Microbial ecological approaches used to investigate DMSP production in Stiffkey salt marsh sediments

by

Beth Teresa Williams

A thesis presented for the degree of Doctor of Philosophy

University of East Anglia

School of Biological Sciences

May 2019

DECLARATION

I, Beth Teresa Williams, declare that this thesis is a record of my own work and that I composed it. It has not been submitted in any previous application for a degree in any other university. All quotations have been distinguished by quotation marks and all sources of information are specifically acknowledged.

ABSTRACT

Dimethylsulfoniopropionate (DMSP) and its catabolite dimethyl sulfide (DMS) are key marine nutrients, with roles in global sulfur cycling, atmospheric chemistry, signalling and, potentially, climate regulation. In the surface layer of salt marsh sediments DMSP concentrations are > three orders of magnitude higher than in the overlying seawater, an environment usually touted as being the most important site for DMSP production. A third of bacterial isolates from salt marsh pond sediment were found to produce DMSP (up to 160 nmol/mg protein) and, furthermore, many more novel DMSP-producing bacteria were identified after performing enrichment microcosm experiments for bacterial DMSP production. Most DMSP-producing isolates contained the *dsyB* gene, but several (*Alteromonas*, *Marinobacter* and *Novosphingobium*), lacked this reporter gene for DMSP synthesis. A *Novosphingobium sp*. MBES04 isolate produced DMSP via a novel bacterial methionine methylation pathway, and a bacterial methionine methyltransferase '*mmtN'* gene was discovered. BLASTp results revealed a diverse range of bacteria that contain it, and both alphaproteobacteria and actinobacteria within that group were shown to produce DMSP. DMSP-producing bacteria, *mmtN* abundance and *dsyB* transcripts were present in all tested seawater samples and *Tara* Oceans bacterioplankton datasets, but were far more abundant in marine surface sediment. Thus, we propose that surface marine sediments are environments with high DMSP productivity and that heterotrophic bacteria are likely important producers in these environments

ACKNOWLEDGMENTS

I would first like to thank my supervisor Jonathan Todd for his support, particularly in the last few months. His encouragement kept me going far past what I thought my limits were. Thank you Colin Murrell for your mentoring and input – despite how busy you are you always had time for me. Also Jennifer Pratscher, who happily went through bioinformatics with me many times, and Andy Smith who was always available.

Thank you to everyone in Lab 1.29 and Lab 1.32. Andy, you were always patient with me, and I really appreciate that at least someone shares my taste in music! Ana you have been a wonderful friend, PhD buddy and part-time therapist to me, and I miss our car journeys together. Ornella, your advice and friendship have been very important to me through the years. Many thanks to Ji, Peter, Jingli and Yanfen for being the loveliest people, you are all kind and cheerful and were always ready to help me, and thank you to Kasha for putting up with the terrible boss that I was – I think we make much better friends!

I want to thank (and also blame) John Munnoch for getting me into this PhD to begin with. Your unswerving loyalty, friendship and general hilariousness are things I am lucky to have experienced, and I genuinely believe I might have given up on many occasions without your encouragement. No acknowledgement would be complete without thanking my platonic soulmate Esme, who has listened to me complain and cry and rage for the past seven years, and hopefully will continue to do so for the next seventy. You're an amazing human being and I cannot express the joy that knowing you has brought me.

Many thanks to Amy and Sam, who I know are proud of me even though they called me a freak for being a science nerd, and especially to my parents for keeping me alive and sane in the last few months as I desperately tried to finish this thesis. I've always felt nothing but love and support from my family, and feel eternally lucky for that.

More thanks to my non-biological family, in Norwich and on the South Coast – from Florence, Beckie and Naomi for being there through my formative years, to all the people at UEA who believed that I would succeed, and commiserated with me if I didn't, and to Chloe and Alec for being wonderful housemates. Thanks to everyone in Saving Faith, which has been a haven of peace and a place of happiness for me. I would especially like to thank Annette, Abigail and Becky for being the best support system a girl could need this year.

Finally, I would like to thank God and the strength that He gives me, without which I would almost certainly fall apart.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

·OH	hydroxyl radical
µM	micromoles per litre or micromolar
16S rRNA	16S ribosomal RNA in bacteria and archaea
3-HP	3-hydroxypropionate
ABC	ATP binding cassette
APS	adenosine phososulfate
aq	aqueous
BCCT	betaine choline carnitine transport
BLAST	basic local alignment search tool
bp	base pairs
BSA	bovine serum albumin
C	carbon
CCN	cloud condensation nuclei
CLAW	Robert **C**harlson, James **L**ovelock, Meinrat **A**ndreae and Stephen **W**atson hypothesis
CoA	Coenzme A
CO	carbon monoxide
CO_2	carbon dioxide
DMS	dimethyl sulfide
DMSO	dimethylsulfoxide
DMSHB	4-dimethylsulfonio-2-hydroxybutyrate
DMSP	dimethylsulfoniopropionate
DMSP-a	3-dimethylsulfoniopropylamine
DMSP-ald	3-dimethylsulfoniopropionaldehyde
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotide triphosphate
EDTA	ethylendiametetraacetic acid

Fe	iron
FH_4	tetrahydrofolate
g	gram
GBT	glycine betaine
GC	gas chromatography
GOS	Global Ocean Sampling expedition
H_2	hydrogen
H_2S	hydrogen sulfide
h	hour
HGT	horizontal gene transfer
IMG	Integrated Microbial Genomes
IPTG	isopropyl β-D-1-thiogalactopyranoside
JGI	Joint Genome Institute
L	litre
LB	Luria Broth
m	meter
M	molar (concentration)
MaS	malonate semialdehyde
MBM	minimal basal media
MeSH	methanethiol
Met	methionine
mM	millimolar
MMETSP	Marine Microbial Eukaryote Transcriptome Sequencing Project
MMPA	methyl mercaptopropionate
MTA	methylthioacryloyl
MTHB	4-methylthio-2-hydroxybutyrate
MTOB	4-methylthio-2-oxobutyrate
MTPA	3-methylthiopropanamine
MPA	3-mercaptoproprionate

MW	molecular weight
N	nitrogen
NaOH	sodium hydroxide
NBP	nucleotide binding protein
NCBI	National Center for Biotechnology Information
nm	nanometers (wavelength)
nM	nanomolar
NO_3	nitrate
NTC	no template control
O_2	oxygen
OM-RGC	Ocean Microbial Reference Gene Catalogue
OTUs	operational taxonomic units
PCR	polymerase chain reaction
PSU	practical salinity units
pmol	picomole
RT	reverse transcription
S	sulfur
SAM	S-adenosyl methionine
SAH	S-adenosyl-L-homocysteine
SBP	substrate binding protein
SDS	sodium dodecyl sulfate
SMM	S-methylmethionine
SO_2	sulfur dioxide
SO_4^{2-}	sulfate
TBE	tris-borate-EDTA
TMP	transmembrane protein
Tris-HCL	tris(hydroxymethyl)aminomethane hydrochloride
UV	ultraviolet
V	voltage

CHAPTER 1

INTRODUCTION

1 INTRODUCTION

1.1 The sulfur cycle

Sulfur is the ninth most abundant element in the universe, and is essential to life on earth (Sievert et al. 2007). Although the majority of the sulfur on earth is fixed in mineral and rock form, it also makes up roughly 1 % of the dry weight of the biomass of an organism, in the form of amino acids such as cysteine and methionine, as well as playing a role in coenzymes and metalloproteins (Sievert et al. 2007). It exists in several different states, the most stable of which is an inorganic sulfate (SO_4^{2-}) **(Figure 1-1)**, and also in various reduced and organic forms (Sievert et al. 2007). Although all organisms require sulfur to survive, not all are able to use it in its inorganic form. Microorganisms are able to use it through a process called 'assimilation' (Oduro et al. 2012), where sulfate is integrated into different organosulfur compounds, including methionine and DMSP **(Figure 1-1)**. Animals that are unable to utilise inorganic sulfur are therefore dependent on these preformed sulfur compounds (Sievert et al. 2007), which can also be used as electron acceptors or donors in sulfur reduction/oxidation reactions.

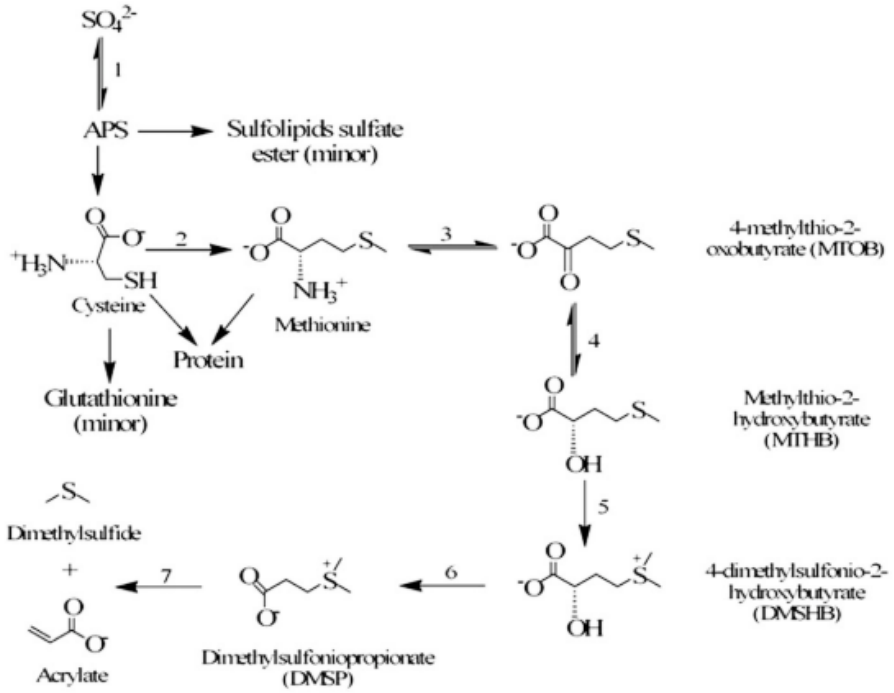

Figure 1-1: The biosynthetic pathway of DMSP/DMS by marine algae through assimilatory sulfate reduction, via methionine enzymatic biotransformation. The reaction processes involved in seawater sulfate assimilation by marine algae species are as follows: (1) carrier-bound sulfate reduction, (2) transsulfuration to methionine biosynthesis, (3) transamination, (4) reduction, (5) methylation, (6) oxidative decarboxylation, and (7) cleavage/degradation. (Oduro et al. 2012)

This assimilation from sulfate to biogenic sulfur and back takes place several times in the sulfur cycle **(Figure 1-2)**. Initially, sulfur dioxide is released from rocks in terrestrial environments through weathering (Schäfer et al. 2010), and oxidises in the air to become sulfate. This is assimilated by various microorganisms and plants and turned into different organosulfur molecules (Andreae 1990), which are consumed by animals that then use the biogenic sulfur, releasing it as sulfate into the soil during death and decomposition. Other emitters of sulfur include volcanic eruptions and biomass burning (Malin 1996). Sulfate eventually ends up in the oceans through deposition (fallout from the atmosphere), as well as in run-off from lakes and rivers (Schäfer et al. 2010). Here it is assimilated into cysteine, methionine, dimethylsulfoniopropionate (DMSP) and, finally dimethyl sulfide (DMS) (Stefels 2000) **(Figure 1-3)**.

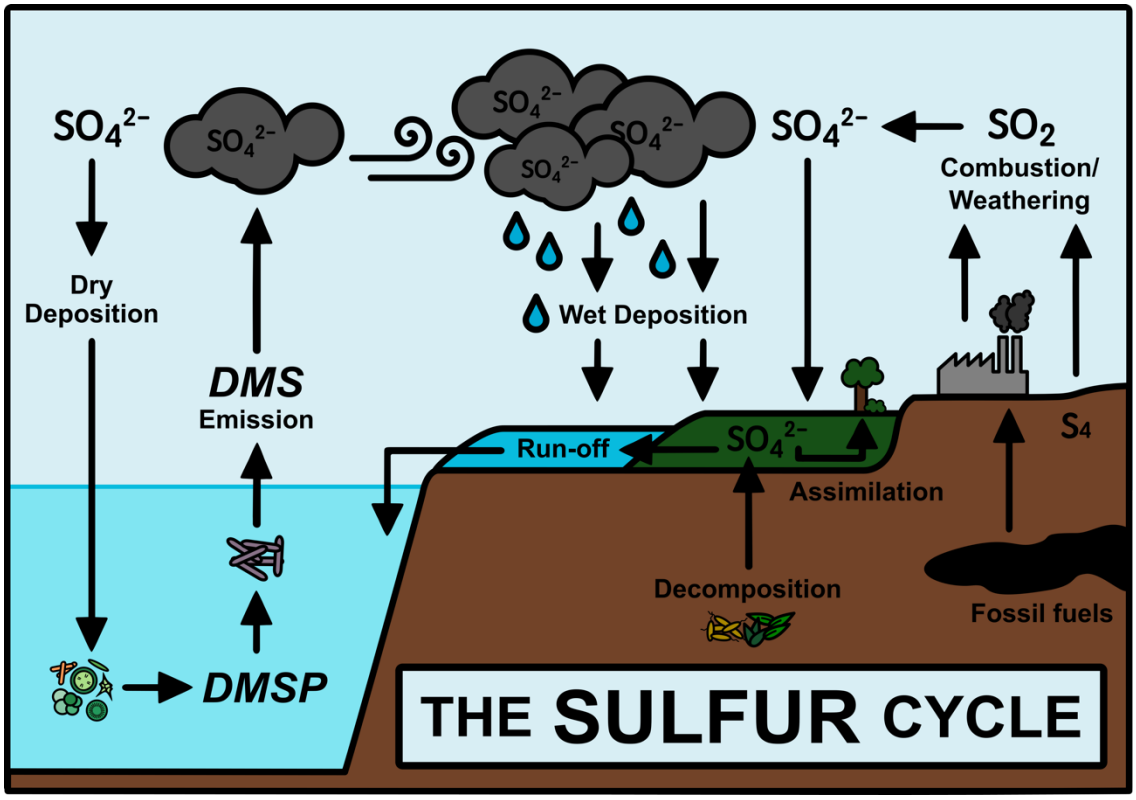

Figure 1-2: The cycle of sulfur in the environment. The majority of the sulfur on earth is trapped within rocks and fossil fuels, but the fraction that is released through combustion of fossil fuels and weathering is cycled through both terrestrial and marine environments. It first oxidises in the air and becomes SO_2, which reacts in the atmosphere, becoming SO_4^2. This is returned to earth as dry deposition, assimilated by plants and microorganisms and reduced to various organosulfur compounds. These are utilised by animals and other bacteria, and released again through decomposition. SO_4^{2-} reaches the ocean through run-off in rivers and through dry deposition from the atmosphere. In the oceans SO_4^{2-} is assimilated and becomes methionine. Marine algae and bacteria turn methionine into DMSP, which is released and broken down to DMS, and emitted into the atmosphere where it oxidises and forms clouds, which bring biogenic sulfur to land through wet deposition. This is the only known movement of biogenic sulfur from the oceans to land.

1.1.1 The assimilation of sulfur

The assimilation of sulfur resulting in DMSP and DMS first requires the production of L-methionine, from which all DMSP is produced. L-Met is formed from cysteine, a product of microbial sulfate assimilation **(Figure 1-1)** (Leustek & Saito, 1999). The uptake of sulfate for this reaction requires energy, which is supplied by ATP (Stefels, 2000). It passes through the cytoplasm and into the chloroplasts, where it is reduced and eventually forms a free sulfide (Stefels 2000). This sulfide combines with *O*-acetylserine (a product of glycolysis), and results in cysteine and acetate (Giovanelli, 1990). Cysteine itself has multiple roles inside the cell, in particular the *de novo* production of methionine. This pathway involves the transfer of the cysteine thiol group to *O*-phosphohomoserine, forming homocysteine, which is then methylated to make methionine (Stefels 2000). This methionine can subsequently be used in various other reactions via its methyl group and the molecule *S*-adenosylmethionine (AdoMet) that acts as a methyl donor.

Throughout all stages of the sulfur cycle these biogenic sulfur compounds are well utilised by marine organisms. Around 10 % (30 million tonnes) of the DMS produced in the ocean is released to the atmosphere (Oduro et al. 2012), and because of the large surface area of the oceans from which it is released, it accounts for around 50 % of the biogenic sulfur (Andreae 1990), and one tenth of the total sulfur flux in the atmosphere (Jackson and Jackson, 2000). Indeed, allowing for a seasonal cycle, the global annual DMS flux from the oceans can range from 13 to 37 TgS yr-1 (Kettle & Andreae 2000). This DMS in the atmosphere becomes sulfate again, in the form of dimethylsulfoxide (DMSO) and other molecules including sulfuric acid, which form cloud condensation nuclei (CCN). Clouds form and blow ashore, returning this large volume of sulfur to the terrestrial environment as deposition **(Figure 1-2)** (Sievert et al. 2007). The land to which this sulfur returns is sulfur-depleted, making it an important step in the cycle. When it falls, the fresh supply of this nutrient positively affects the surface, increasing productivity and therefore weathering, thus completing the cycle and bringing more nutrients into the ocean environment through run-off (Charlson et al. 1987).

1.1.2 The CLAW hypothesis

The clouds produced by DMS emissions not only complete the cycle of sulfur by returning it to terrestrial environments **(Figure 1-2)**, but it has also been theorised that their albedo effect is part of a feedback loop controlling local climate, termed the CLAW hypothesis after the authors who first postulated it (Charlson et al. 1987). This feedback loop was suggested because it had already been established that DMS is one of the major sources of CCN, and therefore the formation of clouds could be regulated by controlling the release and oxidation of DMS (Charlson et al. 1987). Blooms of phytoplankton were

found to produce higher amounts of DMS in warmer conditions (Charlson et al. 1987), likely because increased solar radiation leads to better growth (Schäfer et al. 2010). Higher DMS means increased CCN and therefore cloud formation, reflecting radiation away from the surface and cooling it, which then leads to a decrease in growth and production, causing the cloud cover to ease off and allow more radiation again, maintaining balance (Schäfer et al. 2010). This theory, while often referenced, has not been fully validated, and even though there is evidence for levels of DMS being driven by light dosage (Vallina & Simó 2007), it is now widely thought to be unlikely, or at the very least more complex than previously suggested (Quinn & Bates 2011), namely due to the existence of other sources of CCN that are not DMS-derived.

The amount of sulfur released to the environment through the DMSP/DMS cycle is globally dwarfed by anthropomorphic production in this day and age (Malin 1996), but on a local scale it is still a major player, particularly in the marine environment, and is worthy of being considered a key step in the overall cycling of sulfur (Yoch 2002). The distribution and abundance of these sulfur compounds in the environment, specifically DMS and DMSP, is largely a result of the microorganisms that produce them and break them down, driving biogeochemical cycles on a global scale in the process. Understanding the activity of these microorganisms leads to a better understanding of the cycle as a whole.

1.2 The importance of DMSP

DMSP is not only an important molecule in the sulfur cycle – it is a globally significant organosulfur compound produced by a wide range of marine organisms, with several petagrams of it being released per year (Galí et al. 2015). As previously described, it has been shown to play a significant role in global sulfur cycling (Sievert et al. 2007), and is also a signalling molecule (Seymour et al. 2010) and a key nutrient source for marine organisms (Curson *et al.* 2011b), with osmoprotectant and anti-stress properties also suggested (Stefels 2000). It is the major precursor for the volatile sulfur gas, DMS, which is itself an environmentally important climate-active compound that diffuses into the atmosphere, affecting the climate through the albedo effect (Shaw 1987), completing the global sulfur cycle through the formation of CCN and precipitation (Stefels et al. 2007), and acting as a nutrient source and signalling molecule as well.

1.2.1 DMSP structure

DMSP is a five-carbon compound, synthesised from the amino acid methionine. As the name suggests, its structure includes a twice-methylated sulfur molecule and a carboxylate in the form of propionate (**Figure 1-3**). DMSP is a zwitterion, meaning that the molecule contains both a negative and positive charge, one at either end. The methylated sulfur is positively charged, and the negative charge is on the oxygen in the carboxylate group (Sunda et al. 2002).

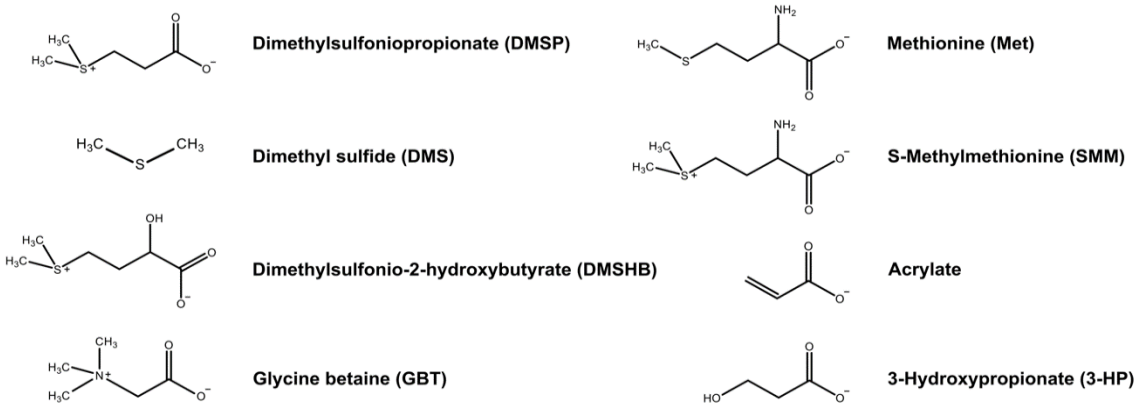

Figure 1-3: DMSP and selected compounds involved in DMSP cycling, including significant intermediates involved in its production (DMSHB and SMM), the products of its breakdown (DMS, Acrylate and 3-Hydroxypropionate), and also the known osmolyte, GBT, demonstrating the similarity in structure as a nitrogen-based homolog to DMSP. The two methyl groups attached to the sulfur class DMSP as a C1 compound, and also mean that a DMS molecule is easily removed from the compound during catabolism. Both DMSHB and SMM also possess the twice-methylated sulfur, and can release DMS under particular conditions.

Elements of this structure go some way to explaining the function of DMSP in the environment: DMS is easily released from the rest of the compound via the cleavage of the sulfur-carbon bond (Curson, et al. 2011b). The methyl groups are accessible sources of carbon, and the overall structure is very similar to other widely-studied compounds such as glycine betaine (GBT) (Otte et al. 2004), differing only in the replacement of the sulfur molecule with nitrogen. GBT is known to play a role in osmoregulation, suggesting that DMSP may also have osmoregulatory function.

DMSP is synthesised or transported into organisms from the environment because it and its catabolites are thought to provide a number of advantages to the organism **(Figure 1-4)**. As previously stated, it acts as a key source of both sulfur and carbon to the organisms that break it down (Simó et al. 2009), and is a potential osmoprotectant, balancing the cell against high saline conditions. Further suggested functions of DMSP/DMS in organisms include use as a signalling molecule (Seymour et al. 2010) and

protection against multiple stresses including oxidative stress and UV damage (Sunda et al. 2002). Research into the synthesis, function and cycling of these molecules is key because they have major impact in the individual organisms that produce them and also in the wider environment and the global sulfur cycle.

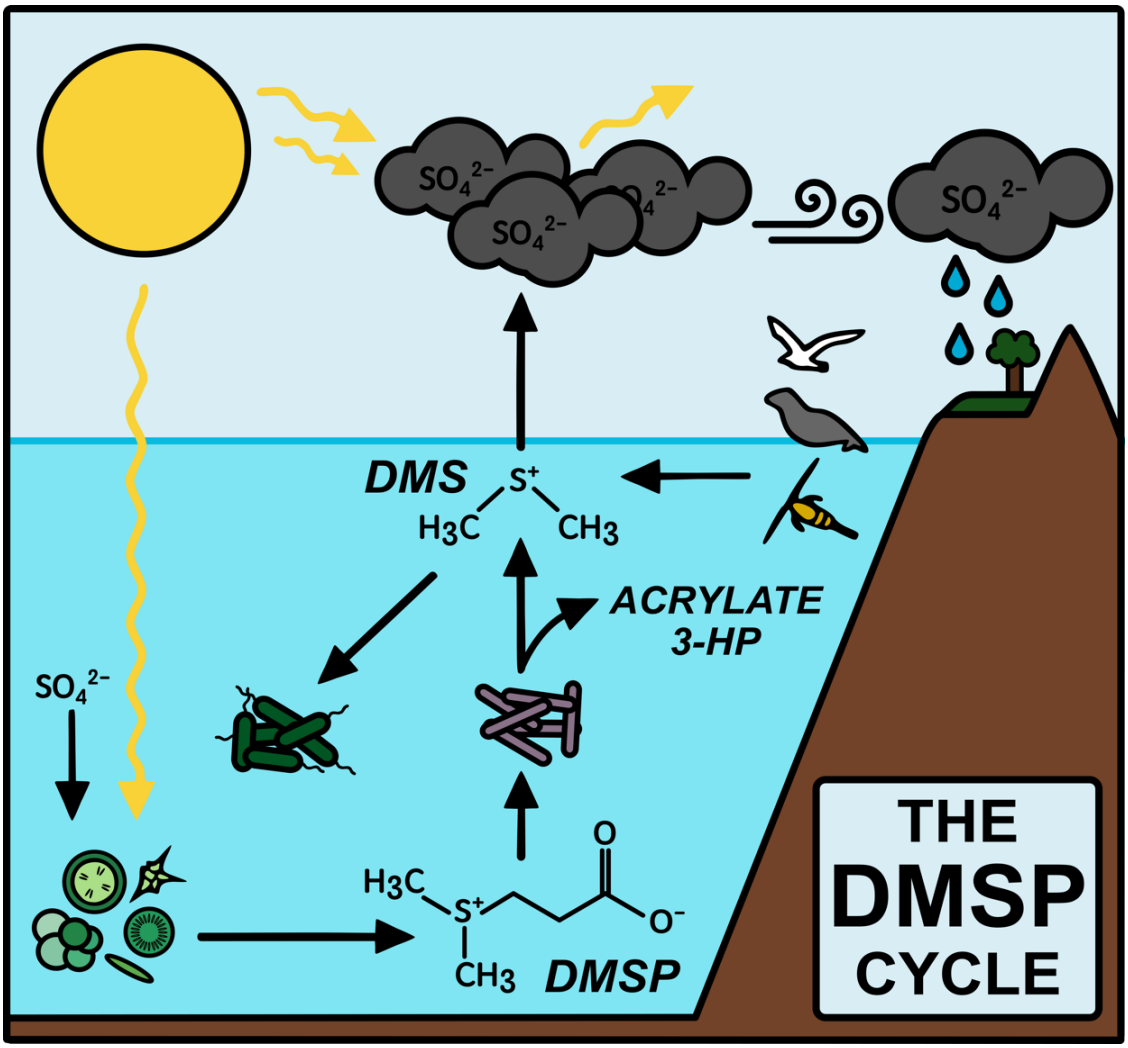

Figure 1-4: The cycle of DMSP production and breakdown in the marine environment, a major step in the sulfur cycle. DMSP is produced in the ocean by marine organisms such as algae and bacteria, and released into the surrounding water upon cell death or lysis. It is taken up by bacteria that lyse it and release DMS (alongside acrylate or 3-HP). The oceanic DMS is used as a carbon source by marine bacteria, and when it reaches the surface 10 % is released as a gas. In both forms it is a chemoattractant for organisms including zooplankton, harbour seals and petrels. In the atmosphere it oxidises to either DMSO or SO_4^{2-} aerosols, which form CCN. These increase the albedo effect (the reflection of the radiation from the sun) which causes local cooling of the waters. When these clouds precipitate they bring biogenic sulfur back to the terrestrial environment.

1.2.2 DMSP as a source of nutrients

Between 30 – 90 % of DMSP released into the ocean is immediately metabolised by marine bacteria (**Figure 1-4**) (Seymour et al. 2010). This is because, as previously mentioned, DMSP and DMS are both excellent sources of nutrients, and therefore energy, in bacteria, bacterioplankton and phytoplankton (Simó et al. 2009), and are well utilised as such. In fact, there are no known single compounds that contribute as much sulfur and carbon to the food web as DMSP does (Yoch 2002). In DMSP-producing algae DMSP can comprise anywhere between 50 – 100 % of the organic sulfur in the cell (Stefels, 2000), and it is often the favoured source of sulfur for marine bacteria such as those in the SAR11 clade (Tripp et al. 2008), who also use it as a carbon source. This is particularly important for this clade because several of them lack the complete set of assimilatory sulfate reduction genes and therefore rely exclusively on the uptake of externally reduced sulfur such as DMSP or methionine (Tripp et al. 2008).

DMSP also contributes to around 10 % of all the fixed carbon in the ocean (Howard et al. 2006), and supports up to 13 % of the bacterial carbon demand in surface waters (Kiene et al. 2000), meaning that it is one of the most substantial single sources of labile carbon in the surface waters. Bacteria possess many ways in which to breakdown DMSP and gain access to the nutrients – not only are there multiple DMSP cleavage enzymes (*ddd* genes) that are very effective in accessing carbon (see below), but the ability to break DMSP down via the demethylation pathway (shown by the presence of the gene *dmdA*) is thought to exist in three in every five bacterioplankton cells tested in the GOS data set (Howard et al. 2008), showing that DMSP is a hugely beneficial carbon source to these surface bacterioplankton. Indeed, in analysis of the OM-RGC database (Sunagawa et al. 2015), an estimated 20 % of the bacterial species in the metagenomic database contain a *ddd* gene (Curson et al. 2018).

1.2.3 DMSP as an antioxidant

In addition to providing a source of carbon and sulfur to various bacteria, DMSP also plays several other suggested roles in the organism and the environment. Many of these functions are predicted based on the observed increase in DMSP production or intake by the organism in response to changed stimuli (Stefels 2000; Sunda et al. 2002; Curson et al. 2017). Production of DMSP is often increased as a response to various oxidative stresses on the cell (Sunda et al. 2002). Oxidative stress occurs because of an imbalance between the production of free radical reactive oxygen species (ROS), and the ability of an organism to remove them (Birben et al. 2012). If left in the cell these free radicals can cause significant damage to cell structures, including proteins and DNA (Birben et al. 2012). Certain environmental conditions increase the amount of oxidative

stress experienced by an organism, including the limitation of particular molecules, such as CO_2 and Fe, an increase in UV radiation or copper levels, or the presence of hydrogen peroxide (Yost et al. 2010), all of which cause the formation of highly reactive hydroxyl radicals in the cell that can damage lipids, proteins and nucleic acids. There are also several natural processes that take place such as respiration and photosynthesis that also create ROS. DMSP and the products of its breakdown (DMS and acrylate) act as protection from these stressors (Yost et al. 2010) by 'scavenging' the harmful hydroxyl radicals by rapidly reacting with them (Sunda et al. 2002).

There are several other conditions that have also been found to increase the levels of DMSP production, in addition to increased oxidative stress. It has been speculated that DMSP acts as a cryoprotectant (antifreeze), protecting protein integrity from excessive temperature decrease, as the levels of DMSP in Antarctic macroalgae increase as the temperature is lowered (Kirst et al. 1991; Ko et al. 1994). As previously discussed, DMSP has a similar structure to GBT, and therefore is likely to have similar functions to this molecule. Furthermore, polar algae has been found to contain significantly higher DMSP levels compared to tropical species (Karsten et al. 1992), and could be acting as an antifreeze by keeping the cytoplasm liquid, and may even be protecting proteins through specific molecular interactions. In Karsten et al. (1996) DMSP was found to not only protect the cells from damage during freezing, but also noticeably improved the activity of specific enzymes, even at 0°C, compared to the controls.

1.2.4 DMSP as a compatible solute

It has also been suggested in numerous instances that DMSP is an osmolyte (Dickson & Kirst 1986; Kirst et al. 1991; Karsten et al. 1992), also known as a constitutive compatible solute (Stefels 2000). DMSP has even been referenced as one of the dominant compatible solutes in marine algae (Kempf & Bremer 1998), although it is usually used in combination with other solutes (Kirst, 1996). In high salinity conditions several organisms have been shown to produce greater amounts of DMSP (Zhuang et al. 2011; Curson et al. 2017), and those that are unable to produce it accumulate it from the environment at a much higher rate than in lower salinity conditions (Cosquer et al. 1999). An increase in salinity is problematic for unicellular organisms because they often lack the cell membrane structure that helps prevent desiccation due to the loss of water from the cell via osmosis (Kempf & Bremer 1998).

As microorganisms are unable to actively transport water molecules into the cell to maintain turgor (Kempf & Bremer 1998), a different solution is required, in the form of production and/or accumulation of osmoprotectants from the environment. These are highly soluble compounds that balance changes in external osmotic potential without

disturbing the functioning of cellular proteins (Kirst et al. 1991). Both DMSP and GBT are effective osmolytes, with DMSP enhancing the salinity tolerance of organisms when present at even nanomolar concentrations (Cosquer et al., 1999). These molecules aid osmotic acclimatisation, adjusting the potential of the cell to match the outside conditions, and enabling it to maintain optimum cell volume and turgor (Stefels 2000). Furthermore, these molecules have a net neutral charge, and can accumulate at high concentrations without affecting cellular processes such as DNA replication, with *Spartina alterniflora* reported as having upwards of 29 µmol/g fresh weight (Kocsis et al. 1998) (Kempf & Bremer 1998). Compatible solutes, including DMSP, are also able to stabilise protein structures and metabolic pathways, protecting them from the adverse effects of high salinity, such as inhibition or denaturation (Arakawa & Timasheff 1985). It has been speculated that the biosynthesis of DMSP by marine organisms could have arisen during the last ice age to combat the more highly saline conditions of the oceans at that time (Charlson et al. 1987). This theory could be supported by the fact that DMSP is also a cryoprotectant, giving organisms better survival rates in the lower temperatures of the water.

1.2.5 DMSP as a chemoattractant

DMSP is utilised as a chemoattractant for several species of bacterioplankton (Seymour et al. 2010) and proteobacteria (Miller et al. 2004), enabling them to find and assimilate it so that it can be catabolised. When DMSP is released from the marine organisms that synthesise it, it is in limited supply, so the ability to sense DMSP confers great advantage to those that possess it (Miller et al. 2004). It is possible that this is not an intended function for the organisms that are synthesising it, and is instead something that marine bacteria have evolved to exploit (Seymour et al. 2010), or it could be that a bacterium-dinoflagellate interaction could be of benefit to both the bacterium and the host (Miller et al. 2004). In either scenario, this chemotaxis and subsequent breakdown of DMSP to DMS plays an important role in the global sulfur budget (Zimmer-Faust et al. 1996). DMSP is also detected by planktivorous reef fishes, sea urchins, various sea birds and harbour seals (Seymour et al. 2010) that use it as an indicator of feeding activity (Debose et al. 2008). It acts as an indirect foraging cue for higher organisms, indicating the presence of algal blooms to reef fish, and consequently, the presence of reef fish to higher predators such as birds and seals (Debose et al. 2008).

1.2.6 DMSP as an antimicrobial/antigrazing molecule

Another suggested purpose for the production or accumulation of DMSP by various marine organisms is that of an antimicrobial or anti-grazing role (Wolfe et al. 1997). It is thought that bacteria and phytoplankton can use the catabolism of DMSP to produce

acrylate as a deterrent to predators such as protozoan herbivores and copepods (Wolfe et al. 1997). This is because acrylate has antimicrobial activity at high concentrations, and can inhibit growth of various bacterial species, depending on their sensitivity (Slezak et al. 1994). It has been hypothesised that the catabolism of DMSP to acrylate is actually activated by grazing (Wolfe et al. 1997; Strom et al. 2003), with acrylate only being produced once the algae was ingested. In these studies it was observed that even though acrylate did not appear to have detrimental effects on the grazing protozoa, other, non-DMSP-containing prey were consumed preferentially and entirely over the DMSP-producing algae (Wolfe et al. 1997). This suggests that the bacterial catabolism of DMSP to DMS may not be exclusively driven by a nutritional need, but could also be as a defensive action (Curson et al. 2008).

It has also been suggested that DMSP can act as a methyl donor during enzymatic transmethylations (Kiene & Taylor 1988).

1.3 Dimethyl Sulfide

DMSP is the major biogenic source of DMS in the environment, which is produced at around 300 million tonnes per year (Curson *et al.* 2011b) via the DMSP lyase pathway (see below). The catabolism of DMSP to DMS is an environmentally significant reaction (Kiene et al. 1999), and DMS plays several important roles in the ocean. It acts as a nutrient source for marine bacteria and it is a volatile sulfur compound that readily diffuses through the sea surface to the air (Kettle et al. 1999). Around 10 % of DMS that is produced in the ocean is transferred to the atmosphere as a gas (Kettle & Andreae, 2000), and in its gaseous form it plays a key role in the sulfur cycle, acting as the main source of natural biogenic sulfur returning to land (Sievert et al. 2007). It is also a small-scale chemoattractant for foraging seabirds (DeBose & Nevitt 2008), and is purported to play a role in local climate control and feedback loops (Charlson et al. 1987).

The bulk of DMS released into marine environments is a result of the catabolism of DMSP (Reisch *et al.* 2011b), although recently it was also found to be produced through the methylation of methanethiol (MeSH) (Carrión et al. 2015), particularly in terrestrial environments, which suggests that the production of DMS is a lot more prolific than previous projections have found. It can also be produced by the reduction of DMSO under anoxic conditions (Schäfer et al. 2010).

1.3.1 DMS as source of nutrients

Of the dissolved DMS in the marine environment, the majority is removed by microbial activity (Kettle & Andreae 2000). DMS is another C1-sulfur compound and is

therefore an excellent source of nutrients, both carbon and sulfur (De Bont et al. 1981). Consequently, much of the DMS in the oceans is degraded by microbial activity. Some species of bacteria are able to grow using DMS as the sole carbon source, and several other microorganisms isolated from a wide range of environments have also been shown to have the ability to degrade it (Schäfer et al. 2010). DMS is degraded through one of two pathways, with the initial oxidation being carried out by either a DMS monooxygenase (De Bont et al. 1981), or a possible methyltransferase (Visscher & Taylor 1993). The oxidation step in the monooxygenase pathway results in the production of formaldehyde and methanethiol, which can then be degraded to formaldehyde, hydrogen peroxide and sulfide (Schäfer et al. 2010). The methyltransferase pathway was theorised when DMS degradation was found to occur without the requirement of oxygen (Visscher & Taylor 1993; Schäfer et al. 2010), with the methyl group being transferred to an acceptor and then further oxidised, leaving methanethiol which is then degraded in the same way as the monooxygenase pathway. DMS can also be photochemically oxidised to DMSO, which can subsequently be used as a carbon source (De Bont et al. 1981).

1.3.2 DMS in the atmosphere

When DMS is in its gaseous form it plays several other roles in the environment. It is a chemoattractant for multiple organisms, including zooplankton (Steinke et al. 2006), harbour seals (Kowalewsky et al. 2006) and sea birds (Nevitt et al. 1995). It is thought that this chemotaxis towards DMS offers the same advantages in terms of foraging cues that chemoattraction to DMSP does (Nevitt et al. 1995).

Once the DMSP is released as a gas into the atmosphere it oxidises to form various products including DMSO and sulfate aerosols such as SO_2 (Malin 1996). These sulfate aerosols are the sulfate particles (SO_4^{2+}) that become CCN (Kettle & Andreae 2000), around which the water droplets condense and form clouds (see above) (**Figure 1-4**). DMS-derived particulates are not the sole source of CCN over oceanic environments, but they are still one of the major contributors to them (Quinn & Bates 2011). As previously mentioned, they cause local cooling through the albedo effect, and are also a vital step in the sulfur cycle once they blow ashore and precipitate, returning biogenic sulfur to terrestrial environments through atmospheric deposition (Sievert et al. 2007).

1.4 DMSP production

Since its purification from *Polysiphonia fastigiata* in 1948 (Challenger & Simpson 1948), it has been thought that DMSP was exclusively synthesised by marine eukaryotes (Dickschat et al. 2015). These include algae and single-celled marine phytoplankton

(Stefels 2000), as well as more complex organisms such as corals (Raina et al. 2013), and several angiosperms (Dickson et al. 1980; Hanson et al. 1994). *Wollastonia biflora* and *Spartina sp.* are among the angiosperms found to produce DMSP (Otte et al. 2004), making them somewhat of an exception, as they are terrestrial rather than marine (although still in saline environments). More surprisingly, DMSP production was recently found to occur within several bacterial species (Curson et al. 2017), deepening our understanding of how widespread this ability is.

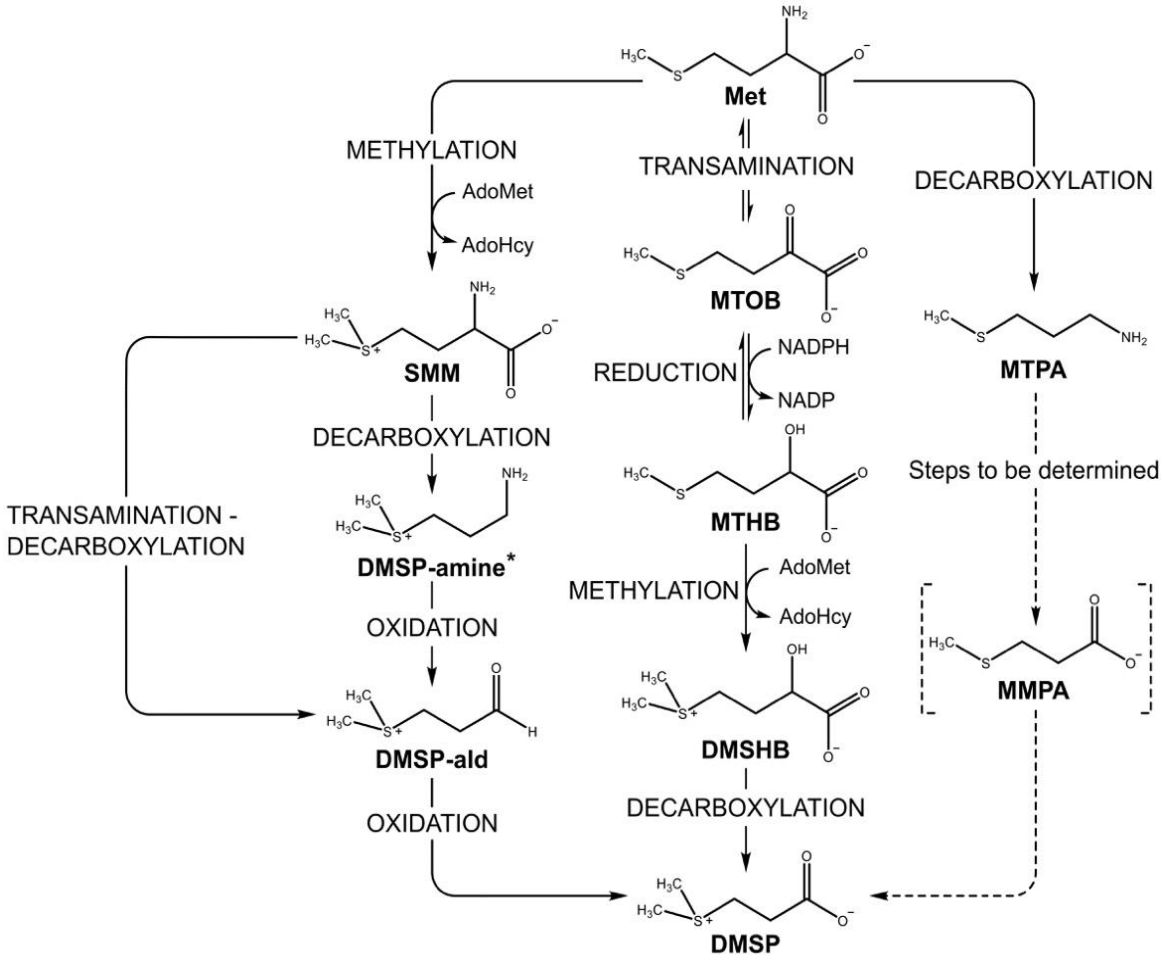

Figure 1-5: The three routes of DMSP production from methionine, named after the first step in the pathway. The methylation pathway is utilised by higher plants, the transamination pathway is used by macroalgae and phytoplankton, and the decarboxylation pathway discovered in one dinoflagellate thus far. The methylation pathway has been found to vary somewhat between Compositae (such as Wollastonia biflora) and Gramineae (Spartina alterniflora and sugarcanes) (Dickschat et al. 2015), with an additional intermediate (DMSP-amine) being produced during DMSP production in Gramineae. It is likely that there is an intermediate formed between SMM and DMSP-aldehyde in Compositae, but this has yet to be observed (Dickschat et al, 2015).

All known DMSP production pathways begin with methionine and then diverge to follow three different pathways, depending on the organism (Dickschat et al. 2015) (**Figure 1-5**). Apart from the initial and final molecules, there is no similarity between the three routes pictured in **Figure 1-5**. The generation of DMSP in higher plants (such as angiosperms) varies yet further between species (Rhodes et al. 1997; Kocsis & Hanson 2000). These differences suggest that the ability to synthesise DMSP from methionine is highly likely to have evolved independently, on at least two if not three separate occasions (Gage et al. 1997). If this is the case, then it implies that the ability to synthesise DMSP is an important function for marine organisms and possibly certain higher plants.

1.4.1 DMSP production in higher plants

One such pathway is that of DMSP production in higher plants. The ability to synthesise DMSP is not widespread among the higher plants, and only a few species of angiosperm thus far have been shown to possess it. It is somewhat of a surprise that these plants appear to have evolved this mechanism, as many of the roles that DMSP is thought to play in single-celled organisms do not appear to be as beneficial in higher plants. Those that are known to produce DMSP are *Wollastonia biflora* (Compositae), some *Spartina* species, and *Saccharum* (Gramineae) (Stefels, 2000). Although the synthesis route used by higher plants is generally thought of as one pathway (the methylation pathway), with several of the intermediates being shared between Compositae and Gramineae, the central steps differ enough to be significant (Dickschat et al. 2015). The initial step (**Figure 1-5**) involves the methylation of *L*-methionine to *S*-methylmethionine (SMM), using the previously mentioned methyl donor, AdoMet (Hanson et al. 1994). This reaction is catalysed by the enzyme S-adenosylmethionine:methionine S-methyltransferase (MMT) (James et al. 1995). This methylation takes place in the cytosol, and SMM is then transported into the chloroplast for the rest of the pathway (Trossat et al. 1996). The conversion of methionine to SMM is not specific to those plants able to produce DMSP – it is a commonplace reaction that seems to occur in all angiosperms (Mudd & Datko 1990; Kocsis & Hanson 2000), as SMM can then be used as the methyl donor for the production of methionine from homocysteine (Ranocha et al. 2000).

Instead, it is the following step in the pathway that appears to be specific to the production of DMSP (Kocsis & Hanson 2000), and, interestingly, it is at this point that the two pathways diverge from each other. In Compositae SMM is converted to DMSP-aldehyde via a pyridoxal 5′-phosphate (PLP) dependent transamination-decarboxylation reaction where the amino group is transferred to 2-oxoglutarate and the CO_2 is released through decarboxylation (Dickschat et al. 2015; Rhodes et al. 1997). As this appears to be a two-step reaction it is assumed that there is an intermediate formed before DMSP-aldehyde, but it is unstable and has yet to be isolated (Dickschat et al. 2015).

Despite both pathways removing the amino and carboxyl groups from SMM to convert it to DMSP-aldehyde, the Gramineae use a different method to carry this out, including an additional intermediate that has been identified as DMSP-amine (Kocsis et al. 1998) (**Figure 1-5**). Firstly, SMM undergoes a PLP-catalysed decarboxylation reaction to form DMSP-amine (Kocsis & Hanson 2000). This intermediate is then converted to DMSP-aldehyde by the removal of its amino group. However, in this case the reaction is not a transamination, but an oxidative deamination that is not dependent on PLP, possibly through an O_2-dependent amine oxidase enzyme (Dickschat et al. 2015). The fact that these steps differ rather dramatically suggests that this ability may have actually evolved independently, on two different occasions – once for Graminae and once for Compositae (Kocsis et al. 1998).

Once DMSP-aldehyde has been formed both pathways converge again. In addition to this, many plants other than those known to produce DMSP from methionine contain dehydrogenases that are able to convert DMSP-aldehyde to DMSP (Trossat et al. 1996). This means that these plants are able to produce DMSP only if supplied with DMSP-aldehyde, as opposed to synthesising it themselves. In *Wollasonia* at least, DMSP is formed via an oxidation reaction that is catalysed by a dehydrogenase using NAD as a cofactor (Trossat et al. 1996; Stefels 2000).

1.4.2 DMSP production in marine algae and diatoms

The second pathway featured in **Figure 1-5** is the transamination pathway. This pathway is used by marine algae, both red and green, as well as by diatoms (Dickschat et al. 2015). It was known that algae used methionine as the initial molecule for DMSP production (Kiene & Visscher 1987), and indeed, DMSP was first purified from an algal species (Charlson et al. 1987), but it took time for the intermediates involved to be fully elucidated. The first species in which these intermediates were properly identified was the Chlorophytum *Enteromorpha intestinalis* (now classified as *Ulva intestinalis*) (Gage et al. 1997), and DMSHB, a key intermediate, was also identified in other diverse phytoplankton species. It was shown that the process of DMSP production in algae followed an entirely different pathway to that used by higher plants, with no steps in common (Rhodes et al. 1997). As the name of this pathway suggests, the initial step is the transamination of methionine, leading to the formation of MTOB (4-methylthio-2-oxobutyrate) (Gage et al. 1997), instead of methylation resulting in SMM. This reversible reaction involves the transfer of an amine group from methionine to a keto acid, in this case 2-oxoglutarate, on which the reaction is dependent (Summers et al. 1998). Following this, MTOB is reduced, gaining a hydrogen to form MTHB (2-hydroxy-4-methylthio butanoic acid), and this is catalysed by an NAD(P)H-linked reductase (Summers et al. 1998).

The subsequent step methylates MTHB, adding a second methyl group to the sulfur molecule to produce DMSHB, via the methyl group donor S-Adenosyl-L-methionine (SAM) (Summers et al. 1998). DMSHB is itself an osmoprotectant, and is the final intermediate before DMSP is produced. The conversion of MTHB to DMSHB is thought to be the committing step in this pathway (Summers et al. 1998), as it is non-reversible and is only found in the context of DMSP production, whereas the other steps in the pathway also exist in species that are not able to synthesise DMSP. It would has also been shown that the regulation of DMSP production in algae is directly defined by the levels of DMSHB synthesis (Ito et al. 2011). The final step in this pathway is the conversion of DMSHB to DMSP which is an oxidative decarboxylation resulting in the loss of carbon dioxide (Dickschat et al. 2015).

It is likely that this pathway is the most widespread of the three described in **Figure 1-5**, as it has also been shown to be the one utilised by diatoms and coccolithophores (Dickschat et al. 2015), corals, *Acropora millepora* and *A. tenuis* (Raina et al. 2013) and, most recently, some marine bacteria (Curson et al. 2017). It was also the pathway from which the first DMSP-synthesis gene, *dsyB*, was identified by Curson et al, (2017) in the bacteria *Labrenzia aggregata*.

1.4.3 DMSP production in dinoflagellates

The final route by which DMSP is synthesised is through decarboxylation. This is the most understudied of the three pathways, and has only been observed in one dinoflagellate species so far (Kitaguchi et al. 1999). Dinoflagellates are some of the major DMSP producers in the marine environment (Miller & Belas, 2004), with some species containing intracellular concentrations of up to 0.5 M DMSP. They form large algal blooms, meaning that a significant amount of DMSP is released from them to the marine environment. Despite how widespread production in dinoflagellates is, the pathway used by them is less well-understood than other pathways, and only one intermediate and the enzyme responsible have been confirmed (Kitaguchi et al. 1999). It was shown that the dinoflagellate *Crypthecodinium cohnii* is able to synthesise DMSP from L-methionine via a decarboxylation reaction that produces MTPA (3-methylthiopropylamine), catalysed by a PLP-dependent L-methionine decarboxylase that was also purified (Kitaguchi et al. 1999). The rest of the pathway has yet to be determined, but it is thought that there is only one other intermediate missing, which has been predicted to be either MMPA (3-methylmercaptopropionate) (Dickschat et al. 2015) or MTP (3-methylthiopropionate) (Kitaguchi et al. 1999).

1.4.4 DMSP production in marine bacteria

Until recently, it was believed that DMSP is exclusively produced by marine algae, phytoplankton and the few terrestrial plants previously listed. However, work by Curson *et al,* (2017) showed that the mechanism for DMSP production also exists within a variety of bacterial species, including the marine alphaproteobacterium *Labrenzia aggregata* LZB033 in which it was first observed. This bacterial DMSP production mechanism appears to utilise the same pathway as that used by marine algae, the transamination pathway (**Figure 1-5**), as DMSP production was shown to increase in the presence of each of the intermediates (MTOB, MTHB, DMSHB), and LC-MS confirmed the production of each of them by LZB033 (Curson et al. 2017). Until this study, only the intermediates formed and enzymes involved in DMSP synthesis were understood, but in LZB033 the first DMSP-synthesis gene in any organism was identified. The gene, *dsyB* (DMSHB synthesis), was found to confer the ability to produce DMSP to *Rhizobium leguminosarum* when subcloned into it, and when it was disrupted in LZB033 it prevented its ability to synthesise DMSP. Furthermore, in the mutant an accumulation of MTHB was observed, suggesting that this gene encodes an enzyme that is able to catalyse the conversion of MTHB to DMSHB (the rate-limiting, committed step). Studying the protein revealed that DsyB is an acetylserotonin O-methyltransferase, belonging to the family of *S*-adenosyl methionine-dependent methyltransferases. When the amino acid sequence was searched for in online databases it was found that homologs of DsyB exist in roughly 80 other alphaproteobacterial species thus far (**Figure 1-6**), several of which were confirmed to be functional. Not only do bacteria possess the ability to synthesise DMSP, but it is more widespread than originally thought.

From this BLAST it was observed that a *dsyB* homolog exists in several marine phytoplankton, including a *Chrysochromulina tobin*. These were termed *DSYB*, for *dsyB* in eukaryotes, and it was found that it exists in a variety of species, from macroalgae, diatoms, prymnesiophytes and prasinophytes (Curson et al. 2018). The function of these genes was confirmed in multiple species, including *Symbiodinium microadriaticum*, *Fragilariopsis cylindrus* and *Prymnesium parvum.*

This discovery was significant because it is the first identified DMSP-producing gene in any eukaryote, and these are some of the most significant contributors to the global DMSP concentrations (Keller et al. 1989). When studying the evolution of the sequences of both prokaryotic and eukaryotic DsyB/DSYB proteins, it was concluded that the first DsyB gene clade was alphaproteobacterial. Therefore, it is theorised that DsyB originated in prokaryotes and was transferred to eukaryotes, through one of two processes – either through the same process by which mitochondria of alphaproteobacterial origin became part of eukaryotic cells (endosymbiosis), or, more recently, through horizontal gene

transfer (HGT) of DsyB back and forth. When DSYB was imaged through immunogold labelling in *P. parvum* it was found to localise most strongly in the chloroplasts, giving more weight to the suggestion that DMSP plays a role as an antioxidant, protecting against oxidative stress (Curson et al. 2018).

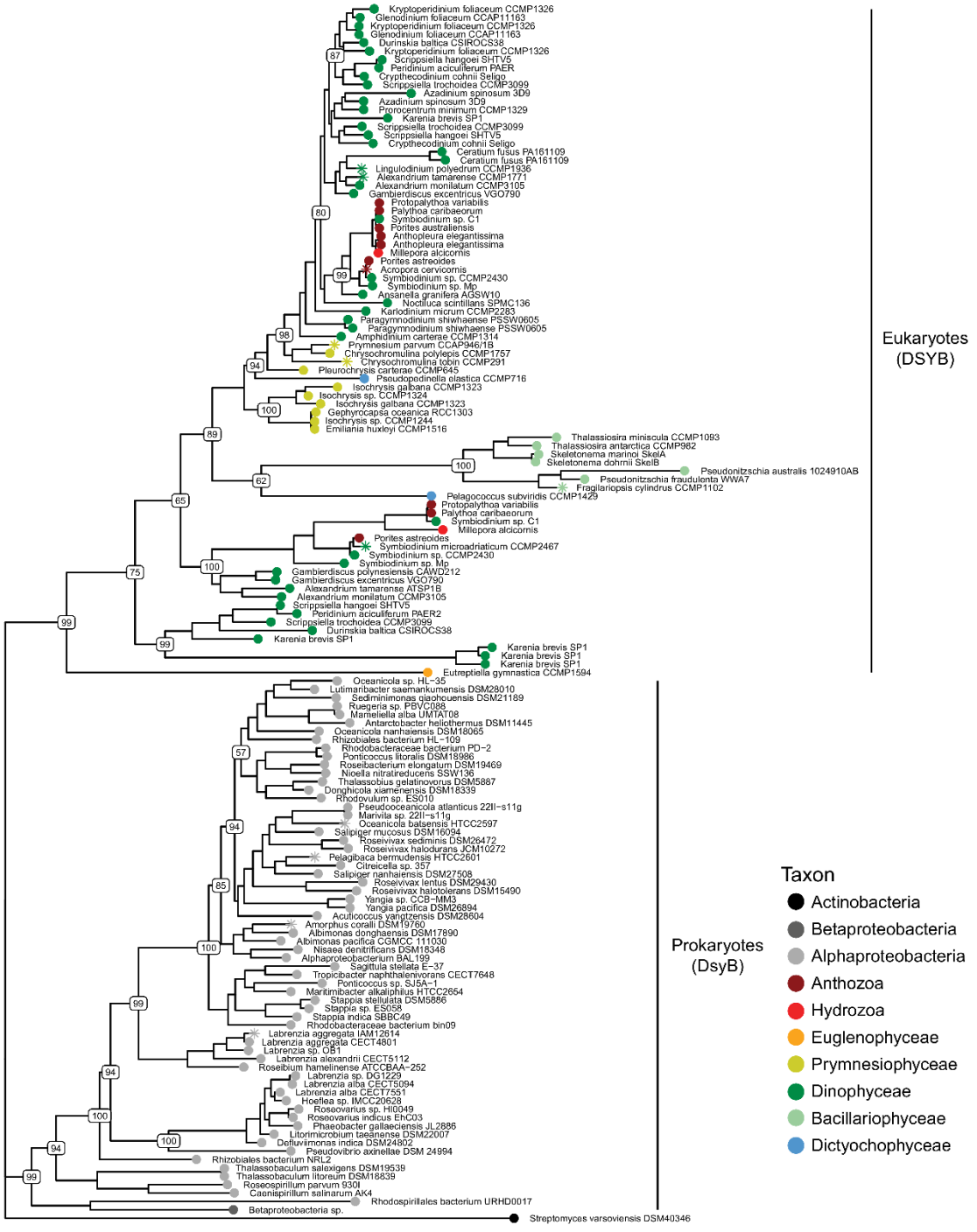

Figure 1-6: A maximum likelihood tree of the 145 known DsyB/DSYB proteins (Curson et al, 2018) identified from sequences in the NCBI, JGI IMG and iMicrobe MMETSP databases. Colour code denotes species according to taxonomic class (see key) and proteins that have been tested and shown functional are indicated with an asterix (*).

1.5 Regulation of DMSP production and/or uptake

As previously discussed, there are a wide range of suggested functions for DMSP in the marine organisms that either produce it or assimilate it. Furthermore, different organisms appear to have very different uses for DMSP. Very few of these suggested functions have been fully confirmed, but there are multiple conditions that appear to either stimulate or inhibit DMSP production, and much can be extrapolated from this.

1.5.1 Regulation by the presence of nutrients

Large areas of the oceans are deficient in various nutrients, and one nutrient in particular that is known to play a role in DMSP regulation is nitrogen (Sunda et al. 2007). Areas of reduced nitrate and silicate in the ocean have been found to correlate with higher DMSP concentrations (Trevena et al. 2003), and it has been shown to increase DMSP production in higher plants (Hanson et al. 1994), marine algae (Grone & Kirst 1992) and, more recently, bacteria (Curson et al. 2017). The reason for this is that DMSP is a sulfur-based compound, instead of several other osmoprotectants including GBT, which are nitrogen-based (Dickschat et al. 2015). Therefore, in nitrogen-limited situations, DMSP is the most viable osmoprotectant to synthesise so that any nitrogen taken up can be routed into more important pathways for survival. Many species have the ability to synthesise multiple osmoprotectants, and therefore in nitrogen-limited environments, organisms produce more DMSP, even to the point of replacing GBT completely (Stefels, 2000). Marine environments are nitrogen-deficient with an abundance of sulfur, and terrestrial environments are relatively the opposite. Considering the different elemental availabilities of both environments, it likely explains why marine environments widely utilise the sulfur-based DMSP as the favoured osmoprotectant, whereas terrestrial environments preferentially use nitrogen-based GBT (Sun et al. 2012).

Another element known to play a role in DMSP regulation is sulfur. Sulfur-deficient conditions were found to inhibit both cellular growth and DMSP production in marine algae (Ito et al. 2011). DMSP production was inhibited by decreasing the activity of the DMSP-producing enzyme, MTHB S-methyltransferase. This reduction in activity is accompanied by an increase in the activity of *O*-acetyl serine sulfhydrylase, a sulfur-assimilation gene. This is most likely due to the fact that limited sulfur conditions means, by extension, limited methionine availability, forcing cells to produce their own. Therefore, methionine synthesis enzyme activity increases, and most of the methionine in the organism is destined for vital pathways such as AdoMet production, instead of the synthesis of DMSP (Ito et al. 2011). Interestingly, both the uptake of DMSP from the environment and the import and conversion of DMSHB to DMSP were seen to increase in response to sulfur deprivation. This suggests that even when methionine is preferentially used in other production

pathways, there is still a need for DMSP to be synthesised/assimilated, and some microorganisms utilise multiple routes by which to accumulate it.

1.5.2 Regulation by salinity

DMSP is an osmoprotectant (or compatible solute), and as such it is often regulated by changes in salinity (Stefels 2000), although some high DMSP-producers such as *Emiliania huxleyi* produce DMSP constitutively (Sunda et al. 2002) and therefore are not particularly regulated. In most organisms, higher levels of salinity require a higher concentration of an osmoprotectant to maintain cell volume. DMSP is still produced by some species at low salinity levels, but is it at the higher salinity levels that production is noticeably increased, through increased activity of the MTHB S-methyltransferase (Ito et al. 2011). Uptake of DMSHB and DMSP also increase under high salinity conditions (Ito et al. 2011), which has also been observed in marine bacteria (Wolfe 1996). Although salinity is often one of the most effective regulators of DMSP production, not all species respond in this way, with concentrations staying the same between salinity changes in organisms living in continuously high salinity conditions, such as *Spartina alterniflora* (Otte et al. 2004). This further suggests that DMSP plays very different roles in different species that produce it. It has also been shown that the addition of alternative osmoprotectants, such as GBT, can dramatically decrease DMSP levels, especially when combined with sulfur-deficient conditions (Ito et al. 2011).

1.5.3 Regulation by temperature

DMSP is suggested to be a cryoprotectant, meaning that temperature is also proposed to be a regulatory condition (Stefels 2000). Decreasing temperatures have been linked to increased DMSP concentrations on multiple occasions (Karsten et al. 1992, Sheets and Rhodes 1996) as protection against the damage that freezing can cause. In low temperatures, the incorporation of carbon into proteins is often seen to reduce, but the production of carbohydrates (which play a role in acetyl-CoA and DMSP production) is relatively unaffected (Stefels 2000).

1.5.4 Regulation by light

In photosynthetic organisms light is also observed to play a regulatory role in the synthesis of DMSP. This is because sulfate reduction is an energy-dependent process and is therefore coupled to cell metabolism and, in turn, is stimulated by light (although not dependent on it) (Stefels 2000). Increased levels of methionine mean more is available for the production of sulfur compounds such as DMSP. Furthermore, in short-day incubations it is thought that carbon fixation is reduced, and therefore reserved for vital metabolic processes, thus reducing the amount of DMSP produced (Stefels 2000). This relationship between increased light levels and DMSP production is seen in multiple species of marine

phytoplankton (Karsten et al, 1996), and fluxes of DMSP production are observed in green algae through annual cycles, with decreasing DMSP content correlating with decreasing daylengths, and vice versa (Karsten et al 1991). *Synechococcus* is also found to assimilate around a 15 % greater proportion of DMSP when incubated in light instead of dark conditions (Malmstrom et al. 2005).

1.6 DMSP transport

In addition to the organisms that can synthesise DMSP, many non-DMSP-producing strains of phytoplankton and bacteria assimilate it from the environment (Vila-Costa et al. 2006). Although DMSP is not the only compatible solute that microorganisms seek to use, it is thought that it is more preferentially used in marine environments than, for example, GBT due to the availability of sulfur in the oceans being much higher than nitrogen, therefore DMSP is more readily produced and taken up, and vice versa in terrestrial environments (Sun et al. 2012). It is important to note that DMSP is not always taken up because of its role as a compatible solute, as uptake is not always regulated or affected by salinity (Otte et al. 2004). Between them, bacteria, phytoplankton and microzooplankton account for between 10 and 50 % of the DMSP assimilated from the marine environment (Vila-Costa et al. 2006), meaning that this assimilation plays a major role in the regulation of sulfur emissions to the atmosphere.

Structurally, DMSP is a zwitterion (**Figure 1-3***)* and this charge means that it is unable to pass through the membrane by simple diffusion, so the organisms must rely on either specific transporters, or make use of other transport systems already in the organism to take it up (Vila-Costa et al. 2006). This is thought to be the case as DMSP and GBT have been shown to have inhibitory effects on the uptake of each other, and both have similar uptake kinetics (Kiene et al. 1998). Furthermore, even terrestrial species that are unlikely to be in regular contact with DMSP demonstrate the ability to take it up when subjected to high salinity environments and are requiring a compatible solute (Cosquer et al. 1999). There are two main families of transporter that are used for the transport of nitrogen-based compatible solutes, and are known to be used by the bacteria *Roseobacter*, SAR11 clade bacteria, cyanobacteria, and also phytoplankton (Dickschat et al. 2015) to also transport DMSP into the cell for use and catabolism. These transporters have been associated with DMSP transport due to the close proximity of genes encoding for their machinery to some of those that are involved in DMSP degradation – the *ddd* genes (DMSP-dependent DMS) (Sun et al. 2012), including *dddD* (Dickschat et al. 2015). This is because DMSP uptake has been shown to be upregulated in *Pseudomonas doudoroffi* cells where DMSP lyase activity was increased (Yoch et al. 1997).

