responsible for the oxidation of hydrocarbons and other compounds in the carbon and nitrogen cycles, are important for the biodegradation of pollutants, and can act as biocatalysts for chemical manufacturing. The soluble di-iron monooxygenases (SDIMOs) are of interest due to their broad substrate range, high enantioselectivity, and ability to oxidise inert substrates like methane. An unusual SDIMO was detected in an earlier study in the genome of the soil organism *Solimonas soli* but was not characterised. This study has shown that the *S. soli* SDIMO is part of a new clade, which is defined as 'Group 7'. These share a conserved gene organisation with alkene monooxygenases but have only low amino acid identity. The *S. soli* genes (named *zmoABCD*) was functionally expressed in *Pseudomonas putida* KT2440 and the recombinants made epoxides from C_2-C_8 alkenes, preferring small linear alkenes (especially propene), but also epoxidising branched, carboxylated and chlorinated substrates. ZmoABCD oxidised the organochlorine pollutants vinyl chloride (VC) and *cis*-1,2-dichloroethene (cDCE) and released inorganic chloride from VC but not cDCE. However, the original host bacterium *S. soli* could not grow on any alkenes tested but grew well on phenol, *n*-octane, isoleucine, leucine, and Tween 80. Δ*zmoABCD* knockout strains of *S. soli* were also able to grow using these substrates as sole carbon sources, suggesting that ZmoABCD is not the sole enzyme responsible for growth on these molecules. The regulation of *zmo* expression was also studied using a plasmid-borne bioreporter construct, where a *gfp* reporter gene was placed under the control of the native *zmo* promoter to identify potential *zmoABCD*-inducing substrates. It was demonstrated in *S. soli* that Tween 80, butane, 1-butene, and 2-propanol induced the expression of *zmoABCD* while *n*-octane and 1-octene repressed expression. Interestingly, the butanol isomers did not induce nor repress expression of the *zmoABCD* cluster. The same construct was used to study expression of the *zmo* promoter in *E. coli* and *S. soli*, with particular focus on the 100 bp untranslated region between the *zmo* promoter and the ribosome binding site of the *zmoABCD* cluster. Deletion of the entire 100 bp region in a multicopy plasmid carrying the *gfp* reporter led to an increase in GFP fluorescence observed in both *E. coli* and *S. soli*. Partial deletions of the region gave similar results and allowed the effect to be localised to the proximal 50 bp of the region. It is hypothesised that an operator region exists in this region. It is also hypothesised that the regulator protein acts as a repressor, but further experiments are required to elucidate its mode of action.

While this study was unable to identify the canonical substrate of ZmoABCD, it has provided a substantial framework to narrow down the substrate range of this novel SDIMO. The characterisation of ZmoABCD will result in the enhanced understanding of the functions of the SDIMO enzymes. This will enable better prediction, increase our understanding of SDIMO evolutionary history, management of biogeochemical processes, and enable new applications of these enzymes for biocatalysis and bioremediation.

Acknowledgements

Firstly, to Nick Coleman, you were my first introduction to the world of microbiology and SynBio. You are the reason I chose to pursue Honours then a PhD in the first place and I cannot begin without thanking you for your years of mentorship, guidance, and friendship. To Michael, thank you for the support and advice during the last leg of the PhD journey. Thank you for being so generous with your time and expertise despite my late arrival to the Kertesz Lab. I could not have asked for two better supervisors. I have learned so much from the both of you these past few years. I also want to extend my appreciations to Associate Professor Victoria Haritos at Monash University and Professor Che Ok Jeon at Chung-Ang University. Thank you to you both for sending us the resources required to begin this project, without assistance from the both of you this project would not have been possible.

To the Coleman Lab members, Scott, Claudia, Ngaio, Marcus, and Brodie, thank you for the fun lab chats and camaraderie. Especially to Mark, thank you for the years of patience, your tips and tricks for lab work, and most of all your wit, humour, and love of musical theatre. I have missed working with you. To the Kertesz Lab members, Rebecca, Amos, Meg, Mahlo, Annabel, and Annabelle, thank you for the supportive environment during a tumultuous time in my PhD journey. Thank you to Matt especially, for being incredibly helpful in the last few months of this journey. In the short time I have spent in the Kertesz Lab I have learned so much about mushrooms and mushroom compost, sometimes against my will, but it is an experience (at times sensory) I will never forget. To the others in LEES, Nellie, Nic, Lachlan, Jasmin, Andrew, and Minh, thank you for chats, teas and baked goods. They made the arduous days in the office much more bearable.

To my family, thank you for loving and supporting me throughout the years of academic pursuit. Linus, it is now your turn.

Finally, to Daniel, thank you for everything. I couldn’t have done it without you.

Publications

Publications

Yang, S. N. N., Haritos, V., Kertesz, M. A., Coleman, N. V. (2024) A novel soluble di-iron monooxygenase from the soil bacterium *Solimonas soli*. Environmental Microbiology, 26, e16567. DOI: 10.1111/1462-2920.16567.

Yang, S. N. N., Kertesz, M. A., Coleman, N. V. (2025) Phylogenetic and functional diversity of soluble di-iron monooxygenases. Environmental Microbiology, 27(2), e70050. DOI: https://doi.org/10.1111/1462-2920.70050

Presentations

Yang, S. N. N., Haritos, V., Coleman, N. V. (2022) Characterisation of a novel monooxygenase in the soil bacterium *Solimonas soli*. Invited seminar presented at the Australian Microbial Ecology Conference. Melbourne, Australia, November 2022.

Yang, S. N. N., Haritos, V., Kertesz, M. A., Coleman, N. V. (2023) Characterisation of a novel monooxygenase in the soil bacterium *Solimonas soli*. Invited seminar presented at the Joint Academic Microbiology Seminars. Sydney, Australia, October 2023.

Yang, S. N. N., Coleman, N. V., Kertesz, M. A. (2024) The regulation of the expression of the *zmoABCD* cluster in the soil organism *Solimonas soli*. Poster presented at the Australian Microbial Ecology Conference. Sydney, Australia, October-November 2024.

List of Figures

List of Tables

List of Abbreviations

°C	Degree Celsius
aa	Amino acid
AGRF	Australian Genome Research Facility
AlkB	Alkane monooxygenase
BLAST	Basic local alignment search tool
BMO	Butane monooxygenase
bp	Base pair
cDCE	*cis*-1,2-dichloroethylene
CuMMO	Copper membrane monooxygenase
DMS	Dimethyl sulphide
DMSO	Dimethyl sulphoxide
DMSP	Dimethylsulfoniopropionate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
fuGFP/*fuGFP*	Free-use green fluorescent protein
GFP/*gfp*	Green fluorescent protein
Gm	Gentamicin
Gm^R	Gentamicin resistant
HGT	Horizontal gene transfer
kb	Kilobases
Km	Kanamycin
Km^R	Kanamycin resistant
KO	Knockout
KOM	Knockout – merodiploid
LAB	Lithium acetate buffer
LB	Lysogeny broth
MCS	Multiple cloning site
MO	Monooxygenase
MOPS	3-(N-Morpholino)propanesulfonic acid

MSM	Minimal salts medium
NBP	4-(4-nitrobenzyl)-pyridine
NCBI	National Centre for Biotechnology Information
OD_{600}	Optical density measured at 600 nm
PCR	Polymerase chain reaction
PH/PhMO	Phenol hydroxylase/Phenol monooxygenase
pMMO	Particulate methane monooxygenase
Pr1MO	Propane-1-monooxygenase
Pr2MO	Propane-2-monooxygenase
R2A	Reasoner's 2A medium
R2A+	Modified Reasoner's 2A medium
RBS	Ribosome binding site
Rif	Rifampicin
Rif^R	Rifampicin resistant
RNA	Ribonucleic acid
RO	Reverse osmosis
SDIMO	Soluble di-iron monooxygenase
sMMO	Soluble methane monooxygenase
Tc	Tetracycline
Tc^R	Tetracycline resistant
THF	Tetrahydrofuran
Thm	Tetrahydrofuran monooxygenase
T_m	Melting temperature
TMO	Toluene monooxygenase
Tris	Tris(hydroxymethyl)aminomethane
VC	Vinyl chloride
WT	Wild type
ZmoABCD/*zmoABCD*	*Solimonas soli* group 7 SDIMO

Chapter 1 – Introduction

Chapter published as: Yang, S. N. N., Kertesz, M. A., Coleman, N. V. (2025) Phylogenetic and functional diversity of soluble di-iron monooxygenases. Environmental Microbiology, 27(2), e70050. DOI: https://doi.org/10.1111/1462-2920.70050

1.1. General introduction

1.1.1. Bacterial monooxygenases

Hydrocarbons are released into the environment by both natural and industrial processes and are of interest as intermediates in global carbon cycles (Koo and Rosenzweig, 2021, Tucci and Rosenzweig, 2024), as persistent pollutants (Shennan, 2005), and as useful substrates for biocatalysis (Donadio *et al.*, 2015). Bacteria can use most hydrocarbons as carbon and energy sources, provided that environmental conditions are appropriate, but they require specialized enzymes for their metabolism. In the case of aromatic hydrocarbons, dioxygenase enzymes are the typical initial catalysts under aerobic conditions (Gibson and Parales, 2000), while for aliphatic hydrocarbons (alkanes and alkenes), monooxygenase (MO) enzymes play this role (Shennan, 2005, Guo *et al.*, 2023). This thesis will focus on the MO enzymes, which have significant roles in biogeochemistry (Greening and Grinter, 2022) and many interesting applications in biotechnology (Canada *et al.*, 2002, Donadio *et al.*, 2015).

Bacteria expressing monooxygenase enzymes are very important for the bioremediation of pollutants. A recent, well-studied example is the breakdown of 1,4-dioxane, a common groundwater pollutant, which occurs via both co-metabolic (Wang *et al.*, 2021b) and growth-linked (Mahendra and Alvarez-Cohen, 2005, He *et al.*, 2017) processes in monooxygenase-containing bacteria. There is also very extensive research on the use of MO enzymes for the biodegradation and bioremediation of chloroethenes such as trichloroethene (TCE), which are also common groundwater pollutants (Shennan, 2005, Mattes *et al.*, 2010). As with 1,4-dioxane, both co-metabolic (Ojo *et al.*, 2023) and growth-linked (Jin and Mattes,

2008) physiologies are associated with the degradation of chloroethenes, depending on the substrate and the bacterial host.

Monooxygenase enzymes are highly prized for their applications in biocatalysis, due to their ability to add oxygen atoms to substrates with high chemo-, regio-, and stereo-specificity under mild reaction conditions (Constable *et al.*, 2007, Que and Tolman, 2008). As a result, there is great interest in discovering and creating novel MOs that can catalyse these reactions for industrial and medicinal chemistry purposes (Constable *et al.*, 2007, Leak *et al.*, 2009, Bryan *et al.*, 2018, Petkevičius *et al.*, 2019). As will become increasingly clear throughout the introduction, a key theme for MO enzymes is 'diversity'. This applies to both the types of MOs that are useful for biocatalysis (these include iron, copper, and flavin-containing enzymes) (Torres Pazmiño *et al.*, 2010) and also the types of reactions that are catalysed (hydroxylation, epoxidation, desaturation) (Keener and Arp, 1994).

Several different classes of MO can be found in Bacteria and Archaea, and a huge diversity of sequence types exists within each of these classes. Bacterial MOs have broad substrate ranges and provide diverse selective benefits. They can therefore be found in physiologically and taxonomically diverse bacteria (Osborne and Haritos, 2019), and are often subject to horizontal gene transfer (HGT) via plasmids (Zou *et al.*, 2021). The major exceptions to this rule are the particulate and soluble methane monooxygenases (pMMO, sMMO), which are specialised for methane oxidation. These are more restricted in their taxonomic distribution and physiological roles (Kalyuzhnaya *et al.*, 2019, Khider *et al.*, 2021). For many years it was believed that pMMO was the only example of a copper-containing membrane-located MO (CuMMO), but it is now clear that both CuMMOs and the related ammonia monooxygenases (AMOs) also exist in non-methanotrophic lineages of Bacteria and Archaea (Coleman *et al.*, 2012, Diamond *et al.*, 2022).

The MOs are typically categorised based on a combination of their cellular location and the cofactors they require (Torres Pazmiño *et al.*, 2010, Coleman *et al.*, 2011), leading to the following classes: haem-containing MOs (cytochrome p450) (Bernhardt, 2006), flavin-dependent MOs (van Berkel *et al.*, 2006), copper-containing membrane-located MOs (Koo *et al.*, 2022, Tucci *et al.*, 2023), pterin-dependent MOs

(Zhao *et al.*, 1994), di-iron membrane-located MOs (AlkB) (Ji *et al.*, 2013, Guo *et al.*, 2023), cofactor-independent MOs (Fetzner, 2002), and soluble di-iron monooxygenases (SDIMOs) (Leahy *et al.*, 2003). Alternatively, the MOs can be categorised based on their applications (biogeochemistry and/or bioremediation and/or biocatalysis, as described above) or their substrate range, but these classifications are less useful, since MOs typically have significant roles across multiple application areas and tend to have broad substrate ranges.

The relationship of MOs to the physiology of the host cell is complex. Some MO reactions are linked to productive metabolism (growth-linked), while others are incidental and do not yield carbon or energy for the cell (cometabolism) (Horvath, 1972, Semprini, 1997). These distinctions are significant for how MO-containing bacteria are deployed for different applications. The MOs covered here almost always act as the first step in a metabolic pathway, and their ‘primary substrate’ (where known) has been defined as the compound used for the enrichment and isolation of the host bacterium. The primary substrate (as defined here) will therefore always be growth-linked, but this does not necessarily imply that this is the ‘best’ substrate in terms of enzyme affinity or turnover rate (for example, sMMO oxidises ethylene more rapidly than methane (Colby *et al.*, 1977), likely due to the chemical lability of ethylene vs. methane). Further complexity is introduced when the broad variety of compounds which can act as inducers of MO enzymes are considered (i.e. inducing expression of the corresponding genes). These are often growth substrates for the respective bacteria, but they can also be downstream metabolites, co-metabolic substrates, or unrelated compounds. These conceptual relationships are summarised in Figure 1.1.1.

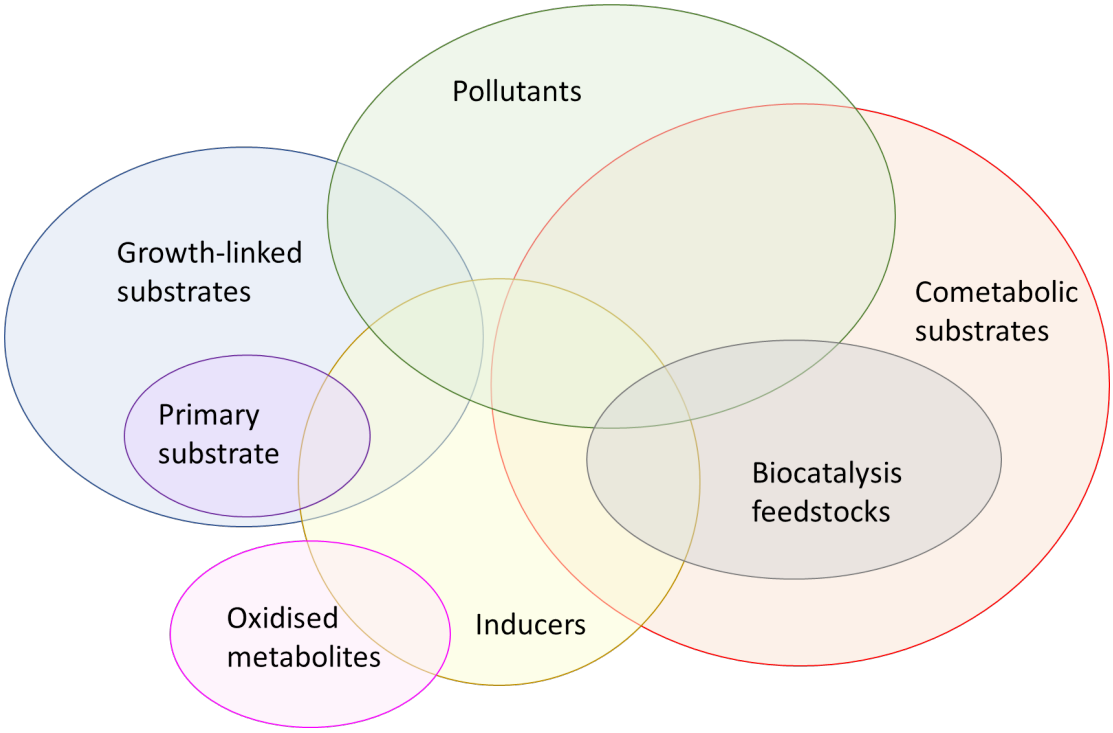


Figure 1.1.1. Relationships between compounds that are substrates for monooxygenase enzymes, with respect to host cell physiology.

These include substrates that support growth of the host (including the "primary substrate" that was used for initial isolation of the host bacterium), and those that are transformed co-metabolically. Environmental pollutants that are MO substrates may fall into either of these groups. Although many biocatalysis feedstocks are growth substrates, in a biomanufacturing situation the MO reaction product is intentionally diverted from further metabolism, and so these are shown as not growth-linked. Inducers of MO gene expression include compounds from both groups of substrates, but also metabolites and other compounds.

1.1.2. The soluble di-iron monooxygenases

The soluble di-iron monooxygenases (SDIMOs) are multicomponent enzymes, requiring at least three components: a hydroxylase, a reductase, and a coupling protein (Leahy *et al.*, 2003, Osborne and Haritos, 2019). The hydroxylase component is itself complex, containing two or three subunits, arranged as $\alpha_2\beta_2$ or $\alpha_2\beta_2\gamma_2$. Some SDIMOs also contain a ferredoxin or other accessory components (Zhou *et al.*, 1998). The reductase extracts electrons from NADH and transfers them to the di-iron core in the hydroxylase α subunit, with assistance from the coupling protein (and ferredoxin, if present) (Wang *et al.*, 2014). The hydroxylase then activates one oxygen atom in O_2 to a high-energy state, and this then attacks the substrate, while the other oxygen atom is reduced to H_2O (Stainthorpe *et al.*, 1990,

Cardy *et al.*, 1991, Sullivan *et al.*, 1998, Banerjee *et al.*, 2019). The product(s) generated are dependent on the substrate and the SDIMO.

The SDIMOs have been reviewed previously (Leahy *et al.*, 2003, Notomista *et al.*, 2003, Coleman *et al.*, 2006, Osborne and Haritos, 2019), and classified into groups according to their protein sequences, the number of subunits, the operon arrangement, and their substrate specificity. Recent massive expansions of sequence databases warrant a re-investigation of these enzymes. An up-to-date overview of SDIMO sequence diversity, groups, and operon structures is given in Figure 1.1.2. A better understanding of the sequence diversity and evolutionary relationships of SDIMOs is needed to enable better predictions of their substrates, physiological roles, and ecological significance. The introduction of this study aims to update the definitions of SDIMO groups, highlight recent discoveries and experimental advances, and pinpoint knowledge gaps for future work. The first focus will be on phylogenetic analyses, with sections organised based on the previously defined groups 1 to 6 (Leahy *et al.*, 2003, Notomista *et al.*, 2003), with the addition of a novel group 7 SDIMO subgroup. After exploring the phylogeny of the SDIMOs, progress on heterologous expression (this is an important tool for the development of biocatalysts) will be summarised. Evidence for HGT of SDIMO genes will also be examined, since this plays a large role in the evolution of hydrocarbon-degrading bacteria (Osborne and Haritos, 2018, Zou *et al.*, 2021). Finally, recent examples of the use of SDIMOs for bioremediation of pollutants will be highlighted.

Some comments on nomenclature are important before beginning the literature review. First, there is a discrepancy in the literature between SDIMO groups 1 and 2. These group numbers introduced by Notomista *et al.* (2003) were switched in a later paper (Holmes and Coleman, 2008). As this error has now been propagated in the literature (Moratti *et al.*, 2022, Ren *et al.*, 2022), the more recent naming scheme will be followed, where group 1 are the “toluene MOs” and group 2 are the “phenol MOs”. Secondly, to simplify language, that protein sequences encoded in genomes and metagenomes will be assumed to be expressed and functional, i.e. they will be referred to as ‘enzymes’. Thirdly, the phylogeny of the alpha subunit will be used as a proxy for the holoenzymes, although typically this only holds true for the two parts of the hydroxylase component (aβ), and less so for the reductase (Leahy *et al.*, 2003).

Finally, the most recent and correct terms for different taxa have been used, so ‘Proteobacteria’ are now ‘Pseudomonadota’, ‘Firmicutes’ are ‘Bacillota’, and ‘Actinobacteria’ are ‘Actinomycetota’ (Oren and Garrity, 2021). The older genus name ‘*Mycobacterium’* has been used (Meehan *et al.*, 2021) rather than splitting this genus into five new genera as previously proposed (Gupta *et al.*, 2018).

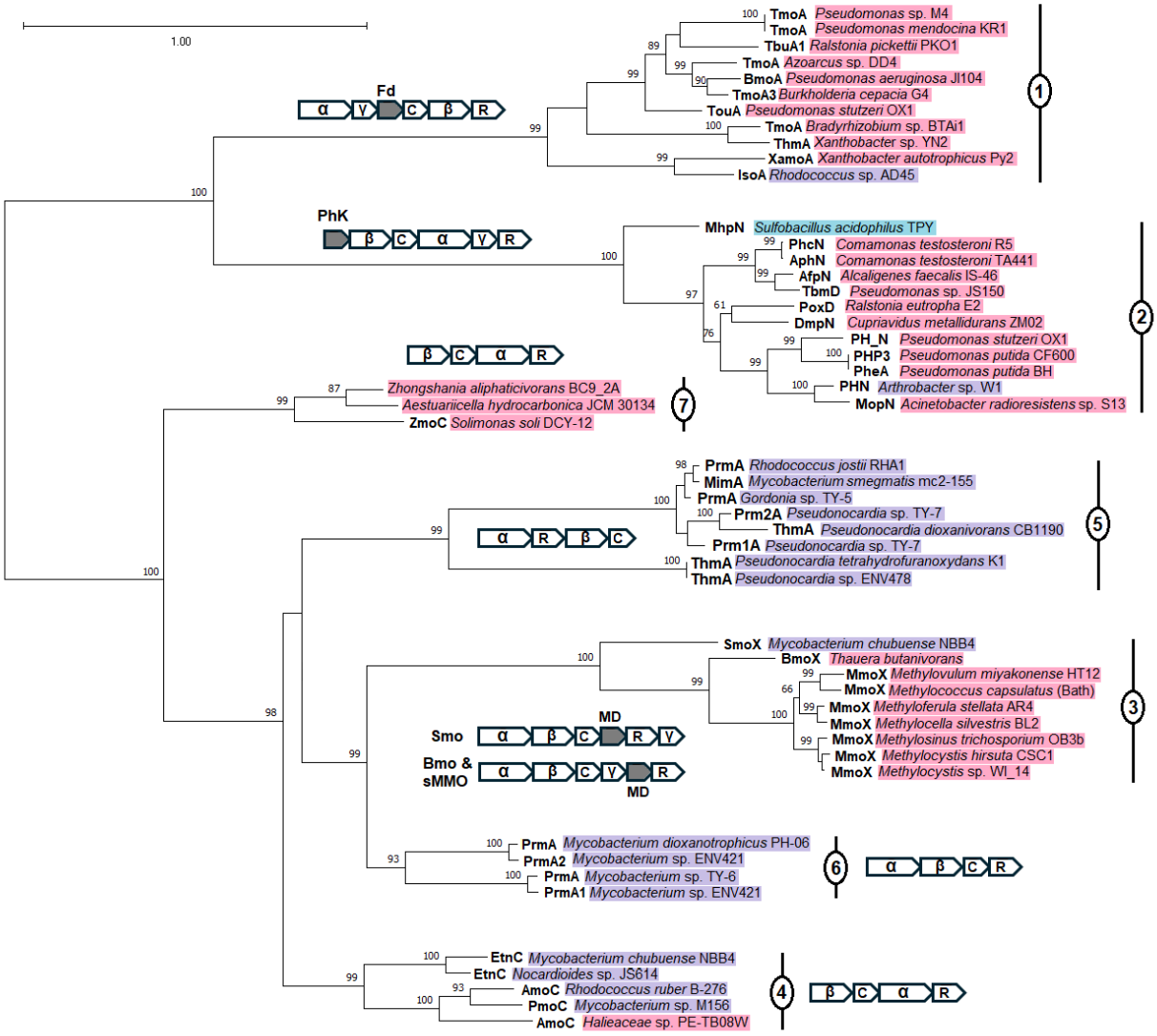


Figure 1.1.2. Evolutionary relationships of SDIMO enzymes.

The maximum likelihood tree was made from a trimmed alignment of alpha-subunit sequences (494 amino acid length), with numbers at nodes indicating percentage bootstrap values. Group numbers are shown in ovals at the right. Only experimentally characterised SDIMOs are shown in the tree, with the exceptions of the predicted enzymes from *Zhongshania aliphativorans* and *Aestuaricella hydrocarbonica* in group 7. The operon structures are shown as block arrow diagrams, with subunits as follows: α = alpha hydroxylase, β = beta hydroxylase, γ = gamma hydroxylase, C = coupling protein, R =

reductase, Fd = ferredoxin, MD = MMOD or MMOD-like, PhK = PhK-like. Coloured shading behind organism names indicates their phylum, as follows: purple = Actinomycetota, pink = Pseudomonadota, light blue = Bacillota. Bootstrap values less than 50% are not shown. Scale bar units are amino acid substitutions per site.

1.2. Methods

A reference sequence from each SDIMO group (Table 1.2.1) was used for protein similarity searches (BLASTp) against the NCBI Protein database (Johnson *et al.*, 2008), retaining all sequence results with over >80% coverage to the query sequence. The results were filtered further by establishing a percent amino acid (aa) identity value cutoff for each group. These cutoffs were set manually after inspection of the BLASTp results and depended on the level of homology within each SDIMO group (Table 1.2.1). In cases where >30 BLAST matches were retrieved, these sequences were clustered using CD-HIT (Fu *et al.*, 2012), and one representative sequence of each cluster was retained. In cases where CD-HIT analysis yielded an uncharacterised SDIMO as a reference sequence, this was replaced with the sequence of a characterised enzyme with sequence as close as possible to the one suggested by CD-HIT. The sequences were assembled in FASTA format, aligned using ClustalX (Larkin *et al.*, 2007), then exported to GeneDoc (Nicholas and Nicholas, 1997) for manual trimming to give uniform lengths. Internal gaps were retained. PhyML 3.0 online execution (Guindon *et al.*, 2010) was used to generate the trees using the Smart Model Selection option (Lefort *et al.*, 2017) and the aLRT SH-like fast-likelihood-based method for branch supports. MEGA11 (Tamura *et al.*, 2021) was then used to visualise and annotate the trees. The accession numbers for all represented alpha hydroxylase sequences in each phylogenetic tree can be found in Table S1.

Table 1.2.1. Details of methods used for phylogenetic tree construction.

SDIMO Group	Reference sequence	Accession number	Minimum amino acid identity for SDIMO group membership	Number of initial BLASTp matches	CD-HIT cutoff for clustering	Number of sequences used for tree construction	Length of amino acid alignment
1	TouA *Pseudomonas* sp. OX1	WP_128709961.1	35%	632	65%	47	533
2	PoxD *Ralstonia* sp. E2	AAC32455.1	35%	1737	70%	52	470
3	MmoX *Methylococcus capsulatus* (Bath)	1FYZ_A	50%	104	95%	44	495
4	EtnC *Mycobacterium chubuense* NBB4	ACZ56346.1	50%	22	Not used	22	497
5	MimA *MycobacteriuM. smegmatis* mc^2-155	WP_003893346.1	40%	1049	80%	35	439
6	Uncharacterised group 6 SDIMO *Mycobacterium chubuense* NBB4	WP_014805366.1	50%	44	90%	24	459
7	ZmoC *Solimonas soli* DCY12	WP_051362144.1	60%	14	Not used	14	455

1.3. Overview of the phylogeny and functions of different SDIMO groups

1.3.1. The group 1 SDIMOs: The toluene monooxygenases

The group 1 SDIMOs (Figure 1.1.2, Figure 1.3.1) are commonly referred to as toluene MOs (TMOs). The abbreviation has been kept for convenience, but it is important to note that their substrate range is very broad, and also includes propene, isoprene, isobutene (2-methyl-propene), phenol, *o*-xylene, and 1,4-dioxane as physiological (i.e. growth-supporting) substrates (Table S2). The TMO operons are composed of 6 genes, arranged as follows: α-hydroxylase, γ-hydroxylase, ferredoxin, coupling protein, β-hydroxylase, reductase. The TMOs are abundantly represented in sequence databases (>600 sequences, Table 1.2.1) and are well-characterised in the scientific literature. High sequence diversity exists in the TMO group, as indicated by the low % aa identity values used in the analyses here as cutoffs for group membership (35%) and CD-HIT clustering (65%) (Table 1.2.1). Further details of TMOs that have been experimentally characterised can be found in Table S2.

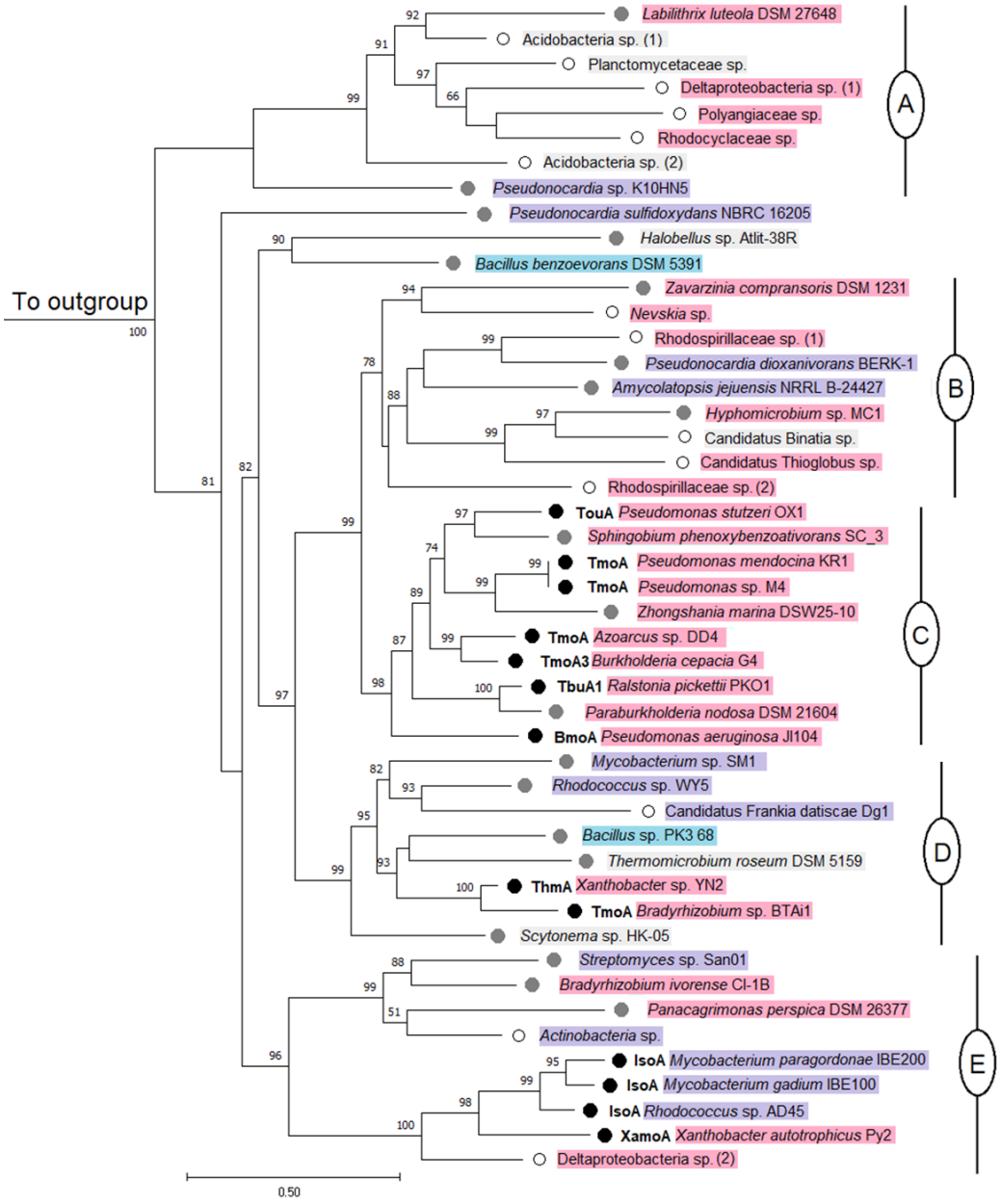


Figure 1.3.1. Evolutionary relationships of the group 1 SDIMOs.

The maximum likelihood tree was made from a trimmed alignment of alpha-subunit sequences (533 aa length), with numbers at nodes indicating percentage bootstrap values. Bootstrap values less than 50% are not shown. Scale bar units are amino acid substitutions per site. ZmoC of *Solimonas soli* was used as the outgroup to root the tree. Dots preceding organism names indicate the following: black circle = characterised SDIMO, grey circle = uncharacterised SDIMO from pure culture genome, white circle = uncharacterised SDIMO from metagenome-assembled genome (MAG). Coloured shading behind organism names indicates their phylum, as follows: purple = Actinomycetota, pink = Pseudomonadota, light blue = Bacillota, grey = other.

Five subgroups within the TMOs (A-E) have been defined in Figure 1.3.1 in order to facilitate discussion regarding patterns of distribution and function. These subgroups are all supported by high bootstrap values (>90%). Subgroups A and B contain uncharacterised enzymes encoded in genomes and metagenome-assembled genomes (MAGs) from Pseudomonadota, Acidobacteria, Actinomycetota, Planctomycetota, and most intriguingly, the candidate phylum Binatota, which contains many putative hydrocarbon-oxidizers (Chuvochina *et al.*, 2019, Murphy *et al.*, 2021). There is no experimental evidence yet for the substrates of the enzymes in subgroups A and B, but given their placement in the group 1 SDIMOs, it is plausible that their substrates are aromatic hydrocarbons. Subgroup C includes most of the well-studied TMOs, which have toluene as the primary substrate and are found typically in the Pseudomonadota. Subgroup D contains dioxane- and toluene-oxidising enzymes, which are found in diverse phyla, i.e. Bacillota, Actinomycetota, Thermomicrobiota, and Cyanobacteriota. Subgroup E contains the isoprene MOs, in addition to uncharacterised enzymes from Pseudomonadota and Actinomycetota. From Figure 1.3.1, it is clear that the majority of TMO sequences in genomes and metagenomes have unknown functions and significance. This is testament to the high sequence diversity of this group and emphasises the wide knowledge gap between sequences and functions within the SDIMOs more generally.

1.3.2. The group 2 SDIMOs: The phenol monooxygenases

The group 2 SDIMOs (Figure 1.1.2, Figure 1.3.2) are commonly referred to as the phenol hydroxylases (PHs) or the phenol monooxygenases (PhMO). The physiological substrates of the PhMOs are diverse, including not only phenol, but also cresols, halogenated aromatic compounds, toluene, tetrahydrofuran (THF), benzene, and xylene (Table S2). The PhMO operons are composed of 6 genes, arranged as follows: accessory protein, β-hydroxylase, coupling protein, α-hydroxylase, γ-hydroxylase, and reductase. The PhMOs are abundant in sequence databases (>1700 sequences, Table 1.2.1) and are very diverse, as indicated by analyses which suggested 35% aa identity for group membership and 70% aa identity as a CD-HIT cutoff for clustering analyses (Table 1.2.1). The details of experimentally characterised PhMOs can be found in Table S2.

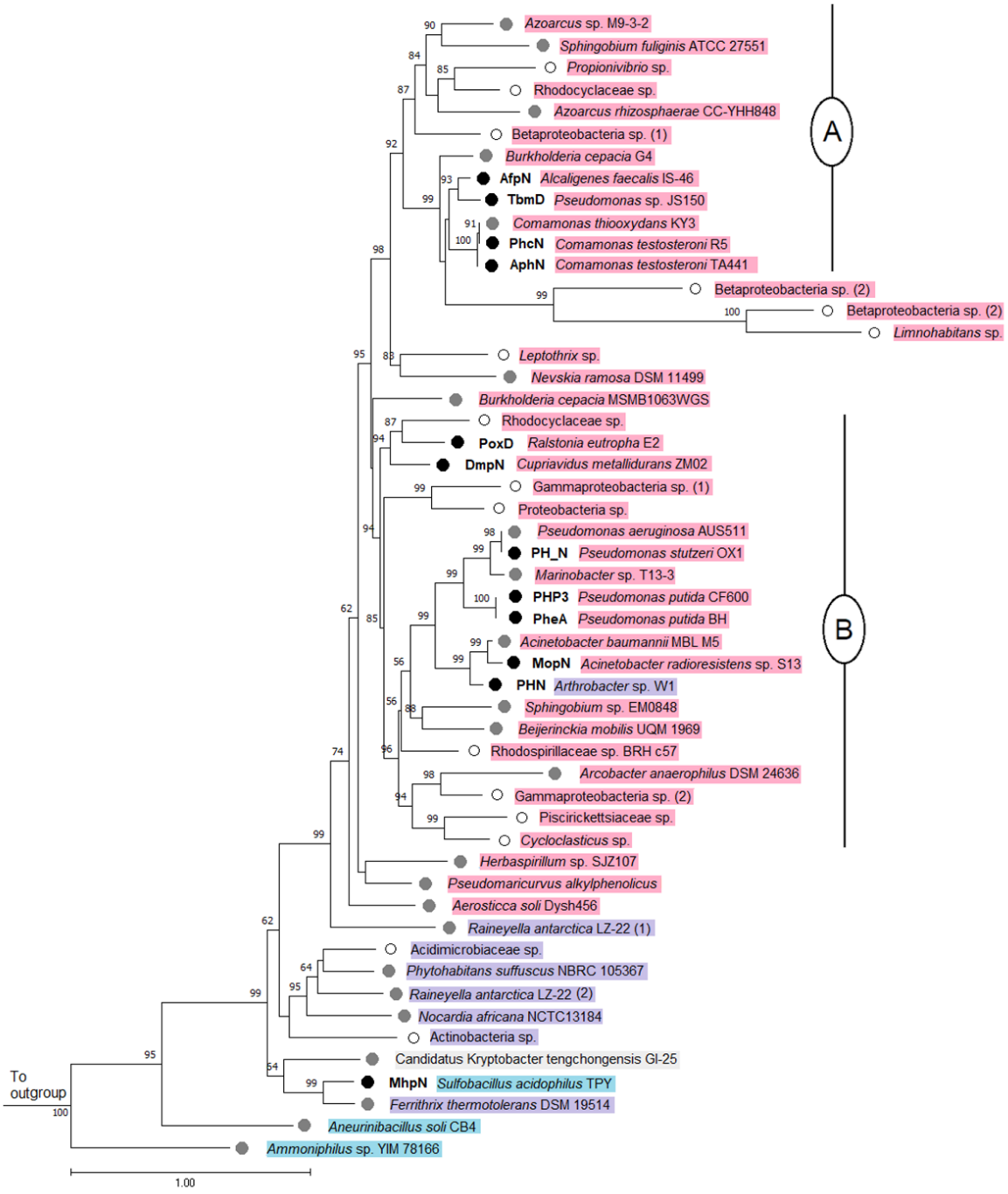


Figure 1.3.2. Evolutionary relationships of the group 2 SDIMOs.

The maximum likelihood tree was made from a trimmed alignment of alpha-subunit sequences (470 aa length), with numbers at nodes indicating percentage bootstrap values. Bootstrap values less than 50% are not shown. Scale bar units are amino acid substitutions per site. ZmoC of *Solimonas soli* was used as the outgroup to root the tree. Dots preceding organism names indicate the following: black circle = characterised SDIMO, grey circle = uncharacterised SDIMO from pure culture genome, white circle = uncharacterised SDIMO from MAG. Coloured shading behind organism names indicates their phylum, as follows: purple = Actinomycetota, pink = Pseudomonadota, light blue = Bacillota, grey = other.

Due to the high diversity of PhMOs and the more gradual transitions seen between the sequences relative to other groups (e.g. compare to the group 3 SDIMO tree, Figure 1.3.3), meaningful subgroups within the PhMO phylogenetic tree were more difficult to establish. The experimentally characterised PhMOs are all represented within the subgroups A and B (these subgroups are supported by bootstrap values >90%), and are harboured by Pseudomonadota, except for MhpLMNOQBP from *S. acidophilus* TPY (Bacillota), and PhmKLMNOP from *Arthrobacter* sp. W1 (Actinomycetota). The PhMO in *S. acidophilus* TPY is unique since it is the only characterised PhMO outside of subgroups A and B, and the only example from a thermoacidophile. Its operon structure is notably different to the other PhMOs (Zhou *et al.*, 2016). Most of the PhMOs have a narrow substrate range compared to their TMO relatives (Table S2), although some members do perform cometabolic oxidations of various pollutants (see Section 1.4 on bioremediation and e.g. *Pseudomonas* CF600 (Shingler *et al.*, 1989, Norlund *et al.*, 1990)). It is intriguing that the deeper branches of the group 2 SDIMO tree are in the Bacillota and Actinomycetota, while the shallower branches are in the Pseudomonadota, suggesting an origin for these enzymes in a Gram-positive lineage followed by a later transfer into the Gram-negatives (see Section 1.6 on horizontal gene transfer).

1.3.3. The group 3 SDIMOs: The soluble methane monooxygenases and relatives

For many years, the group 3 SDIMOs (Figure 1.1.2, Figure 1.3.3) were represented solely by the sMMO enzymes in methanotrophic bacteria, and this is still the most common Group 3 SDIMO found in genomes and metagenomes (two-thirds of the sequences in the NCBI database using our search method). The butane MO (BMO) of *Thauera butanivorans* (BmoXYBZDC) provided the first example of a different group 3 SDIMO, in terms of substrate range and host type (Sluis *et al.*, 2002), later followed by even more divergent examples in the gaseous alkane MOs (Smo) of *Mycobacterium* strain NBB4 (Martin *et al.*, 2014) and *Rhodococcus* strain ZPP (Zou *et al.*, 2021). The sMMOs are arguably the best-characterised SDIMOs, and have well-defined biochemistry, physiology, and structures (Sakai *et al.*, 2023). However, a major barrier that has hindered more detailed molecular studies of the sMMOs has been the lack of a good heterologous expression system (see Section 1.5 on

heterologous expression). This problem also limits their applications in biocatalysis. The physiological substrate range of the sMMOs is limited (methane and short-chain alkanes) but their cometabolic oxidation range is very wide (Table S2).

The sMMO and BMO operons are composed of 6 genes in this order: α-hydroxylase, β-hydroxylase, coupling protein, γ-hydroxylase, accessory protein, and reductase. The Smo operons have 5 genes in this order: α-hydroxylase, β-hydroxylase, coupling protein, γ-hydroxylase, and reductase. The significance of the different operon structures between the different subgroups is not yet known, but it is important to note that the 5-component Smo enzyme of *Mycobacterium* NBB4 was shown to be functional in a heterologous host, despite lacking a homologue of the accessory protein subunit seen in sMMO and BMO (Martin *et al.*, 2014). The sequence diversity of the group 3 SDIMOs is low compared to groups 1 and 2, as demonstrated by the relatively high cutoff for group membership (50% aa identity, Table 1.2.1) and the high clustering value used for CD-HIT analysis (95% aa identity, Table 1.2.1). The details of experimentally characterised group 3 SDIMOs are found in Table S2.

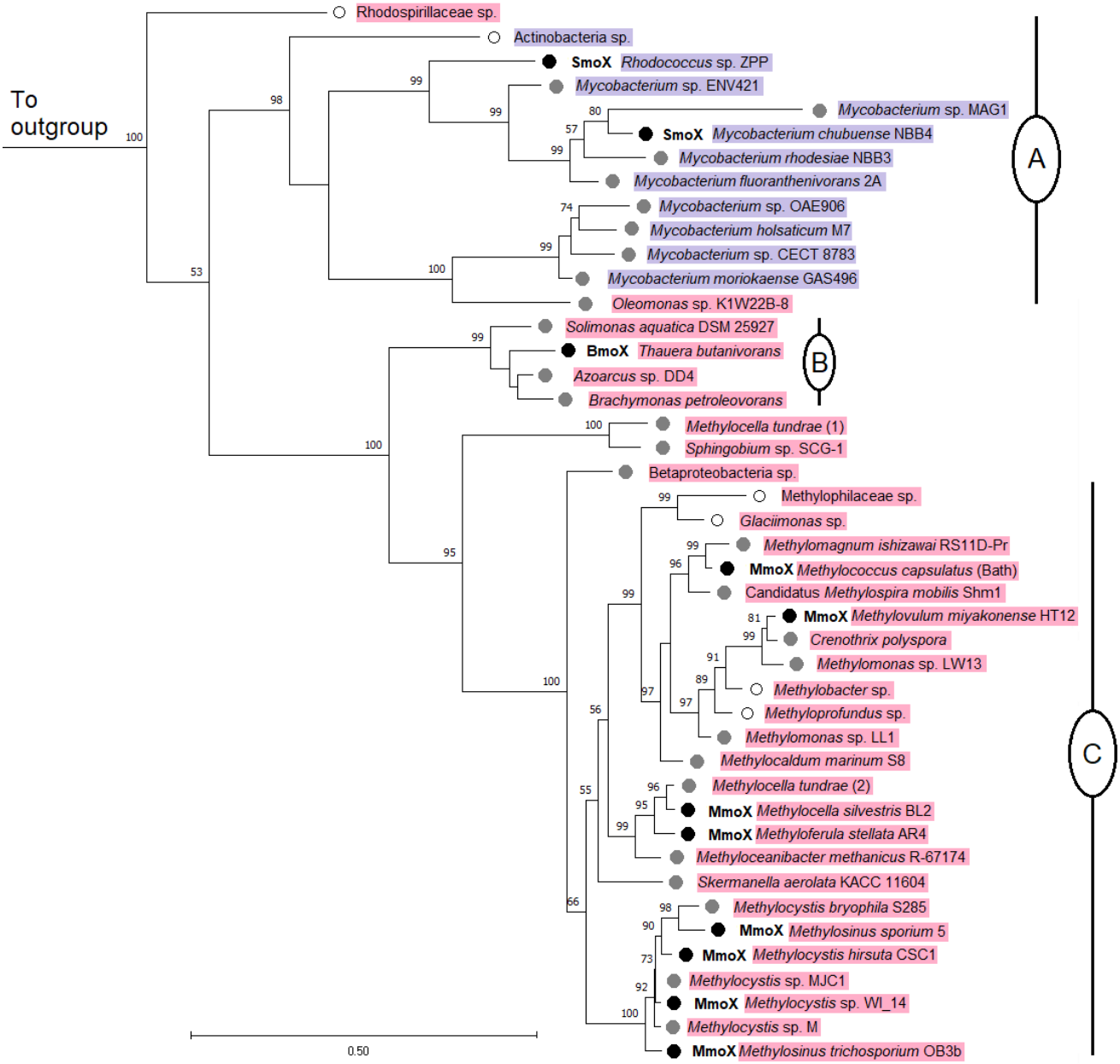


Figure 1.3.3. Evolutionary relationships of the group 3 SDIMOs.

The maximum likelihood tree was made from a trimmed alignment of alpha-subunit sequences (495 aa length), with numbers at nodes indicating percentage bootstrap values. Bootstrap values less than 50% are not shown. Scale bar units are amino acid substitutions per site. ZmoC of *Solimonas soli* was used as the outgroup to root the tree. Dots preceding organism names indicate the following: black circle = characterised SDIMO, grey circle = uncharacterised SDIMO from pure culture genome, white circle = uncharacterised SDIMO from MAG. Coloured shading behind organism names indicates their phylum, as follows: purple = Actinomycetota, pink = Pseudomonadota.

The group 3 SDIMOs are divided into 3 subgroups (A-C) to facilitate discussion (Figure 1.3.3). All of these subgroups are supported by high bootstrap values

(>98%). Subgroup A contains the Smo enzymes and is dominated by Actinomycetota (primarily *Mycobacterium* spp), with one notable exception of a predicted enzyme encoded in the genome of an *Oleomonas* strain (Alphaproteobacteria). Subgroup B enzymes are found in Pseudomonadota and include the BMO, which is thus far the only characterised member of this subgroup. Subgroup C enzymes are also found only in the Pseudomonadota. Most strains harbouring Subgroup C enzymes are canonical methanotrophs that harbour sMMOs and utilise methane as their physiological substrate. The availability of many new sequences in subgroups A and B adds depth to the group 3 SDIMO tree and allows us to make stronger predictions about the origins of methanotrophy. Specifically, this tree strongly supports the previous proposal that the sMMO in methane-oxidising specialists arose from an earlier C_2-C_4 alkane-oxidising Smo-like enzyme in a generalist host (Osborne and Haritos, 2019). The group 3 SDIMO tree shows some similar large-scale patterns to the group 1 SDIMO tree, in that the deeper branches are in the Actinomycetota (with one exception), and the shallower branches are in the Pseudomonadota.

1.3.4. The group 4 SDIMOs: The alkene monooxygenases

The group 4 SDIMOs (Figure 1.1.2, Figure 1.3.4) are commonly referred to as the alkene MOs. Unlike many other SDIMO groups, their physiological substrate range is narrow, and focused on the utilisation of short-chain alkenes, i.e. ethene (Coleman *et al.*, 2011), propene (Saeki and Furuhashi, 1994, Woodland *et al.*, 1995, Saeki *et al.*, 1999), 1-butene (Suzuki *et al.*, 2019), and vinyl chloride (VC) (Mattes *et al.*, 2005) (Table S2). The alkene MO operons are composed of 4 genes which are arranged as follows: β-hydroxylase, coupling protein, α-hydroxylase, reductase. Compared to other SDIMO groups, the alkene MOs are not well-represented in sequence databases (only 22 total sequences, Table 1.2.1) and are not as well-characterised in the scientific literature. The group 4 SDIMOs are exclusively found in the Actinomycetota, with one exception to date (*Halieaceae* strain PE-TB08W, a propene-oxidising member of the Gammaproteobacteria). The alkene MOs have moderate sequence diversity, as shown by the 50% aa identity value used in our analysis here as the cutoff for group membership. Due to the small number of enzymes in group 4, clustering via CD-HIT analysis was not done, and thus Figure 1.3.4 represents all the predicted alkene MOs in sequence databases at the time of

writing. The details of the alkene MOs that have been experimentally characterised are found in Table S2.

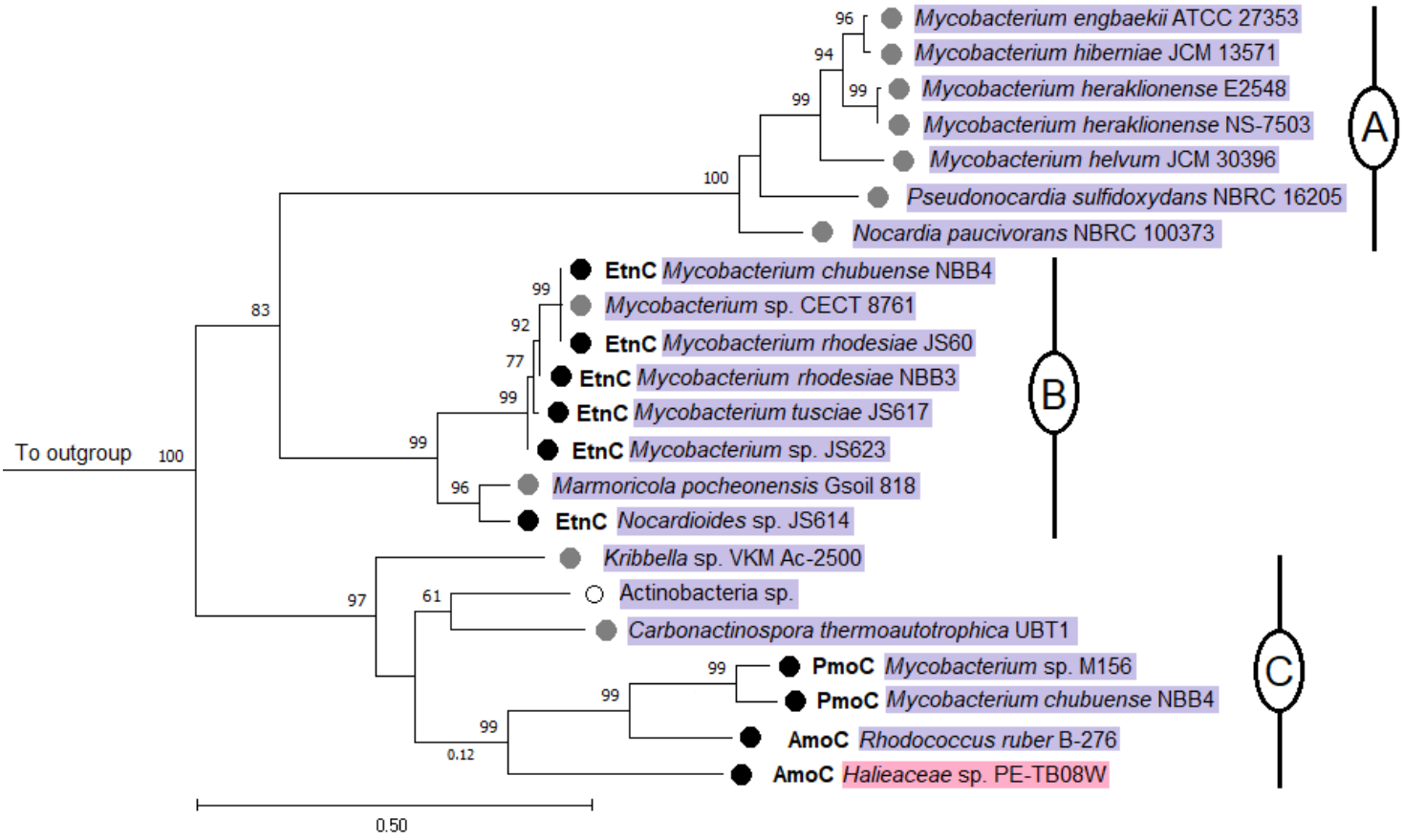


Figure 1.3.4. Evolutionary relationships of the group 4 SDIMOs.

The maximum likelihood tree was made from a trimmed alignment of alpha-subunit sequences (497 aa length), with numbers at nodes indicating percentage bootstrap values. Bootstrap values less than 50% are not shown. Scale bar units are amino acid substitutions per site. ZmoC of *Solimonas soli* was used as the outgroup to root the tree. Dots preceding organism names indicate the following: black circle = characterised SDIMO, grey circle = uncharacterised SDIMO from pure culture genome, white circle = uncharacterised SDIMO from MAG. Coloured shading behind organism names indicates their phylum, as follows: purple = Actinomycetota, pink = Pseudomonadota.

Three distinct subgroups (A-C) within the alkene MOs can be defined. These are supported by bootstrap values >97% (Figure 1.3.4). These correspond quite well to the physiological substrates of the host cells, where this is known. Subgroup A contains no experimentally characterised MOs and is very divergent from subgroups B and C, suggesting that these enzymes may have a different substrate range and

physiological role to the known alkene MOs. Subgroup B contains the ethene MOs (EtnABCD). These have several characterised members in the genera *Mycobacterium* and *Nocardioides* (Coleman and Spain, 2003b, Mattes *et al.*, 2005, Chuang and Mattes, 2007, McCarl *et al.*, 2018). Subgroup C contains the propene MOs (AmoABCD, PmoABCD), with characterised members in *Mycobacterium*, *Rhodococcus*, and *Halieaceae*. The alkene MOs have been characterised primarily by heterologous expression and reverse-transcription quantitative PCR, but these enzymes are less well-studied than the SDIMOs in groups 1, 2 and 3. For example, no knockout mutants have been made, only one of the proteins has been purified (AmoABCD from *Rhodococcus ruber* B-276 (Miuran and Dalton, 1995, Gallagher *et al.*, 1997)), and no crystal structures are available. More work is needed on both the partly characterised group 4 SDIMOs and the unstudied members of this family in order to better understand their ecological significance and potential applications. In the latter case, this is important due to the uniquely high stereoselectivity of oxidations catalysed by the alkene MOs (Cheung *et al.*, 2013).

1.3.5. The group 5 SDIMOs: The propane-2-monooxygenases and relatives

The group 5 SDIMOs (Figure 1.1.2, Figure 1.3.5) are commonly referred to as the propane MOs, or more correctly, the propane-2-MOs (Pr2MOs), but their physiological substrate range covers diverse chemical families including tetrahydrofuran (THF), 1,4-dioxane, acetone, methylethylketone, and *N*-nitrosodimethylamine (Table S2). The group 5 SDIMO operons are composed of 4 genes arranged as follows: α-hydroxylase, reductase, β-hydroxylase, coupling protein. These are represented by a very large number of sequences found in databases (>1000, Table 1.2.1) and some members are moderately well-characterized. The group 5 SDIMOs have high intra-group sequence diversity, as shown by the 40% aa identity value used as a cutoff for group membership and the relatively low clustering cutoff (80%) for CD-HIT analysis (Table 1.2.1). The details of experimentally characterised group 5 SDIMOs can be found in Table S2.

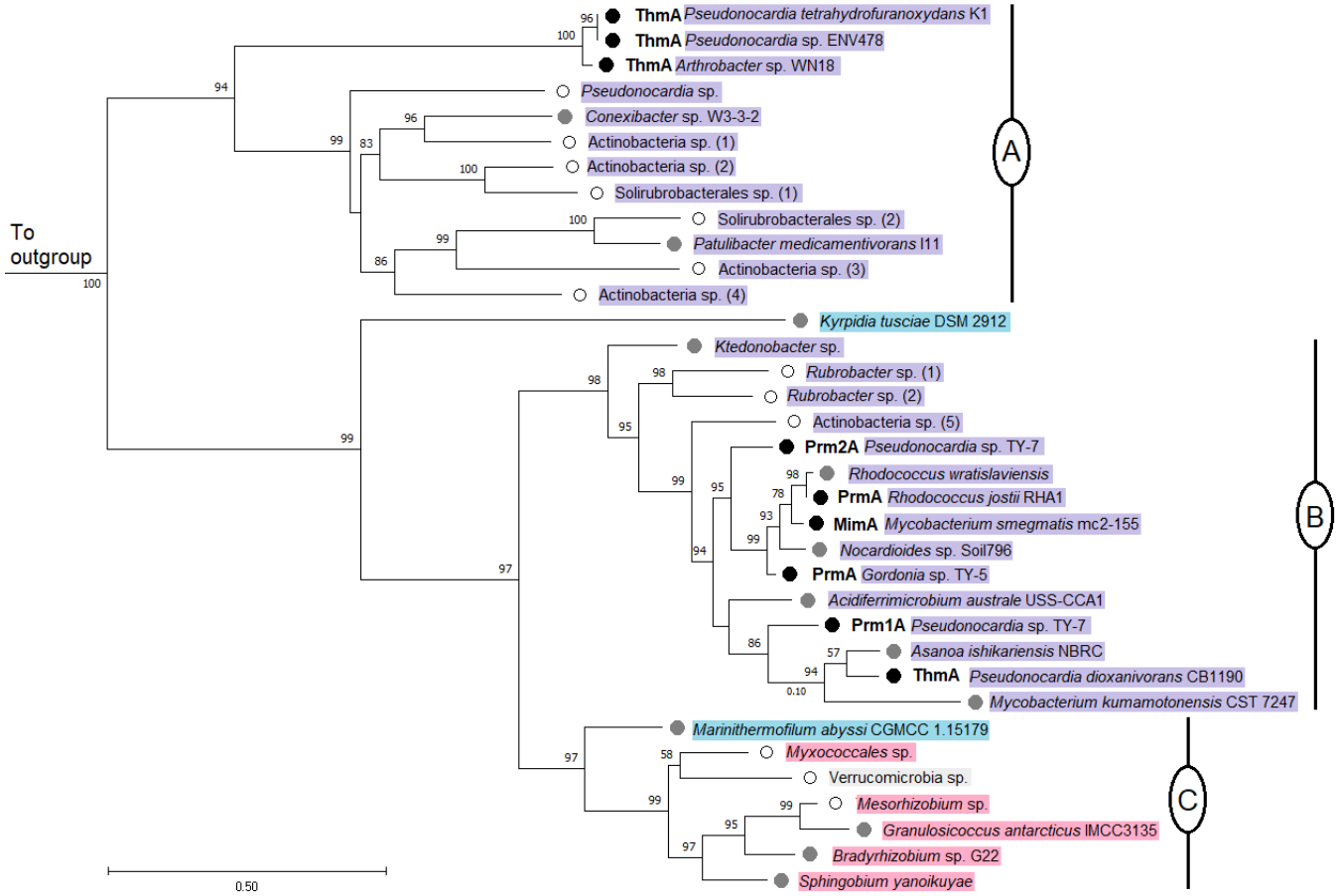


Figure 1.3.5. Evolutionary relationships of the group 5 SDIMOs.

The maximum likelihood tree was made from a trimmed alignment of alpha-subunit sequences (439 aa length), with numbers at nodes indicating percentage bootstrap values. Bootstrap values less than 50% are not shown. Scale bar units are amino acid substitutions per site. ZmoC of *Solimonas soli* was used as the outgroup to root the tree. Dots preceding organism names indicate the following: black circle = characterised SDIMO, grey circle = uncharacterised SDIMO from pure culture genome, white circle = uncharacterised SDIMO from MAG. Coloured shading behind organism names indicates their phylum, as follows: purple = Actinomycetota, pink = Pseudomonadota, light blue = Bacillota, grey = other.

The group 5 SDIMOs are divided here into 3 subgroups (A-C) for the purposes of discussion. Subgroup A contains predominantly sequences from Actinomycetota MAGs, along with three characterised THF-oxidising enzymes (ThmADBC) from *Pseudonocardia* and *Arthrobacter*. Subgroup B is also comprised exclusively of enzymes from Actinomycetota and includes all the characterised Pr2MOs (typically annotated PrmABCD) and one THF-oxidising enzyme (ThmADBC). Subgroup C is found primarily in Pseudomonadota, but also includes one Bacillota host and one Verrucomicrobia host. This subgroup does not contain any experimentally

characterised enzymes. Like all the previously discussed SDIMO groups, there is a large knowledge gap between sequences and functions within the group 5 SDIMOs. In the case of the subgroup B enzymes, this is especially intriguing since they are found in the genomes of many Actinomycetota, regardless of their isolation substrate. Therefore, acetone or related ketones are hypothesised to be the usual physiological substrates of these enzymes, since ketones are common metabolites in many bacteria, unlike hydrocarbons.

1.3.6. The group 6 SDIMOs: The propane-1-monooxygenases and relatives

The group 6 SDIMOs (Figure 1.1.2, Figure 1.3.6) are commonly referred to as propane MOs, or more correctly, propane-1-MOs (Pr1MOs). While their primary substrate is propane (Kotani *et al.*, 2006, Masuda *et al.*, 2012b), the physiological substrates of the group 6 SDIMOs also include *n*-butane (Kotani *et al.*, 2006), 1,4-dioxane, and tetrahydrofuran (THF) (He *et al.*, 2017) (Table S2). A key functional difference between the group 5 and group 6 SDIMOs is that the former oxidise the subterminal carbon of alkanes, whereas the latter oxidise the terminal carbon (Kotani *et al.*, 2006). The group 6 SDIMO operons are composed of 4 genes arranged as follows: α-hydroxylase, β-hydroxylase, coupling protein, reductase. These genes are not common in sequence databases (44 total sequences, Table 1.2.1) and they have low sequence diversity as suggested by the high CD-HIT clustering cutoff (90%) used here for tree construction (Table 1.2.1).

The properties and significance of the group 6 SDIMOs are not well-understood, since only a small number has been experimentally characterised, and even in those cases, only limited evidence for their biochemistry and physiology has been obtained. The details of the Pr1MOs that have been characterised can be found in Table S2. Unlike the group 5 SDIMOs, which tend to be chromosomal, implying a place in the core metabolism of the hosts, the group 6 enzymes in *Mycobacterium* sp. NBB4 (Coleman *et al.*, 2011), *Mycobacterium* sp. ELW1 (Kottegoda *et al.*, 2015), *Mycobacterium gadium* IBE100 (Helbich *et al.*, 2023), and *Mycobacterium paragordonae* IBE200 (Helbich *et al.*, 2023) are encoded on large plasmids. This is consistent with a role for the group 6 SDIMOs in oxidising less-frequently encountered substrates such as gaseous alkanes.

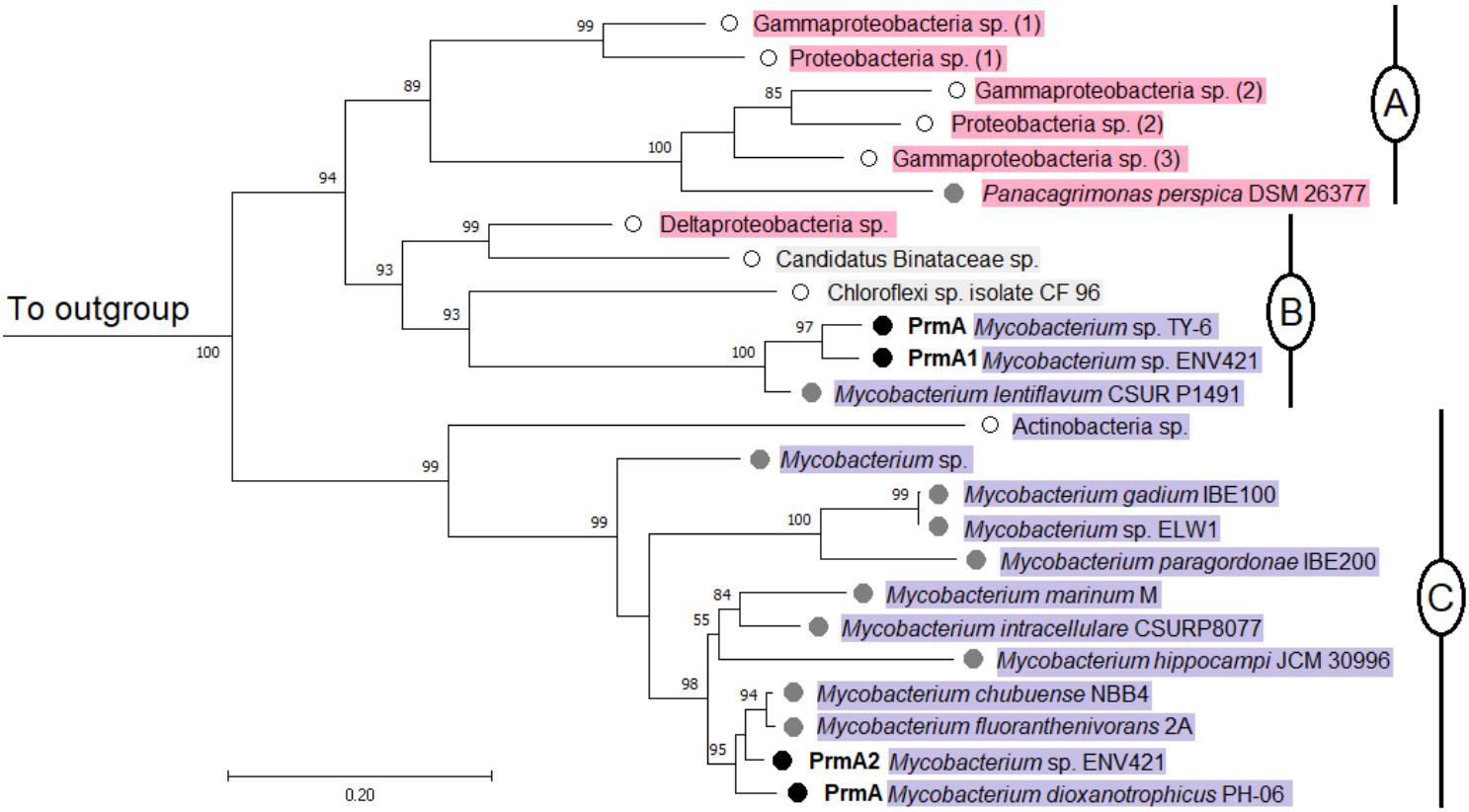


Figure 1.3.6. Evolutionary relationships of the group 6 SDIMOs.