1.6.1 BCCT transporters

One of the transporter types proposed to be utilised by DMSP is the betaine choline carnitine transporter (BCCT) (Sun et al. 2012), which are associated with *dddD* and several other catabolic genes within various species (Curson et al. 2011b). These transporters exist almost ubiquitously in microorganisms, and, as the name suggests, are known to transport GBT across the membrane in species such as *Escherichia coli* (Dickschat et al. 2015). The discovery that they are also able to move DMSP followed later (Ziegler et al. 2010). It was found that although DMSP and GBT appear to use the same transport system, the very similar MMPA does not (Yoch 2002), suggesting that the positive charge on either the sulfur or nitrogen is important in the use of these transporters, indeed, it is a feature shared by many known BCCT carriers (Ziegler et al. 2010). These high-affinity uptake systems are often involved in maintaining osmotic concentrations through the movement of ions and molecules inside and out of the cell, changing the intracellular osmotic potential (and therefore the water flow) accordingly (Ziegler et al. 2010). The genes that encode for these transporters are sometimes found to be regulated by changes in salinity (Ziegler et al. 2010), enabling a rapid response to changes in the salinity of an environment.

BCCT transporters are secondary transporters, and typically have three monomers of 12 predicted transmembrane segments, with variable N- and C-terminal lengths (Ziegler et al. 2010). These terminals protrude into the cytoplasm and play a role in the control of transport activity. BCCT transporters can be either symporters or antiporters, meaning that they transport different solutes in either the same or opposite direction through the membrane (Dickschat et al. 2015). The movement of these solutes (e.g. Na^+ or H^+) is a transmembrane motive force, which is exploited by the transporter to provide the energy required to transport the DMSP into the cell (Ziegler et al. 2010). The nomenclature and amino acid sequences of BCCT transporters varies between species, ranging from CaiT in *E. coli* to BetP in *Corynebacterium glutamicum* (Sun et al. 2012), and DddT in both *Marinomonas* (Todd et al. 2007) and *Halomonas* HTNK1 (Todd et al. 2010). This *dddT* is closely linked to *dddD*, appearing on an operon *dddTBCR* where *dddR* is a transcriptional regulator that activates *dddD* in response to DMSP in both *Marinomonas* and *Halmononas* (Todd et al. 2007; Todd et al. 2010).

1.6.2 ABC transporters

The second transporter family found to carry DMSP across the membrane is the ATP binding cassette (ABC) transporter, a commonly used primary transporter that can be found in all three domains of life (Eitinger et al. 2011). Primary transporters move molecules across the membrane in exchange for an ATP molecule (Eitinger et al. 2011).

Although there are many variations in nature, the ABC transporter used to transport DMSP is of the canonical structure most often found in prokaryotes, although even within this group there is still a sizeable amount of variation in structure (Eitinger et al. 2011). The standard form of an ABC transporter consists of three parts – a transmembrane protein (TMP), a nucleotide binding protein (NBP), and a substrate binding protein (SBP) (Dickschat et al. 2015). Eitinger et al. (2011) writes in detail about the function and design of ABC transporter. The TMP and NBP form a heterotetramer, with two TMPs spanning the membrane forming a translocation pore through it, and two intracellular NBPs that form a dimer and bind ATP molecules, subsequently hydrolysing them. In Gram-negative bacteria the open SBPs freely diffuse through the periplasm between membranes, and have high affinity to the molecules they transport, binding easily. Once bound, the SBP interacts with the TBP structure and causes the NBPs to bind ATP. The NBPs are highly conserved open structures that bind two ATP molecules and close in a tweezer-like' motion, changing the internal conformation of the NBP, and, subsequently, opening the TBP pore to the extracellular environment. The TMP receives the substrate from the SBP and it enters the pore. Once the ATP molecules are hydrolysed to ADP and phosphate, it provides the energy to shift the TBP back to the internal conformation, releasing the substrate into the cell.

Some bacteria contain multiple ABC transporters that are capable of moving DMSP, albeit to differing degrees of success (Dickschat et al. 2015). *Bacillus subtilis* is one such molecule, using its OpuA, OpuC and OpuD ABC transport systems to move not only GBT, but also DMSP, through the membrane (Kempf & Bremer, 1998). Another example of this type of transport is the DMSP transporter encoded for by the *potABCD* genes in *Burkholderia ambifaria* (Dickschat et al. 2015). The genes encoding many of these ABC transporters, like the BCCT transporters, have been linked to the *dddD* gene in multiple species (Sun et al. 2012).

1.7 DMSP catabolism

DMSP is released into the ocean from the marine organisms that synthesise it by cell lysis following senescence (Stefels & van Boekel, 1993), after damage by viruses (Bratbak, 1996), or as a result of grazing by herbivores (Kiene et al. 2000) and microzooplankton (Wolfe & Steinke 1996). When it is in the ocean it becomes an available resource for marine bacteria and phytoplankton that are able to take it up and utilise it (Dickschat et al. 2015). Uptake of DMSP is not simple and requires energy, so the benefits of taking it up must be significant. Most species that transport it do so because they require it as a source of nutrients, which they have access to once they catabolise it. The two

pathways used to do this (demethylation or cleavage) both provide carbon and energy to the cell (Varaljay et al. 2015), and there are many species of marine bacteria that use DMSP as a carbon source (Curson *et al*. 2011b). Interestingly, some bacteria transport DMSP across the membrane, have the mechanisms to catabolise it through demethylation, cleavage, or both, and yet are unable to grow on it as a sole carbon source (Johnston et al. 2008). As many species of marine bacteria are also able to use both DMS (Schäfer et al. 2010) and acrylate (Todd et al. 2010) as carbon sources, this could be the reason behind catabolising DMSP even when it can't be used as a sole carbon source. There are also species that are able to use DMSP as an exogenous sulfur source (Tripp et al. 2008).

Some bacteria take up DMSP because of its use as a compatible solute for osmotic acclimatisation (Dickschat et al. 2015) as described above. It may also be taken up and cleave as an indirect route for scavenging nutrients from phytoplankton, as suggested in the 'messy eater' hypothesis by Johnston *et al*. (2008). It is known that certain species of zooplankton that feed on phytoplankton are attracted to DMS emissions, and when they feed they usually do not consume the entire organism. This means that there are plenty of 'scraps' left that bacteria could utilise, and suggests that these bacteria could be using the lysis of DMSP to produce more DMS and encourage more zooplankton to graze. Furthermore, DMS and acrylate are even more efficient scavengers of hydroxyl radicals than DMSP (Sunda et al. 2002), meaning that their presence within the cell is equally as desirable as an antioxidant.

1.7.1 Demethylation of DMSP

Once DMSP has been transported into the cell, there are two mechanisms by which it can be metabolised (Curson et al. 2011b). The most prevalent route of catabolism is a series of reactions, beginning with a demethylation, that break DMSP down into other useful compounds and allow nutrient (namely carbon and sulfur) assimilation (**Figure 1-7**) (Kiene et al. 2000). This route processes between 50 – 90 % of the DMSP that is taken up by the cells (Kiene et al. 2000). It releases methanethiol (MeSH), enabling the assimilation of biogenic sulfur from DMSP that can then be used for the biosynthesis of amino acids (Dickschat et al. 2015), or it is released and consumed by plankton. This pathway was known to exist for many years before the steps were fully discovered. The first gene associated with this pathway, designated *dmdA*, was discovered by Howard et al. (2006).

The demethylation pathway is actually two pathways – the demethylation of DMSP, followed by the demethiolation of methyl mercaptopropionate (MMPA) (Reisch et al. 2011b). The demethylation of DMSP is catalysed by DmdA, an enzyme with strict substrate specificity, suggesting that this role is its sole purpose (Reisch et al. 2011b). It also requires the presence of FH_4 (tetrahydrofolate), which acts as the methyl group acceptor, becoming Me-FH_4 (Howard et al. 2006). Me-FH_4 can then become the methyl donor in both methionine and *S*-adenosyl-methionine synthesis, or can be oxidised to become Formyl-FH_4, a carbon donor in the synthesis of cysteine from glycine (Reisch et al. 2011b).

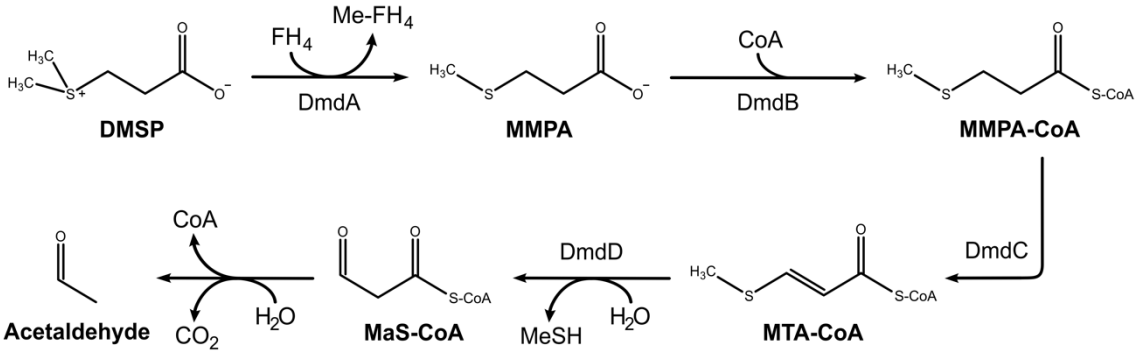

Figure 1-7: The demethylation and demethiolation catabolism of DMSP. DMSP is converted to MMPA via DmdA, removing a methyl group with tetrahydrofolate (FH_4) as the methyl acceptor. DmdB then catalyses the conversion of MMPA to MMPA-CoA through the addition of coenzyme, followed by the oxidation to MTA-CoA via DmdC. MTA-CoA is transformed via DmdD with the addition of H_2O, forming a brief intermediate, followed by the immediate release of MeSH to form MaS-CoA. This is finally converted to acetaldehyde through the addition of another H_2O and the release of CoA and a CO_2 molecule.

The MMPA that results from this demethylation is then demethiolated. For many years the breakdown of MMPA was thought to potentially follow several different routes, but (Reisch et al. 2011a) showed that, in some species at least, the demethiolation pathway resulting in the release of MeSH, CO_2 and acetaldehyde is the one used (**Figure 1-7**). The MMPA-CoA thioester intermediate was discovered in *Ruegeria pomeroyi*, the formation of which is catalysed by a methylmercaptopropionyl-CoA ligase, termed DmdB, and requires one molecule of ATP (Reisch et al. 2011a). The MMPA moiety of this thioester is dehydrogenated, forming a double bond and losing two electrons to FAD, becoming $FADH_2$. This reaction results in a methylthioacryloyl-CoA (MTA-CoA) intermediate, and is catalysed by DmdC, a dehydrogenase (Reisch et al. 2011a). When this function was removed, the mutant *R. pomeroyi* was no longer able to grow on MMPA as a sole carbon source, suggesting that this stage is vital to the breakdown of MMPA (Reisch et al. 2011a). The final enzyme involved in this pathway is DmdD, which belongs

to the crotonase family (Tan et al. 2013). This enzyme catalyses multiple steps in this final reaction, starting with a hydration that incorporates a molecule of H_2O, and liberates MeSH immediately. This forms a malonate semialdehyde-CoA (MaS-CoA) intermediate, which undergoes a hydrolysis with a second H_2O molecule that releases the CoA group from the rest of the molecule (Tan et al. 2013). It is thought that MaS-CoA spontaneously decomposes to acetaldehyde, releasing CO_2. This acetaldehyde can then be converted to acetate via an acetaldehyde dehydrogenase (Reisch et al. 2011b).

Although this pathway is associated with the demethylation of DMSP, it is not restricted to it. This pathway is seen in a wider variety of bacteria than *dmdA,* and homologs of *dmdB* and *dmdC* have also been found in terrestrial bacteria, suggesting that it has other functions (Reisch et al. 2011b).

1.7.2 DMSP cleavage to DMS

The second pathway through which DMSP is catabolised is the cleavage pathway to DMS and either 3-HP or acrylate (Curson et al. 2011b). It does this through enzymatic lysis (Curson *et al*. 2011b). Although the products are the same, DMSP has been found to be lysed through several different routes (**Figure 1-8**).

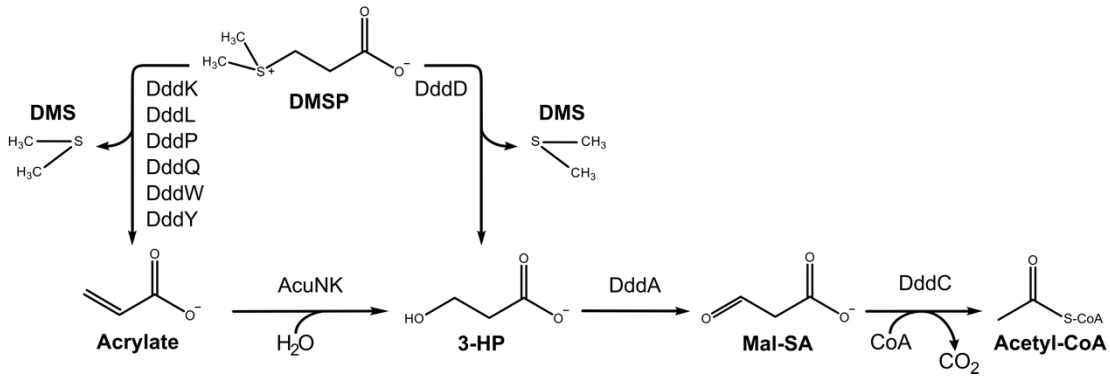

Figure 1-8: The catabolic lysis of DMSP, resulting in the production of DMS through one of two ways. These reactions are controlled by various ddd genes. The direct lysis of DMSP to 3-HP is catalysed by DddD, whereas DddK/L/P/Q/W/Y lyse it to acrylate first, which is then converted into 3-HP via AcuNK. DddA catalyses the oxidation of 3-HP to Mal-SA, and DddC enables the addition of coenzyme A to form acetyl-CoA.

When it is lysed it produces DMS and, depending on the enzyme catalysing the reaction, either acrylate (Kirkwood *et al.* 2010; Curson *et al.* 2011b) or 3-HP (Todd et al. 2007), which are both significant molecules in the environment and also in industry (Werpy & Petersen, 2004).

The breakdown of DMSP to DMS is catalysed by enzymes termed Ddd enzymes in marine bacteria, although there are homologs in other species (Dickschat et al. 2015). There are currently seven different identified *ddd* genes, with organisms often containing

a selection of them, and although ultimately all the Ddd+ enzymes encoded result in the catabolism of DMSP and the release of DMS, they are such a varied group of peptides and processes that it suggests that there is not one overarching, ubiquitous system for DMS production (Curson et al. 2008).

1.7.3 DMSP lysis to DMS via *dddD*

The most notable difference in these enzymes is between DddD, which cleaves DMSP to release DMS and 3-HP (**Figure 1-8**), and the other six Ddd enzymes that release acrylate instead of 3-HP alongside DMS (Curson et al. 2011b). DddD was the first DMSP-catabolising enzyme to be discovered by Todd et al. (2007), in the marine bacterium *Marinomonas* sp MWYL1 that was able to release DMS, but only when the growth media was supplemented with DMSP. The DddD enzyme is from the family of type III acyl coenzyme A (CoA) transferases, and it appears that its expression is controlled by a transcriptional regulator in a second operon, *dddTBCR* (Todd et al. 2007). DddD is similar to an enzyme in *E. coli* called CaiB that transfers CoA to carnitine, and its structure is predicted to contain two CaiB-like intertwined domains that form a long polypeptide linker, with the catalytic aspartate that is required for CoA transfer at the C-terminus (Alcolombri et al. 2014). The presence of this aspartate suggests that DddD uses a similar two step mechanism to CaiB to break down DMSP, transferring CoA to DMSP and forming a 3-HP-CoA intermediate before releasing the 3-HP, although this intermediate has not yet been confirmed (Alcolombri et al. 2014). After this (hypothetical) step, the intermediate is hydrolysed to 3-HP very rapidly, which may be the reason why it has not been possible to detect the intermediate (Dickschat et al. 2015). The *dddD* gene has been found in multiple species of bacteria, including many gammaproteobacteria such as *Halomonas, Pseudomonas* and *Rhizobium leguminosarum* (Sun et al. 2012). As previously mentioned, it is often found close to an operon containing *dddT*, a BCCT transporter, suggesting a contained system dedicated to DMSP uptake and catabolism (Curson et al. 2011b).

1.7.4 DMSP lysis to DMS via *dddL, dddQ, dddW* and *dddK*

Once DddD was shown to be involved in the breakdown of DMSP, homologs were searched for in online databases such as NCBI, and in other species previously known to degrade DMSP. It was found to be present in many *ddd*-containing bacteria, but there were several species where it was not, including the alphaproteobacterium *Sulfitobacter* EE-36 (Curson et al. 2008). This bacterium could be shown to catabolise DMSP, but had no DddD homolog, suggesting that another pathway was being used to break DMSP down. This pathway was unusual because, instead of cleaving DMSP into DMS and 3-HP, *Sulfitobacter* lysed it into DMS and acrylate (Curson et al. 2008). This gene was termed *dddL*, and encodes a small potentially transmembrane peptide DddL that functions in an

entirely different way to DddD (Curson et al. 2008). It is primarily found in roseobacters, but also exists in various other marine alphaproteobacteria (Curson et al. 2011b).

The breakdown pathway catalysed by DddL involves the lysis of DMSP to acrylate (**Figure 1-8**), and is a simple cleavage of the carbon-sulfur bond of DMSP seen in **Figure 1-3**. It was found that the majority of acrylate produced remained outside of the cell, in the growth medium (Curson et al. 2008), leading to the theory that DddL acts on periplasmic DMSP, cleaving it outside of the cell. There are several reasons why DddL would do this. Wolfe et al. (1997) suggest that some bacteria use DddL-mediated acrylate production as a deterrent to other organisms. Acrylate itself has antimicrobial activity at high concentrations, and can deter various predators, such as protozoan herbivores, causing them to select strains with lower acrylate production (Wolfe et al. 1997). It has also been suggested that the release of acrylate could be related to signalling, and that rather than being toxic to predators, acrylate sends some sort of anti-grazing signal to deter them (Wolfe et al. 1997; Strom et al. 2003).

The fact that this lysis may occur outside of the cell could suggest that the bacterial catabolism of DMSP to DMS and acrylate may not be predominantly driven by a nutritional need, but rather as a defensive action (Curson et al. 2008). This is compounded by the fact that several *dddL*+ strains are unable to utilise acrylate as a sole carbon source (Curson et al. 2008). Furthermore, unlike *dddD* and its operon, *dddL* does not appear to be associated with any genes encoding for transporters, which is explained by the transmembrane property of DddL, removing the need for any transporters of DMSP to be linked with this *ddd* gene (Curson et al. 2008).

DddL is one of several similar DMSP lyases. DddL, DddQ, DddW and DddK are all small polypeptides with C-terminal domains that form cupin pockets, and therefore bind to transition metals (Curson et al. 2011b). The rest of the protein structures differ between the three lyases, classing them as different protein families with domains of unknown function (DUFs) that have evolved this cupin separately (Todd et al. 2012). DddQ was discovered because it was observed that some knock-outs of *dddP* did not fully deplete DMSP degradation, initially found in *Roseovarius nubinhibens* (Todd et al. 2011), and later on in other roseobacters. DddW was the third DMSP lyase to be discovered in the roseobacter *Ruegeria pomeroyi* DSS-3, alongside DddP and DddQ (Todd et al. 2012). It is also a small polypeptide, only found in two roseobacter strains so far (Curson et al. 2011b). DddK is the most recent protein to be identified as a DMSP lyase (Sun et al. 2016), in *Pelagibacter* HTCC1062. This discovery took place in a study on the switching between the cleavage and demethylation pathways, where it was found that, in *Pelagibacter* at least, these two pathways can take place simultaneously (Sun et al. 2016), with the balance between them changing with cellular sulfur demands.

1.7.5 DMSP lysis to DMS via *dddP*

Another Ddd protein was discovered by Todd et al (2009) in the bacterium *Roseovarius nubinhiens.* This was the third DMSP lyase to be discovered, following DddD and DddL, and all three were entirely different protein families and structures, demonstrating just how widespread the enzymes catalysing the catabolism of DMSP are. The gene encoding this peptide was named *dddP* and homologs have been found in multiple species (Curson et al. 2011b). The peptide is a homodimer that is part of the family of M24 metallopeptidases, and it splits DMSP to acrylate and DMS (Todd et al. 2009). It is an unusual metallopeptidase because it does not require a metal co-factor in order to be functional, and also because it cleaves the S-C bond instead of an amino group (Kirkwood et al. 2010). This makes it similar to the M24B family, such as creatinase in *Paracoccus sp.* WB1, and not a true metalloprotein (Kirkwood et al. 2010).

It is a very widespread DMSP lyase, possibly the most abundant (Todd et al. 2009), and is found in both marine and terrestrial environments. It has even been identified in several species of fungi (Todd et al. 2009), suggesting that one or more HGT events have occured in the past.

1.7.6 DMSP lysis to DMS via *dddY*

The final Ddd protein that cleaves DMSP to DMS and acrylate to be described here is DddY. This enzyme was identified in the betaproteobacterium *Alcaligenes faecalis* M3A (Curson et al. 2011a), but was previously purified in 1995 by de Souza & Yoch (1995). *Alcaligenes faecalis* M3A, unlike several *dddL*-containing species, is able to utilise both DMSP and acrylate as sole carbon sources (Ansede et al. 1999). The protein family of DddY is unidentified, which means that it is also classed as another DUF (Curson et al. 2011b), but it was strongly predicted to be a periplasmic protein. This prediction was supported by the fact that the previously purified protein from this species was thought to be periplasmic or associated with the outer cell membrane (de Souza & Yoch 1995), and this was later confirmed through fractionation (Curson et al. 2011a). This would make DddY the first DMSP lyase to function outside of the cytoplasm (Curson et al. 2011a). The use of a periplasmic DMSP lyase compared to a cytoplasmic one could be beneficial to an organism in terms of electron transport – there is no need to actively transport DMSP into the cell, making its catabolism a less costly process (de Souza & Yoch, 1996).

DddY is found across multiple species, and is most likely spread through HGT, as this gene is very widespread, ranging between beta- gamma- and epsilonproteobacteria, but not including alphaproteobacteria, making it the first DMSP lyase not associated with this class (Curson et al. 2011a). One feature that almost all *dddY*+ species have in common is that they are all microaerobic, being mostly marine and in sediments.

Interestingly, the *dddY* in *Shewanella* species was found to be close to genes that are involved in anaerobic respiration (membrane-bound cytochromes), suggesting that in some species the catabolism of DMSP is not for the release of DMS, but perhaps for the production of acrylate, which can be used as an electron acceptor (Curson et al. 2011a).

1.7.7 'Switching' between lysis and demethylation

Species that catabolise DMSP are often able to utilise both the lysis and demethylation pathways, switching between them when most appropriate (Kiene & Linn 2000). Several factors have been suggested to regulate this switch, including nutrient supply, light and temperature (Levine et al. 2012). It was found that *Roseobacter* clades preferentially use the cleavage pathway to catabolise DMSP when under higher UV-A conditions, and the demethylation pathway under low UV-A (Levine et al. 2012), which could be explained by the antioxidant function of DMS that protects from reactive oxygen species. It was also seen that elevated temperature conditions led to increased DMSP cleavage compared to demethylation (Levine et al. 2012). Nutrient assimilation also plays a role in regulating DMSP catabolism, indeed, it was observed that the demethylation pathway was favoured when DMSP was the predominant source of organic sulfur in the environment (Varaljay et al. 2015), as organisms need to first and foremost utilise the biogenic sulfur to meet their biosynthetic requirements, rather than losing it in the form of gaseous DMS. When other sources of biogenic sulfur are present, then the DMSP cleavage pathway is also utilised as it is less vital for DMSP to be a source of sulfur (Varaljay et al. 2015).

1.8 Conclusions and Research Gaps

It has long been acknowledged that DMSP plays a vital role in the marine environment, not only in the cycle and provision of nutrients, but also in numerous protective roles within the cell, and, indirectly, in local climate control. Despite this knowledge, evidence that DMSP production is not restricted to marine eukaryotes and does in fact take place in bacteria as well, was only discovered in the last year. This discovery has revealed how understudied these particular aspects of the DMSP cycle are, namely what species are able to produce it, and what environments this production takes place in.

It is important to rectify this, as our understanding of these processes will affect how we model and predict sulfur cycling, and could mean that synthesis of DMSP normally attributed to eukaryotes could be partially the result of bacterial production. This work will attempt to increase our understanding of bacterial DMSP production by looking at how

widespread the ability to produce DMSP is within bacteria, and how significant their contribution is to the total levels of DMSP in a specific the environment, namely Stiffkey salt marsh. The following research aims will be discussed in this thesis:

1. To determine the diversity and abundance of *dsyB* in the environment, in both metagenomes and isolated bacteria.
2. Use culture-independent methods to investigate the importance of bacterial DMSP synthesis in Stiffkey salt marsh.
3. Identify key bacterial DMSP producers and determine the means by which DMSP is synthesised in bacteria, using culture-dependent techniques.

CHAPTER 2

MATERIALS AND METHODS

2 MATERIALS AND METHODS

2.1 Chemical syntheses

DMSP was synthesized from DMS (Sigma-Aldrich) and acrylic acid (Sigma-Aldrich) as described in (Todd et al. 2010). DMSP-amine and SMM were synthesized as described in (Curson et al. 2017). Met, MTOB, MTHB and MTPA are commercially available and were obtained from Sigma-Aldrich.

2.2 Media preparation and growth conditions

Novosphingobium sp. MBES04*, Thalassospira profundimaris*, *Roseovarius indicus* and the rest of the isolated bacterial strains from Stiffkey were grown in YTSS (González et al. 1996), MB (Zobell Marine Broth 2216) (Buck & Cleverdon 1960) medium, or MBM (Marine Basal Medium) 35 PSU (practical salinity units) unless otherwise stated, 10 mM mixed carbon source from a 1 M stock of 200 mM succinate, glucose, pyruvate, sucrose and glycerol, and 0.5 or 10 mM NH_4Cl as nitrogen source at 30°C. *Streptomyces mobaraensis* was grown in in GYM *Streptomyces* medium (4 g glucose, 4 g yeast extract, 10 g malt extract, 2 g calcium carbonate, 12 g agar per litre distilled water) at 25°C and *Nocardiopsis chromatogenes* was grown in MYM medium (4 g glucose, 4 g yeast extract, 10 g malt extract, 2 g calcium carbonate, 10 g NaCl, 12 g agar per litre distilled water) at 37°C. Where indicated, the salinity of MBM was adjusted by altering the amount of sea salts (Sigma-Aldrich) added, and nitrogen levels were altered through the adjustment in volume of NH_4Cl added as the nitrogen source. Methylated sulfur compounds, namely the DMSP pathway intermediates, were only added to MBM where indicated in experiments that specifically addressed the effect of adding such compounds. *Escherichia coli* was grown in Luria-Bertani (LB) (Sambrook et al. 1989) complete medium at 37 °C. *Rhizobium leguminosarum* was grown in tryptone yeast (TY) (Beringer 1974) complete medium or Y (Beringer 1974) minimal medium (with 10 mM succinate as carbon source and 10 mM NH_4Cl as nitrogen source) at 28°C. Where necessary, antibiotics were added to media at the following concentrations: streptomycin (400 µg ml^{-1}), kanamycin (20 µg ml^{-1}), spectinomycin (200 µg ml^{-1}), gentamicin (20 µg ml^{-1}), ampicillin (100 µg ml^{-1}), rifampicin (400 µg ml^{-1}). Strains used in this study are listed in **Table 2-1**.

Table 2-1: A list of strains used in this study

Strain	Description	Reference
Escherichia coli 803	Strain used for routine transformations	Wood (1966)
E. coli BL21	Strain for overexpression of cloned genes in pET vectors	Studier and Moffat (1986)
E. coli JM101	Strain for expression of *lacZ* gene in blue-white screen	Yanisch-Perron et al., (1985),
Rhizobium leguminosarum J391	Streptomycin-resistant derivative of wild type strain 3841 used for library screening and expression of genes cloned in plasmid pLMB509 or pRK415	Young et al. (2006)
Labrenzia aggregata LZB033	Wild type strain, isolated from ME3 site, *dsyB*+	Curson et al. (2017)
Sagittula stellata E-37	Wild type strain, *dsyB*+	Gonzalez et al. (1997)
Oceanicola batsensis DSMZ21189	Wild type strain, *dsyB*+	Cho and Giovannoni (2004)
Amorphus coralli DSMZ19760	Wild type strain, *dsyB*+	Zeevi Ben Yosef et al. (2008)
Pelagibaca bermudensis HTCC2597	Wild type strain, *dsyB*+	Cho and Giovannoni (2006)
Sulfitobacter sp. EE-36	Wild type strain, *dsyB*-	Gonzalez et al. (1996)
Alteromonas genovensis PQQ33	Wild type strain, isolated from Stiffkey Salt marsh	This study
Oceanicola sp. Ar-45	Wild type strain, isolated from Stiffkey Salt marsh	This study
Labrenzia sp. BR-18	Wild type strain, isolated from Stiffkey Salt marsh	This study
Marinobacter sp. Set72	Wild type strain, isolated from Stiffkey Salt marsh	This study
Novosphingobium sp. BW1	Wild type strain, isolated from Stiffkey Salt marsh	This study
Pseudooceanicola sp. 22II1-22F33	Wild type strain, isolated from Stiffkey Salt marsh	This study
Roseobacter sp. ARCTIC-P4	Wild type strain, isolated from Stiffkey Salt marsh	This study
Rhodobacter sp. AB300d	Wild type strain, isolated from Stiffkey Salt marsh	This study
Rhodobacterales bacterium JB-27	Wild type strain, isolated from Stiffkey Salt marsh	This study
Stappia sp. M8	Wild type strain, isolated from Stiffkey Salt marsh	This study
Thalassiospira profundimaris WPO211 (DSM 17430)	Wild type strain used to demonstrate DMSP production by *mmtN* expression	DSMZ Culture Collection
Roseovarius indicus B108 (DSM 26383)	Wild type strain used to demonstrate DMSP production by *mmtN* expression	DSMZ Culture Collection
Nocardiopsis chromatogenes (DSM 44844)	Wild type strain used to demonstrate DMSP production by *mmtN* expression	DSMZ Culture Collection
Streptomyces mobaraensis IPCR16-22 (DSM 40847):	Wild type strain used to demonstrate DMSP production by *mmtN* expression	DSMZ Culture Collection
T. profundimaris WPO211-Rif	Rifampicin-resistant derivative of *T. profundimaris* WPO211	This study
T. profundimaris WPO211-Rif (*mmtN*$^-$)	*T. profundimaris* WPO211-Rif with mutation in *mmtN*	This study

2.3 Transformations into *E. coli*

2.3.1 Making chemically competent cells

A starting culture of 5 ml LB was inoculated with *E.coli* (803/JM101) and incubated overnight at 37°C. This was inoculated in a 1:100 dilution to 100 ml LB and incubated at 37°C, 200 rpm for 2 – 3 h (to OD_{600} 0.2 – 0.4). The culture was transferred into 50 ml sterile falcon tubes, and cells were retrieved using a pre-cooled centrifuge at 4°C, spinning at 4000 rpm for 10 minutes. Falcon tubes were kept on ice and the supernatant removed. Both pellets were carefully mixed with 10 ml ice cold 0.1M $CaCl_2$, and left on ice for 60 minutes. The mix was centrifuged as before, and the supernatant removed. One pellet was resuspended in 2 ml of 0.1M $CaCl_2$, and the second pellet was resuspended in this mixture. Cells were left on ice for at least 3 h, and could be stored in the fridge overnight.

2.3.2 Heat shock transformations

Up to 16 µl of DNA was added to 100 µl competent cells and incubated on ice 1 hour alongside a negative control of cells, and a positive control of vector DNA only. Samples were heat shocked at 42°C for 2 minutes, and transferred to ice for 2 minutes. Cells were mixed with 750 µl LB and incubated at 37°C for 60 – 90 minutes. Aliquots of 100 µl of cells were plated on LB containing selective antibiotics. The remaining mix was centrifuged at max speed for 2 minutes and the majority of the supernatant poured off. The pellet was resuspended in the remaining liquid and also plated on LB. Plates were incubate at 37°C overnight.

2.4 Polymerase chain reaction (PCR)

Genes were amplified using polymerase chain reaction (PCR) in a Thermal Cycler, either 25 µl or 50 µl mixes. Standard 25 µl PCR mixes contained 12.5 µl MyFi™ DNA Polymerase (enzyme/buffer/dNTPs/DMSO), 0.5 µl template (50–100 ng), 0.5 µl of 20 pmol of F and R primers (list of primers in **Table 2-2**) and 11 µl nuclease-free H_2O. Every PCR included a negative control of sterile water instead of template, and a positive control of genomic DNA. DNA was also amplified directly from bacterial colonies for large scale isolate screening. A sterile toothpick was used to pick the edge of a colony, and then stabbed into PCR tubes with 100 µl sterile water. Tubes were microwaved for 10 seconds, and 1 µl of the lysed mixture was used in the PCR mix (with 10.5 µl sterile water).

Table 2-2: A list of the primers used in this study

Primer name	Sequence (5' to 3')*	Use
27F	AGAGTTTGATCCTGGCTCAG	Forward primer used to amplify the 16S rRNA gene for identification
1492R	GGTTACCTTGTTACGACTT	Reverse primer used to amplify the 16S rRNA gene for identification
Eub_338F	ACTCCTACGGGAGGCAGCAG	Reverse primer used to amplify the 16S rRNA gene for RT-qPCR
Eub_518R	ATTACCGCGGCTGCTGG	Reverse primer used to amplify the 16S rRNA gene for RT-qPCR
M13 uni (-43)	AGGGTTTTCCCAGTCACGACGTT	Universal forward primer used to amplify inserts in pLAFR3
M13 rev (-29)	CAGGAAACAGCTATGACC	Universal reverse primer used to amplify inserts in pLAFR3
dsyB_deg1F	CATGGGSTCSAAGGCSCTKTT	Degenerate primer for amplification of *dsyB* in PCR and RT-qPCR
dsyB_deg2R	GCAGRTARTCGCCGAAATCGTA	Degenerate primer for amplification of *dsyB* in PCR and RT-qPCR
dsyB_deg3R	GCCGCCSACRTCSAGCA	Degenerate primer trialled for amplification of *dsyB* in PCR
mmtN_degF	GGCAGYGAYCTYGAYCCSCG	Degenerate primer for amplification of *dsyB* in PCR and RT-qPCR
mmtN_degR	CCAVGGRTARTARTGSGC	Degenerate primer for amplification of *dsyB* in PCR and RT-qPCR
NOmmtN_NdeI-F	CGGATCCCATATGTCTGACGCAGATGACTCC	Cloning of *N. sp* BW1 *mmtN* into pET21a for pET21a-Nov
NOmmtN_EcoRI-R	GGAATTCACTCTACCTTGGGGATACC	Cloning of *N. sp* BW1 *mmtN* into pET21a for pET21a-Nov
TPmmtN_NdeI-F	CGGATCCCATATGCAACATGCTTTAGAAGAGAGC	Cloning of *T. profundimaris mmtN* into pET21a for pET21a-Tprof
TPmmtN_EcoRI-R	CGAATTCTTAGGCCGGTGTGCCGCGAATGAC	Cloning of *T. profundimaris mmtN* into pET21a for pET21a-Tprof
RImmtN_NdeI-F	CGAATTCCATATGACCGATTTCAAAACGCCCG	Cloning of *R. indicus mmtN* into pET21a for pET21a-Rind
RImmtN_EcoRI-R	CCCGGATCCTCAACGATTGGACGGATCGGTTTCC	Cloning of *R. indicus mmtN* into pET21a for pET21a-Rind
NCmmtN_NdeI-F	CGGATCCCATATGCCGTCCGAGCACACGATG	Cloning of *N. chromatogenes mmtN* into pET21a for pET21a-Ncard
NCmmtN_EcoRI-R	CGAATTCATCGCCGGTCCTCCTCGTCGG	Cloning of *N. chromatogenes mmtN* into pET21a for pET21a-Ncard
SMmmtN_NdeI-F	CGGATCCCATATGAGAACAGAGACCGGACCGCC	Cloning of *S. mobaraensis mmtN* into pET21a for pET21a-Smob
SMmmtN_EcoRI-R	CGAATTCTACGTGGCGGGTGTGCCCCTGAC	Cloning of *S. mobaraensis mmtN* into pET21a for pET21a-Smob
TPSCO_BamHI_F	CGGGATCCGTCGCCTTTATCTTGCAAAG	Generating a single crossover *mmtN* knockout in *T. profundimaris*
TPSCO_EcoRI_R	CGGAATTCCGTTCCGGAATGTTGCAG	Generating a single crossover *mmtN* knockout in *T. profundimaris*
TP_OUT_EcoRI_F	CGGAATTCATGCTAGAAGAGAGCAGC	Forward primer to *T. profundimaris mmtN*
TP_OUT_BamHI_R	CGGGATCCTTAGGCCGGTGTGCCGCG	Reverse primer to *T. profundimaris mmtN*

* Restriction sites included in primers underlined

2.5 Visualization and extraction of DNA

2.5.1 Gel Electrophoresis

PCR products were visualised using gel electrophoresis. Gels were made to 1 – 1.5% (w/v) agarose using 1x TAE Buffer (50x stock: 242 g Tris base, 57.1 ml glacial acetic acid, 100 ml 500 mM EDTA (pH 8.0), water to 1 liter. A 1x solution contains 40mM Tris, 20 mM acetic acid, and 1 mM EDTA), melted and cooled to 50°C before adding 3 µl Ethidium Bromide (10 mg/ml) and pouring into gel trays. Samples were loaded into wells alongside a 1 KB Plus DNA ladder (Invitrogen) for reference to size. Gels were typically run at 90 V for 45 – 60 minutes, and the separation of DNA fragments was visualised using a UV gel imaging doc.

2.5.2 PCR purification (Roche)

PCR amplified DNA was recovered using the Roche High Pure PCR Product Purification Kit, using Binding Buffer to five times the PCR mix volume. The purified PCR product was eluted from the column using 35 – 50 µl sterile water, collected in a 1.5 ml microcentrifuge tube.

2.5.3 Gel extraction (QIAGEN)

DNA was extracted from the agarose gel after gel electrophoresis as described in the QIAquick Gel Extraction kit. Dissolved gel samples are precipitated with 1 volume of isopropanol and 10 µl of 3 M Sodium acetate. DNA was eluted using 35 – 50 µl of sterile water added to the centre of the membrane, left to rest for 1 minute and then centrifuged for 1 minute.

2.6 Methods of DNA extraction (linear and plasmid)

2.6.1 Phenol chloroform DNA extractions

A starting culture of 5 ml LB was inoculated and incubated overnight at 28 – 37°C with shaking. Up to 1.5 ml culture was poured into a microcentrifuge tube and cells pelleted by centrifuging at maximum speed for 2 minutes. Supernatant was discarded and the pellet resuspended in 250 µl Buffer P1, mixed by inversion with 250 µl Buffer P2, and 350 µl Buffer P3 immediately after. Samples were left on ice for up to 5 minutes, and then centrifuged for 10 minutes at maximum speed. The supernatant was removed to a clean microcentrifuge tube, mixed with 400 µl Phenol:Chloroform:Isoamyl Alcohol 25:24:1 (v/v) and vortexed for 5 – 10 seconds until the mixture was homogenised and cloudy. Samples were centrifuged for 2 minutes at maximum speed, and the top aqueous layer was removed to a new microcentrifuge tube, to which 800 µl of 100% ethanol was also added.

Tubes were mixed by inversion and spun for 10 minutes at maximum speed. The supernatant was discarded, and 500 µl of 70% ethanol was added over the pellet. Samples were once again spun for 2 minutes at maximum speed, and the ethanol removed. The pellet was air-dried for 5 – 10 before being resuspended in 35 – 40 µl nuclease-free water. DNA was quantified by nanodrop.

Table 2-3: A list of the plasmids used in this study

Plasmid	Description	Reference
pLAFR3	Wide host-range cosmid vector, used for library construction	Staskawicz et al. (1987)
pET21a	Plasmid vector for expression of cloned genes in E. coli	Merck Millipore
pRK2013	Helper plasmid used in triparental matings	Figurski and Helinski (1979)
pK19-Spec	Plasmid used in creating *mmtN* SCO knockout in *T. profundimaris*	Todd *et al* 2011
pLMB509	Alpha expression vector used to express *mmtN* in *T. profundimaris* mutant	Tett *et al* 2012
pBIO2275	*Prymnesium parvum* CCAP946/6 *DSYB* cloned in pRK415	Curson *et al* 2018
pBIO2276	*Symbiodinium microadriaticum* CCMP2467 *DSYB*, codon-optimised, synthesised and cloned in pLMB509	Curson *et al* 2018
pBIO2272	*Chrysochromulina tobin* CCMP291 *DSYB*, codon-optimised, synthesised and cloned in pLMB509	Curson *et al* 2018
pBIO2270	*Acropora cervicornis DSYB*, codon-optimised, synthesised and cloned in pLMB509	Curson *et al* 2018
pBIO0438	pLAFR3 cosmid from a *Novosphingobium sp.* MBES04 library that contains ~21 kb genomic DNA including *mmtN*	This study
pBIO0762	pLAFR3 cosmid from a *Novosphingobium sp.* MBES04 library that contains ~30 kb genomic DNA including *mmtN*	This study
pBIO21N1	*N. sp* MBES04 *mmtN* cloned in pET21a(+)	This study
pBIO509N	*N. sp* MBES04 *mmtN* cloned into pLMB509	This study
pBIO21T2	*T. profundimaris* WPO211 *mmtN* cloned in pET21a(+)	This study
pBIO19TK	Disruption mutant for *T. profundimaris mmtN* created in pK19-Spec plasmid	This study
pBIO21R3	*R. indicus mmtN* cloned in pET21a(+)	This study
pBIO21N4	*N. chromatogenes mmtN* cloned in pET21a(+)	This study
pBIO21S5	*S. mobraensis mmtN* cloned in pET21a(+)	This study

2.6.2 Minipreps (QIAGEN)

Plasmid or cosmid DNA was extracted from starter cultures of 5 ml LB incubated overnight at 28 – 37°C, using the QIAprep Spin Miniprep Kit. Cells were recovered from ~3 ml culture and DNA extracted. The column was eluted with 35 – 50 µl nuclease-free water added to the membrane. This was left to rest for 1 minute, and then centrifuged for 1 minute to elute the DNA. A list of the plasmids used in this study is in **Table 2-3**.

2.6.3 QIAGEN Plasmid Midipreps

For high quality, high concentration plasmid extractions the QIAGEN Plasmid Midiprep kit was used, with the QIAGEN-tip 100 column, on 100 ml culture. The DNA was eluted from the column using 5 ml Buffer QF. To precipitate DNA 3.5 ml of room-temperature isopropanol was added to the eluted DNA and mixed. The mixture was separated into 1.5 microcentrifuge tubes and centrifuged immediately at maximum speed for 30 minutes. Supernatant was discarded and the DNA pellets washed with 500 µl 70% ethanol, centrifuging at maximum speed for 10 minutes. The ethanol was aspirated and the pellet left to air-dry for 5 – 10 minutes, then re-dissolved in a suitable volume of nuclease-free water. Concentration was quantified by nanodrop, and plasmid stored at -20°C.

2.6.4 Genomic DNA extractions (Promega)

Genomic DNA from bacterial isolates was extracted using the Wizard® Genomic DNA Purification Kit. After nuclei lysis, the mix was incubated for 5 min at 80°C, and then cooled to room temperature, and 3 µl of RNase Solution was added to the cell lysate, mixed and incubated at 37°C for ~45 minutesBetween 35 – 50 µl of DNA Rehydration Solution was added to the tube and incubated at 65°C for 1 hour, or at 4°C overnight, after which it was stored at -20°C.

2.7 Restriction digests using FastDigest enzymes

Digestions of DNA were carried out using Thermo Scientific FastDigest restriction enzymes. Up to 16 µl of DNA (depending on the concentration), 1 µl Enzyme 1, 1 µl Enzyme 2 (if required), 2 µl FastDigest Buffer and distilled water were mixed in a microcentrifuge tube to a total volume of 20 µl. The mix was incubated at 37°C for up to 60 minutes, and then inactivated by incubation at either 65°C or 80°C for 5 or 20 minutes. If necessary the digested DNA was dephosphorylated by adding an additional 1µ alkaline phosphatase, 2.5 µl of buffer and 1.5 µl nuclease-free water, and incubating for 60 minutes. Digested DNA was then visualised on a 1% agarose gel, and the correct sized fragment extracted.

2.8 Quantification of DMSP

2.8.1 GC quantification of DMSP/DMS/SMM

To quantify DMS, DMSP and SMM, gas chromatography (GC) assays were utilised. This protocol involved of the measurement of headspace DMS, either directly produced by the sample, or via alkaline lysis of DMSP/SMM, using a flame photometric detector (Agilent 7890A GC fitted with a 7693 autosampler) and a HP-INNOWax 30 m x 0.320 mm capillary column (Agilent Technologies J&W Scientific). All measurements on the GC were performed using 2 ml glass vials containing 300 µl liquid samples and sealed with PTFE/rubber crimp caps. For the measurement of DMSP, it was first lysed to DMS with the addition of 100 µl 10 M NaOH to 200 µl culture. Vials were crimp sealed immediately, incubated at 22°C for 24 h in the dark, and then measured by GC. For quantifying SMM, the same protocol was observed, with an additional step of a 10 minute incubation at 80°C before incubating overnight at 22°C. An eight point calibration curve was also produced by the alkaline lysis of known DMSP standards in water **(Figure 2-1)**, and incubated in the same way. The detection limit for headspace DMS was 0.015 nmol in water.

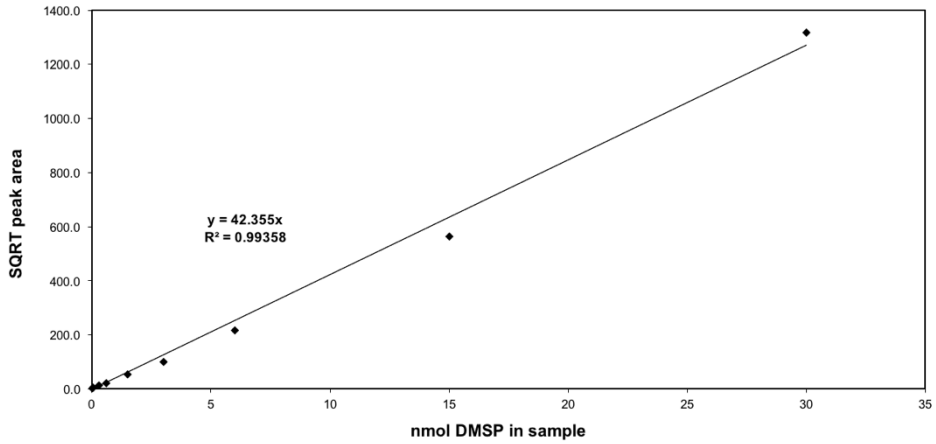

Figure 2-1: The eight-point calibration curve of DMSP, used to calculate DMSP concentrations in samples from DMS released via alkaline lysis. The curve was produced using known concentrations of DMSP ranging from 0.015 nmol to 30 nmol, added to 100 µl 10M NaOH and 200 µl sterile water, then incubated without shaking overnight in the dark.

2.8.2 Quantification of DMSP via LC-MS

LC-MS was required to rule out the possibility that DMS detected by GC was due to some other compound, and not DMSP, as both SMM and DMSHB lyse to DMS after alkaline hydrolysis. Samples were extracted as follows: cells were recovered from 3 ml of culture, and the pellets resuspended in 300 µl of 80% LC-MS grade acetonitrile (extraction solvent), and mixed by pipetting. This was centrifuged at maximum speed for 3 minutes, and 200 µl of the supernatant was collected in a fresh 2 ml screw-cap tube. For a second round of extraction, another 200 µl of the extraction solvent was added to the pellet and remaining 100 µl supernatant, and the pellet resuspended before centrifugation for 3 minutes. Another 200 µl of supernatant was collected, and a third round of extraction was then performed in the same way, to give a total volume of 600 µl of the collected supernatant for LC–MS analysis. LC-MS was carried out using a Shimadzu Ultra High Performance Liquid Chromatography (UHPLC) system formed by a Nexera X2 LC-30AD Pump, a Nexera X2 SIL-30AC Autosampler, a Prominence CTO-20AC Column oven, and a Prominence SPD-M20A Diode array detector; and a Shimadzu LCMS-2020 Single Quadrupole Liquid Chromatograph Mass Spectrometer. Samples were analysed in hydrophilic interaction chromatography (HILIC) mode using a Phenomenex Luna NH2 column (100 x 2 mm with a particle size of 3 µm) at pH 3.75. Mass spectrometry spray chamber conditions were capillary voltage 1.25 kV, oven temperature 30 ºC, desolvation temperature 250 ºC and nebulising gas flow 1.50 L min^{-1}. Solvent A is 5% acetonitrile + 95% 5 mM ammonium formate in water. Solvent B is 95% acetonitrile + 5% 100 mM ammonium formate in water. Flow rate was 0.6 ml min^{-1} and gradient (% solvent A/B) was t = 1 min, 100% B; t = 3.5 min, 70% B; t = 4.1 min, 58% B; t = 4.6 min, 50% B; t = 6.5 min, 100% B; t = 10 min, 100% B. The injection volume was 15 µl. All samples were analysed immediately after being extracted. The targeted mass transition corresponded to $[M+H]^+$ of DMSP (m/z 135) in positive mode. A calibration curve was performed for quantification of DMSP using pure DMSP standards in the extraction solvent.

2.8.3 Quantification of DMSP via Purge-Trap

Measurements of 0.5 g were weighed out and dissolved in 25 ml distilled water, with H_2SO_4 to 0.5 %. This mix was incubated at room temperature for 1 hour and then 5 ml was mixed with 1 ml 10 M NaOH and incubated overnight in the dark, before using the purge and trap method to quantify the DMS produced (Zhang et al. (2008). The samples were purged for 20 minutes and then compounds were detected with an Agilent 7890B gas chromatography (GC) instrument and quantified. The calibration curve was made using the same method, using 5 ml of each gradient DMSP concentration standards with 1 ml of NaOH.

2.9 Protein quantification

2.9.1 Quantification of protein concentration by Bradford assay

In order for DMS/DMSP concentrations to be quantified by cell growth, the protein in culture was measured. This was achieved by recovering cells from 1 ml culture through centrifugation for 1 minute at maximum speed, and resuspending in 500 µl Tris-HCl buffer (50mM, pH 7.5). Following this resuspension, the cells are lysed using sonication, for three repeats of 10 seconds, being kept on ice in between. Following sonication, samples were centrifuged at max speed for 10 minutes, and 20 µl of the supernatant was mixed with 980 µl Bradford Reagent. This was added to a cuvette and the absorption measured using a spectrometer set to OD_{595}. A four point protein standard graph was produced, using known concentrations of BSA **(Figure 2-2)**. Standards include dH_2O alone, and concentrations of 100, 200, and 400 µg/ml. This enables the calculation of the of µg protein in each culture.

2.9.2 Protein estimations with Qubit

For estimating protein concentrations in the sediment incubation experiments a Qubit™ Protein Assay Kit was used as it is more sensitive. Samples were prepped in the same way as for the Bradford method, up until the 10 minutes centrifugation. After the centrifugation of the samples, 20 µl of the supernatant was added to Qubit assay tubes, and mixed with 180 µl Working Solution (199 µl Protein Buffer and 1 µl Protein Reagent) and vortexed for 2 – 3 seconds. Samples were incubated in the dark at room temperature for 15 minutes and measured on a Qubit™ Fluorometer, alongside three standards of 0 ng/µl, 200 ng/µl and 400 ng/µl BSA concentrations.

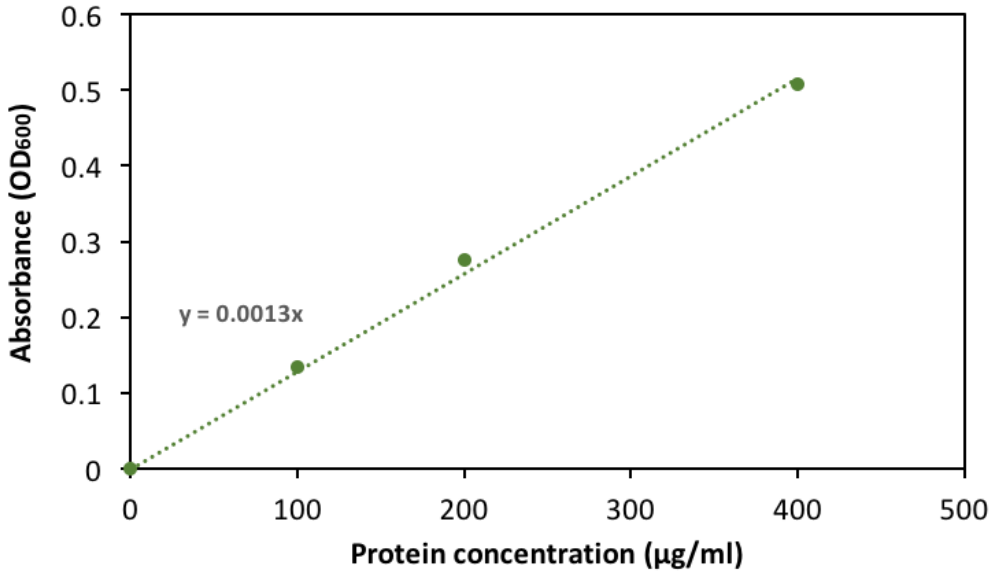

Figure 2-2: The absorbance measured in OD_{600} of four BSA standards of known concentration in Bradford Reagent, plotted with line of best fit to calculate protein concentrations of unknown samples.

2.10 Experiments at Stiffkey salt marsh

2.10.1 Sampling of Stiffkey sediment

Sediment used in this study was sampled from Stiffkey salt marsh (52.9643, 0.9255). Triplicate marine sediment samples were collected using acrylic corers tapered at one end. DMSP concentration was measured from the water (200 µl), the oxic layer (top 1.5 cm) and from three anoxic depths (5 cm, 10 cm and 15 cm). All samples were transferred to the laboratory and analysed immediately. Other measurements of the environment included pH using an electronic pH meter, salinity and temperature. DNA and RNA were extracted in tandem from 0.5 g of sediment for sequencing, qPCR and RT-qPCR. All of the isolate and enrichment work was carried out on Stiffkey sediments, but Time 0 measurements were also taken from samples from Cley salt marsh (52.958649, 1.047364) and Yarmouth Estuary (52.613321, 1.716267) to enable comparison of DMSP levels, abundance and expression of specific genes.

2.10.2 Isolation of DMSP-producing bacteria

DMSP-producing bacteria were isolated from Time 0 Stiffkey salt marsh sediment, as well as from sediment that was treated to 14-day enrichment incubations under MBM media conditions designed to increase the occurrence of DMSP-producing strains (50 PSU, 0.5 mM nitrogen, 0.1 mM MTHB, 140mM sulfur). Samples from Time 0 and enriched samples were serially diluted and plated onto MBM minimal medium. Plates were incubated at 28ºC and after 72 h single colonies were obtained. Colonies with different morphologies were picked and tested for DMSP production. Positive isolates were purified and identified by PCR amplification of the 16S rRNA gene, using the primer set 27F/1492R (Lane et al. 1985). Purified PCR products were sequenced by Eurofins Genomics (https://www.eurofinsgenomics.eu, Munich, Germany), and the isolates were taxonomically identified using BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Isolates were screened for DMSP production in low nitrogen conditions by GC, and normalized to cellular protein content estimated by Bradford assays. Isolates were also checked for the presence of *dsyB* using degenerate primers (described below). To store bacterial isolates pure colonies were inoculated into 5 ml LB and incubated overnight at 28 – 37°C with shaking. Into a 2 ml screw-cap tube 525 µl of culture was mixed with 225 µl xx mM DMSO and 750 µl 50% glycerol, and frozen at -80°C. Viability of those stocks was tested every six months.

2.10.3 Microscopy

Cultures were grown to stationary phase before being examined by Microscopy. Aliquots of 5 µl were placed on glass microscope slides, then covered with glass coverslips, avoiding as many bubbles as possible. Drops of Immersion Oil were placed on the slide and the samples were observed under 100x light magnification using the Olympus BX40 microscope, equipped with an Olympus Camedia C-7070 digital camera. Samples were checked for contamination by searching for variations in morphology of observed cells in multiple areas of the glass slide. If none were observed, samples were declared to be pure.

2.10.4 Whole genome sequencing and analysis

Several isolated strains from Stiffkey were sent for Whole Genome Illumina sequencing to the MicrobesNG sequencing facility (https://microbesng.uk) at the University of Birmingham. These included *Novosphingobium sp.* BW1, *Marinobacter sp.* Set72 and *Alteromonas genovesnis*. The genomes were sequenced with the Illumina MiSeq platform producing 2 x 250 bp paired-end reads. These reads were trimmed using Trimmomatic and the quality was assessed using in-house scripts combined with the following software: Samtools, BedTools and bwa-mem. Annotation was performed with RAST (http://rast.nmpdr.org) (Aziz et al. 2008), the NMPDR, SEED-based, prokaryotic genome annotation service. The trimmed forward and reverse reads were uploaded and annotated against the genomic sequence of a closely-related species (e.g. *Novosphingobium sp* MBES04). For the genomic library screening, the fragment of *Novosphingobium* DNA found to contain the potential DMSP-producing gene was aligned against the sequenced genome, and all the functional genes in that region were analysed for the likelihood that they play a role in DMSP synthesis in *Novosphingobium* by using BLAST alignment to determine possible function.

2.10.5 Growth curves

Growth curves were measured for *Novosphigobium* and *T. profundimaris* to determine exponential phase for other growth experiments. Starter cultures were inoculated in 5 ml MBM and incubated for 16 hours or until reaching 0.6 OD_{600}. Triplicate flasks of 100 ml MBM were then inoculated with 2 ml of the starter culture, and incubated at 30°C with shaking at 200 rpm. Growth by OD_{600} was measured in 1 ml of culture taken every hour until levels reached stationary phase (the same OD reading for at least three hours), and measurements were averaged and plotted on a line graph.

2.10.6 DMSP production pathways induction experiment

Rough induction experiments were set up on several isolated strains. Cultures were inoculated in 10 ml triplicate MBM, in either standard conditions, (20 PSU with 12 mM nitrogen at that time) or MBM containing 5, 35 or 50 PSU salt levels, or lowered nitrogen levels at 0.5 mM nitrogen. Cultures in standard MBM were either incubated with 0.5 mM Met, DMSHB, MTHB or MMPA (intermediates of the transamination pathway). Cultures were incubated overnight at 30°C with shaking and DMSP quantity was measured using GC, and normalised for protein concentration.

An induction experiment specifically for the pathway intermediates was performed on *Novosphingobium*. A starter culture was OD_{600} adjusted to 0.6 and inoculated into 3 x 100 ml MBM, and incubated for 12 hours. The DMSP levels the Time 0 culture was detected, and then the cultures were measured into 5 ml aliquots and mixed with 0.5 mM of each of the intermediates Met, MTOB, MTHB, DMSHB, MMPA, MTPA, SMM and DMSP-amine, and a control mix with nothing else added. These mixed cultures were incubated at 30°C with shaking, and the DMSP concentration of 200 µl of each culture was quantified in triplicate, and protein content were measured at 30, 60, 120 and 240 minutes

2.10.7 DMSP production in environmental conditions

The effect of changing environmental conditions on the production of DMSP by *Novosphingobium* was tested from a starter culture of standard media (35 PSU MBM, 12 mM nitrogen), which was inoculated into 5 ml of either 50 PSU, 35 PSU, 5 PSU MBM with 12 mM nitrogen, 35 PSU MBM with 0.5 mM nitrogen, or standard media, in triplicate. All were incubated at 30°C overnight with shaking at 180 rpm, with the exception of one of the standard media cultures, which was incubated at 16°C instead. Measurements of protein content and DMSP concentration were taken and compared. For *T. profundimaris,* triplicate cultures were inoculated into MBM of salinity levels of 5, 35, 50 and 70 PSU, with 0.5 mM nitrogen (which was set as the 'standard' MBM condition from this point), to test the effect that salinity has on the production of DMSP. Cultures were also grown in 35 PSU MBM with high nitrogen levels of 12 mM to observe the impact. Cultures were incubated overnight, and DMSP levels quantified on 200 µl.

2.10.8 DMSP in seawater incubations

Purge-Trap measurements enable the detection of low levels of DMSP in seawater conditions. Strains were grown overnight in triplicate in MBM. Bacterial cells were harvested, washed three times and resuspended in filter-sterilised seawater (collected from Zhanqiao Pier, Qingdao, January 2018). The resuspended cultures were adjusted to an OD_{600} of 0.4 and then diluted 1:100 into 20 ml filter-sterilised seawater (T0), followed by incubation at 25°C with 90 rpm for 21 h (T1) and 43 h (T2). From the T0, T1 and T2

samples, bacterial cells were spun down and cell-free supernatants collected. The cell pellet was resuspended in Tris-HCl buffer (50mM, pH 7.5), and DMSP in the cells and cell-free supernatants were measured by alkaline-hydrolysis by adding 500 µl of 10 mM NaOH and incubating in dark overnight. Generated DMS was processed by a modified purge and trap method described by Zhang *et al*. (2018) and measured by Agilent 7890B gas chromatography (GC) with a flame photometric detector. An HP-5 (0°C – 325°C) 30 m x 320 µm x 0.25 µm capillary column (Agilent Technologies, Inc) was used to separate sulfur gases under the oven thermal cycle of 50°C to 120°C (20°C/min) to 180°C (30°C/min) to 50°C. The GC detection limit for DMS was ~0.015 nmol.

2.11 Extraction of DNA/RNA from Stiffkey

2.11.1 Extraction of DNA/RNA from sediment

DNA and RNA were extracted together following the protocol in Dumont et al. (2011) from marine sediment samples taken from Stiffkey, Cley and Yarmouth at Time 0 and from samples enriched for DMSP-producing bacteria after 14 days. Sediment samples were measured into 0.5 g aliquots and flash frozen in liquid nitrogen. To extract nucleic acids, 200 µl 0.1 mm silica beads (MP Biomedicals, Cambridge, UK) were added to the sediment alongside 1 ml extraction buffer (sodium dodecyl sulfate 87 mM; sodium phosphate buffer pH 8.0, 200 mM; sodium chloride 100 mM; ethylenediaminetetraacetic acid pH 8.0, 50 mM, Sigma-Aldrich), and bead beaten for 45 seconds at 6 m/s on a Bead Blaster 24 bead beater (Benchmark, Edison, NJ, USA). Samples were then centrifuged at 15 000 x *g* for 5 minutes at 4°C. The supernatant was carefully removed and mixed by vortexing with 850 µl Phenol:Chloroform:Isoamyl alcohol (25:24:1, Sigma-Aldrich), and centrifuged at max speed, 4°C for 5 minutes. The aqueous phase was removed and mixed by vortexing with 800 µl Chloroform:Isoamyl alcohol (24:1, Sigma-Aldrich), and centrifuged again. The aqueous phase was again removed and mixed with 1 ml Precipitation solution (polyethylene glycol 6000 20%; sodium chloride 2.5 M), and incubated for at least 1 h at room temperature. Samples were centrifuged for 30 minutes at max speed, 20°C and washed with 800 µl of ice-cold 75% ethanol, then centrifuged for 10 minutes at 4°C max speed. Ethanol was aspirated and the pellet air-dried and dissolved in 100 µl nuclease-free water. Aliquots of 50 µl were stored -80°C for RNA purification.

2.11.2 RNA purification from sediment extraction

To 50 µl of the DNA/RNA extract 37.5 µl of nuclease-free water, 10 µl of Buffer RDD (Qiagen), and 2.5 µl of DNase (Qiagen) were added. This was mixed gently and incubated at room temperature for 10 minutes. The RNA was cleaned following the

RNeasy Cleanup Protocol kit (Qiagen). The sample was mixed with 350 µl Buffer RLT and then 250 µl 100% EtOH before transferring to an RNeasy Mini spin column and centrifuging at 10,000 rpm for 15 seconds. Flow through was discarded, and 500 µl Buffer RPE was added, and tubes centrifuged as before. Flow through was discarded and a second wash of 500 µl Buffer RPE added, centrifuging for 2 minutes. The column was placed in a fresh collection tube and centrifuged at full speed for 1 minute before placing in a nuclease-free 1.5 ml collection tube. To elute 87.5 µl of nuclease-free water was added to the column and centrifuged for 1 minute at 10,000 rpm. The entire DNase treatment was repeated, only adding Buffer RDD and DNase (no need for water), and the clean-up repeated as well, before eluting in 2 x 30 µl of nuclease-free water. Aliquots were taken to quantify by Qubit 3.0 Fluorometer, following the protocol of the Qubit RNA High Sensitivity Assay Kit (Thermo Fisher Scientific). To confirm removal of gDNA 16S PCR and gel electrophoresis were also carried out. Samples were frozen at -80°C.

2.11.3 Extraction of DNA from pool water

Pool water from Stiffkey salt marsh was collected and 250 ml was passed through a 0.2 µm membrane filter by pump filtration. The membrane was cut into pieces and placed into 2 ml screw-cap tubes containing 200 µl 0.1 mm silica beads (MP Biomedicals, Cambridge, UK). To these tubes was added 600 µl STE buffer (NaCl 100 mM, Tris-HCl (pH 8.0) 10 mM, EDTA 1 mM), and tubes were bead beaten for 45 seconds at 6 m/s, then mixed with 10 µl lysozyme (10 mg/ml) and incubated at 37°C for 30 minutes, mixing every 10 minutes. Following this, 60 µl SDS (10 % w/v) and 6 µl protease (10 mg/ml) were added, and incubated at 65°C for 20 minutes, mixing every 10 minutes. To this 676 µl Phenol:Chloroform:Isoamyl alcohol (25:24:1) were added and mixed well, before centrifuging at 12,000 rpm for 10 minutes at room temperature. The aqueous phase was recovered to a fresh 1.5 microcentrifuge tube and mixed with 676 µl Chloroform:Isoamyl alcohol (24:1), then centrifuged as before. The aqueous phase was recovered to fresh 1.5 microcentrifuge tubes and the Chloroform step was repeated. The aqueous phase was recovered again, and mixed thoroughly with 0.7 x volume isopropanol to precipitate DNA. Samples were left at -20°C for 3 – 4 hours, then centrifuged for 15 minutes at 14,000 rpm, at 4°C. The supernatant was removed and the pellet washed with 800 µl ice-cold 75% EtOH, gently rolling the tube before tipping out the EtOH and repeating. All the EtOH was aspirated, and the pellet air-dried for 5 – 10 minutes, before being dissolved in 50 µl nuclease-free water and stored at -80°C.