The maximum likelihood tree was made from a trimmed alignment of alpha-subunit sequences (459 aa length), with numbers at nodes indicating percentage bootstrap values. Bootstrap values less than 50% are not shown. Scale bar units are amino acid substitutions per site. ZmoC of *Solimonas soli* was used as the outgroup to root the tree. Dots preceding organism names indicate the following: black circle = characterised SDIMO, grey circle = uncharacterised SDIMO from pure culture genome, white circle = uncharacterised SDIMO from MAG. Coloured shading behind organism names indicates their phylum, as follows: purple = Actinomycetota, pink = Pseudomonadota, grey = other.

Three subgroups (A-C) of group 6 SDIMOs have been defined here (Figure 1.3.6) to facilitate discussion. Subgroup A contains predominantly sequences from Pseudomonadota MAGs and does not contain any characterised enzymes. Subgroup B contains a mixture of Actinomycetota (mostly Mycobacteria), an uncharacterised Deltaproteobacteria sp., and most intriguingly, sequences from Candidatus Binataceae and Chloroflexi. Both these phyla are known to harbour putative hydrocarbon degraders (Chuvochina *et al.*, 2019, Dong *et al.*, 2020). Subgroup C is mostly made up of *Mycobacterium* sequences, except for one sequence from an Actinomycetota MAG. Again, it is clear that most of the sequence

diversity of the group 6 SDIMOs represents unexplored functions, especially in subgroup A.

1.3.7. The group 7 SDIMOs: A new subgroup of unclear function and significance

The group 7 SDIMOs (Figure 1.1.2, Figure 1.3.7) make up the most recently identified subgroup and little is known about their biochemistry, physiology, and ecological significance. The first member of this subgroup was identified in 2019 in the genome of *Solimonas soli* (Osborne and Haritos, 2019). The *S. soli* SDIMO from here onwards will be named ZmoABCD. To date, no study has identified the physiological substrate of the enzyme, and at the time of writing, only 14 group 7 SDIMO sequences can be identified in sequence databases. All of these have the same operon structure as the group 4 SDIMOs (β-hydroxylase, coupling protein, α-hydroxylase, reductase). The group 7 SDIMOs form a coherent clade which is distant from the existing SDIMO groups (Figure 1.1.2), justifying the proposal for a new SDIMO group.

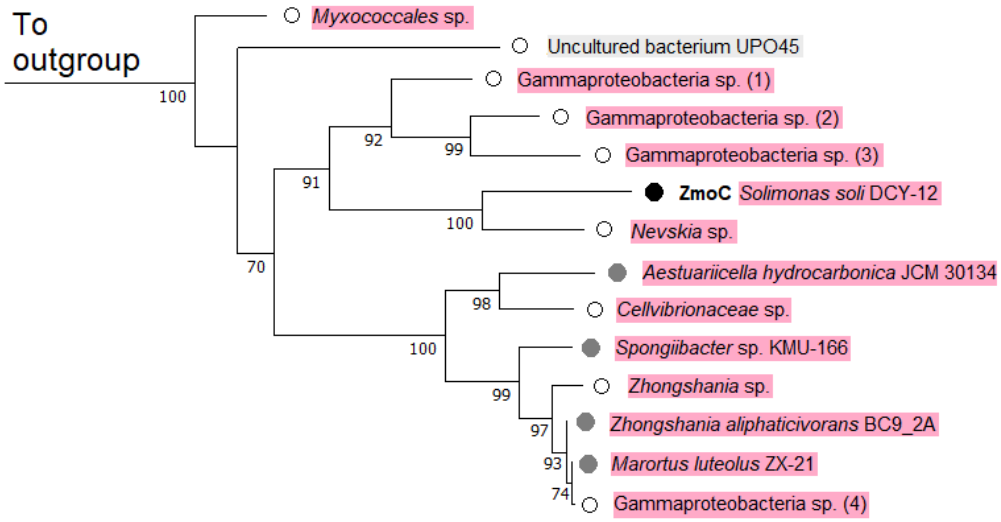


Figure 1.3.7. Evolutionary relationships of the group 7 SDIMOs.

The maximum likelihood tree was made from a trimmed alignment of alpha-subunit sequences (455 aa length), with numbers at nodes indicating percentage bootstrap values. Bootstrap values less than 50% are not shown. Scale bar units are amino acid substitutions per site. PoxD of *Ralstonia* sp. E2 was used as the outgroup to root the tree. Dots preceding organism names indicate the following: black circle = characterised SDIMO, grey circle = uncharacterised SDIMO from pure culture genome, white circle = uncharacterised SDIMO from MAG. Coloured shading behind organism names indicates their phylum, as follows: pink = Pseudomonadota, grey = other.

The group 7 SDIMOs are found almost exclusively in Gammaproteobacteria (Figure 1.3.7), with one exception in the Myxococcota, and one in a taxonomically indeterminate MAG. As mentioned above, the primary substrate of this SDIMO group remains unknown, as do most of the relevant properties of the enzymes, such as the nature of the inducer(s), their protein structures, and their associated metabolic pathways. One interesting trend in this subgroup is that the majority of the bacterial hosts are marine bacteria, which suggests that the physiological substrate(s) of this SDIMO type may be specific to this habitat.

1.4. SDIMOs in biodegradation and bioremediation

Biodegradation is the process by which organic compounds are broken down by microbes, while bioremediation refers to the technology of harnessing these biodegradation processes to cleanup organic pollutants from a contaminated site (Wackett and Hershberger, 2001). SDIMO-containing bacteria are popular candidates for pollutant cleanup due to their wide substrate range. The use of SDIMOs for bioremediation has previously been reviewed, especially with respect to halogenated alkanes and alkenes (Shennan, 2005), halogenated aromatic hydrocarbons (Pimviriyakul *et al.*, 2020) and cyclic ethers (Skinner *et al.*, 2024). Therefore, only a brief overview will be provided in this section, focussing on some recently discovered SDIMOs that have not been reviewed elsewhere. Refer to Table S2 for specific details of pollutants metabolised by different SDIMOs.

1.4.1. Impacts of cometabolism and physiology on bioremediation

Prior to implementation of an SDIMO for bioremediation, it is helpful to first understand the place of the enzyme in its host's physiology, since a purely biochemical understanding of the enzyme's activity may be insufficient for successful deployment of the bacteria for bioremediation (Skinner *et al.*, 2024). It is especially important to understand the process limitations imposed by cometabolism (metabolism of a compound in a manner that provides no growth benefit (Boethling and Alexander, 1979, Schmidt *et al.*, 1985)). In cometabolic bioremediation processes, the growth substrate competes with the pollutant for the enzyme active

site, and this impacts the enzyme kinetics (Alvarez-Cohen and Speitel, 2001). While high rates of turnover are possible using cometabolic reactions, the lack of carbon and energy yield from the pollutant and the need for an auxiliary substrate causes numerous problems for process implementation and sustainability. The growth substrate may need sequential (rather than continuous) addition (Frascari *et al.*, 2012), a different inducer may also need to be added (Suttinun *et al.*, 2009), and the cometabolic substrate may itself be toxic or may give rise to toxic metabolites (van Hylckama Vlieg and Janssen, 2001, Halsey *et al.*, 2005). For more detail, refer to a recent review (Skinner *et al.*, 2024) which covers issues of cometabolism and SDIMOs in depth.

1.4.2. SDIMO-mediated biodegradation of halogenated hydrocarbons

Halogenated alkanes, alkenes, and aromatics are common groundwater pollutants and thus there has been much interest in finding and characterising bacteria that can biodegrade these chemicals (Shennan, 2005, Pimviriyakul *et al.*, 2020). Nearly all of the SDIMO groups contain members that are capable of cometabolic oxidation of haloalkenes (Table S2), with some well-studied examples including the PhMO of *Burkholderia cepacia* G4 (Shim and Wood, 2000) and the TMO of *Pseudomonas stutzeri* OX1 (Ryoo *et al.*, 2001) (Table S2). The alkene monooxygenases (EtnABCD, group 4) found in mycobacteria are noteworthy, since these enable growth on a chlorinated ethene (VC), not just cometabolism. One such strain, *Mycobacterium* sp. JS623, provided the first evidence for the specific mutations associated with adaptation to growth on chlorinated hydrocarbons like vinyl chloride (VC) from a cometabolic ancestor (Jin *et al.*, 2010). The mycobacteria have proved a rich source of unusual SDIMOs. *Mycobacterium chubuense* NBB4, for example, has 4 SDIMOs (Coleman *et al.*, 2011), including SmoXYC1B1Z, an unusual group 3 SDIMO which can attack VC and 1,2-dichloroethane (Martin *et al.*, 2014). This enzyme is notable for being the only group 3 SDIMO to date that is amenable to heterologous expression in a generalist bacterial host (see following section on heterologous expression, Section 1.5).

Two interesting, recently discovered SDIMOs capable of biodegradation of chlorinated ethenes are the TmoABCDEF (group 1) and PrmABCD (group 5) in

Azoarcus sp. DD4 (Deng *et al.*, 2020, Li *et al.*, 2021). These SDIMOs are unusual because they are found in a single bacterium and both have activity against propane. This is unique for a group 1 SDIMO. *Azoarcus* sp. DD4 has been used for bioremediation of trichloroethylene (TCE) in groundwater, by employing an anaerobic consortium to convert TCE to *cis*-dichloroethylene (cDCE) and VC, and then using *Azoarcus* to cometabolise cDCE and VC during growth on propane (Li *et al.*, 2021). The only SDIMO that is known to be involved in growth on a halogenated aromatic is Tbm (group 2 SDIMO) from *Pseudomonas* sp. JS150, which utilises *p*-chlorobenzene (Spain and Nishino, 1987). However, other group 2 SDIMOs can readily oxidise halogenated phenols by cometabolism (e.g. phenol-grown cells of *Pseudomonas putida* CF600, which will attack 4-chlorophenol (Nowak and Mrozik, 2016)).

1.4.3. SDIMO-mediated biodegradation of cyclic ethers

Cyclic ethers are another major category of groundwater pollutants (Miao *et al.*, 2023). Two of the most well-studied are tetrahydrofuran (THF) and 1,4-dioxane, which have been used as isolation substrates and/or cometabolic substrates for many bacteria that contain group 5 SDIMOs (Table S2). A recent interesting ether-degrading bacterium is *Xanthobacter* sp. YN2, which in contrast contains a group 1 SDIMO (ThmABCDEF), and can utilise 1,4-dioxane, 1,3-dioxane, 1,4-dioxene and THF as growth substrates (Ma *et al.*, 2021) (Table S2). This SDIMO is remarkable since it is the first example of an enzyme from group 1 linked to ether metabolism, and also because it enables growth on a broad range of cyclic ethers. The nomenclature of this enzyme is a difficult issue. The sequence of the *Xanthobacter* sp. YN2 enzyme shows it to be more closely related to Tmo enzymes than to other Thm enzymes, but its substrate range looks more like the latter. This problem occurs in numerous places in the SDIMO family tree (Figure 1.1.2). One must be wary of assumptions about genetic relatedness or enzymatic functions based on SDIMO names, and to critically assess the prior literature when naming new SDIMOs. Therefore, new enzymes are recommended to be named based on sequence identity, not substrate range, since the latter is extremely fluid with SDIMOs.

Another recently reported cyclic ether degrader is *Cupriavidus metallidurans* ZM02 (Ren *et al.*, 2022). This isolate is unusual since it oxidises THF using a group 2 SDIMO, DmpKLMNOP (Table S2). In this case, the enzyme has been correctly named based on its sequence identity to dimethylphenol (Dmp) monooxygenases, and it is notable that *C. metallidurans* ZM02 does also grow on phenol. The activity of DmpKLMNOP towards THF was confirmed by knockout experiments (these knockout strains lost the ability to grow on both THF and phenol) and also by heterologous expression in *Cupriavidus* strain JMP134. It was remarkable that the recombinants in the latter experiments could not only degrade THF but could also grow on this compound. This is a powerful reminder that acquisition of SDIMO genes alone can sometimes modify the growth substrate range of the host and impact its fitness in different environments – these considerations are relevant for understanding the selective forces impacting HGT of SDIMO genes (Section 1.6).

Arthrobacter strain WN18 uses a group 5 SDIMO (ThmADBC) to grow on THF and co-metabolically degrade 1,4-dioxane (Wang *et al.*, 2021a, Wang *et al.*, 2021b) (Table S2). Strain WN18 is typical of most previously described dioxane degraders, which are Actinomycetota containing group 5 SDIMOs that use THF or propane as primary substrates. It is worth noting that to date, relatively few organisms have been shown to grow on dioxane (Inoue *et al.*, 2016, He *et al.*, 2017). THF-utilising bacteria provide a good example of the challenges associated with cometabolic biodegradation processes, since the issues relate not just to the supply of carbon and energy, but also the need for THF as an inducer of the genes in the wild-type bacteria. It is notable that the environmental tolerances of the *Arthrobacter* strain WN18 isolate (pH, temperature, salinity, and resistance to other pollutants) were characterised (Wang *et al.*, 2021a). This was vital, as these properties of a potential inoculant are important in determining the success of bioremediation.

Mycobacterium strain PH-06 is another interesting recent isolate (He *et al.*, 2017, Deng *et al.*, 2018). This is distinguished by its possession of a group 6 SDIMO PrmABCD, which attacks both THF and 1,4-dioxane, and enables growth on both of these ethers as carbon and energy sources (Table S2). While the Prm is upregulated by both THF and 1,4-dioxane, *Mycobacterium* strain PH-06 preferentially metabolises THF. This is in accordance with the preference for THF vs. dioxane seen

in other ether-degraders (Table S2) and is consistent with the fact that THF is a naturally-occurring molecule (Parod, 2014), while 1,4-dioxane is xenobiotic. Rigorous evidence for the involvement of the Prm enzyme in ether metabolism was obtained through heterologous expression of the SDIMO in *Mycobacterium smegmatis* mc^2-155. Such recombinants may be useful for bioremediation, since they allow uncoupling of the pollutant-degrading enzymes from the metabolism of the host organism and enable the use of inducers that do not compete with pollutants for the enzyme active site. These approaches however do bring their own technical, legal and ethical issues (de Lorenzo, 2009, Kelle, 2013, Dvořák *et al.*, 2017).

1.5. Heterologous expression: A key tool for understanding and applying SDIMO activities

The expression of SDIMO genes in an alternative host bacterium (heterologous expression) is a very useful method for testing the function of the encoded enzymes and holds promise for applying them in bioremediation and biocatalysis. Ideally, fast-growing and easy-to-manipulate hosts like *Escherichia coli* are used for heterologous expression experiments (Rosano and Ceccarelli, 2014), but in the case of SDIMOs, this is often not possible, for reasons that are not clear. There is evidence that the phylogenetic distance between the native and heterologous hosts is one predictor of success (a more distant relationship typically causes problems), and it also seems that SDIMO expression causes physiological stresses to the heterologous hosts, either due to the proteins themselves misfolding, or due to reactive oxygen species generated in the active site. Below, some relevant strategies that have enabled successful heterologous expression of SDIMOs are highlighted.

The first factor that affects heterologous expression of SDIMOs is the phylogenetic relationship between the native and heterologous hosts, with greater success observed with closely related hosts. There are many reports of successful expression of group 1 and 2 SDIMOs using standard cloning vectors in *E. coli* (Shields *et al.*, 1995, Bertoni *et al.*, 1996, Horinouchi *et al.*, 1999, Tao *et al.*, 2004). These PHs and TMOs are harboured by hosts such as *Pseudomonas*, *Burkholderia*, and *Ralstonia*, which are Gammaproteobacteria like *E. coli*. These group 1 and 2 SDIMOs can also be successfully expressed in *Pseudomonas* hosts (Kukor and

Olsen, 1990, Norlund *et al.*, 1990, Bertoni *et al.*, 1996) (see Table S2 for more details). Similar patterns are seen in other SDIMO groups, for example, the group 5 SDIMOs from *Pseudonocardia tetrahydrofuranoxydans* K1 and *Pseudonocardia dioxanivorans* CB1190 can be expressed in *Rhodococcus jostii* RHA1 (Sales *et al.*, 2013). In these cases, both native and expression hosts are Actinomycetota. The host phylogeny effects are at least partly due to codon usage differences (Gustafsson *et al.*, 2004), and codon optimisation of the target genes can therefore help to improve expression. This was employed to good effect with the group 5 SDIMO MimABCD of *Mycobacterium smegmatis*, enabling expression in *E. coli* (Furuya *et al.*, 2013) (Table S2), although co-expression of chaperone proteins (see below) was also needed. One recent advance that has yet to be applied to the case of SDIMOs is the use of codon harmonisation (Angov *et al.*, 2008) rather than optimisation. In this approach, codons with similar rarity in both native and expression host are used rather than just the most rapidly translated codons. It will be interesting to see if this approach helps SDIMO expression.

Another factor that may impede the success of SDIMO expression is the physiological stress that oxygenase enzymes place on the host cell. This reflects at least two distinct underlying problems, firstly, the challenge of correctly assembling a large multi-subunit enzyme which requires numerous cofactors (Izzo *et al.*, 2011, McCarl *et al.*, 2018), and secondly, the generation of damaging reactive oxygen species in the active sites of oxygenases (Lee, 1999, Sazykin *et al.*, 2019, Bopp *et al.*, 2022). One approach to combat these problems is to express the SDIMO in a host that already has its own native MO, e.g. sMMO of *Methylosinus* was successfully expressed in a pMMO-expressing host (Lloyd *et al.*, 1999), and the SMO from *Mycobacterium chubuense* NBB4 was able to be expressed in *M. smegmatis* mc^2-155 (Martin *et al.*, 2014) (Table S2). Using an MO-containing host is not ideal, however, due to potential difficulties in assigning activities to the relevant MO. In addition, the heterologous expression of MOs in MO-containing hosts while impressive may not be viable as an avenue for deployment of these enzymes, as the hosts are likely less amenable to *in situ* deployment or industrial processes. Another approach is to co-express chaperone proteins that aid protein folding. This approach is informed by the presence of chaperone genes next to many MO gene clusters e.g. *mimG* near *mimABCD* in *M. smegmatis* mc^2-155 (Furuya *et al.*, 2012), and *bmoG*

near *bmoXYBZDC* in *Thauera butanivorans* (Kurth *et al.*, 2008). Co-expression of chaperones has enabled heterologous expression of Actinomycetota SDIMOs in Pseudomonadota hosts (e.g. EtnABCD from *M. chubuense* NBB4 in *P. putida* KT2440 (McCarl *et al.*, 2018), and MimABCD of *M. smegmatis* mc^2-155 in *E. coli* (Furuya *et al.*, 2013)) (Table S2), and the notoriously difficult to heterologously express sMMO in *E. coli* (sMMO from *Methylomonas methanica* MC09) (Zill *et al.*, 2022).

The successes described above, and other examples found in Table S2, have enabled important advances in understanding of SDIMO function and laid the foundations for more detailed molecular work. For example, the ease of expression of the PHs and TMOs in *E. coli* has enabled site-directed mutagenesis and directed evolution studies to create enzymes with altered kinetics and substrate ranges (Canada *et al.*, 2002, Vardar and Wood, 2004, Vardar and Wood, 2005). The co-expression of chaperone proteins and the use of codon optimisation strategies have enabled the SDIMOs from Actinomycetota to be expressed in more industrially useful and genetically tractable hosts like *E. coli* and expanded the potential applications of these enzymes. Another general method used to reduce the stress of MO expression on heterologous hosts was the use of low copy number vectors (Yen *et al.*, 1991). Despite all these advances, there still appear to be unknown factors that are preventing heterologous expression of some SDIMOs (especially sMMO) in ‘workhorse’ bacteria, and even SDIMOs that are amenable to heterologous expression in *E. coli* give recombinants with much lower activities than the original host cells (Martin *et al.*, 2014, McCarl *et al.*, 2018). Strategies such as transposon mutagenesis or TraDIS (Cain *et al.*, 2020) could be used to pinpoint these unknown factors in the native hosts, as a prequel to their co-expression alongside the SDIMOs in heterologous hosts.

1.6. Horizontal gene transfer

Many studies have documented the importance of mobile genetic elements (MGEs) such as plasmids and transposons in the evolution of pollutant-degrading bacteria (Tan, 1999, Top *et al.*, 2002, Top and Springael, 2003), but less is known about the specific HGT events that have involved SDIMO genes. Understanding these events

may help us piece together the missing links in the story of how SDIMO genes have moved between different taxa and evolved different functions suited to different environmental niches. An early seminal review of SDIMO evolution (Leahy *et al.*, 2003) drew several key conclusions. Firstly, the α and β subunits are paralogs (i.e. duplicated from a common ancestor), secondly, the reductase subunits share ancestry with oxidoreductases from diverse oxygenase families, thirdly, the common ancestor of the SDIMOs most likely had a very broad substrate range (both alkenes and aromatics), and finally, HGT played an important role in SDIMO evolution, both within gene clusters (acquisition and shuffling of subunits) and between taxa (movement of whole operons between diverse hosts).

The conclusions from the older literature regarding HGT of SDIMOs are still valid, and it is clear from the examination of the Figures in this chapter that the SDIMO genes have moved over large taxonomic distances (i.e. Pseudomonadota to Actinomycetota) numerous times. This poses questions about the mechanisms of HGT, since the plasmid types that carry SDIMOs are traditionally considered to be specific to either Gram-negative bacteria (e.g. IncP-2 for PhMOs (Bartilson *et al.*, 1990)) or Gram-positive bacteria (e.g. large linear plasmids for alkene MOs (Saeki *et al.*, 1999)). Some resolutions to this conundrum may be found in recent studies which suggest that firstly, novel MGEs with very broad host ranges exist in yet-to-be-cultured environmental bacteria (Smalla *et al.*, 2015) and secondly, that well-studied plasmids may have broader host ranges than previously believed, when these are tested using more innovative and/or systematic approaches (Klümper *et al.*, 2015, Pesce *et al.*, 2019).

The group 1 SDIMOs in particular appear to have been subject to very extensive HGT, since the phylogeny of their α subunit sequences is very incongruent with the phylogeny of their bacterial hosts (Figure 1.3.1). A good example of this can be seen in the isoprene-oxidising enzymes from *Xanthobacter* Py2 (XamoABCDEF) and *Rhodococcus* AD45 (IsoABCDEF), which are closely related (70% aa identity in α subunit) despite the hosts belonging to different phyla (Pseudomonadota vs. Actinomycetota). In both these cases, the SDIMOs are plasmid-encoded (Zhou *et al.*, 1996, Crombie *et al.*, 2015), but little is known of the biology of these plasmids.

There is indirect evidence that large linear catabolic plasmids may be able to move between different phyla (Fetzner *et al.*, 2007) but specific experiments to test this are lacking. The reasons for the apparent higher mobility of the group 1 SDIMOs relative to the other groups are not known, although it could be speculated that this is related to their relative ease of heterologous expression. There is also some evidence for inter-phylum transfer of group 2 SDIMOs in the group 1 SDIMO phylogeny tree (Figure 1.3.2), e.g. the PhMOs from *Acinetobacter* strain S13 and *Arthrobacter* strain W1 are close neighbours (85% aa identity), but most HGT with the group 2 SDIMOs seems to be occurring over smaller taxonomic distances (e.g. between different families and classes within the Pseudomonadota).

While the sMMO genes (group 3 SDIMOs) are considered to be stable chromosomal features of methanotrophs, these enzymes are likely to have had a complex evolutionary history involving numerous HGT events (Osborne and Haritos, 2018). It is especially notable that the more 'divergent' group 3 SDIMOs (i.e. non-sMMO-like) discovered in recent years tend to be encoded on plasmids (Martin *et al.*, 2014, Zou *et al.*, 2021). There is evidence for some major taxonomic divides within the deeper branches of group 3 SDIMO tree (Figure 1.3.2), where it appears that these genes have moved between the Actinomycetota and Pseudomonadata phyla on more than one occasion. It seems very likely from inspection of this tree that the sMMOs evolved from Smo-like ancestors that oxidised larger substrates. Notably, this still allows for the hypothesis that the methane-oxidising enzymes were moved via HGT into methanol-utilising hosts to give rise to the first aerobic methanotrophs (Kang *et al.*, 2019).

A case for HGT in the group 4 SDIMOs can also be made, but in a different way to the groups discussed above. All but one of the known group 4 SDIMOs are found in Actinomycetota (Figure 1.3.4), with only one case of apparent long-range HGT, i.e. AmoABCD in *Halieaceae* sp. PE-TB08W (Gammaproteobacteria), which has 68% aa identity in the alpha subunit to the AmoC from *Rhodococcus* B-276. However, three of the four experimentally characterised alkene MOs are known to be plasmid encoded (Saeki *et al.*, 1999, Chan Kwo Chion *et al.*, 2005, Chuang and Mattes, 2007, Coleman *et al.*, 2011), and there is indirect evidence based on pulsed-field gel electrophoresis experiments that many of the others are too (Coleman and Spain,

2003a). The genome sequences of a set of ethene- and VC-oxidizing mycobacteria further indicate that the alkene assimilation genes are found on large plasmids or plasmid-like contigs in nearly all cases (strains *Nocardioides* sp. JS614 (NCBI Reference Sequence: NC_008697.1), *Mycobacterium rhodesiae* JS60 (NCBI Accession Number: AY243034), *Mycobacterium rhodesiae* NBB3 (Reference Sequence: NC_016604.1), *Mycobacterium chubuense* NBB4 (NCBI Reference Sequence: NC_018022.1)), although knockout or curing experiments are needed to confirm this link. The picture that emerges of the group 4 SDIMOs is that they are highly mobile within the Actinomycetota thanks to the activities of large catabolic plasmids and are also more rarely involved in HGT events over greater taxonomic distances. Much more work is needed to understand the biology of the large catabolic plasmids found in the Actinomycetota, which are poorly studied compared to corresponding catabolic plasmids in Gram-negative bacteria.

As mentioned in Section 1.3.5, group 5 SDIMOs are common components of the chromosomes of many Nocardioform species, and the corresponding SDIMO tree is dominated by the Actinomycetota (Figure 1.3.5). At first glance therefore, these are not strong candidates for HGT. However, looking at the deeper branches in this tree reveals some taxonomic anomalies suggestive of HGT, including the presence of two Bacillota spp (*Kyrpidia tusciae* and *Marinithermofilum abyssi*), alongside several Pseudomonadota species in subgroup C. It would be especially interesting to further probe the differences between the well-studied subgroup B enzymes from Actinomycetota and the related group C enzymes from Pseudomonadota. None of the latter have been experimentally characterised yet, but their phylogeny suggests they may have a different substrate and/or ecological niche relative to the better-studied Prm and Thm enzymes. At least some of the group 5 SDIMOs are plasmid-borne (Thiemer *et al.*, 2003, Sales *et al.*, 2013), providing a mechanism for HGT in at least a few cases.

There is only very limited literature available on the group 6 SDIMOs. However, it is likely that at least some HGT has occurred in this lineage, between closely related genera. For example, the uncharacterized group 6 SDIMOs in *Mycobacterium gadium* IBE100, *Mycobacterium paragordonae* IBE200, and *Mycobacterium* sp. ELW1 appear to have been recently shared among these species by mobile

elements such as plasmid pELW1-1 (Helbich *et al.*, 2023). The group 6 SDIMO of *Mycobacterium chubuense* NBB4 is plasmid-encoded (NCBI Reference Sequence: NC_018022.1), as is the group 6 SDIMO from *Mycobacterium dioxanotrophicus* PH-06 (Deng *et al.*, 2018). The latter is known to be located between two insertion sequences, which is a significant finding since it provides a mechanism for the SDIMO genes to move between different replicons.

1.7. Summary

This chapter presented a comprehensive and updated phylogenetic analysis of the SDIMOs and supports their division into 7 groups based on the sequence identity of the catalytic alpha subunit. The relationship between enzymes, bacterial hosts, and substrates has been examined in detail, and robust links between these have been identified where possible. It is also important to acknowledge that these relationships are in a state of constant flux due to forces such as HGT. An overview of the issue of HGT as it applies to SDIMOs has also been provided. It was also shown via phylogenetic analyses that HGT is common for SDIMO genes, although the patterns of HGT differ among different SDIMO groups. Heterologous expression as a key tool for understanding the functions and significance of SDIMOs have also been discussed. Heterologous expression remains the best approach for investigating the substrate range of these enzymes, since this method separates SDIMOs from the genome of the original host (which may contain multiple monooxygenases) and allows access to enzymes from genomes and metagenomes in the absence of information about growth substrates or inducers. However, the importance of native expression systems and their role in our understanding of the basic assembly of SDIMO subunits and cofactors that are required by the SDIMO in the context that is physiologically relevant must be acknowledged. Heterologous expression and native expression studies should therefore be utilised in tandem to maximise our understanding of SDIMOs where possible. Recent examples of SDIMOs used for pollutant biodegradation have also been highlighted, and a comprehensive list of these biodegradative activities can be found in Table S2. In the cases of both chlorinated ethenes and cyclic ethers, SDIMO-containing bacteria are still the lead candidates for bioremediation efforts under aerobic conditions.

Several large knowledge gaps in the SDIMO literature have become apparent while compiling this introduction, leading to the proposal of several priority research areas. Perhaps most urgent is the need to experimentally characterise representatives of the group 7 SDIMOs and the deep-branching and thus-far-uncharacterised SDIMO lineages found in genomes and metagenomes (e.g. subgroups 1A, 4A, 5C, 6A). These enzymes may have significant environmental impacts and/or useful biotechnology applications. Advances in high-throughput synthetic biology methods

may facilitate this work (Casas *et al.*, 2024), by allowing the rapid construction and screening of large numbers of recombinant clones. These possibilities depend to some extent on closing the second major knowledge gap, which is to understand why heterologous expression of some SDIMOs is so difficult, and to pinpoint the factors in the wild-type hosts that enable high-level expression of functional SDIMO enzymes. The last knowledge gap worth mentioning is the ongoing lack of good genetic tools and methods for making knockouts and other modifications in environmental bacteria. Recent advances in CRISPR (Burbano *et al.*, 2024) and similar technologies combined with the increased availability and affordability of DNA synthesis may yield breakthroughs on this front.

1.8. Aims

The main aim of this study is to identify the canonical substrate of the *Solimonas soli* group 7 SDIMO ZmoABCD. Several approaches have been devised to achieve this:

1. Heterologous expression of the *zmoABCD* operon and characterisation of the resulting enzymes,
2. Knockout of the *zmoABCD* operon in *S. soli*, and
3. Construction of a *zmo* bioreporter construct in *S. soli*.

These experiments are designed to work in tandem to characterise the substrate range of the novel SDIMO harboured by *S. soli*. Additionally, these experiments will expand our understanding of SDIMO heterologous expression, the genetic manipulation of *S. soli*, and the regulation of the ZmoABCD cluster in *S. soli*.

Chapter 2 – General Methods and Materials

The methods described in this chapter are common to the experimental procedures in the later chapters. Methods and materials specific to those chapters are described in their respective sections.

2.1. General Methods

Chemicals were purchased from Sigma-Aldrich/MERCK or ChemSupply and purity of chemicals was >99.0%. Media and solutions were sterilised by autoclaving at 121 ºC, 101.3 kPa for 20 minutes, unless stated otherwise. Ultrapure water was obtained from Milli-Q or Sartorius systems. For solid medium, agar was added to 1.5% (w/v) before sterilisation. If required, the pH of media and solutions were adjusted with 1 M HCl, 1 M NaOH, 1 M K_2HPO_4, or 1 M KH_2PO_4 depending on the nature of the solution before sterilisation. Solutions that contain >50% (v/v) ethanol were not sterilised.

2.2. Cultivation of bacteria

2.2.1. Growth conditions of bacteria used in this study

All *Escherichia coli* strains, *Pseudomonas putida* KT2440, and *Mycobacterium smegmatis* mc^2-155 were cultured in aerobic conditions on solid Lysogeny Broth (LB) medium or in LB cultures (with 0.05% (v/v) Tween 80 supplied to *M. smegmatis* mc^2-155 cultures). Liquid cultures were shaken at 200 rpm. *E. coli* cultures were incubated at 37 ºC while *P. putida* cultures were incubated at 30 ºC. *M. smegmatis* cultures were incubated at either temperature. Antibiotics were added to media when required.

Solimonas soli DCY12 was a gift courtesy of Dr. Che Ok Jeon, Chung-Ang University, Seoul. All *S. soli* strains were cultured in aerobic conditions on solid R2A medium or in R2A+ cultures (Table 2.2.1). Liquid cultures were shaken at 200 rpm. *S. soli* cultures were incubated at 30 ºC and antibiotics were added to media when required.

2.2.2. Growth media

Table 2.2.1. Growth media used in this study.

Media	Composition
Lysogeny Broth (LB)	10 g/L tryptone, 5 g/L yeast extract, 5 g/L sodium chloride, pH 7.0.
R2A	0.5 g/L peptone, 0.5 g/L casamino acids, 0.5 g/L yeast extract, 0.5 g/L glucose, 0.5 g/L soluble starch, 0.3 g/L dipotassium phosphate, 0.03 g/L magnesium sulphate, 0.3 g/L sodium pyruvate, pH 7.0.
R2A+	0.5 g/L peptone, 0.5 g/L casamino acids, 5.5 g/L yeast extract, 0.5 g/L glucose, 0.5 g/L soluble starch, 0.3 g/L dipotassium phosphate, 0.03 g/L magnesium sulphate, 0.3 g/L sodium pyruvate, pH 7.0. MOPS buffer added to 10 mM after sterilisation.
Mineral salts medium (MSM)	2.27 g/L dipotassium hydrogen phosphate, 0.95 g/L potassium dihydrogen phosphate, 0.67 g/L ammonium sulphate, adjusted to pH 7.0 H_2SO_4. 0.2% (v/v) trace element solution (See below) added after sterilisation.
Trace element solution	6.37 g/L $Na_2EDTA.2H_2O$, 1.0 g/L $ZnSO_4.7H_2O$, 0.5 g/L $CaCl_2.2H_2O$, 2.5 g/L $FeSO_4.7H_2O$, 0.1 g/L $NaMoO_4.2H_2O$, 0.1 g/L $CuSO_4.5H_2O$, 0.2 g/L $CoCl_2.6H_2O$, 0.52 g/L $MnSO_4.H_2O$, 60.0 g/L $MgSO_4.7H_2O$, adjusted to pH 6.5 using H_2SO_4. Sterilised by filtration through a 0.22 µm nitrocellulose filter. Stored at 4 ºC, wrapped in foil.

2.2.3. Antibiotics and additives

Table 2.2.2. Antibiotics and additives used in this study.

All antibiotics and additives were sterilised by filtration through a 0.22 µm nitrocellulose filter except for rifampicin, which was filtered using a 0.22 µm nylon filter. All additives were stored at -30 ºC, unless stated otherwise.

Antibiotic/additive	Solvent	Stock concentration	Typical working concentration(s)	Additional notes
4-isopropylbenzoic acid (Cumate) solution for *E. coli* and *M. smegmatis*	Ethanol	100 mM	100 µM	Stored at room temperature.
Cumate solution for *P. putida*	50% (v/v) Ethanol	500 mM	1 mM	1 M cumate added to 1 M Tris-base to make a 1:1 solution. Stored at room temperature.

Acetamide	Ultrapure water	1 g/mL	2 mg/mL	Stored at room temperature.
Gentamicin	Ultrapure water	10 mg/mL	10 µg/mL	N/A
Rifampicin	DMSO	20 mg/mL	20 µg/mL, 40 µg/mL, or 400 µg/mL	N/A
Tetracycline	Ethanol	10 mg/mL	10 µg/mL	N/A
Kanamycin	Ultrapure water	50 mg/mL	50 µg/mL or 20 µg/mL	N/A
Tween 80	Ultrapure water	5.0% (v/v)	0.05% (v/v) or 0.1% (v/v)	Stored at 4 ºC.
MOPS (pH 7.0)	Ultrapure water	1 M	10 mM	Stored at 4 ºC.

2.2.4. Bacterial Strains

Table 2.2.3. Bacterial strains used in this study.

Strain	**Genotype**	**Phenotype**	**Use**	**Supplier or Reference**
Escherichia coli TOP10	F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L(Str^R) *end*A1 *nup*G	Chemically competent, Streptomycin resistant.	Most molecular biology work (Cloning, plasmid propagation, etc.) and as an expression host.	Invitrogen
E. coli CC118λ*pir*	Δ(*ara-leu*) *araD* Δ*lacX74 galEgalK phoA20 thi-1 rpsE rpoB argE*(Am) *recA1* λ*pir*	Chemically competent.	Molecular biology work for plasmids with R6Kγ origin of replication.	Herrero *et al.* (1990)
E. coli S17-1	*hsdR thi pro recA*; RP4 integrated into the chromosome (*kan::Tn*7 *ter::*Mu)	Chemically competent, able to transfer plasmids via conjugation.	Used as conjugation donor.	Simon *et al.* (1983)

E. coli S17-1λ*pir*	*hsdR thi pro recA*; RP4 integrated into the chromosome (*kan::Tn7 ter::Mu*) λ*pir*	Chemically competent, able to transfer plasmids via conjugation.	Used as conjugation donor for plasmids with R6Kγ origin of replication.	de Lorenzo and Timmis (1994)
Solimonas soli DCY12	Wild type	Wild type	Source of group 7 SDIMO, native host growth experiments, and attempted electroporation.	Kim *et al.* (2007)
Solimonas soli DCY12R	As above, but rifampicin resistant	Rifampicin resistant	Host for conjugation and knockout mutant generation.	This study
Solimonas soli 379KO	As DCY12R, with the *zmoABCD* cluster replaced with a *fuGFP* gene.	Rifampicin resistant, *zmoABCD* knockout mutant.	Growth comparisons with wild type, and biosensor.	This study
Solimonas soli 380KO	As 379KO, without the 100 bp putative regulatory region between the *zmo* promoter and *fuGFP* ribosome binding site.	Rifampicin resistant, *zmoABCD* knockout mutant.	Growth comparisons with wild type.	This study
Pseudomonas putida KT2440	*rmo- mod+*	Electrocompetent	Expression host.	Bagdasarian *et al.* (1981)
Mycobacterium smegmatis mc^2-155	Wild type	Electrocompetent, sensitive to kanamycin.	Expression host.	Snapper *et al.* (1990)

2.3. Polymerase chain reactions

Primers were designed using the SnapGene Software Package (Dotmatic) and were analysed using the OligoAnalyzer Tool provided by the Integrated DNA Technologies website (https://www.idtdna.com/calc/analyzer) with the qPCR parameter. The sequence and use of each primer can be found in Table 2.3.1. Primers were purchased from Integrated DNA Technologies Incorporated and Thermo Fisher Scientific and were resuspended to 50 µM in sterilised ultrapure water. PCRs were performed in 20 µL or 50 µL volumes containing commercial master mixes (see following section on polymerases), 1 µM of each primer and 50-250 ng DNA template with thermocycling conditions as follows: 95 ºC initial denaturation for 5 minutes, followed by 25-30 cycles of denaturation at 95 ºC for 30 seconds, annealing for 30 seconds (temperatures in Table 2.3.2), extension at 72 ºC (times in Table 2.3.2), then a final extension at 72 ºC for 5 minutes. The completed reactions were held at 15 ºC in the thermocycler until collection. Either 2X MangoMix (Meridian Bioscience) or Q5® High-Fidelity 2X Master Mix (New England Biolabs) were used for different PCRs as indicated in Table 2.3.2. To screen for correctly assembled plasmids, PCR using whole cells from colonies was employed (i.e. colony PCR). In this case a small portion of each colony picked from solid medium using 10 µL pipette tips was transferred to each PCR tube as the source of template DNA.

2.3.1. Primers

Table 2.3.1. Primers used in this study.

T_m values were calculated by the SnapGene Software Package (Dotmatic), assuming 50 mM Na^+ and 0.25 μM for each primer or 0.5 μM each for two annealed oligomers. Bolded bps in the Sequence column correspond to the restriction site detailed in the "Use" column, and italicised bps correspond to the BsaI cut site. MCS = multiple cloning site.

Name	T_m (ºC)	Sequence (5' → 3')	Use/Binding site	Reference
CFM13	55	CCCGTGATATTGCTGAAGAG	Reverse primer that binds to the *Kan*^{R} gene of pUS250 derived plasmids. Used for junction PCRs and Sanger sequencing of pUS250 derived plasmids.	Nick Coleman, unpublished
MVS48	57	GTTCGGTTCGTAAACTGTAATGC	Forward primer that binds to the Pc promoter upstream of the multiple cloning sites of pUS250 derived plasmids. Used for junction PCRs and Sanger sequencing of pUS250 derived plasmids.	Nick Coleman, unpublished
MVS116	55	AGGTACTGATGATTGGGTCTC	Forward primer that binds to the 5' gBlock adaptor. Used to amplify gBlocks.	Nick Coleman, unpublished
MVS117	54	TCGTAGCTGAACAGGTCTC	Reverse primer that binds to the 3' gBlock adaptor. Used to amplify gBlocks.	Nick Coleman, unpublished
LFE3	57	AAAGCTAGCCGCAAAGAGAAAGCAGGTAG	Upstream of the *km*^{R} gene in pBBR1MCS-2 plasmids.	Nick Coleman, unpublished
LFE4	56	AAACTCGAGATCAGAAGAACTCGTCAAGAAGG	Downstream of the *km*^{R} gene in pBBR1MCS-2 plasmids.	Nick Coleman, unpublished
NIY60	58	GGGTGTACGTGCTGATAATGC	Reverse primer that binds to the native *S. soli zmoC* gene. Used for pUS367 and pUS368 junction PCRs and Sanger sequencing.	This study

NIY61	57	CAATTTGCGTGGCTCGG	Forward primer that binds to the native *S. soli zmoD* gene. Used for pUS367 and pUS368 junction PCRs and Sanger sequencing.	This study
NIY62	59	GCGATGCCTTGTCTGACG	Primer that binds to the native *S. soli zmoA* gene. Used for Sanger sequencing.	This study
NIY63	59	GCATAACGCAGCACCAGC	Primer that binds to the native *S. soli zmoA* gene. Used for Sanger sequencing.	This study
NIY64	56	GACTTCCTCACCGTCCTG	Primer that binds to the native *S. soli zmoB* gene. Used for Sanger sequencing.	This study
NIY65	56	AGGCGAAGTTGGAGAGG	Primer that binds to the native *S. soli zmoC* gene. Used for Sanger sequencing.	This study
NIY66	58	GATCCAGTCCGACGAGGC	Primer that binds to the native *S. soli zmoC* gene. Used for Sanger sequencing.	This study
NIY67	60	AGCCGCCGTAGAACTCG	Primer that binds to the native *S. soli zmoC* gene. Used for Sanger sequencing.	This study
NIY68	55	GTTGAGCCATACGGTCAC	Primer that binds to the native *S. soli zmoC* gene. Used for Sanger sequencing.	This study
NIY69	57	TGAGTGCGACTTCTGATGC	Primer that binds to the native *S. soli zmoD* gene. Used for Sanger sequencing.	This study
NIY70	55	AAAAA**CATATG**ATGCTGATCGGGTATGG	Forward primer used to amplify *S. soli* MO directly from genome. Includes **NdeI** site for cloning.	This study
NIY71	59	TTT**AAGCTT**GCGTTCCTCCGTCCG	Reverse primer used to amplify *S. soli* MO directly from genome. Includes **HindIII** site for cloning.	This study
NIY74	58	CCGATAAGAGAAAGGGAGTCCAC	Forward primer that binds to the acetamidase promoter region in pUS116 derived plasmids. Used in pUS368 junction PCRs to verify construction.	This study

NIY75	58	ACATCAGAGATTTTGAGACACAACG	Reverse primer that binds to the *kan*^{R} gene in pUS116 derived plasmids. Used in pUS368 junction PCRs to verify construction.	This study
NIY76	58	GCGGAAGCGTCCCAAG	Forward primer that binds to the *S. soli* putative epoxide hydrolase gene. Used to verify the identity of *S. soli*.	This study
NIY77	58	GCGATCCAAGCCACCAAG	Reverse primer that binds to the *S. soli* putative epoxide hydrolase gene. Used to verify the identity of *S. soli.*	This study
NIY78	56	AAA**AAGCTT**TCGTCACCCATAACAGATACG	Forward primer that binds upstream of the *groES* gene in pUS229. Includes **HindIII** site for cloning.	This study
NIY79	56	TTT**AGATCT**TTGTTTATTTCTGCGAGGTGC	Reverse primer that binds downstream of the *groEL* gene in pUS229. Includes **BglII** site for cloning.	This study
NIY80	57	CAGCAGATTTAGTTTCAACTTCTTTACG	Primer that binds to the *groES* gene in pUS369. Used to verify construct.	This study
NIY81	56	GCGACGGCAACTACG	Primer that binds to the *groEL* gene in pUS369. Used to verify construct.	This study
NIY82	57	ATCGCTGAGATGGGCTTG	Forward primer that binds upstream of *S. soli* group 7 SDIMO. Used to verify KO of SDIMO.	This study
NIY83	58	CCACTGAGCATGACGAACG	Reverse primer that binds upstream of *S. soli* group 7 SDIMO. Used to verify KO of SDIMO.	This study
NIY94	59	AAA**GGTCTC**A*CAAC*CACTCAAAGGCGGTAATACGG	Forward primer used to amplify pEX18Tc. This included a **BsaI** site for cloning (CAAC).	This study
NIY95	56	AAA**GGTCT**CA*GAGA*CATGCGAGAGTAGGGAACTGC	Reverse primer used to amplify pEX18Tc. This included a **BsaI** site for cloning (GAGA).	This study

NIY96	55	AAA**AAGCTT**AGGAGACATGAACGATGAAC	Forward primer that binds upstream of *sacB* in pEX18Tc. This included a **HindIII** site for cloning.	This study
NIY97	55	TTT**GAGCTC**TTATTTGTTAACTGTTAATTGTCCTTGTTC	Reverse primer that binds downstream of *sacB* in pEX18Tc. This included a **SacI** site for cloning.	This study
NIY98	55	ATGACCATGATTACGCCAAG	Forward primer that binds to *lacZ* in pUS375. Used to verify construct.	This study
NIY99	54	AAATGCCGTATGTTTCCTTATATG	Reverse primer that binds to *sacB* in pUS375. Used to verify construct.	This study
NIY100	54	ACCTTTACTTACTCACACTTCG	Forward primer that binds to *sacB* in pUS375. Used to verify construct.	This study
NIY101	54	GGTTTTCCCAGTCACGAC	Reverse primer that binds to *lacZ* in pUS375. Used to verify construct.	This study
NIY102	55	TGTTCTTTCCTGCGTTATCC	Forward primer that binds to MCS in pUS374. Used to verify construct.	This study
NIY103	55	AGCCCCGTATCACTGAG	Reverse primer that binds to 5’ end of gene upstream of *S. soli* SDIMO in pUS374. Used to verify construct.	This study
NIY104	54	TCTTCATTAACGGTCTAGCATAG	Forward primer that binds to Zmo promoter in pUS374. Used to verify construct.	This study
NIY105	55	CGTCGCCATAACCTTCAC	Reverse primer that binds to 5’ end of *fuGFP* in pUS374, pUS379, and pUS380. Used to verify constructs.	This study
NIY106	55	CCAATAGCGTGCTGAGC	Binds to the 3’ end of *fuGFP* in pUS374, pUS379, and pUS380. Used to verify constructs.	This study
NIY107	54	GGAAGAAGGCGATCTCG	Binds to the 5’ end of *cntI* in pUS374. Used to verify construct.	This study

NIY108	54	CTCGCCAACCTCTACAG	Binds to the 3' end of *cntI* in pUS374. Used to verify construct.	This study
NIY109	55	CCTTTCGTTTTATTTGATGCCTG	MCS in pUS374. Used to verify construct.	This study
NIY110	57	TGAAGCTATGCTAGACCGTTAATG	Zmo promoter in pUS374v2.	This study
NIY111	56	AGATACAAAGGAGGGCAAGG	RBS of fuGFP in pUS374v2.	This study
NIY112	54	ACGGCAGGTATATGTGATG	Upstream of *sacB* gene on pEX18Tc plasmids.	This study
NIY113	53	GGATTCTACGCAGACAAAC	Downstream of *sacB* gene on pEX18Tc plasmids.	This study
NIY114	56	AAA**ACTAGT**CGAACAAAGGGCTGGTG	5' end of KO cassette in pUS374v2 and pUS376 (Includes **SpeI** site).	This study
NIY115	60	AAA**CTCGAG**CACTCCACTTGTCAAGCG	3' end of KO cassette in pUS374v2 and pUS376 (Includes **XhoI** site).	This study
NIY116	55	GATGTAACGCACTGAGAAGC	3' end of R6Kγ ori in pKNOCK-Km.	This study
NIY117	57	CCGCTTCCTTTAGCAGC	Downstream of MCS in pKNOCK-Km.	This study
NIY118	54	ATTGATGCCGAGATCGTC	Amidohydrolase gene of S. soli. Just upstream of where the KO cassette starts.	This study
NIY119	54	TGATTCATTGCTAGGTTCCTC	Cupin-domain binding protein gene of S. soli. Just downstream of where the KO cassette ends.	This study
NIY126	55	AAA**TCTAGA**CAGATGATAGACGAGATGCG	5' end of *zmoC* in *S. soli* genome (Incudes **XbaI** site).	This study
NIY127	54	AAA**GTCGAC**GAATGATGGTCTTGTCGGTC	3' end of *zmoC* in *S. soli* genome (Incudes **SalI** site).	This study
NIY128	54	GGATCGCAGGTCGTG	5' end of *zmoC* ** fragment in pUS382.	This study
NIY129	56	AGATGGACGATCTCTCGC	3' end of *zmoC* ** fragment in pUS382.	This study
NIY130	54	CTCTCCAACTTCGCCTTC	5' end of *zmoC* in *S. soli* genome after pUS382 integration.	This study
NIY131	54	ACCGTGATAGAGATCGTAAAAG	3' end of *zmoC* in *S. soli* genome after pUS382 integration.	This study

NIY132	55	AAA**GGATCC**ATGTATCGCCACATCGC	5' end of *zmoA* in *S. soli* genome (Incudes **BamHI** site).	This study
NIY133	56	AAA**GTCGAC**GCCTGCCGCATTTCG	3' end of *zmoA* in *S. soli* genome (Incudes **SalI** site).	This study
NIY134	53	CACATGGCATTGGCAC	5' end of *zmoA* ** fragment in pUS383.	This study
NIY135	55	CGATCAATCTCGTCGTCG	3' end of *zmoA* ** fragment in pUS383.	This study
NIY136	56	CAACGCAATATGAAGAAGTCACC	5' end of *zmoA* in *S. soli* genome after pUS382 integration.	This study
NIY137	54	GACCAGATCGGGATTCG	5' end of *zmoB* in *S. soli* genome after pUS382 integration.	This study
NIY138	56	CCTCACCAATCCATTGACG	Between Zmo promoter and fuGFP.	This study
NIY139	55	TCCACGGCGGTCC	Between Zmo promoter and fuGFP.	This study
NIY140	55	CGTCAATGGATTGGTGAGG	Between Zmo promoter and fuGFP.	This study
NIY141	57	GGACCGCCGTGGAG	Between Zmo promoter and fuGFP.	This study
NIY142	56	AAA**GAATTC**TGAATGGCGAATGGAAATTG	Downstream of pBBR1MCS-2 MCS. Contains **EcoRI** site.	This study
NIY143	54	AAA**AAGCTT**TGGGGTGCCTAATGAGTG	Upstream of pBBR1MCS-2 MCS. Contains **HindIII** site.	This study
NIY144	56	AAA**AAGCTT**GGGCTGCGGATTTATATTGAC	Upstream of zmo promoter in pUS379 plasmids. Contains **HindIII** site.	This study
NIY145	54	AAA**GAATTC**TCGTGACATGCGTTCC	Downstream of fuGFP in pUS379 plasmids. Contains **EcoRI** site.	This study
NIY146	53	CCCAATACGCAAACCG	Backbone of pBBR1MCS-2_v2.	This study
NIY147	53	GCCCTGCCACTCATC	Backbone of pBBR1MCS-2_v2.	This study

2.3.2. PCR conditions

Table 2.3.2. PCR conditions used in this study.

Both 2X Q5 Master Mix and 2X MangoMix are commercially prepared PCR master mixes that include dNTPs, the appropriate buffer and concentration of magnesium ions, and the relevant polymerase. 2X MangoMix also includes a loading dye for agarose gel electrophoresis in its buffer. For use, only the addition of template and primers along with the dilution of the master mix to 1X concentration are required. MCS = multiple cloning site.

Primer pair	**Target**	**Size (bp)**	**Polymerase used**	**Annealing temperature (ºC)**	**Extension time (s)**
MVS116 & MVS117	DNA sequence between the gBlock adaptors.	Variable	2X Q5 Master Mix	Variable	Variable
NIY60 & MVS48	A region between the pUS250 *kan*^{R} gene and the native *zmoA* gene of the *S. soli* MO in pUS367.	191	2X MangoMix	54	20
NIY61 & CFM13	A region between the pUS250 Pc promoter and the native *zmoC* gene of the *S. soli* MO in pUS367.	380	2X MangoMix	54	20
NIY60 & NIY74	A region between the pUS116 acetamidase promoter region and the native *zmoA* gene of the *S. soli* MO in pUS368.	161	2X MangoMix	54	20
NIY61 & NIY75	A region between the pUS116 *kan*^{R} gene and the native *zmoD* gene of the *S. soli* MO in pUS368.	253	2X MangoMix	54	20
NIY70 & NIY71	The *S. soli zmoABCD* cluster.	4112	2X Q5 Master Mix	60	50
NIY76 & NIY77	A section of the putative *S. soli* epoxide hydrolase gene.	493	2X MangoMix	55	40
NIY78 & NIY79	*groES/EL* genes from pUS229.	2059	2X Q5 Master Mix	59	30
NIY61 & NIY80	A region between the *groES* gene and the native *zmoD* gene of the *S. soli* MO in pUS369.	271	2X MangoMix	52	15

NIY73 & NIY81	A region between the pUS250 *kan*^{R} gene and the *groEL* gene in pUS369.	301	2X MangoMix	52	15
NIY82 & NIY83	The *S. soli zmoABCD* region. Used to determine presence of *fuGFP* or *zmo* cluster.	Variable	2X MangoMix	54	30
NIY94 & NIY95	The pEX18Tc backbone.	5656	2X Q5 Master Mix	61	70
NIY96 & NIY97	*sacB* from pEX18Tc.	1436	2X Q5 Master Mix	58	25
NIY98 & NIY99	A region between the *sacB* gene and the *lacZ* gene in pUS375.	242	2X MangoMix	51	15
NIY100 & NIY101	A region between the *sacB* gene and the *lacZ* gene in pUS375.	281	2X MangoMix	51	15
NIY102 & NIY103	A region between the *ori* region and the CODDBKIM_00035 gene in pUS374.	342	2X MangoMix	51	15
NIY104 & NIY105	A region between the Zmo promoter region and the *fuGFP* gene in pUS374, pUS379, pUS380, pUS384-388.	Variable	2X MangoMix	51	15
NIY106 & NIY107	A region between the *fuGFP* gene and the *cntI* gene in pUS374.	261	2X MangoMix	51	15
NIY108 & NIY109	A region between the *cntI* gene and the pEX18Tc_v2 backbone in pUS374.	224	2X MangoMix	51	15
NIY110 & NIY111	Amplified the pUS374 backbone, excluding the putative operator region between the Zmo promoter and the *fuGFP* gene. Used to generate pUS376	8320	2X Q5 Master Mix	59	90
NIY112 & NIY113	Used to amplify the pUS374 and pUS376 backbone to exclude the *sacB* gene. Used to generate pUS377 and pUS378.	6685 or 6591	2X Q5 Master Mix	56	70

NIY114 & NIY115	Used to amplify the knockout cassette cloned into pUS374 and pUS376. Includes sites that allow cloning into plasmids with the pUC MCS. In this study, used to clone the KO cassette into pKNOCK-Km to form pUS379 and pUS380.	2682 or 2588	2X Q5 Master Mix	59	30
NIY116 & NIY103	A region between the R6Kγ ori and the amidohydrolase gene in pUS379 and pUS380.	399	2X MangoMix	52	15
NIY117 & NIY108	A region between the *cntI* gene and the pKNOCK-Km backbone in pUS379 and pUS380.	205	2X MangoMix	52	15
NIY118 & NIY105	A region between the chromosomal *S. soli* amidohydrolase and the *fuGFP* in the KO cassette. To verify integration of the cassette into the correct position in the *S. soli* genome.	1154	2X MangoMix	51	40
NIY119 & NIY106	A region between the *S. soli* genome and the *fuGFP* in the KO cassette. To verify integration of the cassette into the correct position in the *S. soli* genome.	1157	2X MangoMix	51	40
NIY118 & NIY119	The region that the KO cassette replaces. This should amplify the homology arms and the *fuGFP* in double crossover *S. soli* knockout mutants, or the homology arms and the *zmoABCD* cluster in WT *S. soli*.	2808 (379KO), 2688 (380KO), or 6055 (WT)	2X Q5 Master Mix	57	70

LFE3 & LFE4	The *km*^{R} gene in pBBR1MCS-2 plasmids.	1012 (merodiploid), does not amplify using *S. soli* 379KO and *S. soli* 380KO as templates	2X MangoMix	57	15
NIY126 & NIY127	Used to amplify a portion of the *zmoC* gene from the *S. soli* genome. Includes restriction sites that allow cloning into plasmids with the pUC MCS. In this study, used to clone the partial *zmoC* gene into pKNOCK-Km to form pUS382.	787	2X Q5 Master Mix	57	10
NIY128 & NIY116	A region between the R6Kγ ori and the partial *zmoC* gene in pUS382.	283	2X MangoMix	51	15
NIY129 & NIY117	A region between the partial *zmoC* gene and the pKNOCK-Km backbone in pUS382.	299	2X MangoMix	51	15
NIY130 & NIY117	A region between the chromosomal *S. soli zmoC* gene and the pKNOCK-Km backbone. To verify integration of the pUS382 into the correct position in the *S. soli* genome.	990	2X MangoMix	50	40
NIY131 & NIY116	A region between the R6Kγ ori and the chromosomal *S. soli zmoC gene*. To verify integration of the pUS382 into the correct position in the *S. soli* genome.	1103	2X MangoMix	51	40

NIY132 & NIY133	Used to amplify a portion of the *zmoA* gene from the *S. soli* genome. Includes restriction sites that allow cloning into plasmids with the pUC MCS. In this study, used to clone the partial *zmoA* gene into pKNOCK-Km to form pUS383.	473	2X Q5 Master Mix	58	10
NIY134 & NIY116	A region between the R6Kγ ori and the partial *zmoA* gene in pUS383.	227	2X MangoMix	51	15
NIY135 & NIY117	A region between the partial *zmoA* gene and the pKNOCK-Km backbone in pUS383.	243	2X MangoMix	51	15
NIY136 & NIY117	A region between the chromosomal *S. soli zmoA* gene and the pKNOCK-Km backbone. To verify integration of the pUS383 into the correct position in the *S. soli* genome.	822	2X MangoMix	51	40
NIY137 & NIY116	A region between the R6Kγ ori and the chromosomal *S. soli zmoB gene*. To verify integration of the pUS383 into the correct position in the *S. soli* genome.	994	2X MangoMix	51	40
NIY138 & NIY111	The backbone of pUS379 and pUS386 to generate plasmids pUS384 and pUS388 respectively.	4678 (pUS384) 4653 (pUS388)	2X Q5 Master Mix	59	50
NIY139 & NIY111	The backbone of pUS379 to generate plasmid pUS385.	4708	2X Q5 Master Mix	59	50
NIY140 & NIY110	The backbone of pUS379 to generate plasmid pUS386.	4703	2X Q5 Master Mix	59	50
NIY141 & NIY110	The backbone of pUS379 to generate plasmid pUS387.	4667	2X Q5 Master Mix	59	50

NIY142 & NIY143	The backbone of pBBR1MCS-2 to remove the constitutive promoter and the *lacZ* gene and form pBBR1MCS-2_v2.	4729	2X Q5 Master Mix	58	50
NIY144 & NIY145	To amplify *zmo* promoter and *fuGFP* regions from pUS379 plasmids (i.e. pUS380, pUS384-387). For cloning into pBBR1MCS-2_v2 to generate pUS389-395.	Variable	2X Q5 Master Mix	58	10
NIY146 & NIY105	A region between the 5' end of the *fuGFP* gene and the pBBR1MCS-2_v2 backbone in pUS389-395.	Variable	2X MangoMix	50	15
NIY147 & NIY106	A region between the 3' end of the *fuGFP* gene and the pBBR1MCS-2_v2 backbone in pUS389-395.	268	2X MangoMix	50	15

2.4. General molecular biology techniques

All restriction enzymes, ligases and other enzymes, and all other molecular biological reagents were purchased from New England Biolabs (NEB), unless otherwise specified.

2.4.1. Plasmid isolation

50 mL or 100 mL LB (with the appropriate antibiotic(s)) was inoculated with one *E. coli* colony from solid medium harbouring the desired plasmid and was grown overnight at 37 ºC with 200 rpm shaking. The culture was centrifuged at 3500 rpm for 10 minutes. For each 50 mL of culture, 1 mL Buffer P1 (5 mM EDTA, pH 8.0. 300 µg/mL RNAse added after sterilisation) was used to resuspend the pellet. Then, to each 250 µL aliquot of the P1 cell suspension, 250 µL Buffer P2 (1% (w/v) SDS, 0.2 M NaOH, not sterilised) was added. The mixtures were inverted 10 times and incubated at room temperature for 10 minutes. Then, 350 µL Buffer N3 (4.2 M Guanidine HCl, 0.9 M Potassium acetate, pH 4.8, not sterilised) was added to each P1-P2 mixture, the mixtures were again inverted 10 times and incubated at room temperature for 5 minutes. The mixtures were then centrifuged at 20000 *g* for 5 minutes at room temperature to concentrate cell debris. The supernatant from each mixture was then transferred to silica membrane spin columns (Epoch Life Sciences) and centrifuged at 20000 *g* for 30 seconds. Until the addition of Buffer EB (see following section), the eluate from each column after each centrifugation step was discarded. Then, 700 µL Buffer PB (5 M Guanidine HCl, 100 mM sodium acetate, 30% (v/v) isopropanol, buffered to pH 5.2, not sterilised) was added to each spin column and centrifuged at 20000 *g* for 30 seconds. Then, 700 µL Buffer PE (10 mM Tris-HCl, pH 8.0, 80% (v/v) ethanol) was added to each spin column and centrifuged at 20000 g for 30 seconds, and this wash step was repeated once more. The “empty” spin columns were centrifuged at 20000 *g* for 2 minutes. 20 µL of Buffer EB (10 mM Tris-HCl, pH 8.0) was added to each column directly onto the membrane and was allowed to incubate at room temperature for 2 minutes. The spin columns were centrifuged at 20000 *g* for 1 minute and the eluate from each spin column were pooled for storage. The plasmid quality and quantity were determined using a Nanodrop spectrophotometer (Thermo Scientific) after pooling.

2.4.2. Restriction digests and ligase reactions

Restriction digests were generally prepared in 100 µL reactions with NEB CutSmart buffer or NEB 3.1 buffer, as prescribed by the manufacturer. The reactions included 50-250 ng of DNA and 10 U of the appropriate enzyme(s) (1 µL). The reactions were carried out at 25 ºC or 37 ºC as prescribed by the manufacturer, for 15-30 minutes.

Ligation reactions were performed in 20 µL reactions made up of 2 µL of 10X T4 DNA ligase buffer, and included 8 µL of restriction digested insert DNA, 8 µL of restriction digested plasmid, 400 U of T4 DNA ligase (1 µL), and 1 µL of sterilised ultrapure water. The ligation reactions were incubated at room temperature for 1 hour, or 4 ºC overnight.

2.4.3. DNA purification from enzymatic reactions

The reaction mixtures were made up to 100 µL with sterile ultrapure water. Then, 600 µL Buffer PB+ (5 M Guanidine HCl, 100 mM sodium acetate, 15 µg/mL phenol red, 30% (v/v) isopropanol, buffered to pH 5.2, not sterilised) were added to the reaction mix and were transferred to a silica membrane spin column (Epoch Life Sciences). The column was centrifuged at 20000 *g* for 30 seconds. Until the addition of Buffer EB (see following section), the eluate from the spin column after each centrifugation step was discarded. Then, 700 µL Buffer PE (10 mM Tris-HCl, pH 8.0, 80% (v/v) ethanol) was added to the spin column and centrifuged at 20000 *g* for 30 seconds, and the PE wash step was repeated once. The “empty” spin column was centrifuged at 20000 *g* for 2 minutes. 15 µL of Buffer EB (10 mM Tris-HCl, pH 8.0) was added to the column directly onto the membrane and was allowed to incubate at room temperature for 2 minutes. The spin column was centrifuged at 20000 *g* for 1 minute and the eluate from spin column was retained. After column purification, the quantity and quality of DNA from PCR reactions were accessed using a Nanodrop spectrophotometer (Thermo Scientific). DNA from restriction digests were not accessed.