2.11.4 Extraction of RNA from pool water

RNA was extracted from 200 ml pool water filtered through a 0.2 µm membrane filter by pump filtration, stored in RNAlater and flash frozen in liquid nitrogen. Extraction

was with Direct-zol RNA MiniPrep (Zymo Research, CA, USA). The work space was treated with RNaseZap, and Trizol (Tri Reagent, Sigma, Cat no. T9424) was heated to 65 °C in 1 ml aliquots. Samples were defrosted, the filters taken out, the RNAlater removed and the filter cut into pieces using 70% ethonol and RNaseZap cleaned scissors. The 1 ml pre-heated Trizol was added directly onto the filters along with 426-600 µm, acid-washed glass beads (Sigma). Cells were disrupted using a Mini-Beadbeater and 3 cycles of 30 seconds with 1 minute recovery. Samples were incubated at room temperature for 5 minutes, and centrifuged for 2 minutes at 15,000 rpm, 4°C. The supernatant was transferred into fresh 2 ml screwcap tubes and mixed by vortexing with 1 ml EtOH (95 - 100%). Up to 500 µl of sample was loaded onto the spin column and centrifuged for 1 minute at 15,000 rpm, 4 °C. Flow through was discarded and steps repeated for residual sample. To the column 400 µl RNA preWash was added and centrifuged for 1 minute as above. Flow through was discarded, and the wash step repeated. Aliquots of 700 µl RNA Wash Buffer was added and samples centrifuged as above, and flow through discarded, then centrifuged again for 2 to remove buffer. The spin column was placed in a fresh 1.5 ml RNAse-free tube with 100 µl of nuclease-free water and incubate for 5 min at 4°C. Columns were centrifuged for 1 minute at 4 °C to elute RNA. DNA was removed using Turbo DNase (Ambion). To the RNA samples 10 µl of 10x Turbo DNase buffer was added, with 1 µl Turbo DNase and incubated at 37°C for 40 minutes. This step was repeated to remove all the gDNA contamination, before 10 µl of stop buffer (DNAse Inactivation Reagent) was added, and incubated for 2 minutes at room temperature, shaking occasionally. Samples were centrifuged for 2 minutes, 4°C and the supernatant transferred to a fresh tube. A 5 µl aliquot of RNA was quantified by PCR to check for DNA contamination, and agarose gel analysis. All samples were frozen in liquid nitrogen and stored at -80°C.

2.11.5 RNA purification and reverse transcription

Reverse transcription of RNA was carried out on as close to 100 µg DNA-free RNA as possible. Up to 9 µl RNA (≤ 2 µg) was mixed with 1 µl 10 µM specific reverse primer (dsyB_deg-R for DsyB identification), and incubated for 5 mins at 70 °C, then cooled briefly on ice. Per sample 1 µl 1:1:1:1 mix of dNTPs (10 mM), 4 µl M-MLV 5 x reaction buffer (Promega), 0.4 µl RNase Inhibitor (40 U/µl, Roche), 0.8 µl M-MLV reverse transcriptase (200 U/µl, Promega) and 3.8 µl nuclease-free water were added, sample mixed and incubated at 42 °C for 1 h. The stable cDNA was quantified using a Qubit 3.0 Fluorometer, following the protocol of the Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific), and was stored at -20 °C until needed.

2.12 Degenerate primers

2.12.1 Degenerate primer design and optimisation

Degenerate primers were designed to enable amplification of multiple genera from DNA extracted from environmental samples. The primers were designed by means of an amino acid alignment of the 24 DsyB sequences using the ARB project program (http://www.arb-home.de). Conserved regions between species were identified, and the degeneracy of possible primers calculated, with a cutoff of degeneracy of up to 5 degenerate bases. Options were synthesised by Eurofins Genomics and tested against a set of positive and negative genomic DNA controls. Positive controls included five bacterial strains known to contain *dsyB,* and the negative controls included alphaproteobacterial strains unable to produce DMSP, and known eukaryotic DSYB sequences to test specificity to bacterial DsyB. The most effective primer combination (dsyB_deg1F and dsyB_deg2R, Table 2) amplified a fragment of ~246 bp in size from all positive controls tested, with minimal non-specific bands being amplified. The primers were optimised for the annealing temperature (tested between 60 – 65°C), extension time (from 15 – 60 seconds) and number of cycles (between 30 to 40) to give the most specific amplification, with the final program involving an initial denaturation step of 95°C for 5 minutes, followed by 35 cycles of 95°C for 30 seconds, an annealing step of 61°C for 15 seconds and an elongation step of 72°C for 15 seconds, ending in a final extension of 72°C for 5 minutes.

A similar method of optimisation was adopted when designing degenerate primers for the amplification of *mmtN* from multiple species, although it was found that the sequences were too divergent for a single primer set to amplify from all species, so while the most optimal primers (mmtN_degF and mmtN_degR, **Table 2-2**) successfully amplified a ~281 bp fragment of *mmtN* from *N. sp* BW1, *R. indicus* and *T. profundimaris*, they were unable to amplify from *S. mobaraensis* and *N. chromatogenes* genomic DNA. A solution to this could be to design clade-specific primers, as the *mmtN* sequences have been identified in several different classes. Optimisation of the degenerate primers to amplify no non-specific bands involved testing annealing temperatures between 50 – 60°C, and extension times from 15 – 60 seconds, and the final program had an initial denaturation step of 95°C for 5 minutes, followed by 35 cycles of 95°C for 30 seconds, an annealing step of 60°C for 30 seconds and an elongation step of 72°C for 30 seconds, ending in a final extension of 72°C for 5 minutes.

2.12.2 pGEM-T Easy cloning

For creating clone libraries and standards for qPCR, the pGEM-T Easy Vector System (Promega) was used. Fragments to be cloned were amplified using PCR and then purified by gel extraction. Ligations were set up after calculating the appropriate volumes of PCR products using the insert:vector molar ratio 1:3. This equation was used to calculate volumes:

$$\frac{\text{ng of vector (50 ng) x kb size of insert}}{\text{kb size of vector (3 kb)}} \times \frac{3}{1} = \text{ng insert}$$

A standard ligation mix consisted of 5 µl 2X Rapid Ligation Buffer, 1 µl pGEM-T Easy Vector (50ng), X µl PCR product, 1 µl T4 DNA ligase (3 Weiss units/µl) and nuclease-free water to a final volume of 10 µl. Positive controls were also set up with 2 µl control insert DNA in the place of the PCR product. Reactions were mixed well and incubated overnight at 4°C. Ligations were transformed by heat shock into *E. coli* JM101 competent cells as described above, with 5 µl ligation added to 100 µl competent cells, alongside controls. Transformed cells were plated on LB/ampicillin/IPTG/X-Gal agar plates, with 100 µl of the ligation on one, and the rest on another, alongside the positive control and a negative control of cells only. Plates were incubated overnight at 37° and then checked for successful cloning using a blue-white screen, where white colonies have the *lacZ* gene successfully disrupted. These are picked and checked using restriction digests or PCR.

2.12.3 *dsyB* clone library construction

Clone libraries were prepared from *dsyB* primer gene fragments, PCR-amplified from DNA extracted from Stiffkey salt marsh using the degenerate primers dsyB_deg1F and dsyB_deg2R. Fragments were amplified using the PCR protocol described above, imaged using gel electrophoresis and the single specific bands were extracted by gel extraction. These fragments were then cloned into the pGEM-T plasmid using the pGEM-T Easy Vector System I cloning kit (Promega), described above. Transformants were plated on LB/ampicillin/IPTG/X-Gal agar plates, and 19 clones in total were picked and inoculated to 5 ml LB Ampicilin, and the plasmids were extracted using QIAGEN minipreps, then sent for to Eurofins MWG for sequencing.

2.13 Gene library construction

2.13.1 QIAGEN Genomic DNA extraction

High quality and high volume genomic DNA extractions were carried out using the QIAGEN Genomic DNA extraction kit. Once the genomic DNA was eluted with 1 x 5 ml of

Buffer QF. DNA was precipitated by adding 3.5 ml (0.7 volumes) of room-temperature isopropanol to the eluted DNA, and inverting the tube 10 to 20 times. The DNA was collected using a sterile 5 ml pipette tip and transferred to a microcentrifuge tube containing 1 ml of TE buffer (pH 8.0, or 10 mM Tris·Cl, pH 8.5). The DNA was dissolved on a shaker at 55°C for up to 2 hours.

2.13.2 Gene library construction

To identify the gene or genes involved in the synthesis of DMSP by *Novosphingobium,* a genomic library of *Novosphingobium sp.* BW1 was constructed so that fragments could be screened in the wide-host species *R. leguminosarum* J391. The method followed was essentially described in (Curson et al. 2008). *Novosphingobium* genomic DNA was extracted using a QIAGEN Genomic-tip 100/G kit, and test digestions with EcoRI were carried out to determine the stage at which the genome was partially digested into roughly 25 – 30 kb fragments (usually digestion continued for 5 – 10 minutes), before being flash-frozen in liquid N_2 to halt the digest, and run on a 0.5% agarose gel to assess fragment size. Once a time was confirmed, up to 10 µg of genomic DNA was partially digested, then 100 µl of the digest was transferred to a tube containing 200 µl 100% ethanol and 10 µl 3M sodium acetate (pH4.8), and frozen in liquid N_2 to stop the reaction.

The DNA was ethanol precipitated and quantified, and at least 2.5 µg of genomic DNA was used to ligate into the EcoRI-digested, dephosphorylated cosmid vector pLAFR3. The ligation was ethanol precipitated resuspended in 17 µl nuclease-free water, and then 0.7 µg was packaged, ready to transfect into *E. coli* 803, using the Stratagene Gigapack III XL Packaging mix. The mix was removed from the -80°C freezer until partially thawed, at which point the genomic DNA was added and mixed by stirring with the pipette tip. Tubes were incubated for 2 hours at 22°C, before being mixed with 500 µl SM buffer and 20 µl chloroform. The supernatant containing the phage was removed and stored as glycerol. The packaged genomic DNA fragments were transfected into the *E. coli* strain 803, which was prepared by inoculation in 100 ml LB supplemented with 10 mM $MgSO_4$ and 0.2% (w/v) maltose, and incubated at 37°C for 4 – 6 hours, or 30°C overnight (not reaching above an OD_{600} of 1). The cells were pelleted by centrifuging at 500 x g for 10 minutes, and then resuspended in 10 mM $MgSO_4$ to an OD_{600} of 0.5. To titre the library and determine the number of clones the packaged mix was diluted either 1:10 or 1:50 in SM buffer, and 1 µl was mixed with 200 µl host cells and incubated at 37°C for 15 minutes, topped up with LB (up to 1.5 ml) and plated on LB tetracycline to select for pLAFR3 cosmids. The resulting colonies were counted and the number of clones in the library calculated – in total the library consisted of an estimated 50,000 clones. Another transfection into *E. coli* was set up with a higher volume of packaged cosmids – 25 µl of

cosmid mix and *E. coli* were mixed 1:1 and incubated at 37°C for 30 minutes, then mixed with 200 µl LB and incubated for another hour, gently shaking every 15 minutes. Cells were pelleted and resuspended in 50 µl LB, then plated on LB tetracycline. Up to six colonies were tested for the correct-sized fragments by digestion with EcoRI, BamHI and HindIII. They were also tested to make sure the fragments were different.

Multiple rounds of infection into *E. coli* were set up to ensure a high number and variation of cosmids, then pooled together and stored in glycerol at -80°C. The clones were crossed via triparental mating to *R. leguminosarum* J391, and a total of 750 transconjugants were picked to RM medium containing 0.5 mM Met, incubated overnight at 30°C and then screened by GC for those conferring the ability to produce SMM to *R. leguminosarum* J391 (as a result of MMT activity) by mixing with 100 µl 10 M NaOH and heating at 80°C for 10 minutes, then incubating in the dark overnight before quantifying DMS produced. Positive samples were checked by re-inoculation and repeated screening. The plasmids were extracted and transformed into *E. coli* 803 to be mobilised back into *R. leguminosarum* J391 by tri-parental cross, before re-confirming MMT activity. Positive cosmids were digested with *Eco*RI, *Bam*HI, *Hind*III and *Pst*I to demonstrate the presence of inserted fragments in the pLAFR3 cosmid, and compare fragments.

2.13.3 Tri-parental crossing

Tri-parental crossing was utilised to transfer plasmids or cosmids from *E. coli* to *Rhizobium.* It involves three strains: the host strain of *Rhizobium*, the donor strain of *E. coli* that contains the plasmid or cosmid to conjugate into the Host strain, and the helper strain, which is the kanamycin-resistant *E. coli* strain 803 (pRK2013). A 5 ml universal of TY media was supplemented with appropriate antibiotics and the *Rhizobium* host strain, and incubated at 28°C with shaking. The helper plasmid *E. coli* 803 and donor strain were both inoculated in 5 ml LB with antibiotics and incubated at 37°C with shaking overnight. A 1 ml aliquot of the host was centrifuged at maximum speed for 1 min. The supernatant was removed and the pellet resuspended in 500 µl fresh TY media. This was repeated three times to wash out residual antibiotics, before being centrifuged again and resuspend in 100 µl of TY media. The donor and helper strains were treated in the same way and resuspended in 100 µl TY. A sterile filter was placed on a TY agar plate with no antibiotics using ethanol-sterilised forceps. Aliquots of 100 µl of each strain were added to the filter and mixed using a sterilised loop. Control crosses were also set up with just the helper and host, and just the donor and helper. Plates were incubated at 28°C overnight. Ethanol-sterilised forceps were used to remove the filters and place them in sterile universals. The cells were washed off the filter using 2 ml of 50% glycerol, and then plated at a suitable dilution on selective TY plates containing kanamycin and other selective antibiotics from

the donor DNA. Plates were incubated at 28°C and successful crosses confirmed using colony PCR.

2.14 Identification of MmtN enzymes, phylogenetic trees

BLAST searches to identify homologues of the *Novosphingobium* MmtN protein were performed using BLASTP at NCBI or JGI. MmtN homologues, along with selected other more distantly related methyltransferases in Pfam family PF10672 below the predicted cut-off for MMT functionality (E values ≤ $1e^{-90}$, iIentity = 36 %), were aligned by ClustalW in MEGA v6 and visualised in a maximum-likelihood phylogenetic tree to observe the relatedness of the sequences. Predicted non-functional MmtN sequences that are just below the cut-off are *Candidatus Taylorbacteria bacterium* and *Candidatus Peregrinibacteria bacterium.*

2.15 Quantitative PCR with reverse transcription RT-qPCR

Standards for qPCR were created by PCR amplifying a fragment from either community DNA or genomic DNA using primers to the gene of interest. The amplified fragment was excised from the gel and purified by gel extraction, then cloned into the pGEM-T Easy vector (described above). Once positive clones had been identified the colonies were cultured and the plasmids extracted using the QIAGEN Miniprep kit. These were sequenced to confirm identity and then digested using FastDigest enzymes in a 50 µl reaction, performed on 8 µg DNA with 2 µl of the *Nde*I restriction enzyme to linearise the plasmid. To confirm the linearized plasmid was run on gel electrophoresis for 1 hour, and the band extracted and purified through gel extraction, eluting in 30 µl, resulting in > 50 ng of standard. To calculate the copy numbers in the standard this calculation is used, and samples are diluted to 10^{-8} ready for qPCR, and dilutions were repeated every two months.

$$\text{Number of copies} = \frac{(6.02 \times 10^{23}) \times \text{ng in digest}}{(\text{fragment bp} + \text{plasmid bp}) \times (1 \times 10^{9}) \times 650}$$

To perform quantitative PCR, a master mix was made up for reactions of 20 µl aliquots, with 10 µl 2 × SensiFAST SYBR mastermix, 400 nM of both the forward and reverse primers **(Table 2-2)** and 2 µl cDNA, or 2 µl of 1/10- or 1/100-dilutions of DNA. The 18 µl aliquots were added to a 96-well qPCR plate and 2 µl of either standard or template added, alongside three wells that were the no template control (NTC) with only master mix in. A single gene was quantified per run, with three biological replicates and three technical

replicates. The plates were sealed and centrifuged for a few seconds to ensure bubbles are removed.

Quantitative PCR was performed with a C1000 Thermal cycler equipped with a CFX96 Real-time PCR detection system (BioRad), using a SensiFAST SYBR Hi-ROX Kit (Bioline) as per the manufacturer's instructions for a three-step cycling programme. For 16S rRNA qPCR a 95°C initial denaturation step for 3 minutes was followed by 40 cycles of 95°C for 5 seconds, 53°C for 10 seconds and 72°C for 25 seconds, at which point the fluorescence was quantified. For the melt curve stage, the initial temperature was 95°C for 15 seconds, and then increased from 70°C to 95°C with the data collected every 0.2°C increase. *dsyB* qPCR involved a 95°C initial denaturation step for 3 minutes was followed by 40 cycles of 95°C for 20 seconds, 60°C for 30 seconds and 72°C for 30 seconds, at which point the fluorescence was quantified. For melt curve analysis, the initial temperature was 95°C for 1 minute, and then increased from 60°C to 95°C with data collected every 0.5°C increase. For each condition and gene, the cycle threshold (Ct) values of the technical and biological replicates were averaged and manually detected outliers were excluded from further analysis. Standard curves of control DNA were calculated from five points in 1:10 dilutions, and used to calculate copy numbers in the samples.

2.16 General *in vivo* and *in vitro* genetic manipulations

Plasmids **(Table 2-3)** were transferred to *E. coli* by transformation, or *R. leguminosarum* J391 or *T. profundimaris* by conjugation in a tri-parental mating using the helper plasmid pRK2013. Routine restriction digestions and ligations for cloning were performed essentially as in Downie et al. (1983). The oligonucleotide primers used for molecular cloning were synthesised by Eurofins Genomics and are detailed in **Table 2-2**. Sequencing of plasmids and PCR products was performed by Eurofins Genomics.

2.17 S-methyl methionine transferase assays

The SAM-MMT genes from *Novosphingobium, R. indicus, T. profundimaris, S. mobaraensis* and *N. chromatogenes* were cloned into pET21-a by specific primer-amplifying fragments from genomic DNA **(Table 2-2)** that were digested with *Nde*I and *EcoR*I (*BamH*I for *R. indicus*) restriction enzymes. All plasmid clones are described in **Table 2-3**. To measure SMM activity from pET21a clones expressing the *mmtN* gene in *E. coli* BL21, cultures were grown (in triplicate) overnight in LB complete medium, 1 ml of culture was spun down, resuspended in the same volume of LB medium and diluted 1:100

into 5 ml LB and incubated for 2 h at 37 °C. This was then induced with 0.2 mM IPTG (Sigma-Aldrich) and incubated at 30 °C overnight. For each culture, 1 ml of culture was mixed with 0.5 mM L-Met (Sigma-Aldrich), and incubated for 8 hours at 30°C before sampling for GC analysis to determine the amount of SMM produced (see 'Quantification of DMS/DMSP/SMM by gas chromatography'). Protein concentrations were determined using the Bradford method (BioRad), or using Qubit when higher sensitivity was required. Controls run included the media alone, *E. coli* BL21 and *E. coli* BL21 containing the empty pET21a vector.

2.18 Purification of MmtN and in vitro catalytic assays

2.18.1 SAM Charcoal affinity testing

To confirm the sequestering of SAM by activated charcoal 1.5 ml microcentrifuge tubes containing either 0.5 mM SAM, 0.5 mM SMM, or both, in 2 x 200 µl sterile water were set up. The DMS released from the sample was quantified, then samples were mixed with 400 µl of an activated charcoal solution (38 mg ml−1 in 0.1 M Acetic Acid), incubated for 5 minutes at room temperature, and centrifuged at mazimuum speed for 1 minute to remove the charcoal and the compounds adsorbed to it. The remaining supernatant was carefully removed and the DMS measured and compared to levels before charcoal.

2.18.2 *Novosphingobium* cell lysate activity

For *Novosphingobium* cell lyaste experiments, cultures were inoculated in YTSS then harvested by centrifugation and resuspension into a 50 mM Tris-HCl buffer. Samples were sonicated to lyse the cells, then centrifuged at maximum speed to pellet debris, and the lysate was removed. This lysate was dialysed to remove any pre-existing metabolites, using dialysis tubing (SpectrumLabs) in 2 litres of dialysis buffer (20 mM HEPES, 150 mM NaCl, pH 7.5) at 4°C overnight. From this lysate 2 x 200 µl was mixed with either 1 mM SAM, 1 mM Met, or both, and then incubated for 30 minutes at room temperature, allowing for cell extract activity assay. MMT activity was measured in the samples by adding 100 µl 10 M NaOH and heating for 10 minutes at 80°C before quantifying by GC.

2.18.3 Purifying MmtN

The MmtN protein was expressed in *E. coli* BL21 cultures grown in LB media at 37°C, to an OD_{600} of 0.8 – 1.0, and then induced at 20°C for 16 hours with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The protein was purified first with Ni^{2+}-NTA resin (QIAGEN, Germany), and then fractionated using gel filtration buffer (10 mM Tris-HCl [pH 8.0] and 100 mM NaCl) on a Superdex-200 column (GE Healthcare, America). Purification of the protein took place at 4°C. For the Ni^{2+}-NTA resin purification, wash buffer (50 mM

Tris-HCl [pH 8.0], 250 mM NaCl and 20 mM imidazole) was used to remove protein impurities, followed by the elution buffer (50 mM Tris-HCl [pH 8.0], 250 mM NaCl and 250 mM imidazole) to elute the purified protein from the column. MmtN enzyme activity was measured by monitoring the production of SAH (S-adenosyl homocysteine) produced by the demethylation of SAM, detected by HPLC. Optimal MmtN activity was determined by testing temperature and pH conditions, and comparing enzyme activity, with the highest activity defining 100 % activity, and other tested conditions described as relative to it. The reaction mixtures were incubated at temperature intervals of 10°C, from 0°C to 60°C, for 30 minutes For optimal pH levels MmtN activity was examined using Brtitton–Robinson buffer (40 mM H_3BO_3, 40 mM H_3PO_4 and 40 mM CH_3COOH), at pH values between pH 5.0 and pH 10.0. The kinetic parameters (K_m) were determined by non-linear analysis, based on the initial rates and determined using 3.34 µM MmtN and 0.1 – 4 mM SAM, or 0.1 – 6 mM Met

2.19 Mutagenesis of *mmtN* and phenotyping

A *T. profundimaris* spontaneous Rif-resistant mutant was created to enable selection of *T. profundimaris* away from *E. coli*. A highly concentrated number of cells (10^{10}) was plated on MB-Rif plates and incubated for 48 - 72 hours, until colonies grew. These were picked and tested to confirm resistance. The *T. profundimaris*-Rif strain was then treated as wild-type for all the experiments performed comparing the *mmtN*$^-$ to wild-type. Primers were designed to amplify a fragment of 500 bp internal to the *T. profundimaris* WPO211 *mmtN* open reading frames **(Table 2-2)** as well as containing restriction sites for the enzymes *Bam*HI and *Eco*RI This fragment was cloned into pBIO1879 (Todd et al. 2011), a derivative of the suicide vector pK19mob to form pBIO19TK. This was transferred to *T. profundimaris*-Rif+ by tri-parental conjugation, using the helper strain *E. coli* pRK2013. Mutants in which the plasmids had recombined in the target genes were selected for by growth on YTSS agar containing rifampicin (WPO211), kanamycin (pBIO1879) and spectinomycin (pBIO1879). Potential *T. profundimaris mmtN* mutant colonies were isolated (~200 colonies at 10^0 dilution), and all were checked for DMSP production and were confirmed to have worked by digestion with with *Bam*HI and *Eco*RI to show the insert. PCR was also used with primers designed to either side of where the plasmid inserts – gels that show no DNA likely have the enormous plasmid inserted into the DNA. The mutant was complemented by crossing the pBIO21N1 plasmid containing the *Novosphingobium mmtN* gene back into *T. profundimaris* and observing the return of at least some function.

To identify a phenotype for the mutations in *mmtN*, *T. profundimaris* wild type and *mmtN*− strains were grown in MBM with varying levels of salt and nitrogen, and tested for survival after freezing. To test the effect of salinity on the mutant, the wild type and mutant strains were grown in triplicate MBM minimal medium with 0.5 mM nitrogen (now standard) made with different amounts of sea salts (Sigma-Aldrich) equivalent to salinities of 35 and 50 practical salinity units (PSU), with 35 being the approximate salinity level of sea water, and growth of the strains was monitored spectrophotometrically by the optical density at 600 nm (OD_{600}) until reaching stationary phase. OD values were measured every hour. To test the effect of nitrogen levels, the strains were grown in 35 PSU MBM, with 12 mM nitrogen to determine the effect of higher levels of nitrogen. Growth was monitored by OD_{600}. To test the tolerance to freezing, cultures of the wild type and mutant strains were grown to stationary phase in 35 PSU MBM (0.5 mM NH_4Cl) then adjusted to the same cell density by measuring the OD_{600} of each culture, spinning down an appropriate volume (~1 ml) of culture and then resuspending the cells in 1 ml of the same medium. A 100 µl volume of each culture was removed, serial diluted and then plated on MB agar plates to count the number of colonies that grew after 2–3 days growth and used to calculate the percentage of cell survival for the two strains after exposure to freezing. The remaining 900 µl of culture was placed at −20 °C for 5 days before thawing, serial dilution and plating as above. To further test for any phenotype changes in survival for the two strains competition experiments were performed. Cultures of the wild type and mutant strains were grown to stationary phase in 35 PSU MBM (10 mM NH_4Cl) and mixed in equal parts (500 µl of both). The mixed culture was plated for single colonies on MB agar and these were picked after 2 days of growth at 30 °C and tested for kanamycin/spectinomycin resistance to determine survival.

The *T. profundimaris mmtN*$^-$ mutant was complemented with an *mmtN* gene to observe a return to function. The plasmid that was used was the *Novosphingobium mmtN* gene that was subcloned from pBIO21N1 into the taurine-inducible wide-host range plasmid, pLMB509 using the *Nde*I and *Eco*RI restriction enzymes, creating pBIO509N.. As it was cloned in pLMB509, it was resistant to gentamycin, meaning that it could easily be selected for when mobilised into the mutant. The *mmtN* clone in pBIO509N was mobilised into the *T. profundimaris*-Rif *mmtN*$^-$ mutant through tri-parental mating, and positive colonies growing on genetamycin were tested for the return of DMSP production by GC.

2.20 Statistics

All measurements for DMSP production or DsyB/MmtN enzyme activity (in cell lysate experiments or enzyme assays) are based on the mean of at least three biological

replicates per strain/condition tested, as are the metegenomes and 16S amplicon sequencing.

2.21 Sequencing and analysis

2.21.1 16S rRNA amplicon sequencing

The 16S rRNA gene amplicon sequencing analysis of the DNA extracted from the grassland soil samples was performed by MR DNA (Shallowater, TX, USA). Three biological replicates of each condition were analysed. The primer set 515F/806R of the V4 variable region of the 16S rRNA gene (Caporaso *et al.*, 2012) was used in the PCR reaction, with the former being barcoded. The PCR reaction consisted of an initial step of 94 ºC for 3 min, followed by 28 cycles of 94 ºC for 30 s, 53 ºC for 40 s and 72 ºC for 1 min, after which a final elongation step at 72 ºC for 5 min was performed. Samples were later purified using calibrated Ampure XP beads. Purified products were used to prepare an Illumina DNA library. Sequencing was performed on a MiSeq system according to the manufacter's instructions and data were processed using the MR DNA analysis pipeline, obtaining an average of 47 984 reads per sample with an average length of 300 bp. The data processing included joining the sequences, depleting of the barcodes, removing sequences <150 bp and sequences with ambiguous bases. Resulting sequences were denoised, operational taxonomic units (OTUs) generated and chimeras removed.

Sequencing was run on the hits and files from the runs were converted to OTU tables, joined in Qiime v1.8. The samples with fewer than 150 bp in size or ambiguous bases were filtered out. After running preliminary summary statistics on the data, all samples were rarefied to 36,066 sequence counts per sample. The joined tables were then split according to type of sample; time 0, control or nriched. Each group of samples were analysed separately at the genus level, and the genus-level tables and corresponding meta data files were uploaded to a Calypso bioinformatics program (http://cgenome.net/wiki/index.php/Calypso) (Zakrzewski et al. 2016). Data were normalized using total sum normalisation to convert raw counts to relative abundances. Taxa with less than 0.01% mean relative abundance across all samples were removed. This kept 330 genera, excluding 491 from the original 821 genera. Rarefaction curves were created to demonstrate species richness, with average number of species (richness) plotted against number of reads sampled.

2.21.2 Metagenomic sequencing

Samples for Time 0, Control and Enriched sample groups were combined in equal parts to create pooled samples of the three conditions, in triplicate, on which metagenomic

analysis could be performed. This sequencing was also carried out by Mr DNA, at 2 x 150 bp 10-20 million paired sequences per sample. Libraries of DNA extracted from samples were prepared using the Nextera DNA Sample preparation kit (Illumina, San Diego, CA, USA) following the manufacturer's user guide. The initial concentration of DNA was evaluated using the Qubit® dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, USA). The samples were then diluted to achieve the recommended DNA input of 50 ng at a concentration of 2.5 ng·µl^{-1}. Samples underwent simultaneous fragmentation and addition of adapter sequences. These adapters were incorporated over 5 cycles of PCR. Following the library preparation, the final concentration of the library was measured using the Qubit® dsDNA HS Assay Kit (Life Technologies), and the average library size was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies). The average library size Time 0 samples was 826 bp, 931 bp for Control samples and 1364 bp for Enriched. The library was then pooled in equimolar ratios of 2 nM, and 10.5 pM of the library pool was clustered using the cBot (Illumina) and sequenced paired end for 300 cycles using the HiSeq 2500 system (Illumina). Reads were quality-filtered and trimmed using Trimmomatic (Bolger *et al.*, 2014), obtaining an average of 13 909 226 reads per sample with an average length of 151 bp. Metagenomes were then assembled using SPAdes assembler with kmers 55 to 127 (Bankevich *et al.*, 2012), and assemblies were analysed using Quast (Gurevich *et al.,* 2013). N50 values were ~1 kb for all metagenomes assemblies.

The abundance of functional genes in unassembled metagenomes was determined by tBLASTx (www.ncbi.nlm.nih.gov) of selected ratified gene sequences (*dsyB, mmtN, DSYB, Alma1, ddd* genes) against the raw reads (E≤e^{-4}). Each potential sequence retrieved from the analysis of metagenomes was manually curated by BLASTp against the RefSeq database and discounted as a true sequence of interest if the top hit was not to a known sequence. Only unique hits were counted. Hit numbers were normalised against read number of the smallest sample, to gene length and to hits of *recA*.

2.21.3 Analysis of metagenomes/metatranscriptomes

Hidden Markov Model (HMM)-based searches for *mmtN* homologues in metagenome and metatranscriptome datasets were performed as described in (Curson et al. 2018) using HMMER tools (version 3.1, http://hmmer.janelia.org/). The MmtN protein sequences were used as training sequences to create the HMM profiles. Profile HMM-based searches eliminate the bias associated with single sequence BLAST queries. HMM searches were performed against peptide sequences predicted from OM-RGC database assemblies and all hits with an E value cut-off of 1e^{-30} were retrieved. In the case of metatranscriptome datasets (*Tara* Oceans and GeoMICS metatranscriptomes), homologues with an E value cut-off of 1e^{-5} were retrieved. Each potential MmtN sequence

retrieved from the analysis of metagenomes and metatranscriptomes was manually curated by BLASTP analysis against the RefSeq database, and discounted as a true MmtN sequence if the top hits were not to a recognised MmtN. To estimate the percentage of bacteria containing *mmtN*, the number of unique hits to MmtN in metagenomes was normalized to the number of RecA sequences.

CHAPTER 3

DEVELOPING *dsyB* GENE PROBES

3 DEVELOPING *dsyB* GENE PROBES

3.1 Introduction

3.1.1 Using gene probes to study *dsyB*

One of the aims of this work was to study the diversity and abundance of *dsyB* (the first bacterial DMSP-synthesis gene to be discovered) in marine environments, as well as to identify it in multiple bacterial species. This is because Curson et al. (2017) are confident that it can be utilised as a reliable reporter for bacterial DMSP production, as the presence of *dsyB* in an organism is likely indicative of its ability to produce DMSP. Therefore, by studying its distribution in an environment we are, by extension, studying the distribution of the potential for bacterial DMSP production in any tested environment. Furthermore, studying the transcription of *dsyB* in a given environment would give a good indication that the microorganisms possessing this gene are likely producing DMSP. There are both culture-dependent and culture-independent methods through which we can investigate the abundance, diversity and expression of *dsyB* in the environment or in model organisms, including metagenomics, metatranscriptomic sequencing, metaproteomics and also through the use of gene probes on environmental DNA and RNA. This chapter focuses on the design, optimisation and validation and use of *dsyB* gene probes.

There are multiple ways in which gene probes can be used, including in southern blotting (McDevitt et al. 2002), fluorescence in situ hybridization (FISH) (Mühling et al. 2008), and other fluorescence experiments that involve nucleic acid sequences complementary to a target sequence (McLenon & DiTullio 2012). Indeed, primers designed to the bacterial 16S ribosomal RNA gene are a type of gene probe, be they designed to a phylum/class for use in DGGE analysis (Mühling et al. 2008), or more generalised for phylogenetic identification (Lane 1991). For the purpose of this study however, it was decided that gene probes in the form of degenerate primers designed to functional *dsyB* was the most suitable approach.

Degenerate primers are oligonucleotides specifically designed to anneal to conserved components of a functional gene, in this case *dsyB*. The design process involves aligning known functional gene sequences, identifying conserved regions and then designing primer sets which allow you to specifically target this gene in complex nucleotide preps. The aim is to amplify one gene from a phylogenetically diverse group of bacteria, where the gene sequences are similar but not identical (Ashelford et al. 2002). The primers are designed at the amino acid level as, given the degeneracy of some amino acids, there are more conserved regions to observe. In order for a primer to be classed as

‘degenerate’ it must contain one or more degenerated base pairs in the sequence, accounting for single nucleotide differences in the nucleic acid sequences (Iserte et al. 2013). These bases have varying levels of degeneracy, with some encoding a fifty-fifty chance between two bases at that site, and others offering three or even all four options in equal proportion during manufacture (Iserte et al. 2013).

These degenerate primers can be exploited in several ways that allow us to study multiple aspects of *dsyB* in the environment. Firstly, they can be used in PCR reactions to test isolated species for the presence or absence of the gene, without having to send the whole genome for sequencing. This is cost and time effective, and allows for large-scale screening of bacterial isolates. Furthermore, these primers can be used in diversity assay amplicon sequencing of an environment to observe the distribution of *dsyB*-containing bacteria, and how this distribution changes between environments or when treated under different conditions. Finally, these primers, if designed properly (described below), can also be used in quantitative PCR experiments on DNA and cDNA, enabling us to study both the abundance and transcription of *dsyB*. The latter use is an extremely powerful tool to estimate the significance of a process in an environment, e.g. bacterial DMSP production in Stiffkey salt marsh sediment, the focus of this thesis.

3.1.2 Use of degenerate primers in literature

Degenerate primers (or variations of them known as mixed primers) have been used in research for years and are often utilised as a means to study functional genes. In the literature they have aided in the answering of a variety of questions, from elucidating the full sequence of a gene from a known fragment to identifying entirely new or uncharacterised sequences of a particular gene (Compton 1990).

One example of how degenerate primers have been used to isolate and clone a full-length gene from only partial amino acid sequences is seen in Lee et al. (1988) in the study of a urate oxidase, an enzyme involved in the oxidation of uric acid to allantoin in most mammals (excluding humans and some primates). At the time of publishing, DNA sequencing was more complex and time-consuming than it is today, so this technique of using mixed oligonucleotide primer amplification meant that the sequencing could be more targeted. The identities of the first 32 amino acids of a porcine urate oxidase were determined and, due to the degeneracy of many of these amino acids, mixed oligonucleotide primers were designed to that section. These primers were used to amplify from the reverse transcription of the gene (single-stranded cDNA), and clones that were the expected size were checked using Southern blotting with an internal probe, then sequenced using the dideoxy procedure. This process resulted in rapid generation of a cDNA probe that was then used to screen for the full-length porcine sequence from a

cDNA library. This is an example of the convenience of degenerate primers in that there is less of a requirement for complete specificity, allowing for variation and unknown sequences much more than with regular primer design.

Another example of this early use of mixed primers is seen in Girgis et al. (1988). In this paper primers representing all codon choices for each amino acid were designed to the first and last five amino acids of the diabetes associated peptide (DAP). This method producing a band of the predicted size from genomic DNA, and resulted in probes that could be used to clone the full DAP sequence even with limited amino acid sequence information.

A similar strategy has also been used in the discovery of entirely new or uncharacterised sequences that are related to a known gene family. Using conserved regions of known sequences, Ehlers et al. (1999) designed degenerate and deoxyinosine-substituted primers that could successfully amplify the DNA polymerase gene from multiple herpesvirus species. They went on to use these primers to amplify several DNA polymerase amplicon sequences that were previously uncharacterised, widening the understanding of the spread of this herpesvirus in these animals. The primers were also used in consensus PCR experiments on DNA extracted from blood samples of various species, demonstrating amplification and identity of a particular herpesvirus, the presence of which was a novel finding in these particular animals.

Degenerate primers have been widely utilised as gene probes in the study of genes involved in methanotrophy and methylotrophy. For example, in work published by McDonald et al. (1995), standard primers were designed using the conserved regions of several soluble methane monooxygenase (*mmoX*) sequences, in order to detect the presence of methane-oxidising bacteria in natural environments, without the need to perform enrichment and isolation experiments. *mmoX* is not the only gene involved in this pathway however, as there is also a particulate methane monooxygenase (*pmoA*), which is present almost universally in methanotrophs. There is also an ammonia monooxygenase (*amoA*) that is found in ammonia-oxidising nitrifying bacteria. In work published by Holmes et al. (1995), degenerate primers were designed to both these genes as well, targeting shared conserved regions of their active sites. As neither protein had been purified in active form, the fact that their degenerate primers specifically amplified homologous genes from nitrifiers and methanotrophs, and that they were not detected in species unable to oxidise methane or ammonia, further supports the evidence that *pmoA* and *amoA* are components of these proteins.

It was later discovered that *mmoX* and other genes such as *pmoA* were not ubiquitous within all methanotrophs (specifically proteobacteria) (Lau et al. 2013), so

attention was instead turned to *mxaF*, which encodes for the large subunit of methanol dehydrogenase, and is ubiquitous in all but one phylum. In order to use *mxaF* as a phylogenetic and functional marker, it was necessary to design degenerate primers to incorporate the wider diversity of sequences compared to *mmoX*. These primers were shown to amplify *mxaF* from several diverse environments (from soil to hydrothermal vent mussel tissues) and isolates, and is a useful phylogenetic marker at the family level (Lau et al. 2013). Furthermore, these primers enabled the identification of 13 new putative *mxaF* genes from deep-sea bacteria, showing the usefulness and multi-faceted applications of this type of gene probe.

This technique has also been used in publications looking at aspects of the DMSP/DMS cycle **(Chapter 1)**. McDevitt et al. (2002) employed a similar method for the elucidation of genes involved in the DMS dehydrogenase (*ddh*) cluster in *Rhodovulum sulfidophilum*, which catalyses the oxidation of DMS to DMSO, during photoautotrophic growth. Conserved regions of the N-terminus amino acids of two of the subunits that make up Ddh were identified, and several sets of degenerate primers designed which were then used in PCR experiments on chromosomal DNA to discover the full nucleotide and, thus, amino acid sequences of the genes in the *ddh* operon, termed *ddhABDC*. The amplification products of these PCR experiments were sequenced and subsequently used to generate probes for Southern blotting experiments that enabled the purification of those genes, which were then fully sequenced.

Finally, degenerate primers have also been used to study genes involved in the catabolism of DMSP. One example is work carried out by Peng et al (2012), where degenerate primers were designed to conserved regions of an alignment of *dddP* sequences, and used them to investigate the diversity of sequences within mangrove soil environments in Southern China, through culture-independent PCR-based analysis on the community DNA. Up to 144 clones of *dddP* were produced and identified, falling into seven distinct phylogenetic groups, three of which included sequences belonging to previously known Ddd^{+} bacteria, whereas the other 69% were from novel bacteria. This demonstrated a broad diversity of *dddP* within mangrove soils, the distribution of which appeared to be influenced by external pressures such as pH and availability of nitrogen or sulfur. Since this work numerous other *dddP* sequences have been identified, and it appears that the primers designed in this study, although useful at the time, are highly biased towards *dddP* in species of *Roseobacter*. Another gene involved in DMSP catabolism that was studied in this was *dmdA*, which is involved in the other pathway of DMSP catabolism – the demethylation pathway. These degenerate primers were created to be either universal or clade-specific, from metagenomic reads pulled from the Global Ocean Sampling metagenome (Varaljay et al. 2010). They were designed to cover the natural sequence

heterogeneity in *dmdA*, and were used to compared free-living and particle-associated bacterial communities in the coastal waters of Sapelo Island, as well as in qPCR experiments. It was found that they did not necessarily increase the diversity amplified compared to the specific primers, but equally diverse but they certainly captured a slightly different suite of sequences.

3.1.3 Use of quantitative PCR to study gene abundance and transcription

Another technique that degenerate primer design can be utilised in is quantitative PCR. This method is often used to study functional genes in a given environment, and is a recognised and well-utilised technique in research today (Saleh-Lakha et al. 2005). Instead of regular PCR where amplification occurs and only the end product is analysed, qPCR (or Real-Time PCR) works by monitoring the amplification of the target sequence throughout the reaction (real-time). This amplification is detected either through specific DNA probes that are fluorescently-labelled, or through a non-specific fluorescent dye, such as SYBR green, that inserts into any double-stranded DNA (the output of the PCR). The accumulation of fluorescence is measured every cycle at the extension step, and displayed in a curve. The number of cycles passed before the fluorescence passes the threshold point is the quantification cycle, C_q, and from this number and the curve of standards, the copy number in the sample can be calculated and normalised per gram of sample. Furthermore, in two-step qPCR, RNA can first be transcribed into cDNA using reverse transcriptase and random or sequence-specific primers, then quantified in the same way, to analyse gene activity. Another aspect of analysis the melting curve, which is a program run after the qPCR that gradually increases the temperature of the mix until 50% of the double-stranded DNA is denatured. This can be an indicator of whether or not the specific product has been amplified, and is useful in non-specific fluorescence reactions, as the dye will intercalate into any double-stranded DNA. Furthermore, with primers that may have a degree of non-specific binding, which is especially common in degenerate primers, it is possible to restrict the melt curve to only the region in which the curve is expected to appear if the specific gene has been amplified. It is still important to use this in conjunction with other analysis such as gel electrophoresis to determine a single product.

In order to use degenerate primers in qPCR assays, there are several requirements. The size of the fragment amplified should be as small as possible, ideally 75 – 200 bp (or at least below 250 bp), the GC content should be ~50 – 60 %, and the melting temperature between 60 – 65 °C. Once the primers have been designed, they can be optimised by testing different concentrations of primer and a standard template in the PCR mix, and also by optimising the times and temperatures of the stages of amplification (denaturation, annealing and extension), and detecting a single melt curve.

qPCR has been used to study several aspects of the DMSP/DMS cycle in the past (Levine et al. 2012). It is a relatively inexpensive culture-independent method that does not require large-scale sequencing, unlike techniques like metagenomic or metatranscriptomic analysis, although qPCR is often used in conjunction with some of these sequencing experiments (Yergeau et al. 2010). Often this is because coverage of a particular gene in meta-analysis can be weak, as it is governed by the abundance of that gene, which can sometimes be lost under more dominating processes.

In Yergeau et al. (2010) the fate of methane (production and oxidation) in the permafrost versus the overlying active soil was studied using a combination of sequencing, qPCR and microarrays. qPCR was performed on DNA extracted from these environments, on ribosomal genes for identification (16S rRNA) and functional genes including *pmoA* and *amoA*. The 16S rRNA analysis showed bacterial dominance in both the active layer and permafrost, although there were differences in the dominant phyla. A similarity was also observed between the two samples when the metagenomic sequencing was analysed for functional genes compared to other samples. There were some differences between the qPCR experiments and the sequencing; qPCR was able to amplify (at low concentration) some genes that were not detected in the metagenomes, and type II methanotrophs were detected in large numbers by qPCR, despite not being detected at all in the metagenomic libraries. Despite this, the overarching pattern was similar in that the actual quantification of 16S rRNA indicated that type I methanotrophs are the dominant group. Almost all the genes related to the N-cycle were detected in both metagenomic samples, which was also observed in the copy numbers of all the N-cycle related genes tested by qPCR. Many of the differences observed between the two methods of analysis were suggested to be due to the bias caused by MDA (multiple displacement amplification) treatment on metagenomic samples, demonstrating the importance of applying a combination of methods when analysing DNA extracted from the environment, especially at a time when sequencing data is so easily available. qPCR methods keep environmental analysis targeted, and metagenomic analysis enables us to look at conditions and effects beyond what we would expect. Furthermore, where they both show similar results it lends weight to the conclusions drawn. They both corroborate and expand upon each other

As well as measuring the copy numbers of ribosomal and functional genes, qPCR is also used to analyse RNA, either through one-step RT-qPCR (Levine et al. 2012), or by using a cDNA template that has separately undergone reverse transcription (Saleh-Lakha et al. 2005), which is also the method used in this thesis, as it is more flexible and allows for a limited amount of starting material. This form of qPCR elevates the knowledge that a gene is present in a sample, as it informs us on whether or not this gene is transcribed and may be expressed in that environment. If a gene is transcribed in an environment it

provides a far better indication as to whether the process is active or not. The work carried out on samples from the Sargasso sea (Levine et al. 2012) was an in-depth analysis of the abundance and transcription of DMSP degradation genes, as well as studying the DMSP lyase enzyme rates and consumption and production rates over a 10-month period in the Sargasso sea. qPCR was used to study the variability in abundance of *dmdA* (a DMSP demethylation gene) and *dddP* (a DMSP lyase gene). This was to study to interplay between algal DMSP cleavage (by bacterial DddP) and bacterial DMSP demethylation (DmdA). The abundance of both genes was shown to follow similar patterns seasonally to those observed in previous metatranscriptomic and qPCR studies, (Vila-Costa et al. 2010). It seemed that *dddP* and a few *dmdA* subclades were more abundant in winter and spring, while the other *dmdA* clades were of higher abundance in the summer and autumn.

One-step RT-qPCR was used to quantify the transcription levels of the *dmdA* and *dddP* genes through the 10-month period, finding them to be relatively low all year, with the highest transcript numbers occurring in the summer and early fall, despite *dddP* being more abundant in the winter. Many of the increased transcript numbers coincided with time points found to have 'elevated' DMSP consumption rates, although there were also times DMSP levels were elevated when the transcription of *dddP* and *dmsA* was not, suggesting that the other DMSP degradation genes could be playing a role at those particular times. This could be further studied with degenerate primers designed to other *ddd* genes to observe the interplay between the transcription of all of them through the year. They also looked at effect on the transcription of both when subjected to elevated UV-A levels, finding that bacterial DMSP cleavage is tolerant of it, while DMSP demethylation is not. As there are theories that phytoplankton cleave DMSP to DMS as an anti-oxidant response to UV radiation (Sunda et al. 2002), it is possible that DMSP cleavage plays a similar role in bacteria. It was originally hypothesised that the two pathways would vary , with species choosing one or the other, whereas it appeared that both pathways could take place together, although other conditions such as UV-A levels, can still cause a 'switch' between the two (Levine et al. 2012).

Gene probes combined with qPCR gives a strong analytical tool for gene exploration, allowing abundance and transcription to be measured under a myriad of conditions or time-frames. Once degenerate primers are designed to *dsyB*, qPCR will be invaluable to the study of it.

3.1.4 Chapter aims

As detailed in **Chapter 1**, *dsyB* is a diagnostic gene for potential bacterial DMSP production. The aim of this chapter was to design, test and optimise degenerate primers to *dsyB*, using the amino acid sequences of the 24 known DsyB enzymes that are suitable not only for taxonomy based work, but also for qPCR. This was achieved by aligning the sequences, identifying conserved regions and calculating the optimum degeneracy. Multiple options were trialled before selecting the best forward and reverse primers, which were then optimised to get the most specific amplification. PCR on various controls (positive and negative) demonstrated that the chosen primers amplify *dsyB* from multiple different species. These primer sets were deemed suitable for qPCR and were used on a host of environmental DNA/cDNA samples isolated from a range of different marine environments.

3.2 Methods and Results

3.2.1 Primer design

The process of developing degenerate primers for *dsyB* as gene probes is briefly described in materials and methods, and was carried out with help from Dr Jennifer Pratscher. Although it is now known that DsyB sequences are found in over 200 bacterial species, at the time of primer design only 24 DsyB amino acid sequences were known **(Figure 3-1)**. As detailed in the introduction, these sequences were exclusively from marine alphaproteobacteria, of three different classes; Rhodobacterales, Rhodospirillales and Rhizobiales. These were aligned using ClustalW, alongside DSYB sequences (DMSHB synthase enzymes ~ 33% amino acid identity to bacterial DsyB enzymes, see **Chapter 1**) from eukaryotic algae and corals, and non-functional DsyB-like proteins from terrestrial bacteria. The data outputs can be seen in the form of a phylogenetic tree in **(Figure 3-1)** and a multiple sequence alignment in **(Figure 3-2)**.

From the alignment two primer combination options were designed, with one forward primer and two reverse primer options from different conserved regions. These were a combination of specific and degenerate bases, with degenerate bases coding for an equal proportion of a selection of bases. The degenerate bases used in this design are 'S' which results in either G or C, 'K' that gives G or T, and 'R' that gives A or G **(Table 3-1)**.

Table 3-1: The oligonucleotide sequences for the degenerate primers, designed from three conserved regions of the 24 DsyB amino acid sequences.

Primer	Sequence	GC content	Melting temperature (°C)
dsyB_deg1F	CATGGGSTCSAAGGCSCTKTT	57	64
dsyB_deg2R	GCAGRTARTCGCCGAAATCGTA	45	62
dsyB_deg3R	GCCGCCSACRTCSAGCA	71	61

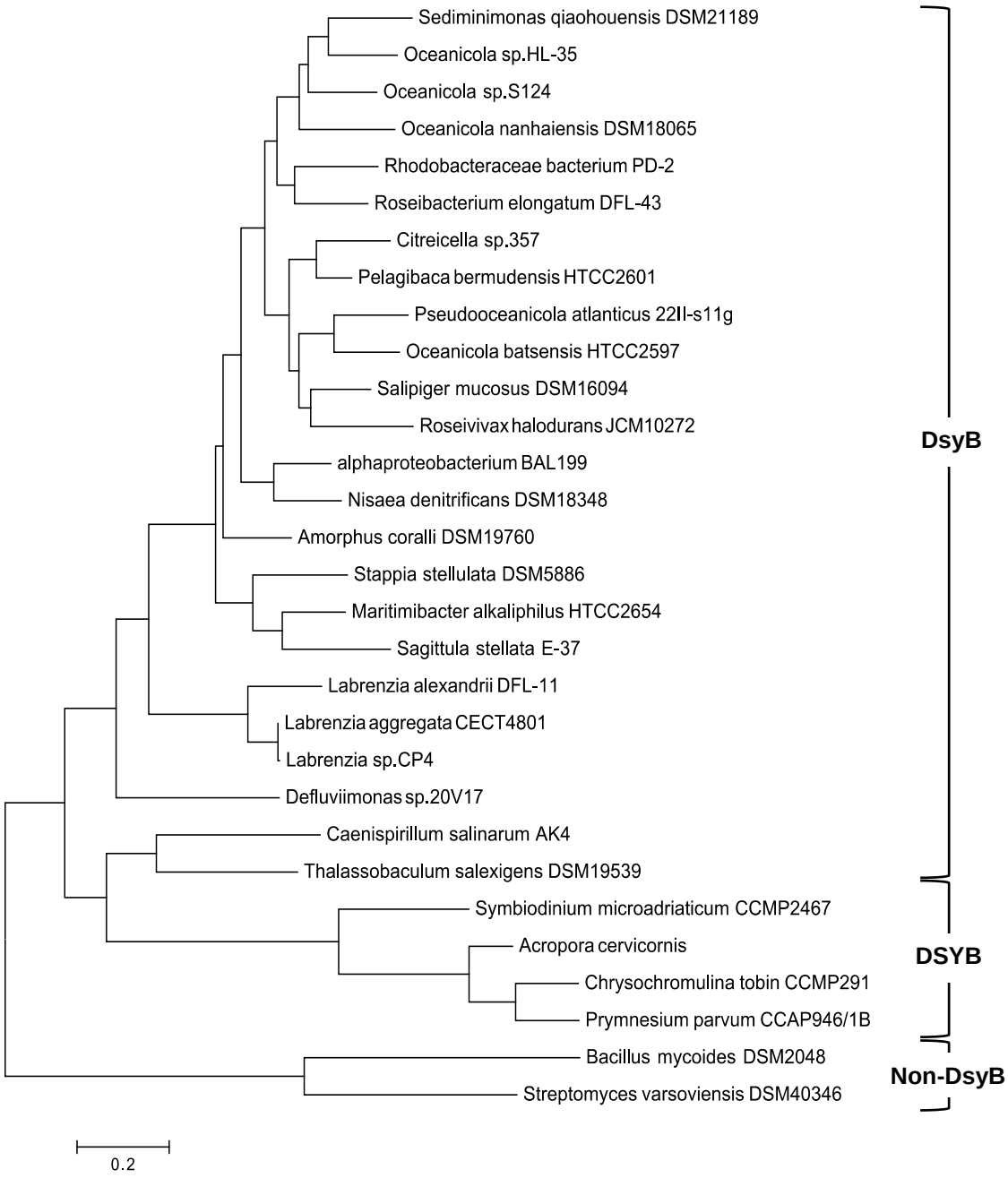

Figure 3-1: Maximum-likelihood phylogenetic tree of 24 of the known DsyB proteins used in the design of degenerate primers, four ratified eukaryotic DSYB sequences and two non-functional 'DsyB' sequences. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site, as indicated on the scale bar.

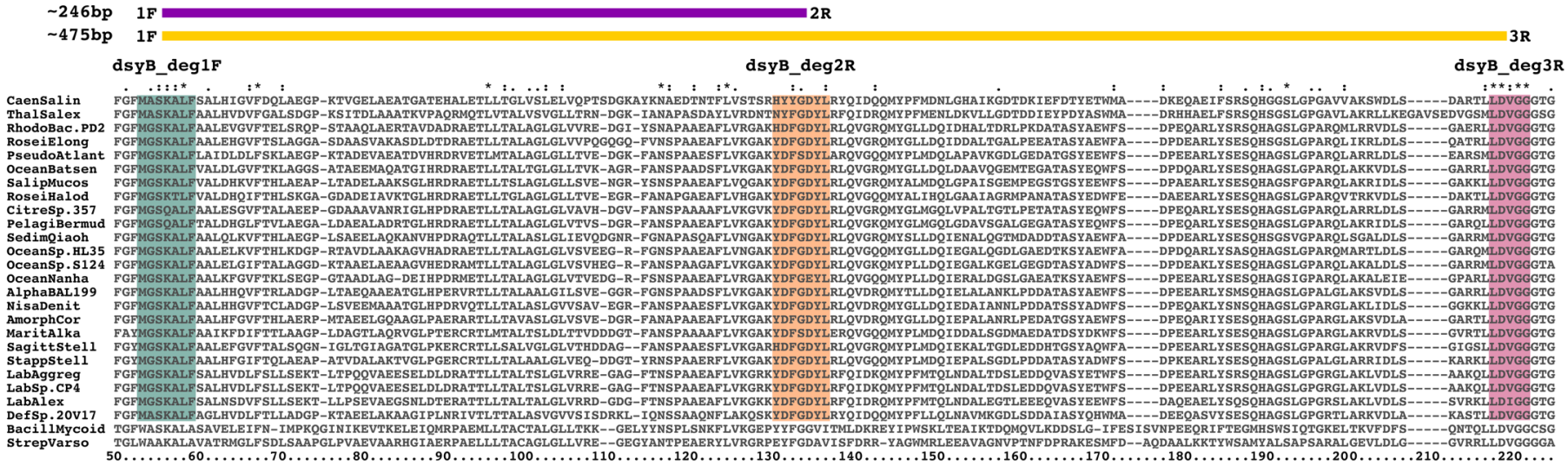

Figure 3-2: A section of the amino acid alignment of the 24 DsyB sequences, aligned using the ClustalW option, including two non-functional 'DsyB' sequences. Conserved regions were identified – fully conserved amino acids between all sequences are marked by an asterix (*), closely similar amino acids are marked by two dots (:), and less closely (but still mostly conserved) amino acids are marked by a single dot (.). The three regions from which the primers were designed, with a maximum of 5 degenerate bases, are indicated by the shaded boxes, with dsyB_deg1F (teal) tested against both dsyB_deg2R (orange) and dsyB_deg3R (pink). The fragments produced are also shown, with the 1F/2R primers producing a ~246bp fragment, and the 1F/3R primers producing a ~475bp fragment.

3.2.2 Primer optimisation

Two primer combination options were designed from the alignments **(Table 3-1)**, and tested under multiple conditions. The first step was to determine which primer pair was most suitable. The two sets were tested for how well they amplified *dsyB* from the genomic DNA of various controls, and how many non-specific bands were also amplified by them.

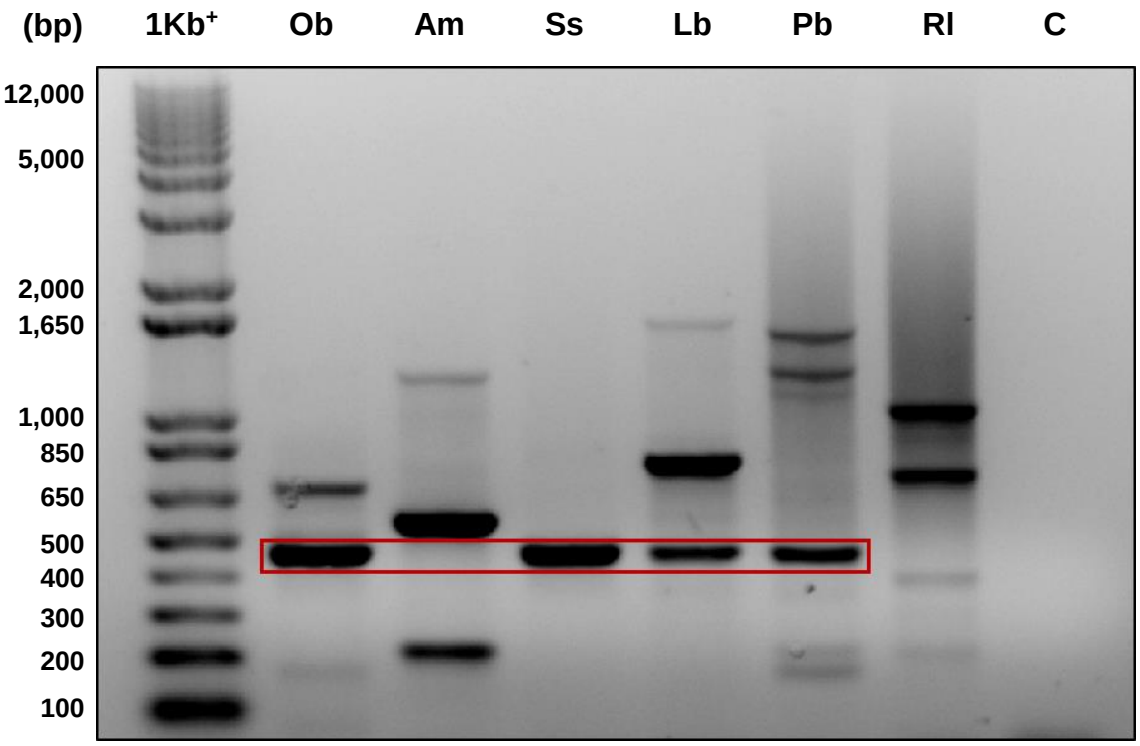

Figure 3-3: Gel electrophoresis of degenerate *dsyB* primer optimisation, showing the PCR amplification of five positive controls, one negative control and a water control (C). Positive controls include *O. batsensis* (Ob), *A. coralli* (Am)*, S. stellata* E-37 (Ss), *L. aggregata* LZB033 (Lb) and *P. bermudensis* HTCC2597 (Pb), and the negative control was *R. leguminosarum* J391 (Rl). This amplification was carried out using the primer set dsyB_deg1F and dsyB_deg3R, amplifying a 475bp fragment (indicated by the red box). Run against a 1Kb Plus ladder

The first primer pair tested (dsyB_deg1F and dsyB_deg3R) was discarded because, despite various condition changes in an attempt to optimise it, there was never any amplification of *dsyB* observed from *A. coralli,* as visible by the different-sized band when samples were run on gel electrophoresis **(Figure 3-3)**. The closest band in size was extracted and sequenced, and found to be a 16S rRNA methyltransferase. The *A. coralli dsyB* was checked using primers designed specifically to the sequence, and confirmed to be present and of the correct size in the genomic DNA that was being used. Although *dsyB* was successfully amplified from the other positive controls tested, this non-amplification of

A. coralli suggested that the primers were not consistently amplifying *dsyB*, which was not acceptable especially since the other pairing of primers were also working and did not have this problem **(Figure 3-4)**. As can be seen in **Figure 3-4**, when using primers dsyB_deg1F and dsyB_deg2R, the correct size product was amplified from all positive samples, but not from the negative control. Furthermore, the smaller size (~246 bp) makes this combination potentially more useful for qPCR amplification. Thus it was decided to continue work with dsyB_deg1F and dsyB_deg2R.

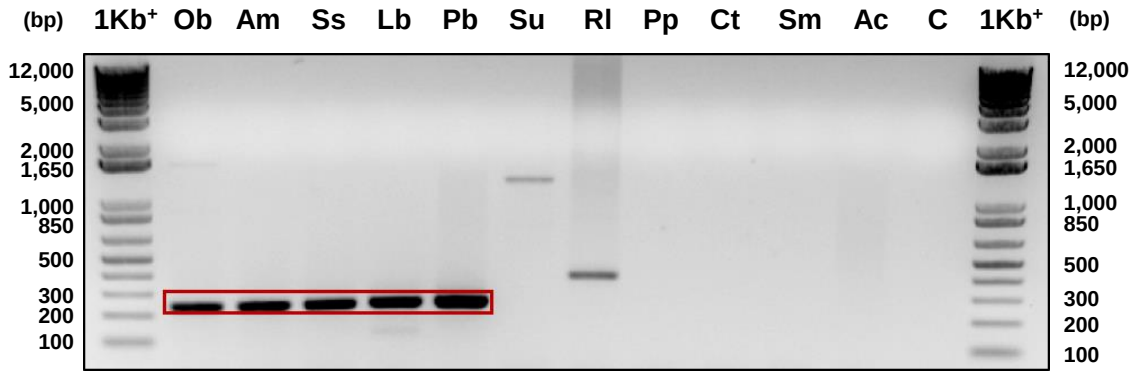

Figure 3-4: Gel electrophoresis of final degenerate *dsyB* primers, showing the optimised PCR amplification of *dsyB* using degenerate primers on five positive controls *O. batsensis* (Ob), *A. coralli* (Am), *S. stellata* E-37 (Ss), *L. aggregata* LZB033 (Lb) and *P. bermudensis* HTCC2597 (Pb), six negative controls *R. leguminosarum* J391 (Rl), *Prymnesium parvum* (Pp), *Chrysochromulina tobin* (Ct), *Symbiodinium microadriaticum* (Sm) and *Acropora cervicornis* (Ac) and a water control (C). This amplification was carried out using the primer set dsyB_deg1F and dsyB_deg2R, amplifying a 245bp fragment (indicated by the red box) in all positive controls. Run against a 1Kb Plus ladder

The next step in optimising the PCR program for the dsyB_deg1F and dsyB_deg2R primers was finding the optimal annealing temperature. This was accomplished using a gradient PCR, testing temperatures between 60°C and 65°C depending on the melting temperature of the primers. The optimal temperature was determined to be 61°C, as it gave the strongest, most specific band. Following this, the extension time and number of cycles were also tested to minimise the non-specific bands that can be amplified. For the 246 bp sized fragment, extension times using the MyFi™ taq was tested between 15 and 60 seconds, and found to be most effective at 15 seconds. Several variations in the number of cycles were also trialled, between 30 and 40, and the optimum was decided to be 35.

Under these conditions, all the tested positive controls amplified a single specific band, while the negative controls showed little or no amplification, even from the negative controls **(Figure 3-4)**. Multiple negative controls were used to test different aspects of the degenerate primers. Four of these controls were the synthesised *DSYB* genes of known eukaryotic DMSP-producers (*P. parvum* CCAP946/6, *S. microadriaticum* CCMP2467, *C. tobin* CCMP291 and *A. cervicornis*). The fact that no correct band was attained with the *dsyB* primers demonstrates that they are specific to the bacterial *dsyB* sequences. Further, no PCR products were detected from the other two negative controls, which were DNA from alphaproteobacterial strains (*R. leguminosarum* J391 and *Sulfitobacter sp.* EE-36) that are unable to produce DMSP, and lack *dsyB*. This demonstrates that the dsyB_deg1F and dsyB_deg3R primers and conditions applied do not amplify non-specific fragments at a similar size as those attained for *dsyB* in positive control strains.

To confirm that these primers indeed amplify *dsyB* and not another gene at the same size, bands were excised and purified using the QIAGEN™ gel extraction kit, then sent to Eurofins Genomics for dideoxy chain termination method sequencing, using the dsyB_deg1F primer. The sequences were checked using BLASTp against the NCBI database, and aligned to the known sequences using the Clutsal Omega website (https://www.ebi.ac.uk/Tools/msa/clustalo/). The sequences were at least 99% identical, confirming that the primers amplify *dsyB* from genomic DNA. Thus, the dsyB_deg1F and dsyB_deg2R primers seem to be suitable for use as *dsyB* gene probes on genomic DNA from pure organisms, enabling us to predict whether an isolate has the genetic potential to synthesise DMSP. These were also deemed as suitable to test on environmental nucleotides (see below).

3.2.3 Utilising the degenerate primers

Now that the degenerate primers were designed, optimised and demonstrated to amplify *dsyB*, they were tested to see if they would also amplify *dsyB* from DNA extracted from an environment **(Figure 3-5)**. This would mean that they would be able to be utilised in qPCR and RT-qPCR experiments, to analyse abundance and transcription easily as well as being used to study *dsyB* diversity. The environment from which the community DNA was extracted for these experiments was Stiffkey salt marsh, specifically sediment sampled from tidal pools in the lower section of the marsh (see **Chapter 4**).

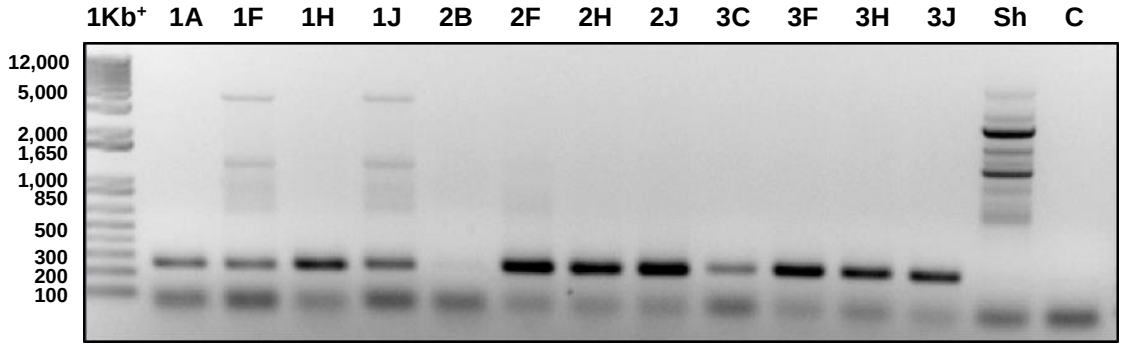

Figure 3-5: Gel electrophoresis showing the PCR amplification of *dsyB* using the primers dsyB_deg1F and dsyB_deg2R on multiple samples of DNA from a salt marsh environment (1A – 3J), a negative control not containing *dsyB* (*S. putrefaciens*), and a water control (C). Run against a 1Kb Plus ladder.

This PCR confirmed that *dsyB* is easily amplified from the salt marsh DNA. Several of the resulting gel fragments were extracted and purified, and then ligated into the pGEM-T Easy plasmid to make clone libraries. This was done to test the diversity of *dsyB* sequences that can be identified in an environment, and to show that multiple sequences of *dsyB* can be amplified by the same primers. The successfully ligated colonies were picked, cultured and the plasmid extracted and sequenced (see **Chapter 2**). The sequences were checked to make sure they were *dsyB* using BLASTp, and out of 19 tested, 17 were confirmed to be *dsyB.* DNA sequences were translated in the correct reading frame, and then added to the ClustalW alignment to be placed in a phylogenetic tree **(Figure 3-6)**, alongside the other 24 sequences that were used to generate the maximum-likelihood tree in **Figure 3-1**.

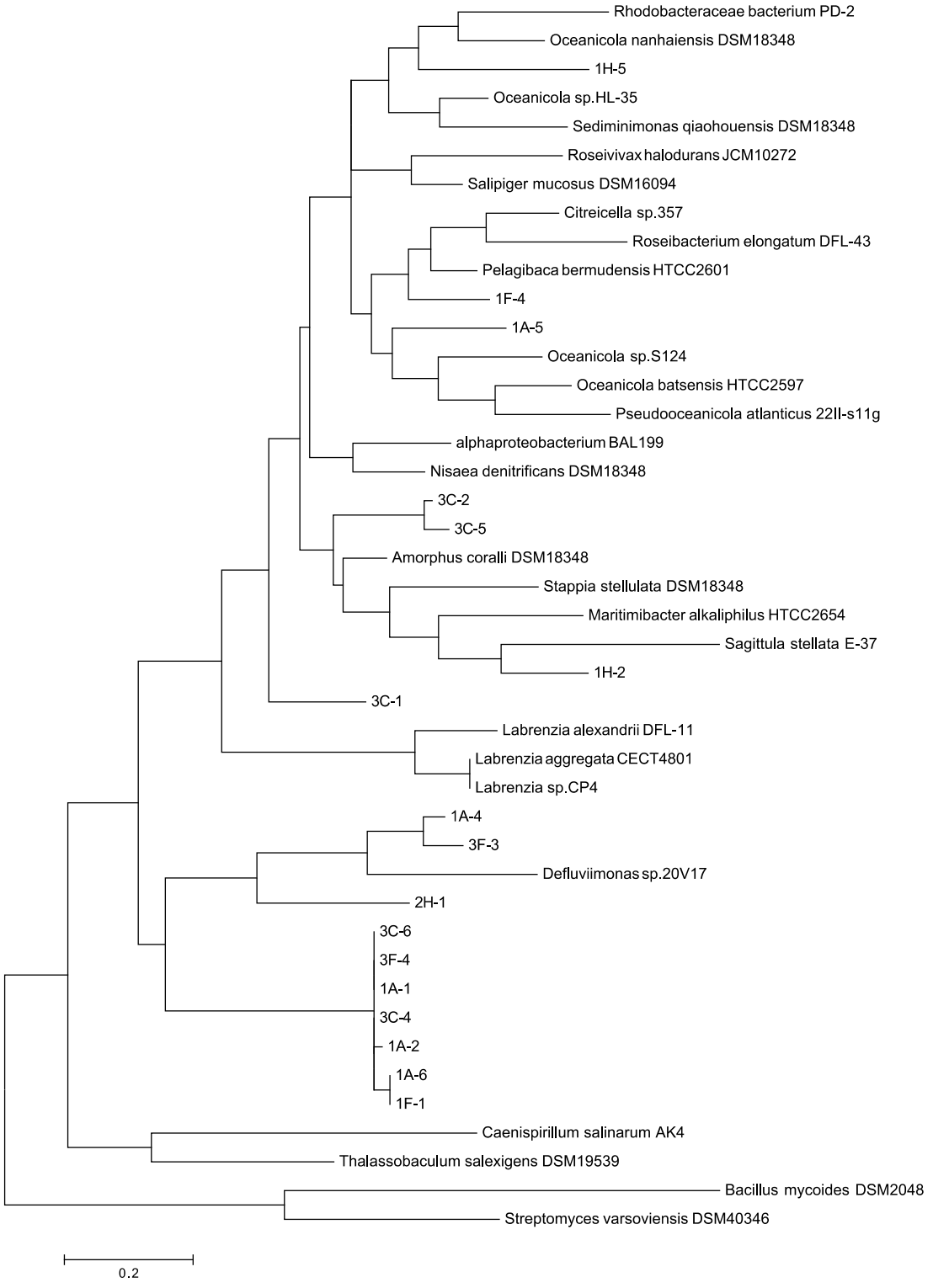

Figure 3-6: Maximum-likelihood phylogenetic tree of the 24 known DsyB proteins used in the design of degenerate primers, including the clone library sequences from the salt marsh environment (1A – 3F). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site, as indicated on the scale bar.

The clone library from the Stiffkey salt marsh environment demonstrated that the dsyB_deg1F and dsyB_deg2R primer set amplify DsyB sequences that span a range of known DsyB sequences. There seems to be a diverse array of DsyB sequences in the salt marsh environment that are easily detected using this combination of degenerate primers. From this information, we can hypothesise that this salt marsh environment contains a variety of bacterial species possessing the ability to synthesise DMSP from Met, and that in this way they contribute to the high DMSP levels known to exist in salt marshes (Steudler & Peterson 1984). Historically it was thought that the high DMSP and DMS production levels in salt marshes was due to the DMSP-producing plant *Spartina* that widely inhabits them (Kocsis et al. 1998). This data suggests that perhaps bacteria may significantly contribute to these levels, given they are likely to always be present (discussed in subsequent Chapters). It is interesting that the *dsyB* primers generate a good proportion of DsyB sequences which cannot easily be classified by organisms whose genomes have been sequenced **(Figure 3-6)**. This provides evidence that the primer set captures a range of DsyB diversity, and that Stiffkey salt marsh may contain a high level of DMSP-producing bacteria with diverse DsyB sequences, and is something that is addressed in subsequent chapters.

The clone library from the Stiffkey salt marsh environment demonstrated that the dsyB_deg1F and dsyB_deg2R primer set amplify DsyB sequences that span a range of known DsyB sequences. There seems to be a diverse array of DsyB sequences in the salt marsh environment that are easily detected using this combination of degenerate primers. From this information, we can hypothesise that this salt marsh environment contains a variety of bacterial species possessing the ability to synthesise DMSP from Met, and that in this way they contribute to the high DMSP levels known to exist in salt marshes (Steudler & Peterson 1984). Historically it was thought that the high DMSP and DMS production levels in salt marshes was due to the DMSP-producing plant *Spartina* that widely inhabits them (Kocsis et al. 1998). These data suggests that perhaps bacteria may significantly contribute to these levels, given they are likely to always be present (discussed in subsequent Chapters). It is interesting that the *dsyB* primers generate a good proportion of DsyB sequences which cannot easily be classified by organisms whose genomes have been sequenced **(Figure 3-6)**. This provides evidence that the primer set captures a range of DsyB diversity, and that Stiffkey salt marsh may contain a high level of DMSP-producing bacteria with diverse DsyB sequences, and is something that is addressed in subsequent chapters.

3.2.4 Optimising degenerate primers for qPCR

As previously mentioned, one of the intended uses for these degenerate primers was in qPCR and RT-qPCR analysis of DNA/RNA extracted from the environment. This

method of analysing functional genes allows for a more in-depth study, as it is not only encompasses the diversity of sequences but it also quantifies the amount of *dsyB* gene/transcripts present in a sample, and is not restricted to single gene analysis, or to those species that are cultivable in the laboratory. As the chosen primer combination amplifies a ~246 bp sized fragment, the primers were already a good size for qPCR, the optimal size being around 200 bp. Although the degenerate primers designed in this chapter do not completely fulfil all the requirements for qPCR primers (see above), they were shown to be functional, and it was more important to maintain the degenerate characteristics. Test runs using a linearised *dsyB* standard cloned into a pGEM-T Easy vector were set up to optimise the program, determine primer concentrations and to check for primer dimers. Although there is a small amount of primer dimerization, it is low enough that it is only observed in the no template control (NTC), and the melt curve is identifiable as non-*dsyB*. The efficiency of this primer pair was also calculated during this test run, and it is 81.86%, which is acceptable for environmental work. The DNA that was used as the template in this *dsyB* qPCR was extracted from Stiffkey, Cley and Yarmouth sediments, which are characterised in **Chapter 4** and were found to contain high concentrations of DMSP. These samples were also used in RT-qPCR analysis, on cDNA that was created from RNA extracted at the same time, which was then reverse transcribed using the specific primer dsyB_deg2R as it was found that random hexamer primers did not produce any detectable *dsyB* amplification (see below). The qPCR reactions were set up inTriplicate samples (biological replicates) were run in triplicate (technical) **(Figure 3-7)**. The conditions for *dsyB* qPCR amplification are described in **Chapter 2**.

Due to the lack of ubiquitous housekeeping gene primer sets that could be used on DNA extracted from the environment, 16S rRNA gene qPCR primers were decided upon as the simplest choice as a rough method of normalisation for the abundance of bacterial species in the environment. There is supposed to be an average of 3.61 copies of this gene per bacterium (Sun et al. 2013), so copy numbers calculated from the qPCR were divided by 3.61 to estimate the cell number in the sample. This then allowed us to estimate the extent of bacteria with the genetic potential to produce DMSP in varied marine samples. The reaction mix was the same as for the *dsyB* qPCR. The conditions for 16S rRNA qPCR amplification are described in **Chapter 2 (Figure 3-7)**.

3.2.5 Using degenerate primers in qPCR

Running qPCR with ribosomal primers as well as with the *dsyB* degenerate primers on the same samples, provides a way to normalise the abundance of *dsyB* in those environments, and calculate a hypothetical percentage of bacteria containing *dsyB*. In samples taken from Stiffkey salt marsh. it was estimated that 0.21% of bacteria possess *dsyB*, in Yarmouth *dsyB* is predicted to be in 0.23% of species, and in Cley it is thought to be 0.1%.The percentage in Stiffkey seems lower than in the predicted percentages calculated from metagenomic work (see **Chapter 6**), which could be due to the *dsyB* degenerate primers not amplifying all possible *dsyB* sequences. All these samples are coastal, and have high salinity, both Cley and Stiffkey being well-known salt marshes, and Yarmouth is an estuary (see **Chapter 4**).

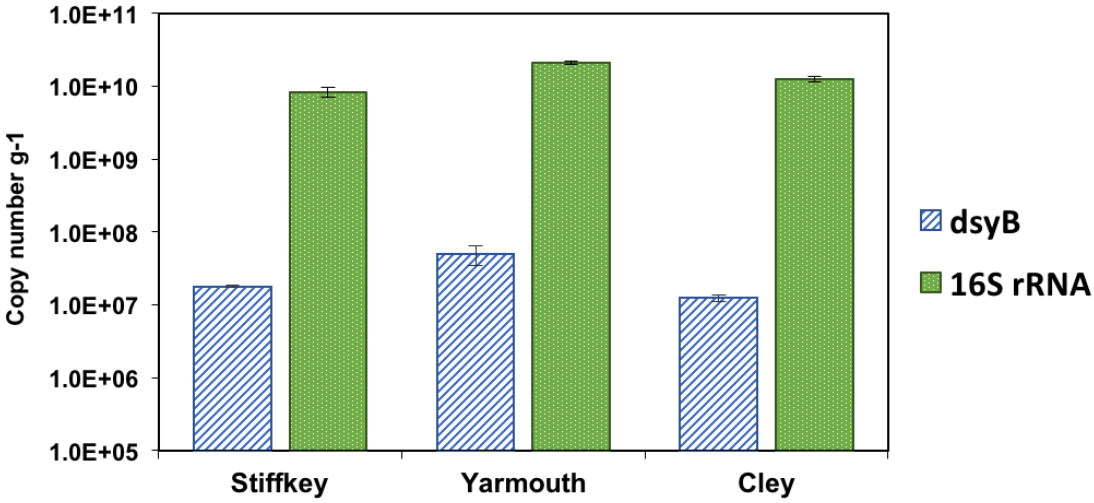

Figure 3-7: A semi-logarithmic plot of the abundance (copies/g) of the functional gene *dsyB* (Blue stripes) and the 16S rRNA gene (green dots) amplified using qPCR from Stiffkey salt marsh, Cley salt marsh and Yarmouth estuary. *dsyB* was amplified using the degenerate primers dsyB_deg1F and dsyB_deg2R, and the 16S rRNA gene amplified using ribosomal primers Eub_338F and Eub_518R. Samples are the average of triplicate data with error bars indicating the standard error of the means.

Despite seeming low in value (< 0.23), it should be kept in mind that the number of bacteria in the sediments are extremely high, being at least 10^9 in number. Thus even at these relatively low predicted percentages, this equates to a huge number of bacteria potentially producing DMSP. In fact it is far more than is present in tested seawater samples (see **Chapter 6**).

After performing qPCR experiments on *dsyB* standards and on DNA extracted from the environment, the next step was to perform reverse transcription on purified RNA to

produce cDNA. After several experiments using a random hexamer primer in the reverse transcription, it seemed that perhaps *dsyB* is too low in abundance to be easily amplified in this manner. Instead, specific primer reverse transcription was utilised, using the dsyB_deg2R primer. RT was performed on purified RNA quantified by Qubit. RNA quality was assumed. Due to the lack of viable housekeeping gene primers, it was difficult to provide much standardisation between samples. Using 16S primers was not appropriate, as samples were not treated to remove rRNA and were therefore dominated by 16S rRNA, masking other genes and therefore not useful for comparison to the *dsyB*-RT samples. Instead all RT reactions were performed on as close to 100 ng RNA as possible from each sample, and then normalised per gram of sediment.

This method proved more successful, and although it limits the claims that can be made about *dsyB* transcription in the natural environment, we were able to demonstrate that *dsyB* RNA is present in these samples **(Figure 3-8)** and not in controls. Therefore, it is possible to state that bacterial DMSP production through the transamination pathway is active in these environments. Although we cannot compare to other gene transcript levels in the tested samples, we can at least claim that *dsyB* transcription might be higher in one sample rather than another.

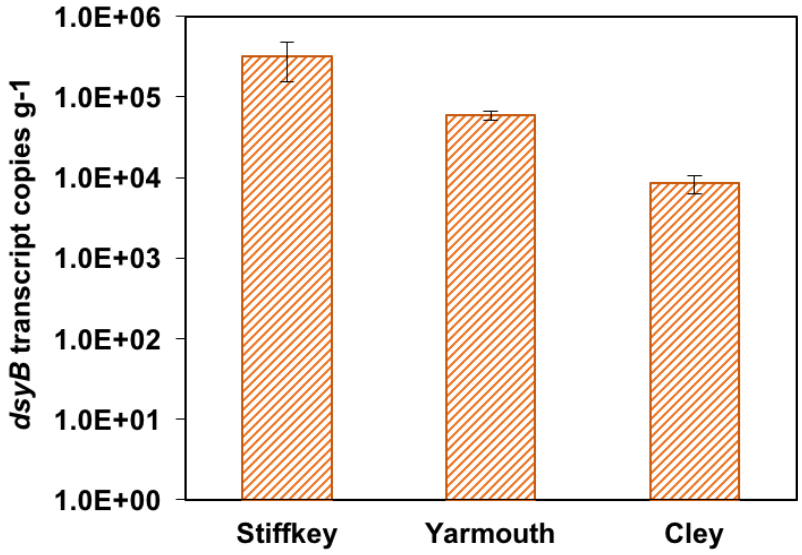

Figure 3-8: A semi-logarithmic plot of the transcription levels (transcript copies/g) of the functional gene *dsyB* RT-qPCR, using specific primer cDNA from Stiffkey salt marsh, Cley salt marsh and Yarmouth estuary. Samples are the average of triplicate data with error bars indicating the standard error of the means.

All three marine sediments show that while *dsyB* appears to be present at fairly similar levels of abundance between them **(Figure 3-7)**, the transcription of this gene at those sites has more variation **(Figure 3-8)**. DsyB transcripts seem most abundant in Stiffkey salt marsh surface sediment, and possible reasons for these differences will be explored in other chapters in this study. Clone libraries were made eith the amplicons and are currently being sequenced. Unfortunately I do not currently have the data to include in this thesis, but it will be important because it will inform us as to the active bacteria likely producing DMSP in these particular environments.

3.3 Discussion

3.3.1 Summary of work

This chapter set out to design and test degenerate primers as an analytical tool for the study of *dsyB* diversity, abundance and transcription in marine sediment environments. Known sequences of the gene were aligned and conserved regions identified. Several primers were designed and tested against both positive and negative genomic DNA controls, and the most specific primer pair chosen.

3.3.2 The primer design process

It was surprising that the first set of primers tested in **Figure 3-3** were unable to amplify *dsyB* from *A. coralli* genomic DNA, as the *A. coralli dsyB*-specific primers amplified it correctly, at the right size. Even looking at the alignments in **Figure 3-2**, in theory the primers should have amplified. It could be because dsyB_deg3R was a slightly smaller primer than dsyB_deg1F and dsyB_deg2R, and therefore was more likely to bind elsewhere in the genome. For whatever reason, this issue was only found with the dsyB_deg1F and dsyB_deg3R combination, as the other set successfully amplified a 246bp-sized fragment from all positive strains tested.

The chosen primers were then tested on DNA extracted from the environment and found to successfully amplify a specific band at the correct size. To confirm that the amplified band was *dsyB*, clone libraries were constructed and several of them sequenced. The majority of them were indeed *dsyB*, demonstrating that the primers are able to amplify from both pure genomic DNA and mixed DNA from an environment. It was important that these primers were able to do both, as a large reason for their design was to use to test isolated bacteria for the presence of *dsyB*, and to study its abundance in various environments.

From these experiments, we can be almost certain that the presence of a band at ~246bp after PCR with dsyB_deg1F and dsyB_deg2R denotes the presence of *dsyB* in a species or environment, showing that these degenerate primers are useful tools in predicting for the ability to produce DMSP. However, it does not necessarily follow that the absence of a band means the absence of the gene – with *dsyB* sequences continually being discovered in new species, even some gammaproteobacteria, it is entirely possible that more divergent *dsyB* sequences exist that still produce a viable enzyme, whilst not being amplified by the same primers. This would be potentially rife with horizontal gene transfer, which we know is likely with *dsyB* (Curson et al. 2017). Therefore, whilst these primers are effective, they are not definitive, and it is important to take this into account when making claims about *dsyB* abundance, as it is definitely underestimated when using

only these primers. This is another reason for combining degenerate primer analysis with other techniques, such as metagenomics and metatranscriptomic sequencing (Yergeau et al. 2010); to ensure that any conclusions drawn aren't affected by lack of coverage. It is also likely that, with any degenerate primers, there is a degree of primer bias towards particular species or sequences (Jin & Mattes 2011), due to PCR conditions, or even because of primer mismatches. This can be remedied somewhat by using more degeneracy when designing the primers, although this can lead to a lack of specificity. Work in the future should look into any potential bias of the *dysB* primers by mixing different standard DNA samples at known concentrations and then observing the relative proportions in the clone libraries generated from the PCR product.

3.3.3 Problems with qPCR

One issue that was encountered when using qPCR on multiple genomes extracted from the environment was the difficulty in finding a housekeeping gene. Although we used the 16S rRNA gene for inferring bacterial abundance in sediment, and by extension the percentage of bacteria containing *dsyB*, there are several problems associated with this. Firstly, only one primer set was utilised for all qPCR amplification. All primers have some degree of bias, meaning that there are undoubtedly species that were not picked up, and others that were over-represented. If time had not been a constraint, it may have been beneficial to experiment more with different primer combinations, or even use a suite of 16S rRNA primers that are designed to different domains, phyla or classes (Yergeau et al. 2010). Furthermore, although the 16S rRNA gene is ubiquitous among bacteria, and is excellent for phylogenetic analysis, there is also no guarantee that all species only contain one copy – indeed, this is known to not always be the case (Rainey et al. 1996). This means that copy numbers could be grossly overestimated, and therefore *dsyB* percentages underestimated compared to the actual number. It is for this reason that other methods of study, including metagenomics, have also been undertaken (see **Chapter 7**). Finally, as previously discussed, 16S rRNA primers were not viable for normalisation of specific primer RT, making it difficult to draw more significant conclusions from the RT-qPCR. In future work, perhaps genes such as *recA* should be looked into as a possible substitution, as it has been used in the past (Giloteaux et al. 2013).

Despite this, much can still be drawn from qPCR and RT-qPCR using *dsyB* degenerate primers and 16S rRNA primers. Even between just three highly saline sediment environments there is much to be observed. Stiffkey, Yarmouth and Cley appear to have similar 16S copy numbers, but the percentage of *dsyB* sequences varies quite dramatically **(Figure 3-7)**. This is mostly echoed in the differences between *dsyB* transcript numbers **(Figure 3-8)**, showing a decrease in *dsyB* transcription in Yarmouth compared to Stiffkey, and an even greater one in Cley. One reason for this difference could be the

type of sediment, and also the site within the salt marsh that they were sampled. Samples were taken from Stiffkey salt marsh from the middle of tidal pools close to the sea, whereas the samples from Cley, whilst still taken from the salt marsh, were taken from a higher site that although still saline, is not as influenced by tides, and can be either much higher salinity or much lower salinity depending on rainfall. This could have an effect on both the presence of *dsyB,* and its activity. Yarmouth appeared to have similar levels of *dsyB* abundance, but significantly lower levels of *dsyB* transcripts. The different conditions between Stiffkey and Yarmouth could explain much about the regulation of *dsyB*, and perhaps even the role of DMSP in bacteria. However, only pH, salinity and temperature were analysed at each of these sites (see **Chapter 4**). Further study and more in-depth environmental measurement is required to properly explore this topic.

3.3.4 Concluding Statements

Degenerate primers as gene probes are an extremely useful molecular tool for the study of functional genes. They enable us to broaden our understanding, moving from looking at species-specific genes in an environment to looking at the entire spectrum of the same gene, as well as giving a wider scope to the search for uncharacterised sequences or novel species containing those genes. They are also easily utilised in both qPCR and sequencing experiments, including customised diversity assay amplicon sequencing using those primers.

Designing degenerate primers to *dsyB* has provided many opportunities to study various aspects of its abundance, distribution and activity in an environment, which has made them a vital tool in answering our questions. These primer sets were extensively used in subsequent chapters to determine if DMSP-producing bacterial isolates contain *dsyB* or not, and to study the diversity of DsyB in marine sediments using amplicon sequencing.

Potentially the most important finding of this chapter is that bacteria with the genetic potential to synthesise DMSP are abundant in tested marine sediment. Furthermore, given that we can detect *dsyB* transcription in these samples, it leads us to hypothesise that bacteria may be important producers of DMSP in these sediment environments, which we show in subsequent chapters to be environments of high DMSP standing stocks.

CHAPTER 4

USING CULTURE-DEPENDENT TECHNIQUES TO IDENTIFY DMSP-PRODUCING SPECIES

4 CULTURE-DEPENDENT IDENTIFICATION

4.1 Introduction

4.1.1 An introduction to Stiffkey salt marsh

Studying the importance of heterotrophic bacterial DMSP production in marine sediments allows us to draw conclusions about the contribution of bacterial DMSP production to the wider environment, and observe gene abundance and activity *in situ*, rather than only in artificial conditions in the lab. To that end, it was decided that bacterial DMSP production should be studied in a salt marsh environment. This would be the first study of its kind, and would enable the development of techniques that could be applied to other environments.

Stiffkey salt marsh on the North Norfolk coast **(Figure 4-2)** was chosen as the site of study for several reasons. Firstly, salt marsh sediments are highly saline and sulfurous, and have DMSP levels several orders of magnitude higher than the overlying seawater (see below). Salt marshes have long been demonstrated to be important sites of DMSP production (Steudler & Peterson 1984), although until now this has been almost entirely attributed to the presence of *Spartina* grass (Dacey et al. 1987), a known DMSP producer (Kocsis et al. 1998). Furthermore, marine sediments cover up to 70 % of the Earth's surface. The tidal pools in the lower marsh of Stiffkey (Davy & Smith 1988) are accessible and easily sampled and we will use them as a representive for marine sediment which is a major component of the Earth's surface.

This sampling site is under an hours drive from the laboratory at UEA, Norwich **(Figure 4-2)** meaning that samples were therefore only a few hours old when they were processed. They were therefore fresh and unlikely to have been affected by removal to the lab, allowing conclusions drawn to also be applied to sediment *in situ*. It was also easy to sample, as the tidal pools were accessible on foot, and easily identifiable by the *Spartina* grass growing around them. This meant that the same pools could be sampled each time, giving confidence to the reproducibilty of the results. The sediment itself was also easily identifiable, as the boundary between microoxic and anoxic sediment was distinguished by a dramatic colour change **(Figure 4-1)**, meaning that it was easy to ensure that there was no contaimination between the two when being processed.

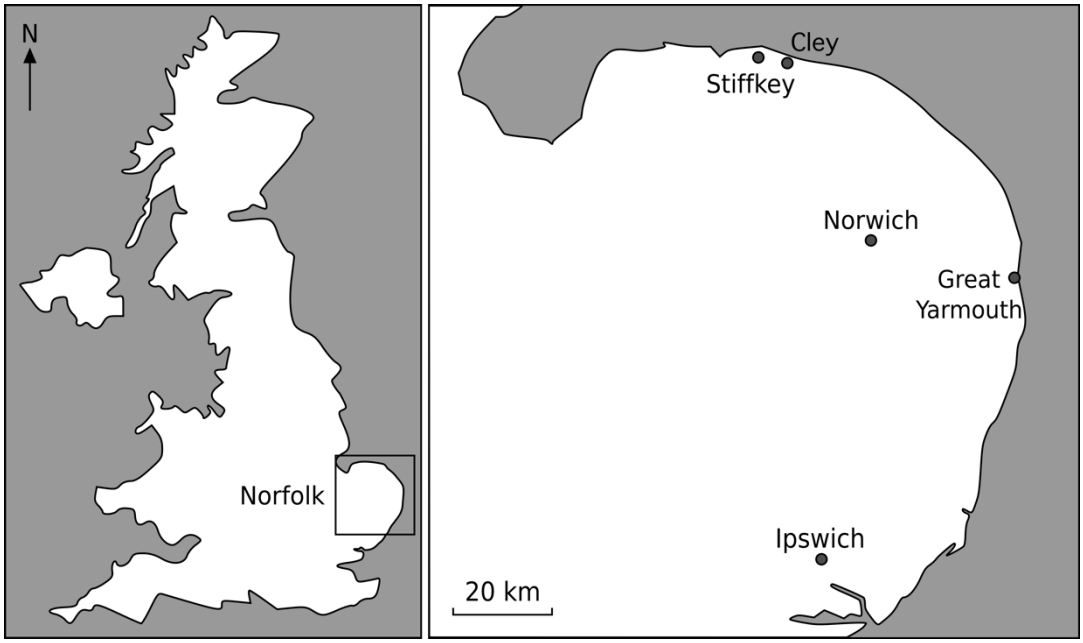

Figure 4-2: The location of Stiffkey salt marsh (latitude 52.964947, longitude 0.925655) in relation to the UK, Norwich and other salt marshes in Norfolk including Cley salt marsh and Yarmouth estuary.

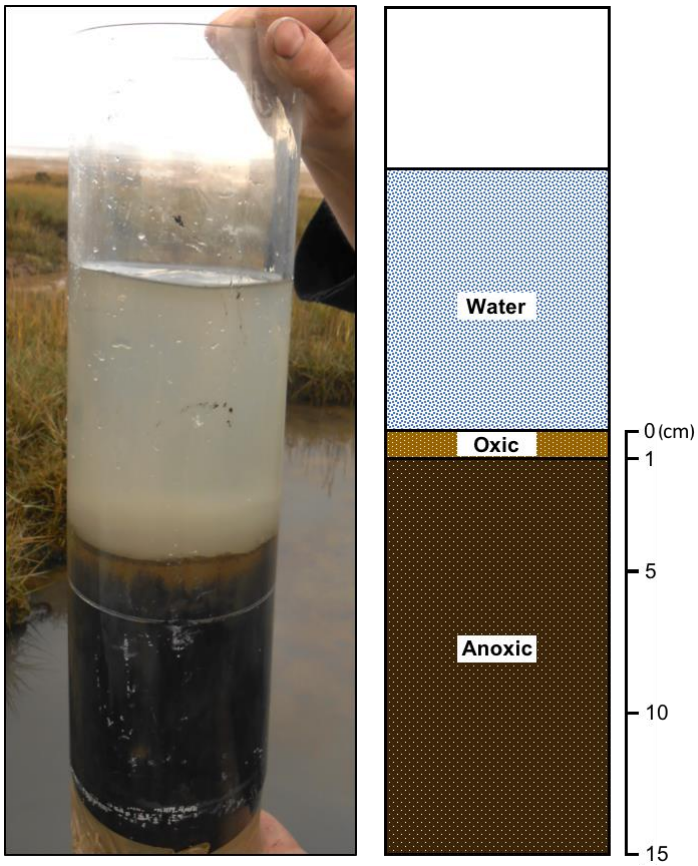

Figure 4-1: A typical acrylic core sample from a Stiffkey tidal pool. The schematic represents the layers seen in the picture, with saline pool water, a top 1 cm layer of oxic sediment, light brown in colour, and the rest a dark brown anoxic layer

4.1.2 Culture-dependent study of an environment

In order to study the bacterial contribution to DMSP production within the Stiffkey salt marsh environment, a suite of techniques had to be decided upon. The first set of experiments carried out were culture-dependent methods including plate isolations and characterisation.

Culture-dependent techniques are some of the oldest, and often simplest, methods used to study the microbes within a community and how they function in an environment. Indeed, it is sometimes so easy to culture some species that they even grow where they are not wanted! Often the isolation of model organisms is the most cost effective way to build up a picture of key microbial players in an environment, and the on-going processes. Prior to this study, nobody had ever attempted to isolate DMSP-producing bacteria, so there was no indication as to how difficult this might be. However, it was decided that isolation work would be an important aspect of this study in order to gain an idea of the microbes that had this capacity. Thus, we began our environmental study by seeing what could be cultured from Stiffkey, and then asking which of these model bacteria might produce DMSP.

The most significant disadvantage to culture-dependent work is the fact that only around 1% of bacteria in an environment are predicted to be cultivable under laboratory conditions (Davis et al. 2005; Saleh-Lakha et al. 2005), meaning that the majority of bacteria are missed. However, often the more prevalent bacteria are the ones most able to adapt and grow, meaning they are usually readily isolated. This indicates that there is still merit to using this type of study, and indeed, there are several other advantages to culturing bacteria from an environment over using only culture-independent methods such as metagenomics, metatranscriptomics and qPCR. From our viewpoint culture-dependent work was especially important because DMSP production by bacteria was novel, and very little was known of the bacteria able to do this, other than what was inferred by genomic predictions based on the occurrence of *dsyB*. Despite the clear importance of culture dependent work, it should also be noted that it is not always the case that the most abundant and/or important bacteria are the ones you can culture; for example, no one would question the importance of SAR11 bacteria in marine biogeochemical cycling (Malmstrom et al. 2004), yet these bacteria are notoriously hard to culture. For this reason it is vital that any culture dependent work is complemented by suitable culture-independent experimentation.

The degenerate primer clone library work in **Chapter 3** demonstrated that bacteria with the genetic potential to produce DMSP (*dsyB*$^+$ bacteria) were present in the three tested marine sediment environments, and that the *dsyB* gene was expressed. Thus far,

this is only predictive, and is based on the presence of a single gene, meaning that the analysis is biased and does not account for bacteria that might be producing DMSP via other pathways, if such bacteria exist. In the same way, metagenomics/metatranscriptomics and 16S rRNA amplicon sequencing, despite being less skewed and much more representative of the natural environment, can only provide information about DMSP production based on what is already known about the genes and species involved. There are likely other, more complex methods that can enable the discovery of novel DMSP-producing species, such as single-cell sorting and sequencing which would enable us to amplify genomes that contain known DMSP-producing genes, or more in-depth analysis of metagenomes under different growth conditions. However, ultimately, the simplest and most often utilised is plate culturing (Steven et al. 2007; Carrión et al. 2017). This method is able to easily isolate bacteria that can be purified and tested for DMSP production by GC and/or LC-MS, and is the only method that results in pure individual cultures of a DMSP-producing species, regardless of what genes it may or may not possess. Although culture-dependent work is biased in as much as you only isolate microbes that will grow under lab conditions, it does not predispose or exclude bacteria based on the presence or absence of *dsyB*.

It is also possible to be reasonably selective when performing plate isolations. In some cases the composition of the agar can be changed to semi-solid to increase isolation of microaerobic species, or plates can even be incubated in fully anaerobic conditions. The salinity or nutrient concentrations can be altered, and any number of additional substances can be added in order to push the isolation to favourably grow particular bacterial species. All of these alterations can encourage a wider range of bacterial species to be isolated.

Another strength of culture-dependent isolations of bacterial species is that by identifying individual species from a site, it confirms the presence of that species in that environment, beyond the level of estimation, especially if the same or similar species are isolated multiple times. These isolated species, if easy to maintain and manipulate, can sometimes become model organisms for the study of a particular process in that environment. This carries more weight than work done on basic models such as *E. coli* in some ways because the species is linked directly to that particular environment, and even though the conditions of growth in the lab are not necessarily comparable to those in the natural environment, it still gives a more realistic picture. Furthermore, by isolating specific bacterial species and sequencing them, it enables a greater level of confidence when declaring the presence or absence of genes like *dsyB*, rather than relying on phylogeny or the closest sequence match.

4.1.3 Culture-dependent methods in literature

Most microbiological research on an environment has involved the use of culture-dependent techniques at some point in the study, often in conjunction with culture-independent methods in order to give a more complete view of the environment. One such study was carried out by Carrión et al (2017), on methanethiol-dependent DMS production in terrestrial environments. This work was following on from the discovery of a novel DMS-producing pathway, the methylation of MeSH to release DMS (Mdd), and the first gene associated with it (Carrión et al. 2015). Having characterised this gene, it could then be used as a reporter for the Mdd process in different environments, including terrestrial soil and marine sediments. Carrión et al. (2017) combines both culture-dependent and –independent work in the study of this process very effectively to identify microbes involved in this process.

Rates of MeSH consumption and DMS production were measured in samples from a variety of environments, and grassland soil was incubated with MeSH in order to enrich for MeSH-methylating species that produce DMS, as well as DMS added to enrich for DMS-consuming bacteria. Species that were isolated were characterised for their 16S rRNA identity, and for the ability to methylate MeSH and/or consume DMS. Alongside these isolations, the community was analysed using 16S rRNA amplicon sequencing of the T=0 and 14-day enrichments. It was found that the species isolated from T=0 samples were consistent with the most abundant classes in the 16S rRNA sequencing, and this was the same at genus level, with *Pseudomonas*, *Streptomyces* and *Bacillus* being present in both. This pattern was also observed with the enriched samples, even at genus level with *Ensifer, Pseudomonas* and *Acinetobacter* being the most abundant. The isolates were tested for DMS production when supplemented with Met (a MeSH precursor) or MeSH, and in T=0 samples ~58% of isolates could generate DMS, demonstrating a diverse natural ability in the soil. However, this number increased to ~96% when isolates from the enriched samples were tested, demonstrating the effectiveness of enrichment cultivation experiments. In this case, both culture-dependent and culture-independent methods were used as confirmation against each other – they both showed similar patterns of enrichment, and therefore validate the findings of both. It would certainly be worthwhile to have such experimentation done, focussed on the process of bacterial DMSP-production. This is exactly what was attempted in this chapter.

There are many other studies that also utilise this mix of culturing and high-throughput sequencing. Steven et al. (2007) is one study that uses the same techniques, but with a different application in order to study the diversity of the microbial community in the Arctic permafrost. In this study, the two techniques were used to cover different aspects of the same site, rather than as confirmation for each other. The culture-independent work

involved using community DNA in the creation of 16S rRNA gene libraries for bacteria and archaea, which were sequenced and analysed for phylogeny. The isolation work was used both to identify previously uncultured organisms from the permafrost samples, and to characterise the abilities of those isolates, including their halotolerance and psychrotolerance. This enabled Steven et al to draw more in-depth conclusions about the salt and temperature tolerance of microbes in that community, as it is the physical demonstration of conclusions drawn from the sequencing.

4.1.4 Culturing bacterial species from Stiffkey salt marsh

An important aspect of culture-dependent work is, naturally, being able to access the sediment from which to culture bacteria. As explained above, Stiffkey is easily accessible and easily sampled, and the oxic sediment is immediately identifiable from its light brown colour, compared to the dark brown of the anoxic layer **(Figure 4-1)**. Although both the oxic and anoxic layers have been shown to produce DMSP (see below), this study only used sediment from the microoxic layer, in order to have as much of a focussed and in-depth approach as was possible in the time and with the funding available. This was also because the microoxic layer has by far the highest standing stock concentration of DMSP (compared to the pool water and the anoxic sediment) (see below).

It is not a complex process to cultivate at least some species of bacteria from this type of sediment, as it is known to be bacteria-rich, with some estimates placing the number of bacterial cells in intertidal sediments between 2×10^8 and 3.5×10^9 per gram (Kuwae & Hosokawa 1999). This is compared to estimates of bacteria in the water column, where even in surface waters bacteria only reach numbers up to 1×10^6 per ml (Hobbie et al. 1977). This knowledge, combined with the work in the previous chapters that demonstrate that Stiffkey has high levels of DMSP, and also contains several versions of the DMSP-producing gene, *dsyB*, meant that we were confident that bacterial DMSP-producing strains would be easily isolated from the sediment, although this still needed to be tested experimentally.

It is common practice, especially in the Todd lab, to maximise the chance of isolating bacteria of interest by carrying out enrichment culturing techniques, such as selecting for bacteria able to degrade DMSP by including DMSP in the media as sole carbon source. This is relatively easy for bacteria that use a substrate as a carbon source, but for DMSP-producing bacteria such substrate enrichments were not appropriate. It was not obvious how one could 'enrich' for bacteria producing DMSP because, unlike DMSP catabolism, there is no single molecule that can be added to increase the activity of those species. We would be attempting to enrich for a *process*, rather than for the use of a molecule, and any substrate added is likely to be used as a carbon source, in addition to

DMSP production. The process of designing an ‘enrichment’ method for this is described below, based on work by (Curson et al. 2017) that showed a variety of growth conditions including external conditions like temperature, and internal changes to the media composition, all of which were found to alter the production of DMSP by *Labrenzia* **(Figure 4-3)**. In the same way, it was proposed that changing some of these aforementioned conditions to create an ‘enrichment media’ to incubate sediment in for a period of time could push the bacteria to increase DMSP production, possibly conferring a survival advantage and eventually skewing the sediment community towards an abundance of bacterial DMSP-producers.

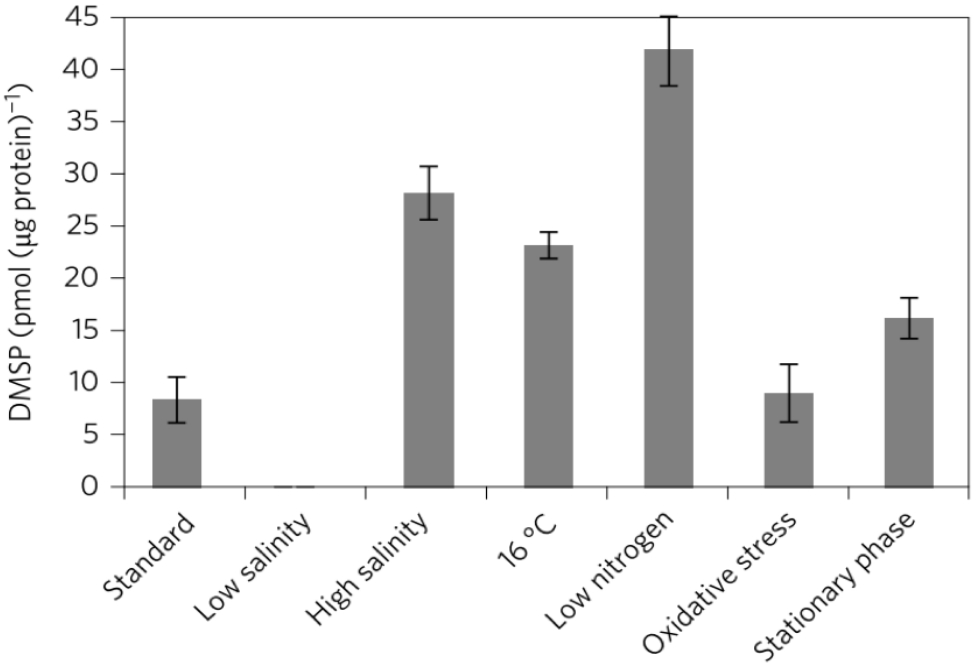

Figure 4-3: DMSP production in *L. aggregata* LZB033 in MBM media with different conditions. Varying growth conditions include salinity, temperature, nitrogen concentration, oxidative stress and growth phase. (Curson et al, 2017).

4.1.5 Chapter Aims

From the degenerate primer PCR amplification and subsequent clone library analysis of sediment from Stiffkey salt marsh, we can be confident that DMSP-producing bacteria are present in this environment, and RT-qPCR showed that *dsyB* is actively transcribed, confirming that bacterial DMSP production takes place in Stiffkey. We hope to show that Stiffkey sediments have higher DMSP standing stocks and production rates compared to the overlying seawater, which would suggest that these are highly productive areas for DMSP synthesis, perhaps by bacteria.

The aim of this chapter was to ascertain whether DMSP-producing bacterial species could be easily isolated from Stiffkey salt marsh, to design enrichment

experiments that maximise DMSP production in the sediment to increase isolation of DMSP-producing species, and to characterise those species, confirming the presence or absence of *dsyB* within them. This would not only demonstrate that bacterial DMSP production likely takes place in the salt marsh environment, but could also result in the identification of novel DMSP-synthesising bacteria, and/or even novel DMSP-producing genes/pathways.

4.2 Methods and Results

4.2.1 Preliminary sediment sampling

The first set of samples taken from Stiffkey were part of a preliminary excursion which also went to Cley salt marsh and Yarmouth estuary, to determine if this type of sediment was appropriate for bacterial DMSP-analysis. Samples were collected on 29-07-2016, alongside measurements of several parameters of the pool water including temperature (measured on site), as well pH and salinity (measured in the laboratory using an electronic pH meter and a handheld analogue refractometer) **(Table 4-1)**. The sediment was sampled using bespoke acrylic corers that were driven into the centre of tidal pools (usually 1.5 – 2 m in size) in the lower section of the marsh, to a depth of ~ 15 cm **(Figure 4-1).**

Samples were carefully transported to the laboratory, ensuring that the layers were not mixed, and after taking samples and pH/salinity measurements from the water layer, this was drained off. The entire surface oxic sediment (the top 1 cm) was removed completely for DMSP quantification, as well as for use in initial DMSP experiments such as incubations and culturing (see below). The core was then split down the middle, and the anoxic sediment from the centre of the core was sampled at distances of 5, 10 and 15 cm in from the surface sediment. For this study the anoxic sediment was only used for DMSP quantification measurements **(Table 4-1)**. Cley and Yarmouth were sampled less comprehensively, with only the oxic layer sediment and pool water taken. Aliquots of 0.5 g of the oxic layer sediment from all three locations were measured into 2 ml screw-cap tubes and flash frozen in liquid nitrogen and stored at -80°C for community DNA/RNA extractions at a later point.

The DMSP quantification for all the sediment sampled in this excursion was performed on 0.1 g sediment weighed into 1.5 ml GC vials, before being mixed with 100 µl sterile water and 100 µl 10 M NaOH, crimp-sealed and vortexed for 5 – 10 seconds. Pool water measurements were performed on 200 µl water, mixed with 100 µl 10 M NaOH. All samples were in triplicate, and were incubated overnight in the dark, before the headspace was measured using an Agilent 7890A GC, fitted with a 7693 autosampler.

Table 4-1: The characteristics of Stiffkey, Cley and Yarmouth sediment.

Sampling site (29/07/16)	Location (Lat, Long)	Depth (cm)	nmol DMSP/g or /ml	Salinity (PSU)	pH	Temp. (°C)
Stiffkey sediment	52.964947, 0.925655	Oxic, 0 – 1	77.1 ± 15.0			
		Anoxic, 1 – 5	9.8 ± 0.8			
		Anoxic, 5 – 10	4.6 ± 0.3			
		Anoxic, 5 – 15	3.9 ± 0.03			
Stiffkey pool water			0.4 ± 0.1	38	7.5	17
Cley sediment	52.957825, 1.046553	Oxic, 0 – 1	91.4 ± 15.6			
Cley pool water			0.3 ± 0.02	32	7.6	17
Yarmouth sediment	52.614855, 1.715255	Oxic, 0 – 1	103.6 ± 30.4			
Yarmouth pool water			0.3 ± 0.01	30	7	17

4.2.2 Site characterization of Stiffkey salt marsh

A more in-depth analysis of the characteristics of the Stiffkey ponds was carried out with help from Andrew Hind, UEA, to provide further site information. Mud and water samples for incubation were collected from a small tidal pool in the same way that sediment was sampled for the culture-dependent and –independent experiments, at low tide. The sediment was taken. Salinity was measured to be 32 PSU, and the water temperature was 19°C. Conductivity and temperature were measured using a Fisherbrand accumet AP75, and salinity was determined from conductivity using a three point calibration, using Fisherbrand Traceable Conductivity Standards that are NIST Certified Reference Materials (CRM).

The oxygen saturation of the sediment was 62% immediately below the water surface, declining to 34% at half depth (80 mm from surface) and 29% immediately above the water/sediment interface (160 mm from surface). Oxygen measurements were made using a Jenway 970. A 2-point calibration was performed in the field at ambient temperature, using filtered seawater in equilibrium with air (100% oxygen saturation) and a 2 M sodium sulphite solution (0% oxygen saturation). Dissolved organic carbon (DOC) was calculated to be 3.60 mg/L. This is the mean of triplicate measurements, the standard error being 0.07 mg/L. Total dissolved nitrogen (TDN) was 0.59 mg/L, the mean of triplicate measurements, standard error 0.01 mg/L. TDN represents the sum of all dissolved

nitrogen-containing species, excluding dinitrogen (N_2), and includes organic nitrogen species as well as nitrate (NO_3^{1-}), nitrite (NO_2^-), ammonium (NH_4^+) and nitrous oxide (N_2O).

DOC and TDN measurements were made using a Skalar Formacs CA15 analyser, employing a six-point calibration. The calibration was validated against Environment Canada Environmental Matrix Reference Material Cranberry-05, lot 0317. Samples were also frozen so that nutrients analysis could be performed (phosphate, nitrate, nitrite, ammonium and silicate), but these are still awaiting analysis at CEFAS, and are not reported here.

The DMSP content of the oxic layer of sediment from tidal pool sediment (the top 1 cm) and the pool water at half depth (~80 mm) was also quantified using the purge and trap method, which is often used in environmental analysis as it provides greater sensitivity compared to standard GC. Measurements of 0.5 g were dissolved in 25 ml distilled water, with H_2SO_4. This mix was incubated at room temperature for 1 hour and then 5 ml was mixed with 1 ml 10 M NaOH and incubated overnight in the dark, before using the purge and trap method to quantify the DMS produced (Zhang et al. (2008). Purge and trap removes all the volatile organic compounds released by the sample through purging with an inert gas and trapping them in an analytical trap, which is a short gas chromatograph column. The compounds are then desorbed from the trap and injected into an Agilent 7890B gas chromatography (GC) instrument and quantified. Oxic Stiffkey sediment was shown to have a DMSP standing stock of 128.4 ± 14.0 nmol/g, which was 3 orders-of-magnitude higher compared to the pool water, which contained 0.7 ± 0.1 nmol/ml.

More sediment was also collected in order to perform a microcosm enrichment experiment on Stiffkey sediment to increase the amount of DMSP production by the sediment, hopefully enriching the community for DMSP-producers. DNA/RNA from Time 0 natural sediment and the end-point of the enrichments could then be sent for 16S rRNA amplicon and metagenomic sequencing. These experiments are described in **Chapter 6.**

4.2.3 Experiments with *Spartina anglica* in Stiffkey

As the high levels of DMSP in salt marshes have always been attributed to the activity of *Spartina anglica,* some preliminary experiments were performed with help from Peter Riviera and Yanfen Zheng, to quantify the DMSP content in *Spartina sp.* plants taken from Stiffkey, as well as measuring the change in the DMSP content of sediment along a transect moving away from *Spartina*.

The DMSP content of 0.1 g of four leaf and four root samples from a *Spartina* plant picked from a Stiffkey tidal pool was measured by pulverizing the sample and treating with methanol (MeOH) to create a methanolic extract, which was quantified using GC and then normalized to 1 g fresh weight (FW). The *Spartina* roots contained 2,568.6 ± 24.5 nmol/g

FW DMSP, and the leaves were found to contain DMSP at a level of 9,579.5 ± 796.9 nmol/g FW sample, which is close to some levels published in the literature (Otte et al. 2004), although much lower than in other papers (Kocsis et al. 1998), suggesting that there is a very variable range in concentrations, with some species of *Spartina* not able to produce any detectable levels.

Although the endogenous DMSP levels of the *Spartina* taken from Stiffkey are clearly much higher than any detected in bacteria so far, it was hypothesised that the DMSP produced by these plants will only influence the sediment that is most closely surrounding them, meaning that while the very high DMSP content in sediment close to the cordgrasses is predominantly due to eukaryotic activity, in sediment that is further away from the plants bacterial and or algal DMSP production is likely to play more of a significant role, as it is unlikely to diffuse through the sediment. To test this, oxic sediment samples were taken from a transect, starting with directly below a *Spartina* plant and then moving increasingly further away, and quantified for DMSP content by GC **(Figure 4-4)**.

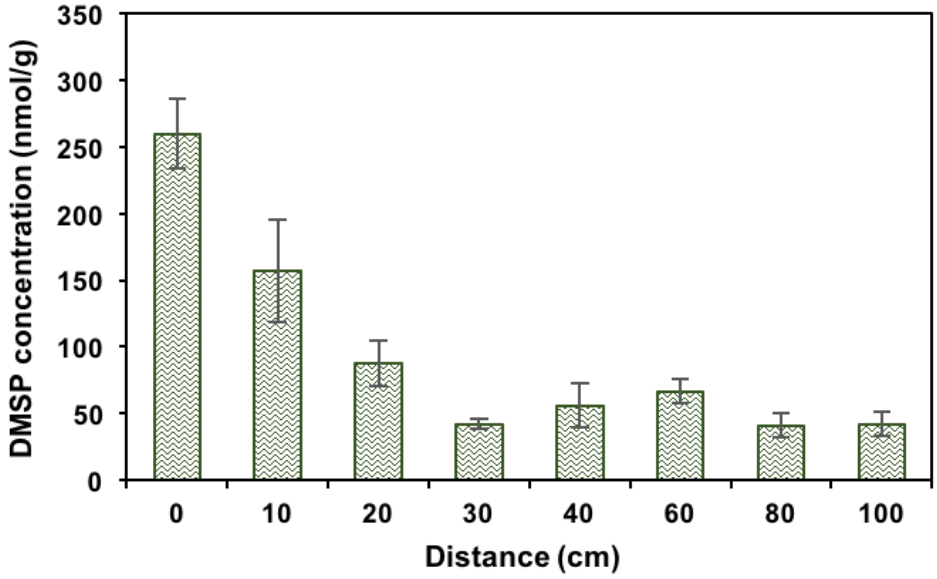

Figure 4-4: The change in DMSP concentration of Stiffkey oxic sediment as distance away from the cordgrass, *Spartina angelica*, increases, with 0 cm being sediment directly beneath the plant. Error bars display standard error.

In the first few samples (0 – 20 cm away) the DMSP concentration was extremely high, and decreased quite dramatically, but after that it appeared to stabilize, reaching a similar level to those measured by GC in other sampling experiments (see above), as the tidal pools were usually sampled in the centre, which would be at least 60 – 70 cm away

from any *Spartina* plants. This suggests that while *Spartina* certainly contributes to the total DMSP content of Stiffkey, there is also likely a major input from other DMSP-producing organisms, such as bacteria.

4.2.4 Isolating DMSP-producing bacteria from Stiffkey

The first experiment performed on Stiffkey salt marsh sediment, extracted on 17/03/15, was to determine that DMSP-producing bacteria could be isolated from the sediment. A serial dilution was performed on 100 µg of sediment in MBM media to a dilution factor of 10^{-6}, then 100 µl was plated on MBM agar containing a mixed carbon source, with no selective pressure other than selection for heterotrophic bacteria. Plates were incubated for a week at 28°C and colonies of different morphologies were purified to single colonies **(Figure 4-6)**, then picked and tested for DMSP production GC. Of the species isolated and checked, 27% (9 of 33 tested colonies) were found to produce detectable peaks of DMS when treated with NaOH (which chemically cleaves DMSP into DMS and acrylate) **(Table 4-2),** and isolates were identified by the sequencing of their 16S rRNA gene **(Table 4-3)**. This data showed that DMSP-producing bacteria are present in the sediment, and are relatively easy to isolate. Although DMS can be released from other compounds upon addition of NaOH, the most likely explanation is that it originated from alkaline lysis of DMSP. Furthermore, for several of these bacteria, DMSP production was confirmed by analytical LC-MS **(Table 4-5).** The DMSP-producing bacteria identified were mainly alphaproteobacterial, but there were also some gammaproteobacteria isolated. The purification and identification of these bacteria is described in detail below.

4.2.5 Optimising conditions for DMSP-producing bacteria

The next step was to perform growth experiments on the sediment from Stiffkey to investigate conditions that potentially enhance DMSP synthesis. The aim of this is to enrich for a variety of DMSP-producing isolates, including some with potentially novel DMSP-synthesis genes/pathways. This was a challenge because enriching for a process is more complicated than enriching for the uptake or degradation of a particular substance, as mentioned above.

Sediment was taken from tidal pools in Stiffkey, and 2 g weighed into 100 ml flasks, then mixed with 30 ml MBM medium of different conditions in triplicate, as well as a control of standard MBM **(Figure 4-5)**. The conditions were chosen based on those observed to increase DMSP synthesis in *L. aggregata* (Curson et al. 2017) **(Figure 4-3)**. Increased salinity has long been known to increase DMSP synthesis (Karsten et al. 1992), as has low nitrogen levels (Sunda et al. 2007). Low nitrogen is thought in part to increase DMSP synthesis because a lack of nitrogen results in decreased production of glycine betaine, a nitrogen-based osmolyte, and therefore in bacteria where both osmolytes are made,

DMSP synthesis is preferentially produced. Furthermore the process of DMSP synthesis actually liberates an amino group from Met in the transamination step, providing extra nitrogen for the organism to use (see **Chapter 1**). For the same reason, it was hypothesised that increased levels of sulfur may increase DMSP production, as extra sulfur means that it is available for use in the formation of compounds such as DMSP. The other conditions trialled were supplementing the media with MTHB, the precursor for the reaction catalysed by *dsyB* gene product (Curson et al. 2017), and finally, using a combination of all four of those conditions. The flasks were incubated at 30°C with shaking at 180 rpm for one week, and the DMSP content analysed by NaOH addition and GC analysis, as above **(Figure 4-5)**.

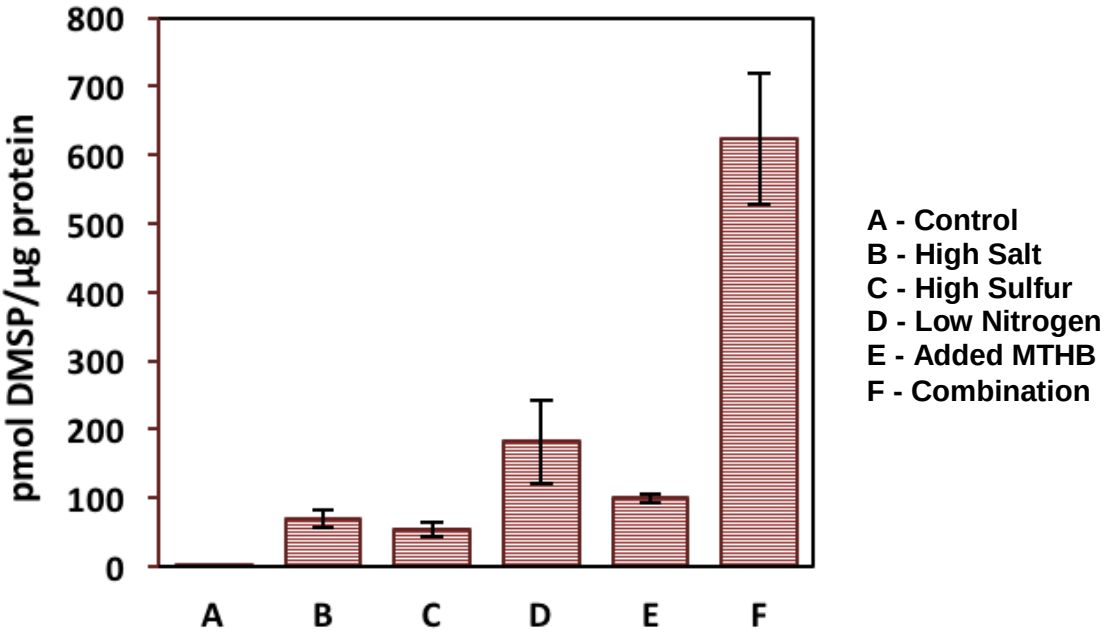

Figure 4-5: The DMSP production by the pelleted sediment taken from Stiffkey and treated with one of six conditions, including increased salinity and sulfur, decreased nitrogen, added MTHB and a combination of all of them. Samples are averaged and bars display standard error.

Unsurprisingly, the most effective enrichment condition by far was the media that combined all four of the other conditions, with DMSP production being increased at least three-fold more than any other condition. Therefore, this was used as the enrichment condition from this point onwards.

Table 4-2: Table of the initial DMSP production of isolated bacteria from Stiffkey salt marsh, with those isolated from Time 0 sediment labelled T1-9, and those isolated from incubated experiments labelled E1-51.

Isolate	Conditions of isolation	Intracellular DMSP concentration in MBM 0.1 mM Met (pmol DMSP µg protein-1)	Isolate	Conditions of isolation	Intracellular DMSP concentration in MBM 0.1 mM Met (pmol DMSP µg protein-1)
S1	T0 sediment	23.1	**E22**	Low Nitrogen	68.8
S2	T0 sediment	4.9	**E23**	Low Nitrogen	1.4
S3	T0 sediment	11.4	**E24**	Low Nitrogen	34.7
S4	T0 sediment	30.9	**E25**	Added MTHB	21.0
S5	T0 sediment	518.7	**E26**	Added MTHB	6.9
S6	T0 sediment	125.4	**E27**	Added MTHB	75.0
S7	T0 sediment	6.8	**E28**	Added MTHB	9.1
S8	T0 sediment	25.7	**E29**	Added MTHB	104.2
S9	T0 sediment	6.7	**E30**	Combination	23.8
E1	Control	28.8	**E31**	Combination	2.6
E2	High Salt	2.0	**E32**	Combination	12.0
E3	High Salt	3.2	**E33**	Combination	2.7
E4	High Salt	2.0	**E34**	Combination	2.4
E5	High Salt	21.2	**E35**	Combination	101.0
E6	High Salt	7.6	**E36**	Combination	2.6
E7	High Salt	3.3	**E37**	Combination	3.3
E8	High Sulfur	2.0	**E38**	Combination	2.4
E9	High Sulfur	3.1	**E39**	Combination	116.3
E10	High Sulfur	2.3	**E40**	Combination	2.3
E11	High Sulfur	2.9	**E41**	Combination	83.9
E12	High Sulfur	2.9	**E42**	Combination	37.7
E13	High Sulfur	5.9	**E43**	Combination	5.5
E14	High Sulfur	3.7	**E44**	Combination	52.6
E15	High Sulfur	2.0	**E45**	Combination	226.8
E16	High Sulfur	2.4	**E46**	Combination	3.9
E17	Low Nitrogen	2.6	**E47**	Combination	2.3
E18	Low Nitrogen	2.1	**E48**	Combination	10.5
E19	Low Nitrogen	5.6	**E49**	Combination	1.9
E20	Low Nitrogen	3.5	**E50**	Combination	2.2
E21	Low Nitrogen	1.5	**E51**	Combination	2.6

To test for the variety and abundance of DMSP-producing species in each condition, the sediment from these enrichment experiments was then used in isolation experiments, similar to the ones described above, plating the sediment on MBM agar plates of the same conditions as they were incubated in **(Figure 4-6)**. After incubation at 28°C for one week, an average of 15 colonies of different morphologies were purified to single colonies from each condition and inoculated in MBM with Met, as above. These cultures were characterised in the same way as above, testing them for the ability to produce DMSP, and the percentage of DMSP-producers compared to non-producers was calculated **(Figure 4-7)**. Once again it was found that sediment treated with the media that combined all four of the conditions provided the highest percentage of DMSP-producing bacteria.

TIME 0 – 10^{-1} dilution **COMBINATION– 10^{-5} dilution**

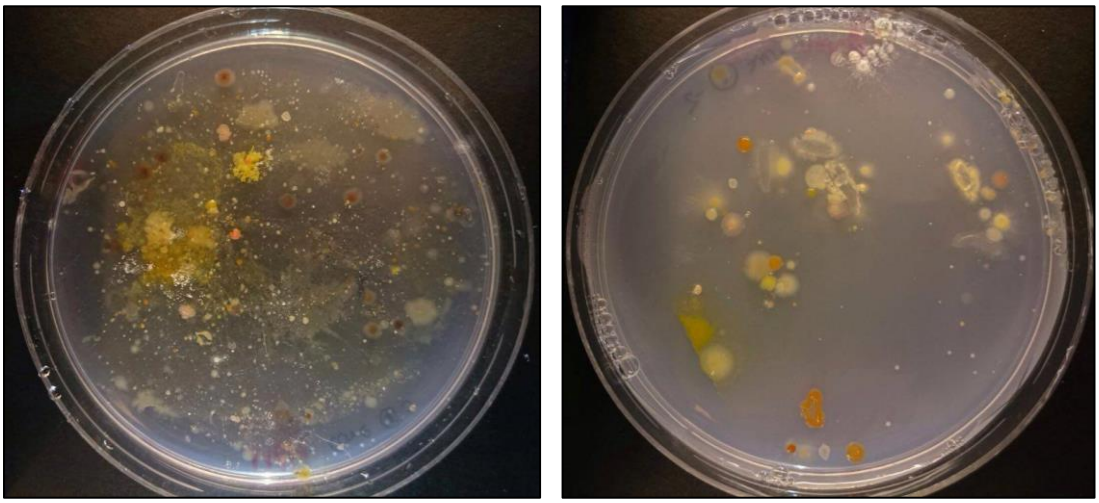

Figure 4-6: Example MBM agar plates of colonies isolated from Stiffkey sediment displaying the variation in morphology. Plates are sediment from before enrichment (T0) at a dilution factor of 10^{-1}, and after incubation in the combination media for one week and diluted at a factor of 10^{-5}.

Together with the observed increased in DMSP levels in **Figure 4-5**, the combination media was judged to be the most effective in optimising Stiffkey salt marsh sediment for increased DMSP production and DMSP-producing isolates.

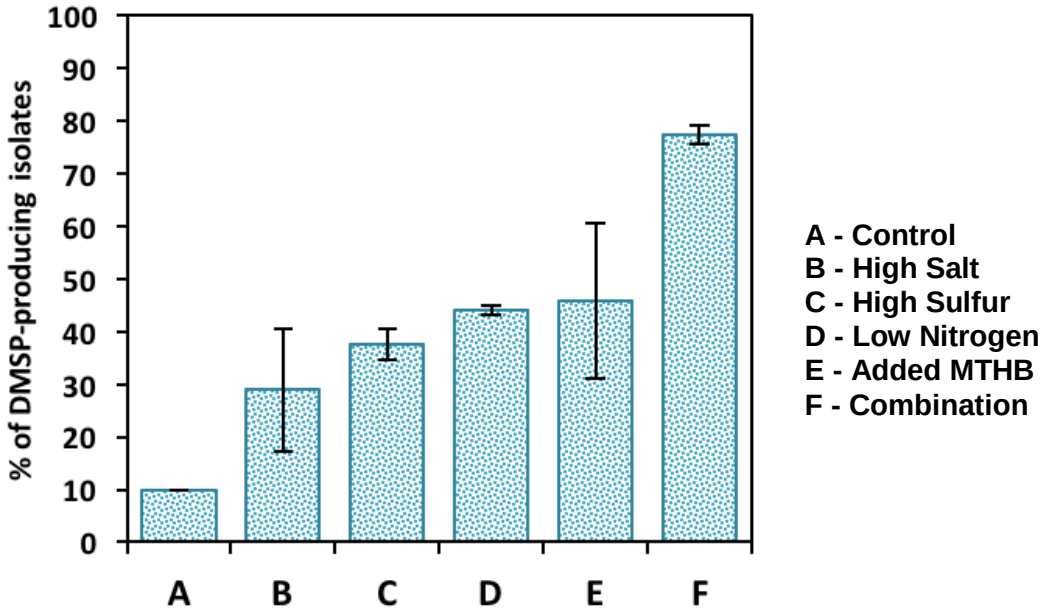

Figure 4-7: The percentage of species isolated from sediment that were able to produce DMSP. Sediment was treated with different media conditions including increased salinity and sulfur, decreased nitrogen, added MTHB and a combination of all of them. Samples are averaged and bars display standard error.