2.4.4. DNA gel electrophoresis

Gel electrophoresis of DNA fragments were performed in 0.8-1.5% (w/v) agarose gels, depending on the size of the DNA fragments. LAB buffer (10 mM lithium acetate and 10 mM boric acid) was used and the gels were run at 250 V for 12-15 minutes, depending on the desired separation of bands. Agarose gels were pre-stained with thiazole orange (1 µL

of 15 mg/mL stock in DMSO per 10 mL of agarose gel) and the DNA were visualised using the Bio-Rad ChemiDoc MP Imaging System.

2.4.5. DNA purification via gel extraction

For gel extractions, DNA bands in the agarose gels were not visualized using the Bio-Rad ChemiDoc MP Imaging System. Instead, the DNA bands in the gel were then visualised using a blue light (wavelength 470 nm) and viewed under an orange filter. The desired band was excised and weighed in an Eppendorf tube to determine the gel volume to be extracted, the weight of the excised agarose gel is “one volume”. Then, 3 volumes of Buffer QG+ (5.5 M guanidine thiocyanate, 20 mM Tris HCl, pH 6.6, not sterilised) was added to the excised gel (i.e. add 600 µL for a 200 mg gel slice) and was heated at 80 ºC in a heat block for 5 minutes to melt the agarose gel. The tube was then cooled to room temperature and 1 volume of 100 % isopropanol was added. This mixture was transferred to a silica membrane spin column and the DNA was purified as described in Section 2.4.3. The quantity and quality of DNA after column purification was accessed using a Nanodrop spectrophotometer (Thermo Scientific).

2.4.6. DNA sequencing

Sanger sequencing reactions were set up according to the Australian Genome Research Facility (AGRF) guidelines. Reactions as follows: 2 µL 5 µM primer, 1 µg plasmid or 60 ng PCR product, and ultrapure water up to a final volume of 12 µL. These were sent to AGRF in Westmead, Sydney, Australia for processing. Full-length plasmid sequencing reactions were set up according to Plasmidsaurus or AGRF guidelines. Samples were sent to Plasmidsaurus in Oregon, USA or AGRF in Westmead, Sydney for processing.

2.4.7. DNA analysis & in silico plasmid design

DNA analysis (from sequencing reactions or annotations) and *in silico* plasmid design were performed in the SnapGene Software Package (Dotmatic). Potential primer binding sites were visualised using SnapGene. The minimum free energy, structure, and dimerisation of primers were analysed using the OligoAnalyzer Tool provided by the Integrated DNA Technologies website (https://www.idtdna.com/calc/analyzer) with the qPCR parameter. Software used for phylogenetic analyses is described in the methods section of Chapter 1.

2.4.8. Plasmids

Table 2.4.1. Plasmids used in this study.

MCS = multiple cloning site.

Name	Selective marker	Characteristics	Source
pBBR1MCS-2	Km^R	5.1 kb plasmid. A broad host range expression vector for Gram negative bacteria. Medium copy number in *E. coli*. The derepressed *lac* promoter gives constitutive expression of cloned genes. Used in this study as an expression vector and a positive control for conjugation experiments.	Kovach *et al.* (1995)
pBBR1MCS-3	Tc^R	5.2 kb plasmid. A broad host range expression vector for Gram negative bacteria. Medium copy number in *E. coli*. The derepressed *lac* promoter gives constitutive expression of cloned genes. Used in this study as a control for conjugation experiments.	Kovach *et al.* (1995)
pBBR1MCS-5	Gm^R	4.8 kb plasmid. A broad host range expression vector for Gram negative bacteria. Medium copy number in *E. coli*. The derepressed *lac* promoter gives constitutive expression of cloned genes. Used in this study as the first trial for conjugation experiments.	Kovach *et al.* (1995)
pEX18Tc	Tc^R	6.3 kb plasmid. A suicide plasmid intended for recombineering. High copy number in *E. coli*. Contains a counter-selectable marker (*sacB*). MCS is contained within the *lacZ* gene, under the same orientation as pUC18. Used in this study as the source of *sacB* and template for the backbone of pUS374, pUS376, pUS377, and pUS378.	Hoang *et al.* (1999)
pKNOCK-Km	Km^R	2.1 kb plasmid. A suicide plasmid intended for recombineering. Medium copy number in *E. coli* and only replication is driven by an R6Kγ origin of replication. Used in this study as the backbone of pUS379, pUS380, pUS382, and pUS383 for recombineering in *S. soli*.	Alexeyev (1999)
pJV53	Km^R	8.8 kb plasmid derived from pJAM2 containing origins of replication for *E. coli* and *Mycobacterium*. High copy number in E. coli, low copy number in *Mycobacterium*. Expresses mycobacteriophage Che9c recombinase genes 60-61 under the control of the acetamide-inducible promoter.	van Kessel and Hatfull (2007)

pUS116	Km^R	6.6 kb plasmid derived from pJV53. The recombineering genes were removed from pJV53 to form a basic *Mycobacterium* expression vector. The induction of gene expression using acetamide functions only in *Mycobacterium*. Contains NdeI, BamHI, PvuII, EcoRI sites for cloning. Used in this study as the expression vector for the native *S. soli* group 7 SDIMO in *M. smegmatis* mc^2-155.	Addgene #84578
pUS229	Gm^R	6.3 kb plasmid derived from pUCP24. It contains origins of replication for *E. coli* and *P. putida*. The plasmid carries the *E. coli* chaperonins GroES/EL and are expressed constitutively. Used in this study for co-expression experiments and the source of chaperonins for cloning.	McCarl *et al.* (2018)
pUS250	Km^R	4.7 kb plasmid derived from pBBR1MCS2. High copy number in Gram negative hosts. The *lac* promoter was replaced with an inducible cumate repressor system (*cymR*). A gene encoding a chromoprotein, *amilCP*, was introduced between the multiple cloning sites for easy visual screening during cloning. Used in this study as the expression vector for native *S. soli* group 7 SDIMO in P*. putida* KT2440.	Addgene #198322
pUS367	Km^R	7.9 kb plasmid derived from pUS250. The native *S. soli zmoABCD* (from PCR) was cloned into pUS250 using NdeI and HindIII. Expression of the native monooxygenase is controlled by the inducible cumate promoter.	This study.
pUS368	Km^R	10.8 kb plasmid derived from pUS116. The native *S. soli zmoABCD* (from PCR) was cloned into pUS116 using NdeI and BglII. Expression of the native monooxygenase is controlled by the inducible acetamidase promoter.	This study.
pUS369	Km^R	10.0 kb plasmid derived from pUS250. The *E. coli* chaperonins *groES/EL* genes were cloned into pUS367 using HindIII and BglII. Expression of the native monooxygenase and the chaperonins are controlled by the inducible cumate promoter.	This study.
pUS370	Km^R	3.9 kb plasmid derived from pUS250. The *amilCP* gene was removed with NotI and re-ligated the plasmid to itself. Used in this study as an empty plasmid control for *P. putida* KT2440 cumate growth experiments.	This study.
pUS374	Tc^R	8.4 kb plasmid derived from pEX18Tc. The backbone was amplified using NIY94 and NIY95 to add BsaI sites for cloning. gBlocks from TWIST Bioscience was assembled using Golden Gate assembly. Contains an *fuGFP* gene under the control of the Zmo promoter, flanked by genes upstream and downstream of the *S. soli* group 7 SDIMO. Used in this study as a suicide vector for recombineering in *S. soli* and the template for KO cassette amplification.	This study.

pUS375	Km^R	6.5 kb plasmid derived from pBBR1MCS-2. The *sacB* gene from pEX18Tc was cloned into pBBR1MCS-2 using HindIII and BglII. Used in this study to verify if *sacB* can be used as a counter-selection marker in *S. soli*.	This study.
pUS376	Tc^R	8.3 kb derived from pUS374. Identical to pUS374 except the putative operator region between the Zmo promoter and the *fuGFP* has been removed.	This study.
pUS377	Tc^R	6.7 kb plasmid derived from pUS374. Identical to pUS374 except the *sacB* gene has been removed.	This study.
pUS378	Tc^R	6.6 kb plasmid derived from pUS376. Identical to pUS376 except *sacB* gene has been removed.	This study.
pUS379	Km^R	4.7 kb plasmid derived from pKNOCK-Km. The amplified KO cassette from pUS374 was cloned into the MCS of pKNOCK-Km. Used in this study as a suicide vector for recombineering in *S. soli* and to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli*.	This study.
pUS380	Km^R	4.6 kb plasmid derived from pUS379. Identical to pUS379 except the putative operator region between the Zmo promoter and the *fuGFP* has been removed. Used in this study as a suicide vector for recombineering in *S. soli* and to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli*.	This study.
pUS382	Km^R	2.8 kb plasmid derived from pKNOCK-Km. A portion of the *S. soli zmoC* gene was amplified by PCR and cloned into the MCS of pKNOCK-Km. Used in this study as a suicide vector for recombineering in *S. soli*.	This study.
pUS383	Km^R	2.5 kb plasmid derived from pKNOCK-Km. A portion of the *S. soli zmoA* gene was amplified by PCR and cloned into the MCS of pKNOCK-Km. Used in this study as a suicide vector for recombineering in *S. soli*.	This study.
pUS384	Km^R	4.7 kb plasmid derived from pUS379. The plasmid was amplified from pUS379 using primers and conditions described in Table 2.3.2. Used in this study to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli*.	This study.
pUS385	Km^R	4.7 kb plasmid derived from pUS379. The plasmid was amplified from pUS379 using primers and conditions described in Table 2.3.2. Used in this study to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli*.	This study.
pUS386	Km^R	4.7 kb plasmid derived from pUS379. The plasmid was amplified from pUS379 using primers and conditions described in Table 2.3.2. Used in this study to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli*.	This study.

pUS387	Km^R	4.7 kb plasmid derived from pUS379. The plasmid was amplified from pUS379 using primers and conditions described in Table 2.3.2. Used in this study to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli*.	This study.
pUS388	Km^R	4.7 kb plasmid derived from pUS379. The plasmid was amplified from pUS379 using primers and conditions described in Table 2.3.2. Used in this study to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli*.	This study.
pUS389	Km^R	5.7 kb plasmid derived from pBBR1MCS-2_v2. The backbone was amplified from pBBR1MCS-2 using primers and conditions described in Table 2.3.2. The *zmo* promoter and *fuGFP* region from pUS379 was cloned into pBBR1MCS-2_v2. Used in this study to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli* and *S. soli*.	This study.
pUS390	Km^R	5.6 kb plasmid derived from pBBR1MCS-2_v2. The backbone was amplified from pBBR1MCS-2 using primers and conditions described in Table 2.3.2. The *zmo* promoter and *fuGFP* region from pUS380 was cloned into pBBR1MCS-2_v2. Used in this study to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli* and *S. soli*.	This study.
pUS391	Km^R	5.6 kb plasmid derived from pBBR1MCS-2_v2. The backbone was amplified from pBBR1MCS-2 using primers and conditions described in Table 2.3.2. The *zmo* promoter and *fuGFP* region from pUS384 was cloned into pBBR1MCS-2_v2. Used in this study to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli* and *S. soli*.	This study.
pUS392	Km^R	5.6 kb plasmid derived from pBBR1MCS-2_v2. The backbone was amplified from pBBR1MCS-2 using primers and conditions described in Table 2.3.2. The *zmo* promoter and *fuGFP* region from pUS385 was cloned into pBBR1MCS-2_v2. Used in this study to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli* and *S. soli*.	This study.
pUS393	Km^R	5.6 kb plasmid derived from pBBR1MCS-2_v2. The backbone was amplified from pBBR1MCS-2 using primers and conditions described in Table 2.3.2. The *zmo* promoter and *fuGFP* region from pUS386 was cloned into pBBR1MCS-2_v2. Used in this study to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli* and *S. soli*.	This study.

pUS394	Km^R	5.6 kb plasmid derived from pBBR1MCS-2_v2. The backbone was amplified from pBBR1MCS-2 using primers and conditions described in Table 2.3.2. The *zmo* promoter and *fuGFP* region from pUS387 was cloned into pBBR1MCS-2_v2. Used in this study to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli* and *S. soli*.	This study.
pUS395	Km^R	5.6 kb plasmid derived from pBBR1MCS-2_v2. The backbone was amplified from pBBR1MCS-2 using primers and conditions described in Table 2.3.2. The *zmo* promoter and *fuGFP* region from pUS388 was cloned into pBBR1MCS-2_v2. Used in this study to examine the putative operator region between the *zmo* promoter and fuGFP in *E. coli* and *S. soli*.	This study.

2.5. Transformation of bacteria

2.5.1. Preparation of chemically competent *E. coli* cells

A single colony of *E. coli* was transferred into 5 mL LB and was grown overnight (16-24 hours). Then, fresh 50 mL LB was inoculated with the overnight culture to an OD_{600} of 0.05 and the culture was incubated at 37 ºC with 200 rpm shaking. When the cells reached an OD_{600} of 0.5 (mid-exponential phase) the 50 mL culture was centrifuged (4 ºC, 3900 *g* for 10 minutes). The pellet was resuspended in 33 mL ice-cold RF1 solution (100 mM rubidium chloride, 50 mM manganese chloride, 30 mM potassium acetate, 10 mM calcium chloride, 15% (w/v) glycerol, pH 5.8. Sterilised by filtration through a 0.22 µm nitrocellulose filter) and the cells were incubated on ice for an hour. The suspension was centrifuged under the same conditions as above. The pellet was resuspended in 8 mL ice-cold RF2 solution (10 mM MOPS, 10 mM rubidium chloride, 75 mM calcium chloride, 15% (w/v) glycerol, pH 6.8, sterilised by filtration through a 0.22 µm nitrocellulose filter) and was incubated on ice for 15 minutes. The cell suspension was divided into 220 µL aliquots and were frozen at -80 ºC until required.

2.5.2. Heat-shock transformation of chemically competent *E. coli* cells

1 µL of purified plasmid or 5-10 µL ligation mixture was added to 50 µL of chemically competent *E. coli* cells. The cells were incubated on ice for 15 minutes and were then subject to heat-shock treatment at 42 ºC for 45 seconds. The cells were incubated on ice again for 2 minutes. After, 1 mL LB was added to the cell suspension and the cells were allowed to incubate at 37 ºC with 200 rpm shaking for an hour. The cell suspension was spread onto solid LB medium supplemented with the appropriate antibiotic(s).

2.5.3. Preparation of electrocompetent *M. smegmatis* mc^2-155

From solid medium, a single colony of *M. smegmatis* mc^2-155 was transferred into 5 mL LB (with 0.05% (v/v) Tween 80) and was grown for 2-3 days (until late stationary phase). Then, three 50 mL LB (with 0.05% (v/v) Tween 80) were inoculated with the 5 mL culture to OD_{600}s of 0.02, 0.05, and 0.10. These were then grown overnight (16-24 hours), or until one of the cultures reached an OD_{600} of approximately 0.80-

1.0. The appropriate culture was centrifuged at 4 ºC, 3900 *g* for 10 minutes. The cells were washed three times in 20 mL ice-cold electroporation buffer (10% (v/v) glycerol, with 0.05% (v/v) Tween 80 added after sterilisation). After the final wash step, the pellet was resuspended in 3 mL ice-cold electroporation buffer. The cell suspension was divided into 160 µL aliquots and the aliquots were frozen at -80 ºC until required.

2.5.4. Preparation of electrocompetent *P. putida* KT2440

From solid medium, a single colony of *P. putida* KT2440 was transferred into 5 mL LB and was grown for 1-2 days (until late stationary phase). Then, 50 mL LB was inoculated with the 5 mL culture to OD_{600} of 0.05 and the culture was grown to OD_{600} of 0.5. The culture was centrifuged at 4 ºC, 3900 *g* for 10 minutes The cells were washed three times in 20 mL ice-cold electroporation buffer (10% (v/v) glycerol). After the final wash step, the pellet was resuspended in 3 mL ice-cold electroporation buffer. The cell suspension was divided into 160 µL aliquots and the aliquots were frozen at -80 ºC until required.

For electroporation conditions used for *P. putida* KT2440 and *M. smegmatis* mc^2-155, see Section 3.2.2.

2.6. Other methods

2.6.1. SDS-PAGE Sample Preparation

A cell pellet from a 1 mL culture was resuspended in 500 µL 1X SDS-PAGE sample buffer (4% (v/v) sodium dodecyl sulphate, 0.004% (w/v) Bromophenol Blue, 125 mM Tris-HCl, 20% (w/v) glycerol, 10% (v/v) ß-mercaptoethanol, pH 6.8, not sterilised) and transferred to a 2 mL screw cap tube containing with 2 large glass beads, 50 mg small glass beads and 50 mg very small glass beads. The cells were lysed in a beadbeater at 3000 rpm for 30 seconds and were rested on ice for 1 minute. This was repeated once for a total of 2 lysis rounds. The tubes were centrifuged at 20000 *g* for 3 minutes to separate the supernatant from cell fragments and glass beads. The supernatant was heated for 5 minutes at 105°C in a heatblock. The supernatant

was cooled to room temperature and the protein concentration was quantified using a NanoDrop spectrophotometer.

2.6.2. SDS-PAGE conditions and visualisation method

An SDS-PAGE tank was set up containing a 8-16% Mini-PROTEAN® TGX Stain-Free™ Precast gel (Bio-Rad). The tank was then filled with cold 1X SDS running buffer (25 mM Tris-HCl, 192 mM glycine, 0.1% (w/v) sodium dodecyl sulphate, pH 8.3, not sterilised) and 30 ng of protein were loaded into each well. 5 µL of the Precision Plus Protein Unstained Standards (Bio-Rad) was used as the protein marker ladder. The gel was run at 120 V for 60 minutes or until the dye front reached the end of the gel. The proteins were visualised using the Bio-Rad ChemiDoc MP Imaging System with the settings for Stain-Free™ Precast gels.

When required, the acrylamide gel was stained with Coomassie Blue stain as follows: Coomassie Blue staining solution (50% (v/v) methanol, 10% (v/v) glacial acetic acid, 0.025% (w/v) Coomassie R250 Brilliant Blue) was incubated with the SDS-PAGE gel at room temperature for 30 minutes with gentle rocking. The solution was decanted and the gel was incubated with High destain solution (40% (v/v) methanol, 10% (v/v) glacial acetic acid) at room temperature for 30 minutes with gentle rocking. The solution was decanted and the gel was incubated with Low destain solution (10% (v/v) methanol, 10% (v/v) glacial acetic acid) at room temperature for 3 hours with gentle rocking. The solution was decanted and the gel was sandwiched between two thin acrylic sheets and kept at 4 ºC until required for mass spectrometry analysis.

Chapter 3 – Heterologous Expression of the *Solimonas soli* Group 7 SDIMO (ZmoABCD)

Published as: Yang, S. N. N., Haritos, V., Kertesz, M. A., Coleman, N. V. (2024) A novel soluble di-iron monooxygenase from the soil bacterium *Solimonas soli*. Environmental Microbiology, 26(2), e16567. DOI: 10.1111/1462-2920.16567.

3.1. Introduction

Monooxygenases (MOs) initiate the metabolism of hydrocarbons in many bacteria and play critical roles in biogeochemistry via the oxidation of methane and ammonia in the carbon and nitrogen cycles, respectively (Lu *et al.*, 2020, Ayub *et al.*, 2022, Greening and Grinter, 2022). Monooxygenases can biodegrade diverse pollutants and xenobiotics, including aliphatic (Abbasian *et al.*, 2015) and aromatic (Seo *et al.*, 2009) hydrocarbons, haloalkenes (Le and Coleman, 2011), haloalkanes (Hage and Hartmans, 1999), cyclic ethers (Masuda *et al.*, 2012a) and other heterocyclic compounds (Thiemer *et al.*, 2003, Casaite *et al.*, 2020). The MO enzymes are valuable for biocatalysis due to their ability to insert oxygen atoms into inert substrates with high regio- and enantioselectivity (Koeller and Wong, 2001, Leak *et al.*, 2009, Torres Pazmiño *et al.*, 2010, Cheung *et al.*, 2013, Desai *et al.*, 2016, Birolli *et al.*, 2019). The major types of MOs can be categorised as cofactor-dependent or cofactor-independent MOs (Leahy *et al.*, 2003). In the former case, the cofactors may be flavin (van Berkel *et al.*, 2006), heme (p450) (Bernhardt, 2006), pterin (Torres Pazmiño *et al.*, 2010), copper (CuMMO) (Coleman *et al.*, 2012), membrane-bound di-iron (AlkB) (Leahy *et al.*, 2003), or soluble di-iron (SDIMO) (Holmes and Coleman, 2008).

The SDIMOs have high sequence diversity, broad substrate range, and extensive distribution among many bacterial lineages. The SDIMOs are best categorised based on amino acid sequence data, but simpler schemes based on the number of enzyme subunits (ranging from four to six), the arrangement of the gene clusters, or the nature of their canonical substrate have also been used (Leahy *et al.*, 2003). While the last of these methods is unreliable due to the broad substrate range of these enzymes, this system has nevertheless been widely used, and organises the SDIMOs into six groups (Notomista *et al.*, 2003, Coleman *et al.*, 2006): toluene MOs

(group 1) (Bertoni *et al.*, 1996, Bertoni *et al.*, 1998, Cafaro *et al.*, 2004), phenol MOs (group 2) (Arai *et al.*, 1998, Hino *et al.*, 1998, Zhou *et al.*, 2016), soluble methane MOs (sMMO), group 3) (Stainthorpe *et al.*, 1990, Cardy *et al.*, 1991, Sayavedra-Soto *et al.*, 2001), alkene MOs (group 4) (Mattes *et al.*, 2005, McCarl *et al.*, 2018, Suzuki *et al.*, 2019), propane-2-MOs (group 5) (Kotani *et al.*, 2003, Furuya *et al.*, 2011, Furuya *et al.*, 2015), and propane-1-MOs (group 6) (Kotani *et al.*, 2006, Masuda *et al.*, 2012b, Deng *et al.*, 2018). The known sequence diversity of SDIMO genes has greatly expanded in recent years with genomic and metagenomic sequencing, but in many cases, the biochemistry of the enzymes and their role in the physiology of the hosts remain unknown.

In a recent review article, an unusual as-yet-uncharacterised SDIMO was reported in the genome of the soil bacterium *Solimonas soli* (Osborne and Haritos, 2019). The *Solimonas* MO was similar to the alkene monooxygenases in its four-component structure and gene cluster organisation, but phylogenetic analysis based on the sequence of the large (alpha) hydroxylase component indicated that it did not fit into any of the pre-existing SDIMO groups and was closer to the root of the SDIMO tree than any previously reported enzyme. No obvious clues to the function of the *Solimonas* SDIMO can be obtained from the environmental niche of the host bacterium (agricultural soil), conditions of isolation (general purpose medium), or initial phenotypic characterisation (no SDIMO substrates were tested) (Kim *et al.*, 2007).

In this chapter, the classification of the *Solimonas* SDIMO will be re-examined, and the genes will be tested to determine whether they encode a functional monooxygenase. The aim of this chapter is to determine the substrate range of the enzyme, in order to infer its role in the physiology of the host and to provide insights into its ecological role and potential applications in biocatalysis and biodegradation.

3.2. Methods and materials

3.2.1. Bacterial strains, media, and culture conditions

Escherichia coli TOP10 (Invitrogen) and *Pseudomonas putida* KT2440 (Bagdasarian *et al.*, 1981) were both cultured in lysogeny broth (LB) medium (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl), with incubation at 37 ºC and 30 ºC, respectively, unless otherwise indicated. *Mycobacterium smegmatis* mc^2-155 was cultured in either LB medium or minimal salts medium (MSM) (Coleman *et al.*, 2002) (see Table 2.2.1 for details on preparation of this medium) with 20 mM glucose as the carbon and energy source. Stock solutions of cumate (0.5 M) were prepared in 50% (v/v) ethanol in mixed with 0.5 M Tris-HCl (pH 8) to a 1:1 solution. Tween 80 (5% v/v) and acetamide (100% (w/v)) stock solutions were both prepared by filter-sterilisation through a 0.22 µm filter (Table 2.2.2). To MSM, Tween 80 was added to 0.05% (v/v) while acetamide was added to 2% (w/v) as an inducer. All cultures were incubated aerobically, with liquid cultures shaken at 200 rpm. Kanamycin (Km) (50 µg/mL) was added to *E. coli* TOP10 and *P. putida* KT2440 cultures, where required. For *M. smegmatis* cultures, Km was added at 25 µg/mL, where required. Gentamicin (Gm) (10 µg/mL) was added to *P. putida* cultures where required.

3.2.2. General molecular biology methods

The *zmoABCD* gene cluster was amplified from *S. soli* DCY12 genomic DNA via PCR using Q5 MasterMix (New England Biolabs) and primers NIY70 (AAAAACATATGATGCTGATCGGGTATGG) and NIY71 (TTTAAGCTTGCGTTCCTCCGTCCG) (For details, refer to Table 2.3.1) PCR products were purified using silica membrane spin columns (Epoch Life Sciences) (See Section 2.4.3). Plasmids were extracted by alkaline lysis (Green and Sambrook, 2016), and purified using the same spin columns as above. Yields and purity of DNA products were assessed using a Nanodrop 1000 spectrophotometer (Thermo Scientific). Sizes of DNA products were tested by agarose gel electrophoresis in 1% (w/v) agarose gels in 1X Lithium-acetate-borate (LAB) buffer (Singhal *et al.*, 2010), stained with thiazole orange (O'Neil *et al.*, 2019), and visualised using a ChemiDoc MP Imaging System (Bio-Rad). Plasmids were transformed into chemically competent *E. coli* TOP10 cells via heat shock (Hanahan, 1983), and transformed into *P. putida* KT2440 and *M. smegmatis* mc^2-155 cells by

electroporation (2500 V, 25 µF, 200 Ω or 800 Ω respectively, Gene Pulser Xcell, Bio-Rad) in 2 mm gap cuvettes. The *E. coli* TOP10 cells were made chemically competent via the rubidium chloride method (Hanahan, 1983) while *P. putida* KT2440 and *M. smegmatis* mc^2-155 cells were made electrocompetent by repeated washing in 10% (v/v) glycerol (Fu *et al.*, 2010) (with the addition of 0.05% (v/v) Tween 80 for *M. smegmatis* mc^2-155 cells). All molecular biology enzymes were from New England Biolabs.

3.2.3. Construction of pUS367 and expression of ZmoABCD in *P. putida* KT2440

Plasmid pUS250 (Addgene #198322, Table 2.4.1) consists of the cumate-inducible regulatory system from *Pseudomonas putida* F1 (Eaton, 1997) and the replication and resistance functions of pBBR1MCS-2 (Kovach *et al.*, 1995).The pUS250 plasmid and *zmoABCD* PCR product were separately digested with NdeI and HindIII-HF, purified, then ligated with Hi-T4 DNA ligase. The ligation mix was transformed into *E. coli* TOP10 cells. The resulting plasmid was designated pUS367 (Figure 3.2.1) was confirmed by whole plasmid sequencing via the Nanopore method (https://www.plasmidsaurus.com/). Cells of *P. putida* KT2440 were transformed with pUS367, then colonies arising after 3 days were homogenised in LB medium, and transferred into 500 mL LB-Km50 to give an initial OD_{600} = 0.1. The culture was grown to mid-exponential phase (OD_{600} = 0.5, 3-4 hours), cumate was added to 1 mM to induce *zmoABCD* expression, then incubation was continued for another 4 hours at 20 ºC. Cells were harvested by centrifugation (3900 *g*, 10 min, 20 ºC) and used immediately for MO activity assays (see Sections 3.2.6 and 3.2.7) or preserved at -30 ºC for protein analysis.

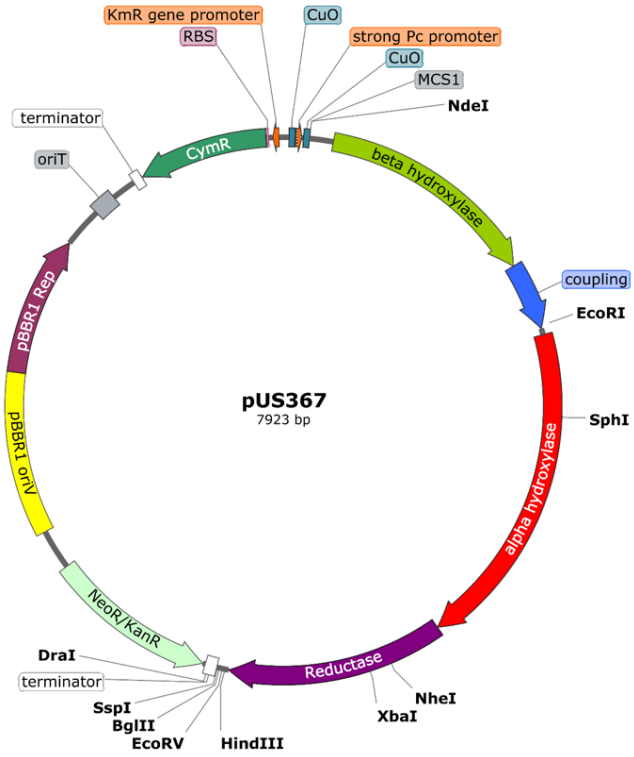


Figure 3.2.1. Map of pUS367.

pUS367 is a 7.9 kb Km^R plasmid derived from pUS250. Unique restriction sites are shown. For more detail refer to Table 2.4.1.

3.2.4. Co-expression of ZmoABCD and *E. coli* chaperonin GroES/EL

Plasmid pUS229 (Figure 3.2.2) (McCarl *et al.*, 2018) was introduced into *P. putida* KT2440 cells by electroporation, as described in Section 3.2.2, with selection on LB-Gm10 agar. The Gm^R transformants were transformed with pUS367 by electroporation as above. Colonies arising after 3 days were grown in 100 mL LB-Km50Gm10 to mid-exponential phase (OD_{600} = 0.5), cumate was added to 1 mM to induce *zmoABCD* expression, then incubation was continued for another 4 hours at 20 ºC. Cells were harvested by centrifugation (3900 *g*, 10 min, 20 ºC) and used immediately for MO activity assays.

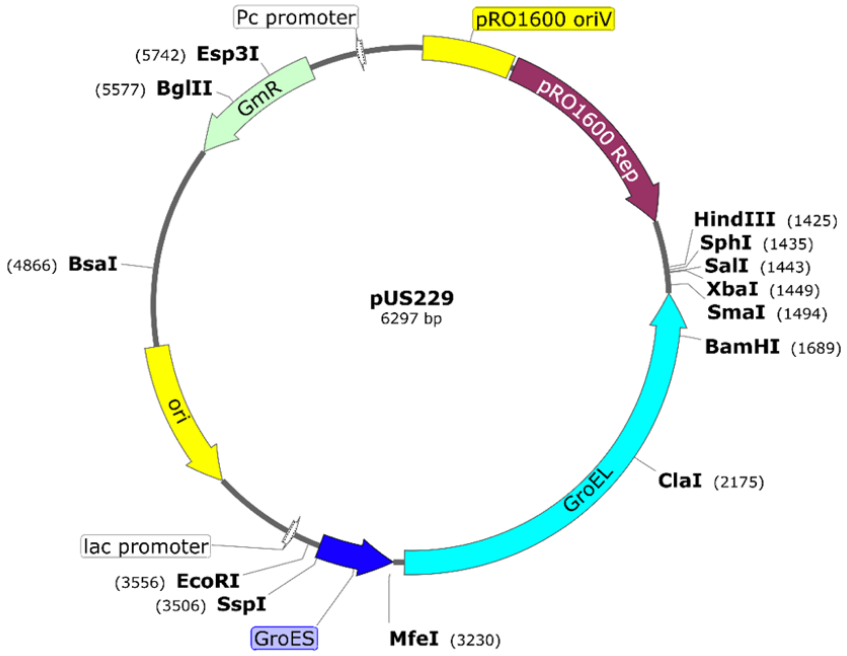


Figure 3.2.2. Map of pUS229.

A 6.3 kb plasmid (Gm^R) derived from pUCP24. It contains origins of replication for *E. coli* and *P. putida*. The plasmid constitutively expresses the *E. coli* chaperonins GroES/EL in both hosts. Unique restriction sites are shown. For more detail refer to Table 2.4.1.

3.2.5. Construction of pUS368 and expression of ZmoABCD in *M. smegmatis* mc^2-155

The *zmoABCD* gene cluster was amplified from pUS367 via PCR using Q5 MasterMix (New England Biolabs) and primers MVS48 (GTTCGGTTCGTAAACTGTAATGC) and CFM13 (CCCGTGATATTGCTGAAGAG) with thermocycling conditions according to Table 2.3.2. Plasmid pUS116 (Addgene #84578, Table 2.4.1) was digested with NdeI and BamHI-HF while the *zmoABCD* PCR product were digested with NdeI and BglII. The digested DNA fragments were purified, then ligated with Hi-T4 DNA ligase, and was transformed into *E. coli* TOP10 cells. The resultant Km^R colonies were replica-plated onto new LB-Km50 agar plates, then screened by PCRs across the ligation junctions, using primers NIY74 (CCGATAAGAGAAAGGGAGTCCAC) and NIY60 (GGGTGTACGTGCTGATAATGC) (left junction) and NIY75 (ACATCAGAGATTTTGAGACACAACG) and NIY61 (CAATTTGCGTGGCTCGG) (right junction) (Table 2.3.2). One clone that was positive for both PCRs was retained for further work. The new plasmid (designated pUS368, Figure 3.2.3) was purified, and its structure was confirmed to be correct by

whole plasmid sequencing via the Nanopore method (https://www.plasmidsaurus.com/). Cells of *M. smegmatis* mc^2-155 were transformed with pUS368 as described above, then colonies arising after 3 days were transferred into 100 mL MSM-glucose medium to give an initial OD_{600} = 0.05. The culture was grown to mid-exponential phase (OD_{600} = 0.5), acetamide was added to 2% (w/v) to induce *zmoABCD* expression, then incubation was continued for another 24 hours at 20 ºC. Cells were harvested by centrifugation (3900 *g*, 10 min, 20 ºC) and used immediately for MO activity assays.

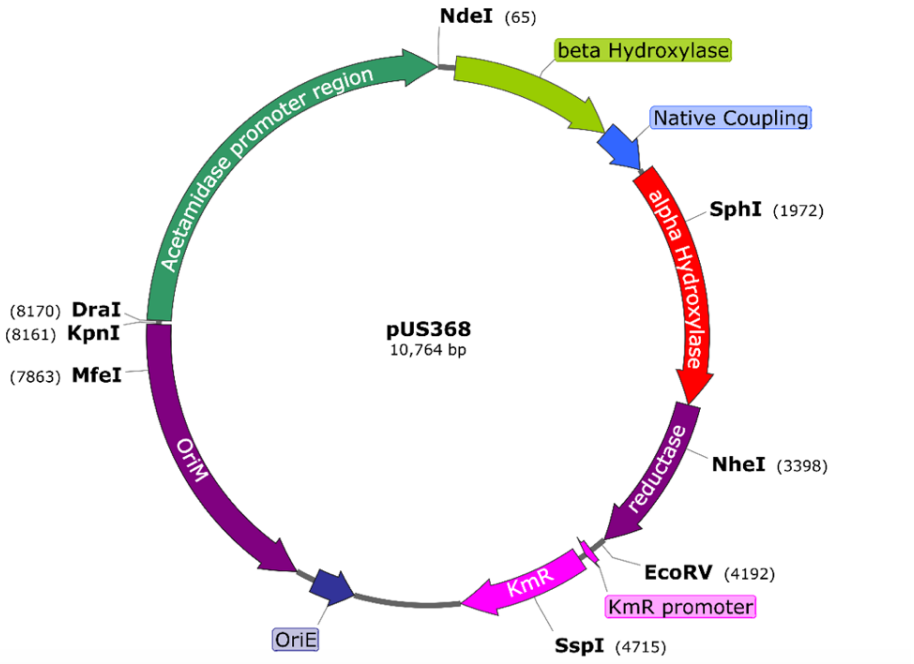


Figure 3.2.3. Map of pUS368.

A 10.8 kb (Km^R) plasmid derived from pUS116. The *zmoABCD* genes were cloned into pUS116 as an NdeI-BglII fragment, under the control of the inducible acetamidase promoter. Unique restriction sites are shown. For more detail refer to Table 2.4.1.

3.2.6. Detection of alkene epoxidation via NBP assay

The 4-(4-nitrobenzyl)pyridine (NBP) reagent reacts with epoxides and other reactive electrophiles to yield conjugates which appear intensely coloured under alkaline conditions (introduced through the addition of triethylamine). Cumate-induced *P. putida* KT2440 or acetamide-induced *M. smegmatis* mc^2-155 cells harbouring the relevant plasmids were washed twice in 20 mM phosphate buffer, pH 7, then

resuspended in the same buffer at OD_{600} = 30 (for *P. putida* KT2440 cells) or OD_{600} = 10 for (for *M. smegmatis* mc^2-155 cells). Aliquots of the cell suspension (0.9 mL) were mixed with glycerol (0.1 mL of 20% (v/v) glycerol) as an energy source in 16 mL vials. Smaller vials (2 mL) containing NBP solution (1 mL of 100 mM NBP) were placed inside the 16 mL vials. The outer vials were crimp-sealed, then alkene substrates (50 µmol) were added through the septum, either as neat liquids or gases, with liquid substrates added to the resting cell suspension and gases added to the headspace. The vials were incubated upright at 30 ºC with 200 rpm shaking for 16 hours, then the inner vials were removed, heated for 1 h at 105 ºC, and triethylamine was added (500 µL of a 50 % solution in acetone). The absorbance of the purple NBP-epoxide conjugate was measured at 555 nm (for ethylene oxide, VC, and cDCE) (Morrill *et al.*, 1981) or 550 nm (for other substrates) (Cheung *et al.*, 2013). The amount of epoxide produced was calculated via extinction coefficients (ethene oxide, VC, and cDCE = 2.11 x 10^3 L mol^{-1} cm^{-1}, propene oxide = 2.08 x 10^3 L mol^{-1} cm^{-1}, 1-butene oxide = 2.48 x 10^3 L mol^{-1} cm^{-1}, styrene oxide = 3.20 x 10^3 L mol^{-1} cm^{-1}. For other substrates tested, an average of the extinction coefficients of propene oxide, 1-butene oxide, and styrene oxide = 2.60 x 10^3 L mol^{-1} cm^{-1} was utilised) (Morrill *et al.*, 1981, McCarl *et al.*, 2018). All assays were performed in triplicate. The negative controls were uninduced *P. putida* KT2440 or *M. smegmatis* mc^2-155 cells with the addition of propene, and cumate-induced *P. putida* KT2440 or acetamide-induced *M. smegmatis* mc^2-155 cells with no alkene added.

3.2.7. Detection of pollutant dechlorination via colorimetric chloride assay

Cells of *P. putida* KT2440 were grown, induced, washed, and mixed with glycerol in 16 mL crimp-seal vials as in Section 3.2.6, except no inner vial was added. Chlorinated ethenes were injected as either neat gas (VC, 200 µL at 99% purity) or aqueous stock solution (cDCE, 32 µL at 250 mM). The vials were incubated at 30 ºC with 200 rpm shaking for 22 hours. After incubation, the cell suspensions were removed from the vials and cells removed by centrifugation (20000 *g*, 1 min). Chloride was detected in the supernatants via a modification of the method of Bergmann & Sanik (Le and Coleman, 2011, Martin *et al.*, 2014), as follows. Supernatant (900 µL) was mixed with iron reagent (200 µL of 8% (w/v) ferric ammonium sulphate in 6 M nitric acid) and mercury reagent (400 µL of 0.1% (w/v)

mercuric thiocyanate in ethanol), and incubated at 20 °C for 10 min, then absorbance was measured at 460 nm. The A_{460} data were converted to chloride concentrations using a standard curve of NaCl, prepared in the same buffer as the test samples (20 mM K_2HPO_4, pH 7) (Figure 3.2.4). All assays were done in triplicate. The negative controls were uninduced *P. putida* KT2440 cells with VC, and induced *P. putida* KT2440 cells with no alkene added.

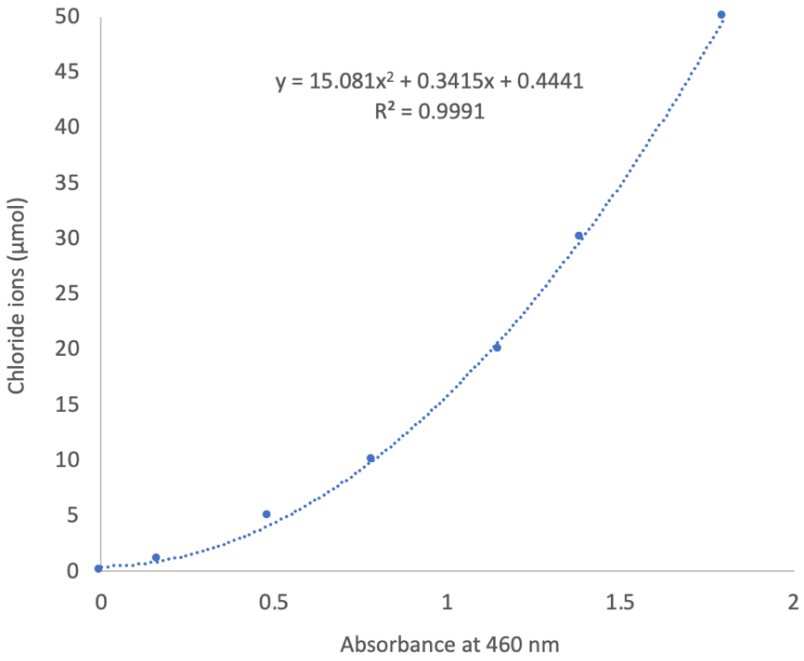


Figure 3.2.4. Chloride assay standard curve.

3.3. Results

3.3.1. Bioinformatic analysis: Proposal for SDIMO group 7

Since the first report of the unusual *Solimonas* SDIMO (Osborne and Haritos, 2019), several similar sequences have been deposited in GenBank (Clade 7, Figures 3.3.1 and 3.3.2), but these are all uncharacterised. Apart from one metagenomic sequence, these SDIMOs were all derived from Gammaproteobacteria, either *Sinobacteraceae* or *Spongiibacteraceae* families, with the majority from marine bacteria. Some of the host bacteria and/or sampling locations were impacted by petroleum hydrocarbons, which might select for MO enzymes, but this was not true in all cases. There is substantial sequence variation in these SDIMOs (60 to 78% amino acid identity to the *Solimonas* alpha hydroxylase subunit). However, they all formed a coherent and well-supported group alongside the *Solimonas* MO (Figure 3.3.1) based on phylogenetic analysis, and this was distinct from the six previously defined SDIMO groups. Analysis of the phylogeny of the beta hydroxylase and reductase subunits also provided strong support (99% bootstrap values) for these SDIMOs as a separate clade (Figures 3.3.3 and 3.3.4). Therefore it can be proposed that these enzymes collectively form a new SDIMO clade, ‘group 7’. The *Solimonas* MO gene cluste has been provisionally named *zmoABCD*, where ‘z’ indicates that the function is thus far unknown (note that the ‘xmo’ designation for a MO of unknown function has already been used (Khadka *et al.*, 2018)).

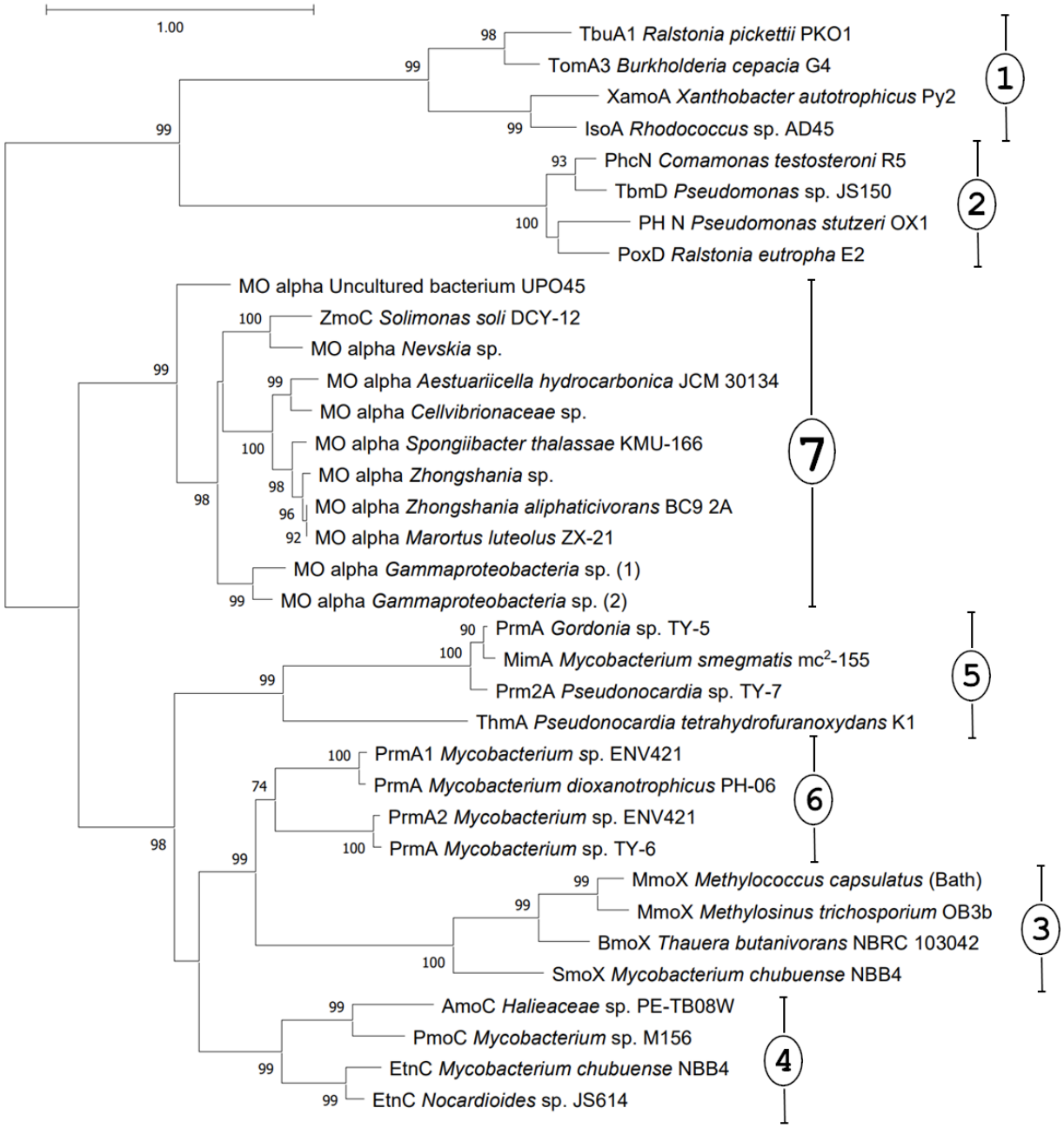


Figure 3.3.1. Phylogeny of SDIMOs based on alpha subunit hydroxylase amino acid sequences.

SDIMO group numbers are shown at the right. Tree was constructed in MEGA11 (Tamura *et al.*, 2021) by the maximum likelihood method using default settings from PhyML 3.0 (Guindon *et al.*, 2010) from an alignment (471 aa) including some internal gaps. Bootstrap values >50% are shown at the nodes. The scale bar denotes amino acid substitutions per sequence site. For clarity, only four representatives from pre-existing SDIMO groups were used for tree construction. These are only a small subset of the total diversity, but this was sufficient to cleanly resolve the group 7 SDIMOs from the pre-existing six groups.

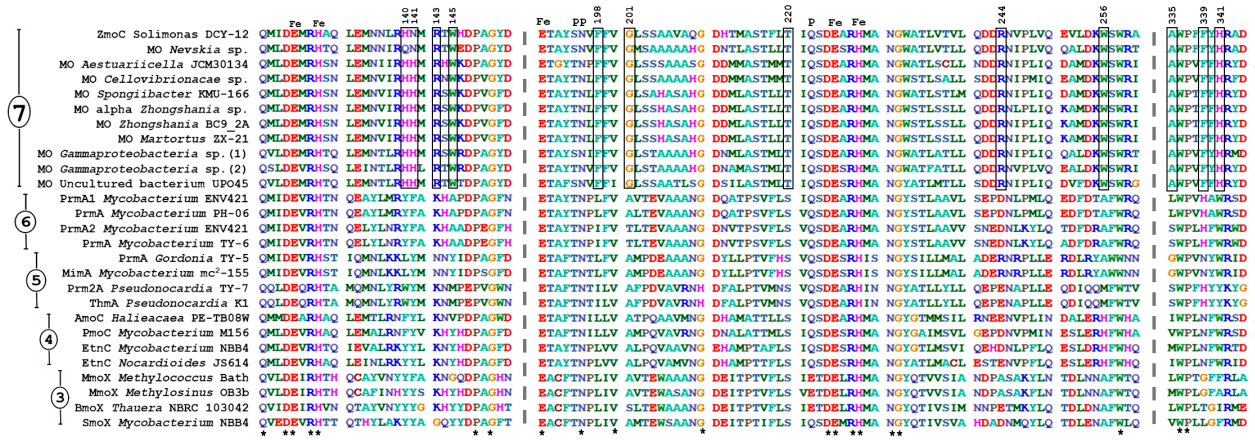


Figure 3.3.2. Abbreviated amino acid sequence alignment of alpha subunit hydroxylase components of SDIMO groups 3-7.

Residues are numbered relative to ZmoC of *S. soli*. Residues characteristic of group 7 are boxed. Fe, iron-binding residues, P, pore residues, asterisks, residues conserved across all SDIMOs in the alignment. Note that this is not a complete alignment of these proteins, it has been trimmed at the ends and there are internal gaps (indicated by dashed grey lines). This work was undertaken by Victoria Haritos at Monash University, Victoria, Australia.

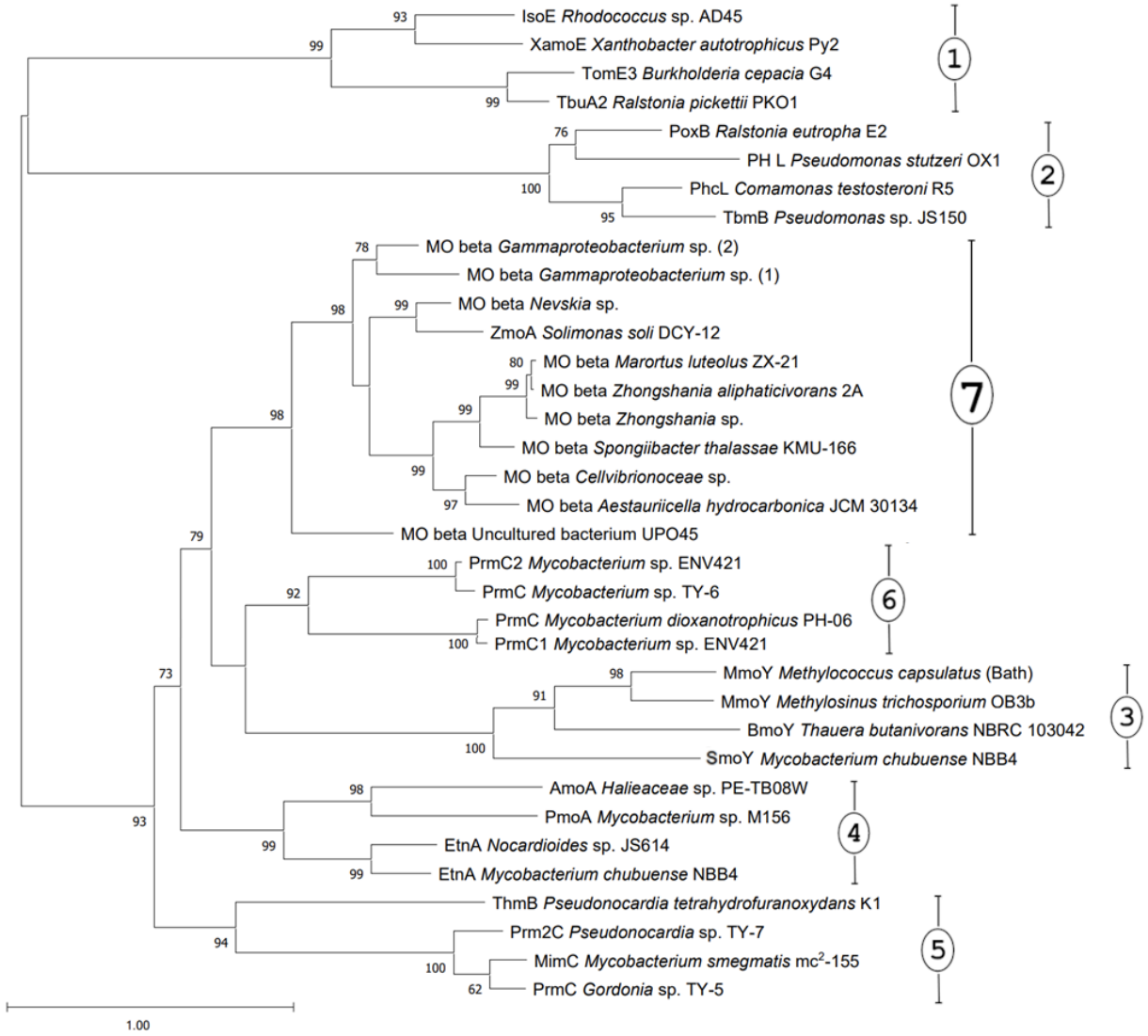


Figure 3.3.3. Phylogeny of SDIMOs based on beta hydroxylase amino acid sequences.

SDIMO group numbers are shown at the right. Tree was constructed in MEGA11 (Tamura *et al.*, 2021) by the maximum likelihood method using default settings from PhyML 3.0 (Guindon *et al.*, 2010) from an alignment (278 aa) including some internal gaps. Bootstrap values >50% are shown at the nodes. The scale bar denotes amino acid substitutions per sequence site. For clarity, only four representatives from each of the pre-existing SDIMO groups were used for tree construction. These are only a small subset of the total diversity, but this was sufficient to cleanly resolve the groups.

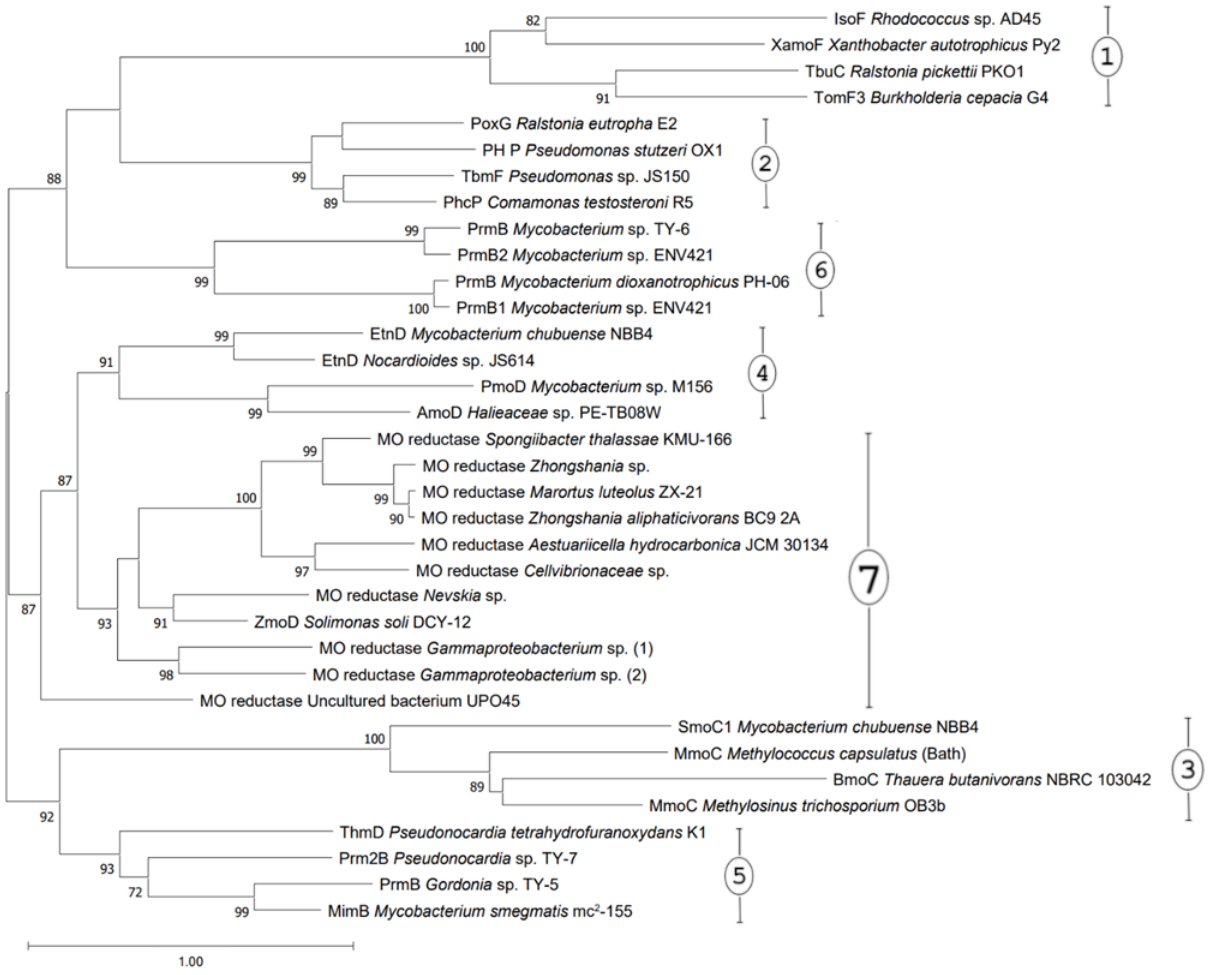


Figure 3.3.4. Phylogeny of SDIMOs based on reductase subunit amino acid sequences.

SDIMO group numbers are shown at the right. Tree was constructed in MEGA11 (Tamura *et al.*, 2021) by the maximum likelihood method using default settings from PhyML 3.0 (Guindon *et al.*, 2010) from an alignment (303 aa) including some internal gaps. Bootstrap values (>50%) are shown at the nodes. The scale bar denotes amino acid substitutions per sequence site. For clarity, only four representatives from each of the pre-existing SDIMO groups were used for tree construction. these are only a small subset of the total diversity, but this was sufficient to cleanly resolve the groups.

Further support for the group 7 SDIMOs as a distinct lineage comes from comparison of their alpha-hydroxylase sequences to those of alkane and alkene MO's, which reveals several residues unique to this group (Figure 3.3.2). A pair of hydrophobic residues (YF, LY, WY, FY, YY) found at positions 140-141 in the other MO's has been replaced with polar amino acids (HN, QN, or HH) in the group 7 enzymes. Other changes include R143, W145, F198, G201, T220, R244, W256,

A335, F339, and H341. Some of these (F198, G201, T220) are close to the 'pore' region (S195, N196, Q222, corresponding to T213, N214 and E240 in *Methylococcus capsulatus* Bath), that extends from the active site pocket to the external solvent and is proposed to be an access route for substrate (Banerjee *et al.*, 2019). The group 7 alpha-hydroxylases have two large aromatic residues F198 and F199 in positions where the sMMO of *M. capsulatus* has L216 and I217 (Sazinsky and Lippard, 2005, McCormick and Lippard, 2011). The conserved T220 residues in group 7 correspond to serine residues in other alpha hydroxylases (e.g. S238 in *mmoX*). This residue is part of the putative electron-transport hydrogen-bonding network (McCormick and Lippard, 2011). The A335, F339 and D344 residues of the group 7 sequences correspond to L353, G357 and A362 in *M. capsulatus* and form part of the network of "cavity 3" residues. Interpretation of the functional significance of these patterns of amino acid changes is hampered by the fact that to date, the only available crystal structures of aliphatic hydrocarbon monooxygenases are of the sMMOs (group 3), and not for any of the other groups in the alignment. Taken together, these observations do suggest that the group 7 SDIMOs are likely to have distinctive substrate specificities and catalytic activities.

Inspection of the genomic context of the group 7 SDIMOs reveals only a few conserved features apart from the four SDIMO subunits (Figure 3.3.5). This contrasts with their nearest relatives, the group 4 SDIMOs, which are embedded in a much larger and well-conserved cluster of catabolic and biosynthetic genes (Mattes *et al.*, 2010). A hydrolase gene (45 % aa identity to the epoxide hydrolase of *Stigmatella aurantiaca* DW4/3-1) is present 7.5 kb downstream of the *Solimonas* SDIMO genes, and homologs of this hydrolase are also present near the group 7 SDIMOs in *Zhongshania, Aestuaricella,* and *Cellvibrionaceae* spp. While this is consistent with a physiological role for the group 7 SDIMOs in oxidising alkenes to epoxides, this hydrolase was not universally present near the *zmo* genes. A *tetR* family regulator was found near the group 7 SDIMO in most of the genomes. This suggests that this SDIMO is inducible in response to a specific signal, presumably its substrate or a metabolite thereof. In the majority of cases, a *tonB* homolog is also found near the SDIMO genes. This is notable since TonB is an energy-transfer protein capable of transferring energy to iron membrane transporters, and all SDIMOs require iron as an essential cofactor.

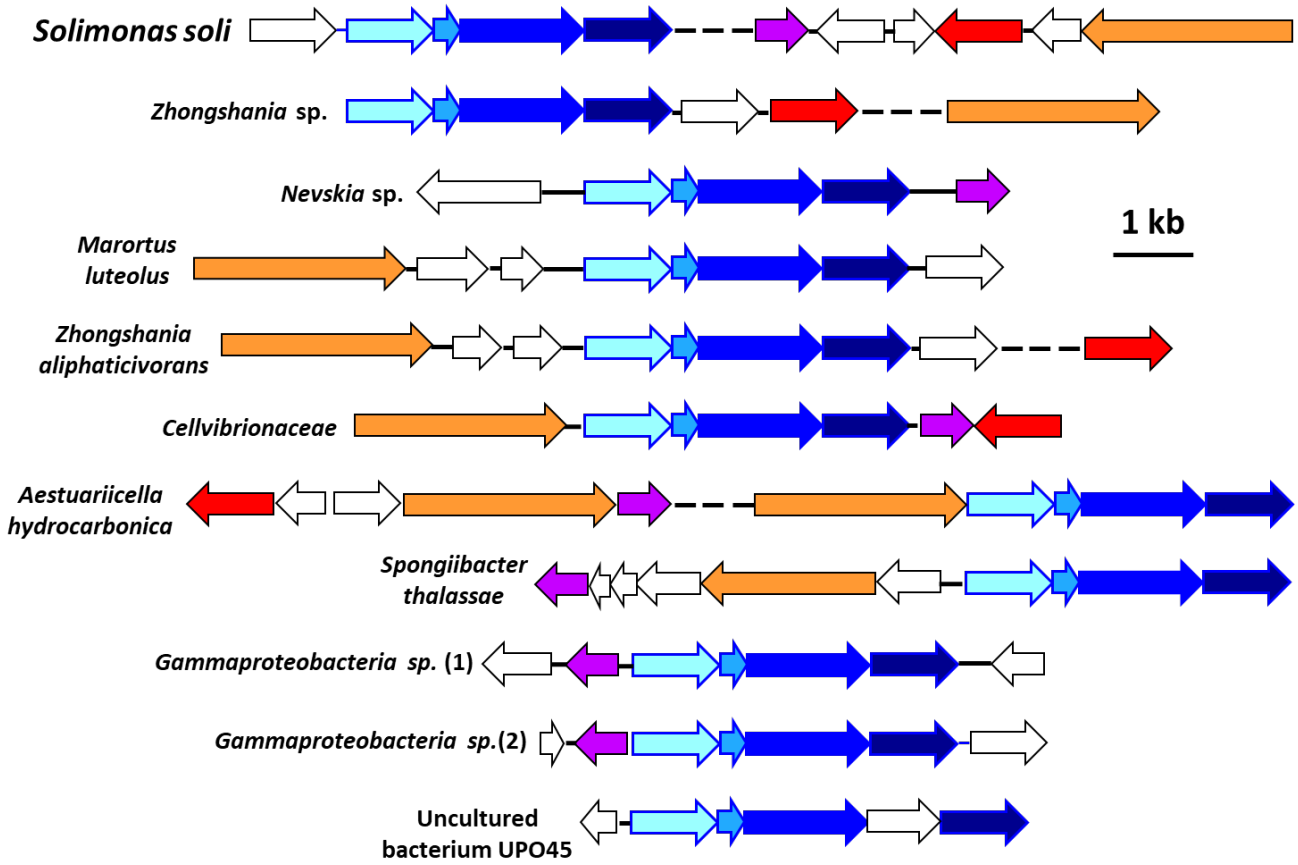


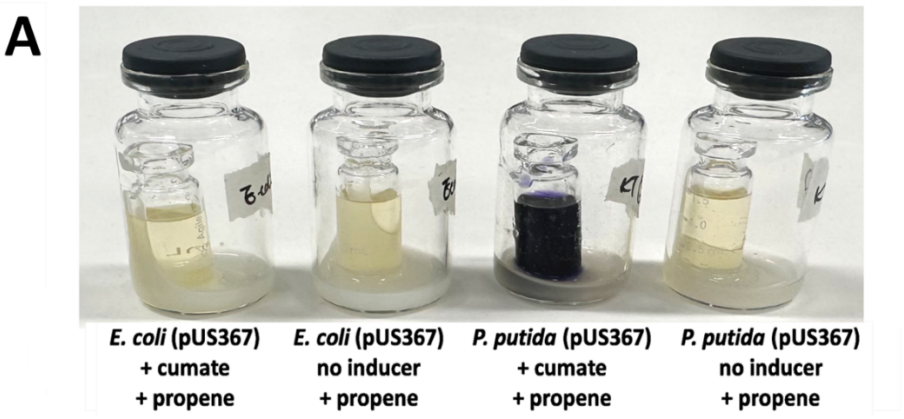
Figure 3.3.5. Comparison of genomic loci containing group 7 SDIMOs in different bacteria.