4.2.6 Purification of DMSP-producing bacteria

The library of isolates were purified and their taxonomy identified using 16S rRNA sequencing. Many of the isolates had the same or similar phylogeny, so only representative isolates for each genera was fully characterised, and these are summarised in **Table 4-3**. Purification of isolates was achieved by streaking a culture for single colonies on MBM agar plates, until no contaminating colonies were observed. These colonies were then inoculated into fresh media, and once grown they were visually checked using Microscopy to look for homogeneity in cell size and shape, ensuring purity **(Figure 4-8).** If purity was not confirmed by microscopy, and more than one cell morphology was observed, then the cultures were serially diluted to dilution factors between 10^{-4} and 10^{-6}, and plated on MBM agar. If, after incubation at 28°C overnight, colonies of multiple morphologies were observed on the plate, both were picked and tested individually for DMSP production. Once purified, isolates were stored at -80°C as well as on agar plates at 4°C, and streaked again every three months. When DMSP content had been tested again and was confirmed, and the samples were pure, they were then classified again to confirm identity **(Table 4-3)**.

The identification of pure isolates was accomplished by using PCR amplification of the 16S rRNA gene on genomic DNA isolated from pure cultures of the isolates, using the primer set 27F/1429R. The resultant PCR product was extracted using a PCR purification kit, after visualising 5 µl of the PCR product using gel electrophoresis, to confirm successful amplification. The purified 16S rRNA fragments were then sent to Eurofins Genomics for sequencing, and the phylogenetic identity was confirmed by submitting the sequences to a nucleotide BLAST against all sequences in the NCBI database, then taking the top hit, which was usually 99% identity. One example isolate of each different strain is represented in **Table 4-3.**

Of the isolates that were identified using 16S rRNA PCR amplification, there were several that were expected to be DMSP producers, including *Labrenzia, Oceanicola, Pseudooceanicola* and *Stappia.* All of these are genera known to contain *dsyB* and produce DMSP, so their identification was not unexpected. In addition to this, *Rhodobacter* and *Rhodobacterales* are both closely related to *Labrenzia*, so it was also unsurprising to find that they produce DMSP. There were three strains where it was surprising to observe DMSP, these were *Novosphingobium*, *Marinobacter* and *Alteromonas. Novosphingobium*, while being an alphaproteobacterium, is from the order Sphingomonodales which has not been implicated in DMSP production. The same was true for both *Marinobacter* and *Alteromonas* which are gammaproteobacteria, and are thus the first of this class to be shown to produce DMSP.

Once pure strains were achieved, cultures were inoculated into 5 ml MBM media containing 0.5 mM nitrogen, which, although not a low enough level to be classed as limited, was significantly lower than the nitrogen conditions used in the initial DMSP quantifications in **Table 4-2**. This low nitrogen condition was used because in both the enrichment conditions **(Figure 4-5)** and in induction experiments on *L. aggregata* **(Figure 4-3)**, low nitrogen levels were found to significantly increase the production of DMSP within the samples. More importantly, this reduced nitrogen level is closer to the normally limiting nitrogen concentrations experienced in many marine environments (see above). DMSP levels in reduced nitrogen conditions were quantified after an overnight incubation, and are reported in **Table 4-3**. It should be noted that the nitrogen levels used are still far higher than those experienced in the field, but the bacterial isolates will not grow to sufficient densities to detect DMSP at any lower amount. Almost all isolates showed an increase in DMSP production when incubated under the lower nitrogen conditions.

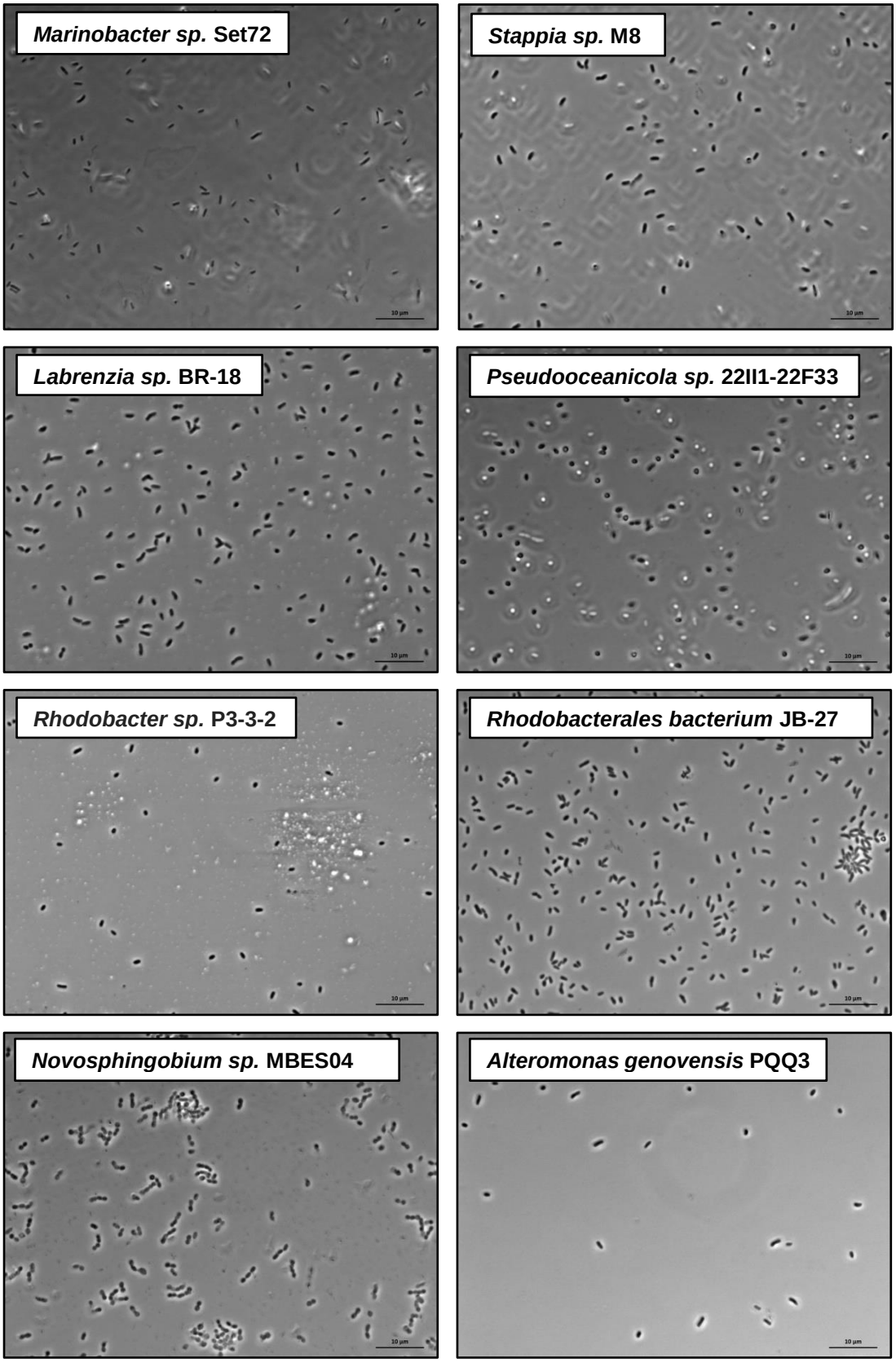

Figure 4-8: Microscopy images of 100x magnification of eight of the ten further characterised and identified strains, showing some variety in morphology. This technique was used to confirm purity of the samples through the presence of only one cell type.

Table 4-3: Characterisation of ten of most abundant and/or novel bacteria isolated from Stiffkey sediment.

Example isolate	Closest 16S rRNA gene identity	Conditions of isolation	Intracellular DMSP concentration, MBM 0.5 mM nitrogen (pmol DMSP µg protein^{-1})	Presence of *dsyB* with degenerate primers	Presence of DMSP indicated by LC-MS
S4	*Marinobacter sp.* Set72	T0 sediment	39.8 ± 1.3	-	YES
S5	*Labrenzia sp*. BR-18	T0 sediment	278.6 ± 6.2	+	YES
S8	*Stappia sp.* M8	T0 sediment	153.1 ± 11.5	-	NT*
E26	*Pseudooceanicola sp.* 22II1-22F33	+ MTHB	64.4 ± 2.3	-	YES
E27	*Rhodobacter sp.* AB300d	+ MTHB	495.4 ± 53.5	-	YES
E30	*Oceanicola sp.* Ar-45	Combination	78.5 ± 0.58	+	NT
E35	*Rhodobacterales bacterium* JB-27	Combination	762.0 ± 403.9	+	YES
E37	*Roseobacter sp.* ARCTIC-P4	Combination	44.2 ± 3.4	NT	NT
E45	*Novosphingobium sp.* MBES04	Combination	665.8 ± 102.3	-	YES
E48	*Alteromonas genovensis* PQQ33	Combination	6.9 ± 3.5	NT	NT

* NT, Not Tested

It was important to confirm that peaks of DMS released when these samples were lysed in alkaline hydrolysis were in fact due to synthesised DMSP. This is because one weakness with using the alkaline hydrolysis method is that other methylated sulfur compounds such as SMM and DMSHB also liberate DMS when treated with NaOH, although not as readily as DMSP (the mixture has to be incubated at 80°C for 10 minutes before they fully lyse). In order to confirm that the bacteria isolated from Stiffkey are synthesising DMSP and not just some of its precursors, LC-MS analysis was utilised with the help of Ana Bermejo-Martinez to identify and confirm the presence of metabolites produced by some of the isolates. Due to time and cost constraints, not all isolates were tested in this manner, but of those that were, all accumulated DMSP at a diagnostic retention time of 4.9, with the appropriate mass/charge ratio **(Table 4-3)** (see **Chapter 5** for example chromatograms).

4.2.7 Experiments to determine potential inducers of DMSP production

Further study was undertaken to characterise some of the isolated strains. It had been shown that changing growth conditions affected the production of DMSP in many of these species, e.g. lower nitrogen levels **(Table 4-3)**. Other induction experiments were set up to observe the effect of variations in media or the addition of intermediates from the transamination pathway (see **Chapter 1**) on DMSP synthesis. Isolates were inoculated in 10 ml triplicate MBM, either standard conditions, which at the time was 20 PSU with 12 mM nitrogen, with 5, 35 or 50 PSU salinity, or lowered nitrogen levels. Met, DMSHB, MTHB or MMPA (intermediates of the transamination pathway) were added separately to a final concentration of 0.5 mM to standard MBM. Cultures were incubated overnight at 30°C with shaking, reaching stationary phase. DMSP quantity was then measured using GC, and normalised for protein concentration **(Figure 4-9)**.

The results of these growth experiments were interesting, with a noticeable variation in DMSP production levels between strains, as well as between growth conditions. Both *Labrenzia* and *Stappia* show an expected pattern of induction, with low nitrogen causing an increase in synthesis compared to the standard, although it was surprising that high salinity did not also increase production as was observed in **(Figure 4-3)**. This could be because some of the bacterial isolates may not survive as well in the raised 50 PSU salinity. For both, the most significant increase was when cultures were incubated with any of the four transamination pathway intermediates, all of which showed similar levels of induction, suggesting that all four are utilised as part of the production pathway.

It was expected that the addition of Met would cause an increase in production, as it is thought to be the initial DMSP synthesis precursor molecule (see **Chapter 1**). Indeed, it was found that Met induced higher DMSP levels than the standard samples, consistent with it being the universal starting precursor for DMSP synthesis, but there seems to be a range of responses. In *Marinobacter* and *Rhodobacterales* there is only a slight increase in DMSP production with the addition of Met. It could be that those concentrations inhibit growth of those species, or it may not be taken up very efficiently. Furthermore, the isolates in question may have had more urgent uses for Met, meaning it was not used to make DMSP.

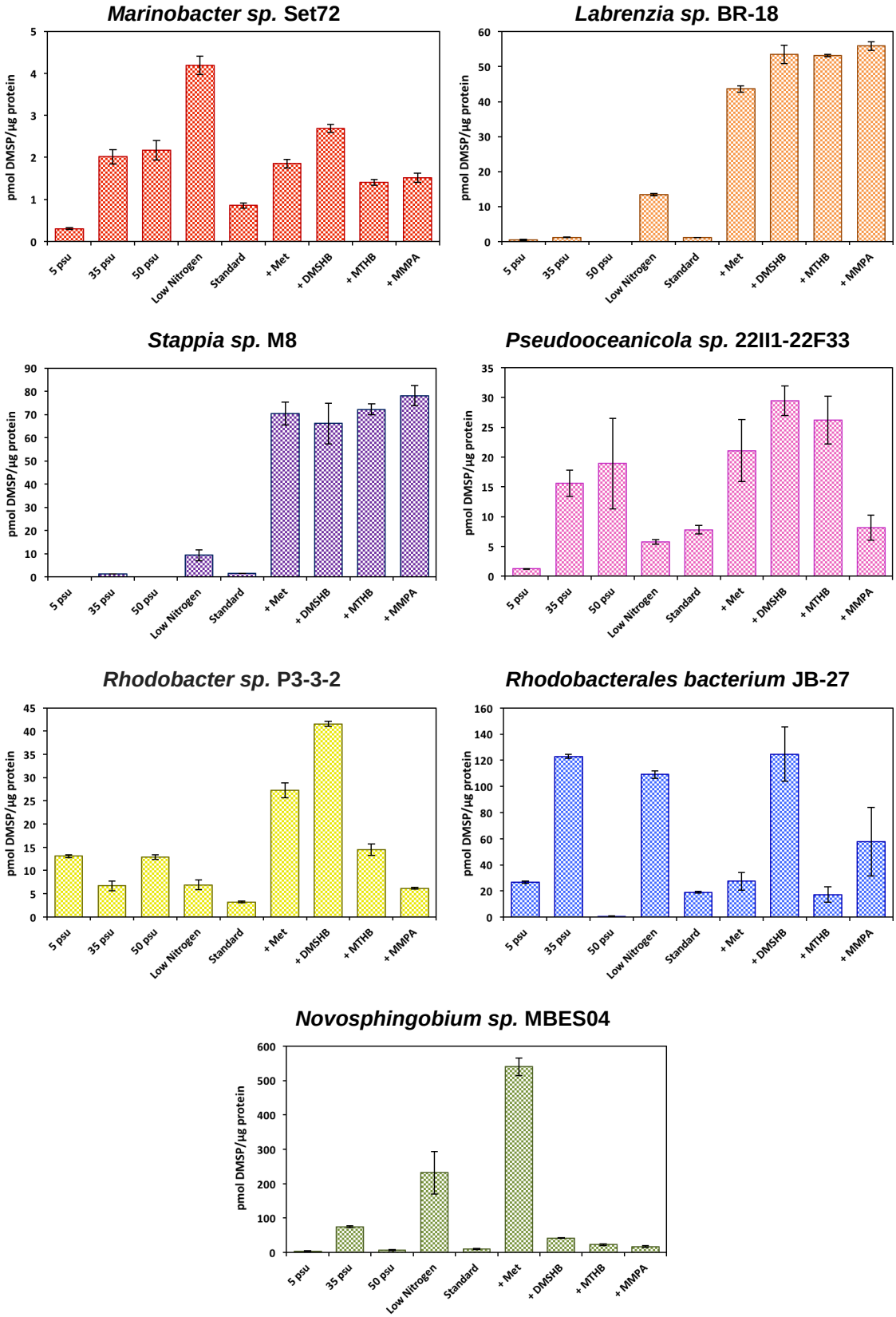

Figure 4-9: Induction experiments performed on seven of the ten isolates from Stiffkey salt marsh. Cultures were grown in standard MBM, or in 5, 35 and 50 psu salinity, low nitrogen, added Met, DMSHB, MTHB, MMPA.

Surprisingly, although *Novosphingobium* shows a dramatic increase in DMSP production once treated with Met, it does not appear to be affected by the addition of other transamination pathway intermediates. The most likely explanation for this is that, while *Novosphingobium* is indeed synthesising DMSP, it is not doing so using the transamination pathway. This is significant because thus far bacteria have only been observed to be using the transamination pathway, believed to be the predominant pathway in marine environments used by marine algae, bacteria with *dsyB* and corals.

In some isolates the addition of DMSHB caused a higher production of DMSP compared to others (*Rhodobacter* and *Rhodobacterales*). This is most likely due to the previously mentioned issue of DMS being released from DMSHB as well as from DMSP, although because the samples were not incubated at 80°C for 10 minutes only a small portion of DMSHB would have lysed.

4.2.8 DMSP seawater incubations

As the conditions that have been used in DMSP quantification experiments and these incubation experiments are far from natural, seawater incubation experiments were also designed in order to demonstrate that DMSP could still be produced *in situ*, and therefore is also likely to be taking place in the environment, not just under laboratory conditions. Cultures of *dsyB*-containing *Pelagibaca bermudensis* and non-*dsyB*-containing *Novosphingobium* (see later) were inoculated into MBM and incubated overnight, then adjusted to an OD_{600} of 0.4 and diluted 1:100 into 20 ml filter-sterilised seawater (T0), followed by incubation at 25°C with 90 rpm for 21 hours (T1) and 43 hours (T2). Bacterial cell pellets were collected and resuspended in Tris-HCl buffer (50mM, pH 7.5), and the supernatant removed. DMSP in the pellets and supernatants were measured by the addition of alkaline-hydrolysis, where DMS was generated and then processed by a modified purge and trap method, and measured by GC **(Figure 4-10)**.

The seawater incubations below demonstrate that *Pelagibaca* and *Novosphingobium* produce DMSP under close to *in situ* conditions, and therefore it is likely they produce DMSP in natural marine environments, such as Stiffkey salt marsh. The DMSP content in the supernatant is much lower than the total (supernatant and pellet), and only slightly increases over time, likely due to the release of DMSP from the cells by export or cell lysis after death. In contrast, the DMSP in the total increases to a much higher level, suggesting that the isolates are synthesising DMSP, not just exchanging it with environmental levels. Perhaps a longer incubation would show even more of a more significant pattern.

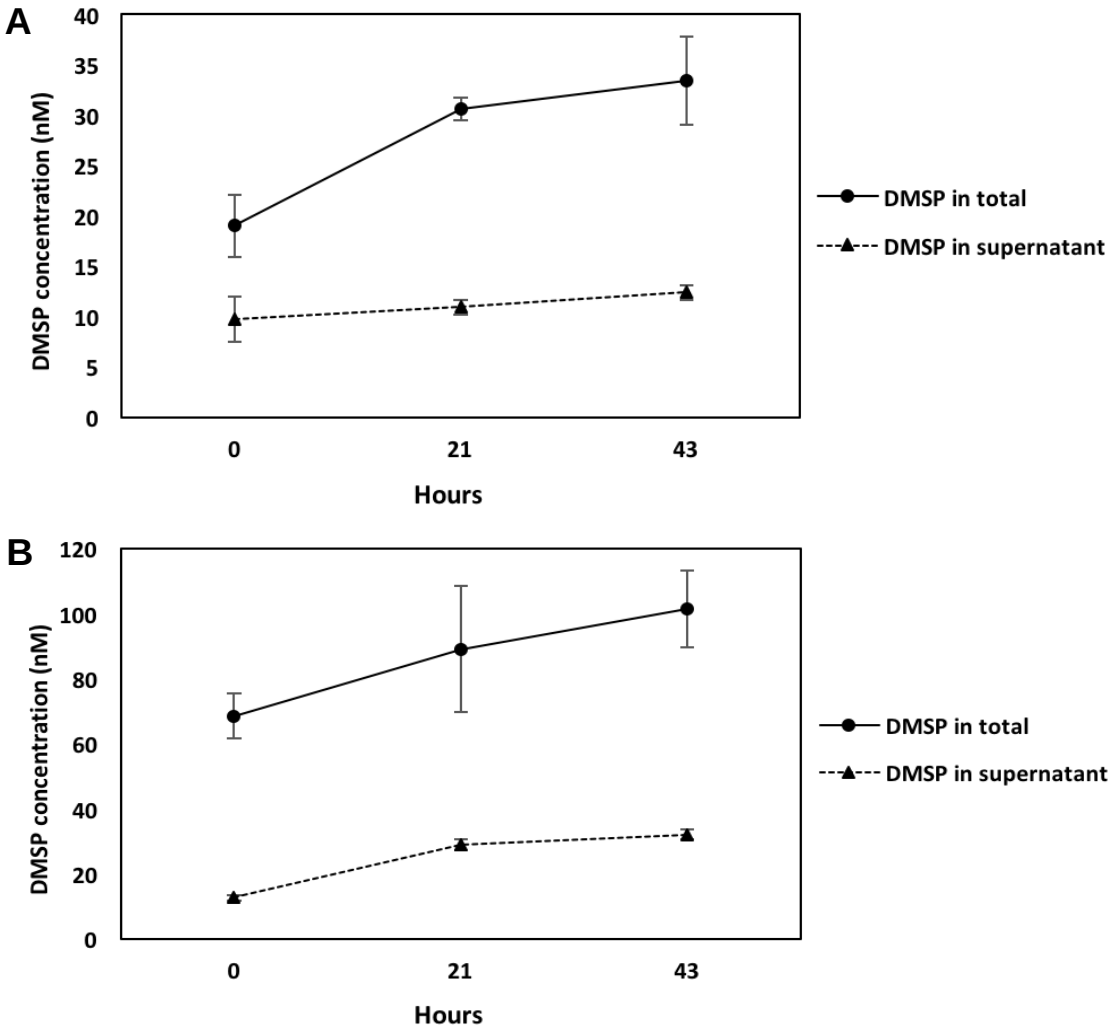

Figure 4-10: Seawater incubation results for the DMSP produced by both the supernatant and total culture of the *dsyB*-containing isolate, *P. bermudensis* (**A**), and for the non-*dsyB* isolate *Novosphingobium sp.* MBES04 (**B**), across an incubation of 21 and 43 hours, processed by purge and trap and measured by GC. Error bars display standard error.

4.2.9 Confirming the presence/absence of *dsyB*

The next logical step for the DMSP-producing isolates was to establish if they contained *dsyB,* and were therefore likely using the transamination pathway to produce DMSP. This was initially accomplished by utilising the degenerate primers, designed in **Chapter 3**, in PCR amplification reactions on several of the isolates. PCR was done on genomic DNA isolated from the pure cultures, where indicated in **Table 4-3**. PCR products were subjected to gel electrophoresis as a preliminary method for the detection of *dsyB* **(Figure 4-11)**.

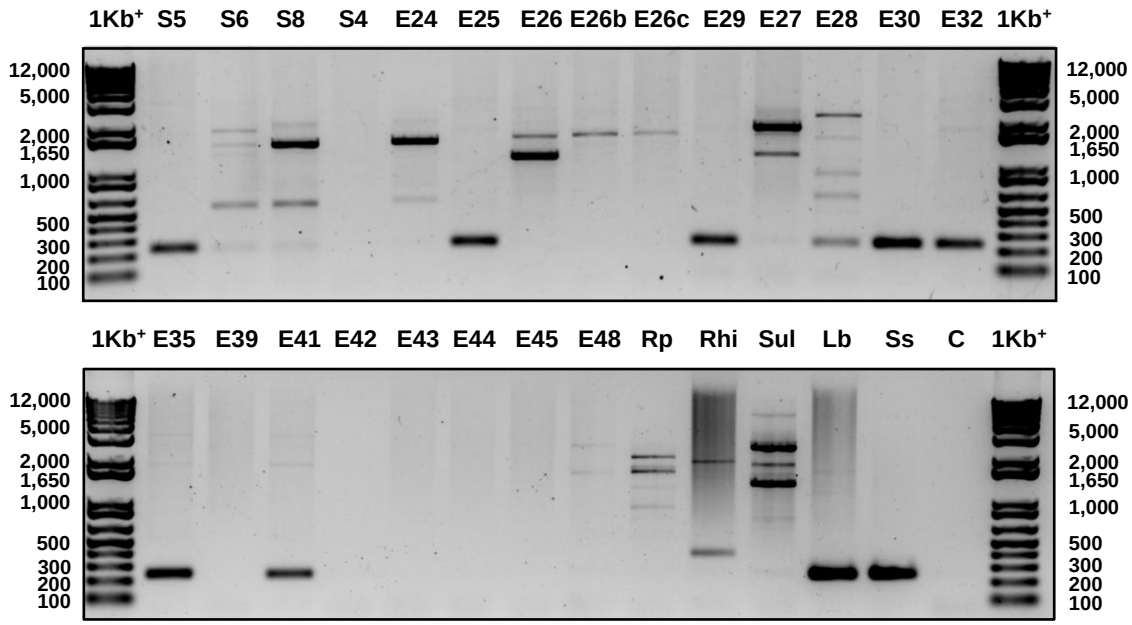

Figure 4-11: Gel electrophoresis of the *dsyB* degenerate primer PCR on genomic DNA, from Stiffkey isolates listed in Table 5.1. Also tested were negative controls (*R. leguminosarum*, *R. pomeroyi* and *S. sp EE-36*) and positive controls for *dsyB* (*L. aggregata* and *S. stellata*)

Although we cannot be certain that the lack of a band at 246 bp indicates a definitive lack of *dsyB*, the presence of a band is almost certainly indicative of it. Unsurprisingly (see above), the *Labrenzia* strain showed the presence of *dsyB* (S5), as did the *Rhodobacterales* (E25, E32, E35, E41) and *Oceanicola* (E30) isolates, of which *Oceanicola* was previously known to contain *dsyB,* and *Rhodobacterales* is similar phylogenetically to *Labrenzia*. The bands were excised and sequenced to confirm *dsyB* identity. However, most of the remaining isolates showed no PCR product. *Pseudooceanicola* (E26) was negative despite closely related strains with genomes being known to contain *dsyB*. Indeed, a *Pseudooceanicola* DsyB sequence was used in the degenerate primer design process. *Rhodobacter* (E27) and *Stappia* (S6, S8, E24), from the same family as *Labrenzia*, were also expected to contain it. This suggests that the degenerate primers, whilst able to amplify many DsyB sequences from genomic and metagenomic DNA, may not cover the full range, perhaps because some are more divergent than those from which the primers were designed, or there were inhibitory factors in the PCR mix. Interestingly, *Novosphingobium* (E39, E42, E43, E44, E45) from the Sphingomonodales order, and *Alteromonas* (E48) and *Marinobacter* (S4) which are both gammaproteobacteria, were all negative. This was expected as *dsyB* is not seen in any sequenced representatives from these genera.

Thus far we have isolated several species of bacteria that have not been previously demonstrated to produce DMSP, several of which do not appear to contain *dsyB*, the only known DMSP-synthesis gene so far. It was hypothesised that, for some of these bacteria at least, this was a weakness of the degenerate primers, and not that the species do not contain *dsyB*. To test this, the sequence of a species of *Pseudooceanicola sp.* La6*,* isolated from the English Channel by Alex Howat in the Murrell lab, was used to design specific primers to the *dsyB* in its genome. Both these and the *dsyB*-degenerate primers were tested on *Pseudooceanicola sp.* La6, the *Pseudooceanicola* isolate, *L. aggregata* as a control alphaproteobacterium, and a water control, then run on gel electrophoresis to determine if the isolate also contains *dsyB* **(Figure 4-12)**.

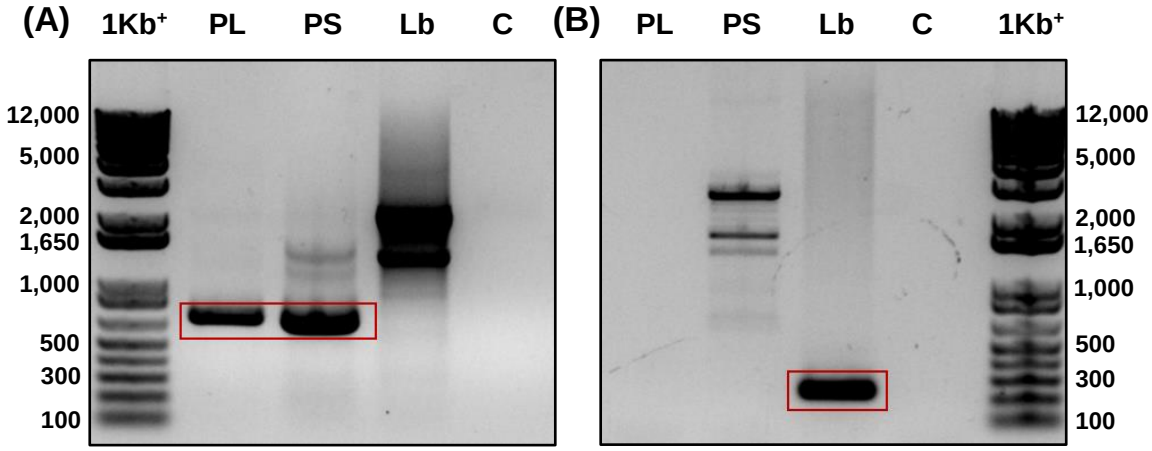

Figure 4-12: Gel electrophoresis to test *Pseudooceanicola,* using both the *dsyB*_degenerate primer PCR (**A**) and *Pseudooceanicola dsyB*_specific PCR (**B**) on *Pseudooceanicola sp.* La6 (PL), *Pseudooceanicola* Stiffkey isolate (PS) and *L. aggregata* IAM12614 (Lb), and a water control (C). The *dsyB*_specific primers should produce a fragment of 700 bp in size, and the *dsyB*_degenerate primers produce a band of ~246 bp.

From this experiment it was clear that not only is it possible to perform PCR amplification on *Pseudooceanicola* genomic extracts, ruling out the possibility that the degenerate primers did not amplify due to some inhibitory substance in the genomic extraction, but also confirmed that the *Pseudooceanicola* isolate contains *dsyB*. This suggests that the degenerate primers designed in **Chapter 3** are not able to amplify all versions of *dsyB* under the conditions used here, even though they were designed from sequences very similar to this isolate. Perhaps more optimisation is required for individual bacteria. Although this is unfortunate in that it lessens the ability to definitively identify the presence or absence of *dsyB,* it also means that any assumptions made about the abundance and activity of *dsyB* in the environment are likely to be vast underestimations of the actual bacterial DMSP production in those environments. When working with degenerate gene probes such as these, this is a common occurrence. Nevertheless these *dsyB* primers are still a valuable environmental resource.

As the degenerate primers were only a rough method of detection, used to rule out *dsyB*-containing sequences, further analysis was needed to confirm whether or not the negative results contain *dsyB.* In order to accomplish this, several species were sent to Microbes NG, Birmingham for Whole Genome Sequencing (WGS). Pure cultures of *Marinobacter, Stappia, Novosphingobium, Rhodobacter, Rhodobacterales* and *Alteromonas* were sent to Birmingham, as all of them show no band at 246 bp **(Figure 4-11)**.These were sequenced using the Illumina MiSeq platform, and then trimmed and checked for quality. Those reads were then annotated against the closest genomic sequence available, using RAST (http://rast.nmpdr.org/) **(Table 4-4)**. Although this sequencing does not generate complete genomes, the completeness of strains done here were greater than 99%.

Table 4-4: Whole Genome Sequenced Stiffkey isolates

Isolate	Closest Identity	Length (Kb)	Contigs	Predicted Proteins	16S Identity (%)	Presence of DsyB Identity (%)	DsyB E Value
S4	*Marinobacter sp*. Set72	4221.7	103	3864	99	/	/
S8	*Stappia sp.* M8	6393.0	77	6043	99	96	0.0
E27	*Rhodobacter sp.* AB300d	6796.2	746	6392	88	60	1.00 E-141
E35	*Rhodobacterales bacterium* JB-27	6105.0	160	5687	99	86	0.0
E45	*Novosphingobium sp.* MBES04	4461.8	136	4216	99	/	/
E48	*Alteromonas genovensis* PQQ33	4930.8	120	4397	98	/	/

From the WGS of the isolates it was shown that the three species that were expected to contain *dsyB* do in fact contain it, but the other three, *Novosphingobium, Alteromonas* and *Marinobacter*, have all been confirmed to not possess a copy. Given the nature of this genome sequencing (i.e. not generating complete genomes) it is possible that the three remaining DMSP-producing bacteria do indeed contain *dsyB* but that it is contained within a missing component of their genomes which was not sequenced. However, I feel this is unlikely given that these bacteria are not closely related to any known bacteria which contains *dsyB*. Further experiments conducted in subsequent chapters support the hypothesis that these bacteria do indeed lack DsyB and likely contain a novel DMSP synthesis pathway.

4.3 Discussion

4.3.1 Summary of work

The work described in this chapter was the culture-dependent analysis of sediment sampled from the tidal pools that make up the lower section of Stiffkey salt marsh, on the North Norfolk coast. These pools are flooded twice daily with the high tide, and therefore maintain seawater levels of salinity all year round (Davy & Smith 1988). This high salinity, combined with the large flux of biogenic sulfur, in particular the high emissions of DMS (Steudler & Peterson 1984), make Stiffkey sediment an ideal environment in which DMSP production is likely to occur. Because of this, we were confident that it would be possible to isolate DMSP-producing species, and through those isolations, potentially find novel isolates and DMSP synthesis genes.

It was discovered that this was indeed possible, with almost a third of the bacterial species isolated from sediment dilutions incubated on agar with no selective pressure were demonstrated to possess the ability to synthesise DMSP **(Table 5.1)**. Furthermore, it was also shown that this number could be dramatically increased (from 27% to 77%) when sediment was treated with an optimised media mix that created conditions favourable for DMSP production. When these bacteria were identified, it was revealed that several of them were of genera that had not previously been shown to produce DMSP or contain DsyB (*Marinobacter, Novosphingobium,* and *Alteromonas*). There were also several species of *Labrenzia*, and the closely related *Stappia*, which was unsurprising. Indeed, it was the reclassification of several species of *Stappia* that were less related to other *Stappia* sequences that created the genus *Labrenzia,* including the species in which DMSP production was first discovered – *Labrenzia aggregata* (Biebl et al. 2007). Other species of *Rhodobacter* and *Rhodobacterales* were also isolated and shown to produce DMSP.

Those bacterial isolates that were shown not to contain *dsyB,* through the use of both degenerate primers and WGS, were most likely either using a *dsyB* isoform gene to synthesise DMSP via the transamination pathway, or they were using an entirely novel bacterial pathway to produce this molecule. For *Marinobacter* it appears likely that it uses the transamination pathway since the intermediates from this pathway enhance DMSP production in it **(Figure 4-9)**. However, in *Novosphingobium,* only the addition of Met had any effect on the levels of DMSP production detected. To test the above hypothesises on *Novosphingobium* incubation and molecular experiments were carried out to explore this as described in subsequent chapters. In brief, known intermediates from the transamination, methylation and decarboxylation pathways were added to cultures of selected isolates and incubated for a short period of time (to limit possible transformations),

and their effect on DMSP production was determined as above. The results from these experiments clearly show that *Novosphingobium* utilises a novel bacterial DMSP synthesis pathway (see **Chapter 5**).

4.3.2 Issues with culture-dependent study

There are many benefits from carrying out culture-dependent microbiological work in environmental biology, but there are also number of problems associated too (see above). Most prominent is the issue of uncultivable bacteria (or as yet uncultivated). Although some bacteria such as *E. coli* and *Micrococcus* are easy to culture in standard conditions in the laboratory, and others like *Novosphingobium* and other alphaproteobacteria grow easily on MBM agar, it is thought that only around 1% of the 10^9 bacterial cells in a gram of sediment form colonies in standard plate isolations (Davis et al. 2005). This could be because other species require a specific combination of nutrients, or a particular media composition, or that the conditions for growth are unusual or difficult to recreate in the laboratory. Furthermore, not many bacterial species grow fast enough to form colonies before the plate is taken over by other, more prolific species. They may also be at a lower abundance in the sediment to begin with, meaning that the chances of them being picked are much reduced.

To a degree, the odds of isolating a variety of species can be improved by altering the variables of growth (Davis et al. 2005). These can include the dilutions of sediment plated, the growth medium and the incubation time, as well as other factors like temperature and whether or not the plates are treated aerobically or subcultured before plating. Even so, it is unlikely and impractical to attempt to culture every species – it is not realistic to trial every possible culture condition, and it is likely that there are many species that will not grow on plates, and instead only replicate in liquid media. There has to be a balance between optimising for a high number of isolates, and trying to culture everything. To that end, in this study culture-dependent plate isolations were used after sediment was treated to an ‘enriched’ media, also plating higher dilutions of sediment and having longer incubation times of up to two weeks (any longer and fungus started to contaminate the plates). Once a good number (~100) of colonies had been picked and tested, it was time to move forward with characterising the most interesting species. These techniques were not used for any kind of cell counting, or to make assumptions about the full community at Stiffkey. Presently (as of work reported in this chapter), we have no way to determine whether our isolates are environmentally relevant (present at a significant frequency in the Stiffkey sediment) or not. In the next chapter we use 16S rRNA gene sequencing and metagenomics to analyse the microbial community within these Stiffkey samples. This allows us to more accurately establish the environmental prevalence of any of the bacteria that we isolate in this chapter.

4.3.3 Other issues with culturing bacteria

As discussed in the designing of the enrichment condition experiments, it was more complicated when attempting to enrich for species with the ability to synthesise DMSP, rather than the ability to use it as a carbon source. This is because the synthesis of DMSP may not necessarily convey enough of a survival advantage to change the community. For this reason, multiple conditions were used in combination, and the levels of DMSP production and percentage of DMSP-producing isolates increased, suggesting that the enrichment was at least partially successful. Clearly these enrichments are not representative of the natural conditions in Stiffkey, but the reason that they were performed was to identify and isolate DMSP-producers, so this is less of a concern. One condition that should perhaps be changed is the addition of Met, instead of MTHB to the mix, as when MTHB was chosen as the intermediate, it was assumed that bacteria were only utilising the transamination pathway. However, it is possible that if Met was added, it would enrich for too many other processes, not just DMSP production.

Another issue that arose from altering these growth conditions to optimise for DMSP-producing isolates is that the conditions are now no longer like the natural environment, limiting the claims that can be made about the species community. This could also be said about the pure cultures of isolates from which DMSP quantification measurements were taken, as they are grown in higher temperatures and with greater access to nutrients than perhaps might be the case *in situ.* This is why the seawater incubation experiment was also included, in order to confirm that DMSP is produced even under *in situ* conditions.

So far this work has only looked at one species at a time, and not how they interact or what the overall community make-up is. Although we can test isolates as close to natural environment as possible, there is always lab bias, so we need to perform more tests under environmental conditions, and look into culture-independent analysis such as community sequencing and metatranscriptomes to make further claims about the function or activity of DMSP producers in the actual environment.

Every day more modern analytical techniques are being developed, including ones that study community DNA, such as metagenomic sequencing and even metatranscriptomics and RT-qPCR. However, there is still a place for the more simple techniques – they need to be used in conjunction with each other. While it would be false to claim that culture-dependent experiments are an entirely accurate measurement of the actual community, they do have a place in the study of an environment. Plate isolations and characterisation are an excellent complement to culture-independent work, as it allows for the practical testing of theoretical conclusions drawn from sequencing analysis.

4.3.4 Concluding Statements

We have demonstrated that it is possible to isolate DMSP-producing bacteria from Stiffkey, and that the proportion of those bacteria within the sediment can be increased through enrichment experiments using specially designed media conditions. Several of these species were the first of that particular genus to be shown to have the ability to produce DMSP, and it was clear that many were not producing DMSP in the same way that *L. aggregata* has previously been observed to. Many of them did not appear to possess *dsyB*, and some were even confirmed not to. Furthermore, their DMSP synthesis levels were not affected with the addition of pathway intermediates from the transamination pathway, as was observed in *L. aggregata*. All this is evidence for the existence of a novel DMSP-producing gene, and possibly the use of a different pathway of production. There is undoubtedly a more interesting, complex story to be told, and bacterial DMSP production is likely not only more widespread, but also more varied than previously thought.

CHAPTER 5

CHARACTERISATION OF DMSP-PRODUCERS LACKING *dsyB*

5 DMSP-PRODUCERS LACKING *dsyB*

5.1 Introduction

5.1.1 Gene discovery

By far the most interesting discovery of the previous chapter was that several bacterial species shown to produce DMSP lacked the only known DMSP-synthesis gene, *dsyB* (Curson et al. 2017). There are two possible explanations for this: i) the existence of multiple DMSP-synthesis isoform genes, which perform same enzymatic function (MTHB methyltransferase) as DsyB in the organisms that lack *dsyB*; or ii) the *dsyB*$^{-}$ microorganisms contain a novel DMSP synthesis pathway, using a whole different and unknown suite of genes.

There is some precedent for both these options, as seen in other species of bacteria that play a role in DMSP metabolism, namely the existence of eight different *ddd* (DMSP-dependent DMS) genes (Curson, Todd, et al. 2011). Most of these encode for proteins that perform almost the exactly the same role, the lysis of DMSP to Acrylate (Alma1, DddK, DddL, DddP, DddQ, DddW and DddY), but several of them are completely different protein families. DddK, DddL, DddQ and DddW are all small proteins with cupin (barrel-shaped) pockets at the C-terminal that bind to transition metals like Fe and Zn (Todd et al. 2011; Johnston et al. 2016). Even though the rest of their structure is markedly different enough to be classed as separate proteins, they all perform the same function (Todd et al. 2012). In contrast, despite also carrying out the lysis of DMSP to Acrylate and DMS, the DddP protein is instead from the M24B metalloprotease family, a larger polypeptide with a 'pitta bread' fold around two active sites (Hehemann et al. 2014). DddD is an example of a different pathway used by particular bacteria for producing DMS from DMSP (Todd et al. 2007). This DddD enzyme, a Class III acetyl CoA-transferase which lyses DMSP through a different route, directly producing 3-HP through the transfer of a CoA molecule, and the likely formation of a 3-HP-CoA intermediate (Alcolombri et al. 2014).

Indeed, even between prokaryotes and eukaryotes there are proteins that carry out the same role in DMSP metabolism – Alma1, a tetrameric protein from the aspartate/glutamate racemase super-family (Alcolombri et al. 2015), is the first eukaryotic DMSP lyase discovered, totally different in structure and family to the previously described Ddd proteins, yet still carrying out the same process of lysis, producing acrylate and DMS from DMSP. It likely carries this out through a different method, not cleaving the C-S bond as seen in the other cupins and metalloproteases, but instead through the abstraction of

a proton close to the DMSP carboxylate, releasing DMS and leaving acrylate (Johnston et al. 2016).

There are several methods by which novel genes can be discovered. For eukaryotes it is more complex than prokaryotes, mainly due to the presence of exons making the genomes significantly larger and more difficult to screen. In order to identify *alma1* biochemical fractionation was combined with shotgun proteomics, where DMSP lyase activity was identified in a ~100kDa protein in the membrane fraction of the chloroplast (Alcolombri et al. 2015). This was further visualised using shotgun liquid chromatography–tandem mass spectrometry–based proteomics analysis in combination with peptide libraries constructed from RNA sequencing. RNA was also utilised in the discovery of *dddW,* which was identified through the use of microarrays that showed greatly enhanced gene expression in a particular gene when *Ruegeria pomeroyi* DSS-3 cells were grown in the presence of DMSP (Todd et al. 2012). This was identified and tested and confirmed to be a DMSP lyase, termed *dddW.*

Although some of these genes were discovered using more unusual techniques, the vast majority of DMSP synthesis and catabolic genes were discovered using genomic libraries and screening for function in suitable heterologous hosts. This involves the creation of a cosmid (or fosmid) library (25-35 kb fragments of contiguous DNA randomly cloned into a vector), screening that library for the phenotype of interest (e.g. DMSP lyase activity) in a suitable heterologous host, sequencing the positive clone fragments (more specifically the cloned genes), and finally the identification of candidate genes within that fragment through the use of bioinformatics combined with sub-cloning and further screening for activity. The DMSP catabolism genes identified in this way were *dddD* (Todd et al. 2007)*, dddL* (Curson et al. 2008)*, dddP* (Todd et al. 2009)*, dddQ* (Todd et al. 2011)*, dddY* (Curson, et al. 2011)*.* Furthermore, *mddA,* the gene recently discovered in *Pseudomonas deceptionensis* M1T that produces DMS from the methylation of MeSH rather than from DMSP (Carrión et al. 2015) was also discovered through this method, as was the first DMSP-synthesis gene, *dsyB* (Curson et al. 2017)*.* This method is especially applicable where the function (e.g. DMSP production or catabolism) has been observed by a particular strain, but, unlike in the discovery of *dddK* (Sun et al. 2016) and DSYB (Curson et al. 2018)*,* there are as yet no candidate genes that can be individually cloned and tested. As this was the case for *Novosphingobium,* it was therefore decided that this method was the most well-suited for searching for the novel DMSP-synthesis gene, as we knew that it was producing DMSP and did not contain *dsyB*, but we did not have any candidates for other potential genes.

5.1.2 The methylation pathway for DMSP production

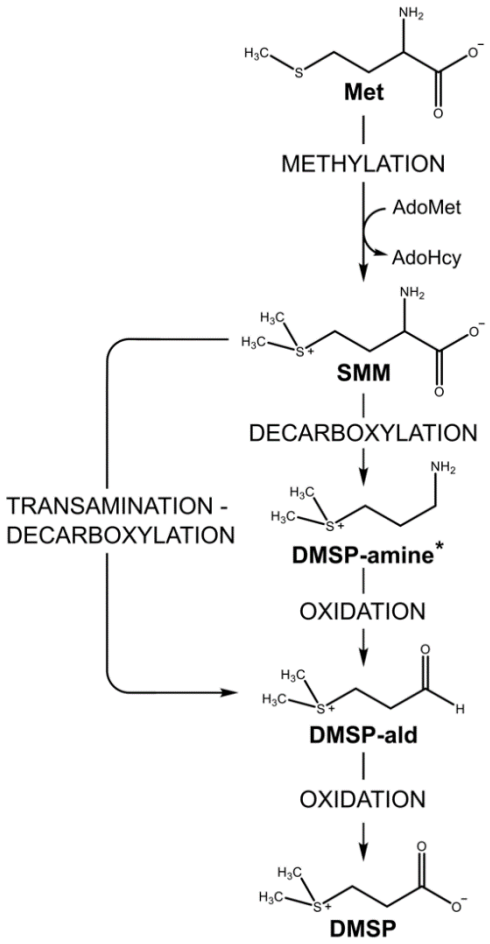

Figure 5-1: The methylation pathway, originally observed in the higher plants *Wollastonia biflora* and *Spartina alterniflora,* by which DMSP is produced. One of two routes can be taken through the pathway after methylation of methionine to SMM, with one forming DMSP-amine through the decarboxylation of SMM, and the other producing an as-yet unidentified intermediate, before both become DMSP-aldehyde.

From the precedent set by the *ddd* genes, it would be unsurprising if *Novosphingobium* or the other strains that appear to lack *dsyB* in fact contain novel DMSP-producing genes and/or pathways. Indeed, from the incubation experiments in **Chapter 4**, it seemed very likely that the transamination pathway is not used in *Novosphingobium*, as DMSP production did not increase with any of the intermediates of this pathway. The most likely pathway to look into would be the methylation pathway in higher plants **(Figure 5-1)**, as there has only been one demonstrated example of the decarboxylation pathway as yet (Kitaguchi et al. 1999), and *Novosphingobium* was isolated from a salt marsh area known to contain large amounts of *Spartina*. Of course, we also cannot entirely rule out the possibility that there is a novel DMSP production pathway

If this is indeed the case it will be an interesting discovery, as the transamination pathway is thought to be the major pathway utilised in the marine environment, with marine algae, diatoms and most dinoflagellates using it. If bacteria isolated from salt marsh sediments are using the methylation pathway, it is likely that they will also be in marine sediments, which are much more abundant than just salt marshes where *Spartina* grow.

As described in **Chapter 1**, this methylation pathway **(Figure 5-1)** is a split pathway, with two different routes being used by either Gramineae (*Spartina* and *Saccharum*) or Compositae (*Wollastonia*). The pathway diverges after the first step, which is a methylation of Met to SMM (S-methylmethionine), and involves either a decarboxylation to the stable intermediate DMSP-amine (Kocsis et al. 1998) followed by oxidation to DMSP-aldehyde, or a transamination-decarboxylation reaction that takes SMM through an unstable intermediate to DMSP-aldehyde (Rhodes et al. 1997). Both pathways result in the synthesis of DMSP-aldehyde, which is in turn oxidised to DMSP **(Figure 5-1)**. The first and last steps in the pathway are shared between Gramineae and Compositae, and can also be found in many other species, with SMM formation occurring in most angiosperms (Mudd & Datko 1990), and although it may not be produced in other species, DMSP-aldehyde is able to be oxidised by dehydrogenases found in several non-DMSP-producing species (Trossat et al. 1997).

The methylation of Met to SMM is catalysed by a Met S-methyltransferase that transfers a methyl group from the co-substrate S-AdoMet (S-Adenosyl Methionine) (James et al. 1995). Indeed, Hanson et al. (1994) demonstrated in *Wollastonia biflora* that SMM produced in this manner was the first intermediate created in the process of producing DMSP from Met, and the enzyme has been found in multiple angiosperms (Mudd & Datko 1990), including *Spartina alterniflora* (Kocsis et al. 1998). The methylation of Met is thought to take place in the cytosol, rather than in the chloroplasts (Trossat et al. 1996) where the conversion to DMSP-aldehyde takes place. There are several plants that contain the *MMT* gene but have not been shown to produce DMSP, including *Sorghum bicolor* and *Zea Mays* (both of which are angiosperms), suggesting that SMM production is important in plants for reasons other than to produce SMM as an intermediate in DMSP production. Indeed, SMM production is part of its own small cycle, where it is synthesised from Met (MMT activity) and then used as a methyl donor for Met synthesis by homocysteine (HMT) **(Figure 5-2)**, which results in two Met molecules (one left after the methyl group is removed, and one formed with the donation of a methyl to Hcy) (Ranocha et al. 2000). It is thought that SMM is also produced as a storage molecule for Met, playing a role in Met regulation (Stefels 2000).

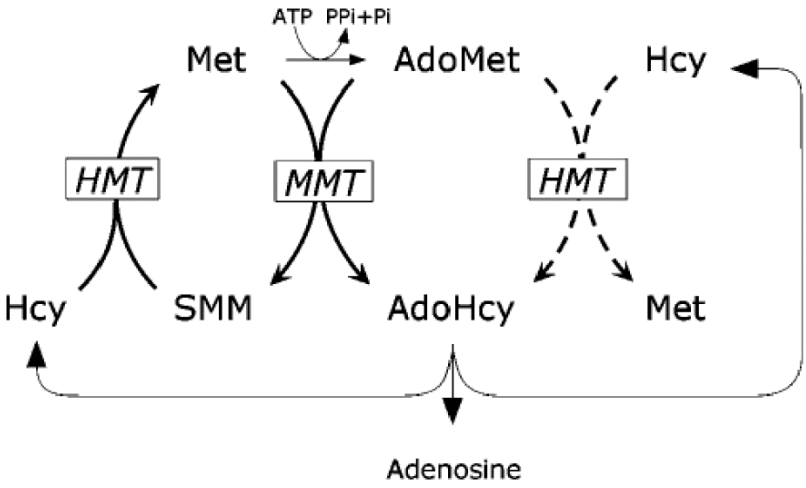

Figure 5-2: The SMM cycle (and some related reactions) in higher plants. Bold lines indicate the core reactions, where SMM is produced from Met (via AdoMet) (**MMT**)and then donates a methyl group to Hcy to produce two Met molecules (**HMT**). The dotted line is a shorter route where AdoMet donates a methyl group to Hcy, producing one Met molecule (**HMT**). Figure from Ranocha et al. (2000).

Interestingly, many species of bacteria as well as yeast (Rouillon et al. 1999) have been known to use HMT to retrieve Met from SMM that is transported into the cell, allowing them to bypass the full Met synthase pathway (Ranocha et al. 2000). However, MMT activity, and the ability to produce SMM from Met, has until now only been attributed to plants (Ranocha et al. 2000). The discovery that it might exist in bacteria, and play a role in bacterial DMSP production, is noteworthy.

5.1.3 Chapter Aims

It was clear from the culture-dependent work in **Chapter 4** that there are several species that possess the ability to synthesise DMSP, but which lack the only known DMSP synthesis gene, *dsyB*. An isolate of *Novosphingobium sp.* MBES04 was isolated from Stiffkey salt marsh, and shown to produce much larger amounts of DMSP than those produced by many other marine bacteria, including the previous model organism *L. aggregata*. In **Chapter 4** it was demonstrated through WGS of the isolate that no *dsyB* homolog existed in the genome, and preliminary induction experiments suggested that DMSP production by *Novosphingobium,* while stimulated by the addition of Met, was largely unaffected by the other intermediates in the transamination pathway (see below).

We hypothesise that *Novosphingobium* contains a novel DMSP-producing gene or gene cluster and is able to synthesise this osmolyte without *dsyB*, likely through a different pathway instead of the transamination pathway. This gene will be identified, sequenced and disrupted within a DMSP-containing species, and the mutant characterised.

5.2 Methods and Results

5.2.1 *Novosphingobium* growth experiments

Before more in-depth experiments could be performed on *Novosphingobium* it was necessary to characterise the growth patterns and conditions under which DMSP production appears to increase the most, in order to better design the following experiments. The first experiment performed was a growth curve, set up using triplicate flasks of MBM. Samples were incubated at 30°C with shaking at 180 rpm, and the OD_{600} reading was measured every hour, until the cultures reached stationary phase **(Figure 5-3)**.

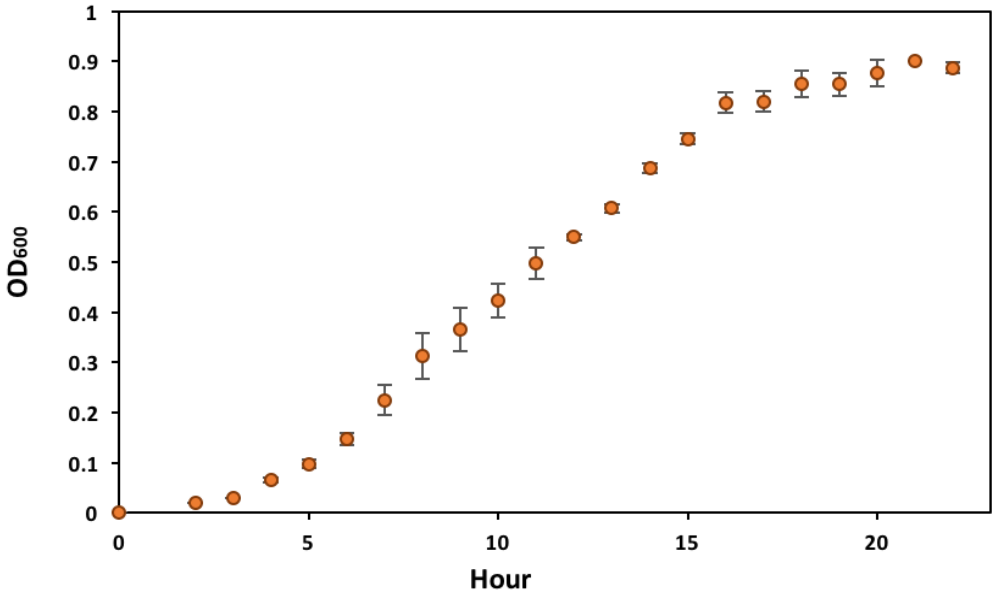

Figure 5-3: Growth curve in triplicate of cultures of the isolate of *Novosphingobium sp.* MBES04, incubated at 30°C, shaken at 180 rpm for 22 hours until cultures reached stationary phase. Error bars display standard error.

Another growth experiment that was carried out on this isolate was measuring the effect of different environmental conditions on the production of DMSP by *Novosphingobium*. From a starter culture of *Novosphingobium* in standard media, flasks were inoculated in triplicate into either standard media, media of salinities varying from 50 PSU to 5 PSU, or media with low nitrogen. All were incubated at 30°C overnight with shaking at 180 rpm, with the exception of one of the standard media cultures, which was incubated at 16°C instead. Measurements of protein content and DMSP concentration were taken and compared **(Figure 5-4)**.

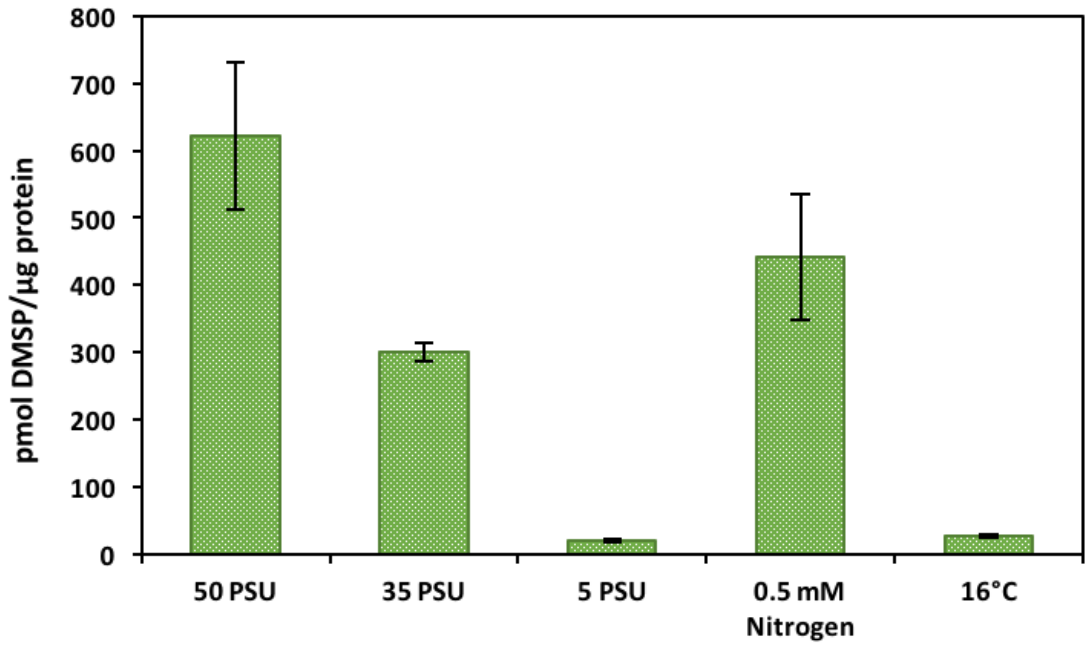

Figure 5-4: The effect of changing salinity, nitrogen availability or temperature on the production of DMSP by *Novosphingobium sp.* MBES04, after incubation overnight. Samples are in triplicate, and error bars display standard error.

This was a more refined version of the work carried out in **Chapter 4**, with more salinity conditions tested, as well as lowered temperature. Compared to that study, we see a similar pattern between 5 PSU and 35 PSU, although there is less of a significant difference in DMSP levels between high and lower nitrogen levels than shown before. It appears that increased salinity does increase DMSP production, as is often observed, lending weight to the theory of DMSP as an osmoprotectant. Lower temperatures seemed to decrease the concentration of DMSP quite dramatically compared to the same media conditions (35 PSU, 12 mM nitrogen) at 30°C, suggesting that, in *Novosphingobium* at least, it may not play a role in protection against lowered temperatures.

5.2.2 *Novosphingobium* intermediate incubation experiment

The growth curve in **Figure 5-3** allowed us to work out the time taken to reach an OD_{600} of 0.5, which was required for the intermediate induction experiment. This experiment was designed specifically to test the DMSP production of the *Novosphingobium* isolate when incubated with intermediates from all three of the pathways of production, described in **Chapter 1**. Its purpose was to identify potential intermediates of the DMSP synthesis pathway used by *Novosphingobium*. The intermediates tested in this experiment were Met, which was expected to increase DMSP production, then the intermediates in the transamination pathway, MTOB, MTHB and DMSHB, as well as MMPA and MTPA which are a part of the decarboxylation pathway,

and SMM and DMSP-amine, which are two of the intermediates from the methylation pathway **(Figure 5-1)**. Increased levels of DMSP production when incubated with a particular set of these intermediates would be suggestive of the use of that pathway by the isolate, as they would be able to take any of the intermediates through the pathway to produce DMSP.

A starter culture of *Novosphingobium* was set up overnight, and cultures were incubated for 12 hours, reaching an OD_{600} of ~ 0.5. The DMSP levels of each were quantified and then separated into 5 ml aliquots, then mixed with 0.5 mM of each of the intermediates individually, including a control mix with nothing else added. These mixed cultures were then incubated at 30°C with shaking, and the DMSP concentration and protein content were measured at 30, 60, 120 and 240 minutes **(Figure 5-5)**.

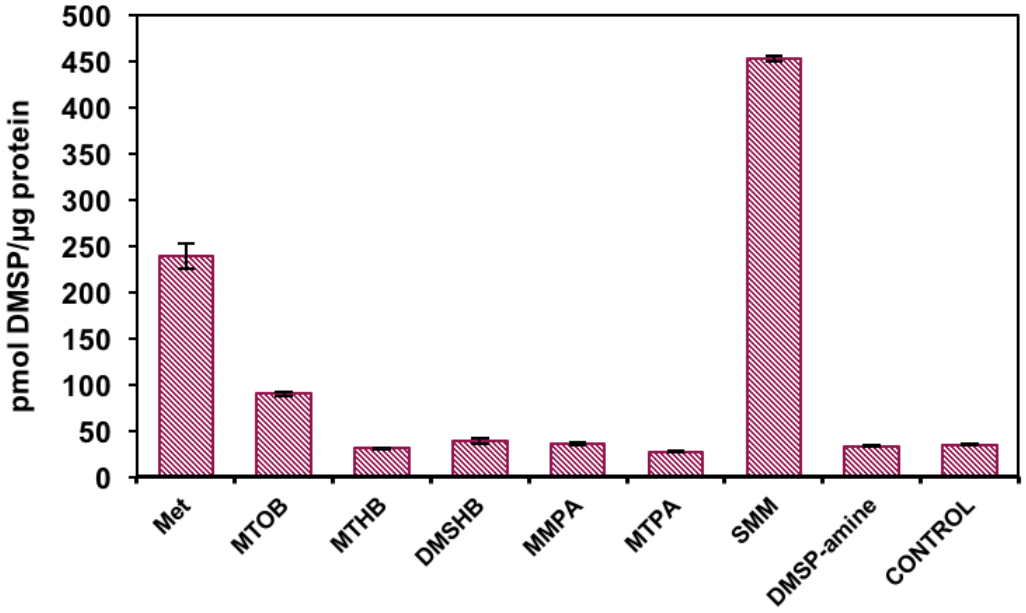

Figure 5-5: The effect on DMSP production by *Novosphingobium* when mixed with selected intermediates from the three pathways of DMSP production, after incubation with the substrates for 4 hours at 30°C. Bars display standard error.

This experiment showed very clearly that *Novosphingobium* produces significantly higher levels of DMSP when incubated with SMM than with any other intermediates, although Met also caused an increase in production. These results suggest that this isolate utilises the methylation pathway used by higher plants, and not the transamination pathway. An unusual result is that of the lack of increase in DMSP when *Novosphingobium* was incubated with DMSP-amine, which is also in the methylation pathway, and would be expected to cause an increase as well. However, this does not negate the previous

conclusion, as the methylation pathway takes two routes, one of which bypasses the formation of DMSP-amine and instead goes from SMM to DMSP-aldehyde, via an unidentified intermediate **(Figure 5-1)**. Unfortunately, DMSP-aldehyde was not tested in this experiment because it is unstable and not commercially available.

5.2.3 Demonstration of MMT activity in *Novosphingobium*

To test the hypothesis that *Novosphingobium* has the ability to methylate Met to SMM, we designed a cell lysate assay for detecting MMT activity. The focal point of this assay required the ability to distinguish Met from SMM, and the simplest way to do so was to use the fact that SMM can liberate DMS upon alkaline hydrolysis and incubation at 80°C, whereas the substrate Met does not, as it does not contain a DMS moiety. DMS can easily be detected by GC (see **Chapter 2**). A complication with using this method of detection was that MMT activity on Met requires a methyl donor in the form of S-AdoMet, which does contain a DMS moiety, and also liberates it when treated with NaOH and heating. This would make the assay ineffective, as peaks of DMS from the production of SMM would be completely masked by peaks of the S-AdoMet added to the mix. A solution to this problem arose from the fact that S-AdoMet can be sequestered from solution by the addition of activated charcoal (Cook & Wagner 1984), due to its nucleotide base region. To test that S-AdoMet would be adsorbed onto the charcoal but SMM would be left in the media, charcoal sequestration experiments were set up using mixes of S-AdoMet, SMM, or both, in sterile water. Samples were then tested for DMS production before and after treatment the samples with activated charcoal, as described in **Chapter 2**, with the DMS released in both instances measured by GC and compared between samples **(Figure 5-6)**.

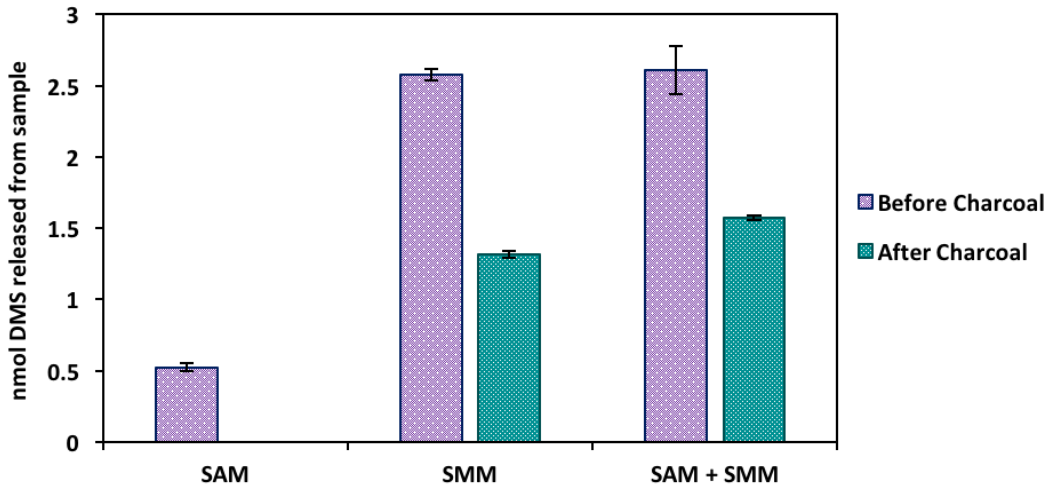

Figure 5-6: DMS released by samples containing either S-AdoMet, SMM or both, before and after treatment with activated charcoal to sequester S-AdoMet.

S-AdoMet was completely removed from the samples when tested on its own, and although there appears to be a decrease in the amount of DMS released by SMM when treated with activated charcoal, it is still detectable and has therefore not been removed along with the S-AdoMet. We can therefore be confident that the S-AdoMet is completely adsorbed, and therefore treatment with charcoal is a useful technique for removing S-AdoMet after the cell lysate assay has been performed, whilst leaving SMM to be quantified by DMS production after alkaline hydrolysis. It should be noted that the levels of SMM synthesised by the samples are likely to be higher than those actually measured by GC, as the charcoal was observed to possibly be removing a portion of it from the solution. Keeping this in mind, we now have a workable assay for the detection of SMM, which was ready to be tested on *Novosphingobium* cell extracts.

To test whether *Novosphingobium* gene products or cell lysate had SAM-dependent Met methyltransferase activity, cell lysate was created (with the help of Ben Pinchbeck). *Novosphingobium* cultures were harvested and resuspended in 50 mM Tris-HCl buffer. Cells were lysed by sonication and centrifuged to pellet debris. The lysate was dialysed at 4°C overnight to remove any pre-existing metabolites. This lysate was mixed with either 1 mM S-AdoMet, 1 mM Met, or both, and then incubated for 30 minutes at room temperature, allowing cell extract activity to take place, before using activated charcoal to stop the reaction by sequestering the S-AdoMet and removing it.

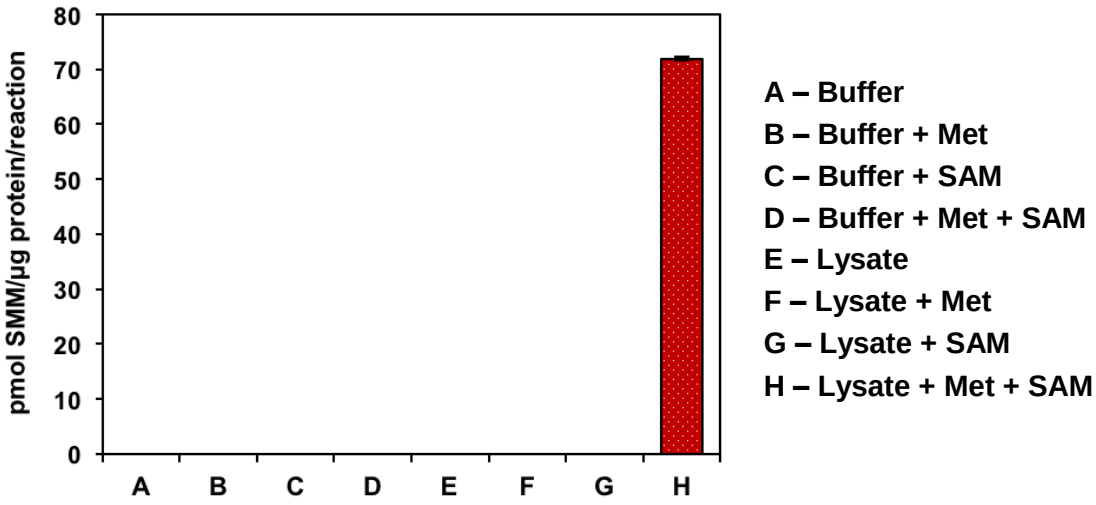

Figure 5-7: SMM production of purified and dialysed cell lysate containing the SAM-MMT protein extracted from *Novosphingobium sp*. MBES04, in duplicate. The buffer was also tested as a negative control, as well as each of the individual compounds, plus a combination of the two. SAM = S-AdoMet. Error bars display standard error.

The MMT activity was measured in the samples as detailed in **Chapter 2 (Figure 5-7)**. As can be clearly seen, MMT activity was only detectable in the *Novosphingobium* lysate when incubated with both Met and S-AdoMet, and not in any of the other controls. Therefore, the *Novosphingobium* lysate has MMT activity. It is possible that the *Novosphingobium* lysate may have further transformed SMM into other intermediates, even potentially all the way to DMSP, but it is unlikely because the downstream enzymes in that pathway likely require PLP, NADPH or other metabolites which would have been removed from the lysates during the dialysis step.

5.2.4 Gene library construction and screening

Given the results above, it seems likely that *Novosphingobium* contains a novel gene encoding for MMT activity (transforming Met to SMM). We have also established that we can screen for the production of SMM by *Novosphingobium* in media by using GC to assay the production of DMS liberated from SMM by the addition of NaOH and heating to 80°C. With this screening method in place, the next logical step was to attempt to identify the novel Met methyltransferase gene/s in this bacterium, using the genomic library and heterologous host approach, screening for the production of DMS from SMM after the addition of Met. A genomic library of the strain was constructed in the cosmid pLAFR3 **(Figure 5-8)** (with the help of Andrew Curson). The genomic library was comprised of fragments of ~25 – 40 kb of the genome cloned into the *Eco*RI site of pLAFR3. To do this, *Novosphingobium* genomic DNA was extracted and partially digested using *Eco*RI, then ligated into the pre-digested vector pLAFR3 **(Figure 5-8)** (Staskawicz et al. 1987) to form concatemers.

These concatomers were packaged using the Stratagene Gigapack III XL Packaging mix, and transfected into *E. coli* 803. The library of clones was then mobilised into *R. leguminosarum* J391 via triparental mating and dilutions were plated on TY for single colonies selecting for tetracycline resistance (pLAFR3). These colonies were individually picked and inoculated in RM media (a minimal *Rhizobium* medium to increase the likelihood of SMM production and to avoid background noise from a rich medium such as TY, which contains yeast extract). Inoculums were mixed with 5 mg/µl tetracyline and 0.5 mM Met, then incubated overnight at 30°C. Colonies were checked individually for the ability to produce SMM, an ability that *R. leguminosarum* J391 does not naturally possess.

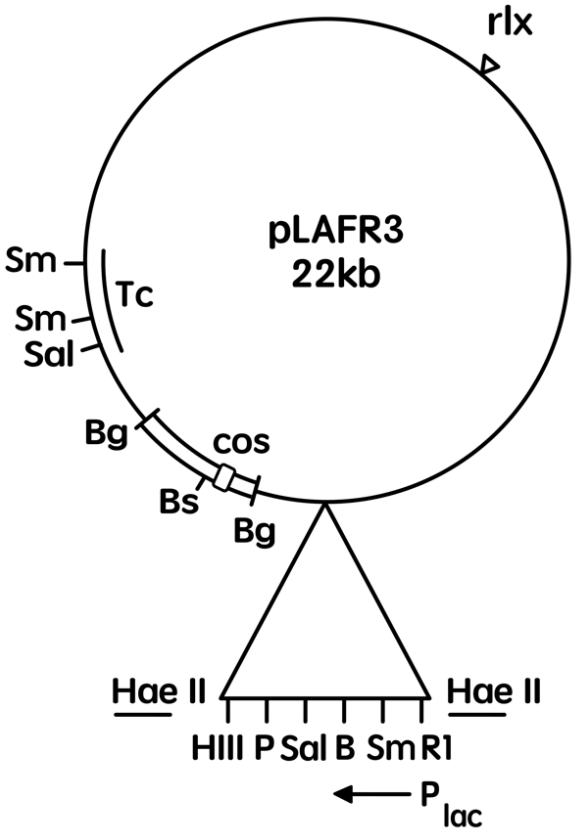

Figure 5-8: The wide-host-range cosmid cloning vector pLAFR3, with restriction digest sites for multiple enzymes, including *Hind*III, *Pst*I, *Sal*I, *Bam*HI, *Sma*I and *Eco*RI. For this study, *Eco*RI was used to construct the clone library of partially digested genomic DNA. Figure was adapted from Staskawicz et al. (1987).

750 cosmids were screened with help from Kasha Sweet, an undergraduate student, and two were found to be positive. This meant that they produced a peak at a retention time that indicates the presence of DMS, showing a peak area that suggested the production of ~ 20.7 nmol/ml of DMS in the headspace, through the alkaline hydrolysis of either DMSP or SMM, while the negative controls and clones that did not contain the fragment including the gene showed no peaks at all. The positive clones were re-inoculated and measured on the GC a second time to confirm that a peak of DMS was indeed produced (from SMM or DMSP). Both cosmids, termed pBIO0438 and pBIO0762, were confirmed to be positive even after extracting the plasmid and cloning back into *R. leguminosarum*. Furthermore, in line with Koch postulates, the pBIO0438 and pBIO0762 plasmids were extracted, transformed into *E. coli* 803 and then mobilised back into *R. leguminosarum* J391, before re-confirming their MMT activity. From this point on we were sure that we had cloned a gene/s that confers MMT activity.

The cosmids containing the positive fragments were then extracted from newly-inoculated cultures, using the phenol-chloroform extraction method. Several restriction digests were set up using the cosmids and four different enzymes – *Eco*RI, *Bam*HI, *Hind*III and *Pst*I. This was to demonstrate the presence of inserted fragments in the pLAFR3 cosmid, and to observe how similar both fragments were **(Figure 5-9)**.

From the restriction digests performed in **Figure 5-9**, it is very clear that pBIO0438 and pBIO0762 contain overlapping DNA, as for example they both contain multiple identical *Eco*RI fragments cloned. These include a large band at ~12 kbp in size, two bands either side of the 5 kb marker, and a faint one at the 2 kb marker, as indicated on the gel. The large band and smear seen in the *Pst*I digest of pBIO0438 suggests that there was no more than one *Pst*I restriction site in the fragment, whereas there was at least two in pBIO0762 as one band can be seen (just below 2 kbp in size) that was separated from the larger fragment. The fact that pBIO0438 and pBIO0762 are not identical, with extra cloned fragments appearing in the pBIO0762 plasmid, likely indicates that the clones likely contain extra *Eco*RI fragments at either end around the essential section that holds the gene or genes conferring MMT activity.

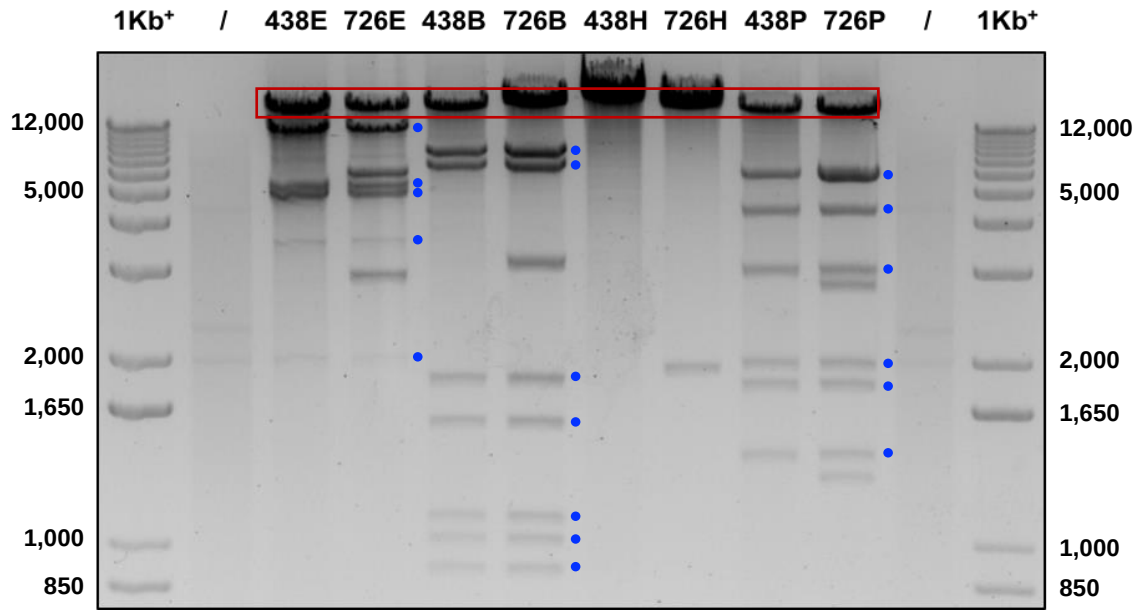

Figure 5-9: Gel electrophoresis showing restriction digests of the two positive *Novosphingobium* genome fragments, pBIO0438 and pBIO0726. Digests were performed using the enzymes *Eco*RI (E), *Bam*HI(B), *Hind*III(H) and *Pst*I (P). The largest band (red box) is likely the 22 kb linearised pLAFR3 cosmid. Fragments in pBIO0726 that are shared with pBIO0438 are indicated by a blue dot.

5.2.5 Candidate DMSP-synthesis gene identification and characterisation

Once it was confirmed that pBIO0438 and pBIO0726 both contain the gene/s involved in the production of SMM, and both appear to be from the same or similar section of the genome (demonstrated by the similarity of band size after the restriction digests, **Figure 5-9**), the next step was to determine which section of the genome was encoded for in these fragments, and what genes exist in that area. To identify the nucleotide sequence cloned in pBIO0438 and pBIO0726, the termini of the cloned DNA were sequenced at Eurofins Genomics, using primers designed to the polylinker in pLAFR3 (M13 uni (-43) and M13 rev (-29)). Sequence identities for ~500 base pairs from the beginning and end of both the fragments were obtained**.** These sequences were then searched for in the whole genome sequence that was obtained in **Chapter 4**, enabling either end of the fragment to be aligned against the known sequence, consequently revealing the sequence of the section between the two ends, which is the fragment cloned into pLAFR3 **(Figure 5-10).** The results clearly confirmed what we had established above, that these clones do indeed contain overlapping fragments.

From the alignment to the sequenced genomic DNA for *Novosphingobium sp*. MBES04 it was possible to calculate the sizes of both the fragments, with pBIO0438 found to be 21.8 kb in size, and pBIO0726 at a larger 30.9 kb. Both are clearly covering the same section of the genome. By using WGS on the *Novosphingobium* isolate it produced the sequence of the bacterial genome, and using RAST meant that this sequence was then annotated, with the predicted coding sequences highlighted as light blue boxes in **Figure 5-10**. These were then analysed to determine if any were likely candidates for playing a role in DMSP production. All the amino acid sequences of the genes within the overlap of the two partially digested fragments were identified using BLASTp **(Table 5-1)**.

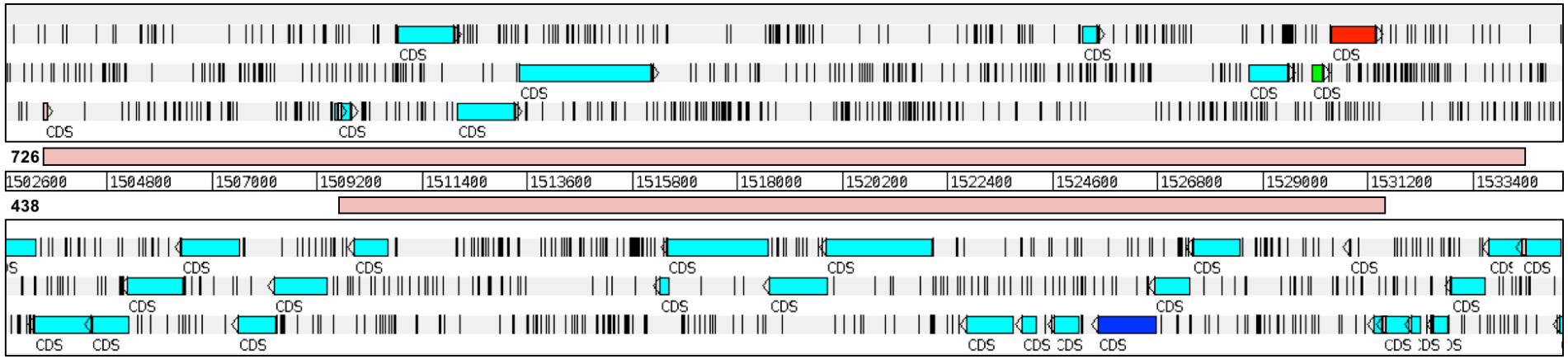

Figure 5-10: An Artemis genome browser display showing the two positive fragments of *Novosphingobium* genomic DNA from the gene library created by the partial digestion of the *Novosphingobium* genome, 438 and 726 (pink bars). These are aligned against the whole genome sequence obtained from MicrobesNG, and from this the sizes of both fragments could be calculated, and the genes encoded within the fragment identified. The gene highlighted in red is annotated as a methionine S-methyltransferase, the gene in dark blue is annotated as an aspartate aminotransferase, and the gene highlighted in green is a hypothetical protein. All other light blue boxes indicate other coding sequences.

The analysis of the genes in this fragment revealed a gene, termed *mmtN* by us, encoding a SAM-dependent methyltransferase (SAM-MMT), specifically belonging to the family of Met S-methyltransferases which require the methyl donor molecule S-AdoMet for function (pFAM; *Methyltrans_SAM* (PF10672)). The *mmtN* gene product is similar to the plant MMT enzyme (E value 2E-18, identity 28%). However, this similarity is only over the N-terminal domain of the plant MMT **(Figure 5-11)**, as MmtN is only 307 amino acids in size, and thus is ~3 times smaller than the MMT of *Arabidopsis thaliana* which is a 1,071 amino acid enzyme **(Figure 5-11)**. The C-terminal domain of MMT that is missing in MmtN contains a conserved PLP binding site of an aminotransferase, which is proposed to be involved in the regulation of MMT in plants, and therefore is not related to or necessary for MMT enzyme function (Bourgis et al. 1999). This is consistent with MmtN lacking this domain, yet still appearing to be functional in methylating methionine. MMT has long been recognized to catalyse the SAM-dependent methylation of Met to generate SMM in plants (Green & Davis, 1960, Ranocha et al. 2000). Given the similarity between these two proteins, MmtN was a strong candidate gene for catalysing the first step of DMSP synthesis in *Novosphingobium,* and perhaps other bacteria.

Table 5-1: The annotation and BLASTp analysis of genes within a specified fragment of the *Novosphingobium* genome. Genes are from both the forward (+) and reverse (-) strands.

Gene location (bp) 1509624 – 1531541	RAST Annotation	BLASTp Identity	E value	Identity (%)
1509579..1509908 (+)	L-fucose mutarotase, type 2	L-rhamnose mutarose	8E-63	90
1509970..1510656 (-)	Transcriptional regulator, GntR family	Transcriptional regulator, FadR family	8E-143	96
1510882..1512051 (+)	Muconate cycloisomerase	Mandelate racemase/ muconate lactonizing enzyme	0.0	93
1512135..1513322 (+)	Major facilitator superfamily MFS_1	MFS transporter	0.0	94
1513424..1516180 (+)	N-acetylglucosamine-regulated TonB-dependent outer membrane receptor	TonB-dependent receptor	0.0	95
1516370..1516543 (-)	Hypothetical small protein yjiX	DUF466 domain-containing protein	6E-27	81
1516540..1518603 (-)	Carbon starvation protein A	Carbon starvation protein A	0.0	97
1518662..1519849 (-)	Putative iron-regulated membrane protein	PepSY domain-containing protein	0.0	94
1519864..1522038 (-)	Ferrichrome-iron receptor	TonB-dependent receptor	0.0	98
1522809..1523741 (-)	Hypothetical protein	Hypothetical protein	-	-
1523952..1524233 (-)	Purple acid phosphatase	Metallophosphoesterase family protein	2E-43	84
1524642..1525136 (-)	Transcriptional regulator, MarR family	Transcriptional regulator, MarR family	4E-103	90
1525222..1525530 (+)	Hypothetical protein	DUF3861 family protein	7E-59	94
1525557..1526756 (-)	Aspartate aminotransferase	Pyridoxal phosphate-dependent aminotransferase	0.0	98
1526738..1527448 (-)	Transcriptional regulator, TetR family	Transcriptional regulator, TetR family	2E-148	93
1527538..1528494 (-)	D-3-phosphoglycerate dehydrogenase	Hydroxyacid dehydrogenase	0.0	96
1528718..1529536 (+)	Ribulose-5-phosphate 4-epimerase and related aldolases	Class II aldolase/adducin family protein	0.0	94
1530026..1530241 (+)	Hypothetical protein	Hypothetical protein	-	-
1530445..1531368 (+)	**Methionine S-methyltransferase**	**SAM-dependent methyltransferase**	**0.0**	**95**

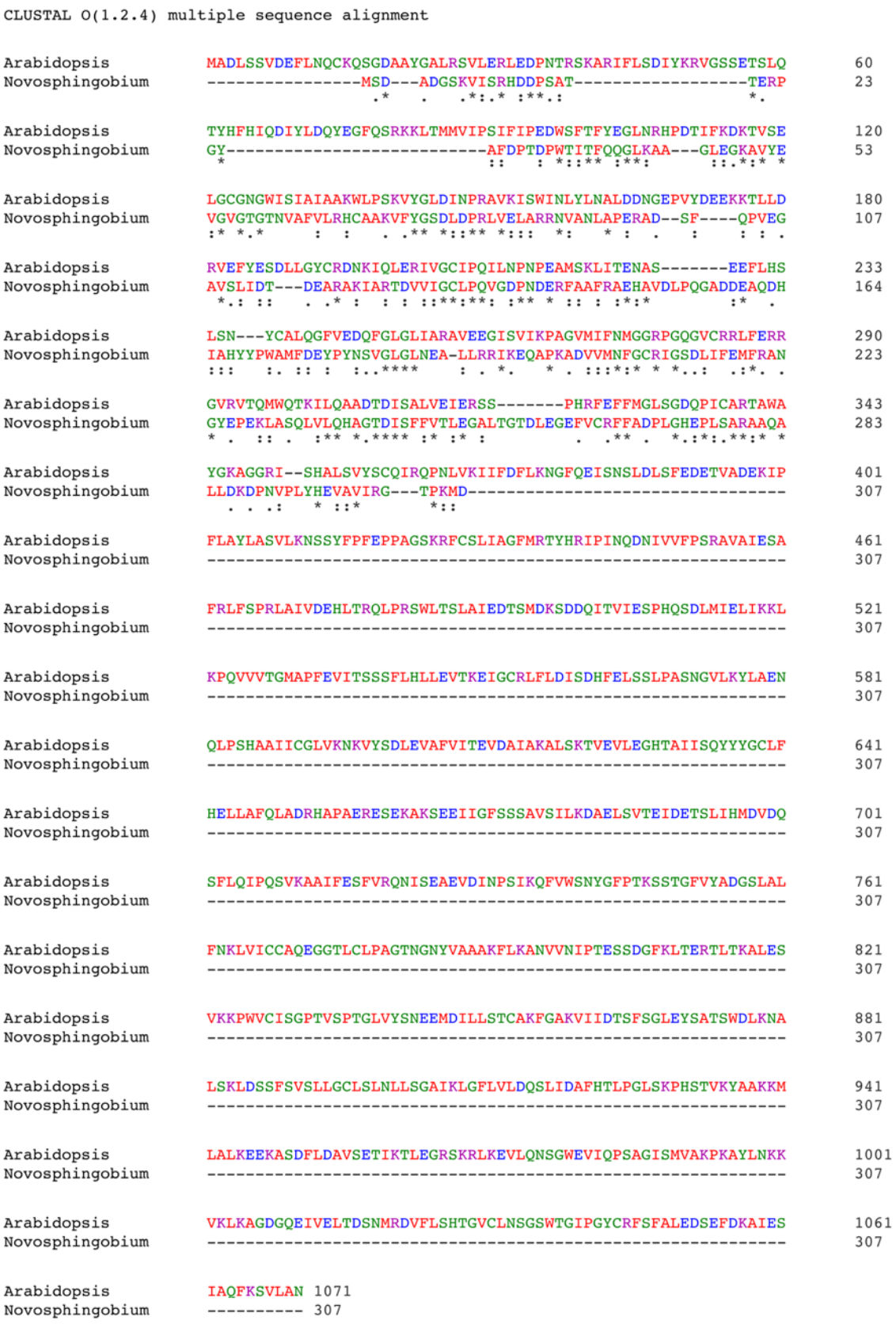

Figure 5-11: Clustal Omega (1.2.4) multiple sequence alignment of the amino acid sequences of the MmtN enzyme in *Novosphingobium sp.* MBES04, compared to the SAM-dependent methyltransferase in *A. thaliana*.

The MmtN sequence was then used as a probe against NCBI and JGI databases using BLASTp, and was found to be in 22 Alphaproteobacteria, 4 Actinobacteria and one Gammaproteobacterium at > 50 % identity, from a variety of genera, all of which appear to be marine in origin **(Figure 5-12)**. Furthermore, these included species of *Labrenzia* and *Sagittula,* both of which are already known to produce DMSP. If *mmtN* was indeed the gene able to encode a DMSP synthesis enzyme in bacteria, it is in a more varied range of bacteria than those found to contain *dsyB*, which were almost entirely alphaproteobacteria.

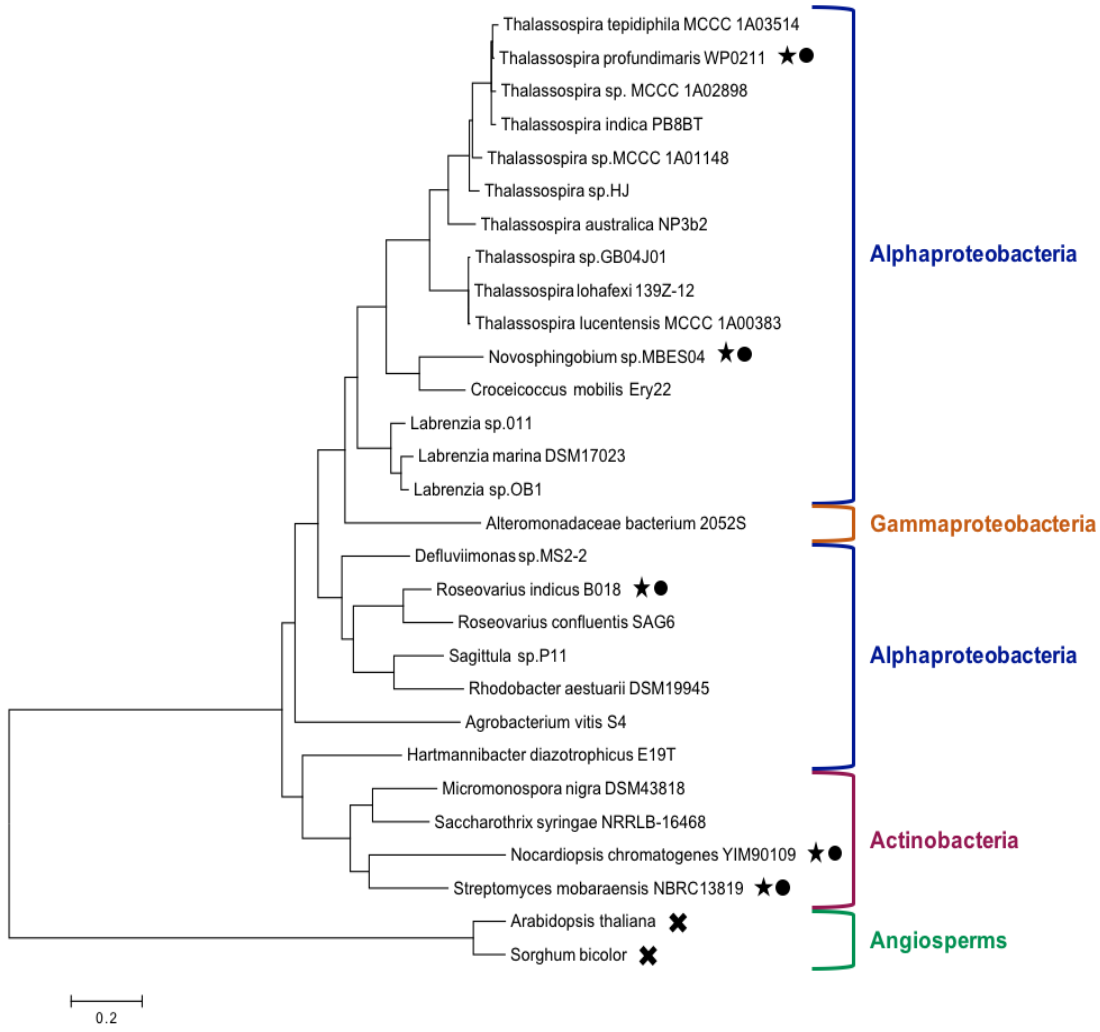

Figure 5-12: Maximum-likelihood phylogenetic tree of the 26 annotated SAM-MMT proteins, retrieved from NCBI and JGI IMG databases. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site, as indicated on the scale bar. Stars indicate strains shown to produce DMSP, circles indicate cloned, functional genes, and crosses indicate the larger MmtN in angiosperms, not linked to DMSP production. Distinctions between bacterial classes are demonstrated as colour-coded brackets, with four separate classes represented.

5.2.6 *Novosphingobium* MmtN is a MMT enzyme

Having shown that *Novosphingobium* lysate produced SMM only when supplied with both Met and S-AdoMet **(Figure 5-7)**, the next step was to demonstrate that MmtN is a Met methyltransferase in *Novosphingobium*. To accomplish this, the *mmtN* gene was amplified using primers designed to either end of the gene (**Chapter 2**) and cloned into the *E. coli* expression plasmid pET21a, creating pBIO21N1 so as to allow for the overexpression of this gene in *E .coli*. The *E. coli* strain BL21 lacks MMT activity, meaning that it does not synthesise DMSP or SMM from Met, even in the presence of S-AdoMet. *E. coli* BL21 containing pBIO21N1 was inoculated into LB media and induced with 0.2 mM IPTG, incubating overnight at 30°C. Cultures were mixed with 0.5 mM Met and incubated for a further 8 hours at 30°C, before determining SMM content by measuring the DMS produced from SMM lysis, and quantifying protein content. Alongside assays of the pBIO21N1 a negative control of *E. coli* BL21 containing the empty pET21a vector was also run**.** *E. coli* expressing MmtN displayed MMT activity compared to negative controls (*E. coli* BL21, *E. coli* with empty vector and the buffer) **(Figure 5-13)**. This confirms that *mmtN* has the expected activity, is likely responsible for SMM production from Met in *Novosphingobium,* and possibly other bacteria containing this gene (see below). To ultimately demonstrate that *mmtN* is involved in DMSP synthesis, the gene needs to be mutated in the host organism and the effects on DMSP production and MMT activity studied.

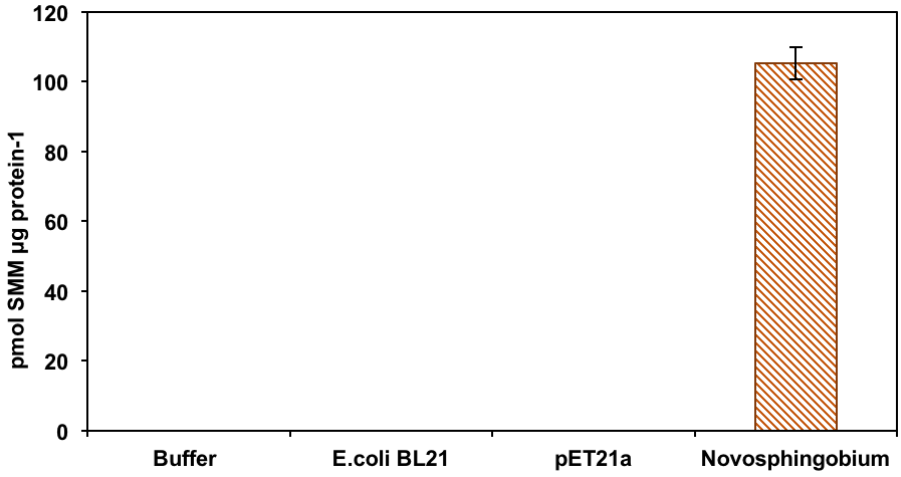

Figure 5-13: SMM production in *E. coli* BL21 containing pET21a with the *Novosphingobium mmtN* clone, alongside the buffer, *E. coli* with the empty vector, and *E. coli* alone. Error bars display standard error.

5.2.7 Bacteria with MmtN make DMSP

Having shown that *Novosphingobium* MmtN has MMT activity, it was important to test whether other marine bacteria that contain MmtN homologues (with > 50 % Identity) produce DMSP, and if these MmtN homologues are also functional. In order to test this, four other bacterial species across the diversity of this novel protein **(Figure 5-12)** were ordered from DSMZ Culture Collection, to study. The species that were chosen were two alphaproteobacteria; *Roseovarius indicus* and *Thalassospira profundimaris,* and two actinobacteria; *Streptomyces mobaraensis* and *Nocardiopsis chromatogenes*. These were cultured and checked for purity and 16S rRNA identity, then quantified for DMSP content **(Table 5-2)**. The full genomes of each species were also mined for the presence of *dsyB*.

DMSP content in the alphaproteobacterial strains was quantified after growth in 35 PSU MBM, with 0.5 mM nitrogen. For the actinobacterial species this quantification was more complex, as they did not grow at all in MBM, nor were they easily cultured in other tested liquid media, including several actinobacteria-specific media recipes (GYM and MYM). However the Actinobacteria strains grew reasonably well on plates, so for these Actinobacterial strains DMSP estimations were based on whole cells extracted from plates, which should be taken into consideration when comparing values.

Table 5-2: Isolates containing the *mmtN* gene originally discovered in *Novosphingobium sp.* MBES04, and the DMSP levels produced by them. Also shown is the confirmation of DMSP production by LC-MS, and the presence/absence of *dsyB*

MmtN-containing species	Similarity to Novo MmtN (BLASTp)		Intracellular DMSP concentration (pmol/ug protein)	Presence of DMSP indicated by LC-MS	Presence of dsyB in genome
	E value	Identity (%)			
Roseovarius indicus	3E-136	63	6.02 ± 1.2	YES	YES
Thalassospira profundimaris	1E-147	68	54.3 ± 3.6	YES	NO
Streptomyces mobaraensis	2E-90	53	*3.9 ± 0.7*	NT	NO
Nocardiopsis chromatogenes	2E-91	51	*1.5 ± 0.05*	YES	NO

Both *R. indicus* and *T. profundimaris* produced DMSP as expected. In the case of these two bacteria, DMSP production was further confirmed by LC-MS (see **Chapter 2**). For the actinobacteria, despite culturing difficulties, both also produced DMSP, as determined by GC, and by LC-MS for *N. chromatogenes*.

Knowing that these bacteria produce DMSP, we next needed to show that the *mmtN* gene in these microbes encoded for an enzyme with MMT activity, as was the case with *Novosphingobium* MmtN. To do this, primers containing restriction sites were designed to amplify from either end of the *mmtN* sequences in all four organisms (as with *Novosphingobium mmtN*, above). An amplified fragment of the gene was produced using these primers in a PCR reaction, purified, digested using the appropriate restriction enzymes and cloned into pre-digested pET21a (see **Chapter 2**). Clones were checked by re-digesting and visualising on gel electrophoresis, and by sequencing the fragment. All the clones were correct, resulting in plasmids pBIO21T2 (*T. profundimaris mmtN*), pBIO21R3 (*R. indicus mmtN*), pBIO21N4 (*N. chromatogenes mmtN*) and pBIO21S5 (*S. mobaraensis mmtN*). These plasmids were tested for MMT function, namely conferring the ability to produce SMM from Met to *E. coli*, as above **(Figure 5-14).**

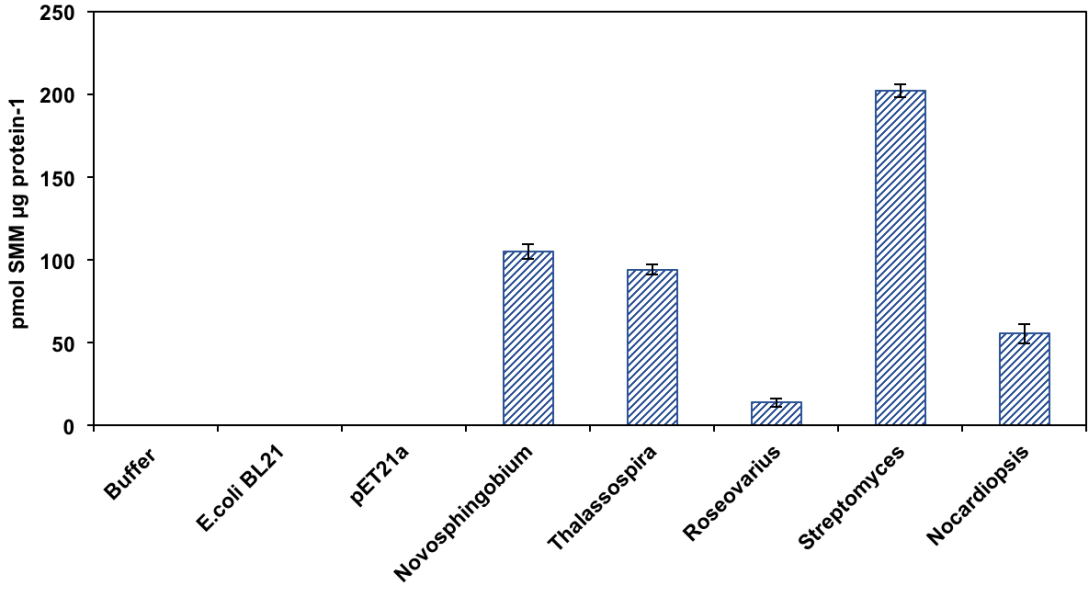

Figure 5-14: SMM production by *E. coli* BL21 containing pET21a with clones of *mmtN* homologs from five species (named on the x-axis), alongside the buffer, *E. coli* with the empty vector, and *E. coli* alone. Error bars display standard error.

We can see that the *mmtN* gene, as with *Novosphingobium above,* is functional in *T. profundimaris, S. mobaraensis* and *N. chromatogenes*. The data on *R. indicus* is inconclusive, and needs more investigation. *S. mobaraensis* did not appear to produce high levels of DMSP **(Table 5-2)**, yet the gene itself appears to be producing the highest amounts of SMM. This could be because in this experiment the gene is expressed at higher levels when it is in *E. coli* BL21, which is easier to culture than *S. mobaraensis.* Alternatively, perhaps SMM and/or DMSP are metabolised to other compounds in *S. mobaraensis*, which do not generate DMS upon NaOH lysis. It could also be that *S. mobaraensis* produces DMSP in the natural environment, but is not able to under lab

conditions, especially considering the difficulties had with culturing. **Figure 5-14** only shows that *mmtN* encodes a functional enzyme able to produce SMM from Met, not that the purpose for production is to synthesise DMSP.

5.2.8 Purifying and testing the MmtN protein

The fact that *mmtN* clones confer MMT activity to *E. coli* and that *Novospingobium* lysates have MMT activity **(Figure 5-14)** suggests that MmtN is responsible for this activity. However it is possible that other components of the cell lysate may be carrying out the MMT activity. To test this, we needed to overexpress and purify MmtN and characterise its activity and enzyme characteristics. The *mmtN* gene cloned into pET21a (pBIO21N1) was subcloned into a second plasmid, pET22b (Novagen, America), which contains a C-terminal His-tag enabling purification of the gene product through an affinity column. This purification work and subsequent characterisation was carried out with Chun-Yang Li at the Shandong University, Jinan, China.

The MmtN protein was expressed in *E. coli* BL21 cultures grown in LB media at 37°C, and then induced at 20°C for 16 hours with 0.5 mM IPTG. The protein was purified first with Ni^{2+}-NTA resin, and then fractionated using gel filtration buffer on a Superdex-200 column. Purification of the protein took place at 4°C. For the Ni^{2+}-NTA resin purification, wash buffer was used to remove protein impurities, followed by the elution buffer to elute the purified protein from the column. The image in **Figure 5-15** shows an example of the purified MmtN protein (33.55 kD) and DsyB (36.94 kD), as seen on an SDS PAGE protein gel, judged to be >95 % pure.

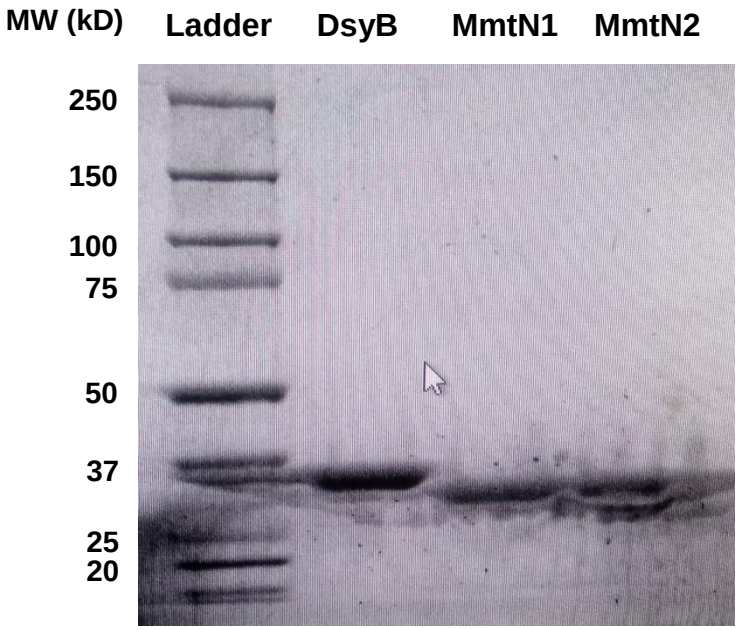

Figure 5-15: Visualisation of the purified proteins, DsyB and MmtN, on an SDS PAGE protein gel, run against a prestained precision protein standard ladder.

Measuring MmtN enzyme activity was complicated, as detection of SMM created by MmtN activity by HPLC and LC-MS is on-going. Instead, it was achieved by monitoring the production of S-AdoHyc (S-adenosyl homocysteine), the molecule produced after S-AdoMet has donated a methyl group to Met (facilitated by the enzyme MmtN) to generate SMM, which can be detected by HPLC. To determine the optimal conditions for MmtN activity, multiple temperature and pH conditions were trialled, with the enzyme activity compared at each stage. Once activity peaks the highest activity detected is defined as 100 % activity, and all the other tested conditions are then described as relative to it. The reaction mixtures were tested against temperature intervals of 10°C between 0°C to 60°C, for 30 minutes **(Figure 5-16-a)**. To determine optimal pH levels, MmtN activity was examined using Brtitton–Robinson at discrete pH values between pH 5.0 and pH 10.0 **(Figure 5-16-b)**. The kinetic parameters (K_m) for each of these experiments were determined by non-linear analysis, based on the initial rates and determined using 3.34 µM MmtN and 0.1 – 4 mM S-AdoMet **(Figure 5-16-c)**, or 0.1 – 6 mM Met **(Figure 5-16-d)**.

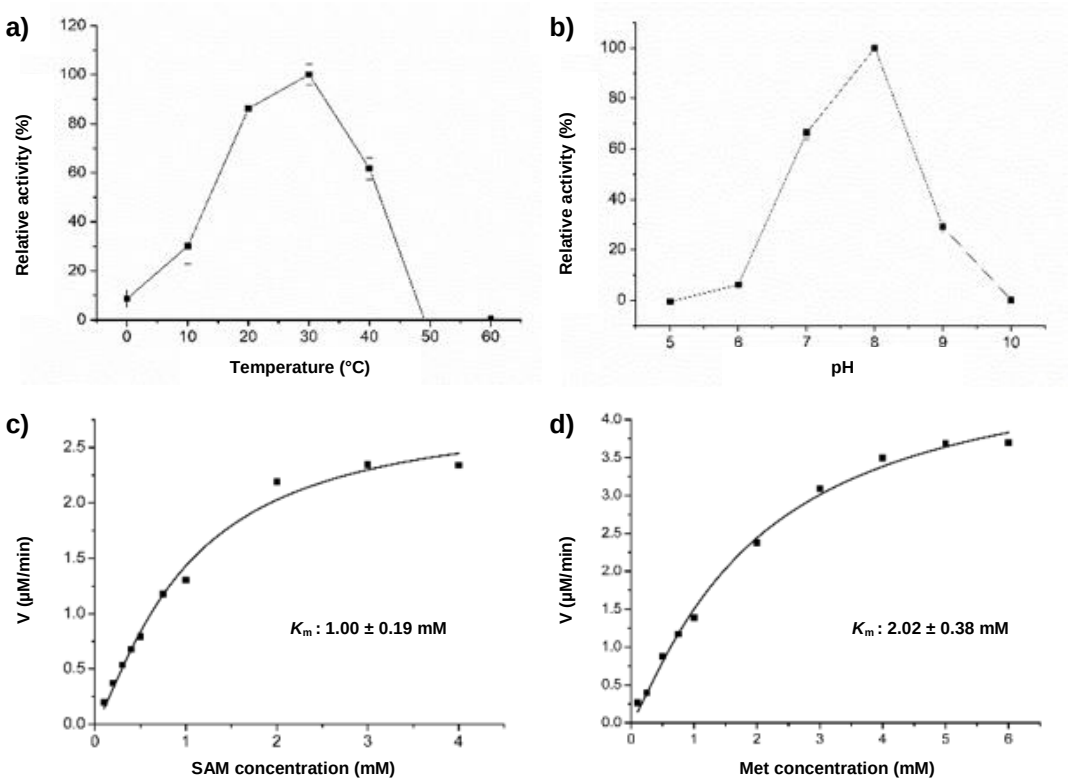

Figure 5-16: Characterization of MmtN. **(a)** Effect of temperature on the enzymatic activity of MmtN. Activity at 30°C was defined as 100%. **(b)** Effect of pH on the enzymatic activity of MmtN. Activity at pH 8.0 was defined as 100%. **(c)** Non-linear fit curve for SAM demethylation by MmtN. Initial rates of SAH generation were determined with 3.34 µM MmtN and 0.1 – 4 mM SAM in the reaction buffer. K_m was 1.00 ± 0.19 mM. **(d)** Non-linear fit curve for Met methylation by MmtN. Initial rates of SAH generation were determined with 2.72 µM MmtN and 0.1 – 6 mM Met in the reaction buffer. K_m was 2.02 ± 0.38 mM.

All experiments following these were performed under optimal pH and temperature for MmtN activity, and the amount of DL-Met and S-AdoMet was always excessive to requirements. The finalised method for testing MmtN activity was set as follows. The purified MmtN protein (3.34 µM), DL-Met (2.5 mM) and S-AdoMet (0.64 mM) were mixed with a reaction buffer containing 50 mM Tris–HCl (pH 8.0), in a total volume of 100 µl. The mixture was incubated at 30°C for 30 minutes, and the reaction was stopped with the addition of 15 µl 20 % hydrochloric acid (HCl). The amount of S-AdoHyc in the reaction mixture at this point was detected by HPLC on a Sunfire C18 column (Waters, Ireland). The MmtN methylation activity on other substrates including MTHB, MMPA and L-Gly was also tested, as well as a control of MmtN on its own **(Figure 5-17)**.

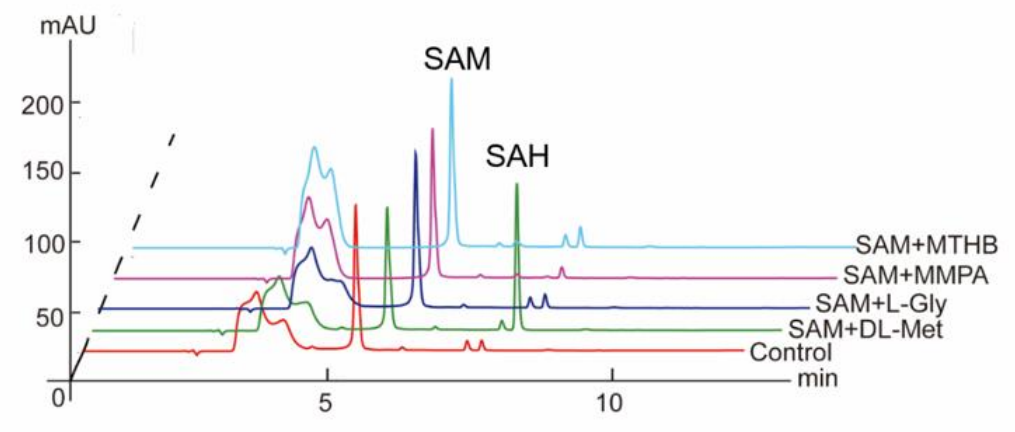

Figure 5-17: Detection of MTHB, MMPA, L-Gly and DL-Met methylation activity of MmtN via the intensity of absorbance on HPLC (wavelength of detection, 260 nm). The different coloured curves represent different reaction systems and include different substrates.

From the HPLC results, it was confirmed that the only reaction conditions from which S-AdoHyc was produced was when MmtN was incubated with S-AdoMet and DL-Met, meaning that we can assume that the MmtN protein is able to methylate Met through the methyl donor S-AdoMet, releasing S-AdoHyc as a by-product.

5.2.9 Selecting a strain for disruption mutation experiments

The next stage in determining gene function is its disruption in a wild-type strain. It was hypothesised that a mutation in *mmtN* would reduce or completely knock out DMSP production in the host strain, and that subsequent phenotyping may show this to be detrimental to the survival of the bacterium under a particular condition. This is necessary to show that *mmtN* is involved the generation of SMM and/or DMSP in *Novosphingobium*.

To determine which bacterium would be most suited for the generation of an *mmtN*$^{-}$ mutant, several were considered. The DMSP-producing Alphaproteobacteria are easier to

grow than the actinobacteria, so the three that were considered were *Novosphingobium*, *R. indicus* and *T. profundimaris.* As *R. indicus* contains both *dsyB* and *mmtN* **(Table 5-2)**, it was not an appropriate choice for a total knock-down in DMSP production because you would be less likely to observe a phenotype, as one pathway may compensate for the other. However, because of this fact it would be interesting to consider *R. indicus* for comparisons between the two genes in the future. In order to select the most appropriate species between *Novosphingobium* and *T. profundimaris*, both were tested against various antibiotics to determine which would be easiest to work with. The reason for testing antibiotic resistance and sensitivity is because the homologous mutation techniques that we utilise all involve antibiotic selection, thus if one organism is multi-drug resistant it would not be suitable for mutagenesis using this methodology. Cultures were grown to stationary phase overnight at 30°C in rich media (MB) before plating on MB agar containing gentamycin (20 µg/ml), kanamycin (200 µg/ml), neomycin (20 µg/ml or 40 µg/ml), rifampicin (20 µg/ml), spectinomycin (200 µg/ml), streptomycin (200 µg/ml) or tetracycline (5 µg/ml or 10 µg/ml), incubating at 28°C and detecting any growth over 48 hours **(Table 5-3).**

Table 5-3: The growth of two species of bacteria, *Novosphingobium sp.* MBES04 and *T. profundimaris* when tested against various antibiotics.

Antibiotic tested	**Growth of *Novosphingobium sp.* MBES04**	**Growth of *T. profundimaris***
Gentamycin	YES	NO
Kanamycin	YES	NO
Neomycin	YES	NO
Rifampicin	YES	NO
Spectinomycin	YES	NO
Streptomycin	YES	NO
Tetracycline	YES	YES

From this experiment, it was clear that *Novosphingobium* would not be easy to work with for mutagenesis, as it was resistant to all the tested antibiotics. Therefore it was decided that *T. profundimaris* was the best strain in which to create an *mmtN* mutant, because of all the tested antibiotics, it was only resistant to tetracycline.

Since all the previous culture dependent work in this chapter had focused on *Novosphingobium,* and not *T. profundimaris,* some extra growth experiments were performed to characterise the optimal growth conditions of *T. profundimaris* before mutagenesis could be attempted. Firstly, a growth curve was performed to determine the timeframe before reaching stationary phase **(Figure 5-18)**. This was performed in the same way as the *Novosphingobium* growth curve **(Figure 5-3)**, in triplicate in 35 PSU MBM.

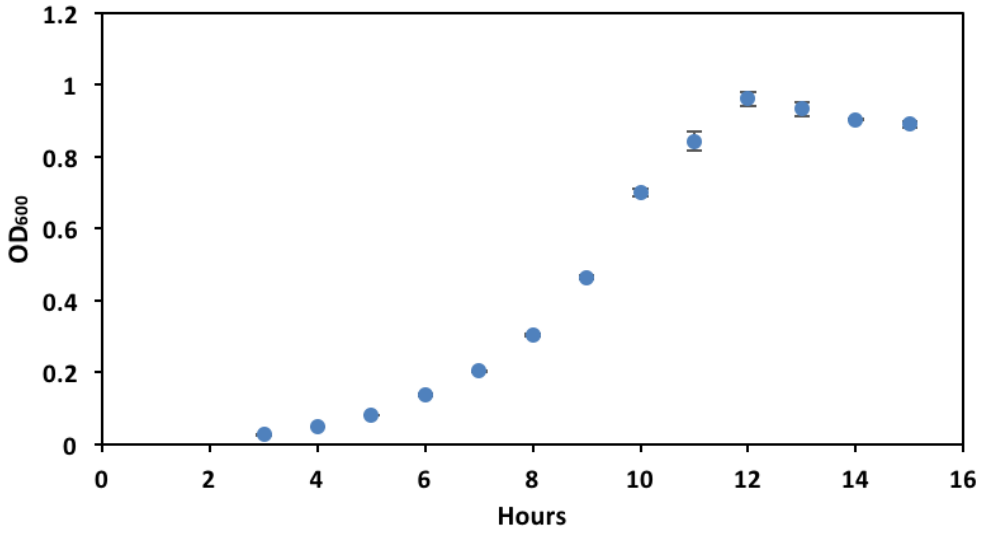

Figure 5-18: Growth curve of *T. profundimaris* in triplicate cultures, incubated at 30°C, shaken at 200 rpm for 15 hours until cultures reached stationary phase. Error bars display standard error.

We also quantified the DMSP production by *T. profundimaris* when cultures were incubated under different salinity and nitrogen conditions. Triplicate cultures were inoculated into MBM of salinity levels between 5 PSU and 70 PSU with 0.5 mM nitrogen, to test the effect that salinity has on the production of DMSP. At this point, it became lab standard to use the low nitrogen levels (0.5 mM nitrogen) in all work that used MBM media, as it was closer to natural conditions compared to the previous standard of 12 mM. Cultures were also grown in 35 PSU MBM with high nitrogen levels of 12 mM to observe the effect. They were incubated overnight, and DMSP levels quantified **(Figure 5-19)**.

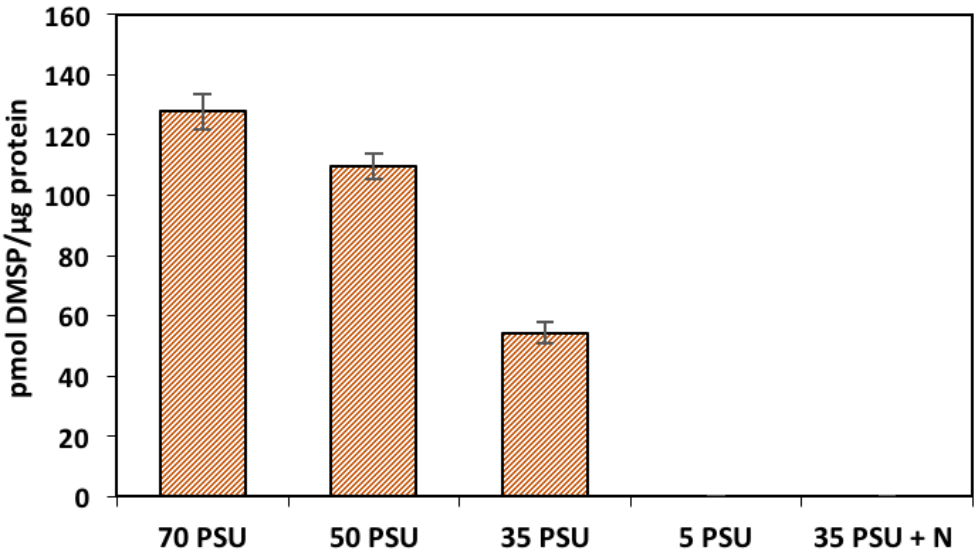

Figure 5-19: The effect of changing salinity or nitrogen availability on the production of DMSP by triplicate samples of *T. profundimaris*. Error bars display standard error.

As expected, when *T. profundimaris* was grown under increased salinity we observed significant increased DMSP production, from no production at 5 PSU to over 100 pmol/µg protein at 70 PSU. It seems that *T. profundimaris* has very high tolerance for salinity, perhaps in part due to the ability to produce DMSP. This would perhaps be expected of a bacterium living in salt marsh sediments. This experiment also confirms that it produces almost no DMSP when grown in nitrogen replete conditions. Although it has always been observed that decreased levels of nitrogen greatly increase DMSP production, it was a surprise that higher levels of nitrogen completely prevented it (at detectable levels). This could be because *T. profundimaris* is able to synthesise GBT as well as DMSP, and therefore when nitrogen levels are high it selectively produces GBT instead (see below).

5.2.10 Disrupting mmtN in T. profundimaris

To generate an *mmtN* mutant in *T. profundimaris,* the single homologous recombination method using the plasmid pBIO1879 (pK19spec,Todd et al. (2011)), was used. The mutagenesis described below was conducted with help from Andrew Curson. The first step in generating a *T. profundimaris mmtN* mutant was to obtain a spontaneous Rif resistant mutant. This is required to allow the selection of *T. profundimaris* away from *E. coli*. This was created by plating high density of cells on plates with Rifampicin in, and incubating them for long enough that spontaneous mutants arise. The *T. profundimaris*-Rif strain was treated as wild-type for all the experiments performed comparing the *mmtN*$^-$ to wild-type.

The single crossover (SCO) gene disruption method involves cloning an internal fragment of a gene (in this case *mmtN*) into pBIO1879 (Todd et al. 2011), which is a derivative of the pK19mob plasmid described in Schäfer et al (1994), with a Spc^R cassette cloned in. This plasmid is a suicide vector, meaning that it only replicates in *E. coli*, thus when it is mobilised out of the donor strain into a host in which the origin of replication does not work, it is not maintained unless it integrates into the genome through homologous recombination. Recombination is most likely to occur between the cloned fragment and complement sequence within the genome of the host. Once the ~5.7 kb plasmid is integrated into the genome it confers spectinomycin resistance, and disrupts the target gene to make it non-functional.

To clone a fragment that was internal to *mmtN* into pBIO1879, primers were designed either side of a ~ 500 bp central region of the *T. profundimaris mmtN* gene. These primers were used to amplify and then digest the fragment, and cloned into pre-digested pBIO1879, essentially as for the cloning of *mmtN* genes into pET21a in **5.2.5**. The plasmid containing the *mmtN* fragment was transformed into *E. coli* 803 competent cells and then mobilised into *T. profundimaris*-Rif using tri-parental crossing, as in methods (Figurski and Helinski, 1979). Potential *T. profundimaris mmtN* mutant colonies were isolated on YTSS media containing Rifampicin and spectinomycin and kanamycin to select for pBIO19TK integration. These potential mutants were checked by PCR using primers exterior to the cloned internal fragment. For any potential mutants that gave no PCR product (expected because the insertion of the plasmid renders the potential product too large for PCR), their DMSP production phenotype was examined by GC, as above. Upon carrying out this screen we identified a $mmtN^-$ mutant which no longer produced DMSP in GC analysis. Furthermore, when examined by LC-MS, this mutant, termed *T. profundimaris*-R ($mmtN^-$), also showed no detectable DMSP **(Figure 5-21)**. Thus, we have generated a *T. profundimaris* $mmtN^-$ deletion mutant, and have demonstrated that mmtN is required for DMSP-synthesis in this marine alphaproteobacterium.

In order to show that the lack of DMSP in the *T. profundimaris* $mmtN^-$ mutant is due to the mutation it was necessary to complement the strain with a W/T *mmtN* gene cloned on a plasmid (see **Chapter 2**). The *mmtN* clone from *Novosphingobium*, termed pBIO21N1, was subcloned into pLMB509 and then mobilised into the *T. profundimaris*-Rif $mmtN^-$ mutant, and tested for DMSP production by GC. As can be seen in **Figure 5-20**, the *mmtN* returned function to the mutant, although not reaching W/T levels, there is a noticeable increase in DMSP production from 0 in the mutant. Thus we can be confident that the SCO mutant in *mmtN* is responsible for the observed phenotypes and not secondary mutations elsewhere in the genome.

Although function was clearly returned to the *T. profundimaris*-Rif *mmtN*$^{-}$ mutant, it is somewhat lower than that in the W/T. This could be due to the fact that pLMB509 possesses a taurine-inducible promoter that is not a natural promoter for *mmtN*, and giving it low expression in comparison to the W/T.

It was notable in the LC-MS work performed on *T. profundimaris* that the wild type strain only produced DMSP, but the *mmtN*$^{-}$ mutant strain produced large quantities of the nitrogenous osmolyte GBT (**Figure 5-21**). This supports the previously suggested theory that there is a switch between the two, where GBT may be being synthesised when DMSP is no longer able to be produced, to take over the role as osmoprotectant. Given this result, one may not expect to see an obvious phenotype for the *mmtN*$^{-}$ mutant, if GBT can compensate by adopting the role of DMSP in *T. profundimaris.*

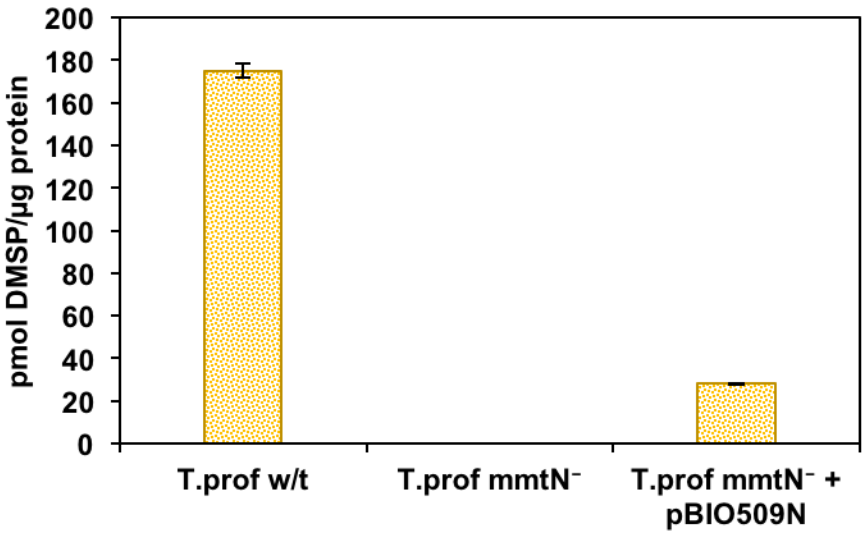

Figure 5-20: DMSP production by the *T. profundimaris* W/T, *mmtN*$^{-}$ mutant and the complemented mutant, expressing *mmtN* from *Novosphingobium* via pBIO509N.

5.2.11 Phenotyping the *mmtN⁻ mutant*

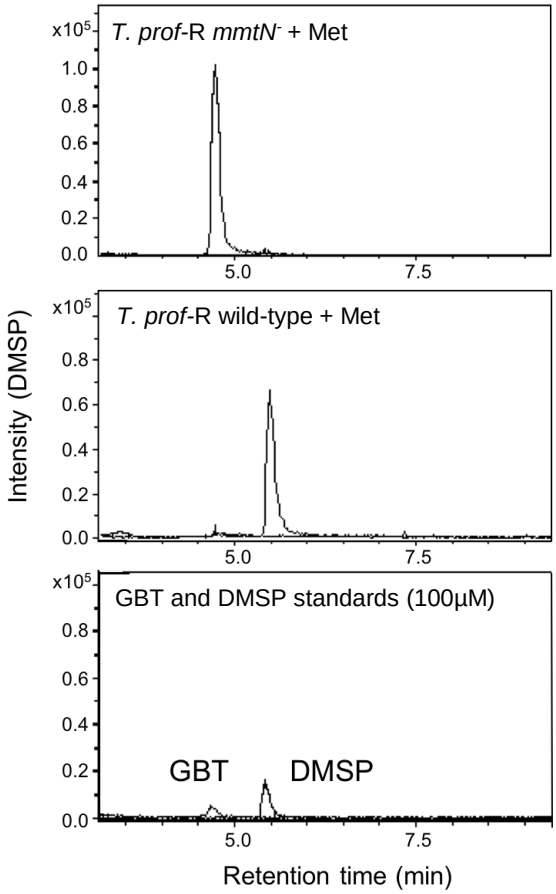

Figure 5-21: LC-MS chromatogram showing DMSP (*m/z* 135) and GBT (*m/z* 118[H]+) in *T. profundimaris*-Rif wild-type and *mmtN⁻* strains in the presence of Met.

Having a *T. profundimaris mmtN⁻* allowed us to examine any phenotype that the lack of DMSP production might cause. Any impact on growth or survival on *T. profundimaris* would suggest that DMSP was either necessary for growth, or at the very least required for optimal cell function. As salinity is clearly tied to DMSP production in *T. profundimaris* **(Figure 5-19)**, it was the first condition used to test the mutant with. Triplicate 100 ml cultures of MBM were inoculated with the *T. profundimaris*-Rif wild-type or *mmtN⁻* mutant. These cultures were either 35 PSU or 50 PSU salinity with 0.5 mM nitrogen, or 35 PSU MBM with 12 mM nitrogen as a control. These were incubated for 14 hours at 30°C, shaking at 200 rpm. The growth was measured every hour by reading the OD_{600} of 1 ml culture, until at least some of the cultures appeared to reach stationary phase, judged by similar OD_{600} levels being measured for at least three hours **(Figure 5-22)**.

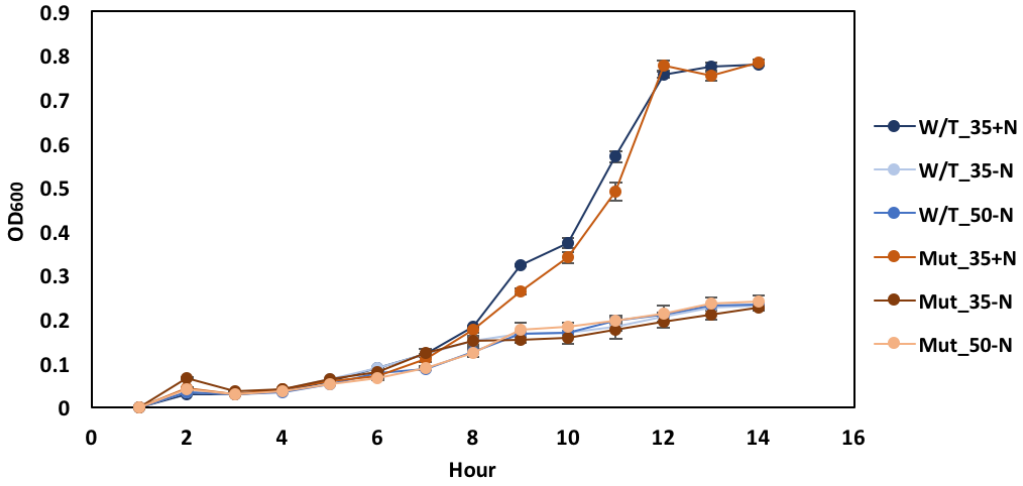

Figure 5-22: Growth curves of *T. profundimaris*-Rif wild-type (W/T) and the *mmtN*$^{-}$ mutant (Mut) under 35 PSU or 50 PSU salinity and 0.5 mM nitrogen (indicated by –N), as well as under 35 PSU salinity with 12 mM nitrogen (indicated by +N). Samples are in triplicate and error bars display standard error.