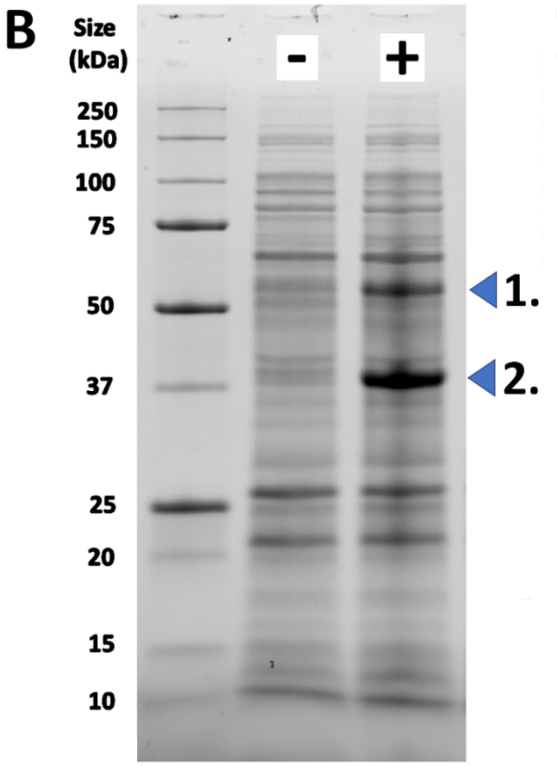
Colour scheme is as follows: blue, group 7 SDIMO; purple, *tetR* family regulator; red, epoxide hydrolase homologue; orange, *tonB* family energy transfer protein; white, all other genes. Accession numbers of each loci are as follows: *Solimonas soli*, NZ_AXDW01000023.1; *Zhongshania* sp., JAGPVP010000005.1; *Nevskia* sp., JADHNH010000061.1; *Marortus luteolus*, NZ_PQGG01000023.1; *Zhongshania aliphaticivorans*, NZ_CACSIJ010000012.1; *Cellvibrionaceae*, PANW01000082.1; *Aestuariicella hydrocarbonica*, NZ_JAAONZ010000002.1; *Spongiibacter thalassae,* NZ_JAAWWK010000007.1; *Gammaproteobacteria* sp. (1), JACPPA010000016.1; *Gammaproteobacteria* sp. (2), SHZJ01000030.1; and Uncultured bacterium UPO45, KU144973.1.

The bioinformatic analysis indicates that the *Solimonas zmoABCD* genes are part of a distinctive and novel SDIMO clade, group 7, with a conserved four-component gene organisation. While some evidence suggests a role in epoxidation of alkenes, the lack of other nearby catabolic genes makes it difficult to hypothesise further about likely substrates or pathways. Therefore, heterologous expression was employed to test the activity of the *Solimonas* enzyme.

3.3.2. *Solimonas* ZmoABCD is a functional MO, and oxidises alkenes to epoxides

The *Solimonas zmoABCD* gene cluster was amplified by PCR and cloned into the broad host range cumate-inducible expression vector pUS250 to make pUS367 (Figure 3.2.1). Note that in this construct, the *zmoABCD* genes retain their original ribosome binding sites but are removed from any native promoter elements and are under the control of the cumate-responsive promoter in the plasmid. The MO activity of cells containing ZmoABCD was tested using a colorimetric assay for epoxides, with ethene as the initial test substrate. Cumate-induced *E. coli* TOP10 cultures carrying pUS367 did not yield any detectable MO activity but transfer of the plasmid into *P. putida* KT2440 yielded recombinants capable of ethene epoxidation (Figure 3.3.6A). SDS-PAGE analysis of lysates from cumate-induced *P. putida* KT2440(pUS367) cells showed strong expression of proteins at sizes consistent with the alpha hydroxylase subunit (ZmoC, 60 kDa) and the beta hydroxylase and/or reductase subunits (ZmoA and ZmoD, 37 and 40 kDa, respectively) (Figure 3.3.6B). These bands were not seen in uninduced controls. LC-MS analysis of excised gel bands confirmed that the 60 kDa band was primarily ZmoC, while the 40 kDa band was mostly ZmoA (Figure 3.3.6B), but with ZmoC and ZmoD also present.





Band 1. LC-MS

Protein identified	Coverage (%)	No. of peptides	Mascot Score
ZmoC (alpha hydroxylase)	74	49	24461
ZmoA (beta hydroxylase)	65	17	1704
ZmoD (reductase)	27	5	225

Band 2. LC-MS

Protein identified	Coverage (%)	No. of peptides	Mascot Score
ZmoA (beta hydroxylase)	72	26	19248
ZmoC (alpha hydroxylase)	70	28	5813
ZmoD (reductase)	63	18	3616

Figure 3.3.6. Expression of *Solimonas* ZmoABCD enzyme in *P. putida* KT2440.

A) Colorimetric assay for propene epoxidation by ZmoABCD. The figure shows the setup of the vial-in-vial method used, and the strong reaction seen with induced cells of *P. putida* KT2440. Note that the purple colour only develops after heating and amine addition. In this image, the smaller vials have been placed inside the larger vials after those steps for the purposes of illustration. B) SDS-PAGE analysis of lysates of *P. putida* KT2440(pUS367) cells, and LC-MS analysis of bands. Arrows indicate two bands in the induced cells (+) which are absent in the uninduced cells (-), and the tables detail the results of LC-MS analysis of these two excised gel bands. Predicted sizes of the SDIMO subunits (kDa) are as follows: ZmoA, 37, ZmoB, 12, ZmoC, 60, ZmoD, 40.

Cumate-induced cells of *P. putida* KT2440(pUS367) oxidised C_2-C_8 alkenes to the corresponding epoxides (Figure 3.3.7A). No epoxides were detected with uninduced cells or in vials lacking added alkenes, confirming that ZmoABCD was responsible, and that the epoxides were derived from the added alkenes. The highest yield of epoxides was seen with propene, with decreasing yields as the size of the alkenes increased above C_3. The activity was highest with simple linear alkenes, but epoxidation activity was also seen with branched (isoprene), carboxyl-substituted (acrylic acid), and chlorinated (VC, cDCE) alkenes. At the time of writing, this is the first report of microbial epoxidation of acrylic acid, as this alkene is normally metabolised via hydrolysis or β-oxidation (He *et al.*, 2020). No significant activity was seen with styrene, allyl alcohol, or 2-methyl-3-buten-2-ol.

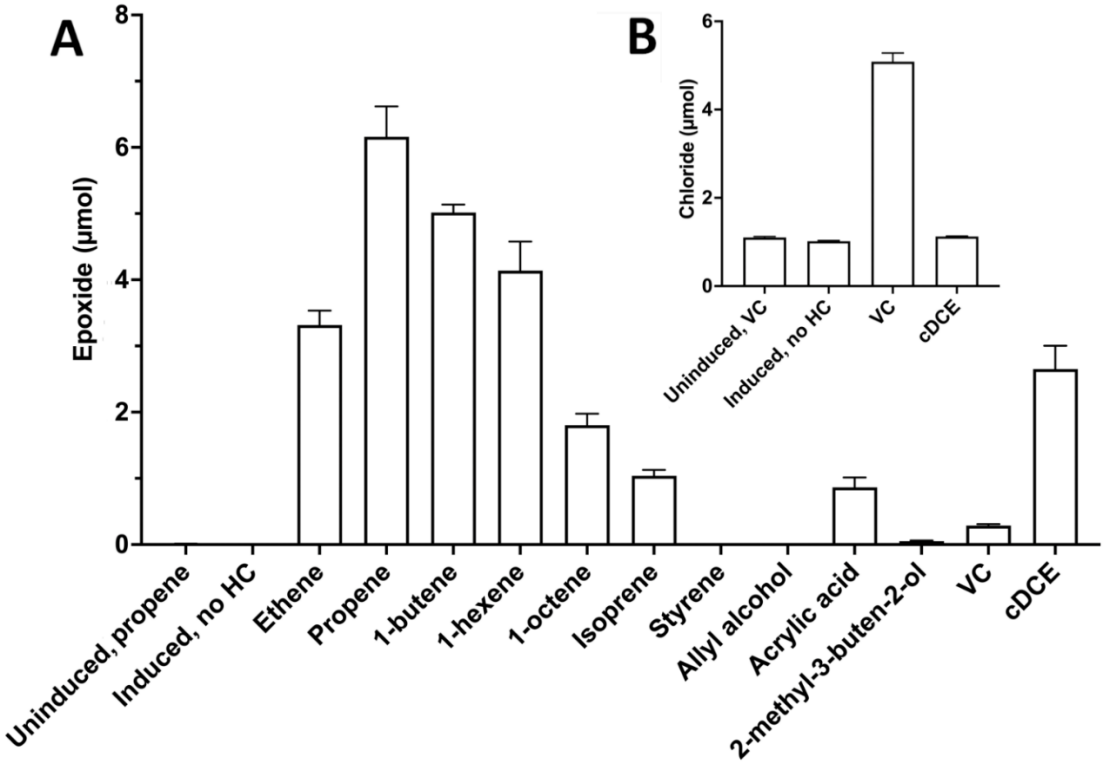


Figure 3.3.7. Oxidation of various alkenes by whole cells of *P. putida* KT2440(pUS367).

A) Epoxidation of alkenes measured via NBP assay. Cumate-induced *P. putida* KT2440(pUS367) cells were used as the catalyst, unless otherwise indicated. B) Production of inorganic chloride from VC and cDCE, as detected by colorimetric assay. For both assays, data shown are the mean of three replicate vials, with error bars indicating the standard deviation.

Based on the end-point epoxide assay, ZmoABCD attacked cDCE at a similar rate to ethene, but the activity against VC appeared to be tenfold lower (Figure 3.3.7A). The biodegradation of the two chlorinated alkenes was investigated in more detail via a colorimetric assay for inorganic chloride release (Figure 3.3.7B). Cumate-induced *P. putida* KT2440(pUS367) cells released substantial amounts of chloride from VC, but not from cDCE, despite the fact that the epoxidation assay indicated much higher activity on cDCE. After correction for background chloride levels in the uninduced control, 3.8 µmol of chloride ions were calculated to be released from 8.3 µmol of VC added, indicating that chloride was released from 46% of this substrate through degradation over the course of the assay.

Several other experiments were also performed to determine if heterologous expression of ZmoABCD has been optimised. The co-expression of *E. coli* chaperones GroES/EL with SDIMOs is a common method to improve SDIMO activity in Gammaproteobacterial heterologous hosts. When this was attempted with ZmoABCD in *P. putida* KT2440, no increase in epoxide production was detected when compared to expression of ZmoABCD alone (Figure 3.3.8). In addition, it is known that *Mycobacterium smegmatis* mc^2-155 is a heterologous host capable of the expression of Actinobacterial SDIMOs. As a result, the expression of the Gammaproteobacterial ZmoABCD in the mycobacterial host was also explored. When ZmoABCD was expressed in the bacteria, no epoxides were produced by *M. smegmatis* mc^2-155 (Figure 3.3.9).

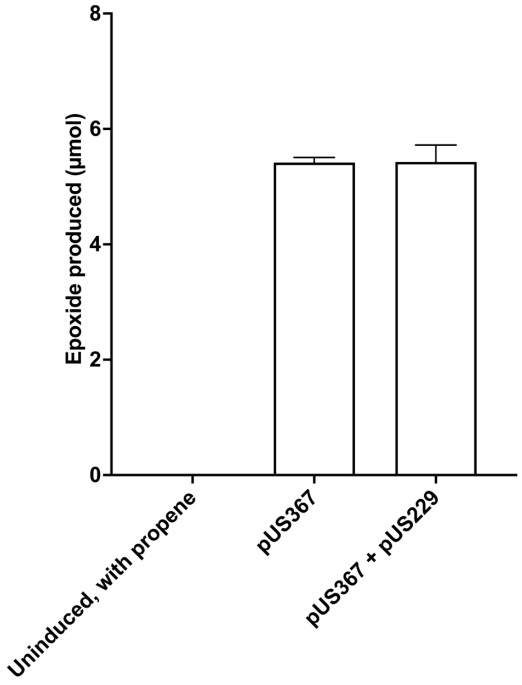


Figure 3.3.8. Effect of chaperonin on propene epoxidation by whole cells of *P. putida* KT2440.

Epoxidation of propene was measured via the NBP assay. Cumate-induced whole cells of *P. putida* KT2440 containing either pUS367 or pUS367+pUS229 were used as the catalyst, except for the control which was uninduced. Data shown are the mean of three replicate vials, with error bars indicating the standard deviation.

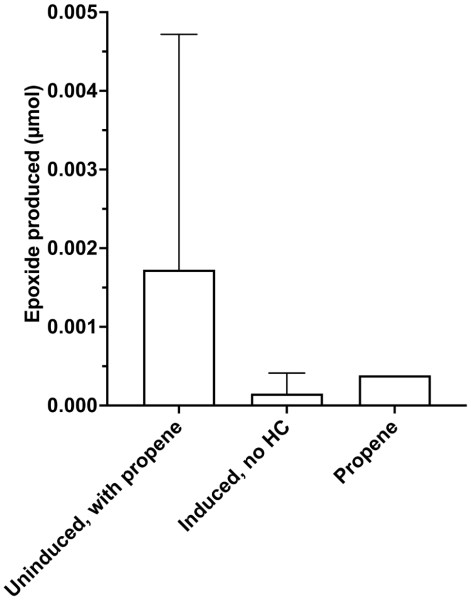


Figure 3.3.9. Test of propene oxidation by whole cells of *M. smegmatis* mc2-155(pUS368).

Epoxidation of alkenes measured via NBP assay. Acetamide-induced whole cells of *M. smegmatis* mc^2-155(pUS368) were used as the catalyst, except in the control, which

contained uninduced cells of the same type. A second control consisted of induced *M. smegmatis* mc^2-155(pUS368) cells with no hydrocarbon (HC) added. Data shown are the mean of three replicate vials, with error bars indicating the standard deviation. The large error bar is due to the lack of epoxide detected by the NBP assay. Note the <0.005 µmol values.

The heterologous expression experiments confirmed that ZmoABCD was a functional alkene monooxygenase, with a preference for small linear substrates, but was also capable of oxidising a diverse range of alkenes, including some priority pollutants. With these clues in hand, the growth substrate range of the original host bacterium *S. soli* was then tested to see if this corresponded to the substrate preferences of the ZmoABCD enzyme (Chapter 4).

3.4. Discussion

In this chapter, a new lineage of SDIMOs (group 7) was defined, which are predominantly found in marine *Gammaproteobacteria*, but which also exist in soil microbes such as *Solimonas soli*. It was demonstrated that the group 7 SDIMO ZmoABCD of *S. soli* is a functional alkene monooxygenase, which can oxidise diverse alkenes to epoxides, and can biodegrade chlorinated ethenes, which are important groundwater pollutants. ZmoABCD had highest activity with small straight-chain gaseous alkenes, and lower levels of activity with branched or substituted alkenes. Thus, the data to date suggest that the canonical substrate of ZmoABCD is likely to be an as-yet-unidentified small volatile molecule.

For reasons that remain unclear, SDIMOs are difficult to express outside of their native hosts (Smith *et al.*, 1999, Chan Kwo Chion *et al.*, 2005). Even the group 1 and group 2 SDIMOs which are functional in *E. coli* (Shields *et al.*, 1995, Canada *et al.*, 2002) give recombinants with approximately 100-fold lower activities than the corresponding wild type hosts (McCarl *et al.*, 2018). This pattern held true for the ZmoABCD enzyme of *Solimonas soli*, which was unable to be expressed in *E. coli*, and only became active after expression in *P. putida*. Since the plasmid construct used in both hosts was identical, there must be trans-acting factors in the *P. putida* KT2440 genome that enable this MO to function. Candidates for this role include chaperonins and oxidative stress defence enzymes (Coleman *et al.*, 2011, Furuya *et al.*, 2012, McCarl *et al.*, 2018), but it is notable that unlike the case for some other

SDIMOs, no such accessory genes were seen near the group 7 SDIMOs in their host genomes, and co-expression of an *E. coli* chaperonin (GroEL) did not stimulate activity in *P. putida* (Figure 3.3.8). *P. putida* has been shown to be a useful host for expression of the ethene MO EtnABCD (McCarl *et al.*, 2018), although that enzyme gave much higher activity in a *Mycobacterium smegmatis* host. Interestingly, the reverse was true for ZmoABCD, which did not yield any significant activity in *M. smegmatis* (Figure 3.3.9). The preference for phylogenetically-related hosts could indicate an issue with codon usage, which has been shown to be important previously for recombinant expression of SDIMOs (Furuya *et al.*, 2013). Note that in this study the codon usage of the *zmoABCD* genes was not modified.

Cells of *P. putida* KT2440(pUS367) formed epoxides from both VC and cDCE, with an apparent preference for cDCE as a substrate, but chloride ions were only released from VC, and not cDCE. There are two unexpected findings here. Firstly, the higher epoxidation activity with the more chlorinated substrate, and secondly, the opposite substrate preferences suggested by each assay. It can be suggested that these apparent inconsistencies are both due to the much higher stability of cDCE epoxide vs. VC epoxide. In an earlier study (Van Hylckama *et al.*, 1996), the half-lives of these metabolites were estimated to be 72 hours (cDCE epoxide) vs. 78 seconds (VC epoxide). Thus, spontaneous breakdown of VC epoxide would compete with the NBP conjugation reaction to a much greater degree than for cDCE epoxide. The breakdown of VC epoxide (to either chloroacetaldehyde or glycolaldehyde (Guengerich, 1992)), would also yield chloride ions much more rapidly than is the case for cDCE epoxide. Therefore, it can be argued first that VC is a better substrate for ZmoABCD than the ‘vial-in-vial’ type epoxidation assay implies, and secondly, that cDCE is also a good substrate for the enzyme, despite the lack of inorganic chloride release seen.

Further work is needed to explore the substrate range and kinetics of the ZmoABCD enzyme e.g. via substrate depletion assays using gas chromatography. These assays would be helpful for resolving the issues discussed above with the differential activities seen with VC and cDCE. Further efforts to improve activity in the recombinant system would aid further characterisation of the enzyme’s activities. Avenues to be pursued could include co-expression of helper proteins (e.g. native *S.*

soli chaperonins), codon optimisation or harmonisation (Angov *et al.*, 2008), modification of the transcription and translation signals, or testing other bacteria as heterologous hosts.

Chapter 4 – Characterisation of ZmoABCD in the *Solimonas soli* Native Host

Partially published as: Yang, S. N. N., Haritos, V., Kertesz, M. A., Coleman, N. V. (2024) A novel soluble di-iron monooxygenase from the soil bacterium *Solimonas soli*. Environmental Microbiology, 26: e16567. DOI: 10.1111/1462-2920.16567.

4.1. Introduction

Environmental organisms harbour suites of genes that encode interesting enzymes. These genes may be involved in reactions that could be beneficial or desired for bioremediation and bioaugmentation (Ayilara and Babalola, 2023), and biocatalysis applications (Sukul *et al.*, 2017). However, environmental organisms are also notoriously difficult to culture in laboratory conditions (Rappé and Giovannoni, 2003). In addition, unlike common laboratory strains, newly discovered environmental isolates are also notoriously difficult to genetically manipulate (Riley and Guss, 2021). As a result, one must first optimise a method to culture the organism of interest then establish a reliable protocol to manipulate its genetics.

Solimonas soli DCY12 is a Gammaproteobacterium isolated from the surface soil of a ginseng field in South Korea (Kim *et al.*, 2007). Much is unknown about this novel organism, and inconsistencies in the reporting of its enzymatic profile and carbon source usage were present (Kim *et al.*, 2007, Sheu *et al.*, 2011). Despite this, analysis of its genome revealed genes encoding a novel soluble di-iron monooxygenase (SDIMO) with its phylogeny between the aromatic hydrocarbon-degrading SDIMOs and the aliphatic hydrocarbon-degrading SDIMOs (Osborne and Haritos, 2019). However, as mentioned in Chapter 3, no obvious clues to the function of the novel SDIMO can be inferred from physiological characteristics reported in the initial publication (Kim *et al.*, 2007). The previous chapter examined the activity of the *S. soli* group 7 SDIMO (ZmoABCD) in heterologous hosts and confirmed its activity against alkenes. Whether *S. soli* is able to utilise alkenes as growth substrates is unknown, and its wider carbon source usage range also remains elusive. As SDIMOs usually facilitate the first step of hydrocarbon metabolism (Chapter 1), it was hypothesised that *S. soli* is able to utilise its ZmoABCD cluster to oxidise short-

chain alkenes and grow on these alkenes and their metabolites (i.e, epoxides and glycols) as sole carbon sources.

The main aim of this chapter is to study the genetics and physiology of *S. soli*, in order to generate ZmoABCD knockout strains of this environmental isolate. To do so, an auxiliary aim of this chapter is to establish reliable protocols to culture and genetically manipulate *S. soli*. In conjunction with the previous heterologous expression experiments, utilising the knockout strains and the rigorous screening of carbon sources, the primary substrate of ZmoABCD can be ascertained.

4.2. Methods and materials

4.2.1. Growth of *Solimonas soli* in microtitre plates

An OD_{600}-normalised (to 1.0) culture of *S. soli* was used as the inoculum for microtitre plate experiments. Further 1 mL cultures were inoculated to starting OD_{600}s of 0.1. Each of these *S. soli* cultures was transferred to a 96-well microtitre plate to 200 µL per well. The lid of the microtitre plate was swabbed with "anti-condensation solution" (0.05% (v/v) Triton-X in 20% ethanol, not sterilised) (Brewster, 2003) which was allowed to dry before returning onto the microtitre plate. The microtitre plate was placed into the CLARIOstar Plus (BMG Labtech) plate reader and was incubated at 30 ºC with double orbital shaking at 300 rpm and OD_{600} measurements were taken hourly for 72 hours.

4.2.2. *Solimonas soli* freezer aliquots generation

To ensure a consistent starting OD_{600} for growth experiments, a large culture of *S. soli* was generated. An *S. soli* colony was used to inoculate 50 mL R2A+ (R2A with additional 5 g/L yeast extract and 10 mM MOPS, Table 2.2.1) and was grown to saturation at 30 °C with 200 rpm shaking (about 7 days). The saturated culture was subcultured into 500 mL R2A+ in a 2.5 L conical flask to an OD_{600} of 0.05 and the culture was grown to mid-exponential phase at 30°C with 200 rpm shaking. The final OD_{600} of the 500 mL culture was measured and the culture was centrifuged at 10000 rcf for 10 minutes. The supernatant was discarded and the cell pellets were resuspended in 50 mL 10% glycerol. The cell suspension was centrifuged again at

10000 rcf for 10 minutes. This was repeated twice more. Then, the pellet was resuspended in 10% glycerol to a theoretical OD_{600} of 30 and were stored in 360 µL aliquots at -80 °C until later use.

4.2.3. Growth of *Solimonas soli* in minimal salts medium with various carbon sources

6.5 mL cultures of MSM (Table 2.2.1) were set up in 27 mL serum bottles inoculated directly with frozen *S. soli* stocks to an initial OD600 of 0.1. Potential carbon and energy sources were added to final concentrations of 5 mM of DMSO stock solutions (for hydrocarbons) and 3 mM of aqueous stock solutions (for amino acids) before crimp-sealing, except for gaseous substrates, which were added at 10 % (v/v) in the headspace after crimp-sealing. After 14 days of incubation, the final OD_{600} of the cultures was measured. All growth tests were performed in triplicate. The negative controls included inoculated bottles with no carbon source added (to test for any growth on residual glycerol), and inoculated bottles with DMSO only added (to test toxicity or growth effects of the solvent).

For growth assessment of knockout strains, colonies of *S. soli* DCY12, 379KO, and 380KO from R2A agar plates were inoculated into 5 mL R2A+. The cultures were grown to saturation at 30 °C with 200 rpm shaking. The final OD_{600} of the cultures were measured, and the OD_{600} of the cultures were normalised to 1 in 1 mL R2A+. 20 mL MSM were inoculated to 1% (v/v) with each of the normalised cultures. Hydrocarbon and amino acid substrates that wildtype *S. soli* was known to utilise as sole carbon sources were supplemented as described above. Then, 6.5 mL of each culture was then transferred to each 27 mL crimp-seal vials which were treated and incubated as described above. The OD_{600} of each vial were measured on the 3rd day of incubation and every 2 days thereafter until the 14th day.

4.2.4. *Solimonas soli* in API 20NE strips

Two API 20NE strips were set up according to the manufacturer's instructions (https://www.biomerieux-usa.com/clinical/api), with minor alterations to the API® AUX Medium of one of the API strips. For the altered medium, 0.1% (v/v) R2A+ was

supplemented and this was used as the inoculum for one of the API strips, as a prior study had done with LB (Kim *et al.*, 2007). The strips were incubated at 30 ºC for 48 hours and were visually inspected for positive results as per the manufacturer's instructions.

4.2.5. *Solimonas soli* in Biolog carbon utilisation assay plates

The Biolog Carbon utilisation assay plates PM1 and PM2 (https://www.biolog.com/) were set up as per the manufacturer's instructions with minor alterations. The Inoculating Fluid-0 (IF-0) with replaced with MSM because *S. soli* was unable to grow in the provided IF-0 medium with 0.1% (v/v) Tween 80 or 5 mM phenol as the sole carbon source. The assay plates were incubated at 30 ºC with 200 rpm shaking for 5 days and were visually inspected for any colour change at 48 hours, 72 hours, and 144 hours.

4.2.6. *Solimonas soli* electroporation trials

Solimonas soli DCY12 was electroporated as follows: 50 µL *S. soli* DCY12 freezer aliquots and 50 ng of plasmid were combined in 2 mm cuvettes, using the conditions in Table 4.2.1. After electroporation, 1 mL room temperature R2A+ was added to each cuvette, the suspensions were transferred to Eppendorf tubes and the cells were allowed to recover for a varying amount of time at 30 ºC with 200 rpm shaking (Table 4.2.1). The suspension was centrifuged at 10000 rcf for 1 minute and the cell pellet was spread onto R2A agar (with the appropriate antibiotics). The agar plates were incubated for 4 days at 30 ºC.

Table 4.2.1. Electroporation conditions used for the transformation of *S. soli.*

Trial	Voltage (V)	Charge (μF)	Resistance (Ω)	Plasmid	Recovery time (hours)	Agar used
1	2500	25	200	pUS250	4	R2A-Km50
2	2500	25	400	pUS250	4	R2A-Km50
3	2500	25	800	pUS250	4	R2A-Km50
4	3000	25	200	pUS250	4	R2A-Km50
5	1500	25	200	pUS250	4	R2A-Km50
6	2500	20	200	pUS250	4	R2A-Km50
7	2500	15	200	pUS250	4	R2A-Km50
8	2500	50	200	pUS250	4	R2A-Km50
9	2500	25	200	pBBR1MCS-5	2	R2A-Gm10
10	2500	25	200	pBBR1MCS-5	4	R2A-Gm10
11	2500	25	200	pBBR1MCS-5	6	R2A-Gm10
12	2500	25	200	pBBR1MCS-2	2	R2A-Km50
13	2500	25	200	pBBR1MCS-2	4	R2A-Km50
14	2500	25	200	pBBR1MCS-2	6	R2A-Km50
15	2500	25	200	pBBR1MCS-2	8	R2A-Km50
16	2500	25	200	pBBR1MCS-2	Overnight	R2A-Km50

4.2.7. Generation of a spontaneous rifampicin resistant *S. soli* mutant

A dense suspension of mid-exponential phase *S. soli* cells was spread onto R2A-Rif400 (R2A with 400 μg/mL rifampicin) agar. The agar plate was wrapped in aluminium foil and incubated at 30 ºC for 1 week. The resulting colonies were then replica-plated onto plain R2A agar and R2A-Rif20 agar and incubated at 30 ºC for 3 days. For colonies that grew well on the R2A-Rif20 agar plates, the replicate colonies from the R2A plate were inoculated into 1 mL R2A+ cultures alongside a culture of wild type *S. soli*. These were transferred to a 96-well plate (as described in Section 4.2.1) for growth analysis and detection of potential growth defects. The colony that grew most similarly to wildtype *S. soli* was chosen and restreaked (from the R2A replicate plate) onto R2A-Rif20 agar, alongside wild-type *S. soli*. The agar plate was incubated for 4 days at 30 ºC. Freezer aliquots of rifampicin resistant (Rif^R) *S. soli* (Named strain DCY12R) were prepared as described in Section 4.2.2.

4.2.8. General conjugation process

E. coli S17-1 or S17-1λ*pir* were transformed with the relevant plasmids as described in Section 2.5.2, and were used as donor strains for conjugation with *S. soli*. 1 mL of each mid-exponential donor strain and mid-exponential *S. soli* DCY12R ("Recipient") were washed three times with 1 mL 20 mM K_2HPO_4 (pH 7) and centrifuged at 10000 rcf for 1 minute. The donor and recipient cell suspensions were then mixed, where the donor:recipient ratio was varied to determine the most efficient ratio (Section 4.3.5). The cell suspension was pipetted onto the surface of R2A agar to form a small "puddle", and the agar plates were incubated at 30 ºC for 2 days in a Tupperware container with wet paper towel to maintain a humid environment.

The resulting "puddle" was resuspended in 1.5 mL of 20 mM K_2HPO_4 (pH 7) and the suspension was transferred to an Eppendorf tube. The cell suspension was serially diluted with 20 mM K_2HPO_4 (pH 7) and plated onto R2A-Rif20 agar with the addition of the relevant antibiotic (plasmid-dependent). The agar plates were incubated at 30 ºC for 3-5 days and putative transconjugant *S. soli* DCY12R colonies were streaked onto R2A-Rif20 agar containing the relevant antibiotic and grown at 30 ºC for 4 days. The presence of the relevant plasmid was confirmed by PCR to ensure that resulting *S. soli* DCY12R colonies were true transconjugants and not spontaneous mutants.

4.2.9. Construction of pUS374

A Golden Gate reaction was set up so that the molar ratio of the synthetic DNA sequences and the pEX18Tc_v2 PCR product (Section 4.3.6) was equal. The reaction was incubated at 37 °C for 1 hour, the enzymes were then heat-inactivated at 60 °C for 5 minutes, and the reaction mixture was chilled at 4 °C. *E. coli* TOP10 cells were chemically transformed with the reaction mixture as described in Section 2.5.2. The correctly constructed plasmid was verified by junction PCRs and was then purified and sequenced. The initial plasmid was named pUS374 (Figure 4.2.1).

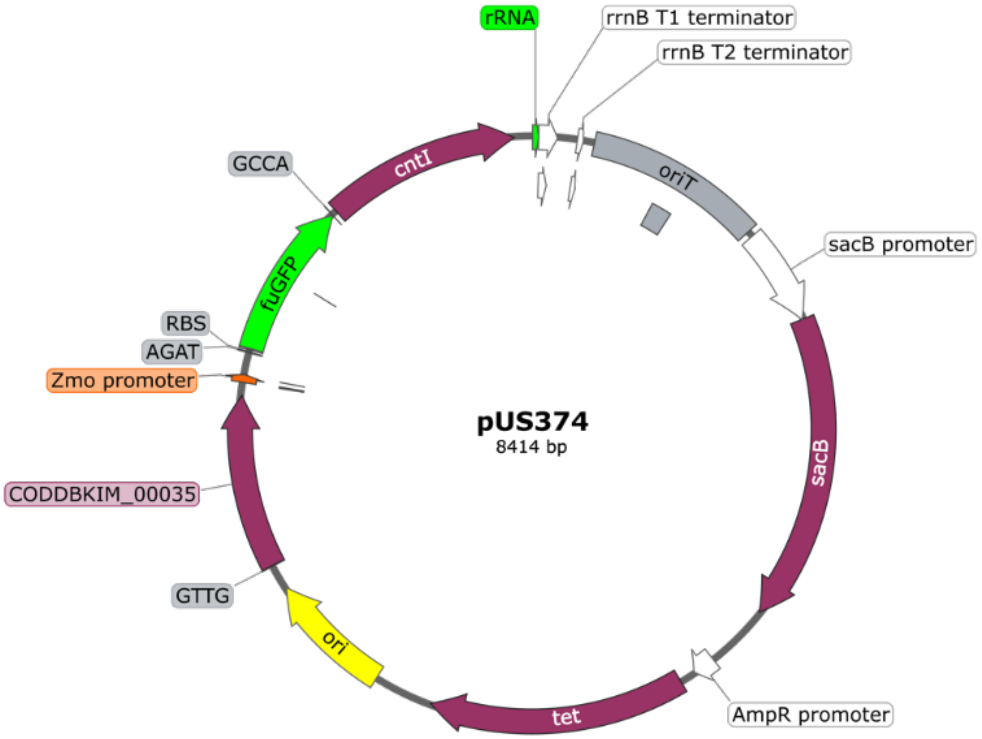


Figure 4.2.1. Map of pUS374.

pUS374 is a 8.4 kb Tc^R plasmid derived from pEX18Tc. Initially planned for use as the allele exchange plasmid to knockout the *zmoABCD* cluster in *S. soli*. For more detail refer to Table 2.4.1.

4.2.10. Construction of pUS379 and pUS380

A new cloning scheme was devised to generate pUS379 (Figure 4.2.2). The knockout construct (the ~2.7 kb span of DNA including the two homology arms and the *fuGFP* gene, termed the "KO cassette") was amplified from pUS374 using primers NIY114 and NIY115 (Table 2.3.2). The KO cassette PCR product and the vector pKNOCK-Km (Table 2.4.1) were digested with SpeI-HF and XhoI and ligated together to form pUS379. pUS380 (Figure 4.2.3) was generated by PCR amplification of the pUS379 backbone (Table 2.3.2), excluding the putative regulatory region between the *zmo* promoter and the *fuGFP* gene (More on this in Chapter 5). These constructs were transformed into *E. coli* CC118λ*pir* and verified by junction PCRs (Tables 2.3.1 and 2.3.2) and were then purified and later sequenced.

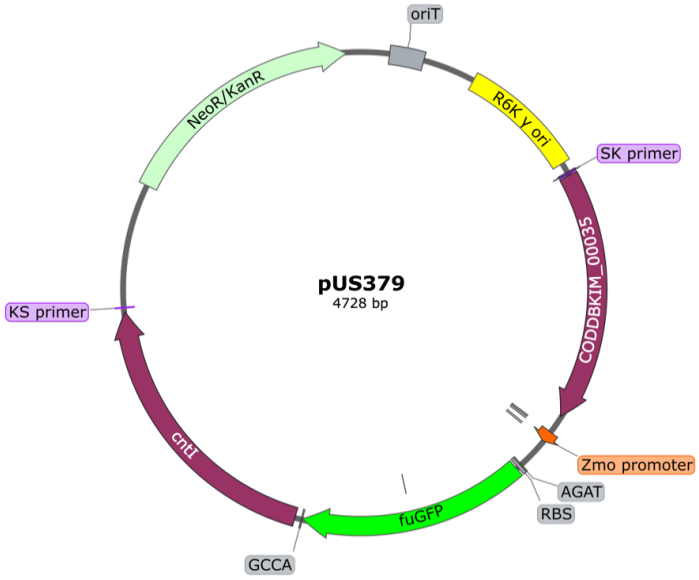


Figure 4.2.2. Map of pUS379.

pUS379 is a 4.7 kb Km^R plasmid derived from pKNOCK-Km. The KO cassette was cloned into pKNOCK-Km. This was used as the allele exchange plasmid to knockout the *zmoABCD* cluster in *S. soli* and replace it with an *fuGFP* gene. For more detail refer to Table 2.4.1.

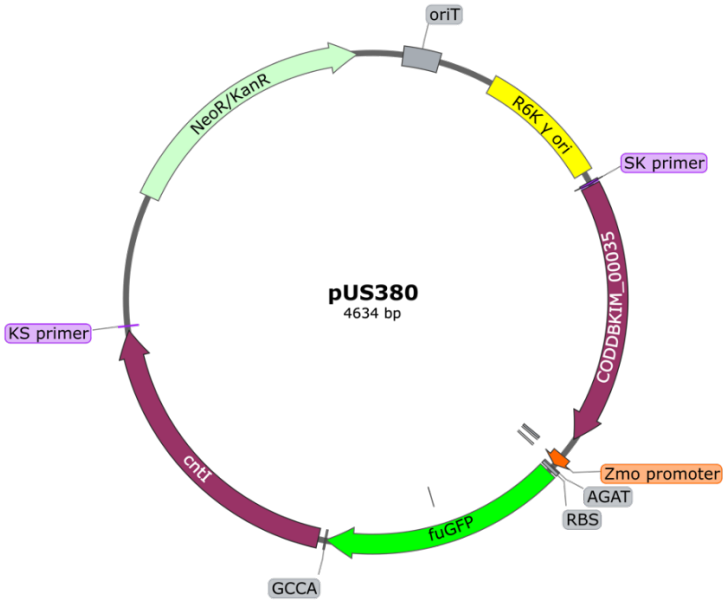


Figure 4.2.3. Map of pUS380.

pUS380 is a 4.6 kb Km^R plasmid derived from pKNOCK-Km. This plasmid is identical to pUS379, except the putative regulatory region between the Zmo promoter and the *fuGFP* gene was removed. This was also used as an allele exchange plasmid to knockout the *zmoABCD* cluster in *S. soli* and replace it with an *fuGFP* gene. For more detail refer to Table 2.4.1.

4.2.11. Construction of pUS382 and pUS383

Two single crossover knockout constructs, pUS382 (Figure 4.2.4) and pUS383 (Figure 4.2.5), were devised to facilitate the truncation of the *zmoC* and *zmoA* genes respectively in *S. soli*. Portions of the *zmoC* and *zmoA* genes were amplified from the *S. soli* genome using primer pairs NIY126 and NIY127, and NIY132 and NIY133 respectively (Table 2.3.2). The *zmoC* gene PCR product was digested with SalI-HF and XbaI while the *zmoA* PCR product was digested with SalI-HF and BamHI-HF. The *zmoC* and *zmoA* PCR products were ligated with pKNOCK-Km (Table 2.4.1) separately to form pUS382 and pUS383 respectively. These constructs were transformed into *E. coli* CC118λ*pir*, verified by junction PCRs (Table 2.3.2), and were then purified and later sequenced to confirm their identities. It should be noted that plasmids that were derived using the pKNOCK backbone are narrow-host range plasmids as they can only replicate in hosts containing the *pir* gene.

Figure 4.2.4. Map of pUS382.

pUS382 is a 2.8 kb Km^R plasmid derived from pKNOCK-Km. A portion of the *zmoC* gene is cloned into pKNOCK-Km. This was used as an allele exchange plasmid to truncate the *zmoC* gene *S. soli* to disrupt the *zmoABCD* cluster. For more detail refer to Table 2.4.1.

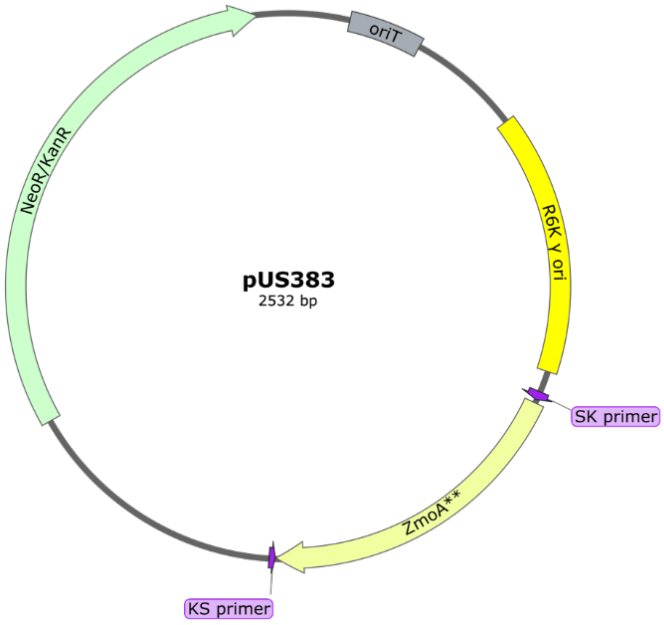


Figure 4.2.5. Map of pUS383.

pUS382 is a 2.5 kb Km^R plasmid derived from pKNOCK-Km. A portion of the *zmoA* gene is cloned into pKNOCK-Km. This was used as an allele exchange plasmid to truncate the *zmoA* gene *S. soli* to disrupt the *zmoABCD* cluster. For more detail refer to Table 2.4.1.

4.2.12. Construction of pUS375

To confirm SacB functionality in *S. soli* and test if SacB can be used as a counterselective marker in the knockout process, plasmid pUS375 (Figure 4.2.6) was constructed. Primers NIY96 and NIY97 were designed to amplify the *sacB* gene from pEX18Tc (Tables 2.3.1 and 2.3.2). The primers included SacI and HindIII sites for cloning into pBBR1MCS-2. *E. coli* TOP10 were chemically transformed as described in Section 2.5.2. The correctly constructed plasmid was verified by junction PCRs and was then purified and the plasmid was sequenced (Table 2.3.2).

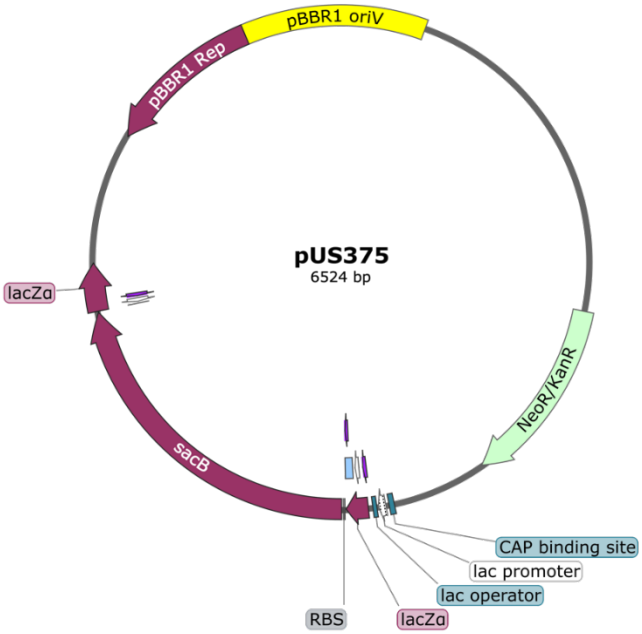


Figure 4.2.6. Map of pUS375.

pUS375 is a 6.5 kb Km^R plasmid derived from pBBR1MCS-2. The *sacB* gene from pEX18Tc was cloned into pBBR1MCS-2 and is expressed constitutively. pUS375 was used to test if SacB is a viable counterselectable marker for allele exchange in *S. soli*. For more detail refer to Table 2.4.1.

4.3. Results

4.3.1. *Solimonas soli* grows best in buffered-yeast-extract-supplemented R2A cultures

When *S. soli* was grown in R2A, the final OD_{600} of the cultures did not exceed 0.3 (Figure 4.3.1). The initial hypothesis was that the bacteria were limited by the minimal amount of carbon in the R2A medium. However, when 20 µg phenol red was added to the saturated culture it was discovered that the pH of the culture was alkaline, suggesting a release of ammonia from amino acid usage (Figure 4.3.2). Concerned about the innate buffering capabilities of the phosphate in R2A, MOPS was added as a buffering agent. Despite this, the growth of the *S. soli* cultures remained low (Figure 4.3.1). An effort to determine if *S. soli* growth could be optimised through increasing certain components in R2A was attempted. It was found that an increased amount of yeast extract (along with 10 mM MOPS) greatly improved the growth yield of *S. soli* (Figure 4.3.1). The supplementation of glucose

was not attempted as *S. soli* was reported to not utilise the sugar by previous studies.

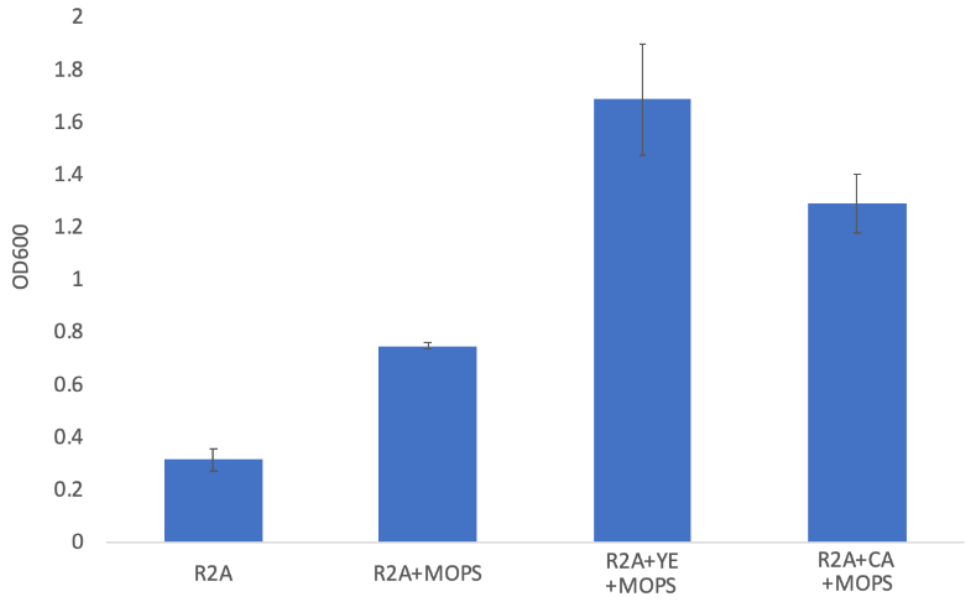


Figure 4.3.1. OD_{600} of *S. soli* cultures after 5 days of growth.

Solimonas soli was inoculated into 50 mL variations of R2A cultures. The OD_{600} of the cultures were measured after 5 days. Each data point is an average of three replicates, each scale bar represents the standard deviation of each data point. MOPS = 10 mM MOPS, YE = 5 g/L yeast extract, CA = 5 g/L casamino acids.

Figure 4.3.2. pH of *S. soli* culture after 5 days of growth.

20 µg phenol red was added as a pH indicator to sterile R2A and to a *S. soli* R2A culture after 5 days of growth.

To determine the exact impact of the addition of 10 mM MOPS and 5 g/L yeast extract, *S. soli* cultures were inoculated into a microtitre plate containing variants of R2A and the growth yield was monitored by OD_{600}. Four variants of R2A were tested: 1 mL R2A, 1 mL R2A with the addition of 10 mM MOPS, 1 mL R2A supplemented with 5 g/L yeast extract, and 1 mL R2A with the addition of 10 mM MOPS and 5 g/L yeast extract. It was evident that the addition of MOPS or yeast extract alone could not support increased *S. soli* growth. Yeast extract must be supplemented and MOPS must be added for pH buffering to achieve optimal growth (Figure 4.3.3). However, as the aim of this experiment was to optimise the growth of *S. soli*, the impact of vitamins through the addition of yeast extract was not assessed.

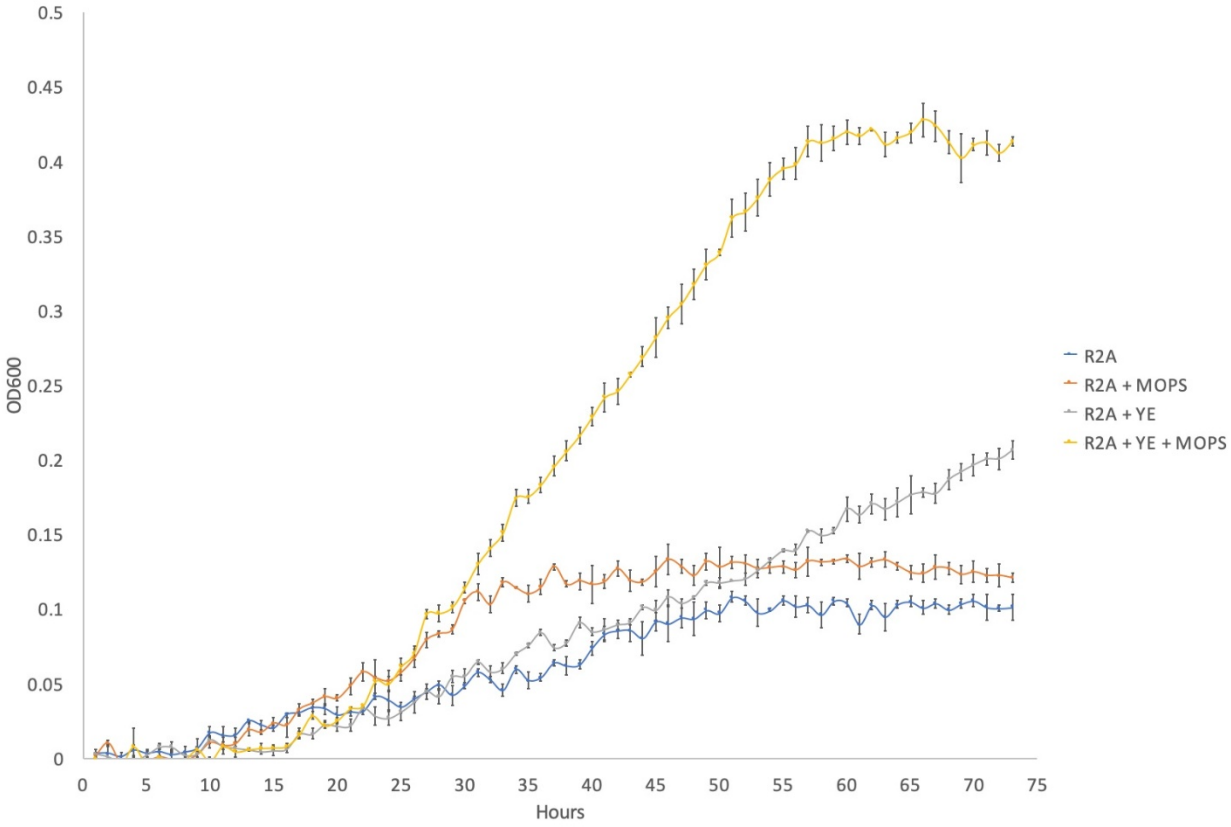


Figure 4.3.3. OD_{600} of *S. soli* cultures after 72 hours of growth in four modified R2A media.

Solimonas soli was inoculated into 200 µL variations of R2A cultures in a 96-well microtitre plate. The OD_{600} of the cultures were measured hourly for 72 hours. Each data point is an average of three replicates, each scale bar represents the standard deviation of each data point. MOPS = 10 mM MOPS, YE = 5 g/L yeast extract.

4.3.2. *Solimonas soli* is able to utilise Tween 80 as a sole carbon source

In order to exclude potential auxotrophic growth requirements, *S. soli* must be able to demonstrate growth using a sole carbon source in minimal medium. Two API strips (API 20NE, bioMerieux) were utilised to test for the substrate range of *S. soli*. One API strip was inoculated as per the manufacturer's instructions, while the medium of the other API strip was supplemented with 0.01% (v/v) R2A+ (Section 4.2.4). Wildtype *S. soli* was able to assimilate mannose in the non-supplemented API strip, and it was capable of hydrolysis using β-glucosidase in both the supplemented and non-supplemented API strips. However, in the single experiment performed, *S. soli* was unable to grow using mannose as a sole carbon source in MSM. As a result, a search for a substrate *S. soli* was able to utilise as a sole carbon source was attempted. Twelve different carbon sources were tested. These substrates were chosen because they ranged from commonly used by microorganisms (e.g. glucose, *N*-acetylglucosamine), their proximity to central metabolism (e.g. sodium acetate, succinate), substrates that alkene oxidising bacteria can utilise (e.g. ethylene glycol), and Tween 80, a detergent that some environmental organisms are able to break down. It was discovered that *S. soli* was able to utilise Tween 80 as a sole carbon source (Figure 4.3.4).

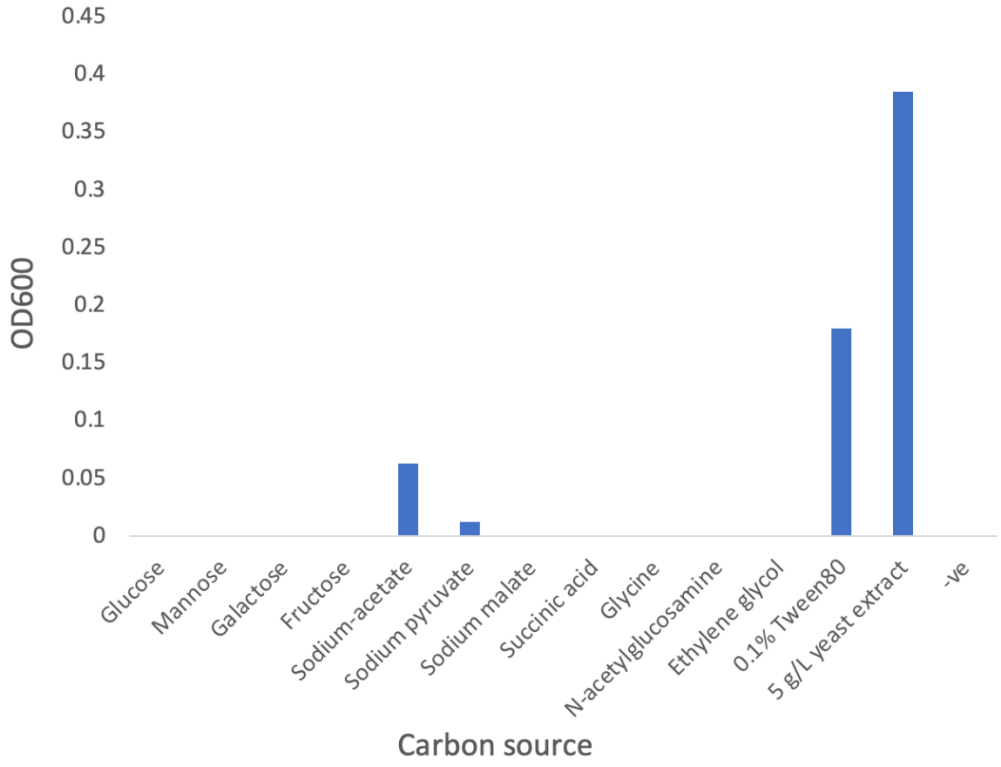


Figure 4.3.4. The growth yield of *S. soli* on 12 different carbon sources in 30 mL MSM cultures after 5 days of growth.

Carbon sources were added to 10 mM, except for Tween 80, which was added to 0.1% (v/v). 5 g/L yeast extract was included as a positive control. -ve: no carbon source added. No replicates were included in this experiment.

4.3.3. *Solimonas soli* grows on *n*-octane, phenol, isoleucine, and leucine but not alkenes

Despite the demonstrated activity of ZmoABCD on alkenes when expressed in *P. putida* KT2440 (Chapter 3), *S. soli* was unable to grow using any of the tested alkenes as sole carbon and energy sources in minimal medium (Figure 4.3.5). While the vials were crimp-sealed, the vials still allowed aerobic growth due to the large headspace volume of air (20.5 mL), therefore access to oxygen should not have limited *S. soli* growth on hydrocarbons. Since the bacterium grew well on Tween 80 as a sole carbon and energy source in minimal media, autotrophy can also be excluded. Perhaps there is another growth requirement that the minimal medium and incubation conditions did not provide and as a result *S. soli* was unable to growth on short-chain alkenes and alkanes. Tween 80 is a detergent based on sorbitan (which contains a tetrahydrofuran-like group) and contains a long chain fatty acid group. While tetrahydrofuran was tested as a growth substrate and did not support growth

of *S. soli* in minimal medium (Figure 4.3.5), neither sorbitan, sorbitol, or fatty acids like oleate were tested as sole carbon sources. Interestingly, *S. soli* was found to be capable of growth on phenol as a carbon source, and this gave the highest growth yield of any compound tested. Phenol is catabolised aerobically via monooxygenases (Furuya *et al.*, 2011, Zhou *et al.*, 2016), so this could be the native substrate for ZmoABCD. Growth was also seen on *n*-octane, which is another substrate that needs an MO for initial oxidation (Baptist *et al.*, 1963). Additionally, *S. soli* was able to utilise isoleucine and leucine as sole carbon sources (Figure 4.3.5). While growth substrate testing revealed that *S. soli* was able to grow on some aliphatic and aromatic hydrocarbons, the bacterium did not grow on any of the alkenes tested. As a result, no obvious links between the activity of ZmoABCD observed in heterologous expression experiments (Chapter 3) and the growth substrate range of the host bacterium were found. Therefore, a plan to knockout the *zmoABCD* cluster from the genome of *S. soli* was devised, as knockout mutants will allow determination of the potential role of the SDIMO in the metabolism of these carbon sources.

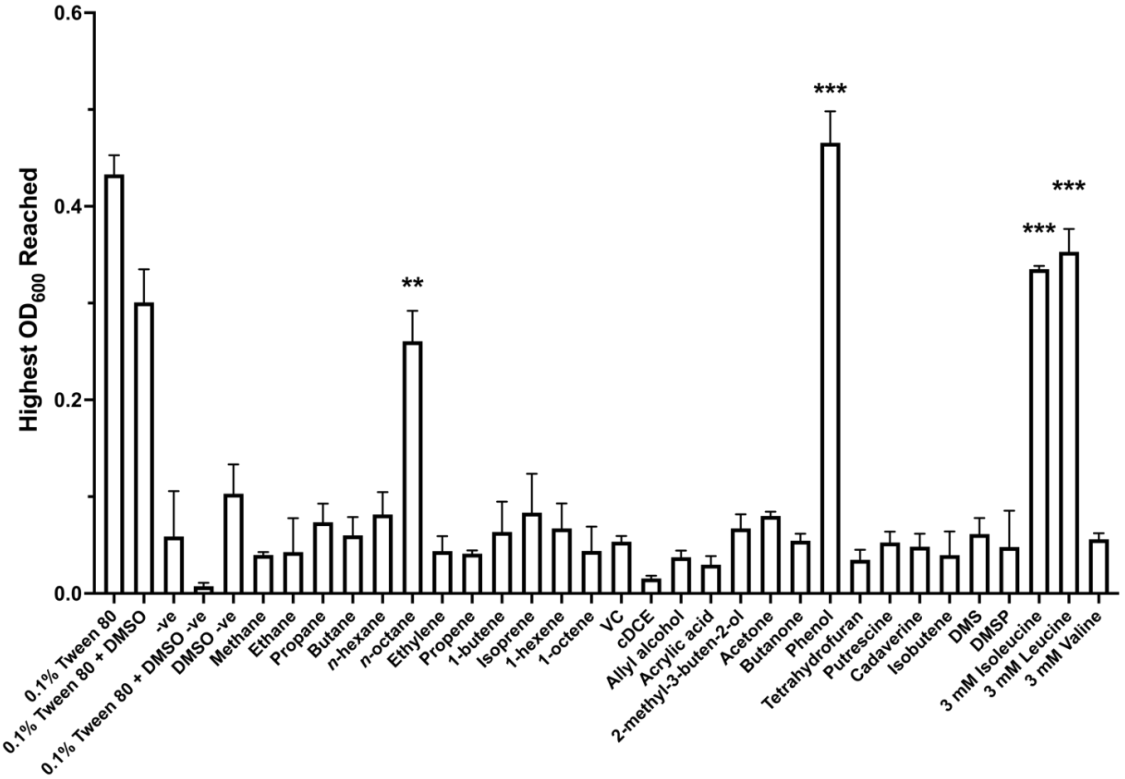


Figure 4.3.5. Growth of *S. soli* on potential SDIMO substrates.

Growth yields were determined as OD_{600} after incubation at varying periods of time. The reported OD_{600} values were the highest OD_{600} reached by the cultures. Liquid substrates were

added as stock solutions in DMSO or water at 5 mM final concentration (except Tween 80 and the amino acids), while gaseous substrates were added as neat gases to 10% of the headspace volume. 0.1% (v/v) Tween 80 was used as a positive control for growth. The negative control ('-ve') had no carbon source added, while DMSO alone ("DMSO -ve") was tested as a solvent control. A no inoculum control (0.1% Tween 80 + DMSO -ve) was also included. Data are the means of three replicates, with error bars indicating the standard deviation. Carbon sources showing significant differences in growth yield compared to the -ve control are shown as either *** ($p<0.001$) or ** ($p<0.01$). VC: vinyl chloride, cDCE: *cis*-1,2-dichloroethylene, DMS: dimethyl sulphide, DMSP, dimethylsulfoniopropionate.

4.3.4. *Solimonas soli* in Biolog carbon utilisation assay plates

The Biolog PM1 and PM2 Carbon Utilisation Assay plates are an effective method to screen for the utilisation of a broad range of carbon sources. Oxidase activity is vital to these assays, as the assay relies on oxidases to interact with a redox dye to produce a colour change. Since *S. soli* DCY12 is an oxidase-positive organism (Sheu *et al.*, 2011), and a high-throughput screening method for carbon utilisation was required, the Biolog PM1 and PM2 plates were employed.

Interestingly, *S. soli* was unable to grow in the supplied Biolog IF-0 medium when it was supplemented with either phenol or Tween 80. Since the composition of the IF-0 medium is proprietary information, it is unknown if a component deterred the growth of *S. soli*, or if a key trace element was absent. As a result, MSM was used as the inoculating fluid in place of the provided IF-0 medium. However, *S. soli* was unable to produce the expected colour change in the Biolog carbon utilisation assay plates, even in the wells containing carbon sources known to be utilised by *S. soli* (Tween 80 and phenol). This contradicts Figure 4.3.5 as *S. soli* is shown to utilise Tween 80 and phenol as a sole carbon sources. As a result, the data from this experiment were inconclusive. *S. soli* DCY12 was also unable to grow in MSM cultures in standard 96-well microtitre plates (200 µL cultures) with Tween 80 and phenol as sole carbon sources (Data not shown), but was able to when R2A+ was the growth medium. However, the cause of this discrepancy was not pursued.

4.3.5. *Solimonas soli* was unable to take up DNA via electroporation but could do so via conjugation

In order to explore the substrate range of the ZmoABCD cluster *in vivo*, a plan to knockout the *zmoABCD* gene cluster from the *S. soli* genome was devised. To attempt allele exchange, a method to introduce DNA into *S. soli* was required. The most common method to transform environmental isolates is via electroporation, therefore, electroporation was the first method attempted. Initially, pUS250, a pBBR1MCS-2 derived broad-host range plasmid for Gram negative organisms, was chosen as the plasmid for the initial trials for its *amilCP* gene. When expressed, the AmilCP chromoprotein would be a distinct marker distinguishing true transformants from spontaneous mutants. When no transformants resulted from electroporation using pUS250 despite many attempts and variations to the electroporation protocol (Section 4.2.6), pBBR1MCS-5 and pBBR1MCS-2 were utilised instead, as they do not harbour a potentially toxic chromoprotein gene. However, no transformants resulted from electroporation with the pBBR1 series of plasmids either. While the ability of *S. soli* to harbour and replicate pBBR1MCS-5 and pBBR1MCS-2 was now in doubt, later conjugation experiments showed *S. soli* can receive and maintain these plasmids.

As an alternative method of transformation, conjugation was trialled, to better success. Using pBBR1MCS-2, different conjugation conditions were tested (Table 4.3.1). It was found that allowing conjugation to proceed for 2 days on R2A agar at 30 °C with a 1:1 donor to recipient ratio of exponential phase cells (both donor and recipient) resulted in the highest conjugation efficiency (7.96×10^{-5} transconjugants per donor cell, Table 4.3.1), with the long conjugation required (2 days) likely due to the slow growth rate of the *S. soli* recipient. Those conditions were selected for future conjugation experiments. In addition, it was discovered that the addition of 50 µg/mL of kanamycin to media was too high for *S. soli* transconjugants, so the concentration of kanamycin was lowered to 20 µg/mL in all subsequent *S. soli* experiments requiring its use.

Table 4.3.1. The conjugation efficiency of pBBR1MCS-2 by *E. coli* S17-1 (donor) to *S. soli* DCY12R (recipient).

Trial	Time on agar (days)	Temperature (°C)	Ratio (Donor:Recipient)	Conjugation efficiency (Transconjugants per donor cell)
1	1	30	1:1	1.32×10^{-6}
2	1	30	1:3	2.20×10^{-7}
3	1	30	3:1	1.60×10^{-6}
4	2	30	1:1	7.96×10^{-5}
5	2	30	1:3	3.43×10^{-5}
6	2	30	3:1	6.47×10^{-5}
7	3	30	1:1	6.22×10^{-5}
8	3	30	1:3	6.37×10^{-7}
9	3	30	3:1	5.26×10^{-5}

4.3.6. Knockout cassette construct design

Primers NIY94 and NIY95 were designed to amplify pEX18Tc (Tables 2.3.1 and 2.3.2), allowing removal of the original multiple cloning site (MCS) along with regions flanking the MCS and the introduction of BsaI sites to enable Golden Gate cloning. The amplified backbone PCR product was named pEX18Tc_v2.

Three synthetic DNA constructs were designed, with compatible BsaI sites to enable Golden Gate cloning into pEX18Tc_v2 and adaptors for amplification if required. These were designed to form a construct with two homology arms that flank an *fuGFP* reporter gene. The left homology arm construct consisted of 1000 bp, of which included part of the gene upstream of *the S. soli* ZmoABCD cluster (CODDBKIM_00035), the putative ZmoABCD promoter and the putative regulatory region. The construct ended before the start codon of *zmoA*. An internal BsaI site was removed by altering the codon from GTC (Val) to GTT (Val). An attempt to maintain the codon usage frequency as much as possible (from 0.40 to 0.22) was made. The codon usage of *S. soli* was calculated using the Biologics Codon Usage

Calculator tool (https://www.biologicscorp.com/tools/CodonUsageCalculator/). The codon usage values were obtained using the ZmoABCD cluster as the entry for the calculator. The right homology arm construct was also 1000 bp, beginning after the stop codon of *zmoD*. The construct included the *cntI* gene, downstream of the *zmoABCD* cluster. The sequence of the *fuGFP* reporter was obtained from pUS251-fuGFP-tag (Mark Somerville, unpublished data), and internal cryptic promoters were removed from the *fuGFP* sequence while preserving the amino acid sequence. The DNA constructs were synthesised by TWIST Bioscience and were resuspended to 10 ng/µL in TE buffer (10 mM Tris, 1 mM EDTA, pH 8). The assembled knockout construct (the ~2.7 kb span of DNA including the two homology arms and the *fuGFP* gene) was termed the "KO cassette" (Figure 4.3.6). pUS374 was the construct these synthetic sequences were designed for (Section 4.2.9), but pUS379 and pUS380 were the constructs that served as the final allele exchange plasmids (Section 4.2.10) used to generate the *S. soli* knockout strains.

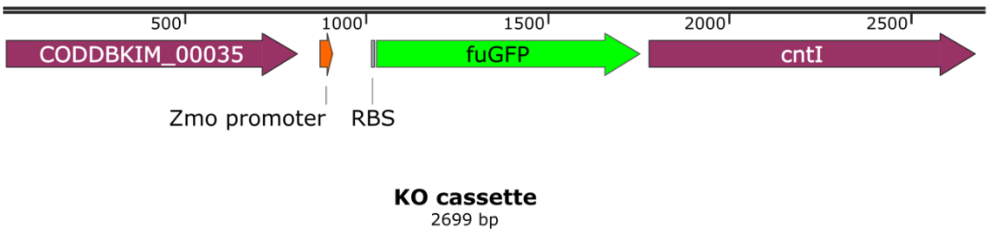


Figure 4.3.6. Map of the knockout cassette used to generate *S. soli* knockout mutants.

The *fuGFP* reporter gene is placed under the control of the native *zmo* promoter. The *fuGFP* gene is situated between partial sequences of genes found upstream (CODDBKIM_00035) and downstream (cntI) of the native *zmoABCD* cluster. This cassette was initially assembled using Golden Gate cloning into a pEX18Tc derived backbone to form pUS374. It was later amplified using PCR and cloned into pKNOCK-Km to form pUS379.

4.3.7. SacB and TetA are toxic or non-functional in *S. soli*

pEX18Tc (Table 2.4.1) was initially chosen as the plasmid backbone for the knockout construct because of its SacB counter-selectable marker and the potential to utilise the tetracycline resistance as a fallback counter-selection screen (Li *et al.*, 2013). When pUS374, a pEX18Tc derived plasmid that harbours the KO cassette (Figure 4.3.6) was conjugated to *S. soli* DCY12R, no *S. soli* colonies were present on R2A-

Rif20-Tc10 agar plates. Therefore, to determine if the constitutive expression of the SacB protein led to the absence of transconjugants, pUS375 was designed. pUS375 is derived from pBBR1MCS-2 and contains the *sacB* gene from pEX18Tc placed under the control of the constitutive promoter upstream of the MCS of pBBR1MCS-2 (Figure 4.2.6). When pBBR1MCS-2 and pUS375 were separately conjugated into *S. soli* DCY12R, no *S. soli* transconjugants harbouring pUS375 were present on R2A-Rif20-Km20 agar plates (Figure 4.3.7). Simultaneously, to determine if the tetracycline resistance gene was non-functional or toxic in *S. soli* DCY12R, pBBR1MCS-3, another broad host range plasmid for Gram negative organisms that confers tetracycline resistance through the *tetA* gene (Table 2.4.1) was conjugated to *S. soli* DCY12R. Note that while the promoters of the tetracycline resistance genes are different in pBBR1MCS-3 and pEX18Tc, the amino acid sequence of the genes are identical. As above, no pBBR1MCS-3 transconjugants were present on R2A-Rif20-Tc10 agar plate. These results suggest that the SacB protein is toxic and the TetA protein is toxic or non-functional in *S. soli*. As a result, pKNOCK-Km (Table 2.4.1) was chosen to be the backbone for the knockout construct since it is known that the kanamycin resistance gene is functional in *S. soli.*

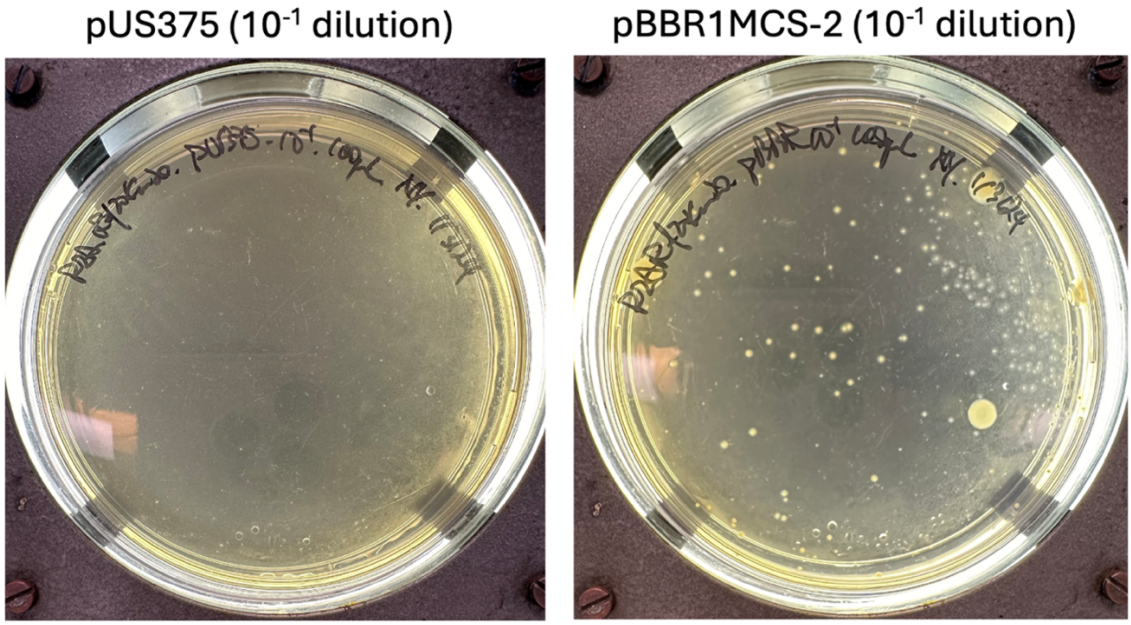


Figure 4.3.7. *Solimonas soli* DCY12R colonies arising from pUS375 and pBBR1MCS-2 conjugations using *E. coli* S17-1 as donors.

100 µL of the 10^{-1} dilution of each conjugation "puddle" cell suspension was spread onto R2A-Rif20-Km20 agar. The large colonies on the pBBR1MCS-2 plate are spontaneous Rif^R *E. coli* S17-1 donor mutants, where the small colonies are *S. soli* DCY12R transconjugants.

4.3.8. Knockout strain generation using allele exchange

For a double crossover knockout, pUS379 and pUS380 were introduced to *S. soli* DCY12R via conjugation as described in Section 4.2.8. The conjugation mixtures were spread onto R2A-Rif20-Km20 agar and Km^R colonies were restreaked until single colonies were obtained. Integration of the plasmid into the correct position in the genome was confirmed by PCR of the left homology arm junction using primers NIY105 and NIY118 (Table 2.3.2). Merodiploid *S. soli* were named *S. soli* 379KOM or 380KOM (KOM for Knockout – Merodiploid), depending on the plasmid that was integrated into the genome. To ensure purity of *S. soli* 379KOM and 380KOM colonies, the colonies were restreaked onto R2A-Rif20-Km20 agar. Once grown, one colony of each merodiploid was selected to induce the second recombination. The merodiploid colonies were inoculated into 5 mL R2A+ and were incubated at 37 °C with 200 rpm shaking for 4 days under non-selective conditions (i.e. no antibiotics). 100 µL of each culture was subcultured into fresh 5 mL R2A+ and were incubated with the same conditions again. This was repeated once more for a total of three growth cycles in 5 mL R2A+ at 37 ºC. Then, the *S. soli* 379KOM and 380KOM cultures were serially diluted in R2A+ to 10^{-7}. 100 µL of the 10^{-5}, 10^{-6}, and 10^{-7} dilutions were spread onto R2A agar and were incubated at 30 °C until colonies were visible. The resulting colonies were screened by PCR to identify successful double crossover knockout mutants using primers NIY106 and NIY119 by amplifying a spanning region across the knockout site (Table 2.3.2). Colonies that exhibited weak bands for the other homology arm junction (the right homology arm, in this case) were assumed to contain cells that had successfully carried out the second recombination event, resulting in a knockout strain. These colonies were inoculated into 5 mL R2A+ cultures, incubated at 30 °C with 200 rpm shaking for 4 days, and serially diluted, spread plated, and incubated as described above. Resulting colonies were replica-plated onto R2A-Km20 agar and R2A agar (in that order) and incubated at 30 °C for 3 days. Kanamycin susceptible *S. soli* replicate patches (on the R2A agar plate) were restreaked onto R2A-Km20 agar and R2A agar (in that order) for confirmation of purity and kanamycin susceptibility. Putative *S. soli* 379KO or 380KO colonies from the streak plates were then screened using PCR for both homology arm junctions as above, *zmoABCD* or *fuGFP* presence (Spanning region, using NIY118 and NIY119), and the pUS379 or pUS380 plasmid backbone (Using the kan^R gene as the marker, primers LFE3 and LFE4) (Table 2.3.2). The spanning region

PCR products of each *S. soli* knockout strain were sequenced to ensure the correct integration of the knockout cassette.

For the single crossover knockouts, the plasmids pUS382 and pUS383 contain an internal fragment of *zmoC* and *zmoA* respectively. Both pUS382 and pUS383 would lead to a single crossover event in the genome of *S. soli* in *zmoC* and *zmoA* respectively, leading to incomplete fragments in the *zmoABCD* operon and therefore resulting in deletion mutants. *S. soli* DCY12R was transformed with pUS382 and pUS383 via conjugation as described in Section 4.2.8. The conjugation mixtures were plated onto R2A-Rif20-Km20 agar, and resulting Km^R colonies were assumed to have successfully integrated the knockout plasmids. Merodiploid *S. soli* were named *S. soli* 382KOM or 383KOM depending on the plasmid that was integrated. To ensure purity of the *S. soli* 382KOM and 383KOM colonies, the colonies were restreaked onto R2A-Rif20-Km20 agar. Integration into the correct position in the genome was confirmed by PCR of the two junctions with primer pairs NIY116 and NIY131, and NIY117 and NIY130 for pUS382, and primer pairs NIY117 and NIY136, and NIY116 and NIY137 for pUS383. (Table 2.3.2). Experiments with these merodiploids were discontinued due to the successful generation of the 379KO and 380KO knockout strains.

4.3.9. Deletion of ZmoABCD is not deleterious to *Solimonas soli*

After successful generation of the *S. soli* 379KO and 380KO knockout mutant strains, the physiology of these mutant strains in R2A+ was tested. No obvious differences in growth rates or final optical density were observed when strains of these strains of *S. soli* were grown in R2A+ medium in a 96-well microtitre plate (Figure 4.3.8). This suggests that the deletion of the *zmoABCD* cluster and replacement with the *fuGFP* gene was not deleterious to *S. soli* during growth in R2A+, as these mutations were not lethal or inhibitory in these growth conditions. Since this experiment confirmed that the knockout mutant strains displayed no obvious phenotypic difference in physiology when grown in complex medium, *S. soli* 379KO and 380KO were provided with the carbon sources able to support growth in

wildtype *S. soli*. This was to elucidate the canonical substrate of the ZmoABCD cluster.

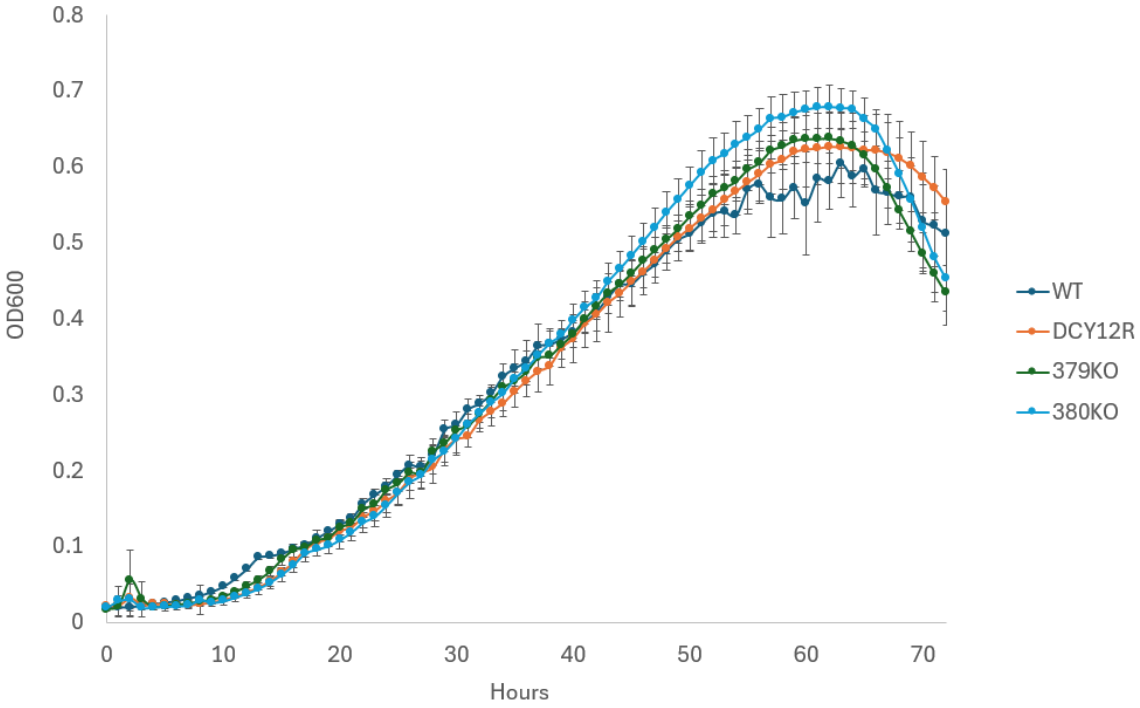


Figure 4.3.8. OD_{600} of different strains of *S. soli* grown in R2A+ liquid cultures over 72 hours.

Strains of *S. soli* were grown in a 96-well microtitre plate. The OD_{600} of the cultures were measured hourly for 72 hours. Each data point is an average of ten replicates, each scale bar represents the standard deviation of each data point. WT = *S. soli* DCY12.

4.3.10. *Solimonas soli* KO mutant growth experiments

The *S. soli* knockout strains (both 379KO and 380KO) showed no difference in growth yield when grown in minimal medium using carbon sources that WT *S. soli* was able to utilise (Figure 4.3.9). This result was unsurprising as upon re-inspection of the *S. soli* genome, it was revealed that the *S. soli* genome also contained a group 1 SDIMO (toluene MO) and a group 2 SDIMO (phenol MO). Based on known substrate preferences of SDIMOs, either of these are more likely than ZmoABCD to be the phenol-oxidising enzyme of *S. soli*. Regarding *n*-octane, another better candidate for octane metabolism (*alkB*) was also seen in the *S. soli* genome, suggesting that octane was also unlikely to be the physiological substrate for ZmoABCD. The genomic contexts of the other MO types detected in the *S. soli* genome (Figure 4.3.10) provide further support for their roles in phenol and octane

oxidation, for example a catechol-2,3-dioxygenase is found immediately adjacent to the group 2 SDIMO, and an alcohol dehydrogenase is co-located with the *alkB* gene. This result suggests that ZmoABCD is not the sole locus responsible for the oxidation of these hydrocarbon substrates. Additionally, it is unlikely that the ZmoABCD enzyme was involved in the breakdown of either branched chain amino acids isoleucine and leucine, as *S. soli* also harbours a branched-chain amino acid transaminase homologue (Accession number: WP_028080876.1, locus NZ_AXDW01000036) that facilitates the metabolism of these branched chain amino acid substrates (Hutson, 2001).

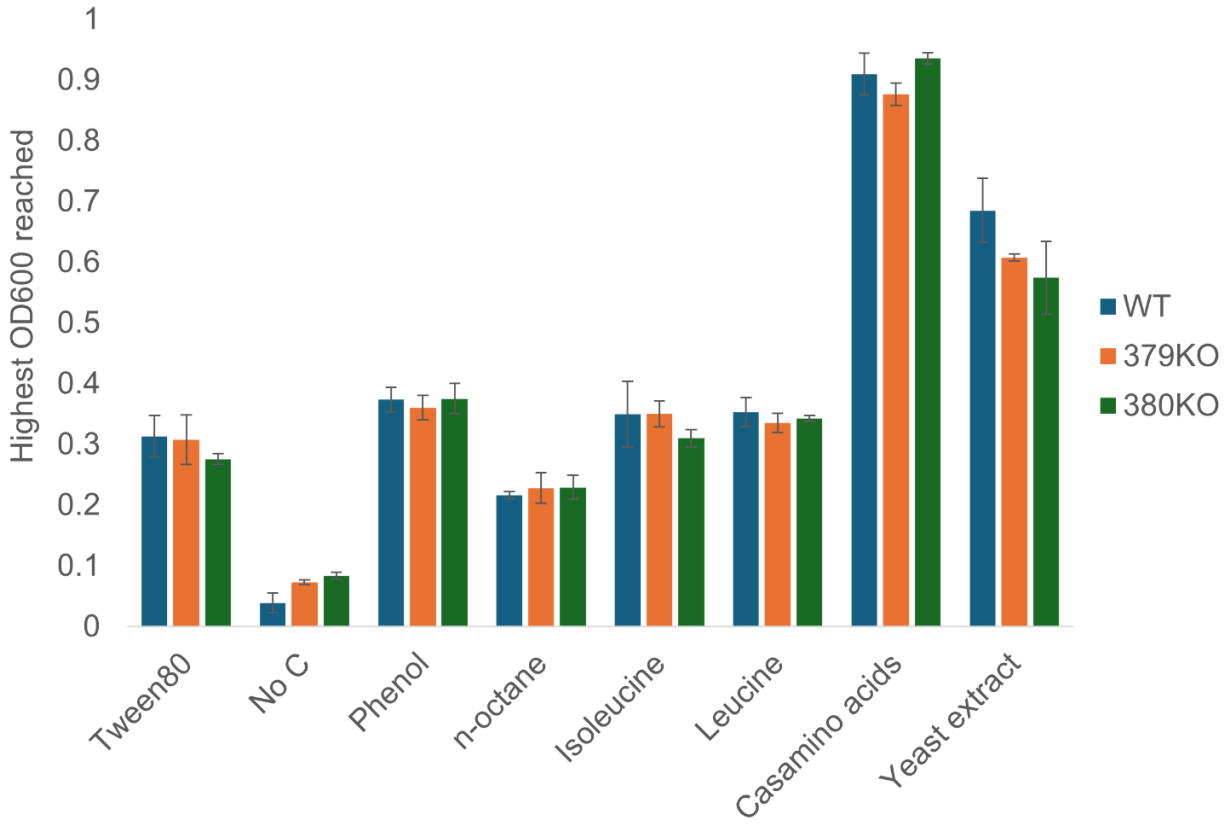


Figure 4.3.9. Growth of *S. soli* strains on several carbon sources.

Growth yields were determined as OD_{600} after incubation at varying periods of time. The reported OD_{600} values were the highest OD_{600} reached by the cultures. Phenol and *n*-octane were added as stock solutions in DMSO at 5 mM final concentration, while the amino acids were added to 3 mM, casamino acids and yeast extracts were added to 0.4% (w/v), and Tween 80 was added to 0.1% (v/v). The No C control had no carbon source added. Data are the means of three replicates, with error bars indicating the standard deviation.

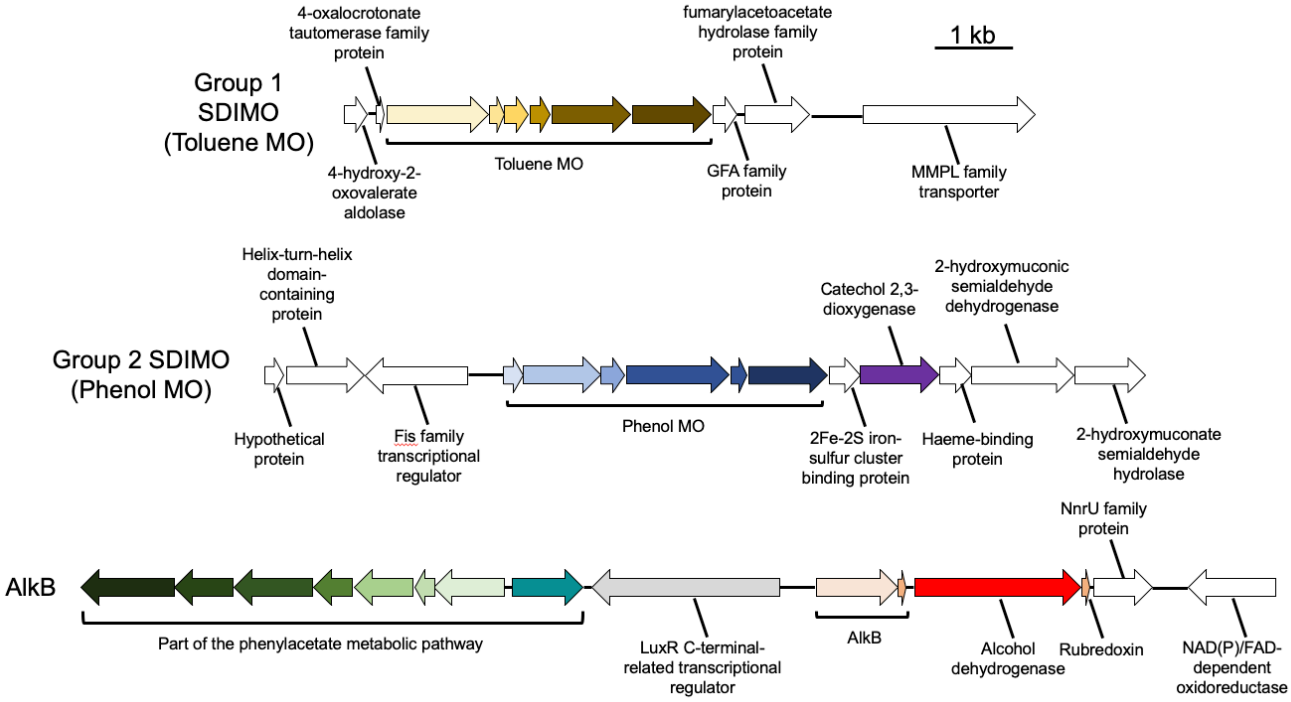


Figure 4.3.10. Genomic context of other monooxygenases in the *S. soli* genome.

The putative toluene MO is found at locus NZ_AXDW01000015 region: 708-2210, the putative phenol MO and AlkB are found at locus NZ_KI440840 region: complement (38592-39767).

4.4. Discussion

In this chapter, a modified R2A medium, R2A+, was developed to increase the growth yield of *S. soli* in liquid cultures. In addition, a reliable protocol to introduce DNA to *S. soli* via conjugation was established. The carbon usage range of *S. soli* DCY12 was explored, and Δ*zmoABCD* knockout strains of *S. soli* were also generated. Furthermore, these knockout strains were utilised to determine ZmoABCD involvement in hydrocarbon degradation.

The development of R2A+, an enhanced R2A medium with increased yeast extract and the addition of MOPS buffer was essential to further experimentation with *S. soli*. The optical density of saturated *S. soli* cultures in unmodified R2A was low and the pH was high. An increased pH in the medium is often the result of amino acid utilisation and is caused by the release of ammonia from those amino acids

(Richardson *et al.*, 2013). In the case of sugar utilisation the pH of the medium decreases (Iwami and Yamada, 1980), strongly suggesting that *S. soli* was utilising the amino acids present in the medium as opposed to the glucose and other sugars. This was supported by the fact that *S. soli* does not utilise glucose as a sole carbon source. Interestingly though, *S. soli* was unable to grow in a 96-well microtitre plate when MSM supplemented with either 5 mM phenol or 0.1% (v/v) Tween 80 were provided but could grow when R2A+ was provided. This could have been due to the limited oxygen levels and the different enzymes responsible (and their oxygen consumption requirements) for growth with different carbon sources.

The development of a reliable protocol to introduce DNA to *S. soli* was not only vital to the generation of the knockout strains but also for future work requiring genetic manipulation of this organism. Electroporation is a very harsh process and bacterial cells subjected to an electric field result in lowered cell survival rates (Dower *et al.*, 1988). *S. soli* was subjected to several trials of electroporation to no avail, likely because the most appropriate electroporation conditions were not applied, or the cells were inherently resistant to electroporation. In hindsight, the concentration of kanamycin was too high when trialling electroporation using the vectors pUS250 and pBBR1MCS-2, but electroporation trials using pBBR1MCS-5 did demonstrate that this process may have been too harsh or not the conditions were suboptimal for *S. soli*. Conjugation on the other hand is much gentler, and since *S. soli* was amenable to conjugation, this process was chosen as the transformation method. In this case, the generation of a rifampicin resistant mutant strain of *S. soli* was required for conjugation, as rifampicin was chosen as the agent to eliminate donor cells in the medium selecting for transconjugants. Most commonly, rifampicin resistance arises from point mutations in the *rpoB* gene, the gene that encodes for the β subunit of the bacterial RNA polymerase (Goldstein, 2014). While no obvious phenotypic changes were observed in Rif^R *S. soli* mutants when grown in complex medium, this does not preclude potential subtle deleterious alterations as the Rif^R mutants were not subjected to whole genome sequencing. It has been previously suggested that mutations in the binding pocket of the RNA polymerase alters how the RNA polymerase interacts with promoters and transcriptional regulators (Maughan *et al.*, 2004), but this risk was accepted in this study as electroporation was not a feasible method to transform *S. soli*. Alternatively, diaminopimelic acid

auxotrophic *E. coli* donor cells could have been chosen to facilitate the conjugation process (Allard *et al.*, 2015). This would eliminate the need to generate mutant strains of the *S. soli* recipient, which would result in a “cleaner” conjugation process. While *E. coli* S17-1 cells are proline auxotrophic, the use of diaminopimelic acid auxotrophic *E. coli* would allow for the selection of *S. soli* transconjugants to be performed using complex medium (in this case R2A agar) as opposed to culturing the recipients using a proline deficient defined medium that may have a negative effect on the conjugation process (i.e. hinder growth).

Several hurdles were encountered in generating the *S. soli* Δ*zmoABCD* knockout mutants. Both the tetracycline resistance gene and the *sacB* gene in the initially chosen suicide vector pEX18Tc were nonfunctional and/or toxic in *S. soli*. Although eventually knockout strains were successfully generated, the lack of a counter-selectable marker massively slowed progress. As there are a selection of other counter-selectable markers for allele exchange purposes (Reyrat *et al.*, 1998), further experiments requiring allele exchange in *S. soli* may be able to make use of them, provided they are non-toxic to this environmental organism. The development of an allele exchange plasmid for *S. soli* would prove useful for the generation future knockout mutants. Difficulties in transforming *S. soli* and in the generation of knockout strains using conventional counter-selectable markers are testaments to the difficulty of utilising environmental organisms in the laboratory. However, since these organisms often harbour interesting enzymes that perform interesting reactions, the effort to search for usable genetic tools in these environmental organisms should be made.

Solimonas soli was unable to grow on C_2-C_8 alkenes that were oxidised by ZmoABCD (Figure 3.3.7), indicating that these activities were co-metabolic rather than growth linked. The bacterium also did not grow on any of the gaseous alkanes which are typical substrates for related four-component SDIMOs (groups 4, 5, and 6). The two hydrocarbons which supported growth of *S. soli* and are also typical MO substrates were phenol and *n*-octane. However, knockout experiments demonstrated that ZmoABCD was not the sole enzyme responsible for growth on phenol, *n*-octane, isoleucine or leucine. While these have yet to be tested for activity with ZmoABCD in recombinant hosts, they are not believed to be the canonical substrates of the

enzyme because there are other MO systems in the *S. soli* genome which are better candidates for these roles (the group 1 and 2 SDIMOs and AlkB). It is conceivable that this bacterium has multiple systems for oxidation of octane and/or phenol, including both the expected enzymes and also ZmoABCD. There is precedent for this kind of redundancy in other MO-containing bacteria where one organism harbours multiple monooxygenases that target the same or similar substrates, e.g. methanotrophs harbouring both soluble and particulate methane monooxygenases that oxidise methane, and *Mycobacterium chubuense* NBB4 harbouring both the ethene and propene monooxygenases which can both oxidise short chain alkenes (Nielsen *et al.*, 1997, Coleman *et al.*, 2011). Terpenes could possibly be the canonical substrate of ZmoABCD (Cho, 2015), as could the prominent greenhouse gas isoprene (Dawson *et al.*, 2023), but the low activity of ZmoABCD against branched-chain alkenes and compounds >C_6 argue against isoprene and terpene as potential substrates. To further explore the involvement of ZmoABCD in the oxidation of phenol and octane, *S. soli* knockout mutants of the other monooxygenases could be generated. In addition, double or triple knockout mutants of ZmoABCD and the other monooxygenases could also be generated to test the necessity of these enzymes for the oxidation of their respective predicted hydrocarbon substrates. Further genomic analyses also confirmed that *S. soli* harbours a homologue to a branched-chain amino acid transaminase, a gene involved in the metabolism of branched-chain amino acids. This strongly suggests that ZmoABCD is an unlikely candidate responsible for growth utilising these substrates.

On the other hand, no alternative enzymes could be found to be directly responsible for growth on Tween 80 as a sole carbon source can were found in the *S. soli* genome. It has been suggested that Tween 80 is first broken down by an esterase (Howe and Ward, 1976), then the oleic acid component undergoes β-oxidation and the products of fatty acid metabolism enter central metabolism (Nguyen, 2018). *S. soli* does harbour multiple esterases and it could be that the bacterium follows the above pathway to breakdown Tween 80 and ZmoABCD may play a role in fatty acid degradation. While oleic acid was not tested as a sole carbon source for *S. soli*, it is likely that the organism could utilise the long chain fatty acid in minimal medium since it was able to utilise *n*-octane as a sole carbon source. In the canonical long chain alkane degradation pathway, the alkane is oxidised to an alcohol which then

enters the β-oxidation pathway (Guo *et al.*, 2023). However, the knockout strains exhibited no difference in growth when Tween 80 was provided as a sole carbon source, perhaps suggesting a redundancy in *S. soli* for fatty acid metabolism that is closely involved in central metabolism. Interestingly, a metabolomics approach demonstrated that the carbon flux in Tween 80-exposed *Mycobacterium tuberculosis* was similar to the carbon flux observed if the bacterium were cultured in a lipid rich environment (Pietersen *et al.*, 2020). The authors also suggested that the oleic acid may have contributed to the induction of a stress response which caused the organism to exhibit these metabolic changes. *Solimonas soli* cultured using Tween 80 as a sole carbon source could be subjected to proteomic and metabolomic studies to examine if the same observations can be seen. These 'omics approaches could also determine if there are any differences in protein expression and metabolite production in *S. soli* Tween 80 degradation versus other organisms, which may play a role in identifying the canonical substrate of ZmoABCD.

One important next step is to develop a high-throughput method to screen for liquid substrates as growth substrate, but this would require the optimisation of *S. soli* growth in 96-well microtitre plates using minimal medium. As this was not explored in this study, the next chapter will focus on the use of the knockout strain (Δ*zmoABCD*::*fugfp*) as a bioreporter. These experiments allow the use of the native host machinery to discover inducers of *zmoABCD* expression while removing the reliance on growth yield as an indicator.

Chapter 5 – The Regulation of *zmoABCD* Expression in *Solimonas soli*

5.1. Introduction

The previous chapters attempted to identify the canonical substrate of the ZmoABCD enzyme through heterologous expression and knockout experiments. While it was shown that the monooxygenase is active against alkenes in *Pseudomonas putida* KT2440 (Chapter 3), the native host *Solimonas soli* was unable to utilise alkenes as growth substrates (Chapter 4). However, knockout experiments eliminated the assumption that ZmoABCD is the sole enzyme responsible for the oxidation of phenol, *n*-octane, isoleucine, leucine, and Tween 80. This chapter will focus on the use of a biosensor to search for inducers that trigger the expression of the monooxygenase utilising GFP as a reporter signal.

A biosensor is a system that utilises biological components to generate a signal in response to the presence of absence of a molecule (Bhalla *et al.*, 2016). Hydrocarbon biosensors are useful in both bioremediation and industrial applications. They have been previously developed to detect dissolved alkanes in contaminated groundwater (Sticher *et al.*, 1997) and marine environments (Sevilla *et al.*, 2015), and also to detect ethylene released from bananas (Moratti *et al.*, 2024). Biosensors can be introduced to chromosomes through knock-in processes (Zhang *et al.*, 2012) or can be carried on plasmids (Sevilla *et al.*, 2015, Moratti *et al.*, 2024), meaning that the host harbouring the biosensor does not necessarily have to be the organism that the biosensor components originated from. Ultimately, the decision between chromosomal biosensors (single-copy) or plasmid-bound biosensors (multiple copies) will depend on the desired signal strength (Carpenter *et al.*, 2018) and the application of the biosensor (e.g. *in situ* detection of hydrocarbons in harsh environments vs. use of industry-friendly hosts for research purposes). Generally, the expression of monooxygenases is inducible as they are often metabolic burdens due to the oxidative stress generated during the enzymatic reactions they catalyse, but the modes of regulation differ among them (Moratti *et al.*, 2022). Therefore, prior to constructing a sensible biosensor, the operator region and the regulatory protein that interacts with said operator region must be identified.

Since neither operator region nor regulatory protein have been characterised for the *zmoABCD* cluster of *S. soli*, a biosensor construct was devised to exploit the native host gene expression machinery. More specifically, this construct is designed to be a bioreporter, a type of biosensor that relies on a reporter gene to produce a quantifiable signal in response to an inducer (Xu *et al.*, 2013). In the genome of *S. soli*, a 100 bp non-coding region is present between the promoter of the *zmoABCD* cluster and the ribosome binding site of *zmoA*, and its function is unknown (Figure 5.1.1). In addition, the 100 bp region is predicted by RNAFold, an RNA structure predictor, to form a hairpin in order to minimise its free energy (Figure 5.1.1). This suggests that this region may play a part in the regulating the translation of the *zmoABCD* cluster, since RNA hairpins are known to be involved in translational regulation in bacteria (Henkin and Yanofsky, 2002). In the proposed bioreporter, the *fuGFP* gene was placed downstream of both the putative promoter that drives the expression of the *zmoABCD* cluster and the 100 bp untranslated region (Figure 5.1.2). This construct is hypothesised to produce a fluorescent signal when an inducer of *zmoABCD* expression is present. The construct was also designed to replace the *zmoABCD* cluster entirely with an *fuGFP* reporter gene, with the remaining genomic context unaltered (i.e. the reporter gene is placed under the control of the native regulatory systems that normally govern the expression of the *zmoABCD* cluster). This construct assumes that the ZmoABCD enzyme and the products of the reaction it catalyses are not directly involved in feedback regulation of this promoter. As briefly mentioned in Chapter 4, two knockout/bioreporter constructs were devised. The difference between the two constructs, pUS379 and pUS380, is the presence or absence of the 100 bp putative regulatory region preceding the ribosome binding site of the *fuGFP* gene (Table 2.4.1, and Figures 4.2.2 and 4.2.3). These were designed to determine if the 100 bp region plays a part in the regulation of *fuGFP* expression, and by proxy the expression of the *zmoABCD* cluster.

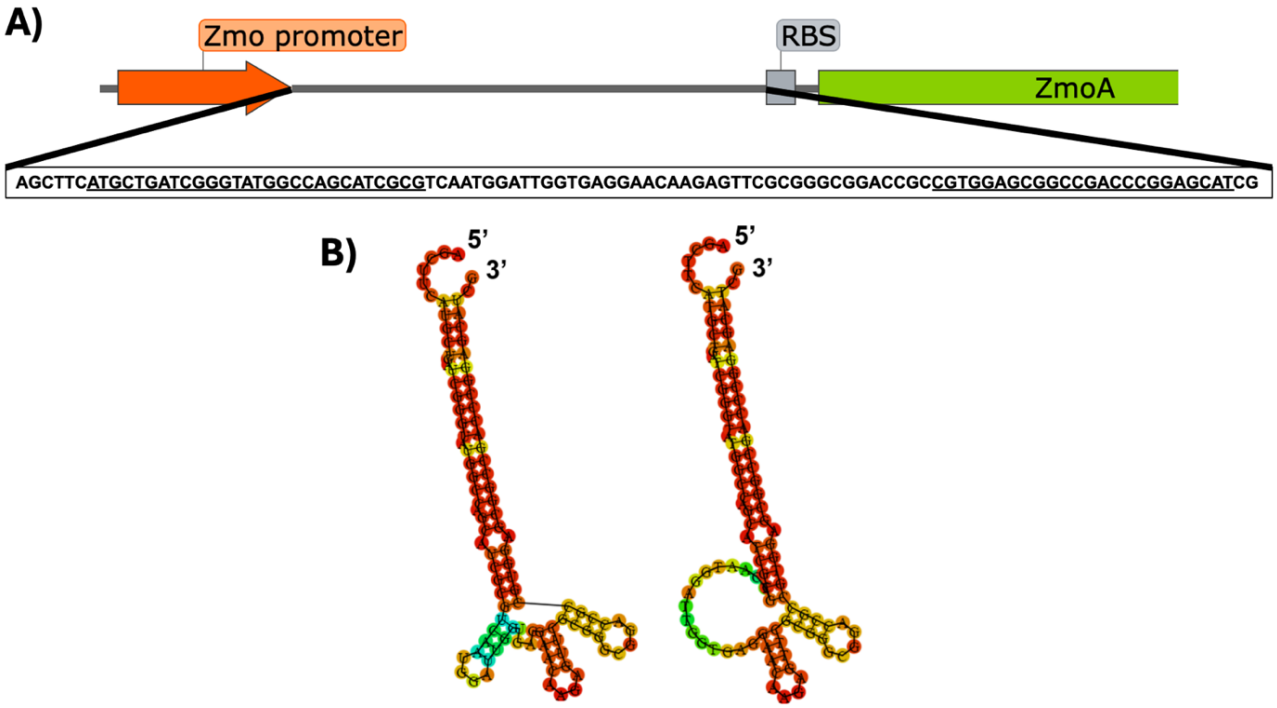


Figure 5.1.1. The 100 bp non-coding region present between the promoter of *zmoABCD* and the ribosome binding site of *zmoA*.

A) The orange arrow represents the *zmo* promoter and the green segment represents the 5' end of the *zmoA* gene. The ribosome binding site (RBS) of *zmoA* is shown as a grey block. The sequence of the 100 bp non-coding region is shown in the text box, and the underlined bps indicate the complementary regions that form the hairpin structure shown in B. B) The hairpin structure of the 100 bp region between the promoter of *zmoABCD* and the ribosome binding site of *zmoA*. The structures shown are the Minimum free energy (left) and centroid (right) secondary structures generated by RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi).

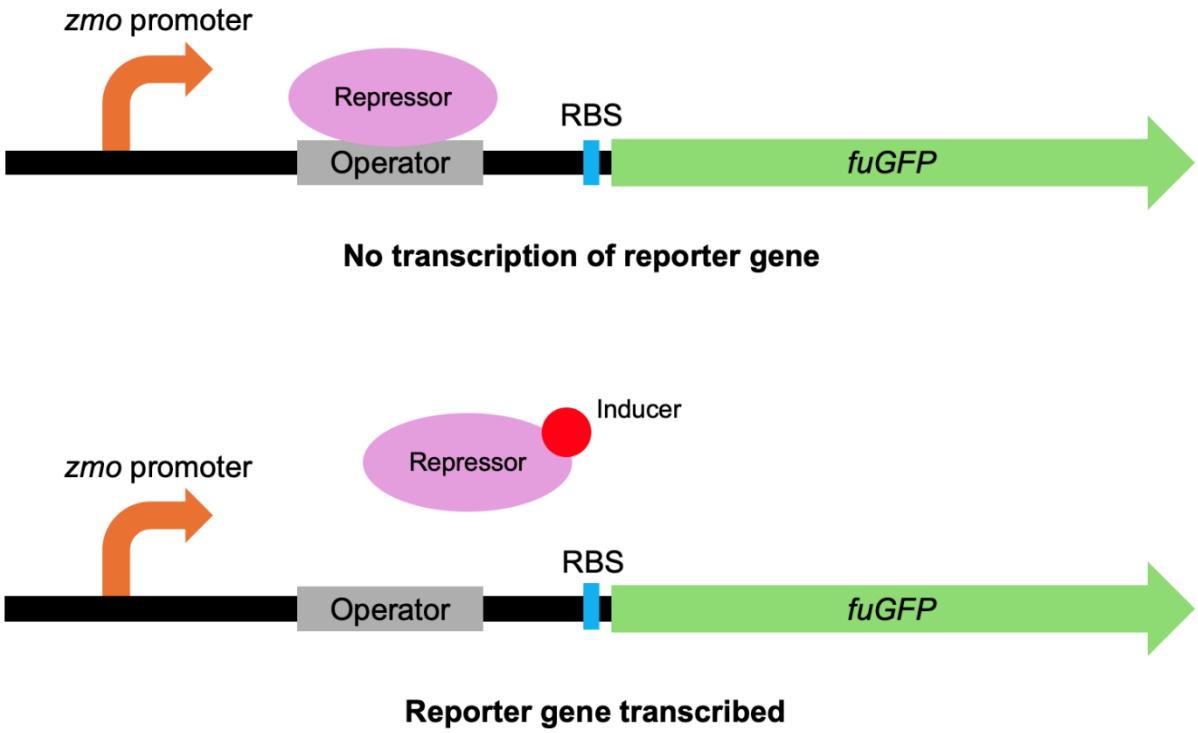


Figure 5.1.2. Graphical representation of the bioreporter construct.

The *fuGFP* reporter gene is placed under the control of the native *S. soli* regulatory systems. The promoter that drives transcription of the *zmoABCD* cluster (*zmo* promoter) is used here, as is the 100 bp region between *zmo* promoter and the ribosome binding site (RBS) of *zmoA*, where a putative operator region is located. The expression of *zmoABCD* is hypothesised to be regulated by a repressor (Pink oval), and therefore it is hypothesised that a repressor protein will bind to the operator region to block transcription of the *fuGFP* reporter gene in the absence of an inducer molecule. Subsequently, when an inducer is present, it will interact with the repressor, release the repressor from the operator region, and allow transcription of the *fuGFP* reporter gene.

The main aim of this chapter is to utilise the knockout strain *S. soli* 379KO as a chromosomal bioreporter to identify inducers of *zmoABCD* expression using fuGFP as a fluorescent signal. Additionally, the 100 bp putative regulatory region will be examined in greater detail in both heterologous and native hosts to identify the putative operator region of the *zmoABCD* gene cluster.

5.2. Methods and Materials

5.2.1. Construction of pUS384-388

Primers binding to pUS379 (Chapter 4, and Table 2.3.1) were designed to generate truncated variants of the putative regulator region (Primer pairs NIY138 and NIY111 for pUS384, NIY139 and NIY111 for pUS385, NIY140 and NIY140 for pUS386, and NIY141 and NIY10 for pUS387) by amplifying the pUS379 backbone using outwardly directed primers (Figure 5.2.1). These primers were used to amplify the pUS379 backbone and generate plasmids pUS384 to pUS387. The purified PCR products (20 ng) were phosphorylated using T4 polynucleotide kinase (New England Biolabs) and were circularised with DNA ligase. *E. coli* CC118λ*pir* was then chemically transformed as described in Section 2.5.2 with the ligated PCR products. The correctly constructed plasmids were verified by junction PCRs (Table 2.3.2) and were then sequenced to confirm their identities. pUS388 was generated by amplification of pUS386 using primers NIY138 and NIY111, and the processes for plasmid construction and verification were repeated as above. For reference, a summary of the differences in the putative regulatory regions between plasmids pUS379, pUS380, and pUS384-388 is provided in Figure 5.2.2. For more details about pUS379, pUS380, and pUS384-388 see Table 2.4.1.

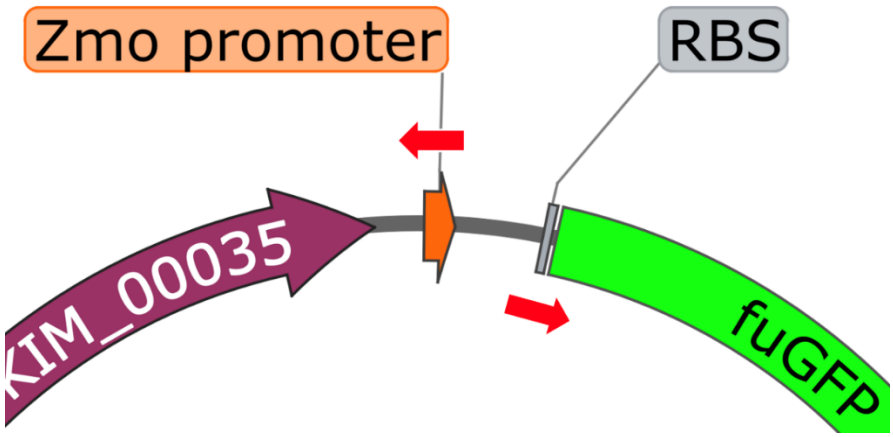


Figure 5.2.1. Graphical representation of the primers and the general amplification scheme used to generate plasmids pUS380, pUS384-388 from pUS379.

The red arrows represent the outwardly directed primers used to amplify pUS379. The resulting PCR products were circularised using DNA ligase to form plasmids pUS380, pUS384-388.

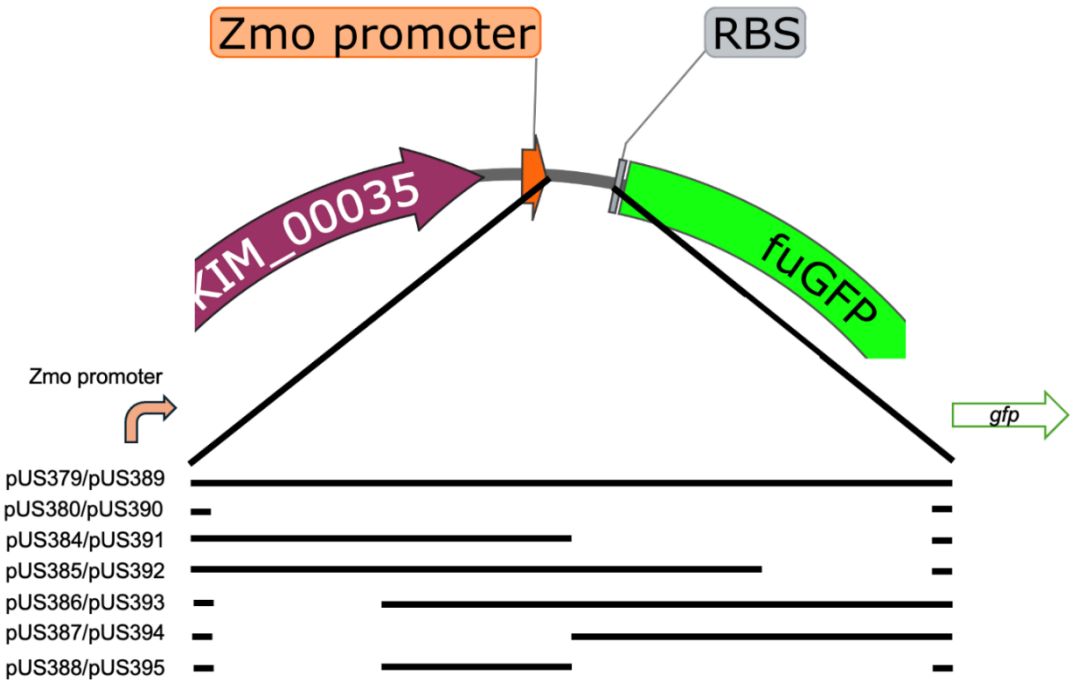


Figure 5.2.2. Graphical summary of differences in the putative regulatory regions between plasmids pUS379, pUS380, pUS384-388 and pUS389-395.

pUS379, pUS380, pUS384-388 were plasmids derived from pKNOCK-Km and were used for experiments in *E. coli* CC118λ*pir*, while pUS389-395 were plasmids derived from pBBR1MCS-2 and were used for experiments in *S. soli* DCY12R and *E. coli* S17-1. In both sets of plasmids, black bars represent the sequence present in each plasmid between the native *zmoABCD* promoter and the ribosome binding site of the *fuGFP* reporter gene, where plasmids pUS379 and pUS389 contain the entire 100 bp region.

5.2.2. Construction of pUS389-395

Since pUS379, pUS380, pUS384-388 are derived from pKNOCK-Km and are therefore non-replicative plasmids in *S. soli*, a set of replicative plasmids with the same putative regulatory regions from pUS379, pUS380, and pUS384-388 were devised. Primers NIY144 and NIY145 were designed to amplify the different *zmo* promoter and *fuGFP* regions harboured by pUS379, pUS380, and pUS384-388 (Table 2.3.1). To remove the constitutive promoter upstream of the multiple cloning site of pBBR1MCS-2, primers NIY42 and NIY143 were designed to amplify the pBBR1MCS-2 backbone (the PCR product was named pBBR1MCS-2_v2) (Table 2.3.2). The primers used contain compatible restriction sites to enable cloning of the biosensor constructs into pBBR1MCS-2_v2 (Table 2.3.1). Each of the *zmo* promoter

and *fuGFP* regions from pUS379, pUS380, and pUS384-388 was cloned into pBBR1MCS-2_v2 using EcoRI-HF and HindIII-HF. *E. coli* TOP10 cells were then chemically transformed as described in Section 2.5.2 with each ligation mix. The correctly constructed plasmids were verified by junction PCRs (Table 2.3.2) and were then sequenced. The resulting plasmids were named pUS389-395. A summary of the differences in the putative regulatory regions between plasmids pUS389-395 is provided in Figure 5.2.2. The plasmid map of pUS389 is provided in Figure 5.2.3. For more details about pUS389-395 see Table 2.4.1. These plasmids were also introduced to *S. soli* DCY12R through conjugation as described in Section 4.2.8 using *E. coli* S17-1 as the donor.

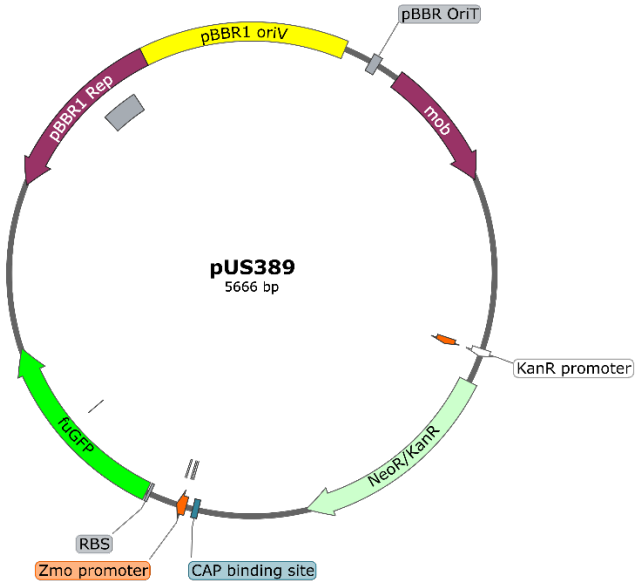


Figure 5.2.3. Map of pUS389.

pUS389 is a 5.7 kb Km^R plasmid derived from pBBR1MCS-2. pUS389 is the biosensor plasmid constructed to probe for the inducers of *zmoABCD* using fuGFP fluorescence as a proxy. For more detail refer to Table 2.4.1.

5.2.3. Fluorescence measurements in the plate reader

Escherichia coli or *S. soli* cells harbouring plasmids containing the *fuGFP* gene were grown to saturation in LB or R2A+ liquid culture respectively, as described in Section 2.2.1. The OD_{600} of each culture was normalised to 1, and the cells were washed twice with 20 mM K_2HPO_4 (pH 7). Then, the cell suspensions were transferred to a black 96-well microtitre plate. The CLARIOstar Plus (BMG Labtech) plate reader was

used to measure the fluorescence intensity of each well at 515-520 nm using an excitation wavelength of 420-470 nm.

5.2.4. Induction of the *Solimonas soli zmoABCD* cluster

Solimonas soli cells harbouring pUS389 were grown in 50 mL R2A+ and were incubated at 30 °C with 200 rpm shaking for 2 days. This pre-culture was used to inoculate 500 mL MSM-Km20-0.4% (w/v) casamino acids to an OD_{600} of 0.05. The MSM culture was grown at 30 °C with 200 rpm shaking until the culture reached mid-exponential phase. The MSM culture was then divided into 5 mL aliquots in 27 mL crimp seal vials. The vials were crimp sealed after the addition of liquid substrates and crimp sealed before the addition of gaseous substrates. Liquid substrates were added to 1 mM (except for Tween 80, which was added to 0.025% (v/v)) and gaseous substrates were added to 10% (v/v) headspace. Liquid substrates were made up as 250 mM stock solutions in either water, ethanol, or DMSO, depending on the substrate. Non-induced and DMSO only controls were included. These cultures were incubated at 30 °C with 200 rpm shaking. At 24 and 96 hours of incubation, 200 µL samples from each vial were extracted at each time point. The OD_{600} and fuGFP fluorescence intensity readings of each vial were measured using black 96-well microtitre plates and the CLARIOstar Plus (BMG Labtech) plate reader as described above.

5.3. Results

5.3.1. Background expression of the reporter gene was observed in *E. coli* S17-1λ*pir* donor colonies

To introduce plasmids to *S. soli* through conjugation, rifampicin was chosen as the agent to select against *E. coli* donor cells (Section 4.2.8). However, due to the relative ease of gaining a point mutation that confers rifampicin resistance (Goldstein, 2014), rifampicin resistant spontaneous mutant *E. coli* donor colonies were commonly seen alongside *S. soli* transconjugants on the rifampicin-containing agar plates during the conjugation process. During the conjugation process in the generation of the *S. soli* knockout strains 379KO and 380KO, the rifampicin resistant mutant *E. coli* S17-1λ*pir* donor colonies carrying pUS379 and pUS380 were found to

be fluorescent on spread plates of each conjugation mix (Figure 5.3.1). Moreover, by the naked eye *E. coli* colonies harbouring pUS380 were more fluorescent than their pUS379-harbouring counterparts (Figure 5.3.1). To quantify the difference in fluorescence output of the two knockout constructs pUS379 and pUS380, *E. coli* CC118λ*pir* cells were transformed with these plasmids. The fluorescence intensity of each *E. coli* strain was determined (Figure 5.3.2). The significantly higher fluorescence shown by *E. coli* cells harbouring pUS380 strongly suggests that the 100 bp putative regulatory region plays a part in regulating expression of the *fuGFP* gene, and that the *zmo* promoter is active in *E. coli*. As a result, further investigation into this putative regulatory region to identify the operator region was undertaken.

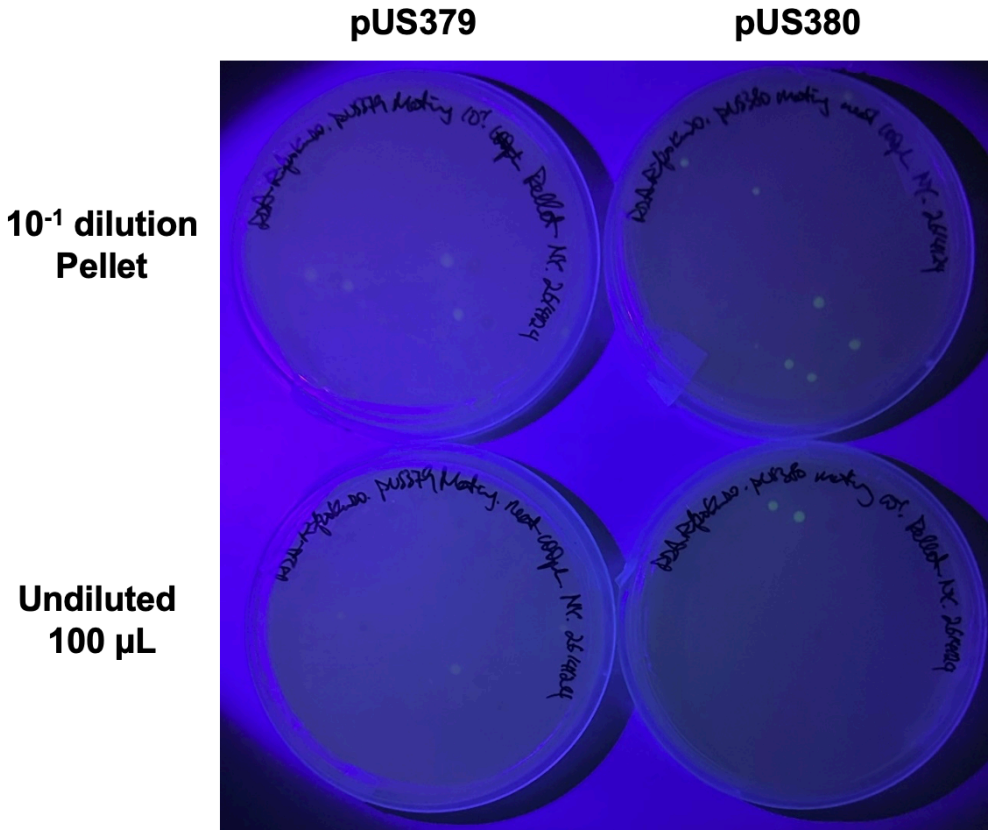


Figure 5.3.1. Spread plates of conjugation mixes on R2A-Rif20-Km20 agar under UV light at 395 nm.

The pellet of the 10^{-1} dilution of the conjugation mixes and 100 µL of the undiluted conjugation mixes were spread onto R2A-Rif20-Km20 agar. The large, fluorescent colonies are spontaneous rifampicin resistant *E. coli* S17-1λ*pir* mutants, while *S. soli* colonies are not visible nor fluorescent in the figure. Note the difference in fluorescence intensity between pUS379 and pUS380 harbouring *E. coli* colonies.

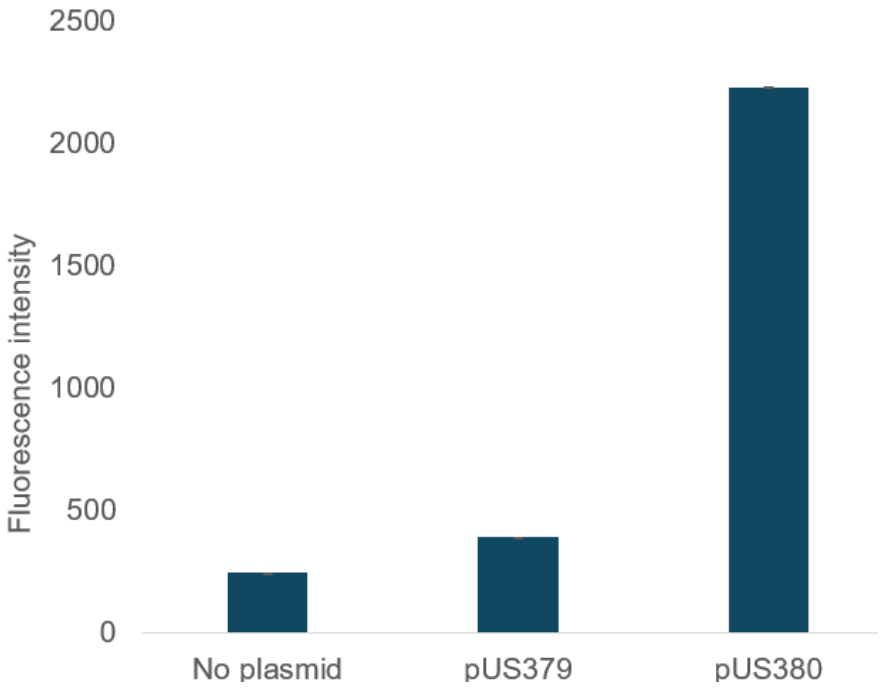


Figure 5.3.2. The fluorescence intensity of fuGFP in *E. coli* CC118λ*pir* when harbouring no plasmid, pUS379, or pUS380.

5.3.2. The *Solimonas soli* Δ*zmoABCD::fuGFP* strains 379KO and 380KO are not fluorescent

The *S. soli* knockout strains generated in Chapter 4 showed no difference in fluorescence intensity versus wildtype *S. soli* when grown in liquid R2A+ (Figure 5.3.3). The expectation was that *S. soli* 380KO would be somewhat fluorescent due to the removal of the putative operator region, since the hypothesised repressor would not be able to bind to the operator region to stop transcription of *fuGFP* (Figure 5.1.2). Several experiments were devised to determine whether the fuGFP protein was non-functional in *S. soli* or if the *fuGFP* gene, present in a single copy, was not strongly enough expressed to generate a fluorescent signal in *S. soli* (Section 5.3.4).

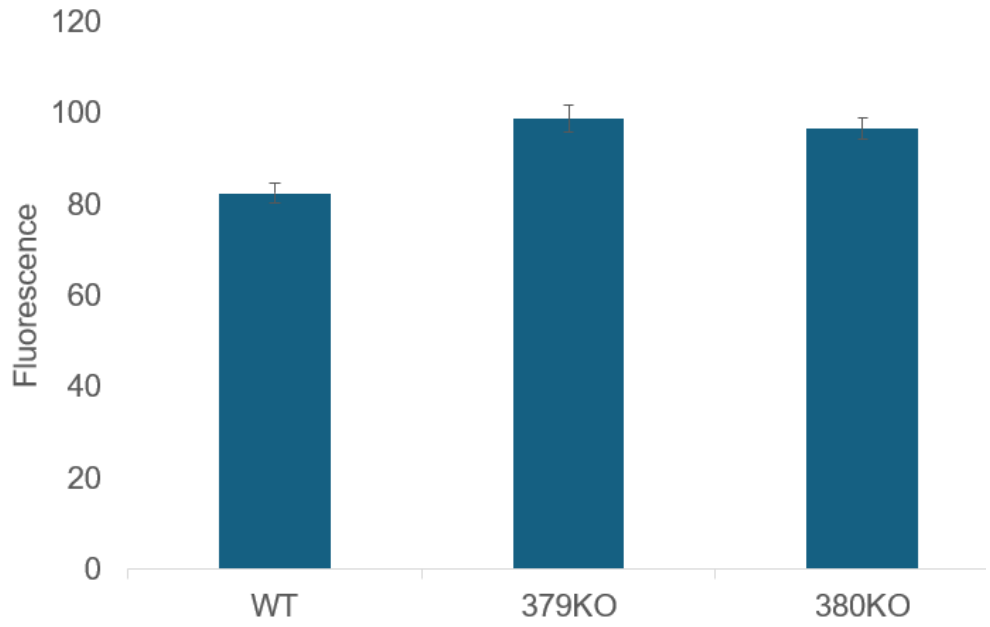


Figure 5.3.3. The fluorescence intensity of different *S. soli* strains.

WT = *S. soli* DCY12.

5.3.3. The putative operator region between the *zmo* promoter and the *fuGFP* gene

To characterise the region that may contain the operator region, truncations of the putative regulatory region were generated by PCR amplification of the pUS379 plasmid (Section 5.2.2). Removal of the first 50 bp of the putative regulatory region increased fluorescence of *E. coli* cells harbouring the pUS387 plasmid construct (pUS387, Figure 5.3.4). The fluorescence intensity of cells harbouring pUS388 returned to levels as if the entire regulatory region was present, suggesting that the regulation is controlled by the 25 bp region present in pUS388 (Figure 5.3.4). When these regions were cloned into a modified pBBR1MCS-2 backbone (pUS389-395) and introduced to *E. coli* S17-1 cells, the same pattern was observed (Figure 5.3.5).

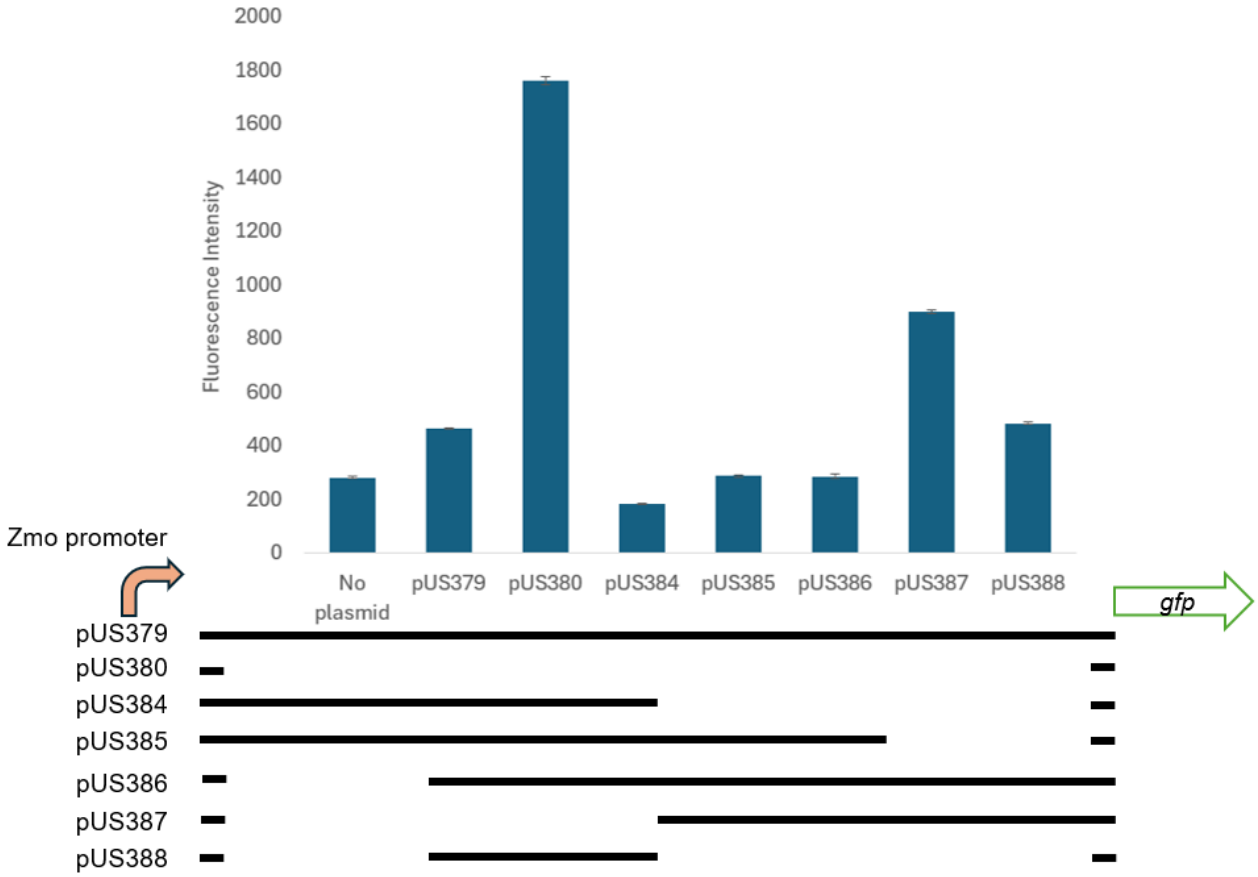


Figure 5.3.4. The fluorescence intensity of *E. coli* CC118λ*pir* when harbouring no plasmid, pUS379, pUS380, or pUS384-388.

The graphical summary as in Figure 5.2.2 is also provided. The black bars represent the sequence present in each plasmid between the native *zmoABCD* promoter and the ribosome binding site of the *fuGFP* reporter gene, where pUS379 contains the entire 100 bp region.