Unfortunately there appears to be no significant difference between growth of the mutant and wild-type in either salinity regime. Under 50 PSU it is clear that growth rate is impeded, but it continues to gradually increase with or without the function of *mmtN*, suggesting that either other osmoprotectants are playing a role (either GBT or others where nitrogen is too low), or that DMSP may not in fact play a role in salinity protection in *T. profundimaris*. It could be that DMSP concentration in *T. profundimaris* is increasing under high salinity as the result of other processes, or is simply not required for growth.

Another method of analysis that was carried out was a competition experiment, observing any change in the ratio between W/T and mutant *T. profundimaris* cells grown under various conditions. Cultures of both the mutant and wild-type were inoculated from fresh plates into 35 PSU MBM with 12 mM nitrogen, incubated at 30°C overnight, then mixed 1:1 (500 µl of each) in a 1.5 ml microcentrifuge tube, leaving one mixed culture in 35 PSU conditions. A serial dilution was performed and plated on MBM agar and incubated at 28°C until single colonies were visible. These were picked and individually streaked in the same place on MBM plates with or without kanamycin, enabling differentiation between the mutant (Kan-resistant) and the wild-type (Kan-sensitive). The percentages of both were calculated. At the same time, the 1:1 mix was also inoculated into high stress conditions, including low nitrogen and high salinity. The mix was inoculated into 50 PSU MBM with 0.5 mM nitrogen and 35 PSU MBM with 0.5 mM nitrogen. Cultures were incubated overnight and the same process of plating and then streaking on MBM with/without kanamycin was followed. Percentages were calculated and compared to those before the stressed conditions **(Table 5-4)**.

Table 5-4: Percentages of *T. profundimaris*-Rif wild-type colonies versus *T. profundimaris*-Rif *mmtN*$^-$ mutant colonies when in competition in a mixed culture after growth in stress conditions.

Growth condition	Replicate	Percentage growth of wild-type	Percentage growth of *mmtN*$^-$ mutant
Before stressed growth		63	37
50 PSU -N	1	64	36
	2	64	36
	3	61	39
	Average	**63**	**37**
35 PSU -N	1	61	39
	2	50	50
	3	50	50
	Average	**54**	**46**

Once again, there does not appear to be an obvious phenotype, except that there is almost consistently a higher proportion of wild-type colonies to mutant colonies in the mix. This doesn't appear to change when the mix was inoculated in 50 PSU conditions, and actually decreases under 35 PSU conditions to almost 50:50.

The effect of temperature was also tested, as it is also thought that DMSP also acts as a cryoprotectant. Cultures of both the mutant and wild-type were grown to stationary phase in 35 PSU MBM with 0.5 mM nitrogen. The OD_{600} was adjusted to 0.3 and 1 ml of cells were then centrifuged at maximum speed, before washing the pellets in 1 ml MBM and performing a serial dilution. These were then plated on YTSS and incubated overnight until colonies were visible. These were then counted in order to calculate cfu/ml and determine cell numbers before freezing. The 1 ml MBM cultures were frozen at -20°C for one week before defrosting, diluting in the same way and plating on YTSS. Colony counts were performed, and the cfu/ml compared to those prior to freezing **(Figure 5-23)**.

Although both the mutant and wild-type had similar cfu/ml counts before freezing, there seemed to be no phenotype of the mutant after freezing. Indeed, while the cfu/ml counts for the mutant remained almost the same after freezing, the wild-type actually seems to have decreased compared to the original counts, which was contrary to what was expected.

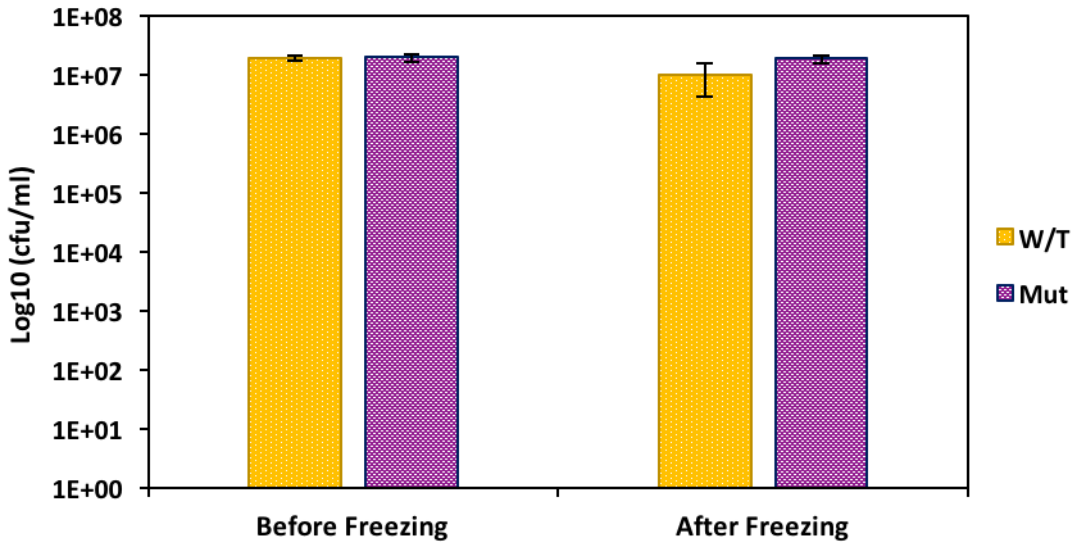

Figure 5-23: The cfu/ml of *T. profundimaris*-Rif wild-type (W/T) and *T. profundimaris*-Rif *mmtN*$^-$ mutant (Mut) before and after freezing at -20ºC for one week. Samples were in triplicate, plotted on a logarithmic scale and error bars are standard error.

From all the characterisation experiments performed in this chapter, we could not identify a phenotype for the *T. profundimaris*-Rif *mmtN*$^-$ mutant. However, there are always other conditions that have yet to be tested, including testing with H_2O_2 to test the reaction to oxidative stress.

5.2.12 Searching for *mmtN* in other Stiffkey isolates

Of the three non-*dsyB*-containing species that were sent for WGS in **Chapter 4**, *Novosphingobium* was the only isolate shown to contain *mmtN*. Both the *Alteromonas* and *Marinobacter* isolates from Stiffkey do not appear to possess it (from BLASTp analysis), although there is an *Alteromonodaceae* bacterium in the JGI database that has an *mmtN* **(Figure 5-12),** which suggests that at least some bacteria in that order may have it. Interestingly, when the other *dsyB*-containing sequences were also mined for *mmtN*, the *Rhodobacterales* also appears to have it (E value 3E-145, identity 69%). This is unusual but not unprecedented, as *R. indicus* also appears to contain both **(Table 5-2),** and many species contain multiple *ddd* genes (Todd et al. 2011; Curson et al. 2012).

5.2.13 Other candidate genes in the methylation pathway

When the *mmtN* gene was identified in *R. indicus* it was observed that there appeared to be several genes upstream that could be part of a DMSP-synthesis operon. This was because the other steps in the methylation pathway likely require some combination of a decarboxylase and a transaminase, both of which appear to be close to the SAM-dependent methyltransferase in *R. indicus* **(Figure 5-24),** with an aspartate

aminotransferase and a diaminopimelate decarboxylase adjacent to it. Where possible, the whole genome sequences of the other *mmtN*-containing species were mined for these genes, and several of them appear to contain both, also adjacent or close to *mmtN*. There were also a number of species that did not contain this 'operon' of genes, for example, *Novosphingobium* appears to contain the same aspartate aminotransferase but no full diaminopimelate decarboxylase (although there is a small hypothetical protein that has ~38% identity to it). Although many of the *Thalassospira* and *Labrenzia* sequences appeared to contain this small hypothetical protein with similarity to the decarboxylase, none of them seem to have the aspartate aminotransferase anywhere near *mmtN.* However, many of them did contain a pyridoxal phosphate-dependent aminotransferase, which could be performing the same or similar role **(Figure 5-24)**.

Both the *R. indicus* aspartate aminotransferase and diaminopimelate decarboxylase, as well as the hypothetical protein between them **(Figure 5-24)** were cloned into pET21a at the same time as the rest of the *mmtN* genes. Although this was all that was accomplished in this body of work, this operon is potentially a very interesting route of study to follow up on, as it could be the first demonstration of the full suite of genes used by a species to take Met all the way through the pathway to DMSP.

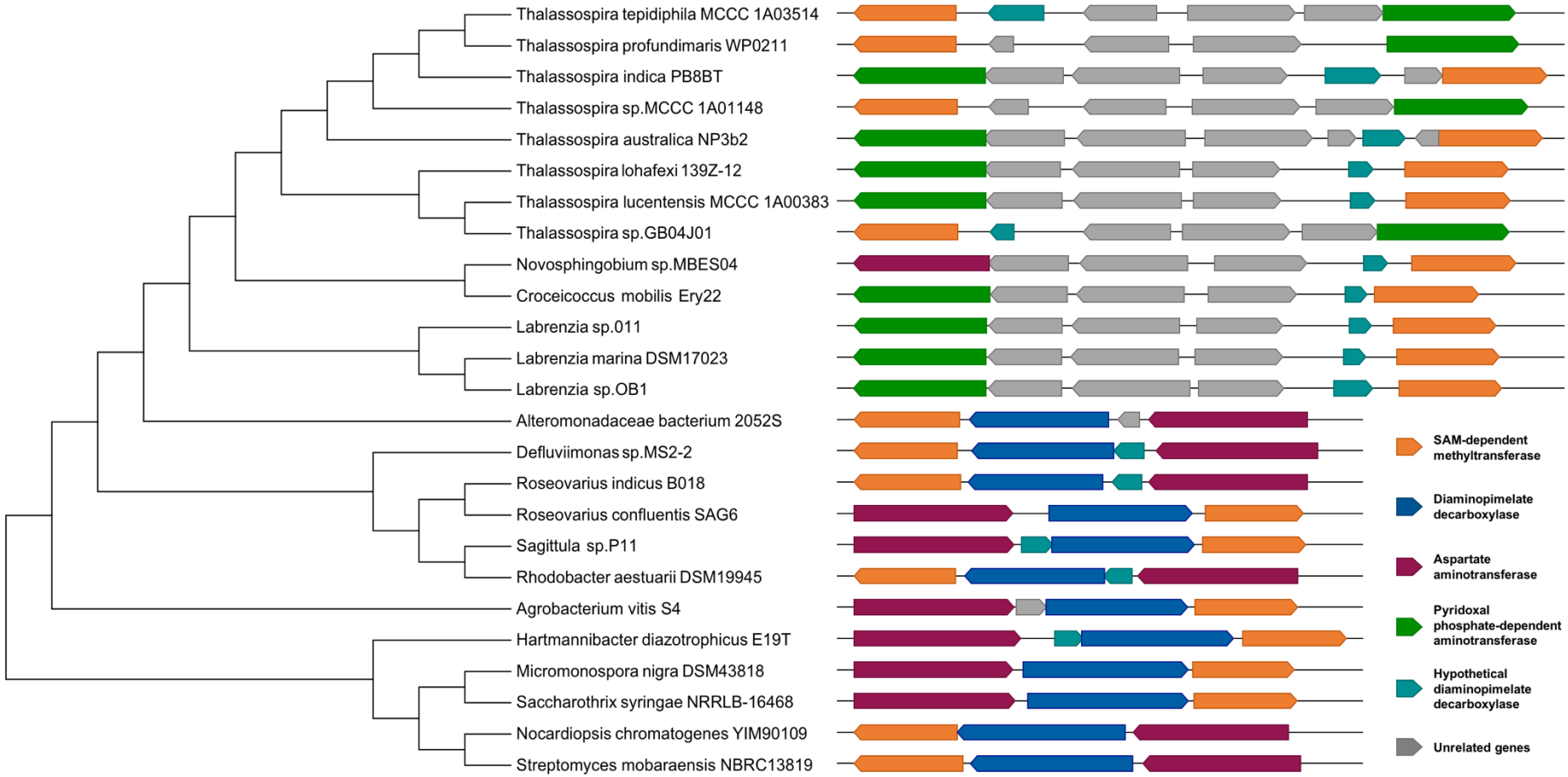

Figure 5-24: The synteny between the genes immediately surrounding *mmtN* in multiple *mmtN*-containing species that are potentially involved in DMSP production, and how those species relate to each other. Genes of interest are highlighted as *mmtN* (SAM-dependent methyltransferase), a diaminopimelate decarboxylase, an aspartate aminotransferase and a pyridoxal phosphate-dependent aminotransferase.

5.3 Discussion

5.3.1 Summary of work

The main aim of the work carried out in this chapter was to identify and characterize novel DMSP-synthesis genes in *Novosphingobium,* which had been shown to lack the only identified DMSP-synthesis gene, *dsyB*. This was accomplished with the discovery of a second DMSP-synthesis gene, *mmtN*, which encodes for a SAM-dependent methyltransferase that methylates Met to become SMM, which is the first step in the methylation DMSP production pathway. This gene was found to be in a variety of other species; it not limited to alphaproteobacteria, but is also in several actinobacterial species as well as potentially one gammaproteobacterium. It was also demonstrated that when this gene is disrupted in *T. profundimaris* WP0211, DMSP production no longer takes place, confirming that this gene plays a role in it. Finally, phenotyping experiments were carried out on the mutant to attempt to draw more conclusions about the role that DMSP synthesis plays in bacteria. Unfortunately, as yet no definitive phenotype has been discovered.

5.3.2 Moving forward with the *mmtN*$^-$ mutant

In tested DMSP-producing bacteria, whether they have *dysB* or *mmtN*, DMSP production, and in case of *L. aggregata* LZB033 (the only bacterium tested for this so far), *dsyB* transcription is enhanced under various conditions, with high salinity causing the highest fold change (Curson et al. (2017). It was therefore expected that the knocking out *dsyB* or *mmtN* would have an impact on survival in general, or at least under regulating conditions. However, this was not the case, both in the published work and in this chapter. This could be because while DMSP production is useful and clearly utilised by the organisms, it may not necessarily impact the survival of them. However, it could also be because the condition in which DMSP is most required has not yet been tested. For the *mmtN*$^-$ mutant at least, phenotyping has only just begun. More conditions have yet to be trialled, including perhaps higher salinity conditions, as the growth curves only tested up to 50 PSU and it has been shown that *T. profundimaris* can grow and produce DMSP even at 70 PSU. Other conditions that could be tested are different types of oxidative stress, such as UV light or H_2O_2 treatments, as it appears that DSYB, the eukaryotic synthesis enzyme, at least, might play a role in oxidative stress protection (Curson et al. 2018). There may also be other impacts that the loss of DMSP production has within the cell, that are not necessarily detrimental to growth. However, a more realistic scenario is that, in terms of salinity tolerance at least, other osmoprotectants including, but not limited to, GBT, are likely produced at higher volume to cover the loss of DMSP. We clearly demonstrate here that GBT accumulation is enhanced in the *T. profundimaris mmtN*$^-$ mutant compared to

the WT supporting this hypothesis. In this case it is still likely that DMSP is playing a significant role in, for example, osmoprotection. However, we do not notice this due to the compensation by e.g., GBT production. The only way to demonstrate this is would be to identify and knock-out genes involved in the synthesis of other osmolytes such as GBT in the *T. profundimaris mmtN*$^-$ mutant, and then screen for phenotypes in such double knock-out mutations. Indeed, it is noteworthy that Ana Bermejo Martinez has created mutants of the genes for GBT synthesis in *L. aggregata* LZB033 and observed no detrimental growth phenotype under the same conditions tested here.

Genetic studies so far are limited to two alphaproteobacteria, both of which also produce GBT. It would be interesting to create disruption mutants in other *mmtN*- and/or *dsyB*-containing strains, as they may lack GBT synthesis pathways and use DMSP as the sole osmolyte. Furthermore, in such microbes the regulation of DMSP synthesis may be different than what is observed in *L. aggregata* LZB033 and *T. profundimaris*. One would imagine that DMSP synthesis may be constitutive in some organisms, as it is in many phytoplankton who use it as their major osmolyte, e.g. a number dinoflgellates (Yoch 2002). The hypothesis would be that in such organisms a DMSP$^-$ mutant would have a more severe detrimental phenotype. As yet no work has been carried out on actinobacterial DMSP producers, which are clearly very divergent to the alphaproteobacteria. These important producers of secondary metabolites may be a good place to start.

It may also be of interest to study DMSP production in bacteria which have multiple DMSP-synthesis genes/pathways, e.g. *R. indicus*. Why would an organism have two pathways for the synthesis of the same molecule? Certainly double mutants would need to be created in this strain to answer that question. It is possible that the two different pathways exist in the same organism as a redundancy or failsafe, suggesting that it might be more important for survival in those particular species.

It may also be worth looking at whether it is the loss of SMM synthesis in bacteria with *mmtN* that might have an effect on growth or function. This is because SMM production in plants is not necessarily specific to DMSP production, and this may also be the case in bacteria.

5.3.3 Other characterization of *mmtN* and bacterial DMSP production

Even though there is no discernible growth phenotype for the *mmtN*$^-$ mutant, there are still other routes of study left to pursue to understand the role that this gene plays within the organism. Firstly, now that the protein has been successfully purified, there are many experiments that can be performed, such as X-ray crystallography to determine the structure and then further understand its reaction mechanism. This may yield important

information as to the environmental context if, for example, *mmtN* is a metalloenzyme. We know that some metals, e.g., Fe, are very rare in the open ocean and thus if an enzyme is dependent on such a metal then Fe availability may affect DMSP productivity. We have not characterised the transcriptional or translational regulation of *mmtN* in any organism. This is essential to gain a better understanding of how DMSP is produced in these organisms. This could be done via the generation of lac fusions and/or RT-qPCR as was done in Curson et al. (2017) and (2018).

If we were to gain a better understanding of the regulation of DMSP production and *mmtN* expression a logical progression and interesting question would be to examine the molecular control points of this regulation. Is there a master regulator that controls the transcription of DMSP synthesis genes and/or other salinity regulated genes? To ask this question one would have to identify the promoter region of *mmtN* and look for conserved potential regulator binding sites in this DNA that are common to other *mmtN* genes form other bacteria. There are a huge number of possible ways forward on this topic and it is an important set of questions that should be addressed in the near future.

Given the fact that MmtN is homologous (~ 30 % identity to the N-terminus of MMT) to the plant MMT eukaryotic, it could be proposed that one is the progenitor for the other. There is a precedent for this with DSYB and DsyB in algae and bacteria respectively (Curson et al. 2018). In the case of DSYB/DsyB, evolutionary analysis suggested that alphaproteobacteria is the sister clade to the eukaryotic gene. Further analysis led to the conclusion that this particular gene actually appeared in prokaryotes first, followed by transfer into eukaryotes, potentially on multiple occasions. It was thought that this transfer either took place through endosymbiosis (at the time of mitochondrial origin), or more recently by horizontal gene transfer (HGT). It would be interesting to perform the same analysis on *mmtN* to determine if perhaps a similar occurrence took place between *mmtN* in prokaryotes and eukaryotes in the past. This would reveal much about the importance of bacterial and eukaryotic DMSP production both historically and in the modern day. Initial evolutionary analysis of MmtN by Lewis Spurgeon suggests that a different scenario may have taken place than what was the case with DSYB/DsyB. Here, the data suggests that early on in the evolution of this protein family there was a gene duplication, where one developed into the ~ 1000 aa MMT which is present in all flowering plants and some bacteria (many deltaproteobacteria), and the other became the shorter MmtN (lacking the MMT C-terminal PLP-binding domain), which is found in exclusively in bacteria. Both these proteins have retained the same enzyme function, and it is likely that the C-terminal domain has a role that is additional to MMT activity, perhaps in regulation as proposed in Bourgis et al. (1999). A further point worth noting is that only *mmtN* is a confident reporter of DMSP synthesis, since all tested bacteria that contain it have been shown to produce

DMSP. In contrast, only one of the four tested bacteria with MMT made DMSP. This is consistent with the fact that all plants have MMT but few make DMSP.

The discovery of a possible full suite of genes close together in the genome of *R. indicus* and other bacterial genomes was also noteworthy. There are many possible steps that could be taken to advance this work. The genes could be cloned into expression vectors such pET21 and screened for the expected transformation of SMM into e.g. DMSP amine. However, as yet we have no evidence that any bacterium with *mmtN* uses the plant methylation DMSP synthesis pathway. All we know is that MMT is involved, and that SMM is an intermediate. Thus it would be prudent to first characterise the DMSP synthesis pathway in *T. profundimaris.* This would involve the feeding of labelled Met (stable or radioactive) and the tracing of intermediates via HPLC or LC-MS/NMR. This is essentially the method used by Hanson to establish the known DMSP synthetic pathways (Gage et al. 1997). Although this metabolomics approach is essential to provide knowledge on the pathway, there are other approaches that also could be used. One such approach is to mutate the candidate DMSP synthesis genes in *T. profundimaris* using methods similar to those used to generate the *mmtN* mutant, being careful not to generate polar mutations. In recent work Andrew Curson has knocked out the aminotransferase gene and shown that the resultant mutant no longer produces DMSP. The metabolites produced by this aminotransferase mutant could be analysed by the above metabolomics techniques using labelled Met, to identify the reaction the enzyme catalyses by virtue of identifying labelled metabolites accumulating in the mutant vs the W/T.

5.3.4 Concluding Remarks

The discovery of not only a novel DMSP-synthesis gene, but one that utilises an entirely new pathway, is very significant, particularly because it means that estimates of bacterial DMSP production based on *dsyB* as a reporter are in fact conservative, and it is likely that even more species are producing DMSP than previously thought. If there are two bacterial DMSP-synthesis genes, it is likely that there are more, especially considering there are still DMSP-producing strains that have been isolated that do not appear to contain either.

CHAPTER 6

USING CULTURE-INDEPENDENT TECHNIQUES TO IDENTIFY DMSP-PRODUCING BACTERIA

6 CULTURE-INDEPENDENT IDENTIFICATION

6.1 Introduction

6.1.1 Work describing bacterial DMSP production so far

The work carried out in this chapter is the complement to that described in **Chapter 4**, which demonstrated that several different genera of DMSP-producing bacteria can be isolated from Stiffkey salt marsh. Many of these bacteria contain one of the DMSP-synthesis genes – either *dsyB* or *mmtN* (or in some cases, both). In this body of work so far, we have shown through culture-dependent experiments that bacteria can be easily isolated from Stiffkey salt marsh sediment, and that it is possible to 'enrich' for DMSP production within the sediment. Primers were designed to amplify multiple homologs of *dsyB*, and were used as a preliminary screening method for the gene in unknown isolates, as well as in qPCR experiments on DNA and cDNA. During this process, it was also discovered that bacteria are able to utilize not one but two of the known DMSP-production pathways, the transamination pathway (involving *dsyB*) and a pathway involving a novel methylation pathway which involves the newly identified bacterial DMSP-synthesis gene, *mmtN*, which carries out the methylation of Met to SMM. This information is very informative as to the rough makeup of the bacterial community in Stiffkey, but does not enable large-scale, refined community analysis.

6.1.2 Combining culture-dependent and –independent analysis

The discovery of *mmtN*, and its likely role as a reporter of DMSP production in bacteria, means that previously published work describing bacterial DMSP production through the analysis of *dsyB* alone are in fact underestimating the role of DMSP synthesis (Curson et al. 2017; Curson et al. 2018). This is an example of how culture-dependent and culture-independent analysis can be utilized to complement each other very effectively. Culture-independent techniques (metagenomics, 16S rRNA amplicon, qPCR, etc.) are valuable tools for widespread analysis of abundance and transcription of functional genes, as well as enabling microbial community analysis. However, they are limited in their potential usefulness unless they have been informed by culture-dependent study – we can only search for genes that we already know are linked to DMSP production (which until this point, has only been *dsyB*). It is necessary to identify key functional genes in model organisms in order to better understand how these organisms and processes of interest function, at least under controlled laboratory conditions before we can draw conclusions about the function that these genes might play *in situ*. Therefore, culture-dependent techniques such as isolation work identifying novel species able to synthesise DMSP,

gene discovery and characterization, are all vital to improve the depth of analysis performed by culture-independent sequencing.

Conversely, culture-independent work is also needed in this type of study, because while the culture-dependent study revealed that DMSP-producers exist in Stiffkey salt marsh, the methods of culturing used typically only identify ~1 % of bacteria in an environment (Davis et al. 2005; Saleh-Lakha et al. 2005), meaning that the analysis is heavily skewed towards that phylogeny of bacteria, potentially missing a significant amount in the natural environment. Culture-independent experiments are able to account for the other 99% of uncultivable bacteria in the sediment, and display the true abundance of particular species in the community. It is also able to analyse abundance and possibly transcription of functional genes without the bias of lab conditions. Metatranscriptomics in particular enables the quantification of bacterial gene activity without removing it from its natural environment, which could potentially be removing it from a number of interactions and factors that are impossible to account for or recreate in the laboratory. Results from this type of work are therefore much closer to the true levels of transcription in the environment. It has long been recognized that plate culturing is an inaccurate method of community analysis (Skinner et al. 1952), although it is acknowledged that other methods such as direct counting using a microscope also has disadvantages. For this type of study, if culture-dependent methods were the only type of method used, then it results in numerous phylogenetic groups being vastly understudied because they are either not cultured at all, or are slow growing or complicated to grow. When performing plate culturing experiments many species are unable to grow on the agar at all, likely due to missing requirements such as particular vitamins or carbon sources. Some would be able to grow if given a longer incubation time, but they are often out-competed as plates are overtaken by faster-growing bacterial species. If they do not form visible colonies on the plate it means that they cannot be picked and tested (Davis et al. 2005). Despite these limitations, this method of culturing is still the prevailing method of culturing novel isolates from an environment. There are ways in which isolations from sediment can be improved, but for the purpose of this study, and taking into account the culture-independent work that was also planned, the depth and breadth of species retrieved from Stiffkey salt marsh was sufficient.

Another issue associated with culture-dependent work is the bias associated with analysis based on what is cultivable (Torsvik & Øvreås 2002), and the fact that, often out of necessity, experiments quantifying gene or protein activity are usually performed under conditions that are nowhere near the *in situ* conditions of an environment. This is often unavoidable, and, in the example of nitrogen levels being higher experimentally than in seawater environments, a balance has to be found between being able to culture and

produce results, and being close to accurate levels. Culture-independent sequencing can be performed on Time 0 (natural) samples, meaning that the abundance and activity of genes measured by this method are measured in as close to *in situ* conditions as possible.

6.1.3 Culture-independent methods in literature

The strength of combining these two methods is demonstrated by how often they are both used in other studies. Not all culture-independent techniques involve large-scale sequencing like metagenomics. Indeed, even just the incorporation of 16S rRNA clone libraries into community analysis can improve the analysis considerably. In work carried out by Steven et al. (2007) it was commented on that previous studies of permafrost environments had been performed solely on cultivable cells, missing > 99 % of the total microbial community. Therefore, both culturing experiments and 16S clone libraries were used to study microbial diversity in an Arctic permafrost sample. Culturing experiments were still used to determine viable heterotrophic bacteria in the sample after different growth conditions. Indeed, it was found that incubation at lower temperatures (around 5°C instead of 25°C) increased cell counts three-fold. Colonies that were different in morphology were picked and identified, with some appearing to be entirely novel species. Many of the cultured species were spore-forming Firmicutes, with Actinobacteria and Proteobacteria also being isolated. As well as growing well in low temperature conditions, many species were also found to be halotolerant, and three were actually able to grow even at subzero temperatures (at least –5°C). There are characteristics that it would not have been possible to detect or confirm through culture-independent work alone. The 16S clone libraries were created using PCR amplification from community DNA, and calculated to have ~69% Bacterial coverage. The likelihood of PCR bias was also taken into account when making claims about the estimates of actual abundance in the permafrost. Actinobacteria and Proteobacteria were detected through this method, but Firmicutes were found to be much less abundant than predicted through culturing experiments. Many of the 16S rRNA genes amplified were only ~97% sequence similarity to species published in GenBank, suggesting a high number of novel species within this environment. The clone library contained species related to *Gemmatimonadetes* and *Planctomyces*, both of which had not been previously associated with permafrost samples. It was also possible to draw conclusions about the diversity of the community in permafrost compared to the active layer as even within a small selection of clones more diverse species are detected in the active layer, likely because the permafrost is a more extreme environment and therefore more selective. Comparing the two methods of community analysis, the culture experiments suggested a very different community composition to the 16S clone libraries, although there is the possibility that dead or dormant cells are also included in the clone

library, while they are not in plate culturing. This study is an example of both methods being analysed in tandem for a more comprehensive study of an environment.

Another study that uses both methods to effectively study an environment was carried out by Carrión et al. (2017), when studying the abundance and diversity of MeSH-dependent DMS production in soil environments. Having previously discovered that bacteria possess a novel DMS-production pathway via MeSH, the first gene associated with it (*mddA*) was also characterized, and found it to be in up to 76% of bacteria in terrestrial environments (Carrión et al. 2015). It was present in multiple species that were not previously thought to produce DMS, such as several Actinobacterial and Rhizobiales species. It was therefore important to analyse the functionality of this gene (and therefore the abundance of this pathway) in a range of environments, including terrestrial soil and marine sediment. Carrión et al. (2017) focused specifically on grassland soil for a comprehensive analysis combining both culture-dependent and –independent techniques to study MeSH and DMS cycling in that environment. Soil incubation experiments confirmed that microorganisms within the soil turned the MeSH into DMS but only when addition MeSH was added. This activity was predicted to be bacterial after treatments with either a cocktail of antibiotics reduced DMS production, whereas treatment with cycloheximide, a eukaryotic inhibitor, had no effect. An enrichment of soil samples for MeSH-consuming species through the addition of MeSH was also carried out. This is a widely utilized method that overcomes the difficulty of studying meta-data with low frequencies of genes or species of interest (Schloss & Handelsman 2003), and is often very effective. The enrichments performed by Carrión et al. (2017) showed an increase in the rates of both DMS production and consumption towards the end of the experiment, although only a small proportion of the MeSH added appeared to be turned into DMS. Community DNA and RNA were extracted and analysed using both 16S rRNA amplicon sequencing and metagenomic sequencing.

The 16S amplicon sequencing showed a change in the diversity of the community after enrichment, although not necessarily associated with DMS production. This was because MeSH and DMS are both carbon sources, and MeSH can be synthesized from DMS as well, meaning that the interactions between microorganism and the two substrates are likely more complicated than can be determined in this type of enrichment. Despite not revealing potential bacteria involved in MeSH-dependent DMS production, the 16S amplicon sequencing did show increased abundance of *Methylotenera* where DMS consumption rates were highest, suggesting it is able to degrade it. To test this, cultures of a model strain were tested and demonstrated to be able to consume both MeSH and DMS, the first example of this ability in the Methylophilaceae family. Increases in the abundance of *Massilia* were also observed in both 16S and metagenomic analysis, which

actually does contain *mdd*-like genes, and the culturing experiments identified several microorganisms containing the Mdd pathway, including several that have not been previously expected to contain it (e.g. *Ensifer* and *Sinorhizobium*). This suggests that *mdd* is likely more widespread than originally assumed. Metagenomic analysis on natural soil also revealed that a high percentage of species contain Mdd (35.9%), although this did not appear to increase in the enrichment samples. The work carried out in this Chapter is modelled closely on the methods that were performed in this study, which gave a comprehensive analysis of the cycling of MeSH and DMS in soil environments.

Metagenomic-analysis on community DNA in a study is not limited to only those data generated during the study, as there are already many large publically available datasets to analyse. These were well utilized by Curson et al. (2018) in the study of the interplay between *dsyB* and the newly-identified eukaryotic homolog, *DSYB.* The ocean microbial reference gene catalogue (OM-RGC) metagenomic dataset was mined for the presence of many functional genes including *dsyB, DSYB* and the full suite of *ddd* genes, as well as *recA* (a single copy gene) for normalization, and to enable abundance of the genes to be expressed as percentages of bacteria present. Bacterial *dsyB* was found in 0.35% of species, with only a small number of *DSYB* genes detected, likely in picoeukaryotes as this metagenome was only performed on the < 3 µm fraction. Metatranscriptomes were also analysed, both the *Tara* Ocean metatranscriptomic dataset, sampled across various oceanic locations (solely apportioned to bacteria), and a smaller-scale metatranscriptome project in the North Pacific Ocean (GeoMICS) that contained fractionation allowing for the study of both bacterial and eukaryotic transcription. *dsyB* and *DSYB* transcription was observed in both datasets. *DSYB* was lower than *dsyB* in both in the OM-RGC dataset (abundance) and *Tara* metatranscriptomes (transcript numbers). However, the difference between the two it was much wider in abundance levels, being ~25-fold lower than *dsyB* in metagenomic hits, but only 3-fold lower in transcript numbers, suggesting that it may be more active than *dsyB* in this environment, and therefore likely plays an important role in oceanic DMSP production. The GeoMICS dataset also revealed the expected pattern of higher *DSYB* transcription levels in the larger-sized fraction (2–53 µm), and higher *dsyB* transcription in the smaller-sized fraction (0.2–2 µm) suggesting much about the distribution and diversity of bacterial and eukaryotic DMSP production.

Tools from the studies described above have either been used (clone libraries, see **Chapter 3**) or are about to be reported below (16S amplicon and metagenomic sequencing), as part of the culture-independent bioinformatics analysis of the bacterial contribution to DMSP production by Stiffkey salt marsh.

6.1.4 Chapter Aims

The aim of this chapter is to examine the community and functional genes in the natural Stiffkey sediment, alongside sediment that has been used in a microcosm experiment treated with the enrichment media conditions designed in **Chapter 4**. The 16S rRNA amplicon sequences of the sediment will be analysed, enabling the identification of the bacteria in the samples, not just those that have been cultured previously. These results will be studied in tandem with the metagenomic analysis of the same samples, which can give information on both the abundance of species (based on other markers than the 16S rRNA gene), as well as any other functional genes including *dsyB, mmtN* and all the *ddd* genes denoting DMSP catabolism, as a comparison. Furthermore, the *dsyB* degenerate primers can be used on RNA extracted from all the samples to detect changes in transcription of the gene between them. Sequencing and RT-qPCR will give a more in-depth and unbiased picture of the community of DMSP-producers in Stiffkey salt marsh. Furthermore, some of these techniques were also used on marine sediment samples from the other sites, described in **Chapter 4**, i.e. Yarmouth Estuary and Cley salt marsh samples.

6.2 Methods and Results

6.2.1 Enriching the bacterial community of Stiffkey for DMSP-producers

In **Chapter 4** experiments were carried out in order to determine the enrichment conditions that would optimise for increased numbers of DMSP-producing bacteria within sediment taken from Stiffkey salt marsh. An enrichment (in triplicate) was performed on 3 g of the oxic layer of Stiffkey sediment, with 45 ml of Combination Media MBM (see **Chapter 4**; 50 PSU MBM, 0.5 mM nitrogen, added MTHB) incubated for two weeks at 25°C. Alongside this was a control enrichment which was set up because it is likely that any microcosm experiment will affect the bacterial community even without any selective pressures. This control media was 35 PSU MBM with 12 mM nitrogen. Both media conditions used a mixed carbon source detailed in **Chapter 2.** The DMSP content of both sediments was quantified daily by agitating the sediment in the media for 30 seconds and allowing it to settle for 5 seconds before taking triplicate aliquots of 300 µl into a 1.5 ml microcentrifuge tube. These were centrifuged at maximum speed for one minute and the supernatant removed, 200 µl of which was added to GC vials. The pellet was resuspended in 200 µl sterile water and also transferred into GC vials, and both supernatant and pellet were mixed with 100 µl 10M NaOH and sealed, before incubating overnight in the dark at 30°C. The GC levels of DMSP were quantified and calculated per mg protein (which was measured using Qubit as spectrometry was not sensitive enough) **(Figure 6-1)**.

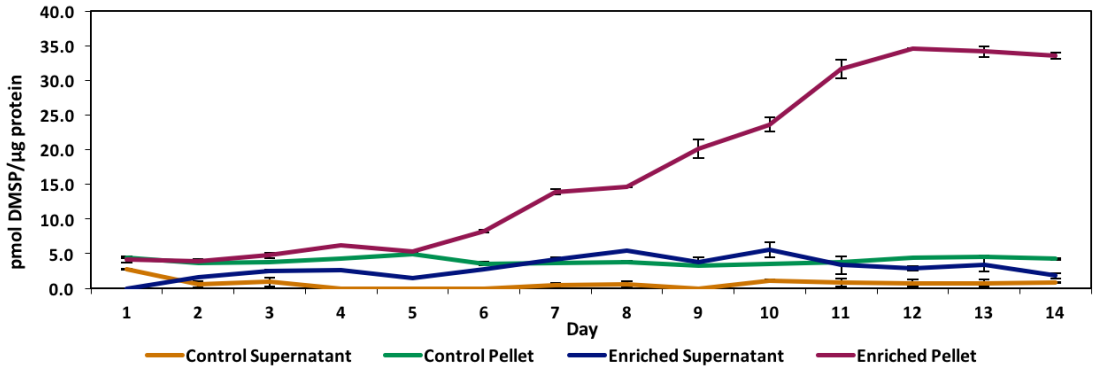

Figure 6-1: The DMSP production after enrichment of Stiffkey sediment in either 'control' or 'enriched' MBM conditions. Both the pellet (DMSPp) and supernatant (DMSPd) are quantified. Samples were in duplicate, error bars denote standard error.

From the microcosm enrichment it was clear that more DMSP was being produced/accumulating in the samples incubated under the 'combination' media, which could be due to either a higher abundance of DMSP-producers in the community, or higher activity by a smaller number of producers. This microcosm experiment was the basis for the work in the rest of this chapter, which focuses on the bacterial community and

functional gene abundance in Stiffkey salt marsh at time 0, and how it changes throughout the enrichment experiment. This allows us to look at the contribution of bacteria not cultured under the conditions used in **Chapter 4**, which could be ~99 % of the bacteria present in the sample.

6.2.2 Extracting DNA and RNA from Stiffkey sediment

Community DNA and RNA were extracted in tandem following the phenol-chloroform extraction method outlined in (Dumont et al, 2011), from 0.5 g of sediment from T0, enriched and control samples. Once extracted the precipitated pellets were washed in 70 % EtOH and air-dried for 5 – 10 minutes, then resuspended in 100 µl of nuclease-free water. Once resuspended, the samples were separated into 50 µl aliquots, with some being stored at -80°C for RNA purification, and the others being stored at –20°C as DNA. The samples were quantified by nanodrop and also run on a 1.5% agarose gel for 20 minutes. This was to determine the presence of RNA in the samples, which is seen as two bands below the larger band of DNA, roughly either side of the 1Kb marker **(Figure 6-2)** (not denoting actual size as dsDNA ladders are not representative of ssRNA), which are the 23S and 16S rRNA bands. If they are clear and not smeared, it suggests a good portion of the RNA is intact and not degraded.

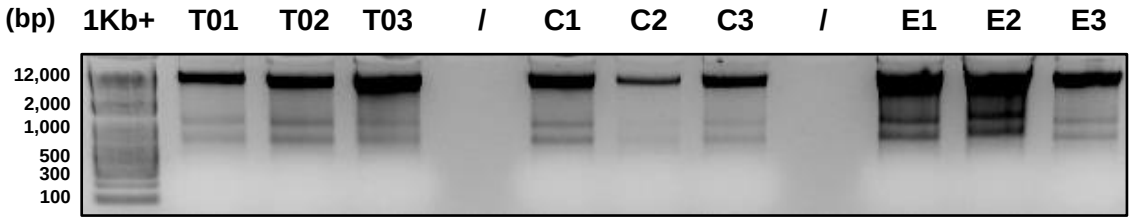

Figure 6-2: Gel electrophoresis of community DNA/RNA extracted from samples of natural Stiffkey sediment (T0), sediment incubated under control conditions (C) and sediment incubated in enriched conditions (E).

The gel electrophoresis demonstrates that, even in samples with seemingly lower concentrations, distinct bands of 23/16S rRNA can be seen, denoting a relatively intact RNA extraction and meaning that RNA analysis can be likely be performed on these samples to good effect.

6.2.3 16S rRNA gene amplicon sequencing

Following the extraction and quantification of DNA/RNA from the environment, the first and most basic method of community analysis is that of 16S rRNA gene amplicon sequencing. This enables the taxonomic identification of bacterial (and plastid) 16S rRNA sequences within an environment, and allows us to see general changes or patterns between the three conditions – if particular species are much more abundant in the enriched samples than in the control and time 0, it suggests that they may play a role in DMSP production in Stiffkey, and have been enriched by the microcosm experiment. The rarerefraction curves give an indication on the depth of bacteria covered within these sequencing experiments. DNA samples were sent in triplicate to MR DNA (Shallowater, TX, USA) and analysed using the 515F/806R primers that amplify the V4 variable region in the 16S rRNA gene. The 515-F primer was barcoded and PCR amplification of the samples was followed by purification using calibrated Ampure XP beads, the products of which were used to prepare an Illumina DNA library which was sequenced on a MiSeq system and processed using the MR DNA analysis pipeline, then checked. The resulting OTUs were identified taxonomically by BLASTn.

Analysis of the results was carried out with help from Brett Wagner (University of Auckland). Sequencing was run and files from the runs were converted to OTU tables and joined in Qiime v1.8. The samples that were not needed, such as those with fewer than 150 bp in size or with ambiguous bases were filtered out. After running preliminary summary statistics on the data, all samples were rarefied to 36,066 sequence counts per sample. The joined tables were then split according to type of sample; time 0, control or enriched. Each group of samples were analysed separately at the genus level, and the genus-level tables and corresponding meta data files were uploaded to a Calypso bioinformatics program (http://cgenome.net/wiki/index.php/Calypso) (Zakrzewski et al. 2016).

Data were normalized using total sum normalisation to convert raw counts to relative abundances. Taxa with less than 0.01% mean relative abundance across all samples were removed. This kept 330 genera, excluding 491 from the original 821 genera. Rarefaction curves were created to demonstrate species richness, with average number of species (richness) plotted against number of reads sampled **(Figure 6-3)**. This curve represents the number of reads that need to be sampled before the diversity of species identified is saturated (shown when the curve plateaus).

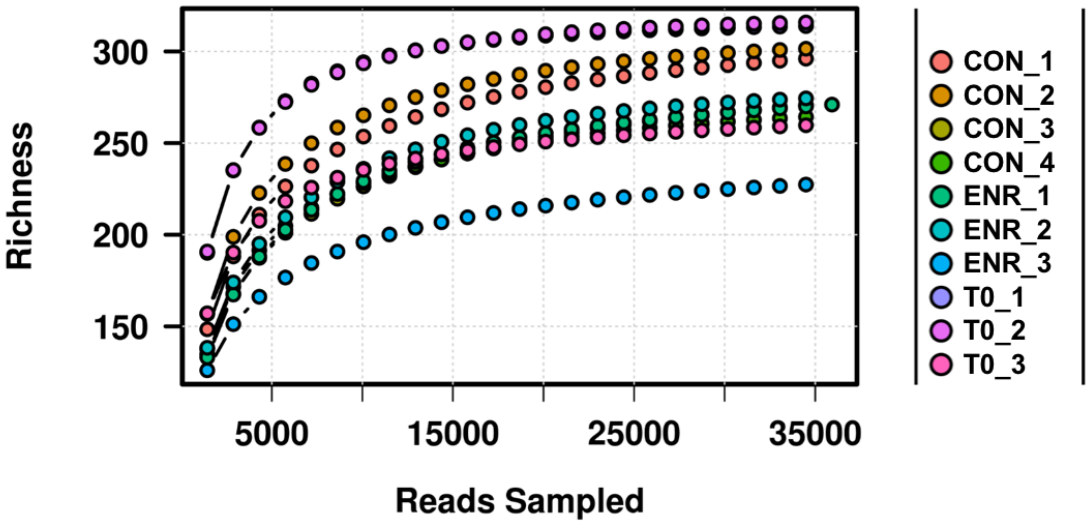

Figure 6-3: Rarefaction plot of all samples rarefied to 36,066 counts per sample. Richness is measured at genus level. All the following analyses were conducted at the rarefied level.

While most of these curves have not fully levelled off, or reached the 'asymptote', suggesting that the sequencing could have been more in depth, they are all almost plateaued, meaning that the sequencing was close to being saturated. From the graph it appears that the ENR_3 sample, while appearing to almost reach sequencing saturation, does not have the same level of species richness compared to the other samples, as it seems unlikely that it would reach the full 330 species no matter how many reads are taken. This could be due to a less in-depth or lower quality sequencing run, but was still worth analysing along with the rest of the samples.

6.2.4 Analysis of the 16S rRNA sequencing before and after enrichment

The actual number of the bacterial species found in these three groups of samples was converted into the relative (proportion) percentage of sequences within each sample that map to the designated taxonomic classification. To keep analysis simple, species that were not found to be above 0.5% abundance in any of the samples were removed and listed as 'others'. The rest were averaged to leave three datasets – one for Time 0, one for Enriched, and one for Control samples, and the 16S rRNA taxonomy was represented in Krona plots – multi-layered pie charts that display the taxonomy of species from domain to genus-level in one plot **(Figure 6-5, Figure 6-4)**.

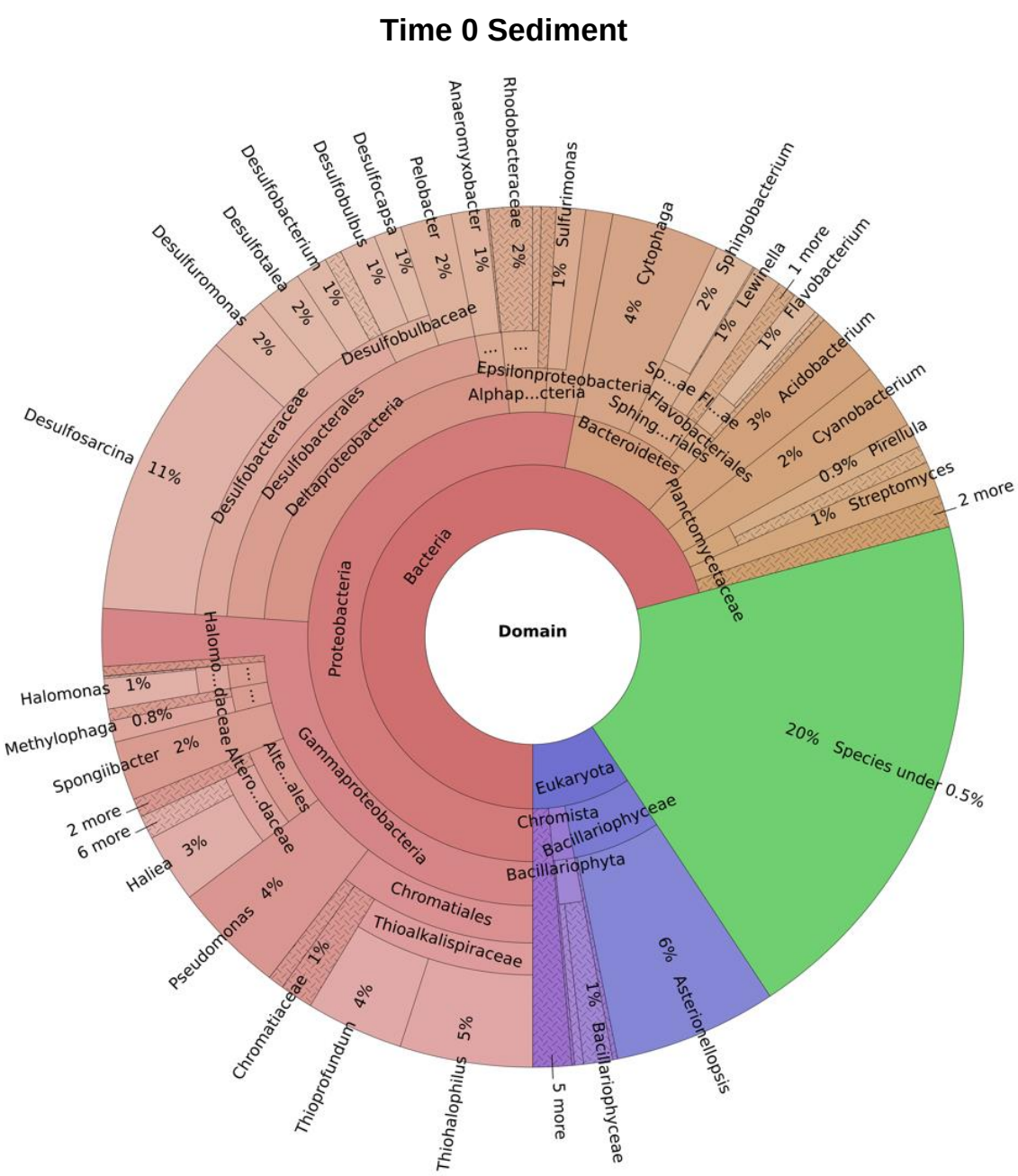

Figure 6-4: Krona plot of the **Time 0** sediment sampled from Stiffkey salt marsh, showing the 16S rRNA gene taxonomy of bacteria in the sample. Plot displays taxonomy of all the species above 0.5% abundance in any of the three samples.

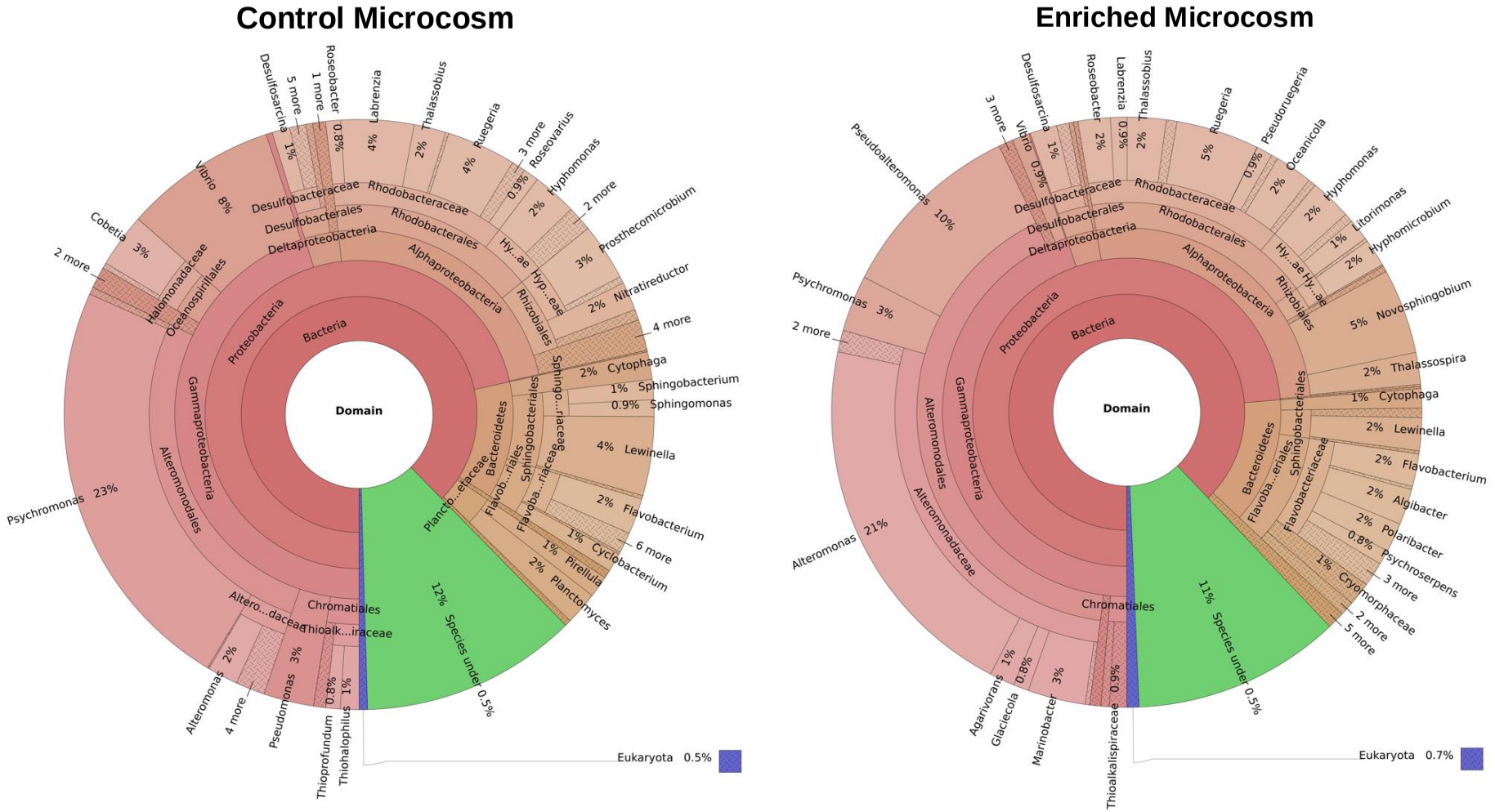

Figure 6-5: Krona plots of the percentage abundance of 16S rRNA gene sequences in DNA, extracted from Stiffkey sediment after incubation for two weeks in either **Control** or **Enriched** conditions. Plots display taxonomy of all species above 0.5% abundance in any of the three samples.

Comparing these Krona plots revealed several interesting changes to the community. An obvious difference is the dramatic decrease in eukaryotes, with the percentages decreasing 10-fold from 9% in Time 0 sediment to 0.5 and 0.7% in the Control and Enriched samples. This was unsurprising, as the media conditions the sediments were incubated in were selective towards heterotrophic bacteria (MBM media), as it was bacteria that we wanted to study. This means that the DMSP increase observed in the Enriched sample is due to the activity of bacteria, not eukaryotes. Furthermore, the percentage of low-abundance species decreased in both the Control and Enriched samples compared to Time 0. This was likely due to the selective pressure of the enrichment, which would mean that some bacterial species either became more abundant, or the most uncultivable ones were lost. Another potential explanation for this is the addition of the mixed carbon source to the incubation experiments. Looking specifically at the bacterial portion of the plots, the taxonomy is dominated by proteobacteria, being 75% of the total bacterial species in Time 0, and increasing to 82% and 84% in the Control and Enriched samples. Previous studies of salt marsh environments have found that proteobacteria are easily isolated from that environment (Ansede et al. 2001), and appear to dominate the culture-dependent method of analysis. Part of the reason for this could be that almost all bacteria linked with DMSP degradation are found to belong to the proteobacterial phylum (Curson, Todd, et al. 2011), but it could also be that a degree of them are also linked to DMSP production. Indeed, the percentage of alphaproteobacteria increases from 3% of total species in Time 0 to 26% in the Enriched, although it also increases to 23% in the Control. Although there is a small (2%) percentage of Actinobacteria (namely *Streptomyces,* which has been shown to contain DMSP-producing species) in the Time 0 sediment, this almost disappears after incubation, again likely due to the incubation conditions being unfavorable. It is unsurprising then, that the majority of the species of interest fall in the proteobacteria group. These include *Alteromonas, Labrenzia, Novosphingobium* and *Thalassospira,* and they are quite dramatically pronounced in the Enriched samples versus the Control and natural samples (Time 0).

In order to look more closely at the species that are likely playing a role in these sediments, the abundance of the 330 species was calculated, and the 50 most abundant genus-level taxa across all the samples were represented in a bubble plot **(Figure 6-6)**. This representation of the sequencing data enabled direct comparison between samples, and was used to observe if there were any major changes in the abundance of particular genes of interest between the three sets of samples. We were specifically looking to see if there was an increased abundance of any species within the enriched sample that was not also enriched in the control, as this would suggest that they may be playing a role in the demonstrated increase of DMSP production seen in that sample **(Figure 6-1)**.

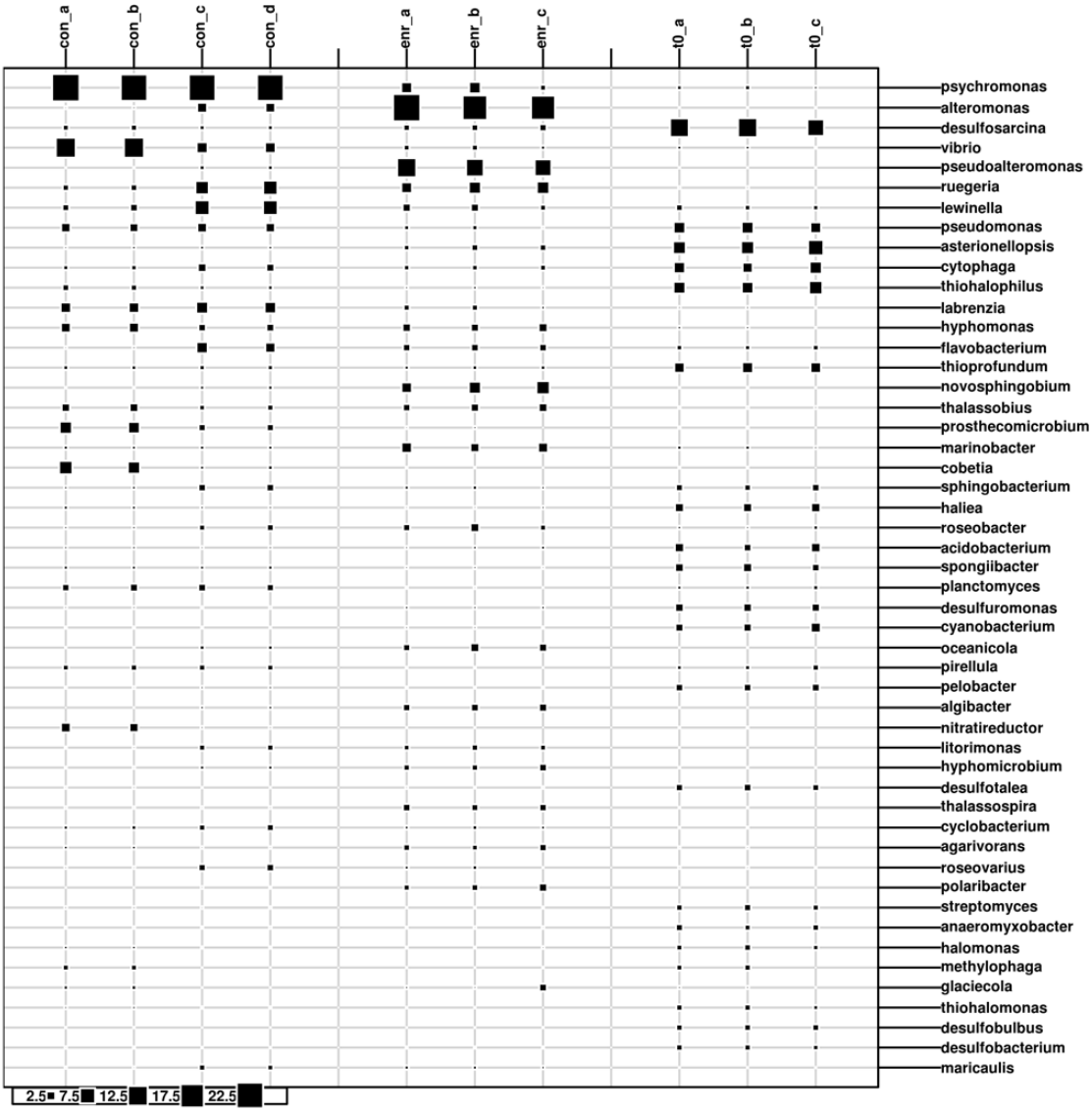

Figure 6-6: A bubble plot of top 50 most abundant genus-level taxa found in all samples, with the size of the box comparative to the percentage abundance of the species within the individual sample.

The percentages discussed below are the averaged results of the replicate sequencing. Firstly, within the three experimental sets dominant bacterial genera in each are *Psychromonas* (23%) in the Control sediment, *Alteromonas* (21%) in the Enriched sediment samples, and *Desulfosarcina* (20%) in the Time 0 sediments. In the Enriched samples *Alteromonas* is closely followed by *Pseudoalteromonas* (10%), and then *Novosphingobium* and *Rugeria* follow with 5% and 6% respectively, making *Novosphingobium* the fourth most abundant species in the Enriched samples. Both *Alteromonas* and *Novosphingobium* are barely visible in the other two sample sets, and have been demonstrated to produce DMSP (see **Chapter 4**), suggesting that they both may play significant roles in the increased production of DMSP by the Enriched samples.

It is no surprise that *Novosphingobium* and *Alteromonas* were also isolated in the culture-dependent work (see **Chapter 4**). Indeed, *Marinobacter* is also present only in significant numbers in the Enriched sample, in addition to *Oceanicola* and *Thalassospira,* all of which represent known DMSP-producing strains. These results go some way to explaining the increase in DMSP levels. More surprisingly, *Labrenzia* was actually higher in the Control samples than the Enriched, with 4% in the Control compared to 0.9% in the Enriched. Without the discovery of *mmtN*-containing species in **Chapter 5,** this 16S abundance would not appear to account for the increased DMSP production at all.

If we follow methods used in work carried out by Curson et al (2017), and use the work carried out in this thesis so far, we can make a degree of assumption about the percentage of DMSP-producing bacteria in the Control, Enriched and Time 0 sediments. These predictions are based on culture-dependent demonstrations of function, alongside estimations based on the sequence homology of both *mmtN* and *dsyB.* These assumptions are used to predict the total percentage abundance of DMSP-producing sequences **(Figure 6-7)**, based on genera that have been shown to contain DMSP-producing species either by isolations, or by the fact that they contain either *dsyB* or *mmtN*.

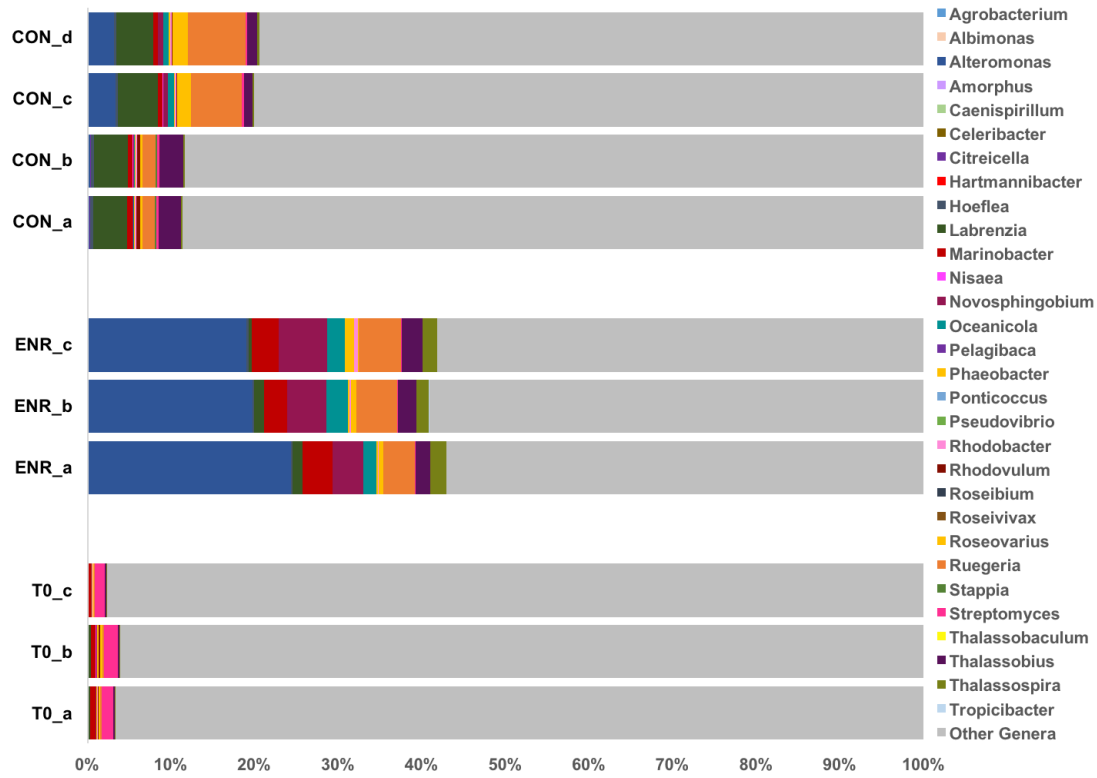

Figure 6-7: The abundance of all the predicted DMSP-producing species within samples from Control, Enriched and Time 0 sediment. Genus-level taxonomy is listed and colour coded.

Looking at **Figure 6-7** it would appear that there is quite a big difference between the bacterial abundance of those we predict to be able to produce DMSP in the Enriched samples, compared to both the Control and Time 0. Indeed, the statistical analysis shows that there is a significant difference between Time 0 and Enriched samples. This seems to explain the differences in DMSP production between the samples, particularly with the increase in *Alteromonas* and *Novosphingobium*. From **Figure 6-7** which shows the hits of each genera as a percentage we can calculate the average number of predicted DMSP producers in each of the sample groups, with Time 0 sediment containing 3.5%, Control containing 14.8%, and the Enriched samples potentially containing up to 42.4% DMSP-producing bacteria. It should be noted that these values are only estimations; there is much room for error, for example, not all genera members will contain *dsyB* or *mmtN*, and there are likely other DMSP synthesis genes or novel pathways to be discovered. However, it does demonstrate the potential role heterotrophic bacteria could be playing in this environment.

Interestingly, although the Control sediment did not show the dramatic increase in DMSP production seen in the Enriched sample **(Figure 6-1)**, it also did not appear to decrease in production either, remaining at roughly 5 pmol/µg protein all the way through. This could be partly explained by the presence of other DMSP producers such as *Labrenzia, Rugeria, Roseovarius* and *Thalassobius* that are as prevalent, if not more so in the case of *Labrenzia*, in the Control compared to the Enriched samples, although there is a fair amount of variation between samples. It is worth noting that these *dsyB*$^{+}$ bacteria are ones that are also known to contain *ddd* genes, thus it is possible that DMSP lyase activity also increases with synthesis, and thus may mask an observed increase in DMSP standing stock in the Control incubations.

It is also important to note that although the presence of these species is potentially indicative of DMSP production taking place, it is not guaranteed, as the activity of those species can often vary quite dramatically, as seen in **Chapter 4 and 5**. This could be another reason why the Enriched DMSP production was so much higher – not only is there a higher abundance of DMSP-producers, but they may also be more active in the low nitrogen conditions and in the presence of the MTHB substrate that is lacking in the other experiments. This could also be the case for Time 0 sediment, which we know has high levels of DMSP (see **Chapter 4**) even though they aren't comparable to the enrichment microcosm DMSP levels as the dilutions of sediment in the media are much higher, and sediment was not weighed out, but was instead resuspended in solution before being spun down and measured as the 'pellet'. Furthermore, protein estimations were not taken of Time 0 samples, as other substances in the sediment make it difficult to measure.