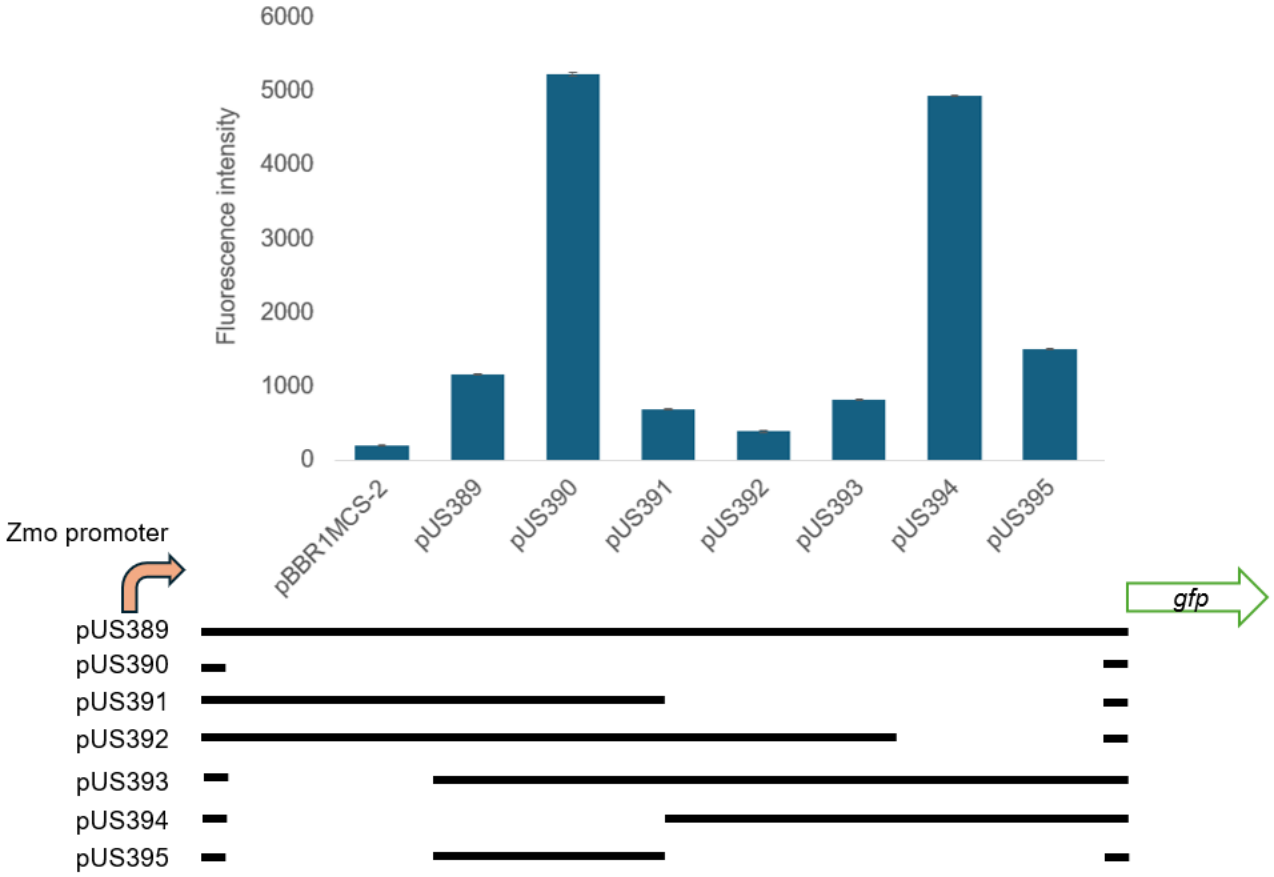


Figure 5.3.5. The fluorescence intensity of *E. coli* S17-1 when harbouring pBBR1MCS-2 or pUS389-395.

The graphical summary as in Figure 5.2.2 is also provided. The black bars represent the sequence present in each plasmid between the native *zmoABCD* promoter and the ribosome binding site of the *fuGFP* reporter gene, where pUS389 contains the entire 100 bp region.

5.3.4. Fluorescence pattern in *S. soli* harbouring pUS389-395 mimics the pattern observed in *E. coli*

pUS389-395 were conjugated to *S. soli* DCY12R and the fluorescence intensity of transconjugants was measured. Firstly, this experiment confirmed that the *fuGFP* gene can be expressed in *S. soli* and was fluorescent in the soil organism when the gene was present in multiple copies, even though fluorescence was below detection when the gene was only present in a single copy on the chromosome (Figure 5.3.3). While a similar fluorescence pattern as *E. coli* cells were observed, the baseline fluorescence of *S. soli* cells harbouring pUS389 was much higher than for *E. coli* harbouring the same plasmid (Figure 5.3.6). In addition, the loss of the first 25 bp of the putative regulatory region increased fluorescence and the presence of the 25 bp

region in pUS395 was not adequate to suppress fluorescence to pUS389 levels (Figure 5.3.6). This suggests that the operator region of the *zmoABCD* cluster in *S. soli* is present in the first 50 bp of the putative regulatory region.

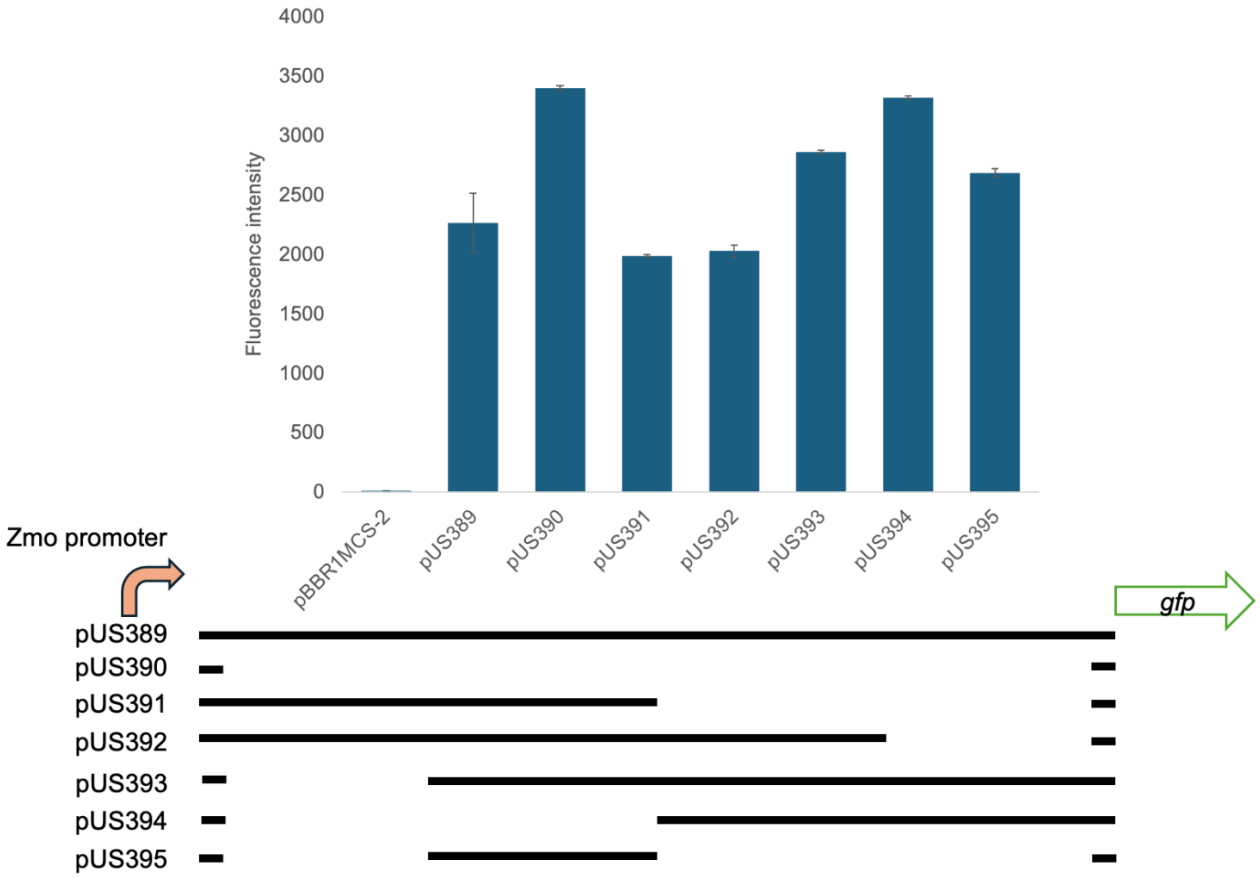


Figure 5.3.6. The fluorescence intensity of *S. soli* DCY12R when harbouring pBBR1MCS-2 or pUS389-395.

The graphical summary as in Figure 5.2.2 is also provided. The black bars represent the sequence present in each plasmid between the native *zmoABCD* promoter and the ribosome binding site of the *fuGFP* reporter gene, where pUS389 contains the entire 100 bp region.

5.3.5. Short-chain hydrocarbons induce expression of the *Solimonas soli zmoABCD* cluster

S. soli(pUS389) cells were exposed to a range of hydrocarbons, and to a range of their expected metabolites for 96 hours. The fluorescence intensity of cells exposed to butane, 1-butene, 2-propanol, *n*-octane, and 1-octene was significantly different compared to the no inducer control (Figure 5.3.7). This suggests that the inducer for the *zmoABCD* genes is a short-chain hydrocarbon. Interestingly, while the enzyme

was active against propene in the *P. putida* KT2440 heterologous host (Figure 3.3.7), propene did not induce the expression of the *zmo* promoter in *S. soli* (Figure 5.3.7). Additionally, 2-propanol, but not 1-propanol, induced the expression of the *zmo* promoter, strongly suggesting that ZmoABCD oxidises short-chain hydrocarbons at the subterminal carbon as opposed to the terminal carbon (Figure 5.3.7). The fluorescence intensity of *S. soli*(pUS389) also decreased when exposed to *n*-octane and 1-octene (Figure 5.3.7), suggesting that the long chain hydrocarbons or their related metabolites contributed to the repression of the *zmo* promoter. The growth yields of *S. soli*(pUS389) were not impacted by the addition of substrates (Figure 5.3.8). Unsurprisingly, Tween 80 promoted the growth of *S. soli* DCY12R(pUS389) (Figure 5.3.8). While the fluorescence intensity of cells exposed to Tween 80 was higher than the no inducer control, butane remains the best inducer of the biosensor when both the growth yield and fluorescence intensity are considered in conjunction, i.e. the fluorescence intensity pattern is the same when normalised for cell density.

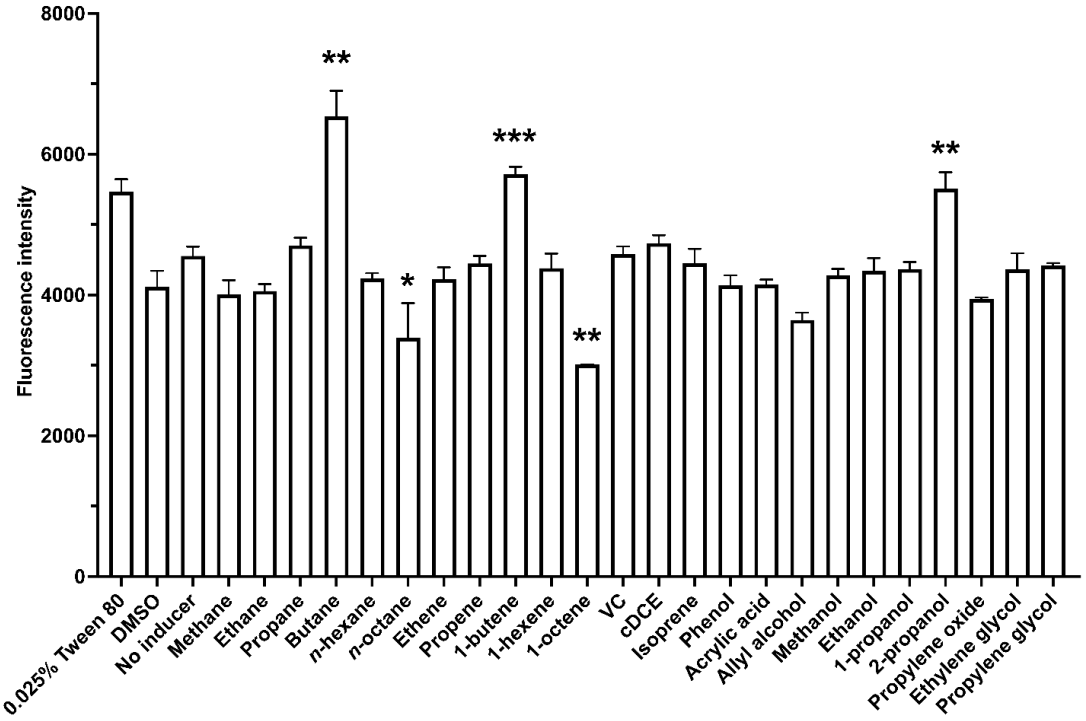


Figure 5.3.7. The fluorescence intensity of *S. soli* DCY12R(pUS389) 96 hours after exposure to hydrocarbons and their metabolites.

The fluorescence intensity of cells was measured at 515-520 nm using an excitation wavelength of 420-470 nm. Liquid substrates were added as stock solutions in DMSO or water at 1 mM final concentration (except Tween 80, which was added to 0.025% (v/v)), while gaseous substrates were added as neat gases to 10% of the headspace volume. DMSO

alone was tested as a solvent control. Data are the means of three replicates, with error bars indicating the standard deviation. Treatments showing significant differences in fluorescence intensity compared to no inducer control are shown as either *** ($p<0.001$), ** ($p<0.01$), or * ($p<0.05$). Statistical analysis was conducted using Welch's t-test. VC: vinyl chloride, cDCE: *cis*-1,2-dichloroethylene.

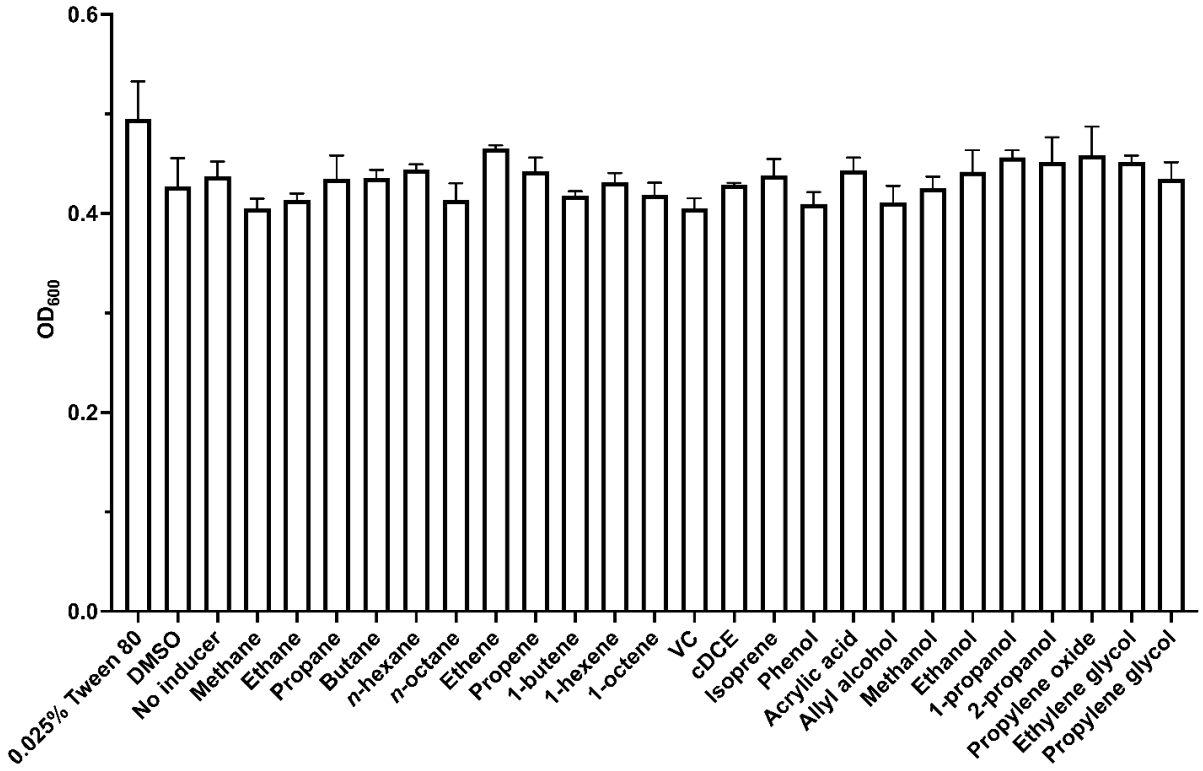


Figure 5.3.8. The growth yield of *S. soli* DCY12R(pUS389) 96 hours after exposure to hydrocarbons and their metabolites.

Growth yields were determined as OD_{600} after incubation with the substrates for 96 hours. Liquid substrates were added as stock solutions in DMSO or water at 1 mM final concentration (except Tween 80, which was added to 0.025% (v/v)), while gaseous substrates were added as neat gases to 10% of the headspace volume. DMSO alone was tested as a solvent control. Data are the means of three replicates, with error bars indicating the standard deviation. VC: vinyl chloride, cDCE: *cis*-1,2-dichloroethylene.

The obvious next step after the above findings was to expose *S. soli*(pUS389) to the four butanol isomers, 1-butanol, 2-butanol, 2-methyl-1-propanol (isobutanol), and *tert*-butanol, since they are the potential metabolites of butane oxidation. Interestingly, none of the butanol isomers induced the expression of the *zmo* promoter (Figure 5.3.9). Exposure to Tween 80 for 24 hours did increase the fluorescence of *S. soli*(pUS389) (Data not shown), but at 96 hours post-exposure

Tween 80 did not have an effect on the expression of the *zmo* promoter (Figure 5.3.9), unlike what was seen in Figure 5.3.7. This suggests that potentially the inducer has been metabolised and resulted a loss in its inductive ability, but that is unusual as GFP protein is remarkably stable so there should be minimal loss in the fluorescent signal, and fuGFP is not known to have different properties to other GFP proteins. Since butane continued to induce the expression of the *zmo* promoter at the 96 hour time point (Figure 5.3.9), consistent with Figure 5.3.7, the Tween 80 result was considered to be an outlier.

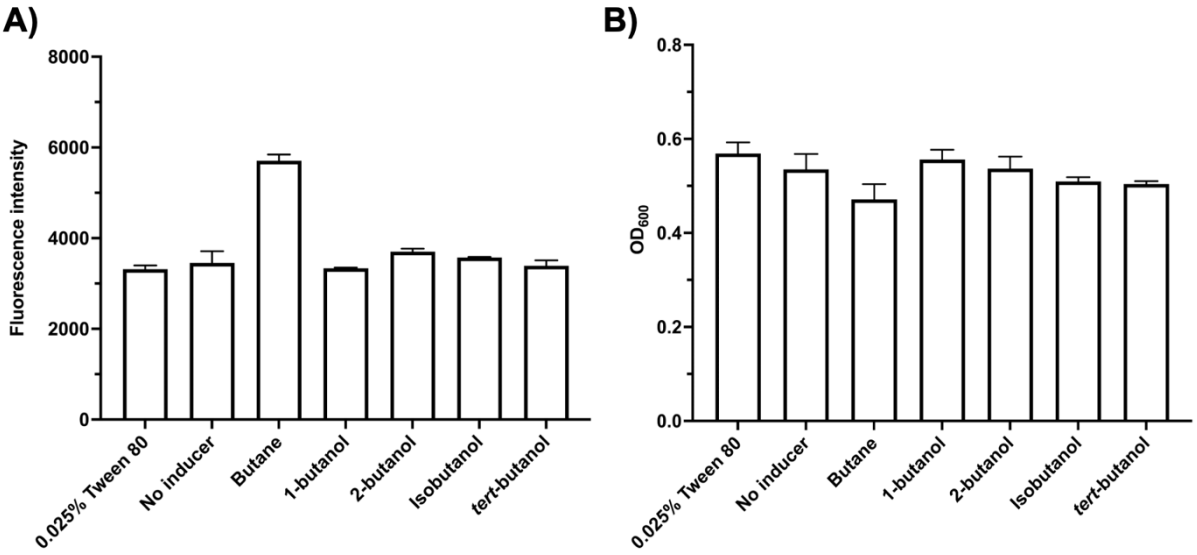


Figure 5.3.9. The fluorescence intensity and growth yield of *S. soli*(pUS389) 96 hours after exposure to butanol isomers.

A) The fluorescence intensity of cells was measured at 515-520 nm using an excitation wavelength of 420-470 nm. Liquid substrates were added as stock solutions in water at 1 mM final concentration (except Tween 80), while gaseous substrates were added as neat gases to 10% of the headspace volume. **B) Growth yields were determined as OD_{600} after incubation with the substrates for 96 hours.** Liquid substrates were added as stock solutions in water at 1 mM final concentration (except Tween 80, which was added to 0.025% (v/v)), while gaseous substrates were added as neat gases to 10% of the headspace volume. For both A) and B) data are the means of three replicates, with error bars indicating the standard deviation.

5.4. Discussion

In this chapter, the regulation of the *zmoABCD* cluster was examined. A putative operator region between the *zmo* promoter and *zmoA* was identified through truncation of the 100 bp putative regulatory region. Inducers of *zmoABCD* expression were also discovered through the use of a plasmid-borne fuGFP bioreporter.

When transformed with the pUS389-395 constructs, the fluorescence pattern in both *E. coli* and *S. soli* were similar (Figures 5.3.5 and 5.3.6). Deletion of the proximal 50 bp of the 100 bp putative regulatory region gave the same increase in the fluorescence signal in both hosts as deletion of the entire 100 bp region. It is highly possible that the hairpin structure that forms in this 100 bp region was not the only factor that played a role in the repression of *fuGFP* expression, as the deletion of the distal 50 bp of the putative regulatory region was inadequate to fully eliminate the expression of the *fuGFP* in both *E. coli* and *S. soli*. This strongly suggests that the proximal 50 bp regions contains a putative operator region that binds a repressor protein in both *E. coli* and *S. soli,* to control the expression of *fuGFP* and hence *zmoABCD* in *S. soli*. Interestingly, since the effect was demonstrated in both *E. coli* and *S. soli*, it is likely that related repressor proteins recognise and interact with this region in these Gammaproteobacterial hosts. However, *E. coli* does not contain genes coding for hydrocarbon-oxidising oxygenase enzymes, so if the regulation is in fact governed by a repressor protein, it would likely be a repressor protein that recognises a related operator region.

The use of the plasmid-borne bioreporter in *S. soli* was not ideal. Multiple copies of the *fuGFP* gene were present in each cell and as a result experiments utilising this bioreporter would not be fully representative of the regulatory system of the single copy, chromosomal *zmoABCD* cluster of wild type *S. soli*. However, since *S. soli* cells harbouring the plasmid-borne bioreporter was able to yield detectable levels of fluorescent protein expression (Figure 5.3.6) and the chromosomal bioreporter was unable to (Figure 5.3.2), experiments utilising the former for the identification of the operator region and the inducer screening continued.

The complete removal of the putative regulatory region in the plasmid-borne bioreporter increased fluorescence in *S. soli* cells (Figure 5.3.6), strongly indicating that a putative operator region has been removed. However, the baseline fluorescence of the *S. soli* cells was much higher than the *E. coli* cells harbouring the same constructs (Figures 5.3.5 and 5.3.6), suggesting that the multi-copy nature of plasmid-borne bioreporters could be overwhelming the natural regulation system in *S. soli* cells. Interestingly, the same was not seen in *E. coli* hosts, perhaps demonstrating that expression of the *zmoABCD* cluster is not tightly regulated in *S. soli* but a related repressor protein more tightly regulates a similar operator region in *E. coli*.

fuGFP expression in *S. soli*(pUS389) cells was induced by butane, propane, 1-butene, and 2-propanol, with the cells exposed to butane displaying the highest fluorescence intensity (Figure 5.3.7). The cells exposed to the alkene metabolites (i.e. propylene oxide, ethylene glycol, and propylene glycol) exhibited the same fluorescence as non-induced cells, but expression of the *zmo* promoter was induced by 2-propanol, an alkane metabolite (Figure 5.3.7). Since SDIMO gene expression is commonly induced by the respective substrates of the SDIMO and also the respective metabolites (Kurth *et al.*, 2008, Moratti *et al.*, 2024), this suggests that ZmoABCD could be a short-chain alkane monooxygenase, in particular a butane monooxygenase that oxidises the subterminal carbon of alkanes. Conversely, 2-butanol was expected to also induce *fuGFP* expression but it was unable to (Figure 5.3.8), suggesting that the butane may not be the canonical substrate of ZmoABCD, further confounding the characterisation of the substrate range of this novel SDIMO. One possibility is that the expression of *zmoABCD* is not induced by its metabolite. This is effect is not unique to the *zmoABCD* cluster though, and is an indicator of the limitations of using inducers of SDIMO gene expression to identify canonical substrates. For example, while *Mycobacterium dioxanotrophicus* PH-06 is able to use both propane and 1-propanol as growth substrates, the expression of its propane monooxygenase (Group 6 SDIMO) is induced by propane, but not 1-propanol (Deng *et al.*, 2018). Another example is the butane monooxygenase (Group 3 SDIMO) from *Thauera butanivorans*. Its butane monooxygenase is induced by and is responsible for the oxidation of C_2-C_9 alkanes (Kurth *et al.*, 2008). Interestingly

though, the monooxygenase is only induced by alcohols with an even, not odd, number of carbons (i.e. the monooxygenase is induced by ethanol, butanol, hexanol, and octanol, but not propanol, pentanol, and heptanol) (Kurth *et al.*, 2008). Another possibility is that the alkanol was simply immediately metabolised by the host cell, resulting in the loss the inductive ability of the alcohol metabolite and hence no fluorescent signal observed. In addition, exposure to either *n*-octane or 1-octene reduced the fluorescence intensity of *S. soli*(pUS389) cells (Figure 5.3.7). This strongly suggests that long chain hydrocarbons repress the expression of the *zmoABCD* cluster. However, it cannot be excluded that it was the metabolites of these long chain hydrocarbons that were responsible for this effect, as the *zmoABCD* cluster was still present in the chromosome of *S. soli*, since the bioreporter was carried by a plasmid-borne system. This argues that there is a much more complex regulatory system at play here. In combination, these results further emphasise the role of ZmoABCD in the metabolism of short chain hydrocarbons or small C_3-C_4 molecules.

On the other hand, Tween 80, a large molecule, also induced the expression of the *zmo* promoter (Figure 5.3.7). A likely scenario was that during Tween 80 metabolism in *S. soli*, the oleic acid component of the detergent was broken down by β-oxidation to C_3-C_4 molecules. These small molecules were then recognised by the repressor protein which then allowed for the expression of the *zmo* promoter. Interestingly though, both *n*-octane and Tween 80 are large molecules with long hydrophobic carbon chains, yet the former represses *zmoABCD* while the latter induces its expression. Again, note that in this study the Δ*zmoABCD* knockout strain of *S. soli* was not used, so the ZmoABCD enzyme would still be expressed if the cells were exposed to the inducer. This may result in the potential oxidation of the inducer and/or substrate to their respective products and could mean the loss of the inducer. It is interesting that Tween 80 also induced expression of the *zmo* promoter, but it is unlikely that ZmoABCD was responsible for the hydrolysis of Tween 80 as demonstrated in Chapter 4, so the enzyme responsible must be elsewhere. More likely, the downstream metabolites of Tween 80 hydrolysis, such as oleic acid, triggered the expression of the *zmo* promoter. The sorbitan component of Tween 80 seems to be an unlikely candidate to trigger the expression of the *zmo* promoter,

since neither the branched chain hydrocarbons nor phenol triggered an increase in fluorescence (Figure 5.3.7). Unfortunately, due to time constraints, tetrahydrofuran, carboxylic acids or fatty acids were unable to be tested as inducers of the bioreporter.

If ZmoABCD is to be considered as a carboxylic acid or fatty acid monooxygenase like the P450 MOs, there are multiple candidate families of repressor proteins that regulate gene expression of enzymes responsible for these metabolic processes, the GntR, TetR/AcrR, and LysR families. The first two families are known to regulate fatty acid degradation as well as alkane degradation (van Beilen *et al.*, 2004, Liang *et al.*, 2017). FadR in *E. coli* belongs to the GntR family of repressors, and in the presence of long chain acyl-CoA the repressor allows for the transcription of genes involved in fatty acid degradation (Cronan, 2021). GntR repressors are also involved in regulating alkane metabolism, especially the expression of AlkB MOs (Pan *et al.*, 2022). Unfortunately, there is no neighbouring GntR repressor homologue near the *zmoABCD* cluster in *S. soli.* On the other hand, downstream of the *zmoABCD* cluster in *S. soli*, there exists a homologue to the TetR family of repressors, but it is 4.5 kb downstream of the stop codon of the *zmoD* gene, a great distance to be its potential regulator (Figure 5.4.1). Nonetheless, across many phyla the TetR family of repressors have diverse functions which include the regulation of stress responses, fatty acid degradation (Cuthbertson and Nodwell, 2013), and AlkB MO expression (Liang *et al.*, 2017). Interestingly, in the proximity of the *tetR r*epressor homologue is also a cluster of genes related to fatty acid degradation. However, due to its distance (6.5 kb upstream) from the CoA transferase homologue, it is also unlikely that this TetR repressor homologue is the regulator for this cluster of genes (Figure 5.4.1). A future experiment could be to utilise a *tetR* homologue deletion mutant of *S. soli* to test the relevance of this repressor protein to the regulation of the expression of either gene cluster. However, it is more likely that a FadR homologue would be responsible for the regulation of the fatty acid metabolism genes and the TetR regulator is responsible for the regulation of other genes. There is also a homologue to the LysR family of repressor proteins downstream of the SDIMO cluster but is even further downstream (6.5 kb) of the *zmoABCD* cluster compared to the *tetR* repressor homologue (Figure 5.4.1). The LysR regulators have notoriously diverse

functions, able to act as activators as well as repressors (Maddocks and Oyston, 2008), and are found in a diverse range of bacteria (Baugh *et al.*, 2023). Since the LysR family of regulators are found in many species of bacteria and can be local or global transcriptional regulators (Hernández-Lucas *et al.*, 2008), it is possible that a conserved LysR homologue in *E. coli* was able to bind to the putative operator region between the *zmo* promoter and the *fuGFP* gene in a similar fashion to the regulation of *zmo::fuGFP* in *S. soli* . This may explain the similar fluorescence pattern exhibited by *E. coli* and *S. soli* cells harbouring the pUS389-pUS395 constructs. Likewise, a *lysR* homologue deletion mutant strain of *S. soli* could be generated to determine if this regulatory protein is responsible for the regulation of *zmoABCD* expression.

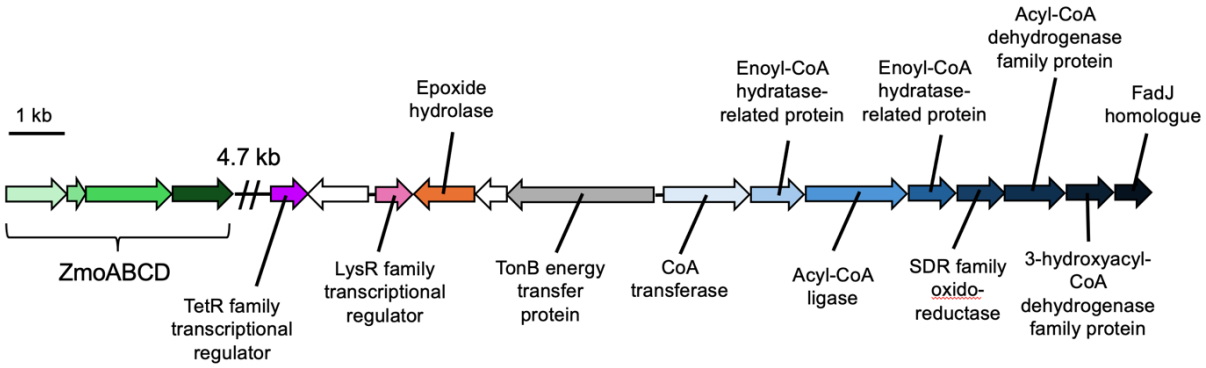


Figure 5.4.1. Genomic context of the regulatory proteins neighbouring the *zmoABCD* cluster in the *S. soli* genome.

Colour scheme is as follows: Green, the *zmoABCD* cluster; blue, fatty acid metabolism genes; purple, *tetR* family regulator; pink, *lysR* family regulator; orange, epoxide hydrolase homologue; grey, *tonB* family energy transfer protein; white, all other genes. The accession number of the locus is NZ_AXDW01000023.1.

While evidence points towards the ZmoABCD being a monooxygenase that oxidises C_3-C_4 compounds, more experimental work is needed to verify this. Short chain fatty acids and carboxylic acids could be tested as potential inducers using the bioreporter system. A transcriptomics approach through RT-qPCR could also be utilised to determine if *zmoABCD* is expressed at a low level and upregulated when exposed to inducers. In addition, a DNA electrophoretic mobility shift assay (Hellman and Fried, 2007) combined with mass spectrometry should be utilised to identify the repressor protein controlling the expression of *zmoABCD* genes. Once the repressor protein has been identified, further bioreporters could be constructed outside of the *S. soli* native host to determine the full inducer range of the *zmoABCD*.

Chapter 6 – General Discussion

6.1. General Discussion

This study has expanded the phylogeny of the SDIMO family of monooxygenases to include a novel group, the group 7 SDIMOs (Chapter 1). While the substrate range of this novel group remains to be fully characterised, several experiments have narrowed the search for the potential canonical substrate of the *S. soli* group 7 SDIMO, ZmoABCD. Heterologous expression of the ZmoABCD monooxygenase in *P. putida* KT2440 showed that the monooxygenase is biologically active against alkenes, especially short-chain alkenes where most epoxide was produced from propene (Chapter 3). Recombinant ZmoABCD also epoxidised the organochlorines vinyl chloride (VC) and 1,2-*cis*-dichloroethylene (cDCE), suggesting a potential bioremediation application of this SDIMO against these groundwater pollutants. However, *S. soli* is unable to utilise short-chain alkanes or alkenes as growth substrates, and *S. soli* Δ*zmoABCD* knockout mutants were used to confirm that ZmoABCD was not the sole locus responsible for the growth on phenol, *n*-octane, isoleucine, leucine or Tween 80 (Chapter 4). Interestingly, expression of *zmoABCD* was induced when *S. soli* cells were exposed to butane, 1-butene, 2-propanol, and Tween 80, but not after exposure to C_3 hydrocarbons like propane, propene or any of the butanol isomers (Chapter 5). However, several questions remain unanswered, including 1. if the alkene epoxidation observed in the *P. putida* heterologous host is an “off-target”, co-metabolic effect, 2. what the lack of conserved genes around the *zmoABCD* cluster means for its substrate range, 3. if it is possible that ZmoABCD acts like the P450 MOs and is responsible for the production of secondary metabolites instead of primary metabolism, and 4. what the inducers of *zmoABCD* gene expression means for the canonical substrate of the group 7 SDIMO.

Recombinant P. putida KT2440 was able to produce epoxides from several alkene substrates, including C_2-C_8 alkenes, the branched chain alkene isoprene, the carboxylic acid acrylic acid, and chlorinated alkenes like vinyl chloride (VC) and *cis*-1,2-dichloroethylene (cDCE) (Figure 3.3.7). The native *host S. soli,* on the other hand, was unable to utilise the above as growth substrates (Figure 4.3.5). As a result, it is evident that alkene epoxidation is an “off-target”, co-metabolic effect.

However, it is confounding that ZmoABCD in recombinant *P. putida* KT2440 produced the most epoxide from propene but not 1-butene while *zmoABCD* expression was induced by the latter and not the former (Figure 5.3.7). It could be that *S. soli* struggled to uptake shorter-chain hydrocarbons (i.e. $<C_4$) or that there was a competing metabolic pathway in its genome. Either of these situations seem unlikely, since smaller molecules should be able to diffuse through the cell membrane more easily, and *S. soli* lacks another obvious short-chain alkane/alkene oxidising enzyme in its genome. Another possibility is that the half-lives of 1,2-butylene oxide and propylene oxide are different and were therefore was not detected as well by the NBP assay, similar to the VC epoxide and cDCE epoxide situation (Chapter 3). However, this seems unlikely as the half-lives of these epoxides are quite long, unlike the VC epoxide and cDCE epoxide (Van Hylckama *et al.*, 1996). For 1,2-butylene oxide, its half-life is 6.5 days in aqueous solution, (NCBI, 2025) and the half-life of propylene oxide is 11.5 days in aqueous solution (Kahlich *et al.*, 1993). Alternatively, the position of the carbon-carbon double bond may play a role. Knowing that *zmoABCD* expression is induced by 2-propanol but not 1-propanol suggests that ZmoABCD is unable to oxidise 1-butene as effectively as it could 2-butene. It would be interesting to determine if the position of the carbon-carbon double bond impacts the epoxidation ability of ZmoABCD.

The lack of conserved genes surrounding the *zmoABCD* cluster and the related group 7 SDIMOs is unusual (Figure 3.3.5), since other SDIMOs usually have neighbouring conserved gene clusters that support the metabolism of the product of SDIMO oxidation. For example, this is seen in the regions neighbouring the group 1 and group 2 SDIMOs in *S. soli*, where catechol degradation genes are located neighbouring the toluene monooxygenase (TMO) and phenol monooxygenase (PMO) (Figure 4.3.10). While the *S. soli* TMO and PMO are not on the same operon, there is one example of interplay between the group 1 and group 2 SDIMOs in *Pseudomonas stutzeri* OX1, where its TMO oxidises benzene to phenol, and its PMO oxidises phenol to catechol, limiting the accumulation of phenol in the medium (Cafaro *et al.*, 2004). This suggests that the TMO and PMO in *S. soli* may also play a concerted role to oxidise aromatic hydrocarbons, reduce toxicity of these molecules, and allow growth utilising these substrates. Another example is the group 4 alkene

SDIMOs. The alkene monooxygenases have conserved genes for Coenzyme M biosynthesis and epoxide metabolism flanking the SDIMO clusters (Mattes *et al.*, 2010). The lack of conserved genes around the *zmoABCD* cluster may suggest that the substrate of ZmoABCD is either a central metabolite, or not involved in metabolism at all, and could be involved in the production of secondary metabolites like the P450 monooxygenases instead (Greule *et al.*, 2018). One piece of evidence supporting the former argument is that the fluorescence intensity patterns of the pUS389-395 constructs were quite similar in both *E. coli* and *S. soli*, suggesting that both organisms harbour repressor proteins that are able to bind to the putative operator region and regulate the expression of the *zmo* promoter in a similar fashion. *E. coli* (pUS389) cells could be exposed to Tween 80, butane, 1-butene, and 2-propanol to determine if these substrates can induce expression of the *zmo* promoter in this Gammaproteobacterial host. The fact that the fluorescence intensity patterns are remarkably similar in these somewhat related hosts could suggest that a universal regulator, rather than a hydrocarbon-specific regulator, is responsible for the control of *zmoABCD* expression. To further confirm this, the use of the electrophoretic mobility shift assay (EMSA) (Hellman and Fried, 2007) and mass spectrometry to identify the *S. soli* repressor protein is vital. This approach was taken to identify CrgA in *Pseudomonas aeruginosa* SJTD-1, one of the LysR family of transcriptional regulators, and demonstrated that CrgA regulated the expression of an *alkB* monooxygenase (Ji *et al.*, 2019). Once identified, the closest relative of the repressor protein can be used to infer the family or type of the repressor protein and potentially uncover previously missed information about the inducer and hence the canonical substrate of ZmoABCD.

The narrow growth substrate range of *S. soli* is interesting (Chapter 4). Perhaps this indicates that the organism is oligotrophic and cannot withstand high levels of nutrients, but the increased growth yield when cultured in a modified R2A medium with an additional 5 g/L yeast extract and MOPS buffer argues against this (Chapter 4). It is possible that trace elements and vitamins present in the additional yeast extract were responsible for the increase in growth yield, but this facet was not pursued in this study. Future studies may assess the impact of vitamins on *S. soli* growth by supplementing a vitamin solution to R2A medium instead of yeast extract.

Solimonas soli is part the Nevskiaceae family which includes the hydrocarbon-utilising genera *Hydrocarboniphaga* and *Alkanibacter*. The former is able to grow using linear alkanes and aromatic hydrocarbons (Palleroni *et al.*, 2004) while the latter only grows on certain alkanes (pentane, hexane, and decane, weak growth with nonane, and no growth with heptane or octane) (Friedrich and Lipski, 2008). Though only two representatives from the genus *Hydrocarboniphaga* have been sequenced and at least one AlkB MO has been identified through sequence data (Chang *et al.*, 2012), it is highly likely that *Alkanibacter difficilis*, the sole representative of the *Alkanibacter* genus, also contains at least one AlkB MO responsible for the oxidation of alkanes as growth substrates. *Solimonas soli* and these Nevskiaceae members are quite close relatives (Sheu *et al.*, 2011). This suggests that perhaps not unlike these other Nevskiaceae genera, *S. soli* utilises its AlkB MO to grow on alkanes and relies on its TMO and PMO for growth on aromatic hydrocarbons, and ZmoABCD is not involved in hydrocarbon metabolism at all, but is utilised in the production of secondary metabolites instead.

One major group of oxygenases involved in the biosynthesis of secondary metabolites is the cytochrome P450 monooxygenases (P450s). They are a diverse family of enzymes found in all domains of life that can facilitate a diverse range of reactions (Nelson, 2018). In bacteria, these reactions include and are not limited to secondary metabolite biosynthesis, terpene metabolism, and steroid degradation (Greule *et al.*, 2018). The P450s are also known to be involved in alkane and fatty acid metabolism (Van Bogaert *et al.*, 2011), and since *S. soli* itself has at least 9 P450 homologues (based on a search of the NCBI database using the BLASTp function), the P450s may be the enzymes responsible for growth on Tween 80 and *n*-octane, as the metabolism of both of these compounds generates fatty acid intermediates. However, it is unlikely that ZmoABCD catalyses secondary metabolite biosynthesis reactions like the P450s, since many of the characteristics of ZmoABCD and the *zmoABCD* gene cluster do not correspond to characteristics of the P450 MOs. Firstly, the P450s are usually co-located with other genes required for biosynthesis (Greule *et al.*, 2018), but this is not the case for the group 7 SDIMOs (Figure 3.3.5). In addition, the biosynthesis reactions that P450s catalyse almost always involve large substrates (Greule *et al.*, 2018), unlike the small alkene

substrates epoxidised by recombinant *P. putida* KT2440 expressing ZmoABCD (Figure 3.3.7). Finally, as mentioned above, *S. soli* itself has at least 9 P450 homologues. Since only 18% of Gammaproteobacterial P450s are involved in secondary metabolism in the known and sequenced Gammaproteobacteria database (Msomi *et al.*, 2021), it is unlikely that *S. soli* has a non-P450 MO to facilitate biosynthesis reactions. However, to test if ZmoABCD is not involved in the production of secondary metabolites, a metabolomic approach (Breitling *et al.*, 2013) should be employed to determine the metabolites produced by wildtype S. soli versus Δ*zmoABCD S. soli* mutants grown in different media and exposed to different carbon sources.

In Chapter 3, fatty acids were not tested as potential substrates due to limitations with the NBP assay, as the assay is only able to detect epoxide products. However, the induction results in Chapter 5 suggest that candidate substrates that may be recognised by common regulators are carboxylic acids, more specifically fatty acids. Heterologously expressed ZmoABCD in *P. putida* KT2440 exhibited the ability to epoxidise acrylic acid, suggesting that the *S. soli* group 7 SDIMO is able to epoxidise functional groups neighbouring a carboxylic acid group (Figure 3.3.7). This is interesting because while other bacteria have been found to be able to utilise acrylic acid as a growth substrate (He *et al.*, 2020), no other SDIMO has demonstrated the ability to epoxidise it. The fact that the detergent Tween 80 was able to induce the expression of the *zmo* promoter suggests that either the downstream products of Tween 80 metabolism (which are likely fatty acids and sorbitan (Pietersen *et al.*, 2020)) or Tween 80 itself may be a substrate of ZmoABCD. The former is more likely as Δ*zmoABCD* knockout strains of *S. soli* retained the ability to utilise Tween 80 as a sole carbon source, unless, while highly unlikely, there are redundant pathways of Tween 80 metabolism in the *S. soli* genome. One possibility is that the repressor protein responsible for the regulation of *zmoABCD* expression is a stress response protein that is induced by long chain fatty acids like oleic acid, a metabolite of Tween 80. Regarding the sorbitan moiety of Tween 80, since *S. soli* was unable to grow using structurally similar tetrahydrofuran (THF) as a sole carbon source (Figure 4.3.5), it is unlikely that the sorbitan moiety of Tween 80 played a role in either the promotion of *S. soli* growth or the induction of *zmoABCD* expression. However, more

experiments that focus on THF and sorbitan (e.g. test if recombinant *P. putida* KT2440 can oxidise THF, and to include THF in the inducer screen experiment) should be considered before a conclusion can be made.

Fatty acid metabolism is known to increase stress in microorganisms. Exposure to and metabolism of long chain fatty acids has been shown to cause stress responses in both *E. coli* (Jaswal *et al.*, 2021) and *Mycobacterium tuberculosis* (Pietersen et al., 2020). In addition, since the TetR/AcrR family of regulators are known to be involved in responses to fatty acid stress (Cuthbertson and Nodwell, 2013), it could be that the TetR repressor homologue near the *zmoABCD* cluster is responsible for the regulation of the operon. A stress response is generated in *E. coli* when the organism is exposed to sodium oleate at 5 mM (Agrawal *et al.*, 2017). Likewise, a stress response is also generated in *M. tuberculosis* after exposure to 0.003% (w/v) long chain fatty acids (equal amounts of oleic acid, stearic acid, and palmitic acid, about 100 μM of fatty acids) (Rodriguez *et al.*, 2014). While it has been suggested that the stress response in these organisms is upregulated due to processes related to fatty acid metabolism, it is unknown if the stress response also arose due to membrane disruption by these long chain fatty acids. *S. soli* does have a TetR repressor homologue near the *zmoABCD* cluster (Figure 5.4.1) and since TetR/AcrR-family regulators are known to be involved in responses to fatty acid metabolism stress (Cuthbertson and Nodwell, 2013), it could be that this TetR repressor homologue is responsible for the regulation of the *zmoABCD* operon. However, butane, 1-butene, and 2-propanol also induced the expression of the *zmo* promoter, and these substrates more closely resemble short chain fatty acids than their long chain counterparts. Perhaps the true substrate of ZmoABCD is a metabolite of long chain fatty acids that has partially undergone β-oxidation and when the fatty acid chain reaches C_4 ZmoABCD oxidises the molecule and the product enters central metabolism. One piece of evidence supporting this great deviation of substrate between the *S. soli* group 7 SDIMO and the other existing SDIMOs is that many of the conserved amino acids in the other aliphatic SDIMO groups are substituted by non-synonymous amino acids in the group 7 SDIMOs (Figure 3.3.2). However, it is difficult to infer information from the amino acid sequence as no crystal structure of ZmoABCD is available. While protein structures can be predicted through

computational approaches using software like AlphaFold, this approach relies on efficacy and the robustness of the prediction software, which in turn relies on the robustness of the database of previously identified and characterised SDIMOs. In addition, a computational approach overlooks the enzyme “dark matter” of uncharacterised enzymes and will likely be unreliable when it comes to novel SDIMOs like the group 7 SDIMOs. While incredible advances have been made regarding protein structure predictors, there are still limits to the number of chains and residues the AlphaFold programme can fold and assemble (Terwilliger *et al.*, 2024). The fact that SDIMOs are multimeric enzymes further complicates the use of the computational approach.

6.2. Conclusions and outlook

The combined results gathered in this study strongly suggest that the canonical substrate of ZmoABCD is a non-hydrocarbon C_4 molecule that will be oxidised at a subterminal carbon. The following next steps should be pursued to identify the canonical substrate of the ZmoABCD. Recombinant *P. putida* KT2440 expressing ZmoABCD should be tested for activity against alkanes, especially butane for alcohol production, activity against 2-butene and a series of carboxylic acids as highlighted above. Since *zmo* was shown to be most strongly induced by butane in *S. soli*, it would be interesting to determine if butane-exposed *S. soli* can transform butane even though it cannot grow utilising the gaseous hydrocarbon. Gas chromatography can be used to measure butane depletion in tandem with high performance liquid chromatography to quantify possible butanol production in order to determine *S. soli* activity towards butane. Gas chromatography also can be utilised to detect substrate depletion by recombinant *P. putida* KT2440 expressing ZmoABCD, in addition to the aforementioned *S. soli* native host experiments. Resultingly, gas chromatography can be used to complement the NBP epoxide detection assay which measures product formation, while gas chromatography enables the detection of substrate depletion, even if the epoxide produced in unstable (e.g. VC epoxide). Gas chromatography also allows for experiments using the heterologous host targeting other volatile substrates like alkanes without the need for a colorimetric assay for product detection, further enabling testing of a wider substrate range of ZmoABCD. Identification of the repressor protein responsible for the regulation of *zmoABCD*

expression will also prove worthwhile to guide future research towards the canonical substrate of this novel SDIMO. Another major step that could be taken is to generate a crystal structure of ZmoABCD, as this would allow us to go beyond sequence gazing and examine the substrate channel of this SDIMO in detail. In addition, the generation of the ZmoABCD crystal structure would prove massively beneficial to those studying SDIMOs, as to date only the sMMO and the toluene MO has had their crystal structures explored. This would not only benefit future research regarding ZmoABCD but also contribute to the robustness of the protein sequence and structure databases that computational analyses heavily rely upon.

While this study was unable to confirm whether the canonical substrate of ZmoABCD is a growth substrate for *S. soli* or if ZmoABCD is involved in the production of secondary metabolites, this study has provided a framework to narrow the potential substrate range of ZmoABCD, the first of the group 7 SDIMOs to be experimentally characterised. This study has also generated data, protocols, plasmids, and bacterial strains which will prove be useful in future studies exploring and identifying the canonical substrate range of ZmoABCD. Once the canonical substrate range of this novel group of SDIMOs is found, our understanding of the evolutionary history of the SDIMOs will be expanded, as will our ability to find potential applications for this new group of SDIMOs, be it for biocatalysis or for bioremediation purposes.

Chapter 7 – References

Abbasian, F., Lockington, R., Mallavarapu, M., Naidu, R. (2015) A comprehensive review of aliphatic hydrocarbon biodegradation by bacteria. Applied Biochemistry and Biotechnology, 176: 670-699.

Agrawal, S., Jaswal, K., Shiver, A. L., Balecha, H., Patra, T., Chaba, R. (2017) A genome-wide screen in *Escherichia coli* reveals that ubiquinone is a key antioxidant for metabolism of long-chain fatty acids. Journal of Biological Chemistry, 292: 20086-20099.

Alexeyev, M. F. (1999) The pKNOCK series of broad-host-range mobilizable suicide vectors for gene knockout and targeted DNA insertion into the chromosome of gram-negative bacteria. Biotechniques, 26: 824-828.

Allard, N., Garneau, D., Poulin-Laprade, D., Burrus, V., Brzezinski, R., Roy, S. (2015) A diaminopimelic acid auxotrophic *Escherichia coli* donor provides improved counterselection following intergeneric conjugation with actinomycetes. Canadian Journal of Microbiology, 61: 565-574.

Alvarez-Cohen, L., Speitel, G. E., Jr. (2001) Kinetics of aerobic cometabolism of chlorinated solvents. Biodegradation, 12: 105-126.

Angov, E., Hillier, C. J., Kincaid, R. L., Lyon, J. A. (2008) Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. PLoS One, 3: e2189.

Arai, H., Akahira, S., Ohishi, T., Maeda, M., Kudo, T. (1998) Adaptation of *Comamonas testosteroni* TA441 to utilize phenol: Organization and regulation of the genes involved in phenol degradation. Microbiology, 144: 2895-2903.

Ayilara, M. S., Babalola, O. O. (2023) Bioremediation of environmental wastes: the role of microorganisms. Frontiers in Agronomy, 5: 1183691.

Ayub, H., Kang, M. J., Farooq, A., Jung, M. Y. (2022) Ecological aerobic ammonia and methane oxidation involved key metal compounds, Fe and Cu. Life, 12: 1806.

Bagdasarian, M., R. Lurz, B. R., Franklin, F. C. H., Bagdasarian, M. M., Frey, J., Timmis, K. N. (1981) Specific-purpose plasmid cloning vectors II. Broad host range, high copy number, RSF 1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas*. Gene, 16: 237-247.

Banerjee, R., Jones, J. C., Lipscomb, J. D. (2019) Soluble methane monooxygenase. Annual Review of Biochemistry, 88: 409-431.

Baptist, J. N., Gholson, R. K., Coon, M. J. (1963) Hydrocarbon oxidation by a bacterial enzyme system. I. Products of octane oxidation. Biochimica et Biophysica Acta, 69: 40-47.

Bartilson, M., Nordlund, I., Shingler, V. (1990) Location and organization of the dimethylphenol catabolic genes of *Pseudomonas* CF600. Molecular and General Genetics, 220: 294-300.

Baugh, A. C., Momany, C., Neidle, E. L. (2023) Versatility and complexity: Common and uncommon facets of LysR-type transcriptional regulators. Annual Review of Microbiology, 77: 317-339.

Bernhardt, R. (2006) Cytochromes P450 as versatile biocatalysts. Journal of Biotechnology, 124: 128-145.

Bertoni, G., Bolognese, F., Galli, E., Barbieri, P. (1996) Cloning of the genes for and characterization of the early stages of toluene and *o*-xylene catabolism in *Pseudomonas stutzeri* OX1. Applied and Environmental Microbiology, 62: 3704-3711.

Bertoni, G., Martino, M., Galli, E., Barbieri, P. (1998) Analysis of the gene cluster encoding toluene/*o*-xylene monooxygenase from *Pseudomonas stutzeri* OX1. Applied and Environmental Microbiology, 64: 3626–3632.

Bhalla, N., Jolly, P., Formisano, N., Estrela, P. (2016) Introduction to biosensors. Essays in Biochemistry, 60: 1-8.

Birolli, W. G., Lima, R. N., Porto, A. L. M. (2019) Applications of marine-derived microorganisms and their enzymes in biocatalysis and biotransformation, the underexplored potentials. Frontiers in Microbiology, 10: 1453.

Boethling, R. S., Alexander, M. (1979) Microbial degradation of organic compounds at trace levels. Environmental Science & Technology, 13: 989-991.

Bopp, C. E., Bernet, N. M., Kohler, H. E., Hofstetter, T. B. (2022) Elucidating the role of O(2) uncoupling in the oxidative biodegradation of organic contaminants by Rieske non-heme iron dioxygenases. ACS Environmental Au, 2: 428-440.

Breitling, R., Ceniceros, A., Jankevics, A., Takano, E. (2013) Metabolomics for secondary metabolite research. Metabolites, 3: 1076-1083.

Brewster, J. D. (2003) A simple micro-growth assay for enumerating bacteria. Journal of Microbiological Methods, 53: 77-86.

Bryan, M. C., Dunn, P. J., Entwistle, D., Gallou, F., Koenig, S. G., Hayler, J. D., Hickey, M. R., Hughes, S., Kopach, M. E., Moine, G., Richardson, P., Roschangar, F., Steven, A., Weiberth, F. J. (2018) Key Green Chemistry research areas from a pharmaceutical manufacturers' perspective revisited. Green Chemistry, 20: 5082-5103.

Burbano, D. A., Kiattisewee, C., Karanjia, A. V., Cardiff, R. A. L., Faulkner, I. D., Sugianto, W., Carothers, J. M. (2024) CRISPR tools for engineering prokaryotic systems: recent advances and new applications. Annual Review of Chemical and Biomolecular Engineering, 15: 389-430.

Cafaro, V., Izzo, V., Scognamiglio, R., Notomista, E., Capasso, P., Casbarra, A., Pucci, P., Di Donato, A. (2004) Phenol hydroxylase and toluene/*o*-xylene monooxygenase from *Pseudomonas stutzeri* OX1: interplay between two enzymes. Applied and Environmental Microbiology, 70: 2211-2219.

Cain, A. K., Barquist, L., Goodman, A. L., Paulsen, I. T., Parkhill, J., van Opijnen, T. (2020) A decade of advances in transposon-insertion sequencing. Nature Reviews Genetics, 21: 526-540.

Canada, K. A., Iwashita, S., Shim, H., Wood, T. K. (2002) Directed evolution of toluene *ortho*-monooxygenase for enhanced 1-naphthol synthesis and chlorinated ethene degradation. Journal of Bacteriology, 184: 344-349.

Cardy, D. L. N., Laidler, V., Salmond, G. P. C., Murrell, J. C. (1991) Molecular analysis of the methane monooxygenase (MMO) gene cluster of *Methyiosinus trichosporium* OB3b. Molecular Biotechnology, 5: 335-342.

Carpenter, A. C., Paulsen, I. T., Williams, T. C. (2018) Blueprints for biosensors: Design, limitations, and applications. Genes (Basel), 9: 375.

Casaite, V., Stanislauskiene, R., Vaitekunas, J., Tauraite, D., Rutkiene, R., Gasparaviciute, R., Meskys, R. (2020) Microbial Degradation of Pyridine: a Complete Pathway in *Arthrobacter* sp. Strain 68b Deciphered. Applied and Environmental Microbiology, 86: e00902-20.

Casas, A., Bultelle, M., Kitney, R. (2024) An engineering biology approach to automated workflow and biodesign. Synthetic Biology, 9: ysae009.

Chan Kwo Chion, C. K., Askew, S. E., Leak, D. J. (2005) Cloning, expression, and site-directed mutagenesis of the propene monooxygenase genes from *Mycobacterium* sp. strain M156. Applied and Environmental Microbiology, 71: 1909-1914.

Chang, H.-K., Zylstra, G. J., Chae, J.-C. (2012) Genome sequence of *n*-Alkane-degrading *Hydrocarboniphaga effusa* strain AP103T (ATCC BAA-332T). Journal of Bacteriology, 194: 5120.

Cheung, S., McCarl, V., Holmes, A. J., Coleman, N. V., Rutledge, P. J. (2013) Substrate range and enantioselectivity of epoxidation reactions mediated by the ethene-oxidising *Mycobacterium* strain NBB4. Applied and Environmental Microbiology, 97: 1131-1140.

Cho, I. H. (2015) Volatile compounds of ginseng (*Panax* sp.): a review. Journal of the Korean Society for Applied Biological Chemistry, 58: 67-75.

Chuang, A. S., Mattes, T. E. (2007) Identification of polypeptides expressed in response to vinyl chloride, ethene, and epoxyethane in *Nocardioides* sp. strain JS614 by using peptide mass fingerprinting. Applied and Environmental Microbiology, 73: 4368-4372.

Chuvochina, M., Rinke, C., Parks, D. H., Rappe, M. S., Tyson, G. W., Yilmaz, P., Whitman, W. B., Hugenholtz, P. (2019) The importance of designating type material for uncultured taxa. Systematic and Applied Microbiology, 42: 15-21.

Colby, J., Stirling, D. I., Dalton, H. (1977) The soluble methane mono-oxygenase of *Methylococcus capsulatus* (Bath). Its ability to oxygenate *n*-alkanes, *n*-alkenes, ethers, and alicyclic, aromatic and heterocyclic compounds. Biochemical Journal, 165: 395-402.

Coleman, N. V., Bui, N. B., Holmes, A. J. (2006) Soluble di-iron monooxygenase gene diversity in soils, sediments and ethene enrichments. Environmental Microbiology, 8: 1228-1239.

Coleman, N. V., Le, N. B., Ly, M. A., Ogawa, H. E., McCarl, V., Wilson, N. L., Holmes, A. J. (2012) Hydrocarbon monooxygenase in *Mycobacterium*: recombinant expression of a member of the ammonia monooxygenase superfamily. The ISME Journal, 6: 171-182.

Coleman, N. V., Mattes, T. E., Gossett, J. M., Spain, J. C. (2002) Biodegradation of *cis*-dichloroethene as the sole carbon source by a ß-proteobacterium. Applied and Environmental Microbiology, 68: 2726-2730.

Coleman, N. V., Spain, J. C. (2003a) Distribution of the coenzyme M pathway of epoxide metabolism among ethene- and vinyl chloride-degrading *Mycobacterium* strains. Applied and Environmental Microbiology, 69: 6041-6046.

Coleman, N. V., Spain, J. C. (2003b) Epoxyalkane:coenzyme M transferase in the ethene and vinyl chloride biodegradation pathways of *Mycobacterium* strain JS60. Journal of Bacteriology, 185: 5536-5545.

Coleman, N. V., Yau, S., Wilson, N. L., Nolan, L. M., Migocki, M. D., Ly, M. A., Crossett, B., Holmes, A. J. (2011) Untangling the multiple monooxygenases of *Mycobacterium* chubuense strain NBB4, a versatile hydrocarbon degrader. Environmental Microbiology Reports, 3: 297-307.

Constable, D. J. C., Dunn, P. J., Hayler, J. D., Humphrey, G. R., Leazer, J. J. L., Linderman, R. J., Lorenz, K., Manley, J., Pearlman, B. A., Wells, A., Zaks, A., Zhang, T. Y. (2007) Key green chemistry research areas—a perspective from pharmaceutical manufacturers. Green Chemistry, 9: 411-420.

Crombie, A. T., Khawand, M. E., Rhodius, V. A., Fengler, K. A., Miller, M. C., Whited, G. M., McGenity, T. J., Murrell, J. C. (2015) Regulation of plasmid-encoded isoprene metabolism in *Rhodococcus*, a representative of an important link in the global isoprene cycle. Environmental Microbiology, 17: 3314-3329.

Cronan, J. E. (2021) The *Escherichia coli* FadR transcription factor: Too much of a good thing? Molecular Microbiology, 115: 1080-1085.

Cuthbertson, L., Nodwell, J. R. (2013) The TetR family of regulators. Microbiology and Molecular Biology Reviews, 77: 440-475.

Dawson, R. A., Crombie, A. T., Jansen, R. S., Smith, T. J., Nichol, T., Murrell, C. (2023) Peering down the sink: A review of isoprene metabolism by bacteria. Environmental Microbiology, 25: 786-799.

de Lorenzo, V. (2009) Recombinant bacteria for environmental release: what went wrong and what we have learnt from it. Clinical Microbiology and Infection, 15: 63-65.

de Lorenzo, V., Timmis, K. N. (1994) Analysis and construction of stable phenotypes in gram-negative bacteria with Tn*5*- and Tn*10*-derived minitransposons. Methods in Enzymology, 235: 386-405.

Deng, D., Li, F., Li, M. (2018) A novel propane monooxygenase initiating degradation of 1,4-dioxane by *Mycobacterium dioxanotrophicus* PH-06. Environmental Science & Technology Letters, 5: 86-91.

Deng, D., Pham, D. N., Li, F., Li, M. (2020) Discovery of an inducible toluene monooxygenase that cooxidizes 1,4-dioxane and 1,1-dichloroethylene in propanotrophic *Azoarcus* sp. strain DD4. Applied and Environmental Microbiology, 86: e01163-20.

Desai, S. H., Koryakina, I., Case, A. E., Toney, M. D., Atsumi, S. (2016) Biological conversion of gaseous alkenes to liquid chemicals. Metabolic Engineering, 38: 98-104.

Diamond, S., Lavy, A., Crits-Christoph, A., Carnevali, P. B. M., Sharrar, A., Williams, K. H., Banfield, J. F. (2022) Soils and sediments host Thermoplasmata archaea encoding novel copper membrane monooxygenases (CuMMOs). The ISME Journal, 16: 1348-1362.

Donadio, G., Sarcinelli, C., Pizzo, E., Notomista, E., Pezzella, A., Di Cristo, C., De Lise, F., Di Donato, A., Izzo, V. (2015) The toluene *o*-xylene monooxygenase enzymatic activity for the biosynthesis of aromatic antioxidants. PLoS One, 10: e0124427.

Dong, X., Rattray, J. E., Campbell, D. C., Webb, J., Chakraborty, A., Adebayo, O., Matthews, S., Li, C., Fowler, M., Morrison, N. M., MacDonald, A., Groves, R. A., Lewis, I. A., Wang, S. H., Mayumi, D., Greening, C., Hubert, C. R. J. (2020) Thermogenic hydrocarbon biodegradation by diverse depth-stratified microbial populations at a Scotian Basin cold seep. Nature Communications, 11: 5825.

Dower, W. J., Miller, J. F., Ragsdale, C. W. (1988) High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Research, 16: 6127–6145.

Dvořák, P., Nikel, P. I., Damborsky, J., de Lorenzo, V. (2017) Bioremediation 3.0: Engineering pollutant-removing bacteria in the times of systemic biology. Biotechnology Advances, 35: 845-866.

Eaton, R. W. (1997) *p*-Cymene catabolic pathway in *Pseudomonas putida* F1: cloning and characterization of DNA encoding conversion of *p*-Cymene to *p*-cumate. Journal of Bacteriology, 179: 3171-3180.

Fetzner, S. (2002) Oxygenases without requirement for cofactors or metal ions. Applied Microbiology and Biotechnology, 60: 243-257.

Fetzner, S., Kolkenbrock, S., Parschat, K. (2007). Catabolic linear plasmids. In Microbial Linear Plasmids. Meinhardt, F. & Klassen, R. (Ed.) Springer, New York, NY, United States. **7:** 63-98.

Frascari, D., Cappelletti, M., Fedi, S., Verboschi, A., Ciavarelli, R., Nocentini, M., Pinelli, D. (2012) Application of the growth substrate pulsed feeding technique to a process of chloroform aerobic cometabolism in a continuous-flow sand-filled reactor. Process Biochemistry, 47: 1656-1664.

Friedrich, M. M., Lipski, A. (2008) *Alkanibacter difficilis* gen. nov., sp. nov. and *Singularimonas variicoloris* gen. nov., sp. nov., hexane-degrading bacteria isolated from a hexane-treated biofilter. International Journal of Systematic and Evolutionary Microbiology, 58: 2324-2329.

Fu, J., Teucher, M., Anastassiadis, K., Skarnes, W., Stewart, A. F. (2010). A recombineering pipeline to make conditional targeting constructs. In Guide to Techniques in Mouse Development, Part B: Mouse Molecular Genetics, 2nd Edition. Wassarman, P. M. & Soriano, P. M. (Ed.) Academic Press, Cambridge, MA, United States. 125-144.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W. (2012) CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics, 28: 3150-3152.

Furuya, T., Hayashi, M., Kino, K. (2013) Reconstitution of active mycobacterial binuclear iron monooxygenase complex in *Escherichia coli*. Applied and Environmental Microbiology, 79: 6033-6039.

Furuya, T., Hayashi, M., Semba, H., Kino, K. (2012) The mycobacterial binuclear iron monooxygenases require a specific chaperonin-like protein for functional expression in a heterologous host. The FEBS Journal, 280: 817-826.

Furuya, T., Hirose, S., Osanai, H., Semba, H., Kino, K. (2011) Identification of the monooxygenase gene clusters responsible for the regioselective oxidation of phenol to hydroquinone in mycobacteria. Applied and Environmental Microbiology, 77: 1214-1220.

Furuya, T., Nakao, T., Kino, K. (2015) Catalytic function of the mycobacterial binuclear iron monooxygenase in acetone metabolism. FEMS Microbiology Letters, 362: fnv136.

Gallagher, S. C., Cammack, R., Dalton, H. (1997) Alkene monooxygenase from *Nocardia corallina* B-276 is a member of the class of dinuclear iron proteins capable of stereospecific epoxygenation reactions. European Journal of Biochemistry, 247: 635-641.

Gibson, D. T., Parales, R. E. (2000) Aromatic hydrocarbon dioxygenases in environmental biotechnology. Current Opinion in Biotechnology, 11: 236-243.

Goldstein, B. P. (2014) Resistance to rifampicin: A review. The Journal of Antibiotics, 67: 625-630.

Green, M. R., Sambrook, J. (2016) Preparation of plasmid DNA by alkaline lysis with sodium dodecyl sulfate: minipreps. Cold Spring Harbor Protocols, 2016: 911-916.

Greening, C., Grinter, R. (2022) Microbial oxidation of atmospheric trace gases. Nature Reviews Microbiololgy, 20: 513-528.

Greule, A., Stok, J. E., De Voss, J. J., Cryle, M. J. (2018) Unrivalled diversity: The many roles and reactions of bacterial cytochromes P450 in secondary metabolism. Natural Product Reports, 35: 757-791.

Guengerich, F. P. (1992) Roles of the vinyl chloride oxidation products 2-chlorooxirane and 2-chloroacetaldehyde in the in vitro formation of etheno adducts of nucleic acid bases. Chemical Research in Toxicology, 5: 2-5.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology, 59: 307-321.

Guo, X., Zhang, J., Han, L., Lee, J., Williams, S. C., Forsberg, A., Xu, Y., Austin, R. N., Feng, L. (2023) Structure and mechanism of the alkane-oxidizing enzyme AlkB. Nature Communications, 14: 2180.

Gupta, R. S., Lo, B., Son, J. (2018) Phylogenomics and comparative genomic studies robustly support division of the genus *Mycobacterium* into an emended genus *Mycobacterium* and four novel genera. Frontiers in Microbiology, 9: 67.

Gustafsson, C., Govindarajan, S., Minshull, J. (2004) Codon bias and heterologous protein expression. Trends in Biotechnology, 22: 346-353.

Hage, J. C., Hartmans, S. (1999) Monooxygenase-mediated 1,2-dichloroethane degradation by *Pseudomonas* sp. strain DCA1. Applied and Environmental Microbiology, 65: 2466-2470.

Halsey, K. H., Sayavedra-Soto, L. A., Bottomley, P. J., Arp, D. J. (2005) Trichloroethylene degradation by butane-oxidizing bacteria causes a spectrum of toxic effects. Applied Microbiology and Biotechnology, 68: 794-801.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology, 166: 557-580.

He, J., Chen, Y., Dai, L., Yao, J., Mei, Y., Hrynsphan, D., Tatsiana, S., Chen, J. (2020) Rapid and complete biodegradation of acrylic acid by a novel strain *Rhodococcus ruber* JJ-3: Kinetics, carbon balance, and degradation pathways. Biotechnology and Bioprocess Engineering, 25: 589-598.

He, Y., Mathieu, J., Yang, Y., Yu, P., da Silva, M. L. B., Alvarez, P. J. J. (2017) 1,4-dioxane biodegradation by *Mycobacterium dioxanotrophicus* pH-06 is associated with a group-6 soluble di-iron monooxygenase. Environmental Science & Technology Letters, 4: 494-499.

Helbich, S., Barrantes, I., Dos Anjos Borges, L. G., Pieper, D. H., Vainshtein, Y., Sohn, K., Engesser, K. H. (2023) The 2-methylpropene degradation pathway in *Mycobacteriaceae* family strains. Environmental Microbiology, 25: 2163-2181.

Hellman, L. M., Fried, M. G. (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nature Protocols, 2: 1849-1861.

Henkin, T. M., Yanofsky, C. (2002) Regulation by transcription attenuation in bacteria: How RNA provides instructions for transcription termination/antitermination decisions. BioEssays, 24: 700-707.

Hernández-Lucas, I., Gallego-Hernández, A. L., Encarnación, S., Fernández-Mora, M., Martínez-Batallar, A. G., Salgado, H., Oropeza, R., Calva, E. (2008) The LysR-type transcriptional regulator LeuO controls expression of several genes in *Salmonella enterica* serovar Typhi. Journal of Bacteriology, 190: 1658-1670.

Herrero, M., de Lorenzo, V., Timmis, K. N. (1990) Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. Journal of Bacteriology, 172: 6557-6567.

Hino, S., Watanabe, K., Takahashi, N. (1998) Phenol hydroxylase cloned from *Ralstonia eutropha* strain E2 exhibits novel kinetic properties. Microbiology, 144: 1765-1772.

Hoang, T. T., Karkhoff-Schweizer, R. R., Kutchma, A. J., Schweizer, H. P. (1999) A broad-host-range Flp-*FRT* recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. Gene, 212: 77-86.

Holmes, A. J., Coleman, N. V. (2008) Evolutionary ecology and multidisciplinary approaches to prospecting for monooxygenases as biocatalysts. Antonie Van Leeuwenhoek, 94: 75-84.

Horinouchi, M., Yoshida, T., Nojiri, H., Yamane, H., Omori, T. (1999) Oxidation of dimethyl sulfide by various aromatic compound oxygenases from bacteria. Biotechnology Letters, 21: 929–933.

Horvath, R. S. (1972) Microbial co-metabolism and the degradation of organic compounds in nature. Bacteriology Reviews, 36: 146-155.

Howe, T. G. B., Ward, J. M. (1976) The utilization of Tween 80 as carbon source by *Pseudomonas*. Journal of General Microbiology, 92: 234-235.

Hutson, S. (2001). Structure and function of branched chain aminotransferase. In Progress in Nucleic Acid Research and Molecular Biology. Moldave, K. (Ed.) Academic Press, Cambridge, MA, United States. **70:** 175-206.

Inoue, D., Tsunoda, T., Sawada, K., Yamamoto, N., Saito, Y., Sei, K., Ike, M. (2016) 1,4-Dioxane degradation potential of members of the genera *Pseudonocardia* and *Rhodococcus*. Biodegradation, 27: 277-286.

Iwami, Y., Yamada, T. (1980) Rate-limiting steps of the glycolytic pathway in the oral bacteria *Streptococcus mutans* and *Streptococcus sanguis* and the influence of acidic pH on the glucose metabolism. Archives of Oral Biology, 25: 163-169.

Izzo, V., Leo, G., Scognamiglio, R., Troncone, L., Birolo, L., Di Donato, A. (2011) PHK from phenol hydroxylase of *Pseudomonas* sp. OX1. Insight into the role of an accessory protein in bacterial multicomponent monooxygenases. Archives of Biochemistry and Biophysics, 505: 48-59.

Jaswal, K., Shrivastava, M., Chaba, R. (2021) Revisiting long-chain fatty acid metabolism in *Escherichia coli*: Integration with stress responses. Current Genetics, 67: 573-582.

Ji, N., Wang, X., Yin, C., Peng, W., Liang, R. (2019) CrgA protein represses AlkB2 monooxygenase and regulates the degradation of medium-to-long-chain *n*-alkanes in *Pseudomonas aeruginosa* SJTD-1. Frontiers in Microbiology, 10: 400.

Ji, Y., Mao, G., Wang, Y., Bartlam, M. (2013) Structural insights into diversity and *n*-alkane biodegradation mechanisms of alkane hydroxylases. Frontiers in Microbiology, 4: 58.

Jin, Y. O., Cheung, S., Coleman, N. V., Mattes, T. E. (2010) Association of missense mutations in epoxyalkane coenzyme M transferase with adaptation of *Mycobacterium* sp. strain JS623 to growth on vinyl chloride. Applied and Environmental Microbiology, 76: 3413-3419.

Jin, Y. O., Mattes, T. E. (2008) Adaptation of aerobic, ethene-assimilating *Mycobacterium* strains to vinyl chloride as a growth substrate. Environmental Science & Technology, 42: 4784–4789.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., Madden, T. L. (2008) NCBI BLAST: a better web interface. Nucleic Acids Research, 36: W5-9.

Kahlich, D., Wiechern, U., Lindner, J. (1993). Propylene oxide. In Ullmann's Encyclopedia of Industrial Chemistry. Elvers, B., Hawklins, S. & Schulz, G. (Ed.) Wiley-VCH, Hoboken, NJ, United States. **A22:** 239–260.

Kalyuzhnaya, M. G., Gomez, O. A., Murrell, J. C. (2019). The methane-oxidizing bacteria (methanotrophs). In Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes. McGenity, T. J. (Ed.) Springer, New York, NY, United States. 1-34.

Kang, C. S., Dunfield, P. F., Semrau, J. D. (2019) The origin of aerobic methanotrophy within the Proteobacteria. FEMS Microbiology Letters, 366: fnz096.

Keener, W. K., Arp, D. J. (1994) Transformations of Aromatic Compounds by *Nitrosomonas europaea*. Applied and Environmental Microbiology, 60: 1914-1920.

Kelle, A. (2013) Beyond patchwork precaution in the dual-use governance of synthetic biology. Science and Engineering Ethics, 19: 1121-1139.

Khadka, R., Clothier, L., Wang, L., Lim, C. K., Klotz, M. G., Dunfield, P. F. (2018) Evolutionary history of copper membrane monooxygenases. Frontiers in Microbiology, 9: 2493.

Khider, M. L. K., Brautaset, T., Irla, M. (2021) Methane monooxygenases: Central enzymes in methanotrophy with promising biotechnological applications. World Journal of Microbiology and Biotechnology, 37: 72.

Kim, M. K., Kim, Y. J., Cho, D. H., Yi, T. H., Soung, N. K., Yang, D. C. (2007) *Solimonas soli* gen. nov., sp. nov., isolated from soil of a ginseng field. International Journal of Systematic and Evolutionary Microbiology, 57: 2591-2594.

Klümper, U., Riber, L., Dechesne, A., Sannazzarro, A., Hansen, L. H., Sørensen, S. J., Smets, B. F. (2015) Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. The ISME Journal, 9: 934-945.

Koeller, K. M., Wong, C. H. (2001) Enzymes for chemical synthesis. Nature, 409: 232-240.

Koo, C. W., Rosenzweig, A. C. (2021) Biochemistry of aerobic biological methane oxidation. Chemical Society Reviews, 50: 3424-3436.

Koo, C. W., Tucci, F. J., He, Y., Rosenzweig, A. C. (2022) Recovery of particulate methane monooxygenase structure and activity in a lipid bilayer. Science, 375: 1287-1291.

Kotani, T., Kawashima, Y., Yurimoto, H., Kato, N., Sakai, Y. (2006) Gene structure and regulation of alkane monooxygenases in propane-utilizing *Mycobacterium* sp. TY-6 and *Pseudonocardia* sp. TY-7. Journal of Bioscience and Bioengineering, 102: 184-192.

Kotani, T., Yamamoto, T., Yurimoto, H., Sakai, Y., Kato, N. (2003) Propane monooxygenase and NAD+-dependent secondary alcohol dehydrogenase in propane metabolism by *Gordonia* sp. strain TY-5. Journal of Bacteriology, 185: 7120-7128.

Kottegoda, S., Waligora, E., Hyman, M. (2015) Metabolism of 2-methylpropene (isobutylene) by the aerobic bacterium *Mycobacterium* sp. strain ELW1. Applied and Environmental Microbiology, 81: 1966-1976.

Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., Peterson, K. M. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene, 166: 175-176.

Kukor, J. J., Olsen, R. H. (1990) Molecular cloning, characterization, and regulation of a *Pseudomonas pickettii* PKO1 gene encoding phenol hydroxylase and expression of the gene in *Pseudomonas aeruginosa* PAO1c. Journal of Bacteriology, 172: 4624-4630.

Kurth, E. G., Doughty, D. M., Bottomley, P. J., Arp, D. J., Sayavedra-Soto, L. A. (2008) Involvement of BmoR and BmoG in *n*-alkane metabolism in '*Pseudomonas butanovora*'. Microbiology, 154: 139-147.

Le, N. B., Coleman, N. V. (2011) Biodegradation of vinyl chloride, *cis*-dichloroethene and 1,2-dichloroethane in the alkene/alkane-oxidising *Mycobacterium* strain NBB4. Biodegradation, 22: 1095-1108.

Leahy, J. G., Batchelor, P. J., Morcomb, S. M. (2003) Evolution of the soluble diiron monooxygenases. FEMS Microbiology Reviews, 27: 449-479.

Leak, D. J., Sheldon, R. A., Woodley, J. M., Adlercreutz, P. (2009) Biocatalysts for selective introduction of oxygen. Biocatalysis and Biotransformation, 27: 1-26.

Lee, K. (1999) Benzene-induced uncoupling of naphthalene dioxygenase activity and enzyme inactivation by production of hydrogen peroxide. Journal of Bacteriology, 181: 2719-2725.

Lefort, V., Longueville, J. E., Gascuel, O. (2017) SMS: Smart Model Selection in PhyML. Molecular Biology and Evolution, 34: 2422-2424.

Li, F., Deng, D., Zeng, L., Abrams, S., Li, M. (2021) Sequential anaerobic and aerobic bioaugmentation for commingled groundwater contamination of trichloroethene and 1,4-dioxane. Science of the Total Environment, 774: 145118.

Li, X. T., Thomason, L. C., Sawitzke, J. A., Costantino, N., Court, D. L. (2013) Positive and negative selection using the *tetA-sacB* cassette: recombineering and P1 transduction in *Escherichia coli*. Nucleic Acids Research, 41: e204.

Liang, J. L., Gao, Y., He, Z., Nie, Y., Wang, M., JiangYang, J. H., Zhang, X. C., Shu, W. S., Wu, X. L. (2017) Crystal structure of TetR family repressor AlkX from *Dietzia* sp. strain DQ12-45-1b implicated in biodegradation of *n*-alkanes. Applied and Environmental Microbiology, 83: e01447-17.

Lloyd, J. S., De Marco, P., Dalton, H., Murrell, J. C. (1999) Heterologous expression of soluble methane monooxygenase genes in methanotrophs containing only particulate methane monooxygenase. Archives of Microbiology, 171: 364–370.

Lu, X., Taylor, A. E., Myrold, D. D., Neufeld, J. D. (2020) Expanding perspectives of soil nitrification to include ammonia‐oxidizing archaea and comammox bacteria. Soil Science Society of America Journal, 84: 287-302.

Ma, F., Wang, Y., Yang, J., Guo, H., Su, D., Yu, L. (2021) Degradation of 1,4-dioxane by *Xanthobacter* sp. YN2. Current Microbiology, 78: 992-1005.

Maddocks, S. E., Oyston, P. C. F. (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. Microbiology, 154: 3609-3623.

Mahendra, S., Alvarez-Cohen, L. (2005) *Pseudonocardia dioxanivorans* sp. nov., a novel actinomycete that grows on 1,4-dioxane. International Journal of Systematic and Evolutionary Microbiology, 55: 593-598.

Martin, K. E., Ozsvar, J., Coleman, N. V. (2014) SmoXYB1C1Z of *Mycobacterium* sp. strain NBB4: a soluble methane monooxygenase (sMMO)-like enzyme, active on C2 to C4 alkanes and alkenes. Applied and Environmental Microbiology, 80: 5801-5806.

Masuda, H., McClay, K., Steffan, R. J., Zylstra, G. J. (2012a) Biodegradation of tetrahydrofuran and 1,4-dioxane by soluble diiron monooxygenase in *Pseudonocardia* sp. strain ENV478. Journal of Molecular Microbiology and Biotechnology, 22: 312-316.

Masuda, H., McClay, K., Steffan, R. J., Zylstra, G. J. (2012b) Characterization of three propane-inducible oxygenases in *Mycobacterium* sp. strain ENV421. Letters in Applied Microbiology, 55: 175-181.

Mattes, T. E., Alexander, A. K., Coleman, N. V. (2010) Aerobic biodegradation of the chloroethenes: Pathways, enzymes, ecology, and evolution. FEMS Microbiology Reviews, 34: 445-475.

Mattes, T. E., Coleman, N. V., Spain, J. C., Gossett, J. M. (2005) Physiological and molecular genetic analyses of vinyl chloride and ethene biodegradation in *Nocardioides* sp. strain JS614. Archives of Microbiology, 183: 95-106.

Maughan, H., Galeano, B., Nicholson, W. L. (2004) Novel *rpoB* mutations conferring rifampin resistance on *Bacillus subtilis*: Global effects on growth, competence, sporulation, and germination. Journal of Bacteriology, 186: 2481-2486.

McCarl, V., Somerville, M. V., Ly, M. A., Henry, R., Liew, E. F., Wilson, N. L., Holmes, A. J., Coleman, N. V. (2018) Heterologous expression of *Mycobacterium* alkene monooxygenases in Gram-positive and Gram-negative bacterial hosts. Applied and Environmental Microbiology, 84: e00397-18.

McCormick, M. S., Lippard, S. J. (2011) Analysis of substrate access to active sites in bacterial multicomponent monooxygenase hydroxylases: X-ray crystal structure of xenon-pressurized phenol hydroxylase from *Pseudomonas* sp. OX1. Biochemistry, 50: 11058-11069.

Meehan, C. J., Barco, R. A., Loh, Y. E., Cogneau, S., Rigouts, L. (2021) Reconstituting the genus *Mycobacterium*. International Journal of Systematic and Evolutionary Microbiology, 71: 004922.

Miao, Y., Zhou, T., Zheng, X., Mahendra, S. (2023) Investigating biodegradation of 1,4-dioxane by groundwater and soil microbiomes: Insights into microbial ecology and process prediction. ACS ES&T Water, 4: 1046-1060.

Miuran, A., Dalton, H. (1995) Purification and characterization of the alkene monooxygenase from *Nocardia corallina* B-276. Bioscience, Biotechnology, and Biochemistry, 59: 853-859.

Moratti, C. F., Scott, C., Coleman, N. V. (2022) Synthetic biology approaches to hydrocarbon biosensors: a review. Frontiers in Bioengineering and Biotechnololgy, 9: 804234.

Moratti, C. F., Yang, S. N. N., Scott, C., Coleman, N. V. (2024) Development of a whole-cell biosensor for ethylene oxide and ethylene. Microbial Biotechnology, 17: e14511.

Morrill, T. C., Friedrich, L. E., Machonkin, M. A., Whitbourne, J. E., Eastman, C. A. (1981) *N*-(2-hydroxyethyl)-4-(*p*-nitrobenzylidene)-1,4-dihydropyridine (DHP) from the reaction of 4-(*p*-nitrobenzyl)pyridine (NBP) with ethylene oxide. Journal of Heterocyclic Chemistry, 18: 1645-1647.

Msomi, N. N., Padayachee, T., Nzuza, N., Syed, P. R., Krys, J. D., Chen, W., Gront, D., Nelson, D. R., Syed, K. (2021) *In silico* analysis of P450s and their role in secondary metabolism in the bacterial class Gammaproteobacteria. Molecules, 26: 1538.

Murphy, C. L., Sheremet, A., Dunfield, P. F., Spear, J. R., Stepanauskas, R., Woyke, T., Elshahed, M. S., Youssef, N. H. (2021) Genomic analysis of the yet-uncultured Binatota reveals broad methylotrophic, alkane-degradation, and pigment production capacities. mBio, 12: e00985-21.

NCBI (2025, Last updated 18th January, 2025). "PubChem Compound summary for CID 7834, 1,2-epoxybutane." Retrieved on 22nd January, 2025 from https://pubchem.ncbi.nlm.nih.gov/compound/1_2-Epoxybutane.

Nelson, D. R. (2018) Cytochrome P450 diversity in the tree of life. Biochimica et Biophysica Acta - Proteins and Proteomics, 1866: 141-154.

Nguyen, N. T. (2018) *Acinetobacter soli* SP2 capable of high-efficiency degradation of food emulsifier Polysorbate 80. Current Microbiology, 75: 896-900.

Nicholas, K. B., Nicholas, H. B. J. (1997). GeneDoc: A tool for editing and annotating multiple sequence alignments. Distributed by the authors.

Nielsen, A. K., Gerdes, K., Murrell, J. C. (1997) Copper-dependent reciprocal transcriptional regulation of methane monooxygenase genes in *Methylococcus capsulatus* and *Methylosinus trichosporium*. Molecular Microbiology, 25: 399-409.

Norlund, I., Powlowski, J., Shingler, V. (1990) Complete nucleotide sequence and polypeptide analysis of multicomponent phenol hydroxylase from *Pseudomonas* sp. strain CF600. Journal of Bacteriology, 172: 6823-6833.

Notomista, E., Lahm, A., Di Donato, A., Tramontano, A. (2003) Evolution of bacterial and archaeal multicomponent monooxygenases. Journal of Molecular Evolution, 56: 435-445.

Nowak, A., Mrozik, A. (2016) Facilitation of co-metabolic transformation and degradation of monochlorophenols by *Pseudomonas* sp. CF600 and changes in its fatty acid composition. Water, Air, & Soil Pollution, 227: 83.

O'Neil, C. S., Beach, J. L., Gruber, T. D. (2019) DNA electrophoresis using thiazole orange instead of ethidium bromide or alternative dyes. Journal of Visualized Experiments, 145: 10.3791/59341.

Ojo, A. O., Castillo, J., Cason, E. D., Valverde, A. (2023) Biodegradation of chloroethene compounds under microoxic conditions. Biotechnology and Bioengineering, 121: 1036-1049.

Oren, A., Garrity, G. M. (2021) Valid publication of the names of forty-two phyla of prokaryotes. International Journal of Systematic and Evolutionary Microbiology, 71: 005056.

Osborne, C. D., Haritos, V. S. (2018) Horizontal gene transfer of three co-inherited methane monooxygenase systems gave rise to methanotrophy in the Proteobacteria. Molecular Phylogenetics and Evolution, 129: 171-181.

Osborne, C. D., Haritos, V. S. (2019) Beneath the surface: Evolution of methane activity in the bacterial multicomponent monooxygenases. Molecular Phylogenetics and Evolution, 139: 106527.

Palleroni, N. J., Port, A. M., Chang, H. K., Zylstra, G. J. (2004) *Hydrocarboniphaga effusa* gen. nov., sp. nov., a novel member of the γ-Proteobacteria active in alkane and aromatic hydrocarbon degradation. International Journal of Systematic and Evolutionary Microbiology, 54: 1203-1207.

Pan, J., Wei, F., Liu, Y., Xu, Y., Ma, Y. (2022) Unravelling the role of GntR on the regulation of alkane hydroxylase AlkB2 in *Pseudomonas aeruginosa* DN1 based on transcriptome analysis. Journal of Applied Microbiology, 132: 2812-2822.

Parod, R. J. (2014). Tetrahydrofuran. In Encyclopedia of Toxicology. Wexler, P. (Ed.) Academic Press, Cambridge, MA, United States. **4:** 505-508.

Pesce, C., Oshone, R., Hurst IV, S. G., Kleiner, V. A., Tisaa, L. S. (2019) Stable transformation of the Actinobacteria *Frankia* spp. Applied and Environmental Microbiology, 85: e00957-19.

Petkevičius, V., Vaitekūnas, J., Vaitkus, D., Čėnas, N., Meškys, R. (2019) Tailoring a soluble diiron monooxygenase for synthesis of aromatic *N*-oxides. Catalysts, 9: 356.

Pietersen, R. D., du Preez, I., Loots, D. T., van Reenen, M., Beukes, D., Leisching, G., Baker, B. (2020) Tween 80 induces a carbon flux rerouting in *Mycobacterium tuberculosis*. Journal of Microbiological Methods, 170: 105795.

Pimviriyakul, P., Wongnate, T., Tinikul, R., Chaiyen, P. (2020) Microbial degradation of halogenated aromatics: Molecular mechanisms and enzymatic reactions. Microbial Biotechnology, 13: 67-86.

Que, L., Jr., Tolman, W. B. (2008) Biologically inspired oxidation catalysis. Nature, 455: 333-340.

Rappé, M. S., Giovannoni, S. J. (2003) The uncultured microbial majority. Annual Review of Microbiology, 57: 369-394.

Ren, H., Wang, H., Wang, Y., Chen, Y., Lu, Z. (2022) A multicomponent THF hydroxylase initiates tetrahydrofuran degradation in *Cupriavidus metallidurans* ZM02. Applied and Environmental Microbiology, 88: e01880-21.

Reyrat, J.-M., Pelicic, V., Gicquel, B., Rappuoli, R. (1998) Counterselectable markers: Untapped tools for bacterial genetics and pathogenesis. Infection and Immunity, 66: 4011–4017.

Richardson, A. J., McKain, N., Wallace, R. J. (2013) Ammonia production by human faecal bacteria, and the enumeration, isolation and characterization of bacteria capable of growth on peptides and amino acids. BMC Microbiology, 13: 6.

Riley, L. A., Guss, A. M. (2021) Approaches to genetic tool development for rapid domestication of non-model microorganisms. Biotechnology for Biofuels, 14: 30.

Rodriguez, J. G., Hernandez, A. C., Helguera-Repetto, C., Aguilar Ayala, D., Guadarrama-Medina, R., Anzola, J. M., Bustos, J. R., Zambrano, M. M., Gonzalez-Y-Merchand, J., Garcia, M. J., Del Portillo, P. (2014) Global adaptation to a lipid environment triggers the dormancy-related phenotype of *Mycobacterium tuberculosis*. mBio, 5: e01125-14.

Rosano, G. L., Ceccarelli, E. A. (2014) Recombinant protein expression in *Escherichia coli*: Advances and challenges. Frontiers in Microbiology, 5: 172.

Ryoo, D., Shim, H., Arenghi, F. L., Barbieri, P., Wood, T. K. (2001) Tetrachloroethylene, trichloroethylene, and chlorinated phenols induce toluene-*o*-xylene monooxygenase activity in *Pseudomonas stutzeri* OX1. Applied and Environmental Microbiology, 56: 545-549.

Saeki, H., Akira, M., Furuhashi, K., Averhoff, B., Gottschalk, G. (1999) Degradation of trichloroethene by a linear-plasmid-encoded alkene monooxygenase in *Rhodococcus corallinus* (*Nocardia corallina*) B-276. Microbiology, 145: 1721-1730.

Saeki, H., Furuhashi, K. (1994) Cloning and characterization of a *Nocardia corallina* B-276 gene cluster encoding alkene monooxygenase. Journal of Fermentation and Bioengineering, 78: 399-406.

Sakai, Y., Yurimoto, H., Shima, S. (2023) Methane monooxygenases; physiology, biochemistry and structure. Catalysis Science & Technology, 13: 6342-6354.

Sales, C. M., Grostern, A., Parales, J. V., Parales, R. E., Alvarez-Cohen, L. (2013) Oxidation of the cyclic ethers 1,4-dioxane and tetrahydrofuran by a monooxygenase in two *Pseudonocardia* species. Applied and Environmental Microbiology, 79: 7702-7708.

Sayavedra-Soto, L. A., Byrd, C. M., Arp, D. J. (2001) Induction of butane consumption in *Pseudomonas butanovora*. Archives of Microbiology, 176: 114-120.

Sazinsky, M. H., Lippard, S. J. (2005) Product bound structures of the soluble methane monooxygenase hydroxylase from *Methylococcus capsulatus* (Bath): protein motion in the α-subunit. Journal of the American Chemical Society, 127: 5814-5825.

Sazykin, I., Makarenko, M., Khmelevtsova, L., Seliverstova, E., Rakin, A., Sazykina, M. (2019) Cyclohexane, naphthalene, and diesel fuel increase oxidative stress, *CYP153*, *sodA*, and *recA* gene expression in *Rhodococcus erythropolis*. MicrobiologyOpen, 8: e00855.

Schmidt, S. K., Simkins, S., Alexander, M. (1985) Models for the kinetics of biodegradation of organic compounds not supporting growth. Applied and Environmental Microbiology, 50: 323-331.

Semprini, L. (1997) Strategies for the aerobic co-metabolism of chlorinated solvents. Current Opinion in Biotechnology, 8: 296-308.

Seo, J. S., Keum, Y. S., Li, Q. X. (2009) Bacterial degradation of aromatic compounds. International Journal of Environmental Research and Public Health, 6: 278-309.

Sevilla, E., Yuste, L., Rojo, F. (2015) Marine hydrocarbonoclastic bacteria as whole-cell biosensors for *n*-alkanes. Microbial Biotechnology, 8: 693-706.

Shennan, J. L. (2005) Utilisation of C2–C4 gaseous hydrocarbons and isoprene by microorganisms. Journal of Chemical Technology & Biotechnology, 81: 237-256.

Sheu, S. Y., Cho, N. T., Arun, A. B., Chen, W. M. (2011) Proposal of *Solimonas aquatica* sp. nov., reclassification of *Sinobacter flavus* Zhou *et al*. 2008 as *Solimonas flava* comb. nov. and *Singularimonas variicoloris* Friedrich and Lipski 2008 as S*olimonas variicoloris* comb. nov. and emended descriptions of the genus *Solimonas* and its type species *Solimonas soli*. International Journal of Systematic and Evolutionary Microbiology, 61: 2284-2291.

Shields, M. S., Reagin, M. J., Gerger, R. R., Campbell, R., Somerville, C. (1995) TOM, a new aromatic degradative plasmid from *Burkholderia* (*Pseudomonas*) *cepacia* G4. Applied and Environmental Microbiology, 61: 1352-1356.

Shim, H., Wood, T. K. (2000) Aerobic degradation of mixtures of chlorinated aliphatics by cloned toluene-*o*-xylene monooxygenase and toluene *o*-monooxygenase in resting cells. Biotechnology and Bioengineering, 70: 693-698.

Shingler, V., Franklin, F. C., Tsuda, M., Holroyd, D., Bagdasarian, M. (1989) Molecular analysis of a plasmid-encoded phenol hydroxylase from *Pseudomonas* CF600. Journal of General Microbiology, 135: 1083-1092.

Simon, R., Priefer, U., Pühler, A. (1983) A broad host range mobilization system for *in vivo* genetic engineering: Transposon mutagenesis in gram negative bacteria. Bio/Technology, 1: 784–791

Singhal, H., Ren, Y. R., Kern, S. E. (2010) Improved DNA electrophoresis in conditions favoring polyborates and lewis acid complexation. PLoS One, 5: e11318.

Skinner, J., Delgado, A. G., Hyman, M., Chu, M. J. (2024) Implementation of *in situ* aerobic cometabolism for groundwater treatment: State of the knowledge and important factors for field operation. Science of the Total Environment, 925: 171667.

Sluis, M. K., Sayavedra-Soto, L. A., Arp, D. J. (2002) Molecular analysis of the soluble butane monooxygenase from '*Pseudomonas butanovora* '. Microbiology, 148: 3617–3629.

Smalla, K., Jechalke, S., Top, E. M. (2015) Plasmid detection, characterization, and ecology. Microbiology Spectrum, 3: PLAS-0038-2014.

Smith, T. J., Lloyd, J. S., Gallagher, S. C., Fosdike, W. L. J., Murrell, J. C., Dalton, H. (1999) Heterologous expression of alkene monooxygenase from *Rhodococcus rhodochrous* B-276. European Journal of Biochemistry, 260: 446-452.

Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T., Jacobs, W. R., Jr. (1990) Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. Molecular Biology, 4: 1911-1919.

Spain, J. C., Nishino, S. F. (1987) Degradation of 1,4-dichlorobenzene by a *Pseudomonas* sp. Applied and Environmental Microbiology, 53: 1010-1019.

Stainthorpe, A. C., Lees, V., Salmond, G. P. C., Dalton, H., Murrell, J. C. (1990) The methane monooxygenase gene cluster of *Methylococcus capsulatus* (Bath). Gene, 91: 27-34.

Sticher, P., Jaspers, M. C., Stemmler, K., Harms, H., Zehnder, A. J., van der Meer, J. R. (1997) Development and characterization of a whole-cell bioluminescent sensor for bioavailable middle-chain alkanes in contaminated groundwater samples. Applied and Environmental Microbiology, 63: 4053–4060.

Sukul, P., Schakermann, S., Bandow, J. E., Kusnezowa, A., Nowrousian, M., Leichert, L. I. (2017) Simple discovery of bacterial biocatalysts from environmental samples through functional metaproteomics. Microbiome, 5: 28.

Sullivan, J. P., Dickinson, D., Chase, H. A. (1998) Methanotrophs, *Methylosinus trichosporium* OB3b, sMMO, and their application to bioremediation. Critical Reviews in Microbiology, 24: 335-373.

Suttinun, O., Muller, R., Luepromchai, E. (2009) Trichloroethylene cometabolic degradation by *Rhodococcus* sp. L4 induced with plant essential oils. Biodegradation, 20: 281-291.

Suzuki, T., Yazawa, T., Morishita, N., Maruyama, A., Fuse, H. (2019) Genetic and physiological characteristics of a novel marine propylene-assimilating *Halieaceae* bacterium isolated from seawater and the diversity of its alkene and epoxide metabolism genes. Microbes and Environment, 34: 33-42.

Tamura, K., Stecher, G., Kumar, S. (2021) MEGA11: Molecular Evolutionary Genetics Analysis version 11. Molecular Biololgy and Evolution, 38: 3022-3027.

Tan, H.-M. (1999) Bacterial catabolic transposons. Applied Microbiology and Biotechnology, 51: 1-12.

Tao, Y., Fishman, A., Bentley, W. E., Wood, T. K. (2004) Oxidation of benzene to phenol, catechol, and 1,2,3-trihydroxybenzene by toluene 4-monooxygenase of *Pseudomonas mendocina* KR1 and toluene 3-monooxygenase of *Ralstonia pickettii* PKO1. Applied and Environmental Microbiology, 70: 3814-3820.

Terwilliger, T. C., Liebschner, D., Croll, T. I., Williams, C. J., McCoy, A. J., Poon, B. K., Afonine, P. V., Oeffner, R. D., Richardson, J. S., Read, R. J., Adams, P. D. (2024) AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. Nature Methods, 21: 110-116.

Thiemer, B., Andreesen, J. R., Schrader, T. (2003) Cloning and characterization of a gene cluster involved in tetrahydrofuran degradation in *Pseudonocardia* sp. strain K1. Archives of Microbiology, 179: 266-277.

Top, E. M., Springael, D. (2003) The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. Current Opinion in Biotechnology, 14: 262-269.

Top, E. M., Springael, D., Boon, N. (2002) Catabolic mobile genetic elements and their potential use in bioaugmentation of polluted soils and waters. FEMS Microbiology Ecology, 42: 199-208.

Torres Pazmiño, D. E., Winkler, M., Glieder, A., Fraaije, M. W. (2010) Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications. Journal of Biotechnology, 146: 9-24.

Tucci, F. J., Jodts, R. J., Hoffman, B. M., Rosenzweig, A. C. (2023) Product analogue binding identifies the copper active site of particulate methane monooxygenase. Nature Catalysis, 6: 1194-1204.

Tucci, F. J., Rosenzweig, A. C. (2024) Direct methane oxidation by copper- and iron-dependent methane monooxygenases. Chemical Reviews, 124: 1288–1320.

van Beilen, J. B., Marin, M. M., Smits, T. H., Rothlisberger, M., Franchini, A. G., Witholt, B., Rojo, F. (2004) Characterization of two alkane hydroxylase genes from the marine hydrocarbonoclastic bacterium *Alcanivorax borkumensis*. Environmental Microbiology, 6: 264-273.

van Berkel, W. J., Kamerbeek, N. M., Fraaije, M. W. (2006) Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. Journal of Biotechnology, 124: 670-689.

Van Bogaert, I. N., Groeneboer, S., Saerens, K., Soetaert, W. (2011) The role of cytochrome P450 monooxygenases in microbial fatty acid metabolism. The FEBS Journal, 278: 206-221.

Van Hylckama, J. T., De Koning, W., Janssen, D. B. (1996) Transformation Kinetics of Chlorinated Ethenes by *Methylosinus trichosporium* OB3b and Detection of Unstable Epoxides by On-Line Gas Chromatography. Applied and Environmental Microbiology, 62: 3304-3312.

van Hylckama Vlieg, J. E. T., Janssen, D. B. (2001) Formation and detoxification of reactive intermediates in the metabolism of chlorinated ethenes. Journal of Biotechnology, 85: 81-102.

van Kessel, J. C., Hatfull, G. F. (2007) Recombineering in *Mycobacterium tuberculosis*. Nature Methods, 4: 147-152.

Vardar, G., Wood, T. K. (2004) Protein engineering of toluene-*o*-xylene monooxygenase from *Pseudomonas stutzeri* OX1 for synthesizing 4-methylresorcinol, methylhydroquinone, and pyrogallol. Applied and Environmental Microbiology, 70: 3253-3262.

Vardar, G., Wood, T. K. (2005) Protein engineering of toluene-*o*-xylene monooxygenase from *Pseudomonas stutzeri* OX1 for enhanced chlorinated ethene degradation and *o*-xylene oxidation. Applied Genetics and Molecular Biotechnology, 68: 510-517.

Wackett, L. P., Hershberger, C. D. (2001). General concepts in biodegradation and biocatalysis. In Biocatalysis and Biodegradation: Microbial Transformation of Organic Compounds. Wackett, L. P. & Hershberger, C. D. (Ed.) ASM Press, Washington, DC, United States. **1:** 1-5.

Wang, P., Cui, T., Li, J., Liu, N., Hong, M. (2021a) 1,4-dioxane degradation performance of tetrahydrofuran-grown *Arthrobacter* sp. WN18. Water, Air, & Soil Pollution, 232: 473.

Wang, P., Li, F., Wang, W., Wang, R., Yang, Y., Cui, T., Liu, N., Li, M. (2021b) Cometabolic degradation of 1,4-dioxane by a tetrahydrofuran-growing *Arthrobacter* sp. WN18. Ecotoxicology and Environmental Safety, 217: 112206.

Wang, W., Iacob, R. E., Luoh, R. P., Engen, J. R., Lippard, S. J. (2014) Electron transfer control in soluble methane monooxygenase. Journal of the American Chemical Society, 136: 9754-9762.

Woodland, M. P., Matthews, C. S., Leak, D. J. (1995) Properties of a soluble propene monooxygenase from *Mycobacterium* sp. (strain M156). Archives of Microbiology, 163: 231-234.

Xu, T., Close, D. M., Sayler, G. S., Ripp, S. (2013) Genetically modified whole-cell bioreporters for environmental assessment. Ecological Indicators, 28: 125-141.

Yen, K. M., Karl, M. R., Blatt, L. M., Simon, M. J., Winter, R. B., Fausset, P. R., Lu, H. S., Harcourt, A. A., Chen, K. K. (1991) Cloning and characterization of a *Pseudomonas mendocina* KR1 gene cluster encoding toluene-4-monooxygenase. Journal of Bacteriology, 173: 5315-5327.

Zhang, D., He, Y., Wang, Y., Wang, H., Wu, L., Aries, E., Huang, W. E. (2012) Whole-cell bacterial bioreporter for actively searching and sensing of alkanes and oil spills. Microbial Biotechnology, 5: 87-97.

Zhao, G., Xia, T., Song, J., Jensen, R. A. (1994) *Pseudomonas aeruginosa* possesses homologues of mammalian phenylalanine hydroxylase and 4 alpha-carbinolamine dehydratase/DCoH as part of a

three-component gene cluster. Proceedings of the National Academy of Sciences of the United States of America, 91: 1366-1370.

Zhou, N.-Y., Chan Kwo Chion, C. K., Leak, D. J. (1996) Cloning and expression of the genes encoding the propene monooxygenase from *Xanthobacter* Py2. Applied and Environmental Microbiology, 44: 582-588.

Zhou, N.-Y., Jenkins, A., Chan Kwo Chion, C. K., Leak, D. J. (1998) The alkene monooxygenase from *Xanthobacter* Py2 is a binuclear non-haem iron protein closely related to toluene 4-monooxygenase. FEMS Letters, 430: 181-185.

Zhou, W., Guo, W., Zhou, H., Chen, X. (2016) Phenol degradation by *Sulfobacillus acidophilus* TPY via the *meta*-pathway. Microbiological Research, 190: 37-45.

Zill, D., Lettau, E., Lorent, C., Seifert, F., Singh, P. K., Lauterbach, L. (2022) Crucial role of the chaperonin GroES/EL for heterologous production of the soluble methane monooxygenase from *Methylomonas methanica* MC09. ChemBioChem, e202200195.

Zou, B., Huang, Y., Zhang, P. P., Ding, X. M., Op den Camp, H. J. M., Quan, Z. X. (2021) Horizontal gene transfer of genes encoding copper-containing membrane-bound monooxygenase (CuMMO) and soluble di-iron monooxygenase (SDIMO) in ethane- and propane-oxidizing *Rhodococcus* bacteria. Applied and Environmental Microbiology, 87: e0022721.

Supplementary Tables

Table S1. Accession numbers and host bacteria of SDIMOs used for phylogenetic analysis in Chapter 1.

Accession number	SDIMO Group	Organism
WP_070156811.1	1	*Sphingobium phenoxybenzoativorans* SC 3
RNL57703.1		*Zhongshania marina* DSW25-10
WP_051482134.1		*Paraburkholderia nodosa* DSM 21604
WP_033292595.1		*Amycolatopsis jejuensis* NRRL B-24427
WP_118605632.1		*Rhodococcus* sp. WY5
WP_109923380.1		*Zavarzinia compransoris* DSM 1231
AEH09306.1		Candidatus Frankia datiscae Dg1
MBL6751253.1		*Nevskia* sp.
WP_103383421.1		*Pseudonocardia dioxanivorans* BERK-1
MBV9794320.1		Actinobacteria sp.
PCJ61408.1		Rhodospirillaceae sp. (2)
MBI05542.1		Rhodospirillaceae sp. (1)
WP_120038486.1		*Bacillus* sp. PK3 68
WP_012642669.1		*Thermomicrobium roseum* DSM 5159
MBY0279729.1		Candidatus Binatia sp.
WP_073632860.1		*Scytonema* sp. HK-05
WP_184522947.1		*Bacillus benzoevorans* DSM 5391
RZO17141.1		Candidatus Thioglobus sp.
WP_213445351.1		*Mycobacterium* sp. SM1
MBT9160403.1		Dehalococcoidia sp.
GEL24807.1		*Pseudonocardia sulfidoxydans* NBRC 16205
WP_013947866.1		*Hyphomicrobium* sp. MC1
TMA11358.1		Deltaproteobacteria sp. (2)
WP_172627877.1		*Bradyrhizobium ivorense* CI-1B
WP_127826834.1		*Streptomyces* sp. San01
WP_121570089.1		*Halobellus* sp. Atlit-38R
WP_169380166.1		*Pseudonocardia* sp. K10HN5
WP_133882394.1		*Panacagrimonas perspica* DSM 26377
PYX42310.1		Acidobacteria sp. (2)
NUP06582.1		Polyangiaceae sp.
PYV37852.1		Acidobacteria sp. (1)
WP_146645982.1		*Labilithrix luteola* DSM 27648
MBX3668327.1		Rhodocyclaceae sp.
NUQ34196.1		Planctomycetaceae sp.
MBI4816915.1		Deltaproteobacteria sp. (1)
O87082.1		*Xanthobacter autotrophicus* Py2
WP_203195512.1		*Xanthobacter* sp. YN2
CAA06654.1		*Pseudomonas stutzeri* OX1
WP_176131917.1		*Burkholderia cepacia* G4

WP_141018170.1		*Azoarcus* sp. DD4
WP_182557825.1		*Ralstonia pickettii* PKO1
Q00456.3		*Pseudomonas mendocina* KR1
AJO68015.1		*Pseudomonas* sp. M4
BAA11761.1		*Pseudomonas aeruginosa* JI104
WP_095413456.1		*Rhodococcus* sp. AD45
WP_012041847.1		*Bradyrhizobium* sp. BTAi1
WP_278219883.1		*Mycobacterium gadium* IBE100
WP_156296907.1		*Mycobacterium paragordonae* IBE200
WP_153041231.1	2	*Paraburkholderia monticola* JC2948
MBS1188323.1		Rhodocyclaceae sp.
KJS45504.1		Rhodospirillaceae sp. BRH c57
WP_060217674.1		*Burkholderia cepacia* MSMB1063WGS
WP_166995784.1		*Pseudomaricurvus alkylphenolicus* KCTC 32386
WP_046462198.1		*Comamonas thiooxydans* KY3
WP_126536818.1		*Aerosticca soli* Dysh456
PCJ47005.1		Gammaproteobacteria sp. (2)
WP_218036763.1		*Sphingobium* sp. EM0848
WP_141839014.1		*Herbaspirillum* sp. SJZ107
WP_034998950.1		*Beijerinckia mobilis* UQM 1969
WP_099755496.1		*Acinetobacter baumannii* MBL M5
MBK7462717.1		Betaproteobacteria sp. (1)
TDJ60115.1		Proteobacteria sp.
MSQ68644.1		Gammaproteobacteria sp. (1)
MAV30580.1		*Cycloclasticus* sp.
KXS54906.1		*Marinobacter* sp. T13-3
RTL47487.1		Rhodocyclaceae sp.
WP_173763396.1		*Azoarcus* sp. M9-3-2
MBP7521906.1		*Leptothrix* sp.
WP_197946093.1		*Phytohabitans suffuscus* NBRC 105367
WP_022976858.1		*Nevskia ramosa* DSM 11499
WP_136385674.1		*Azoarcus rhizosphaerae* CC-YHH848
WP_072211031.1		Candidatus Kryptobacter tengchongensis GI-25
MBP8866971.1		*Propionivibrio* sp.
WP_092607966.1		*Raineyella antarctica* LZ-22 (1)
PCI69579.1		Piscirickettsiaceae sp.
WP_072789905.1		*Ferrithrix thermotolerans* DSM 19514
WP_129082673.1		*Arcobacter anaerophilus* DSM 24636
HAN35823.1		Acidimicrobiaceae sp.
WP_140043082.1		*Sphingobium fuliginis* ATCC 27551
SUA44835.1		*Nocardia africana* NCTC13184
WP_092606539.1		*Raineyella antarctica* LZ-22 (2)
TMK50329.1		Actinobacteria sp.
MBM3364274.1		Betaproteobacteria sp. (2)

WP_096466037.1		*Aneurinibacillus soli* CB4
NBW54534.1		Betaproteobacteria sp. (3)
WP_134703373.1		*Ammoniphilus* sp. YIM 78166
MSQ65499.1		Limnohabitans sp.
WP_149356725.1		*Comamonas testosteroni* TA441
BAA87871.1		*Comamonas testosteroni* R5
2INP_A		*Pseudomonas stutzeri* OX1
AAC32455.1		*Ralstonia* sp. E2
AFQ51794.1		*Burkholderia cepacia* G4
AAM77905.1		*Acinetobacter radioresistens* sp. S13
ACM66934.1		*Arthrobacter* sp. W1
BAA06017.1		*Pseudomonas putida* BH
AAA25942.1		*Pseudomonas putida* CF600
AEJ38828.1		*Sulfobacillus acidophilus* TPY
ABQ14522.1		*Alcaligenes faecalis* IS-46
AAA88459.1		*Pseudomonas* sp. JS150
WP_011516082.1		*Cupriavidus metallidurans* ZM02
WP_221064835.1	3	*Methylomagnum ishizawai* RS11D-Pr
WP_153249048.1		Candidatus Methylospira mobilis Shm1
WP_119632605.1		*Methylocaldum marinum* S8
WP_196434761.1		*Methylomonas* sp. LL1
TAN65936.1		*Methylobacter* sp.
NOQ14620.1		*Methyloprofundus* sp.
WP_033157495.1		*Methylomonas* sp. LW13
NMM29078.1		*Glaciimonas* sp.
WP_087143657.1		*Crenothrix polyspora*
NOT67204.1		Methylophilaceae sp.
WP_069436906.1		*Methyloceanibacter methanicus* R-67174
WP_020174569.1		*Methyloferula stellata* AR4
WP_165048247.1		*Methylocystis* sp. MJC1
WP_134488637.1		*Methylocella tundrae* (2)
WP_044435962.1		*Skermanella aerolata* KACC 11604
PKO92487.1		Betaproteobacteria sp.
WP_085772150.1		*Methylocystis bryophila* S285
WP_142863786.1		*Methylosinus sporium* 5
WP_104955546.1		*Sphingobium* sp. SCG-1
VTZ27606.1		*Methylocella tundrae* (1)
WP_217999562.1		*Thauera butanivorans* NBRC 103042
QDF96990.1		*Azoarcus* sp. DD4
AAR98534.1		*Brachymonas petroleovorans*
WP_093285538.1		*Solimonas aquatica* DSM 25927
MCA1845033.1		Actinobacteria sp.
MBN4016286.1		Rhodospirillaceae sp. AH-315-P19
WP_110318564.1		*Mycobacterium moriokaense* GAS496

WP_119782044.1		*Oleomonas* sp. K1W22B-8
WP_123029559.1		*Mycobacterium* sp. CECT 8783
WP_219069719.1		*Mycobacterium* sp. MAG1
WP_208964386.1		*Rhodococcus* sp. ZPP
WP_187099619.1		*Mycobacterium fluoranthenivorans* 2A
WP_102810290.1		*Mycobacterium* sp. ENV421
WP_014211362.1		*Mycobacterium rhodesiae* NBB3
ODQ85219.1		*Mycobacterium holsaticum* M7
MBE1422054.1		*Mycobacterium* sp. OAE906
1FYZ_A		*Methylococcus capsulatus* (Bath)
ACZ56334.1		*Mycobacterium chubuense* NBB4
7M8Q_A		*Methylosinus trichosporium* OB3b
WP_012590301.1		*Methylocella silvestris* BL2
BAJ17645.1		*Methylovulum miyakonense* HT12
WP_064006727.1		*Methylomonas methanica* R-45363
WP_123175023.1		*Methylocystis hirsuta* CSC1
AAC45289.1		*Methylocystis* sp. M
AAF01268.1		*Methylocystis* sp. WI_14
ACZ56346.1	4	*Mycobacterium chubuense* NBB4 (EtnC)
WP_005148464.1		*Mycobacterium rhodesiae* JS60
WP_179475495.1		*Mycobacterium* sp. CECT 8761
WP_014211290.1		*Mycobacterium rhodesiae* NBB3
WP_015305852.1		*Mycobacterium* sp. JS623
WP_006247675.1		*Mycobacterium tusciae* JS617
WP_123224789.1		*Marmoricola pocheonensis* Gsoil 818
WP_011751516.1		*Nocardioides* sp. JS614
AAS19484.1		*Mycobacterium* sp. M156
WP_132191191.1		*Kribbella* sp. VKM Ac-2500
QOE77414.1		*Rhodococcus ruber* B276
HCU95584.1		Actinobacteria sp.
WP_066884131.1		*Carbonactinospora thermoautotrophica* UBT1
WP_014805422.1		*Mycobacterium chubuense* NBB4 (PmoC)
WP_040789699.1		*Nocardia paucivorans* NBRC 100373
WP_147115713.1		*Pseudonocardia sulfidoxydans* NBRC 16205
WP_163747266.1		*Mycobacterium helvum* JCM 30396
WP_085136508.1		*Mycobacterium hiberniae* JCM 13571
WP_085127614.1		*Mycobacterium engbaekii* ATCC 27353
WP_064996631.1		*Mycobacterium heraklionense* E2548
BBD49902.1		*Halieaceae* sp. PE-TB08W
WP_076048437.1		*Mycobacterium heraklionense* NS-7503
WP_003893346.1	5	*MycobacteriuM. smegmatis* mc2-155
WP_057295027.1		*Nocardioides* sp. Soil796
MST31772.1		*Acidiferrimicrobium australe* USS-CCA1
WP_162113186.1		*Mycobacterium kumamotonensis* CST 7247

MSW50896.1		Actinobacteria sp. (5)
GIF69211.1		*Asanoa ishikariensis* NBRC 14551
OLB40648.1		*Ktedonobacter* sp. 13 2 20CM 53 11
MBA2443197.1		*Rubrobacter* sp. (2)
MBA4115726.1		*Rubrobacter* sp. (1)
WP_188649144.1		*Marinithermofilum abyssi* CGMCC 1.15179
WP_088922033.1		*Granulosicoccus antarcticus* IMCC3135
NNL84659.1		*Myxococcales* sp.
CUU15636.1		*Bradyrhizobium* sp. G22
MBV8416305.1		*Verrucomicrobia* sp.
WP_125999137.1		*Sphingobium yanoikuyae* SK-NIH.Env6 1116
TIX66714.1		*Mesorhizobium* sp.
WP_218917988.1		*Kyrpidia tusciae* DSM 2912
TMM16568.1		Actinobacteria sp. (4)
MSW51085.1		Actinobacteria sp. (1)
WP_154734104.1		*Conexibacter* sp. W3-3-2
MBO0768776.1		Actinobacteria sp. (1)
MBV9313307.1		*Pseudonocardia* sp.
TMM00089.1		Actinobacteria sp. (2)
TML97272.1		Actinobacteria sp. (3)
MBA2349234.1		Solirubrobacterales sp. (2)
WP_007576508.1		*Patulibacter medicamentivorans* I11
BAF34308.1		*Pseudonocardia* sp. TY-7 (Prm2A)
WP_192582011.1		*Rhodococcus wratislaviensis*
QCY50162.1		uncultured *Pseudonocardia* sp.
WP_006370700.1		*Gordonia* sp. TY-5
BAF34304.1		*Pseudonocardia* sp. TY-7 (Prm1A)
AEI99544.1		*Pseudonocardia* sp. ENV478
CAC10506.1		*Pseudonocardia tetrahydrofuranoxydans* K1
AEA22892.1		*Pseudonocardia dioxanivorans* CB1190
WP_011593714.1		*Rhodococcus jostii* RHA1
QPK32437.1		*Arthrobacter* sp. WN18
WP_014805366.1	6	*Mycobacterium chubuense* NBB4
WP_187099575.1		*Mycobacterium fluoranthenivorans* 2A
WP_136246236.1		*Mycobacterium intracellulare* CSURP8077
WP_012392133.1		*Mycobacterium marinum* M
RUP06991.1		*Mycobacterium* sp.
WP_163888788.1		*Mycobacterium hippocampi* JCM 30996
WP_149383891.1		*Mycobacterium* sp. ELW1
WP_278219893.1		*Mycobacterium gadium* IBE100
WP_156296942.1		*Mycobacterium paragordonae* IBE200
MBI2964566.1		Deltaproteobacteria sp.
MCA1842167.1		Actinobacteria sp.
MBF6567560.1		Candidatus Binataceae sp.

TMD95229.1		*Chloroflexi* sp.
WP_090609799.1		*Mycobacterium lentiflavum* CSUR P1491
MBI2803265.1		Gammaproteobacteria sp. (1)
MBI2802102.1		Gammaproteobacteria sp. (3)
MBK6659879.1		Proteobacteria sp. (2)
WP_133879578.1		*Panacagrimonas perspica* DSM 26377
MBK6660972.1		Proteobacteria sp. (1)
MBX9609471.1		Gammaproteobacteria sp. (2)
BAF34294.1		*Mycobacterium* sp. TY-6
WP_102810306.1		*Mycobacterium* sp. ENV421 (PrmA2)
WP_102810202.1		*Mycobacterium* sp. ENV421 (PrmA1)
WP_087083743.1		*Mycobacterium dioxanotrophicus* PH-06
WP_051362144.1	7	*Solimonas soli* DCY12
MBL6751106.1		*Nevskia* sp.
MBI2800181.1		Gammaproteobacteria sp. (1)
NNL85854.1		*Myxococcales* sp.
MSQ68733.1		Gammaproteobacteria sp. (2)
MBX9605089.1		Gammaproteobacteria sp. (3)
MBQ0758084.1		*Zhongshania* sp.
WP_103684344.1		*Marortus luteolus* ZX-21
WP_159239404.1		*Zhongshania aliphaticivorans* BC9 2A
MBU1831811.1		Gammaproteobacteria sp. (4)
WP_168451796.1		*Spongiibacter* sp. KMU-166
MAZ89983.1		Cellvibrionaceae sp.
WP_167182103.1		*Aestuariicella hydrocarbonica* JCM 30134
AMK59197.1		uncultured bacterium UPO45

Table S2. Details of SDIMOs that have been characterised in previous studies: Hosts, substrates, experimental methods, and notable features.

In vivo experiments are defined by whole cell assay experiments while *in vitro* experiments are defined by assays conducted through heterologous expression or experiments utilising purified protein. TMO = toluene MO, PH = phenol hydroxylase, VC = vinyl chloride, DCE = Dichloroethylene, TCE = Trichloroethylene, PCE = tetrachloroethylene or perchloroethylene, THF = tetrahydrofuran, ND = no data. Substrates in bold can act as growth substrates, while all others are oxidised by cometabolism.

SDIMO Group	Bacterial host	Enzyme name	Substrates oxidised	Heterologous expression	Notes
1	*Pseudomonas mendocina* KR1	TmoABCDEF	**Toluene**, **benzene** (Tao *et al.*, 2004), **phenol** (Tao *et al.*, 2004), xylene (Pikus *et al.*, 1997), C_4-C_6 alkenes (McClay *et al.*, 2000), naphthalene (Tao *et al.*, 2005), fluorene (Tao *et al.*, 2005), nitrobenzene (Haigler and Spain, 1991), *N*-nitrosodimethylamine (Sharp *et al.*, 2005), dioxane (Mahendra and Alvarez-Cohen, 2006), chloroform (McClay *et al.*, 1996), TCE (McClay *et al.*, 1995)	- Can be expressed in *E. coli,* confers ability to oxidise toluene to *p*-cresol, and indole to indigo (Yen *et al.*, 1991)	- Enzyme also known as toluene-4-monooxygenase or toluene-*para*-monooxygenase
1	*Ralstonia picketti* PKO1	$TbuA_1UBVA_2C$	**Toluene**, **benzene**, **phenol** (Kukor and Olsen, 1990, Olsen *et al.*, 1994), ***m*-cresol** (Kukor and Olsen, 1990), catechol (Kukor and Olsen, 1990), ethylbenzene (Olsen *et al.*, 1994), *o*-xylene, *p*-xylene, *m*-xylene (Olsen *et al.*, 1994), *o*-cresol, *p*-cresol, styrene	- Can be expressed in *P. aeruginosa* PAO1 (Kukor and Olsen, 1990, Olsen *et al.*, 1994) and *E. coli* (Tao *et al.*, 2004) - Recombinant PAO1 can grow on aromatics (Kukor and Olsen, 1990, Olsen *et al.*, 1994)	- Enzyme also known as toluene-3-monooxygenase or toluene-*para*-monooxygenase - Host previously known as *Pseudomonas picketti* PKO1

			(Olsen *et al.*, 1994), phenylacetylene (Olsen *et al.*, 1994), *N*-nitrosodimethylamine (Sharp *et al.*, 2005), nitrobenzene (Haigler and Spain, 1991), naphthalene (Fishman *et al.*, 2005), 1,4-dioxane (Mahendra and Alvarez-Cohen, 2006), TCE (Park *et al.*, 2002),	- Recombinant *E. coli* can sequentially hydroxylate benzene (Tao *et al.*, 2004)	
1	*Burkholderia cepacia* G4	$TomA_0A_1A_2A_3A_4A_5$	**Toluene**, **phenol**, ***o*-cresol**, ***m*-cresol** (Nelson *et al.*, 1987), ethene, propene (Yeager *et al.*, 1999), benzene (Radway *et al.*, 1998), diethyl ether (Hur *et al.*, 1998), 1,4-dioxane (Mahendra and Alvarez-Cohen, 2006), VC, DCE, TCE (Nelson *et al.*, 1987, Shields and Reagin, 1992, Shields *et al.*, 1995)	- Can be expressed in *E. coli* (Shields *et al.*, 1995), but requires vector promoter. Native promoter is inactive - Recombinants oxidised toluene, phenol, cresols, TCE (Shields *et al.*, 1995)	- Enzyme also known as toluene-2-monooxygenase or toluene-*ortho*-monooxygenase - Host previously known as *Pseudomonas cepacia* G4 - G4 was first bacterium shown to be capable of aerobic TCE oxidation (Nelson *et al.*, 1987) - Genes are on a plasmid (Shields *et al.*, 1995)
1	*Pseudomonas stutzeri* OX1	TouABCDEF	***o*-xylene, toluene, *o*-cresol, *p*-cresol, *m*-cresol, 2,3-dimethylphenol, 3,4-dimethylphenol** (Baggi *et al.*, 1987, Bertoni *et al.*, 1998), indigo (Bertoni *et al.*, 1996), benzene (Bertoni *et al.*, 1996), ethylbenzene (Bertoni *et al.*, 1996), styrene (Bertoni *et al.*, 1996), xylenes (Bertoni *et al.*, 1996), naphthalene (Bertoni *et al.*, 1996), phenol (Vardar and Wood, 2004),	- Can be expressed in *E. coli* and *Pseudomonas putida* KT2440 (Bertoni *et al.*, 1996) - Recombinant KT2440 grew on toluene and o-xylene - Recombinant *E. coli* oxidised benzene, styrene, naphthalene, xylenes (Bertoni *et al.*, 1996)	- Tou is the only enzyme known to date that can oxidise PCE (Ryoo *et al.*, 2000) - Toluene is oxidised to all three isomers of cresol, indicating low regioselectivity (Vardar and Wood, 2004) - Site-directed mutagenesis created variants with modified regioselectivity and substrate range (Ryoo *et al.*, 2000, Vardar and Wood, 2004)

			catechol (Vardar and Wood, 2004), hydroquinone (Vardar and Wood, 2004), resorcinol (Vardar and Wood, 2004), VC, DCE, TCE (Shim and Wood, 2000, Vardar *et al.*, 2005), PCE (Ryoo *et al.*, 2000)		- Strain OX1 also harbours a PhMO (see below)
1	*Azoarcus sp.* DD4	TmoABCDEF	**Propane**, toluene, 1,4-dioxane, 1,1-DCE (Deng *et al.*, 2020)	- Can be expressed in *E. coli* (Deng *et al.*, 2020) - Recombinants oxidised toluene, propane, 1,4-dioxane and 1,1-DCE (Deng *et al.*, 2020)	- Strain DD4 possesses five distinct SDIMO gene clusters (Deng *et al.*, 2019) - DD4 knockout mutant lacking group 3 and 5 SDIMOs was able to degrade 1,4-dioxane and 1,1-DCE (Deng *et al.*, 2020)
1	*Pseudomonas aeruginosa* JI104	BmoABCD$_1$D$_2$EF	**Benzene**, toluene, *o*-xylene, *m*-xylene, ethylbenzene, propyl-benzene (Kitayama *et al.*, 1996), TCE (Han *et al.*, 2001)	ND	- Toluene is oxidised to all three isomers of cresol, indicating low regioselectivity (Kitayama *et al.*, 1996) - Bmo activity has been exploited to create a novel TCE biosensor (Han *et al.*, 2001)
1	*Pseudomonas* sp. M4	TmoABCDEF	**Toluene** and indole (Wongsaroj *et al.*, 2015)	- Can be expressed in *E. coli* (Wongsaroj *et al.*, 2015) - Recombinants oxidised indole to both indigo and indirubin (Wongsaroj *et al.*, 2015)	- Can hydroxylate indole at both the 2- and 3- positions (Wongsaroj *et al.*, 2015)
1	*Bradyrhizobium* sp. BTAi1	TmoABCDEF	Toluene, phenol, benzene, nitrobenzene, naphthalene, indole, and TCE (Yanik-Yildirim and Vardar-Schara, 2014)	- Can be expressed in *E. coli* (Yanik-Yildirim and Vardar-Schara, 2014) - Recombinants oxidised toluene, phenol, benzene, nitrobenzene,	- Discovered in genome of BTAi1, primary substrate unknown (Yanik-Yildirim and Vardar-Schara, 2014)

				naphthalene, indole, and TCE (Yanik-Yildirim and Vardar-Schara, 2014)	- Site-directed mutagenesis showed P101 and P103 in TmoA were critical to regiospecificity (Yanik-Yildirim and Vardar-Schara, 2014)
1	*Xanthobacter autotrophicus* Py2	XamoABCDEF	**Propene, phenol** (van Ginkel and de Bont, 1986, Small and Ensign, 1997, Zhou *et al.*, 1999), C_2-C_6 alkenes (Van Keulen and Da Fonseca, 1994, Small and Ensign, 1997), toluene (Zhou *et al.*, 1999), benzene (Zhou *et al.*, 1999), chlorinated ethenes and propenes (Ensign *et al.*, 1992), TCE (Reij *et al.*, 1995)	- Unable to express in *E. coli,* problems localised to reductase and hydroxylase components (Champreda *et al.*, 2004) - Can be expressed in *Xanthobacter* strains (Zhou *et al.*, 1996)	- Genes are on a megaplasmid (Zhou *et al.*, 1996)
1	*Rhodococcus* sp. AD45	IsoABCDEF	**Isoprene** (van Hylckama Vlieg *et al.*, 2000, Crombie *et al.*, 2015, Sims *et al.*, 2022) *cis*-DCE, *trans*-DCE, toluene, styrene, propene (van Hylckama Vlieg *et al.*, 1998)	ND	- Expression is repressed by preferred carbon sources (succinate) (Crombie *et al.*, 2015) - Inducer probably epoxyisoprene (Crombie *et al.*, 2015) - Plasmid encoded (Crombie *et al.*, 2015)
1	*Xanthobacter* sp. YN2	ThmABCDEF	**1,4-dioxane**, **1,4-dioxene**, **1,3-dioxane**, and **THF** (Ma *et al.*, 2021)	ND	- RT-qPCR indicates the SDIMO is constitutively expressed but further upregulated when grown on 1,4-dioxane (Ma *et al.*, 2021)
1	*Mycobacterium gadium* IBE100	IsoABCDEF	**Isobutene** (Helbich *et al.*, 2023)	ND	- Genes are on a plasmid (Helbich *et al.*, 2023)
1	*Mycobacterium paragordonae* IBE200	IsoABCDEF	**Isobutene** (Helbich *et al.*, 2023)	ND	- Genes are on a plasmid (Helbich *et al.*, 2023)

2	*Pseudomonas* sp. JS150	TbmABCDEF	**Phenol** (Mrozik *et al.*, 2011), ***p*-chlorobenzene** (Spain and Nishino, 1987), benzene (Johnson and Olsen, 1995), toluene (Johnson and Olsen, 1995), catechol (Haigler *et al.*, 1992)	- Can be expressed in *Pseudomonas aeruginosa* PAO1c - Recombinants can grow on phenol and benzene, and hydroxylate toluene and chlorobenzene (Johnson and Olsen, 1995)	- JS150 is a non-encapsulated variant of *Pseudomonas* sp. JS1 (Spain and Nishino, 1987) - Isolate also has two other phenol hydroxylase gene clusters (Tbc1, Tbc2) (Johnson and Olsen, 1995, Kahng *et al.*, 2001)
2	*Pseudomonas putida* CF600	DmpKLMNOP (PHP3 in trees)	**Phenol** (Shingler *et al.*, 1989), ***m*-cresol**, ***o*-cresol**, ***p*-cresol**, **3,4-dimethylphenol** (Shingler *et al.*, 1989, Norlund *et al.*, 1990), catechol, benzoate, *p*-hydroxybenzoate (Norlund *et al.*, 1990), 4-chlorophenol (Nowak and Mrozik, 2016), dimethyl sulphide (Horinouchi *et al.*, 1999)	- Can be expressed in *E. coli* (Horinouchi *et al.*, 1999) and *Pseudomonas* strain PB2701 (Norlund *et al.*, 1990) - Recombinant PB2701 can grow on phenol (Norlund *et al.*, 1990) - Recombinant *E. coli* oxidised dimethyl sulfide to dimethyl sulfoxide (Horinouchi *et al.*, 1999)	- Controlled by regulatory protein DmpR, as shown with luciferase reporter plasmid in *E. coli* This biosensor responds to phenol, *m*-cresol, *o*-cresol, *p*-cresol, 2,3-dimethylphenol, 3,4-di-methylphenol, catechol, 2-ethyl-phenol, 3-ethyl-phenol, and 3-methyl-catechol, 4-methyl-catechol, but not benzoate (Shingler and Moore, 1994)
2	*Pseudomonas stutzeri* OX1	$PhKL_2MN_2O_2P$ (PH_N in trees)	**Phenol**, *o*-cresol, *m*-cresol, *p*-cresol, 2,3- dimethylphenol, 3,5-dimethyl-phenol, 2,5- dimethylphenol (Arenghi *et al.*, 2001), benzene (Cafaro *et al.*, 2004), catechol, 4-hydroxybenzoic acid, and resorcinol (Wang *et al.*, 2018)	- Can be expressed in *E. coli* (Izzo *et al.*, 2011) - PhK component not essential but increased hydroxylase activity in recombinants (Izzo *et al.*, 2011)	- Strain OX1 also harbours a TMO (see above) - The TMO and PhMO in OX1 have opposite regioselectivities (Cafaro *et al.*, 2005)
2	*Pseudomonas putida* BH	PheAB	**Phenol**, ***o*-cresol**, ***m*-cresol**, ***p*-cresol**, **benzoate**, **protocatechuate**, ***p*-hydroxy-benzoate**, ***m*-toluate** (Fujita *et al.*, 1995), TCE (Fujita *et al.*, 1995)	- Can be expressed in *E. coli* (Fujita *et al.*, 1995, Takeo *et al.*, 1995), *P. putida* KT2440 (Fujita *et al.*, 1995, Takeo *et al.*, 1995), and *P.*	

				aeruginosa PAO1c (Teramoto *et al.*, 1999) - Recombinant KT2440 can grow on phenol (Takeo *et al.*, 1995), *o*-cresol, *m*-cresol, and *p*-cresol (Fujita *et al.*, 1995), and degrade TCE (Fujita *et al.*, 1995) - Recombinant *E. coli* can degrade phenol and TCE (Takeo *et al.*, 1995)	
2	*Ralstonia eutropha* sp. E2	PoxABCDEF	**Phenol** (Watanabe *et al.*, 1996), *o*-cresol, *m*-cresol, *p*-cresol, resorcinol, quinol, orcinol, *o*-chlorophenol, *m*-chlorophenol, *p*-chlorophenol, 2,3-dimethylphenol, 3,4-dimethylphenol, benzene, toluene, and 2-ethylphenol (Teramoto *et al.*, 1999)	- Can be expressed in *Pseudomonas aeruginosa* PAO1c (Hino *et al.*, 1998, Teramoto *et al.*, 1999) - Recombinant PAO1c can grow on phenol (Hino *et al.*, 1998)	- May also be expressed in *E. coli* JM109 based on yellow colony colour in presence of phenol (Hino *et al.*, 1998)
2	*Comamonas testosteroni* R5	PhcKLMNOP	**Phenol** (Watanabe *et al.*, 1996), substituted phenols (Teramoto *et al.*, 1999), TCE (Futamata *et al.*, 2001)	- Can be expressed in *Pseudomonas aeruginosa* PAO1c (Teramoto *et al.*, 1999) - Recombinant PAO1c preferentially metabolised *para*-substituted phenols (Teramoto *et al.*, 1999)	- Co-metabolic degradation of TCE only occurs after phenol is depleted (Futamata *et al.*, 2001)
2	*Comamonas testosteroni* TA441	AhpKLMNOP	**Phenol** (Arai *et al.*, 1998) and dimethyl sulfide (Horinouchi *et al.*, 1999)	- Can be expressed in *E. coli* (Horinouchi *et al.*, 1999) - Recombinant *E. coli* oxidised dimethyl sulfide to the sulfoxide (Horinouchi *et al.*, 1999)	- Strain initially unable to grow on phenol, but adapted due to regulatory mutations (Arai *et al.*, 1998, Arai *et al.*, 1999, Arai and Ishii, 2019)

2	*Acinetobacter radioresistens* sp. S13	Mop$KL_2MN_2O_2P$	**Phenol** (Griva *et al.*, 2003) and **benzoate** (Mazzoli *et al.*, 2007)	ND	- SDIMO genes on chromosome (Mazzoli *et al.*, 2007) - Biochemically well-characterised, subunit organisation and interactions (Griva *et al.*, 2003), pH, temperature, ionic strength preferences (Divari *et al.*, 2003) - S13 preferentially utilises acetate, then benzoate, then phenol (Mazzoli *et al.*, 2007)
2	*Cupriavidus metallidurans* ZM02	DmpKLMNOP	**THF** and **phenol** (Ren *et al.*, 2022)	- Can be expressed in *Cupriavidus* JMP134 (Ren *et al.*, 2022) - Recombinant JMP134 can grow on THF, although cell extracts were inactive (Ren *et al.*, 2022) - DmpL, DmpN, DmpP subunits insoluble in *E. coli*, this was improved by coexpression of GroES/EL (Ren *et al.*, 2022)	- Knockout of *dmpL*, *dmpN*, *dmpP*, and *dmpR* prevented strain ZM02 growth on phenol and THF (Ren *et al.*, 2022)
2	*Arthrobacter* sp. W1	PhKLMNOP	**Phenol**, **catechol**, **salicyclic acid** (Wang *et al.*, 2009), benzene, toluene, *m*-cresol, *o*-cresol, *p*-cresol (Ma *et al.*, 2013), indole (Shi *et al.*, 2013), carbazole (Shi *et al.*, 2015)	- Can be expressed in *E. coli* BL21(DE3) (Ma *et al.*, 2013, Shi *et al.*, 2013) - Recombinant *E. coli* oxidised indole to indigo (Shi *et al.*, 2013)	
2	*Alcaligenes faecalis* IS-46	AfpKLMNOP	**Phenol** (Zhang *et al.*, 2004)	- Can be expressed in *E. coli* (Zhu *et al.*, 2008) - Recombinant *E. coli* transform phenol to 2-hydroxymuconic semialdehyde (Zhu *et al.*, 2008)	

2	*Sulfobacillus acidophilus* TPY	MhpKLMNOP	ND	- Reductase MhpP can be expressed in *E. coli* (Li *et al.*, 2016)	- Involvement of PH in phenol oxidation indicated by RT-qPCR (Zhou *et al.*, 2016)
3	*Methylococcus capsulatus* (Bath)	$MmoX_2Y_2BZ_2DC$	**Methane** (Sullivan *et al.*, 1998, Jiang *et al.*, 2010)	ND	- sMMOs are relatively well-studied and reviewed (Stainthorpe *et al.*, 1990, Cardy *et al.*, 1991, Sullivan *et al.*, 1998, Banerjee *et al.*, 2019) - sMMOs are regulated by cofactor availability rather than substrates or metabolites (Nielsen *et al.*, 1996, Scanlan *et al.*, 2009) - Electrochemistry and redox properties of the (Bath) sMMO have been studied (Kazlauskaite *et al.*, 1996) - Crystal structure of the OB3b sMMO is known (Jones *et al.*, 2020)
3	*Methylosinus trichosporium* OB3b	$MmoX_2Y_2BZ_2DC$		- Can be expressed in pMMO-containing hosts (Lloyd *et al.*, 1999) and in *Pseudomonas putida* F1 (Jahng and Wood, 1994), although the latter finding is controversial (Lloyd *et al.*, 1999)	
3	*Methylosinus sporium* 5	MmoXYBZDC	**Methane**, pentane and heptane, dichloroethane, 1,2-dichloropropane, propene, toluene, benzene, naphthalene (Park *et al.*, 2018)	ND	- Mixing the hydroxylase, reductase, and coupling proteins between Bath, OB3b and *M sporium* 5 can produce functional sMMOs (Pilkington and Dalton, 1990) - Has duplicated *mmoX* gene (*mmoX2*) (Ali *et al.*, 2006)
3	*Methylobacterium* sp. strain CRL-26	MmoXYBZDC	**Methane**, C_2-C_8 alkanes, chloromethane, bromomethane, 1-nitropropane, isobutane, carbon monoxide, butadiene, isobutylene,	ND	

			cis-but-2-ene *trans*-but-2-ene, isoprene,2-methyl-1-butene, 2-methyl-2-butene, 2-bromo-1-butene, 2-bromo-2-butene, dimethyl ether, butyl ether, cyclohexane, toluene, benzene (Patel *et al.*, 1982)		
3	*Methylocystis* sp. strain M	MmoXYBZDC	**Methane**, propane, *n*-butane, *n*-pentane (Shinohara *et al.*, 1998a), propene (Nakajima *et al.*, 1992), 1,1-DCE, *cis*-1,2-DCE, *trans*-1,2-DCE, 1,2-dibromoethylene, 1,1,2-trichloroethane, 1,2-dichloroethane, chloroform (Uchiyama *et al.*, 1989), chloral hydrate, TCE (Saeki *et al.*, 1999b)	ND	- sMMO components have been purified (Nakajima *et al.*, 1992) - Coupling protein interacts with the hydroxylase for stabilisation rather than electron transfer (Shinohara *et al.*, 1998b)
3	*Methylocystis* sp. WI 14	MmoXYBZDC	**Methane**, C_2-C_4 alkenes, C_3-C_7 alkanes, diethyl ether, cyclohexane, benzene, toluene, xylene, naphthalene, chloropentane, chlorobenzene, fluorobenzene, bromobenzene, chloro-naphthalene (Grosse *et al.*, 1999)	ND	- Chlorinated compounds are hydroxylated but not dehalogenated (Grosse *et al.*, 1999)
3	*Methylocella silvestris* BL2	MmoXYBZDC	**Methane** and naphthalene (Theisen *et al.*, 2005)	ND	- Only has sMMO, lacks pMMO (Theisen *et al.*, 2005) - sMMO is absent in acetate-grown cultures (Theisen *et al.*, 2005) - sMMO transcription is dependent on the presence of methane (Smirnova and Dunfield, 2018)

3	*Methyloferula stellata* AR4	MmoXYBZDC	**Methane** (Vorobev *et al.*, 2011)	ND	- Only has sMMO, lacks pMMO (Smirnova and Dunfield, 2018) - sMMO is constitutively transcribed (Smirnova and Dunfield, 2018)
3	*Methylovulum miyakonense* HT12	MmoXYBZDC	**Methane** (Iguchi *et al.*, 2010)	ND	- Transcription of sMMO only occurs under copper-limiting conditions in the presence of methane (Iguchi *et al.*, 2010)
3	*Methylomonas methanica* 68-1	MmoXYBZDC	**Methane**, TCE, naphthalene (Koh *et al.*, 1993)	ND	- Activity of sMMO quantified via naphthalene assay (Koh *et al.*, 1993)
3	*Methylomonas* sp. strain GYJ3	MmoXYBZDC	**Methane** and propene (Shen *et al.*, 1997)	ND	- Activity of sMMO quantified via oxidation of propene (Shen *et al.*, 1997)
3	*Methylocystis hirsuta* CSC1	MmoXYBZDC	**Methane** and naphthalene (Lindner *et al.*, 2007)	ND	- sMMO only expressed under copper-limiting conditions (Lindner *et al.*, 2007) - Activity of sMMO quantified via naphthalene oxidation assay (Lindner *et al.*, 2007)
3	*Thauera butanivorans* (Known as *Pseudomonas butanovora* until 2009 (Dubbels *et al.*, 2009))	BmoXYBZDC	**ethane**, **propane, butane,** C5-C9 alkanes (Takahashi *et al.*, 1980), ethene (Sayavedra-Soto *et al.*, 2001), isobutane, isopentane (Dubbels *et al.*, 2007), chloroform, TCE, 1,2-*cis*-DCE, VC (Hamamura *et al.*, 1997), methane (Halsey *et al.*, 2006)	ND	- *T butanivorans* preferentially grows on C_2-C_4 alkanes, is able to grow on alcohols and carboxylic acids but is unable to utilise sugars (Takahashi *et al.*, 1980) - *T butanivorans* can grow on C_2-C_5 alcohols, preferring 1-alcohols over 2-alcohols (Arp, 1999) - BMO is constitutively expressed (Sayavedra-Soto *et al.*, 2001), but

					upregulated by 1-butanol and butyraldehyde (Sayavedra-Soto *et al.*, 2005) - Propionate represses BMO (Doughty *et al.*, 2006) - Chaperonin BmoG is vital to BMO function. *bmoG* deletion mutants cannot grow on alkanes (Kurth *et al.*, 2008)
3	*Mycobacterium chubuense* NBB4	SmoXYB1C1Z	**Ethane**, **propane**, **butane**, **ethene**, 1-propene, 1-butene, VC, and dichloroethane (Martin *et al.*, 2014)	- Can be expressed in *M. smegmatis* mc^2-155 (Martin *et al.*, 2014)	- RT-PCR indicates SMO is constitutively expressed (Coleman *et al.*, 2011) - Genes are on a plasmid (Coleman *et al.*, 2011) - Has 5 components, unlike all other group 3 SDIMOs, which have 6 (Martin *et al.*, 2014)
3	*Rhodococcus* sp. ZPP	SmoXYB1C1Z	**Ethane** (Zou *et al.*, 2021)	- Can be expressed in *Rhodococcus erythropolis* ATCC 25544 (Zou *et al.*, 2021) - Recombinant *R erythropolis* oxidised ethane (Zou *et al.*, 2021)	- Smo genes are on a conjugative plasmid (Zou *et al.*, 2021)
4	*Rhodococcus rhodochrous* B-276	AmoABCD	**1,2-butadiene** (Furuhashi *et al.*, 1981), **ethene**, **propene**, **1-butene**, C_5-C_{18} alkenes (Furuhashi *et al.*, 1986), 2-alkenes, styrene (Furuhashi *et al.*, 1986, Saeki and Furuhashi, 1994, Gallagher *et al.*, 1997), DCE, TCE (Saeki *et al.*, 1999a), trifluoro-	- Can be expressed in *Streptomyces lividans* TK24 (Smith *et al.*, 1999) - Not clear if functional in *E. coli*. One study with recombinant *E. coli* showed trifluoropropylene oxidation (Saeki and Furuhashi, 1994, Smith *et al.*, 1999), but later study failed to	- AMO has been purified (Gallagher *et al.*, 1997) and crystal structure solved (Gallagher *et al.*, 1998) - Constitutively expressed in glucose-grown cells (Furuhashi *et al.*, 1986) - Genes are on a large plasmid (Saeki *et al.*, 1999a). Deletion in

			propylene (Saeki and Furuhashi, 1994)	show propene oxidation (Smith *et al.*, 1999)	plasmid occurs when passaged in rich medium (Saitoh *et al.*, 2013)
4	*Mycobacterium chubuense* NBB4	EtnABCD	**Ethene**, **propene**, C_4-C_{10} alkenes, isoprene, allyl alcohol, styrene cyclopentene, cyclohexene, dihydropyran, and indene (Cheung *et al.*, 2013, McCarl *et al.*, 2018)	- Can be expressed in *M. smegmatis* mc^2-155 and *P. putida* KT2440 (McCarl *et al.*, 2018) - Recombinant mc^2-155 oxidised ethene, propene, 1-butene, 1-octene, cyclopentene, and styrene (McCarl *et al.*, 2018) - Recombinant KT2440 oxidised ethene (McCarl *et al.*, 2018)	- Co-expression of chaperonins increased EtnABCD activity in recombinant KT2440 (McCarl *et al.*, 2018) - Genes are on a plasmid (Coleman *et al.*, 2011)
4	*Nocardioides* sp. JS614	EtnABCD	**Ethene**, **VC**, **vinyl fluoride** (Coleman *et al.*, 2002, Mattes *et al.*, 2005, Taylor *et al.*, 2007), propene, 1-butene, *trans*-2-butene, *cis*-2-butene, isobutylene (Owens *et al.*, 2009)	ND	- Epoxyethane inducer is needed to allow growth on propene (Taylor *et al.*, 2010) - JS614 Etn has a higher activity against VC than against ethene (Mattes *et al.*, 2005) - Genes are on a plasmid (Chuang and Mattes, 2007)
4	*Mycobacterium* sp. M156	PmoABCD	**Propene**, **1-butene**, styrene (Rigby *et al.*, 1994, Woodland *et al.*, 1995, Chan Kwo Chion *et al.*, 2005)	- Can be expressed in *M. smegmatis* mc^2-155 (Chan Kwo Chion *et al.*, 2005) - Recombinant mc^2-155 oxidised styrene (Chan Kwo Chion *et al.*, 2005)	- Expression unsuccessful in *E. coli,* attributed to gene overlap in *pmoABCD* (Chan Kwo Chion *et al.*, 2005) - Heterologous expression in mc^2-155 required weak promoter and growth in minimal medium (Chan Kwo Chion *et al.*, 2005) - Pmo genes are chromosomal (Chan Kwo Chion *et al.*, 2005)

4	*Halieaceae* sp. PE-TB08W	AmoABCD	**Propene**, **1-butene** and ***n*-butane** (Suzuki *et al.*, 2019)	ND	- RT-PCR confirmed AmoABCD involvement in growth on propene and *n*-butane (Suzuki *et al.*, 2019)
5	*Gordonia* sp. TY-5	PrmABCD	**Propane** (Kotani *et al.*, 2003), and butane (Kotani *et al.*, 2006)	ND	- Prm expression induced strongly by propane, weakly by ethane or butane (Kotani *et al.*, 2003) - Oxidises butane to 2-butanol (Kotani *et al.*, 2006)
5	*Pseudonocardia* sp.TY-7	Prm1ABCD	C_2–C_6 alkanes (**propane**) (Kotani *et al.*, 2006)	ND	- Prm1 induced by C_2–C_6 alkanes. Butane and pentane are best (Kotani *et al.*, 2006) - TY-7 metabolises butane into both 1-butanol and 2-butanol (Kotani *et al.*, 2006)
5	*Pseudonocardia* sp.TY-7	Prm2ABCD	C_2–C_6 alkanes (**propane**) (Kotani *et al.*, 2006)		
5	*Pseudonocardia tetrahydrofuranoxydans* K1	ThmADBC	**THF** (Kohlweyer *et al.*, 2000), 1,4-dioxane (Mahendra and Alvarez-Cohen, 2006)	- Can be expressed in *Rhodococcus jostii* RHA1 (Sales *et al.*, 2013) - Recombinant RHA1 oxidised THF and 1,4-dioxane (Sales *et al.*, 2013)	- K1 can mineralise dioxane but does not grow on it (Mahendra *et al.*, 2007) - Thm genes are on a plasmid (Thiemer *et al.*, 2003) - Reductase ThmD insoluble when expressed in *E. coli* (Oppenheimer *et al.*, 2010)
5	*Pseudonocardia dioxanivorans* CB1190	ThmADBC	**THF, dioxane** (Mahendra and Alvarez-Cohen, 2005)	- Can be expressed in *Rhodococcus jostii* RHA1 (Sales *et al.*, 2013) - Recombinant RHA1 oxidised THF and 1,4-dioxane (Sales *et al.*, 2013)	- Thm genes are on a plasmid (Sales *et al.*, 2013) - β-hydroxyethoxyacetic acid induces Thm expression (Sales *et al.*, 2013)
5	*Pseudonocardia* sp. ENV478	ThmADBC	**THF**, 1,4-dioxane, 1,3-dioxolane, bis-2-chloroethylether, methyl tert-butyl ether (Vainberg *et al.*, 2006)	ND	- Thm is constitutively expressed in ENV478 (Vainberg *et al.*, 2006, Masuda *et al.*, 2012a)

					- Knockdown of *thm* transcripts prevents growth on THF (Masuda *et al.*, 2012a)
5	*Rhodococcus jostii* RHA1	PrmABCD	**Propane** (Sharp *et al.*, 2007), *N*-nitroso-dimethylamine (Sharp *et al.*, 2007), *N*-nitroso-diethylamine, *N*-nitroso-di-*n*-propyl-amine, *N*-nitroso-pyrrolidine, *N*-nitroso-morpholine (Homme and Sharp, 2013), 1,2,3-tricholoro-propane (Wang and Chu, 2017)	ND	- PrmABCD shown to be involved in propane metabolism through microarray, RT-PCR, knockout (Sharp *et al.*, 2007)
5	*Mycobacterium smegmatis* mc^2-155	MimABCD	**Propane**, **acetone**, phenol (Furuya *et al.*, 2011, Furuya *et al.*, 2015)	- Can be expressed in *Rhodococcus opacus* B-4 (Furuya *et al.*, 2012) and *E. coli* (Furuya *et al.*, 2013) - Recombinant B-4 and *E. coli* both oxidised phenol to hydroquinone (Furuya *et al.*, 2012, Furuya *et al.*, 2013)	- Chaperonin MimG found downstream of MimABCD cluster, similar in other group 5 SDIMOs (Furuya *et al.*, 2012) - Successful expression in B-4 required co-expression of MimG (Furuya *et al.*, 2012) - Successful expression in *E. coli* required co-expression of MimG and codon optimisation (Furuya *et al.*, 2013)
5	Arthrobacter sp. WN18	ThmADBC	**THF**, 1,4-dioxane (Wang *et al.*, 2021a, Wang *et al.*, 2021b)	ND	- Thm is inducible by THF (Wang *et al.*, 2021b)
6	*Mycobacterium* sp. TY-6	PrmABCD	**propane**, **butane,** pentane, hexane (Kotani *et al.*, 2006)	ND	- Prm only upregulated by propane and butane, not ethane, pentane, or hexane (Kotani *et al.*, 2006)
6	*Mycobacterium* sp. ENV421	PrmA1BCD	**Propane**, possibly 1,4-dioxane (Masuda *et al.*, 2012b)	ND	

6	*Mycobacterium* sp. ENV421	PrmA2BCD			- RT-PCR showed propane induced expression of *prmA* (Masuda *et al.*, 2012b) - Unknown which Prm is involved in 1,4-dioxane oxidation (Masuda *et al.*, 2012b)
6	*Mycobacterium dioxanotrophicus* PH-06	PrmABCD	**Propane, 1,4-dioxane**, THF (He *et al.*, 2017, Deng *et al.*, 2018)	- Can be expressed in *M. smegmatis* mc^2-155 (Deng *et al.*, 2018) - Recombinant mc^2-155 oxidised 1,4-dioxane, THF, propane (Deng *et al.*, 2018)	- Plasmid encoded, between two insertion sequences (Deng *et al.*, 2018) - Upregulated by THF and 1,4-dioxane. THF metabolised preferentially when both supplied (He *et al.*, 2017)
6	*Mycobacterium* sp. ELW1	ND	ND	ND	- Genes are on a plasmid (Helbich *et al.*, 2023)
6	*Mycobacterium gadium* IBE100				
6	*Mycobacterium paragordonae* IBE200				
7	*Solimonas soli* DCY12	ZmoABCD	C_2-C_4 alkenes, 1-hexene, 1-octene, isoprene, acrylic acid, VC, 1,2-*cis*-DCE (Yang *et al.*, 2024)	- Can be expressed in *P. putida* KT2440 but not *E. coli* or *M. smegmatis* (Yang *et al.*, 2024) - Recombinant KT2440 oxidised C_2-C_4 alkenes, 1-hexene, 1-octene, isoprene, acrylic acid, VC, and *cis*-1,2-DCE (Yang *et al.*, 2024)	- Native substrate unknown - only SDIMO to date known to oxidise acrylic acid

Supplementary Material - References for Table S2.

Ali, H., Scanlan, J., Dumont, M. G., Murrell, J. C. (2006) Duplication of the *mmoX* gene in *Methylosinus sporium*: cloning, sequencing and mutational analysis. Microbiology, 152: 2931-2942.

Arai, H., Akahira, S., Ohishi, T., Kudo, T. (1999) Adaptation of *Comamonas testosteroni* TA441 to utilization of phenol by spontaneous mutation of the gene for a *trans*-acting factor. Molecular Microbiology 33: 1132-1140.

Arai, H., Akahira, S., Ohishi, T., Maeda, M., Kudo, T. (1998) Adaptation of *Comamonas testosteroni* TA441 to utilize phenol: Organization and regulation of the genes involved in phenol degradation. Microbiology, 144: 2895-2903.

Arai, H., Ishii, M. (2019) Draft genome sequence of *Comamonas testosteroni* TA441, a bacterium that has a cryptic phenol degradation gene cluster. Microbiology Resource Announcements, 8: e00946-19.

Arenghi, F. L., Berlanda, D., Galli, E., Sello, G., Barbieri, P. (2001) Organization and regulation of meta cleavage pathway genes for toluene and *o*-xylene derivative degradation in *Pseudomonas stutzeri* OX1. Applied and Environmental Microbiology, 67: 3304-3308.

Arp, D. J. (1999) Butane metabolism by butane-grown '*Pseudomonas butanovora*'. Microbiology, 145: 1173-1180.

Baggi, G., Barbieri, P., Galli, E., Tollari, S. (1987) Isolation of a *Pseudomonas stutzeri* strain that degrades *o*-xylene. Applied and Environmental Microbiology, 53: 2129-2132.

Banerjee, R., Jones, J. C., Lipscomb, J. D. (2019) Soluble methane monooxygenase. Annual Review of Biochemistry, 88: 409-431.

Bertoni, G., Bolognese, F., Galli, E., Barbieri, P. (1996) Cloning of the genes for and characterization of the early stages of toluene and *o*-xylene catabolism in *Pseudomonas stutzeri* OX1. Applied and Environmental Microbiology, 62: 3704-3711.

Bertoni, G., Martino, M., Galli, E., Barbieri, P. (1998) Analysis of the gene cluster encoding toluene/*o*-xylene monooxygenase from *Pseudomonas stutzeri* OX1. Applied and Environmental Microbiology, 64: 3626–3632.

Cafaro, V., Izzo, V., Scognamiglio, R., Notomista, E., Capasso, P., Casbarra, A., Pucci, P., Di Donato, A. (2004) Phenol hydroxylase and toluene/*o*-xylene monooxygenase from *Pseudomonas stutzeri* OX1: interplay between two enzymes. Applied and Environmental Microbiology, 70: 2211-2219.

Cafaro, V., Notomista, E., Capasso, P., Di Donato, A. (2005) Regiospecificity of two multicomponent monooxygenases from *Pseudomonas stutzeri* OX1: molecular basis for catabolic adaptation of this microorganism to methylated aromatic compounds. Applied and Environmental Microbiology, 71: 4736-4743.

Cardy, D. L. N., Laidler, V., Salmond, G. P. C., Murrell, J. C. (1991) Molecular analysis of the methane monooxygenase (MMO) gene cluster of *Methyiosinus trichosporium* OB3b. Molecular Biotechnology, 5: 335-342.

Champreda, V., Zhou, N. Y., Leak, D. J. (2004) Heterologous expression of alkene monooxygenase components from *Xanthobacter autotrophicus* Py2 and reconstitution of the active complex. FEMS Microbiology Letters, 239: 309-318.

Chan Kwo Chion, C. K., Askew, S. E., Leak, D. J. (2005) Cloning, expression, and site-directed mutagenesis of the propene monooxygenase genes from *Mycobacterium* sp. strain M156. Applied and Environmental Microbiology, 71: 1909-1914.

Cheung, S., McCarl, V., Holmes, A. J., Coleman, N. V., Rutledge, P. J. (2013) Substrate range and enantioselectivity of epoxidation reactions mediated by the ethene-oxidising *Mycobacterium* strain NBB4. Applied and Environmental Microbiology, 97: 1131-1140.

Chuang, A. S., Mattes, T. E. (2007) Identification of polypeptides expressed in response to vinyl chloride, ethene, and epoxyethane in *Nocardioides* sp. strain JS614 by using peptide mass fingerprinting. Applied and Environmental Microbiology, 73: 4368-4372.

Coleman, N. V., Mattes, T. E., Gossett, J. M., Spain, J. C. (2002) Phylogenetic and kinetic diversity of aerobic vinyl chloride-assimilating bacteria from contaminated sites. Applied and Environmental Microbiology, 68: 6162-6171.

Coleman, N. V., Yau, S., Wilson, N. L., Nolan, L. M., Migocki, M. D., Ly, M. A., Crossett, B., Holmes, A. J. (2011) Untangling the multiple monooxygenases of *Mycobacterium* chubuense strain NBB4, a versatile hydrocarbon degrader. Environmental Microbiology Reports, 3: 297-307.

Crombie, A. T., Khawand, M. E., Rhodius, V. A., Fengler, K. A., Miller, M. C., Whited, G. M., McGenity, T. J., Murrell, J. C. (2015) Regulation of plasmid-encoded isoprene metabolism in *Rhodococcus*, a representative of an important link in the global isoprene cycle. Environmental Microbiology, 17: 3314-3329.

Deng, D., Li, F., Li, M. (2018) A novel propane monooxygenase initiating degradation of 1,4-dioxane by *Mycobacterium dioxanotrophicus* PH-06. Environmental Science & Technology Letters, 5: 86-91.

Deng, D., Li, F., Ye, L., Li, M. (2019) Complete genome sequence of *Azoarcus* sp. strain DD4, a Gram-negative propanotroph that degrades 1,4-dioxane and 1,1-dichloroethylene. Microbiology Resource Announcements, 8: e00775-19.

Deng, D., Pham, D. N., Li, F., Li, M. (2020) Discovery of an inducible toluene monooxygenase that cooxidizes 1,4-dioxane and 1,1-dichloroethylene in propanotrophic *Azoarcus* sp. strain DD4. Applied and Environmental Microbiology, 86: e01163-20.

Divari, S., Valetti, F., Caposio, P., Pessione, E., Cavaletto, M., Griva, E., Gribaudo, G., Gilardi, G., Giunta, C. (2003) The oxygenase component of phenol hydroxylase from *Acinetobacter radioresistens* S13. European Journal of Biochemistry, 270: 2244-2253.

Doughty, D. M., Sayavedra-Soto, L. A., Arp, D. J., Bottomley, P. J. (2006) Product repression of alkane monooxygenase expression in *Pseudomonas butanovora*. Journal of Bacteriology, 188: 2586-2592.

Dubbels, B. L., Sayavedra-Soto, L. A., Arp, D. J. (2007) Butane monooxygenase of '*Pseudomonas butanovora*': purification and biochemical characterization of a terminal-alkane hydroxylating diiron monooxygenase. Microbiology, 153: 1808-1816.

Dubbels, B. L., Sayavedra-Soto, L. A., Bottomley, P. J., Arp, D. J. (2009) *Thauera butanivorans* sp. nov., a C2-C9 alkane-oxidizing bacterium previously referred to as '*Pseudomonas butanovora*'. International Journal of Systematic and Evolutionary Microbiology, 59: 1576-1578.

Ensign, S. A., Hyman, M. R., Arp, D. J. (1992) Cometabolic degradation of chlorinated alkenes by alkene monooxygenase in a propylene-grown *Xanthobacter* strain. Applied and Environmental Microbiology, 58: 3038-3046.

Fishman, A., Tao, Y., Rui, L., Wood, T. K. (2005) Controlling the regiospecific oxidation of aromatics via active site engineering of toluene *para*-monooxygenase of *Ralstonia pickettii* PKO1. Journal of Biological Chemistry, 280: 506-514.

Fujita, M., Ike, M., Hioki, J.-I., Kataoka, K., Takeo, M. (1995) Trichloroethylene degradation by genetically engineered bacteria carrying cloned phenol catabolic genes. Journal of Fermentation and Bioengineering, 79: 100-106.

Furuhashi, K., Shintani, M., Takagi, M. (1986) Effects of solvents on the production of epoxides by *Nocardia corallina* B-276. Applied Microbiology and Biotechnology, 23: 218-223.

Furuhashi, K., Taoka, A., Uchida, S., Karube, I., Suzuki, S. (1981) Production of 1,2-epoxyalkanes from 1-alkenes by *Nocardia corallina* B-276. Applied Microbiology and Biotechnology, 12: 39-45.

Furuya, T., Hayashi, M., Kino, K. (2013) Reconstitution of active mycobacterial binuclear iron monooxygenase complex in *Escherichia coli*. Applied and Environmental Microbiology, 79: 6033-6039.

Furuya, T., Hayashi, M., Semba, H., Kino, K. (2012) The mycobacterial binuclear iron monooxygenases require a specific chaperonin-like protein for functional expression in a heterologous host. FEBS Journal, 280: 817-826.

Furuya, T., Hirose, S., Osanai, H., Semba, H., Kino, K. (2011) Identification of the monooxygenase gene clusters responsible for the regioselective oxidation of phenol to hydroquinone in mycobacteria. Applied and Environmental Microbiology, 77: 1214-1220.

Furuya, T., Nakao, T., Kino, K. (2015) Catalytic function of the mycobacterial binuclear iron monooxygenase in acetone metabolism. FEMS Microbiology Letters, 362: fnv136.

Futamata, H., Harayama, S., Watanabe, K. (2001) Diversity in kinetics of trichloroethylene-degrading activities exhibited by phenol-degrading bacteria. Applied Microbiology and Biotechnology, 55: 248-253.

Gallagher, S. C., Cammack, R., Dalton, H. (1997) Alkene monooxygenase from *Nocardia corallina* B-276 is a member of the class of dinuclear iron proteins capable of stereospecific epoxygenation reactions. European Journal of Biochemistry, 247: 635-641.

Gallagher, S. C., George, A., Dalton, H. (1998) Sequence-alignment modelling and molecular docking studies of the epoxygenase component of alkene monooxygenase from *Nocardia corallina* B-276. European Journal of Biochemistry, 254: 480-489.

Griva, E., Pessione, E., Divari, S., Valetti, F., Cavaletto, M., Rossi, G. L., Giunta, C. (2003) Phenol hydroxylase from *Acinetobacter radioresistens* S13. Isolation and characterization of the regulatory component. European Journal of Biochemistry, 270: 1434-1440.

Grosse, S., Laramee, L., Wendlandt, K. D., McDonald, I. R., Miguez, C. B., Kleber, H. P. (1999) Purification and characterization of the soluble methane monooxygenase of the type II methanotrophic bacterium *Methylocystis* sp. strain WI 14. Applied and Environmental Microbiology, 65: 3929-3935.

Haigler, B. E., Pettigrew, C. A., Spain, J. C. (1992) Biodegradation of mixtures of substituted benzenes by *Pseudomonas* sp. strain JS150. Applied and Environmental Microbiology, 58: 2237-2244.

Haigler, B. E., Spain, J. C. (1991) Biotransformation of nitrobenzene by bacteria containing toluene degradative pathways. Applied and Environmental Microbiology, 57: 3156-3162.

Halsey, K. H., Sayavedra-Soto, L. A., Bottomley, P. J., Arp, D. J. (2006) Site-directed amino acid substitutions in the hydroxylase alpha subunit of butane monooxygenase from *Pseudomonas butanovora*: Implications for substrates knocking at the gate. Journal of Bacteriology, 188: 4962-4969.

Hamamura, N., Page, C., Long, T., Semprini, L., Arp, D. J. (1997) Chloroform cometabolism by butane-grown CF8, *Pseudomonas butanovora*, and *Mycobacterium vaccae* JOB5 and methane- grown *Methylosinus trichosporium* OB3b. Applied and Environmental Microbiology, 63: 3607-3613.

Han, T. S., Kim, Y. C., Sasaki, S., Yano, K., Ikebukuro, K., Kitayama, A., Nagamune, T., Karube, I. (2001) Microbial sensor for trichloroethylene determination. Analytica Chimica Acta, 431: 225-230.

He, Y., Mathieu, J., Yang, Y., Yu, P., da Silva, M. L. B., Alvarez, P. J. J. (2017) 1,4-dioxane biodegradation by *Mycobacterium dioxanotrophicus* pH-06 is associated with a group-6 soluble di-iron monooxygenase. Environmental Science & Technology Letters, 4: 494-499.

Helbich, S., Barrantes, I., Dos Anjos Borges, L. G., Pieper, D. H., Vainshtein, Y., Sohn, K., Engesser, K. H. (2023) The 2-methylpropene degradation pathway in *Mycobacteriaceae* family strains. Environmental Microbiology, 25: 2163-2181.

Hino, S., Watanabe, K., Takahashi, N. (1998) Phenol hydroxylase cloned from *Ralstonia eutropha* strain E2 exhibits novel kinetic properties. Microbiology, 144: 1765-1772.

Homme, C. L., Sharp, J. O. (2013) Differential microbial transformation of nitrosamines by an inducible propane monooxygenase. Environmental Science & Technology, 47: 7388-7395.

Horinouchi, M., Yoshida, T., Nojiri, H., Yamane, H., Omori, T. (1999) Oxidation of dimethyl sulfide by various aromatic compound oxygenases from bacteria. Biotechnology Letters, 21: 929–933.

Hur, H., Newman, L. M., Wackett, L. P., Sadowsky, M. J. (1998) Toluene 2-monooxygenase-dependent growth of *Burkholderia cepacia* G4/PR1 on diethyl ether. Applied and Environmental Microbiology, 63: 1606-1609.

Iguchi, H., Yurimoto, H., Sakai, Y. (2010) Soluble and particulate methane monooxygenase gene clusters of the type I methanotroph *Methylovulum miyakonense* HT12. FEMS Microbiology Letters, 312: 71-76.

Izzo, V., Leo, G., Scognamiglio, R., Troncone, L., Birolo, L., Di Donato, A. (2011) PHK from phenol hydroxylase of *Pseudomonas* sp. OX1. Insight into the role of an accessory protein in bacterial multicomponent monooxygenases. Archives of Biochemistry and Biophysics, 505: 48-59.

Jahng, D., Wood, T. K. (1994) Trichloroethylene and chloroform degradation by a recombinant pseudomonad expressing soluble methane monooxygenase from *Methylosinus trichosporium* OB3b. Applied and Environmental Microbiology, 60: 2473-2482.

Jiang, H., Chen, Y., Jiang, P., Zhang, C., Smith, T. J., Murrell, J. C., Xing, X.-H. (2010) Methanotrophs: Multifunctional bacteria with promising applications in environmental bioengineering. Biochemical Engineering Journal, 49: 277-288.

Johnson, G. R., Olsen, R. H. (1995) Nucleotide sequence analysis of genes encoding a toluene/benzene-2-monooxygenase from *Pseudomonas* sp. strain JS150. Applied and Environmental Microbiology, 61: 3336-3346.

Jones, J. C., Banerjee, R., Shi, K., Aihara, H., Lipscomb, J. D. (2020) Structural studies of the *Methylosinus trichosporium* OB3b soluble methane monooxygenase hydroxylase and regulatory component complex reveal a transient substrate tunnel. Biochemistry, 59: 2946-2961.

Kahng, H. Y., Malinverni, J. C., Majko, M. M., Kukor, J. J. (2001) Genetic and functional analysis of the tbc operons for catabolism of alkyl- and chloroaromatic compounds in *Burkholderia* sp. strain JS150. Applied and Environmental Microbiology, 67: 4805-4816.

Kazlauskaite, J., Hill, H. A., Wilkins, P. C., Dalton, H. (1996) Direct electrochemistry of the hydroxylase of soluble methane monooxygenase from *Methylococcus capsulatus* (Bath). European Journal of Biochemistry, 241: 552-556.

Kitayama, A., Suzuki, E., Kawakami, Y., Nagamune, T. (1996) Gene organization and low regiospecificity in aromatic-ring hydroxylation of a benzene monooxygenase of *Pseudomonas aeruginosa* JI104. Journal of Fermentation and Bioengineering, 82: 421-425.

Koh, S. C., Bowman, J. P., Sayler, G. S. (1993) Soluble methane monooxygenase production and trichloroethylene degradation by a type I methanotroph, *Methylomonas methanica* 68–1. Applied and Environmental Microbiology, 59: 960-967.

Kohlweyer, U., Thiemer, B., Schräder, T., Andreesen, J. R. (2000) Tetrahydrofuran degradation by a newly isolated culture of *Pseudonocardia* sp. strain K1. FEMS Microbiology Letters, 186: 301-306.

Kotani, T., Kawashima, Y., Yurimoto, H., Kato, N., Sakai, Y. (2006) Gene structure and regulation of alkane monooxygenases in propane-utilizing *Mycobacterium* sp. TY-6 and *Pseudonocardia* sp. TY-7. Journal of Bioscience and Bioengineering, 102: 184-192.

Kotani, T., Yamamoto, T., Yurimoto, H., Sakai, Y., Kato, N. (2003) Propane monooxygenase and NAD+-dependent secondary alcohol dehydrogenase in propane metabolism by *Gordonia* sp. strain TY-5. Journal of Bacteriology, 185: 7120-7128.

Kukor, J. J., Olsen, R. H. (1990) Molecular cloning, characterization, and regulation of a *Pseudomonas pickettii* PKO1 gene encoding phenol hydroxylase and expression of the gene in *Pseudomonas aeruginosa* PAO1c. Journal of Bacteriology, 172: 4624-4630.

Kurth, E. G., Doughty, D. M., Bottomley, P. J., Arp, D. J., Sayavedra-Soto, L. A. (2008) Involvement of BmoR and BmoG in *n*-alkane metabolism in '*Pseudomonas butanovora*'. Microbiology, 154: 139-147.

Li, M., Guo, W., Chen, X. (2016) A novel NADPH-dependent reductase of *Sulfobacillus acidophilus* TPY phenol hydroxylase: Expression, characterization, and functional analysis. Applied Microbiology and Biotechnology, 100: 10417-10428.

Lindner, A. S., Pacheco, A., Aldrich, H. C., Costello Staniec, A., Uz, I., Hodson, D. J. (2007) *Methylocystis hirsuta* sp. nov., a novel methanotroph isolated from a groundwater aquifer. International Journal of Systematic and Evolutionary Microbiology, 57: 1891-1900.

Lloyd, J. S., De Marco, P., Dalton, H., Murrell, J. C. (1999) Heterologous expression of soluble methane monooxygenase genes in methanotrophs containing only particulate methane monooxygenase. Archives of Microbiology, 171: 364–370.

Ma, F., Shi, S. N., Sun, T. H., Li, A., Zhou, J. T., Qu, Y. Y. (2013) Biotransformation of benzene and toluene to catechols by phenol hydroxylase from *Arthrobacter* sp. W1. Applied Microbiology and Biotechnology, 97: 5097-5103.

Ma, F., Wang, Y., Yang, J., Guo, H., Su, D., Yu, L. (2021) Degradation of 1,4-dioxane by *Xanthobacter* sp. YN2. Current Microbiology, 78: 992-1005.

Mahendra, S., Alvarez-Cohen, L. (2005) *Pseudonocardia dioxanivorans* sp. nov., a novel actinomycete that grows on 1,4-dioxane. International Journal of Systematic and Evolutionary Microbiology, 55: 593-598.

Mahendra, S., Alvarez-Cohen, L. (2006) Kinetics of 1,4-dioxane biodegradation by monooxygenase-expressing bacteria. Environmental Science & Technology, 40: 5435-42.

Mahendra, S., Petzold, C. J., Baidoo, E. E., Keasling, J. D., Alvarez-Cohen, L. (2007) Identification of the intermediates of *in vivo* oxidation of 1,4-dioxane by monooxygenase-containing bacteria. Environmental Science & Technology, 21: 7330–7336.

Martin, K. E., Ozsvar, J., Coleman, N. V. (2014) SmoXYB1C1Z of *Mycobacterium* sp. strain NBB4: a soluble methane monooxygenase (sMMO)-like enzyme, active on C2 to C4 alkanes and alkenes. Applied and Environmental Microbiology, 80: 5801-5806.

Masuda, H., McClay, K., Steffan, R. J., Zylstra, G. J. (2012a) Biodegradation of tetrahydrofuran and 1,4-dioxane by soluble diiron monooxygenase in *Pseudonocardia* sp. strain ENV478. Journal of Molecular Microbiology and Biotechnology, 22: 312-316.

Masuda, H., McClay, K., Steffan, R. J., Zylstra, G. J. (2012b) Characterization of three propane-inducible oxygenases in *Mycobacterium* sp. strain ENV421. Letters in Applied Microbiology, 55: 175-181.

Mattes, T. E., Coleman, N. V., Spain, J. C., Gossett, J. M. (2005) Physiological and molecular genetic analyses of vinyl chloride and ethene biodegradation in *Nocardioides* sp. strain JS614. Archives of Microbiology, 183: 95-106.

Mazzoli, R., Pessione, E., Giuffrida, M. G., Fattori, P., Barello, C., Giunta, C., Lindley, N. D. (2007) Degradation of aromatic compounds by *Acinetobacter radioresistens* S13: Growth characteristics on single substrates and mixtures. Archives of Microbiology, 188: 55-68.

McCarl, V., Somerville, M. V., Ly, M. A., Henry, R., Liew, E. F., Wilson, N. L., Holmes, A. J., Coleman, N. V. (2018) Heterologous expression of *Mycobacterium* alkene monooxygenases in Gram-positive and Gram-negative bacterial hosts. Applied and Environmental Microbiology, 84: e00397-18.

McClay, K., Fox, B. G., Steffan, R. J. (1996) Chloroform mineralization by toluene-oxidizing bacteria. Applied and Environmental Microbiology, 62: 2716-2722.

McClay, K., Fox, B. G., Steffan, R. J. (2000) Toluene monooxygenase-catalyzed epoxidation of alkenes. Applied and Environmental Microbiology, 66: 1877-1882.

McClay, K., Streger, S. H., Steffan, R. J. (1995) Induction of toluene oxidation activity in *Pseudomonas mendocina* KR1 and *Pseudomonas* sp. strain ENVPC5 by chlorinated solvents and alkanes. Applied and Environmental Microbiology, 61: 3479-3481.

Mrozik, A., Miga, S., Piotrowska-Seget, Z. (2011) Enhancement of phenol degradation by soil bioaugmentation with *Pseudomonas* sp. JS150. Journal of Applied Microbiology, 111: 1357-1370.

Nakajima, T., Uchiyama, H., Yagi, O., Nakahara, T. (1992) Purification and properties of a soluble methane monooxygenase from *Methylocystis* sp. M. Bioscience, Biotechnology, and Biochemistry, 56: 736-740.

Nelson, M. J., Montgomery, S. O., Mahaffey, W. R., Pritchard, P. H. (1987) Biodegradation of trichloroethylene and involvement of an aromatic biodegradative pathway. Applied and Environmental Microbiology, 53: 949-954.

Nielsen, A. K., Gerdes, K., Degn, H., Murrell, J. C. (1996) Regulation of bacterial methane oxidation: Transcription of the soluble methane monooxygenase operon of *Methylococcus capsulatus* (Bath) is repressed by copper ions. Microbiology, 142: 1289-1296.

Norlund, I., Powlowski, J., Shingler, V. (1990) Complete nucleotide sequence and polypeptide analysis of multicomponent phenol hydroxylase from *Pseudomonas* sp. strain CF600. Journal of Bacteriology, 172: 6823-6833.

Nowak, A., Mrozik, A. (2016) Facilitation of co-metabolic transformation and degradation of monochlorophenols by *Pseudomonas* sp. CF600 and changes in its fatty acid composition. Water, Air, & Soil Pollution, 227: 83.

Olsen, R. H., Kukor, J. J., Kaphammer, B. (1994) A novel toluene-3-monooxygenase pathway cloned from *Pseudomonas pickettii* PKO1. Journal of Bacteriology, 176: 3749-3756.

Oppenheimer, M., Pierce, B. S., Crawford, J. A., Ray, K., Helm, R. F., Sobrado, P. (2010) Recombinant expression, purification, and characterization of ThmD, the oxidoreductase component of tetrahydrofuran monooxygenase. Archives of Biochemistry and Biophysics, 496: 123-131.

Owens, C. R., Karceski, J. K., Mattes, T. E. (2009) Gaseous alkene biotransformation and enantioselective epoxyalkane formation by *Nocardioides* sp. strain JS614. Applied Microbiology and Biotechnology, 84: 685-692.

Park, J., Kukor, J. J., Abriola, L. M. (2002) Characterization of the adaptive response to trichloroethylene-mediated stresses in *Ralstonia pickettii* PKO1. Applied and Environmental Microbiology, 68: 5231-5240.

Park, Y. R., Yoo, H. S., Song, M. Y., Lee, D.-H., Lee, S. J. (2018) Biocatalytic oxidations of substrates through soluble methane monooxygenase from *Methylosinus sporium* 5. Catalysts, 8: 582-598.

Patel, R. N., Hou, C. T., Laskin, A. I., Felix, A. (1982) Microbial oxidation of hydrocarbons: Properties of a soluble methane monooxygenase from a facultative methane-utilizing organism, *Methylobacterium* sp. strain CRL-26. Applied and Environmental Microbiology, 44: 1130-1137.

Pikus, J. D., Studts, J. M., McClay, K., Steffan, R. J., Fox, B. G. (1997) Changes in the regiospecificity of aromatic hydroxylation produced by active site engineering in the diiron enzyme toluene 4-monooxygenase. Biochemistry, 36: 9283-9289.

Pilkington, S. J., Dalton, H. (1990) Purification and characterisation of the soluble methane monooxygenase from *Methylosinussporium* 5 demonstrates the highly conserved nature of this enzyme in methanotrophs. FEMS Microbiology Letters, 78: 103-108.

Radway, J. C., Santo-Domingo, J. W., Hazen, T. C., Wilde, E. W. (1998) Evaluation of biodegradation potential of foam embedded *Burkholderia cepacia* G4. Biotechnology Letters, 20: 663-686.

Reij, M. W., Kieboom, J., de Bont, J. A., Hartmans, S. (1995) Continuous degradation of trichloroethylene by *Xanthobacter* sp. strain Py2 during growth on propene. Applied and Environmental Microbiology, 61: 2936-2942.

Ren, H., Wang, H., Wang, Y., Chen, Y., Lu, Z. (2022) A multicomponent THF hydroxylase initiates tetrahydrofuran degradation in *Cupriavidus metallidurans* ZM02. Applied and Environmental Microbiology, 88: e01880-21.

Rigby, S. R., Matthews, C. S., Leak, D. J. (1994) Epoxidation of styrene and substituted styrenes by whole cells of *Mycobacterium* sp. M156. Bioorganic & Medicinal Chemistry, 2: 553-556.

Ryoo, D., Shim, H., Canada, K., Barbieri, P., Wood, T. K. (2000) Aerobic degradation of tetrachloroethylene by toluene-*o*-xylene monooxygenase of *Pseudomonas stutzeri* OX1. Nature Biotechnology, 18: 775-778.

Saeki, H., Akira, M., Furuhashi, K., Averhoff, B., Gottschalk, G. (1999a) Degradation of trichloroethene by a linear-plasmid-encoded alkene monooxygenase in *Rhodococcus corallinus* (*Nocardia corallina*) B-276. Microbiology, 145: 1721-1730.

Saeki, H., Furuhashi, K. (1994) Cloning and characterization of a *Nocardia corallina* B-276 gene cluster encoding alkene monooxygenase. Journal of Fermentation and Bioengineering, 78: 399-406.

Saeki, S., Mukai, S., Iwasaki, K., Yagi, O. (1999b) Production of trichloroacetic acid, trichloroethanol and dichloroacetic acid from trichloroethylene degradation by *Methylocystis* sp. strain M. Biocatalysis and Biotransformation, 17: 347-357.

Saitoh, S., Aoyama, H., Akutsu, M., Nakano, K., Shinzato, N., Matsui, T. (2013) Genomic sequencing-based detection of large deletions in *Rhodococcus rhodochrous* strain B-276. Journal of Bioscience and Bioengineering, 116: 309-312.

Sales, C. M., Grostern, A., Parales, J. V., Parales, R. E., Alvarez-Cohen, L. (2013) Oxidation of the cyclic ethers 1,4-dioxane and tetrahydrofuran by a monooxygenase in two *Pseudonocardia* species. Applied and Environmental Microbiology, 79: 7702-7708.

Sayavedra-Soto, L. A., Byrd, C. M., Arp, D. J. (2001) Induction of butane consumption in *Pseudomonas butanovora*. Archives of Microbiology, 176: 114-120.

Sayavedra-Soto, L. A., Doughty, D. M., Kurth, E. G., Bottomley, P. J., Arp, D. J. (2005) Product and product-independent induction of butane oxidation in *Pseudomonas butanovora*. FEMS Microbiology Letters, 250: 111-116.

Scanlan, J., Dumont, M. G., Murrell, J. C. (2009) Involvement of MmoR and MmoG in the transcriptional activation of soluble methane monooxygenase genes in *Methylosinus trichosporium* OB3b. FEMS Microbiology Letters, 301: 181-187.

Sharp, J. O., Sales, C. M., LeBlanc, J. C., Liu, J., Wood, T. K., Eltis, L. D., Mohn, W. W., Alvarez-Cohen, L. (2007) An inducible propane monooxygenase is responsible for *N*-nitrosodimethylamine degradation by *Rhodococcus* sp. strain RHA1. Applied and Environmental Microbiology, 73: 6930-6938.

Sharp, J. O., Wood, T. K., Alvarez-Cohen, L. (2005) Aerobic biodegradation of *N*-nitrosodimethylamine (NDMA) by axenic bacterial strains. Biotechnology and Bioengineering, 89: 608-618.

Shen, R.-N., Yu, C.-L., Ma, Q.-Q., Li, S.-B. (1997) Direct evidence for a soluble methane monooxygenase from type I methanotrophic bacteria: Purification and properties of a soluble methane monooxygenase from *Methylomonas* sp. GYJ3. Archives of Biochemistry and Biophysics, 345: 223-229.

Shi, S., Ma, F., Sun, T., Li, A., Zhou, J., Qu, Y. (2013) Biotransformation of chloro-substituted indoles to indigoids by phenol hydroxylase from *Arthrobacter* sp. W1. Applied Biochemistry and Biotechnology, 170: 951-961.

Shi, S., Qu, Y., Zhou, H., Ma, Q., Ma, F. (2015) Characterization of a novel cometabolic degradation carbazole pathway by a phenol-cultivated *Arthrobacter* sp. W1. Bioresource Technology, 193: 281-287.

Shields, M. S., Reagin, M. J. (1992) Selection of a *Pseudomonas cepacia* strain constitutive for the degradation of trichloroethylene. Applied and Environmental Microbiology, 58: 3977-3983.

Shields, M. S., Reagin, M. J., Gerger, R. R., Campbell, R., Somerville, C. (1995) TOM, a new aromatic degradative plasmid from *Burkholderia* (*Pseudomonas*) *cepacia* G4. Applied and Environmental Microbiology, 61: 1352-1356.

Shim, H., Wood, T. K. (2000) Aerobic degradation of mixtures of chlorinated aliphatics by cloned toluene-*o*-xylene monooxygenase and toluene *o*-monooxygenase in resting cells. Biotechnology and Bioengineering, 70: 693-698.

Shingler, V., Franklin, F. C., Tsuda, M., Holroyd, D., Bagdasarian, M. (1989) Molecular analysis of a plasmid-encoded phenol hydroxylase from *Pseudomonas* CF600. Journal of General Microbiology, 135: 1083-1092.

Shingler, V., Moore, T. (1994) Sensing of aromatic compounds by the DmpR transcriptional activator of phenol-catabolizing *Pseudomonas* sp. strain CF600. Journal of Bacteriology, 176: 1555-1560.

Shinohara, Y., Uchiyama, H., Kusakabe, I. (1998a) Oxidation of some alkanes and trichloroethylene by H2O2/hydroxylase system of soluble methane monooxygenase from *Methylocystis* sp. M. Journal of Fermentation and Bioengineering, 85: 266-270.

Shinohara, Y., Uchiyama, H., Yagi, O., Kusakabe, I. (1998b) Purification and characterization of component B of a soluble methane monooxygenase from *Methylocystis* sp. M. Journal of Fermentation and Bioengineering, 85: 37-42.

Sims, L. P., Lockwood, C. W. J., Crombie, A. T., Bradley, J. M., Brun, N. E. L., Murrell, J. C. (2022) Purification and characterization of the isoprene monooxygenase from *Rhodococcus* sp. strain AD45. Applied and Environmental Microbiology, 88: e0002922.

Small, F. J., Ensign, S. A. (1997) Alkene monooxygenase from *Xanthobacter* strain Py2. Purification and characterization of a four-component system central to the bacterial metabolism of aliphatic alkenes. Journal of Biological Chemistry, 272: 24913-24920.

Smirnova, A. V., Dunfield, P. F. (2018) Differential transcriptional activation of genes encoding soluble methane monooxygenase in a facultative versus an obligate methanotroph. Microorganisms, 6: 20.

Smith, T. J., Lloyd, J. S., Gallagher, S. C., Fosdike, W. L. J., Murrell, J. C., Dalton, H. (1999) Heterologous expression of alkene monooxygenase from *Rhodococcus rhodochrous* B-276. European Journal of Biochemistry, 260: 446-452.

Spain, J. C., Nishino, S. F. (1987) Degradation of 1,4-dichlorobenzene by a *Pseudomonas* sp. Applied and Environmental Microbiology, 53: 1010-1019.

Stainthorpe, A. C., Lees, V., Salmond, G. P. C., Dalton, H., Murrell, J. C. (1990) The methane monooxygenase gene cluster of *Methylococcus capsulatus* (Bath). Gene, 91: 27-34.

Sullivan, J. P., Dickinson, D., Chase, H. A. (1998) Methanotrophs, *Methylosinus trichosporium* OB3b, sMMO, and their application to bioremediation. Critical Reviews in Microbiology, 24: 335-373.

Suzuki, T., Yazawa, T., Morishita, N., Maruyama, A., Fuse, H. (2019) Genetic and physiological characteristics of a novel marine propylene-assimilating *Halieaceae* bacterium isolated from seawater and the diversity of its alkene and epoxide metabolism genes. Microbes and Environment, 34: 33-42.

Takahashi, J., Ichikawa, Y., Sagae, H., Komura, I., Kanou, H., Yamada, K. (1980) Isolation and identification of *n*-butane-assimilating bacterium. Agricultural and Biological Chemistry, 44: 1835-1840.

Takeo, M., Maeda, Y., Okada, H., Miyama, K., Mori, K., Ike, M., Fujita, M. (1995) Molecular cloning and sequencing of the phenol hydroxylase gene from *Pseudomonas putida* BH. Journal of Fermentation and Bioengineering, 79: 485-488.

Tao, Y., Bentley, W. E., Wood, T. K. (2005) Regiospecific oxidation of naphthalene and fluorene by toluene monooxygenases and engineered toluene 4-monooxygenases of *Pseudomonas mendocina* KR1. Biotechnology and Bioengineering, 90: 85-94.

Tao, Y., Fishman, A., Bentley, W. E., Wood, T. K. (2004) Oxidation of benzene to phenol, catechol, and 1,2,3-trihydroxybenzene by toluene 4-monooxygenase of *Pseudomonas mendocina* KR1 and toluene 3-monooxygenase of *Ralstonia pickettii* PKO1. Applied and Environmental Microbiology, 70: 3814-3820.

Taylor, A. E., Arp, D. J., Bottomley, P. J., Semprini, L. (2010) Extending the alkene substrate range of vinyl chloride utilizing *Nocardioides* sp. strain JS614 with ethene oxide. Applied Microbiology and Biotechnology, 87: 2293-2302.

Taylor, A. E., Dolan, M. E., Bottomley, P. J., Semprini, L. (2007) Utilization of Fluoroethene as a surrogate for aerobic vinyl chloride transformation. Environmental Science & Technology, 41: 6378-6383.

Teramoto, M., Futamata, H., Harayama, S., Watanabe, K. (1999) Characterization of a high-affinity phenol hydroxylase from *Comamonas testosteroni* R5 by gene cloning, and expression in *Pseudomonas aeruginosa* PAO1c. Molecular Genetics and Genomics, 262: 552-558.

Theisen, A. R., Ali, M. H., Radajewski, S., Dumont, M. G., Dunfield, P. F., McDonald, I. R., Dedysh, S. N., Miguez, C. B., Murrell, J. C. (2005) Regulation of methane oxidation in the facultative methanotroph *Methylocella silvestris* BL2. Molecular Microbiology, 58: 682-692.

Thiemer, B., Andreesen, J. R., Schrader, T. (2003) Cloning and characterization of a gene cluster involved in tetrahydrofuran degradation in *Pseudonocardia* sp. strain K1. Archives of Microbiology, 179: 266-277.

Uchiyama, H., Nakajima, T., Yagi, O., Tabuchi, T. (1989) Aerobic degradation of trichloroethylene by a new type II methane-utilizing bacterium, strain M. Agricultural and Biological Chemistry, 53: 2903-2907.

Vainberg, S., McClay, K., Masuda, H., Root, D., Condee, C., Zylstra, G. J., Steffan, R. J. (2006) Biodegradation of ether pollutants by *Pseudonocardia* sp. strain ENV478. Applied and Environmental Microbiology, 72: 5218-5224.

van Ginkel, C. G., de Bont, J. A. M. (1986) Isolation and characterization of alkene-utilizing *Xanthobacter* spp. Archives of Microbiology, 145: 403-407.

van Hylckama Vlieg, J. E., Leemhuis, H., Spelberg, J. H., Janssen, D. B. (2000) Characterization of the gene cluster involved in isoprene metabolism in *Rhodococcus* sp. strain AD45. Journal of Bacteriology, 182: 1956-1963.

van Hylckama Vlieg, J. E. T., Kingma, J., van den Wijngaard, A. J., Janssen, D. B. (1998) A glutathione S-transferase with activity towards *cis*-1,2-dichloroepoxyethane is involved in isoprene utilization by *Rhodococcus* sp. strain AD45. Applied and Environmental Microbiology, 64: 2800-2805.

Van Keulen, F., Da Fonseca, M. M. R. (1994) Batch cultivation of *Xanthobacter* Py2 on 1-pentene. Biotechnology Letters, 16: 989-994.

Vardar, G., Barbieri, P., Wood, T. K. (2005) Chemotaxis of *Pseudomonas stutzeri* OX1 and *Burkholderia cepacia* G4 toward chlorinated ethenes. Applied Microbiology and Biotechnology, 66: 696-701.

Vardar, G., Wood, T. K. (2004) Protein engineering of toluene-*o*-xylene monooxygenase from *Pseudomonas stutzeri* OX1 for synthesizing 4-methylresorcinol, methylhydroquinone, and pyrogallol. Applied and Environmental Microbiology, 70: 3253-3262.

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

Wang, B., Chu, K. H. (2017) Cometabolic biodegradation of 1,2,3-trichloropropane by propane-oxidizing bacteria. Chemosphere, 168: 1494-1497.

Wang, J., Shen, X., Wang, J., Yang, Y., Yuan, Q., Yan, Y. (2018) Exploring the promiscuity of phenol hydroxylase from *Pseudomonas stutzeri* OX1 for the biosynthesis of phenolic compounds. ACS Synthetic Biology, 7: 1238-1243.

Wang, P., Cui, T., Li, J., Liu, N., Hong, M. (2021a) 1,4-dioxane degradation performance of tetrahydrofuran-grown *Arthrobacter* sp. WN18. Water, Air, & Soil Pollution, 232: 473.

Wang, P., Li, F., Wang, W., Wang, R., Yang, Y., Cui, T., Liu, N., Li, M. (2021b) Cometabolic degradation of 1,4-dioxane by a tetrahydrofuran-growing *Arthrobacter* sp. WN18. Ecotoxicology and Environmental Safety, 217: 112206.

Wang, P., Qu, Y., Zhou, J. (2009) Biodegradation of mixed phenolic compounds under high salt conditions and salinity fluctuations by *Arthrobacter* sp. W1. Applied Biochemistry and Biotechnology, 159: 623-633.

Watanabe, K., Hino, S., Onodera, K., Kajie, S.-I., Takahashi, N. (1996) Diversity in kinetics of bacterial phenol-oxygenating activity. Journal of Fermentation and Bioengineering, 81: 560-563.

Wongsaroj, L., Sallabhan, R., Dubbs, J. M., Mongkolsuk, S., Loprasert, S. (2015) Cloning of toluene 4-monooxygenase genes and application of two-phase system to the production of the anticancer agent, indirubin. Molecular Biotechnology, 57: 720-726.

Woodland, M. P., Matthews, C. S., Leak, D. J. (1995) Properties of a soluble propene monooxygenase from *Mycobacterium* sp. (strain M156). Archives of Microbiology, 163: 231-234.

Yang, S. N. N., Haritos, V., Kertesz, M. A., Coleman, N. V. (2024) A novel soluble di-iron monooxygenase from the soil bacterium *Solimonas soli*. Environmental Microbiology, 26: e16567.

Yanik-Yildirim, K. C., Vardar-Schara, G. (2014) Saturation mutagenesis of *Bradyrhizobium* sp. BTAi1 toluene 4-monooxygenase at alpha-subunit residues proline 101, proline 103, and histidine 214 for regiospecific oxidation of aromatics. Applied Microbiology and Biotechnology, 98: 8975-8986.

Yeager, C. M., Bottomley, P. J., Arp, D. J., Hyman, M. R. (1999) Inactivation of toluene 2-monooxygenase in *Burkholderia cepacia* G4 by alkynes. Applied and Environmental Microbiology, 65: 632-639.

Yen, K. M., Karl, M. R., Blatt, L. M., Simon, M. J., Winter, R. B., Fausset, P. R., Lu, H. S., Harcourt, A. A., Chen, K. K. (1991) Cloning and characterization of a *Pseudomonas mendocina* KR1 gene cluster encoding toluene-4-monooxygenase. Journal of Bacteriology, 173: 5315-5327.

Zhang, X., Gao, P., Chao, Q., Wang, L., Senior, E., Zhao, L. (2004) Microdiversity of phenol hydroxylase genes among phenol-degrading isolates of *Alcaligenes* sp. from an activated sludge system. FEMS Microbiology Letters, 237: 369-375.

Zhou, N.-Y., Chan Kwo Chion, C. K., Leak, D. J. (1996) Cloning and expression of the genes encoding the propene monooxygenase from *Xanthobacter* Py2. Applied and Environmental Microbiology, 44: 582-588.

Zhou, N.-Y., Jenkins, A., Chan Kwo Chion, C. K., Leak, D. J. (1999) The alkene monooxygenase from *Xanthobacter* strain Py2 is closely related to aromatic monooxygenases and catalyzes aromatic monohydroxylation of benzene, toluene, and phenol. Applied and Environmental Microbiology, 65: 1589-1595.

Zhou, W., Guo, W., Zhou, H., Chen, X. (2016) Phenol degradation by *Sulfobacillus acidophilus* TPY via the *meta*-pathway. Microbiological Research, 190: 37-45.

Zhu, C., Zhang, L., Zhao, L. (2008) Molecular cloning, genetic organization of gene cluster encoding phenol hydroxylase and catechol 2,3-dioxygenase in *Alcaligenes faecalis* IS-46. World Journal of Microbiology and Biotechnology, 24: 1687-1695.

Zou, B., Huang, Y., Zhang, P. P., Ding, X. M., Op den Camp, H. J. M., Quan, Z. X. (2021) Horizontal gene transfer of genes encoding copper-containing membrane-bound monooxygenase (CuMMO) and soluble di-iron monooxygenase (SDIMO) in ethane- and propane-oxidizing *Rhodococcus* bacteria. Applied and Environmental Microbiology, 87: e0022721.