One final aspect of the 16S rRNA gene analysis that needed to be addressed was the significant eukaryotic component of the Time 0 sediment. Even though the analysis of these sites was performed using 16S-specific primers, these can also be used to analyse eukaryotes with plastids in the sediment due to the plastid 16S rRNA sequences that exist in chloroplasts. Both the Krona plot **(Figure 6-4)** and the Bubble plot **(Figure 6-6)** show that the most significant eukaryotic sequence in Time 0 sediment is *Asterionellopsis,* at around 6% of the total hits. Diatoms from this genus have been analysed in several studies, with some suggesting that it doesn't produce DMSP (Keller et al. 1989), and others finding it to produce it at low levels (Speeckaert et al. 2018). Peter Rivera has since isolated a culture of *Astrionellopsis* from Stiffkey sediment that is 99% identical to the sequences amplified in the 16S amplicon sequencing, and shown that it does produce DMSP but at extremely low levels (0.863 fmol per cell). Thus, it is likely that diatoms of this genus contribute to the DMSP standing stocks in the surface Stiffkey sediment. Also at very low abundance are species from the genera *Phaeodactylum* (0.4%), *Thalassiosira* (0.3%) and *Skeletonema* (0.7%), which also produce DMSP, at varying levels that are generally quite low, as diatoms are generally thought to produce relatively low intracellular concentrations of DMSP. It is very likely that there is an element of eukaryotic DMSP production contributing to the overall levels of production, but it is also equally as likely that the bacterial contribution is important. An indication of comparison between the eukaryotic and bacterial DMSP producers can be inferred by the abundance of DMSP-synthesis genes in the corresponding metagenomics data from the Time 0 samples (see below).

6.2.5 Diversity Assay amplicon sequencing of *dsyB*

Having designed degenerate primers in **Chapter 3**, another sequencing experiment was carried out to study the diversity of *dsyB* sequences that can be amplified using the dsyB_deg1F and dsyB_deg2R primers. A targeted diversity assay was set up with Mr DNA and sequencing of the community DNA samples was performed using the same method as the 16S rRNA gene amplicon sequencing, including primer barcoding, PCR amplification and Illumina DNA library MiSeq sequencing, but optimised for the use of the degenerate primer set. Sequences were analysed using Qiime (Caporaso et al. 2010: Macqiime,version 1.9.0) to map the reads to a reference database of 113 known DsyB amino acid sequences at 55% identity to ratified sequences, and the combined OTU table produced was sorted using an ID-mapping file identifying the phylogeny for each sequence. Taxonomy was assigned to an average of 15,128 counts per sample, and count taxonomy was represented as a percentage bar chart **(Figure 6-8).**

The diversity assay below shows that *dsyB* can be amplified from all tested samples. Although all are presented as a percentage, the total number of *dsyB* counts retrieved from the samples varies, with 9,038 counts in Time 0, increasing to 16,504 in

Enriched samples and 16,797 in Control (although this is a biased method of sequencing, and therefore is not entirely representative). There seems to be a significant change in composition of *dsyB* diversity between Time 0 and the microcosm samples. The most dominant *dsyB* sequence in the Time 0 samples appears to be a *Hyphomicrobiaceae* (68%), followed by *Defluuivimonas* (8%), with smaller contributions from *Labrenzia, Phaeobacter, Roseivivax* and *Rhodospirillalles.* Strangely, in the 16S rRNA data *Hyphomicrobiaceae* appears in **(Figure 6-6)** most abundant in Enriched samples, and not in the Control or Time 0 samples at high numbers.

In regards to the low number of *Hyphomicrobiaceae* in the Time 0 sediment, it is important to note that this *dsyB* assay is displaying proportion, and not actual numbers. Furthermore, just because the most abundant hits in the natural samples align most closely to the *dsyB* of *Hyphomicrobiaceae*, this does not necessarily follow that *Hyphomicrobiaceae* is actually the species in that environment containing the particular *dsyB* gene, as we know it can be transferred by horizontal gene transfer (Curson et al. 2018).

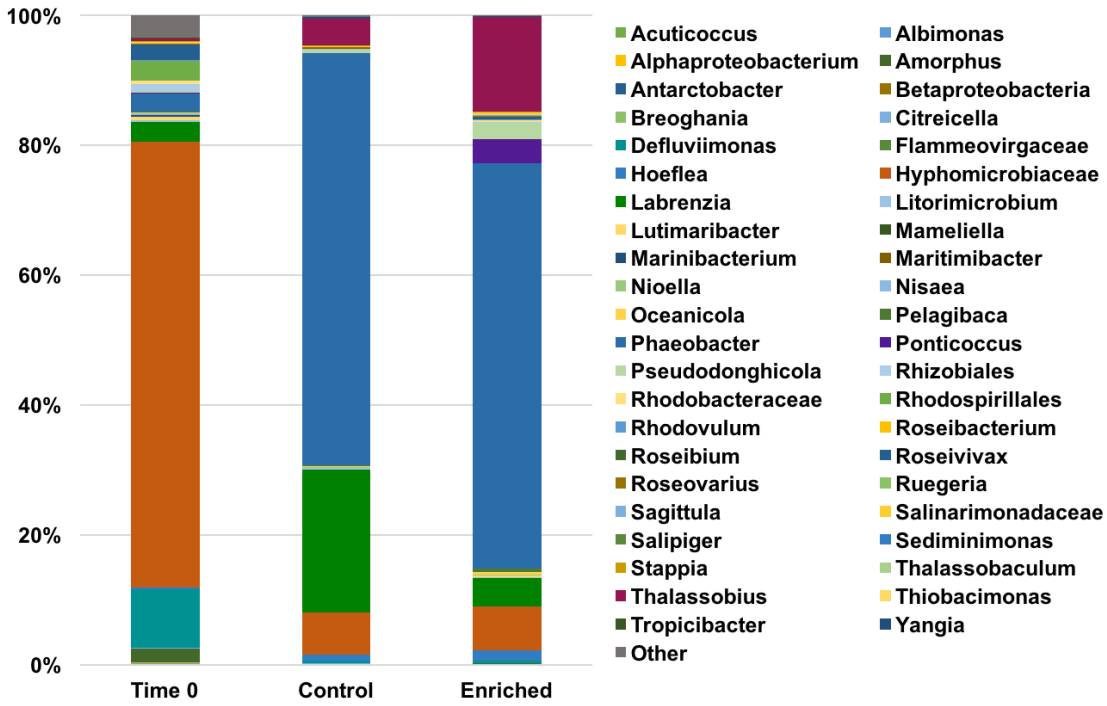

Figure 6-8: The diversity of *dsyB* sequences found within Stiffkey sediment after *dsyB* amplicon sequencing at Time 0 and after incubation in Control or Enriched condition, expressed as percentages.

Control and Enriched samples actually show very similar proportions of *dsyB* diversity, dominated by *Phaeobacter* (~ 60%), although there is a higher proportion of *Labrenzia* in Control samples (21% compared to 4%), which is echoed by the higher abundance of *Labrenzia* seen in the 16S rRNA sequencing **(Figure 6-6)**. There is more variation in *dsyB* diversity in the Enriched samples, with *Thalassobius* (14%), *Ponticoccus* (3.6%) and *Pseudodonghicola* (2.4%) also detected, compared to only 4% *Thalassobius* in the Control incubation, with all other *dsyB* sequences almost undetectable.

As stated in **Chapter 3,** there are limitations to the degenerate primers, with some *dsyB* sequences not amplified as well as others, meaning that there is probable bias towards particular sequences during PCR amplification, and indeed PCR amplification itself also introduces bias. There are likely *dsyB* sequences that are not represented, or that align more closely to those in the Diversity Assay still to be sequenced. However, while the Diversity Assay in **Figure 6-8** is not a full analysis of the *dsyB* diversity within the samples, it certainly adds to the overall picture of the variety in Stiffkey sediment. It will be interesting to look at the abundance and transcription data of *dsyB,* to gauge the importance of this gene and pathway for DMSP synthesis in Stiffkey surface sediment.

Since the initial degenerate primer design and *dsyB* clone library production were performed in **Chapter 3**, many more sequenced homologs of *dsyB* have been published on NCBI and JGI databases –110 have been identified thus far. As a complement to the Diversity Assay described above, the clone library sequences that were created using *dsyB* degenerate primers on sediment sampled from tidal pools at Stiffkey salt marsh (effectively Time 0), were aligned against the full suite of 110 DsyB sequences and represented in a Maximum-likelihood phylogenetic tree **(Figure 6-9).**

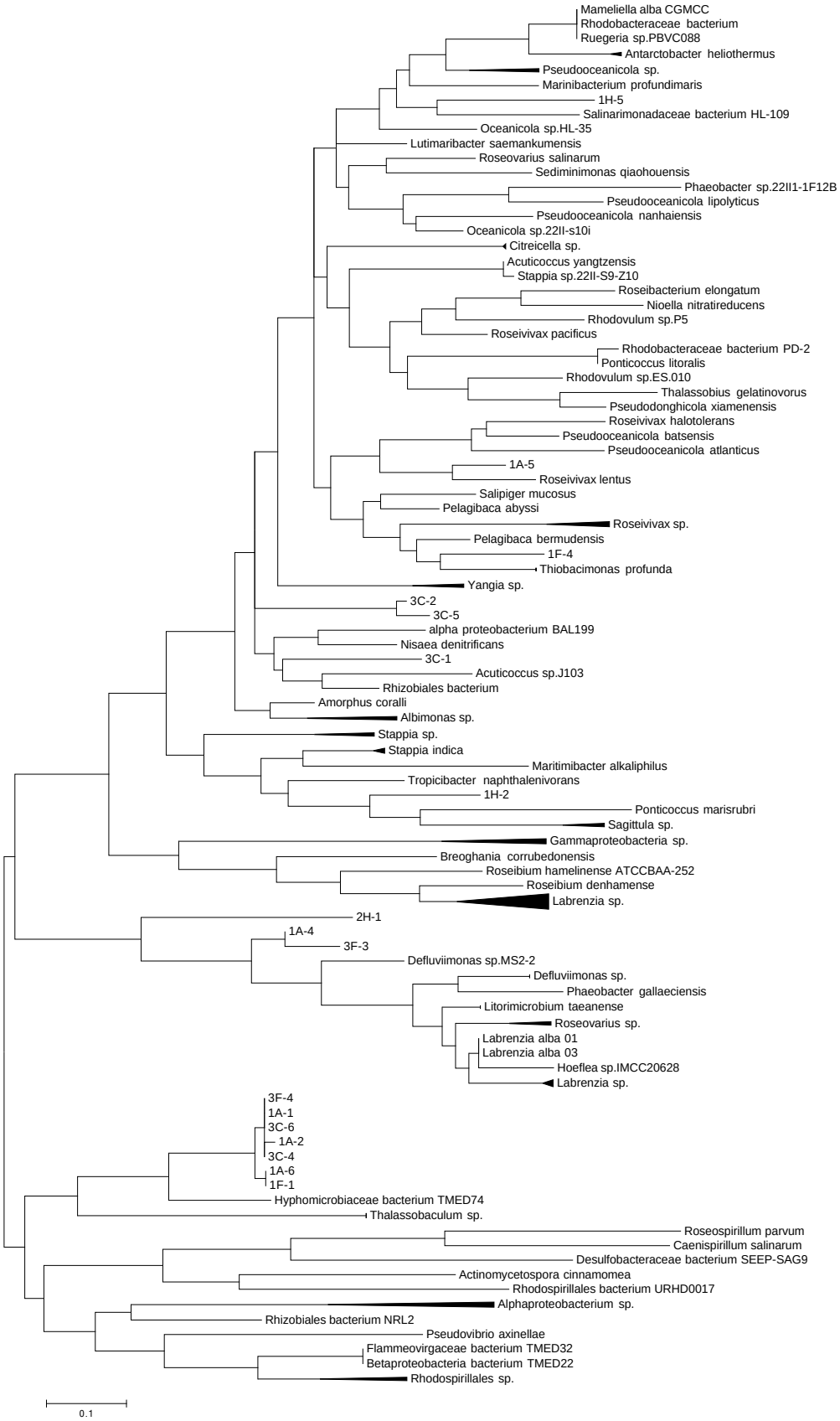

Figure 6-9: Maximum-likelihood phylogenetic tree of all currently known DsyB proteins alongside *dsyB* clone library sequences from Stiffkey, **Chapter 3** (1A – 3F). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site, as indicated on the scale bar.

Despite the fact that the Stiffkey sediment from which the clone library was created was sampled at a different time to the sediment used in the analysis and sequencing in this chapter, the phylogenetic tree reveals many of the same patterns as those observed in **Figure 6-8**. The majority of clones from Stiffkey appear to be most closely aligned to the *Hyphomicrobiaceae dsyB* sequence (7 clones), with several others aligning close to *Defluuivimonas* (3 clones), and also one clone aligning to the *Roseivivax dsyB*. While the closest relative to the group of 7 clones might be *Hyphomicrobiaceae,* the tree also shows that they are not so closely aligned that they could be called identical. This suggests that there are still more *dsyB* sequences to be found.

From this tree it is clear that despite being designed from only 24 sequences, the degenerate primers are able to amplify a broad diversity of *dsyB* sequences from the sediment, with clones spread throughout the DsyB sequences in the tree.

6.2.6 Metagenomic analysis of Stiffkey sediment and enrichments

Another method of sequencing analysis that was performed on the Time 0, Control and Enriched sediment samples was metagenomic sequencing (also in triplicate). Instead of restricting the analysis to the 16S rRNA gene as the previous 16S amplicon sequencing does, the metagenomic sequencing (in theory) covers all the genes in all organisms present in an environment. The only dependent factor is the sequencing depth. As such, metagenomics is an incredibly powerful tool for the study of functional genes in an environment (genetic potential), as well as looking at the abundance of species based on markers other than the 16S rRNA gene identity. Samples for Time 0, Control and Enriched sample groups were combined in equal parts to create pooled samples of the three conditions, in triplicate, on which metagenomic analysis could be performed. This sequencing was also carried out by Mr DNA, Texas, across three separate sequencing runs. Metagenomic sequencing involves creating libraries of DNA that was extracted from the samples, using the Nextera DNA Sample Preparation Kit. Library adapters were incorporated over 5 cycles of PCR. The final library concentration was quantified and average library size was determined, and found to be 826 bp for Time 0 samples, 931 bp for Control samples and 1364 bp for Enriched. Libraries were pooled in equimolar ratios and 10.5 pM of the pool was clustered using the cBot (Illumina) and sequenced paired end for 300 cycles, on the HiSeq system.

Analysis of the metagenomes was carried out with the help of Dr Jennifer Pratscher, and involved the trimming of samples using the Trimmomatic program (Bolger et al. 2014) to obtain ~13 909 226 reads per sample with an average read length of 150 bp. The genome taxonomy within these unassembled metagenomes was analysed using MetaPhlAn (Segata et al. 2012), and represented in **Figure 6-10**.

Assembly was carried out using the SPAdes assembler, with kmers 55 and 127 (Bankevich et al. 2012), and these assemblies were then analysed using Quast (Gurevich et al. 2013). N50 values were ~1 Kb for all the assemblies.

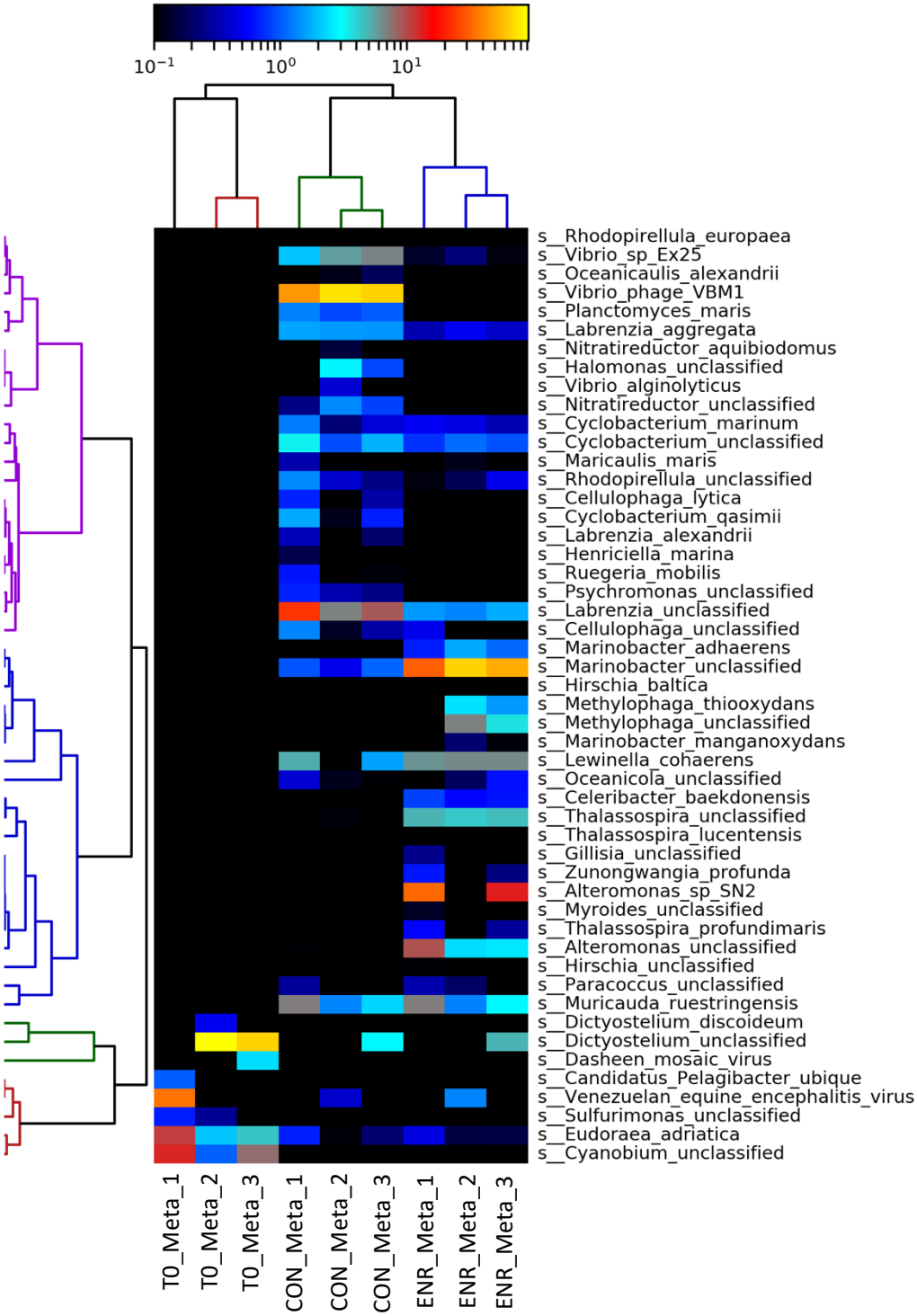

Figure 6-10: Heatmap of the phylogenetic analysis of metagenomes from the Stiffkey sediment and control/enriched microcosms, created using MetaPhlAn. Abundance is represented as a logarithmic scale, reporting the 50 most abundant clades, according to the 90th percentile of the abundance of each clade. Clustering is performed with average linkage, using Bray-Curtis distance for clades and correlation for samples.

There are many similarities between the 16S and metagenomics data in terms of species abundance. The high levels of *Alteromonas* and *Marinobacter*, *Thalassospira* and *Oceanicola* are seen in the Enriched samples, being much more prevalent than in the Control or Time 0 samples, which is entirely consistent with the 16S data analysis. Furthermore, *Labrenzia*, *Psychromonas* and *Vibrio* genera are more abundant in the Control incubations as is also observed in **(Figure 6-6)**. Whilst there is much that is common between the metagenomics and 16S taxonomy data, there are also some notable differences/omissions. Despite being at 5% in the 16S rRNA amplicon sequencing of the Enriched samples, *Novosphingobium* does not appear in the 50 most abundant taxa, according to MetaPhlAn. Similar cases were seen with *Pseudoaletromonas* and *Algibacter,* which appear far more abundant in the enriched samples judged by 16S data, but are not represented in the metagenomics top 50 genera.

Discrepancies such as these may arise from the taxonomic analysis of metagenomics data. For example, 16S rRNA sequencing is highly specific, working from a single gene PCR amplified from each organism, producing longer reads that can be easily assigned at OTU level (if it exists in the RefSeq database). In contrast, MetaPhlAn uses raw reads, which are very short sequences (~150 bp), from a metagenome that contains varying levels of coverage and genome fragments. Raw reads are compared to a genome database, searching for strong markers for each respective genome. Depending on the genetic properties of genomes within the metagenome and database, some species might not have strong markers, or are not represented well enough in the metagenome to match.

For the study of functional genes, analysis was performed on the unassembled, trimmed sequences, as annotation enables identification but is not reflective of the actual abundance within samples. To create peptide databases, the raw reads were translated using the translate function in Sean Eddy's squid package (http://eddylab.org/software.html), generating all open reading frames above 20 amino acids in length. These were mined for functional genes using HMMER (version 3.1b2) at a cutoff of $1e^{-5}$, as described in (Curson et al. 2018).

Hidden Markov Models were created from amino acid alignments of the genes of interest. HMMER finds remote homologs as sensitively as possible, using probability models, and is therefore considered less bias than single-sequence BLAST. Results were dereplicated and manually curated using BLASTp against the RefSeq database, and counted if the top hit was aligned to any of the known sequences. All hits were normalised against read number of the smallest sample and to smallest gene length, and bacterial genes were normalised to *recA* hits, to give the percentage of bacteria containing the functional gene **(Table 6-1).**

Table 6-1: The number of reads of functional genes retrieved from metagenomes of Stiffkey sediment (Time 0, Control and Enriched), normalized to smallest reads and smallest gene length (in brackets, 1dp), with bacterial genes also expressed as a function of *recA*.

Metagenome	**% of bacteria (normalised reads)**											
	DsyB	**MmtN**	**DSYB**	**Alma1**	**DddD**	**DddK**	**DddL**	**DddP**	**DddQ**	**DddW**	**DddY**	**RecA**
T0_Meta_1	1.03 (2.6)	0.18 (0.4)	0	0	1.40 (3.5)	0	5.37 (13.4)	6.80 (17.0)	1.12 (2.8)	0	0.27 (0.7)	(249.3)
T0_Meta_2	1.13 (3.5)	0.07 (0.2)	(0.7)	0	0.85 (2.7)	0.35 (1.1)	5.45 (17.0)	5.37 (16.7)	0.71 (2.2)	0	0.17 (0.5)	(312.0)
T0_Meta_3	0.61 (1.7)	0.14 (0.4)	0	0	1.01 (2.8)	0	3.41 (9.3)	7.68 (21.0)	0.65 (1.8)	0	0 (0.0)	(273.1)
T0_Average	**0.92 (2.6)**	**0.13 (0.3)**	**(0.23)**	**0**	**1.09 (3.0)**	**0.12 (0.4)**	**4.75 (13.2)**	**6.62 (18.2)**	**0.82 (2.3)**	**0**	**0.15 (0.4)**	**(278.2)**
CON_Meta_1	4.99 (21.9)	2.98 (13.1)	0	0	1.76 (7.7)	0	8.75 (38.4)	9.49 (41.6)	1.06 (4.6)	0	0.29 (1.3)	(438.3)
CON_Meta_2	5.11 (21.9)	1.68 (7.2)	0	0	4.16 (17.8)	0	6.40 (27.4)	10.13 (43.4)	0.40 (1.7)	0	0.06 (0.3)	(428.4)
CON_Meta_3	6.30 (28.1)	1.62 (7.2)	0	0	2.08 (9.3)	0	7.22 (32.2)	7.53 (33.7)	1.02 (4.6)	0	0.12 (0.6)	(446.7)
CON_Average	**5.47 (24.0)**	**2.09 (9.2)**	**0**	**0**	**2.67 (11.6)**	**0**	**7.46 (32.7)**	**9.05 (39.6)**	**0.83 (3.6)**	**0**	**0.16 (0.7)**	**(437.8)**
ENR_Meta_1	3.72 (15.7)	3.59 (15.1)	0	(0.2)	2.80 (11.8)	0	6.41 (27.1)	3.32 (14.0)	0.25 (1.1)	0	0.32 (1.4)	(422.0)
ENR_Meta_2	3.68 (16.7)	0.99 (4.5)	0	0	0.93 (4.2)	0	6.32 (28.7)	4.59 (20.8)	0.60 (2.7)	0	0.06 (0.3)	(453.4)
ENR_Meta_3	4.15 (17.1)	2.66 (10.9)	0	(0.2)	2.40 (9.9)	0	6.79 (27.9)	4.37 (18.0)	0.84 (3.5)	0	0.15 (0.6)	(411.5)
ENR_Average	**3.85 (16.5)**	**2.41 (10.2)**	**0**	**(0.1)**	**2.04 (8.6)**	**0**	**6.51 (27.9)**	**4.09 (17.6)**	**0.56 (2.4)**	**0**	**0.18 (0.8)**	**(429.0)**

Interestingly, previous studies on the presence of *dsyB* in large marine metagenomes (*Tara* and GOS) predicted that around 0.5% of bacterial species contain it (Curson et al. 2017). In comparison, the percentage predicated in this study is almost double in this salt marsh metagenome, with an abundance of 0.92% in the Time 0 sediment. Together with the 0.13% predicted to contain *mmtN* in the Time 0 samples, we can estimate that the genetic potential to produce DMSP *in situ* exists in a minimum of 1.05% of bacteria in the marine salt marsh sediment. This value is not so dissimilar to the predictions made in **(Figure 6-7)**, of ~ 3% of species in the natural Time 0 sediment being of genera linked to DMSP production (including isolates in which neither gene has yet been found). Given that not all representatives of these genera likely carry out this process, we feel 1.05% is a realistic value. In comparison to known DMSP catabolic genes, the *dsyB* gene is more abundant than most DMSP lyase genes, the exceptions being *dddD*, *dddL* and *dddP* (present in ~1.90, 4.75 and 6.62 % of bacteria respectively), which are likely important in DMS production in these sediments. The number of hits to the eukaryotic DMSP-synthesis gene *DSYB* were very low, even at Time 0 where 9% of the 16S rRNA hits were apportioned to eukaryotes, suggesting that there is little eukaryotic DMSP production taking place in the salt marsh surface sediment. Alternatively, it could be that there are other DSYB isoform enzymes and/or novel DMSP synthesis pathways in diatoms such as *Astrionellopsis*, which would also be interesting to investigate.

The abundance of both *dsyB* and *mmtN* increase after microcosm experiments, although the increase does little to explain the increase in DMSP production by Enriched samples, because the levels are very similar between these two microcosm sample groups **(Figure 6-1).** For the Control samples *dsyB* increases to 5.47%, with *mmtN* at 2.09%, in comparison to the Enriched samples where *dsyB* only increases to 3.85%, and *mmtN* to 2.41%, which is slightly higher than the Control, but likely not enough to account for the dramatic DMSP increase. It is important to note that abundance does not necessarily result in activity, and it is therefore important to also consider RNA (and therefore transcription) as well, either through the use of RT-qPCR (see below), or through more comprehensive methods such as metatranscriptomics.

As it was not possible, due to time constraints, to perform diversity assays on MmtN like the ones that were carried out on DsyB **(Figure 6-8),** instead, the identity of the top hit for each of the MmtN homologs detected in each metagenome **(Table 6-1)** were recorded when the hits were being manually curated using BLASTp, and the abundance of each genus within each metagenome is listed in **Table 6-2.** This was to produce a rough representation of the diversity of *mmtN* sequences within the Stiffkey sediment samples.

The table below shows that while the identities of MmtN in Time 0 samples vary between each of the replicates, the most abundant MmtN sequence (at least half of the total hits) in the Enriched metagenomes closely aligns to *Novosphingobium* in all three of the replicates, always followed by *Thalassospira* as the second most abundant. In the Control samples *Labrenzia* is consistently the most abundant, with a variety of different MmtN sequences also appearing, including some more unusual sequences such as *Croceicoccus* and *Saccharothrix.* Both observations seem to support conclusions drawn from the 16S rRNA sequencing that shows higher abundance of *Novosphingobium* in the Enriched samples compared to the Control, where *Labrenzia* is more abundant **(Figure 6-6)**.

Table 6-2. The identity of the closest MmtN homologs to sequences extracted from Stiffkey metagenomes (Time 0, Control and Enriched).

Metagenome	MmtN identity	Number of hits
T0_Meta_1	*Thalassospira*	2
T0_Meta_2	*Micromonospora*	1
T0_Meta_3	*Nocardiopsis*	1
CON_Meta_1	*Labrenzia*	29
	Novosphingobium	8
	Thalassospira	2
	Croceicoccus	1
CON_Meta_2	*Labrenzia*	16
	Thalassospira	4
CON_Meta_3	*Labrenzia*	19
	Novosphingobium	7
	Thalassospira	1
	Croceicoccus	1
	Saccharothrix	1
	Rhodobacter	1
ENR_Meta_1	*Novosphingobium*	48
	Thalassospira	15
	Labrenzia	4
	Rhodobacter	1
ENR_Meta_2	*Novosphingobium*	7
	Thalassospira	4
	Labrenzia	2
ENR_Meta_2	*Novosphingobium*	27
	Thalassospira	7
	Labrenzia	5
	Rhodobacter	1

6.2.7 Designing degenerate primers for *mmtN*

As part of the culture-independent analysis of bacterial DMSP-production in Stiffkey sediment, it was decided that degenerate primers should also be designed to *mmtN*, for use in the screening of new libraries of isolates, qPCR on DNA and cDNA to assess gene abundance and transcription, and for the production of clone libraries or other Diversity Assay sequencing.

Degenerate primers to the *mmtN* gene were designed following the same method as the *dsyB* primers in **Chapter 3.** There are currently 23 species we predict contain functional MmtN homologs, including two Actinobacteria (*Nocardiopsis chromatogenes* and *Streptomyces mobaraensis*). Both the amino acid and nucleotide sequences were aligned using ClustalW **(Figure 6-11),** including a more divergent, supposed non-functional *Candidatus Taylorbacteria bacterium* that has only 30% identity to MmtN. When the amino acid sequences were aligned, there were several regions that were reasonably well conserved, but the nucleotide sequences were more divergent. Degenerate primers should not contain more than five degenerate bases or else they become too degenerate and there is too much non-specific amplification. However, looking at the nucleotide alignments, there were more than five divergent nucleotides. This meant that any primer set designed would likely have one or two mismatches to a number of the sequences, even with five degenerate bases. Several sets were designed with little amplification, and even the most successful pair **(Table 6-3)** was not able to amplify from the two species of Actinobacteria **(Figure 6-12)**.

Table 6-3: The oligonucleotide sequences for degenerate primers, designed from two conserved regions of the 23 MmtN amino acid sequences.

Primer	Sequence	GC content	Melting temperature (°C)
mmtN_degF	GGCAGYGAYCTYGAYCCSCG	60	65.5
mmtN_degR	CCAVGGRTARTARTGSGC	44	56.3

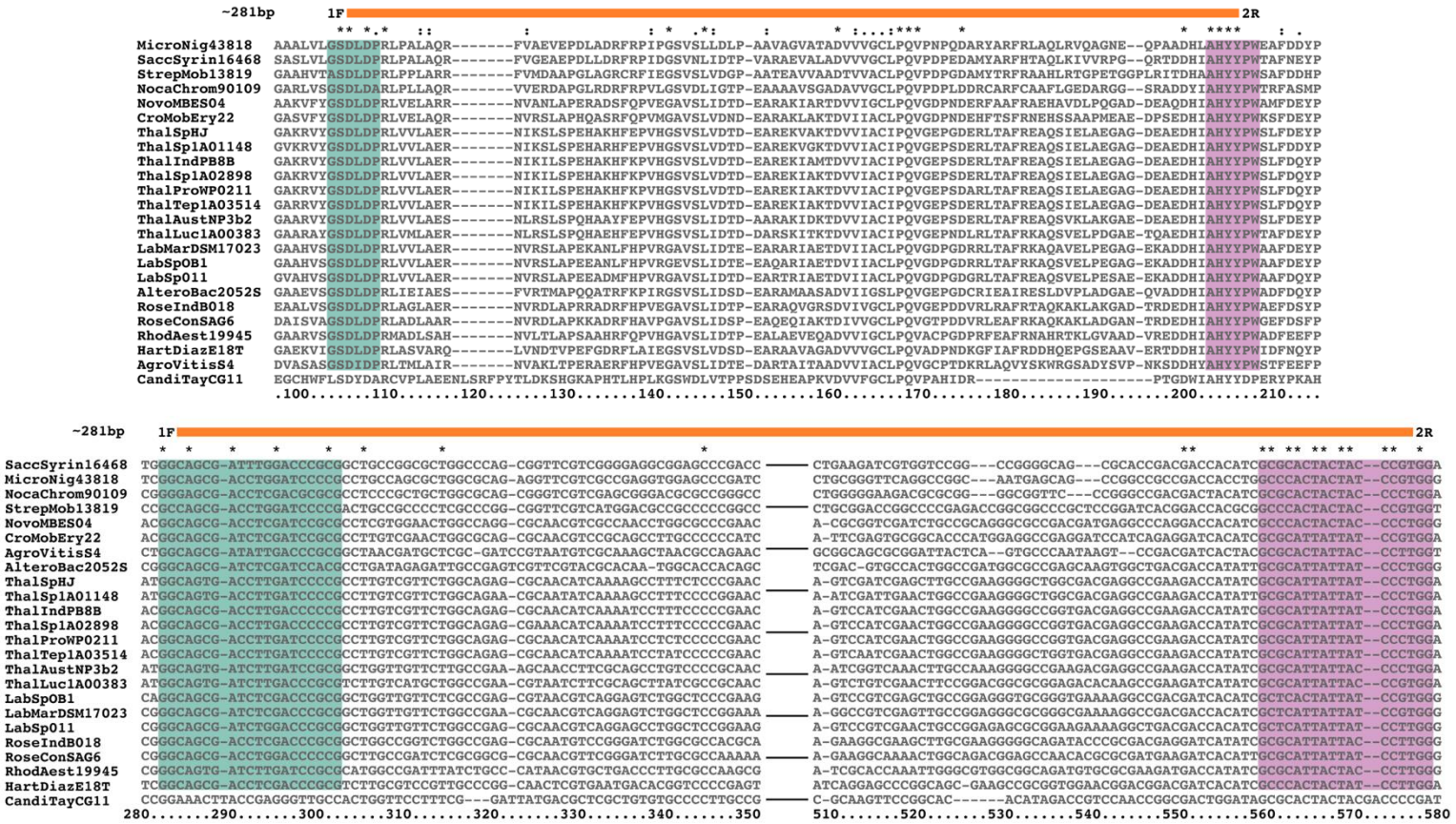

Figure 6-11: A section of the amino acid alignment of 23 MmtN sequences above a section of the nucleotide alignment of *mmtN,* both aligned using ClustalW, alongside a non-functional sequence (CandiTayCG11). Fully conserved amino acids or nucleotides are marked by an asterix (*), closely similar amino acids are marked by two dots (**:**), and less similar amino acids are marked by a single dot (**.**). Two primers with a maximum of 5 degenerate bases are shown; mmtN_degF (teal) and mmtN_degR (purple). The fragment produced is ~281 bp,. Only the first and last 70 bp of the nucleotide region amplified by the primers is shown.

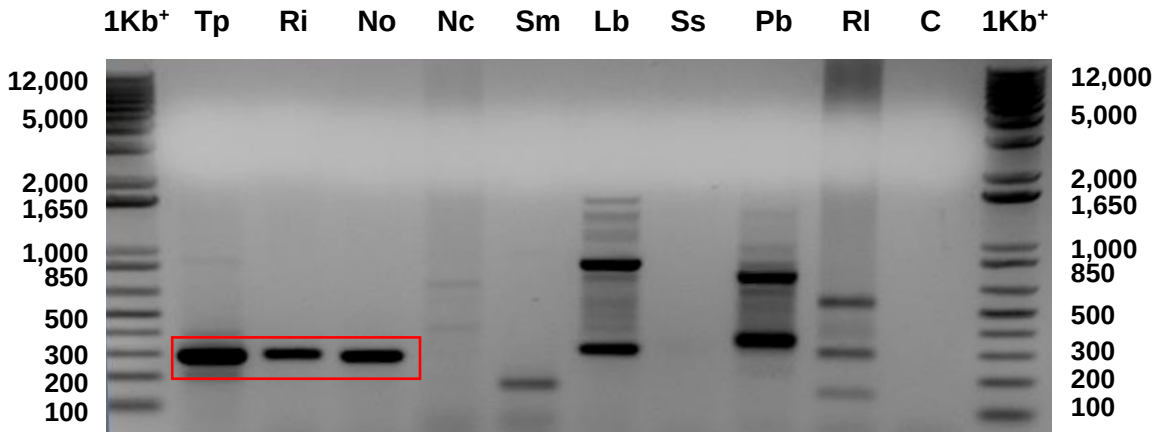

Figure 6-12: Gel electrophoresis with a 1Kb Plus ladder, showing the optimised PCR amplification of *mmtN* from genomic DNA using degenerate primers on five positive controls *T. profundimaris* (Tp), *R, indicus* (Ri)*, Novosphinbobium* (No), *N. chromatogenes* (Nc) and *S. mobaraensis* (Sm), and four negative controls *L. aggregata* (Lb), *S. stellata* (Ss), *P. bermudensis* (Pb) and *R. leguminosarum* (Rl), as well as a water control (C). This amplification was carried out using the primer set mmtN_degF and mmtN_degR, amplifying a ~281bp fragment (the red box).

The optimised program for these primers had an initial denaturation step of 95°C for 5 minutes, followed by 35 cycles of 95°C for 30 seconds, an annealing step of 55°C for 30 seconds and an elongation step of 72°C for 30 seconds, ending in a final extension of 72°C for 5 minutes **(Figure 6-12)**. There was good amplification from the three alphaproteobacterial strains tested, with bands excised and sequenced and confirmed to be *mmtN*, but despite trialling multiple conditions and dilutions of genomic DNA, it was not possible to amplify from either *Streptomyces* or *Nocardiopsis*. It would seem that attempting to incorporate all the sequences under one set of primers may not be an option for degenerate *mmtN* primers, as they are a more divergent group of sequences than *dsyB*. Therefore, the next step will be to design clade-specific degenerate primers instead. Unfortunately, this has not been accomplished in time for this submission, but will be an important piece of work in the future. The primers were still tested for amplification from community DNA from the environment and for use in qPCR amplification, but amplification did not occur from environmental DNA, and although amplification was observed in qPCR, when clone libraries were created from the amplified products, they were not *mmtN*. Therefore, this primer design still needs optimisation.

6.2.8 qPCR analysis of Time 0 and Control/Enriched samples

As previously mentioned, it is important to pair abundance analysis with analysis of the transcription that is taking place in an environment, as the presence of a gene is not necessarily indicative of the activity of said gene in every condition. This is why the

transcription of functional genes should also be measured, through the RNA produced in the samples, as this gives a more realistic estimate of the actual contribution of a gene to its environment. As metatranscriptomic analysis is expensive and time-consuming, it was not a viable option for this study, although it is a method to consider in future analysis of this type. Therefore, qPCR was chosen as a preliminary method to study the DNA/RNA content of Stiffkey salt marsh and the microcosm experiments.

Unfortunately, the *mmtN* primers were not functional in qPCR, so only *dsyB* copies and transcripts were able to be analysed. The protocol for performing this analysis on DNA/RNA was previously established in **Chapter 3**, with DNA and RNA extracted in tandem, and visualised by gel electrophoresis **(Figure 6-2)**. DNA samples were stored at -20°C and used as templates for qPCR in dilutions of either 1/10 (for Time 0 samples) or 1/100 (for microcosm samples) **(Figure 6-13)**. RNA was purified and quantified, and concentrations were normalised so that as close to 100 ng as possible was used in reverse transcription experiments. These took place using gene-specific primers, namely dsyB_deg2R, and the cDNA was quantified and stored at -20°C, then used as the template for RT-qPCR **(Figure 6-14)**.

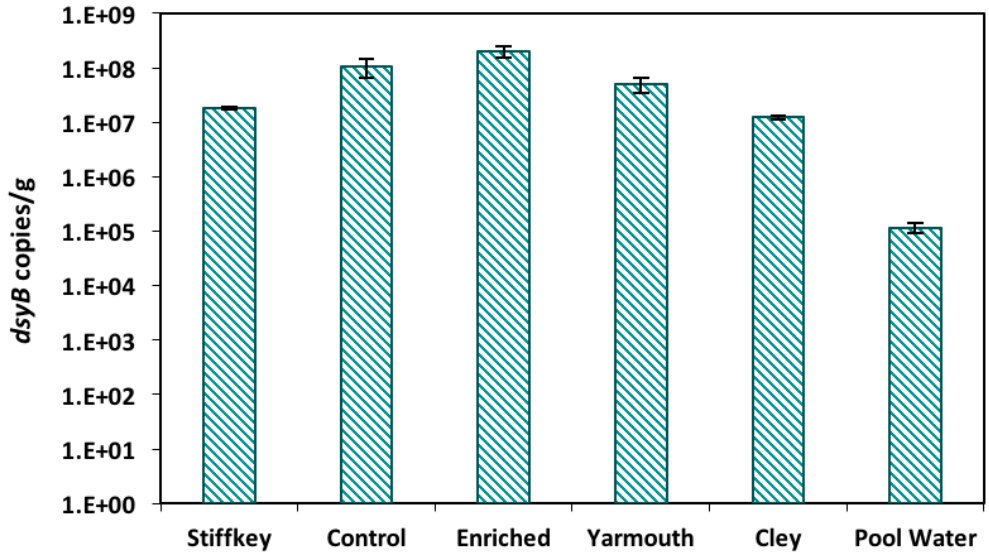

Figure 6-13: A semi-logarithmic plot of the abundance (copies/g or copies/ml) of the functional gene *dsyB,* amplified using qPCR from Stiffkey, Cley Yarmouth sediment, Control and Enriched microcosm experiments, and Stiffkey Pool Water. *dsyB* was amplified using the degenerate primers dsyB_deg1F and dsyB_deg2R. Samples are the average of triplicate data with error bars indicating standard error of the means.

qPCR on *dsyB* and 16S rRNA genes was performed on Yarmouth and Cley sediment (again), Time 0 Stiffkey sediment, the Control and Enriched incubations, and also Pool water sampled from the same tidal pools at Stiffkey salt marsh (see **Chapter 2**). This was to compare the bacterial contribution in sediment and water, as the water column is typically considered the major environment in which DMSP production occurs, particularly surface waters (Bates 1994), likely because they are the main environment in which eukaryotic species dwell. It would therefore be interesting to compare *dsyB* abundance and transcription between these two environments. All qPCR experiments were performed in triplicate (biological and technical), and clone libraries of the products are currently being sequenced.

The abundance of *dsyB* increased in both Control and Enriched samples compared to all the Time 0 sediments **(Figure 6-13)**, with Enriched *dsyB* abundance appearing to be slightly higher than the Control, which was surprising as it seemed contrary to most other evidence previously discussed. However, once the percentage of bacteria containing *dsyB* was calculated, using the 16S rRNA abundance results from qPCR, it was seen that 1.74% of the bacteria in the Control sediment contained *dsyB*, compared to 0.67% in Enriched samples. Although these percentages are lower than those predicted by metagenomic analysis **(Table 6-1)**, they appear to be in a similar proportion to each other, with Control sediment containing roughly twice the percentage of *dsyB* species compared to Enriched samples.

The abundances of Time 0, Cley and Yarmouth sediments were all significantly higher than the abundance of *dsyB* in the Pool Water sample, which is unsurprising considering that species are more dispersed in water than sediment. When the percentages of *dsyB*-containing species were calculated, 0.21% of bacteria in Stiffkey sediment possessed *dsyB*, in Yarmouth *dsyB* is predicted to be in 0.23% of species, and in Cley it is thought to be 0.1%, all of which were shown to be lower than the percentages in Pool water, which is predicted to contain 0.54% *dsyB*-possessing bacteria.

Finally, the cDNA produced by all these samples was calculated **(Figure 6-14)**. Interestingly, *dsyB* activity appears to be roughly the same between Time 0, Control and Enriched samples, with all of them producing similar numbers of transcript copies. This is a much more even balance compared to the abundance of *dsyB*, which confirms that the activity of a gene is not necessarily linked to function within that environment.

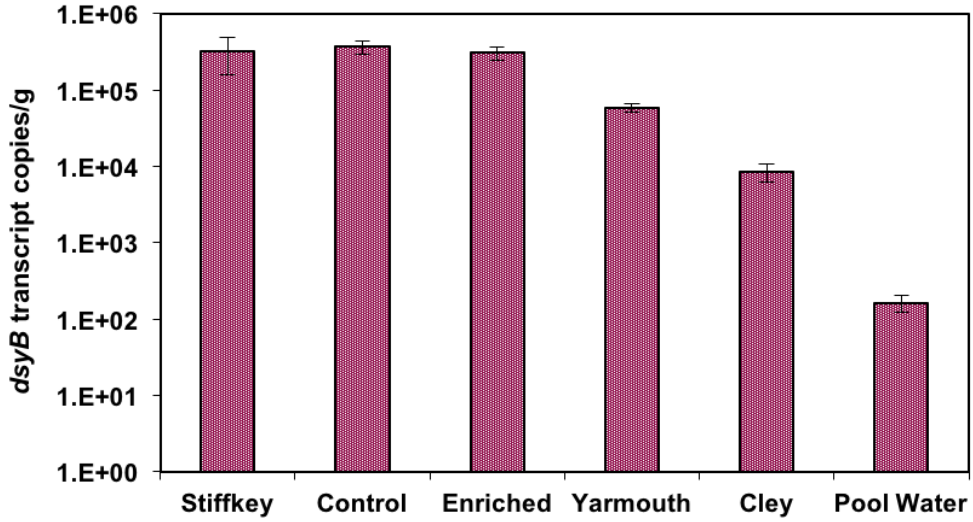

Figure 6-14: A semi-logarithmic plot of transcription levels (transcript copies/g or copies/ml) of the functional gene *dsyB,* using specific primer cDNA from Stiffkey, Cley and Yarmouth sediment, Control and Enriched microcosm experiments, and Stiffkey Pool Water. Primers used were dsyB_deg1F and dsyB_deg2R. Samples are the average of triplicate data with error bars indicating the standard error of the means.

Stiffkey sediment seems to be higher in activity compared to Yarmouth and Cley, which could be due to the slightly higher salinity levels of Stiffkey compared to Yarmouth, and the fact that the area of Stiffkey that was sampled is fully tidal, being submerged twice a day, which means that it maintains a continually high level of salinity (35-40 PSU). In comparison, Cley was sampled much further from the water's edge, and therefore the salinity of the sediment varies much more depending on rainfall and flooding (Silvestri et al, 2005). In comparison to all tested sediment samples, the pool water samples showed very low transcript levels, at least three orders of magnitude lower than Stiffkey Time 0 sediment, which was as expected with so little DMSP production (see **Chapter 4**) and such low *dsyB* abundance. These much reduced transcript levels are in keeping with the ~3 orders-of-magnitude higher levels of DMSP standing stock in the sediments compared to the pool water.

6.2.9 Mining for *dsyB* and *mmtN* in global metagenomes/transcriptomes

One final bioinformatics tool that was utilised in the study of bacterial contributions to DMSP production in marine environments was the mining for *mmtN* in large ocean metagenomes and metatranscriptomes, using the techniques described Curson et al. (2018), namely creating Hidden Markov Models, using hmm search to identify all possible sequences, and then manually curating them. The first search for *mmtN* was in the ocean microbial reference gene catalogue (OM-RGC) **(Table 7-4),** which was a large-scale metagenomic dataset, sampling at a multitude of depths and locations (Sungawa et al,

2015). This database was generated by sequencing the < 3 µm fraction, meaning that many of the eukaryotic sequences were likely removed.

From the work carried out by Curson et al (2018), it was predicted that *dsyB* exists in roughly 0.35 % of bacterial species in the marine environment **(Table 6-4)**, which is almost a third of the size of the percentage predicted from the metagenomes of Stiffkey Time 0 sediment **(Table 6-1)**. This pattern is also seen for the percentages of *mmtN,* with a predicted 0.03 % containing *mmtN* in the OM-RGC database, compared to 0.13 % in Stiffkey metagenomes. Conversely, many of the *ddd* genes are at a higher abundance in the marine metagenomes than in the sediment, with *dddP* predicted to be in 12.53 % of species in the OM-RGC but only in 6.62 % in Stiffkey, and *dddD* with 5.56 % compared to 1.09 %. The only genes that are lower in abundance than *mmtN* in the OM-RGC database are *DSYB* and *dddW*.

The *Tara* Oceans metatranscriptomes were also mined in order to detect the transcription of MmtN in the oceans. MmtN was only detected in 11 of the 26 samples analysed in Curson et al (2018), so only those samples are reported below **(Table 6-5),** although the calculations of transcripts per million sequences was still calculated from the total number of samples. It was found that *mmtN* transcript abundance is far less than *dsyB* and all the *ddd* genes apart from *dddW* and *Alma1*. In addition to the selection of *Tara* Ocean metatranscriptomes that were analysed in Curson et al (2018), there were also a number of other samples in which MmtN transcription seemed more abundant, with an average of 3.7 transcripts per sample, and some samples containing transcripts as high as 14 and 23 hits. These are not fully reported in this work as the full analysis of the other DMSP-synthesis and –catabolic genes has not yet been performed, but it suggests that while still being lower than *dsyB* abundance, it is at least transcribed in most environments, being detected in 48 of the 59 samples available online.

It is not possible to compare the transcription of *dsyB* in Stiffkey to the *Tara* metatranscriptomes, but both show that *dsyB* is transcribed in both environments, and it is likely that *mmtN* does the same, although we have not been able to perform experiments to demonstrate this yet.

Table 6-4: Metagenome information and results of MmtN mining, with published hits for DsyB and other genes, Curson et al (2018)

Project	Biome	Location	Total genes	Number of sequences											Estimated % of bacteria with DsyB	Estimated % of bacteria with MmtN
				MmtN	DsyB	DSYB	DddD	DddK	DddL	DddP	DddQ	DddW	DddY	Alma1	RecA		
OM-RGC	Marine	Various	40,154,882	18	230	9	3,660	425	96	8,242	845	3	28	60	65,727	0.35	0.03

Table 6-5: The samples of the *Tara* Oceans metatranscriptomes apportioned to bacteria in which MmtN was detected, alongside published transcripts abundances for DsyB and other genes in marine seawater samples (Curson et al 2018)

Sample ID	Depth (m)	Location, Latitude. Longitude	Total sequences	Number of transcripts										
				MmtN	DsyB	DSYB	DddD	DddK	DddL	DddP	DddQ	DddW	DddY	Alma1
ERR598953	177	South Pacific Ocean, -12.9794, -96.0232	23,584,453	2	17	0	1	5	24	3599	122	0	6	0
ERR598956	150	South Pacific Ocean, -8.9109, -140.2845	24,881,437	1	49	1	16	1	2	3227	30	0	2	0
ERR598976	5	North Atlantic Ocean, 36.1715, -29.0230	102,893,943	2	29	38	22	448	12	7775	829	0	21	0
ERR598977	60	Arabian Sea, 14.5536, 70.0128	25,301,577	3	7	0	1	8	2	1759	69	0	0	0
ERR598988	120	South Pacific Ocean, -9.0714, -140.5973	24,203,323	2	136	1	5	12	2	2491	113	0	0	0
ERR598999	600	South Pacific Ocean, 8.9729, -139.2393	27,948,130	1	17	1	35	1	11	3788	77	0	1	0
ERR599004	450	North Pacific Ocean, 6.3599, -103.0598	23,015,118	1	29	0	33	1	11	2586	134	0	0	0
ERR599010	5	South Atlantic Ocean, -20.9354, -35.1803	19,979,067	1	14	19	9	8	18	517	56	0	4	0
ERR599011	5	Arabian Sea, 14.6059, 69.9776	19,967,844	3	4	0	6	0	3	1506	109	0	1	0
ERR599113	50	South Pacific Ocean, -12.9723, -96.0122	14,927,388	2	29	0	3	24	4	1630	86	0	1	0
ERR599169	5	South Pacific Ocean, -13.0023, -95.9759	22,146,064	1	1	0	7	9	2	2525	325	0	0	0
	Total *Tara* sequences/transcripts (all samples)		**681,752,622**	**19**	**332**	**60**	**138**	**517**	**91**	**31403**	**1950**	**0**	**36**	**0**
	Total transcripts per million sequences			**0.027**	**0.691**	**0.261**	**0.994**	**5.32**	**0.220**	**89.0**	**5.13**	**0**	**0.090**	**0.0001**

6.3 Discussion

6.3.1 Summary of work

The work carried out in this chapter was the culture-independent counterpart to the culture-dependent work previously carried out in **Chapter 4**, and also builds off the discoveries made in **Chapter 5** where the novel DMSP-producing gene, *mmtN*, was identified. This chapter focused on the bioinformatic analysis of 16S rRNA amplicon and metagenomic sequencing of three treatments of Stiffkey surface sediment. The first was the natural Time 0 sediment with no treatment, the second and third were from microcosm experiments where they were incubated under conditions that either enriched for DMSP production within the environment (Enriched), or were standard growth conditions as a control for the effect that incubating sediment will have on natural sediment (Control). Also performed were a *dsyB* amplicon Diversity Assay, qPCR experiments and mining for *mmtN* in publically available metagenome and metatranscriptome datasets.

The culture-independent work was carried out to look at the bacterial contribution to DMSP-production in the natural Stiffkey salt marsh sediment through identifying known DMSP-producers in the sample, alongside the abundance of the two DMSP-synthesis genes, *dsyB* and *mmtN*, as well as their diversity and transcription (where possible). This analysis was also used on the two microcosm sediment groups, Enriched and Control, to observe differences between them and the Time 0 sediment in the abundance of DMSP-producers or the functional genes themselves. This was because there were increased levels of DMSP production in the Enriched sediment compared to the Control, suggesting that any differences in the bioinformatic characterization of the two could be linked to DMSP production.

6.3.2 Problems associated with various methods used

Although there are numerous strengths to the use of culture-independent experiments, many of which are described in the introduction, there are also issues with various aspects of some of these techniques. For instance, the metagenomic analysis that was carried out by Mr DNA did not contain as full a coverage as it could have – the metagenomes produced were only 2 or 3 GB in size, whereas many companies now offer up to 6 GB of sequencing data, which could have revealed a greater abundance and diversity of the genes and species of interest than is currently described. The 16S rRNA amplicon sequencing appeared to have good coverage, with rarefaction curves seeming almost horizontal in some samples, which means that we can have greater confidence in the analysis performed on it. However, any form of sequencing that involves PCR amplification in library preparation (such as 16S amplicon, metagenomic sequencing and gene-specific Diversity Assays) will also contain a degree of bias. This is because PCR

amplification routinely leads to the under-representation of sequences with extreme base compositions (GC content) (Aird et al. 2011), and most DNA polymerases actually introduce errors through base substitution roughly every $10^5 – 10^6$ bases (Cline et al. 1996). Furthermore, the 16S amplification genes used, 515F/806R, do not necessarily amplify all species equally, meaning that certain clades may be represented less than they are in the actual samples (Walters et al. 2015). To determine if particular species of interest are indeed in the Enriched or Time 0 samples, such as the *Novosphingobium* that appears in 16S sequencing but not in the metagenomes, it may be worth designing specific primers to those species and using either qPCR or ddPCR to confirm their presence. Droplet digital PCR (ddPCR) is a variation of PCR that separates the solution into discrete, defined water-in-oil droplets in which the PCR takes place (Pinheiro et al. 2012). This method enables more reliable and sensitive measurement of nucleic acid amounts, making it useful in the study of variations in gene sequences, and is potentially a method that could be utilized alongside qPCR, as it gives an absolute quantification of fluorescence by the number of positive droplets observed, as opposed to the intensity of fluorescence, although it only gives end-point data. As previously discussed, there is also the issue with multiple 16S rRNA genes existing within most species, but for the amplicon analysis it is all represented as relative abundance, and therefore is corrected for as much as possible, although it assumes that they all have roughly the same number of copies.

Bias can be introduced at several other stages of the process as well, from the choice of DNA extraction methods to the actual sequencing stage, where high cluster densities on the flow-cells used in Illumina sequencing can suppress GC-rich reads. It can even vary depending on the sequencing centre used (Schloss et al. 2011). Analysis of metagenomes can also be skewed depending on the programs used to clean and assemble it, and even gene abundance could be bias depending on the size of the different genomes in the samples (Beszteri et al. 2010)

The weaknesses of qPCR have previously been described in the discussion of **Chapter 3**, so there is little to add here, other than the need for primer optimisation. As mentioned several times throughout this thesis, the *dsyB* degenerate primers dsyB_deg1F and dsyB_deg2R are not perfect. The primer efficiency, although acceptable, is lower than it could be, and we already know that they do not amplify all the known *dsyB* sequences – they have been demonstrated to not amplify certain already identified sequences, and there are also undoubtedly many others that have not been published yet. Future work could involve the redesigning of these primers, with alignments that utilize the increased size of the database of *dsyB* sequences. There could also be more time spent on producing qPCR-suitable primers so that more confidence could be placed on the results. However, although these primers are not ideal, it does mean that we can be confident that

all analysis linked to these primers is likely an underestimation of the true abundance or transcription of *dsyB* in any given environment.

It was also unfortunate that, while the *mmtN* degenerate primers could amplify from genomic DNA, they were not able amplify from community DNA, and therefore were not utilized in qPCR experiments that could have revealed much about the transcription of *mmtN* in the Enriched and Control samples. Degenerate primers have been an important tool in this analysis so far, so it would be beneficial to properly design *mmtN* degenerate primers that can be used in the same way as the *dsyB* ones.

6.3.3 Culture-independent analysis of the bacterial contribution

The analysis of the community and functional gene abundance within the samples described above revealed many interesting results. Most importantly, through the metagenomic sequencing both *dsyB* and *mmtN* are confirmed to exist in the natural Stiffkey sediment. Indeed, it appears to have over twice the percentage of *dsyB*-containing species compared to ocean-based datasets, as well as several species with the potential to produce DMSP including *Streptomyces* and *Marinobacter* species. In addition to this, RNA extracted from the sediment contains enough *dsyB* sequences for the reverse transcription using the degenerate primer to amplify a product, compared to water controls. This confirms that *dsyB* transcription takes place in the natural environment, which we assume leads to DsyB enzyme activity, thus enabling bacteria to synthesise DMSP. This activity is also much higher than that observed in samples taken from the Pool water at Stiffkey, which has dramatically lower levels of DMSP production (see **Chapter 4**) compared to the sediment. The evidence is compelling that bacterial DMSP production takes place in the natural surface sediment taken from Stiffkey salt marsh. The fact that lower values for *dsyB* and *mmtN* are observed in the ocean OM-RGC and *Tara* datasets compared to the sediment samples suggests that while algae are important DMSP producers, especially in the euphotic section of the water column, bacteria are likely key producers of DMSP in salt marsh environments. Although the analysis is mostly performed on sediment from Stiffkey, we also have strong evidence to suggest similar levels of *dsyB* abundance in other salt marsh environments, through the qPCR experiments.

In regards to the Enriched and Control sediments, the picture is less clear. Although the Enriched sample appeared to contain a much higher number of DMSP-producers compared to the other samples, the presence of *dsyB*containing species predicted by metagenomic analysis suggests otherwise, being higher in the Control metagenomes than in the Enriched. Even though *mmtN* abundance is slightly higher in Enriched samples than in the Control, it does not seem likely that this would account for the huge increase in DMSP production, although it could be possible, as several of the mmtN-containing

species analysed so far have the potential to produce large amounts of DMSP. As both the diversity of *dsyB* sequences within the samples and the *dsyB* activity by transcripts amplified by RT-qPCR were almost the same, it suggests that the real difference between these two samples may be in the activity of *mmtN*, which has yet to be quantified, or else through the activity of an entirely unknown species, utilizing an unknown gene involved in the DMSP-synthesis pathway.

There are several steps to be taken in order to complete the study of bacterial DMSP production in Stiffkey sediment. Firstly, the degenerate primers to *mmtN* should be optimised through redesigning and testing more PCR conditions so that they are able to amplify from community DNA, and are also qPCR-compatible, and then used to create another Diversity Assay to show the gene variation within Stiffkey sediment as well as in qPCR and RT-qPCR experiments to confirm copy number and transcription. Once these two primer sets have been designed they can both be used on other sediment samples, as has been done already on samples from Cley and Yarmouth, which would allow claims made about Stiffkey to be applied to a wider range of environments.

Furthermore, doubtless there are still several publically available metagenomic and perhaps metatranscriptomic datasets that could be mined for all the genes of interest. It would be particularly interesting if some of those datasets came from sediment environments.

There is another, less biased method of studying gene activity that should be considered for future analysis of this type, namely the use of metatranscriptomic sequencing. This measures the community RNA extracted from a sediment sample, and would give the most unbiased estimate of the transcription of both *dsyB* and *mmtN.* However, even transcription does not always result in gene activity, as translation does not always take place after a gene has been transcribed. Therefore, to truly study the expression of these two genes in the environment, proteomics or even metaproteomics should be considered as another option.

6.3.4 Concluding Remarks

DMSP-producing bacteria and their *dsyB* and/or *mmtN* transcripts were present in Stiffkey, Cley, Yarmouth and all tested seawater samples and *Tara* Oceans bacterioplankton datasets. It seems that *dsyB* and possibly *mmtN* are far more abundant in marine surface sediment compared to ocean environments. Furthermore, DMSP synthesis rates have been found to be higher in surface sediment samples than seawater

samples (work carried out by Andy Hind, not included). Thus, it seems probable that surface marine sediments are environments with high DMSP productivity, and that heterotrophic bacteria are likely important producers in these environments. Nevertheless, it is also possible that diatoms, like bacteria, are important DMSP producers in these Stiffkey pond surface sediments, and likely other photic surface marine sediments as well.

CHAPTER 7

DISCUSSION AND CONCLUDING REMARKS

7 DISCUSSION AND CONCLUDING REMARKS

7.1 Aims and research gaps

DMSP is an environmentally important molecule in marine environments with several petagrams predicted to be produced by Earth's surface oceans (Ksionzek et al. 2016). DMSP impacts nutrient supply, atmospheric chemistry signalling and sulfur cycling (Kiene et al. 2000). Endogenously, DMSP is purported to play several protective roles against conditions of stress such as high salinity, low temperatures and/or oxidative stress. Since the discovery that the production of this molecule is not restricted to marine eukaryotes, and does in fact take place in heterotrophic bacteria as well (Curson et al. 2017), many assumptions that had previously been made about its distribution, function and source had to be called into question. The possible bacterial contribution to global levels of DMSP is completely disregarded, and because the habitats in which eukaryotes are able to grow are limited, it has therefore limited the environments in which DMSP production has been studied in. The work carried out in this thesis was predominantly aimed at attempting to address this information deficit, through setting a precedent for studying the role that bacteria play in DMSP production, in any environment. The work was roughly divided into several different avenues of study:

1. Determine the diversity and abundance of *dsyB,* the first known bacterial DMSP-production gene, in the environment, both in metagenomes and in bacteria isolated from that environment.
2. Use culture-independent methods to observe the importance of bacterial DMSP synthesis in Stiffkey salt marsh.
3. Identify key bacterial DMSP producers and determine the means by which DMSP is synthesised in bacteria.

It was important to cover all these aspects of DMSP production, so that we can improve our understanding of the mechanics and distribution of this environmentally important compound. The work was designed to analyse bacterial DMSP production on several different levels, from the wider picture of the bacterial community, looking at how diversity of species changes under conditions designed to increase DMSP production, and determining the community potential for DMSP production, down to the abundance of specific functional genes (such as *dsyB*) in that environment, and even more specifically to studying the role of DMSP production in just a single strain, using genetic manipulations to analyse it. Although different combinations of these methods have been used to study aspects of DMSP cycling previously, this body of work is the first comprehensive,

exploratory investigation into bacterial DMSP production in a salt marsh environment. Indeed, many protocols that have been developed and optimised, including the process-based incubation experiments and use of mixed carbon sources with bacterial culturing can easily be utilised in the study DMSP-producing bacteria in any other environment, steadily adding to our knowledge on how widespread and significant this ability might be.

From the work performed by Curson et al (2017) and (2018), we knew that the reporter gene for bacterial DMSP production, *dsyB*, exists in a large number of alphaproteobacterial species, as well as in large marine ocean metagenome and metatranscriptome datasets (OM-RGC and *Tara* Ocean), suggesting that it is not only present but also transcribed under marine conditions. This work expanded upon this foundation, further quantifying *dsyB* diversity and abundance in a specific environment, namely Stiffkey salt marsh, as well as searching for non-*dsyB* containing species that produce DMSP.

7.2 Major findings described in this thesis

7.2.1 *dsyB* degenerate primers reveal diversity of sequences in Stiffkey sediment, confirmed by Amplicon sequencing

Degenerate primer gene probes were designed from alignments of known DsyB sequences, and were utilised in the study of the diversity, abundance and transcription of *dsyB* in unidentified bacterial isolates and community DNA from marine sediments (**Chapter 3).** The primers were useful as a preliminary screen for the presence of *dsyB* species in isolates cultured from Stiffkey, but were even more important in the amplification of *dsyB* DNA and cDNA (from mRNA) in qPCR and RT-qPCR experiments. Although the abundance of *dsyB*-containing bacteria in Stiffkey was predicted to be quite low compared to that calculated from metagenomic sequencing (**Chapter 6**) (0.21 % by qPCR, 0.92 % by metagenome), this is partly because the copy numbers of *dsyB* were normalised for % bacteria using 16S rRNA copy numbers, which are far from accurate owing to the intragenomic heterogeneity resulting in some species having many more copies compared to others (Sun et al. 2013). Furthermore, the primers are not all-encompassing in terms of sequencing amplification, and have been shown not to amplify from genomes that contain *dsyB*. However, this does mean that we can assume that although the numbers of both abundance and transcript qPCR demonstrate that *dsyB* is present and likely transcribed in Stiffkey, they are in fact underestimations of the true value, with transcripts perhaps being missed due to primer bias. In future work, it would be interesting to attempt to re-design *dsyB* degenerate primers now that many more sequences have been discovered

and to specifically optimise them for qPCR to ensure that as many copies or transcripts are being amplified as possible from community DNA.

In terms of diversity, the degenerate primers were used to create a clone library when there were only ~24 ratified sequences available. When the clones were sequenced and phylogeny was displayed in a tree, it was clear that apart from one large grouping of clones, the majority were widely spaced throughout the tree. When this alignment was repeated against the latest species on NCBI and JGI datasets, a while after the original tree was made, there were over 100 sequences to include. This more comprehensive phylogenetic tree, despite having many more sequences, still showed the same high level of diversity in clones. When the community DNA was sequenced using the custom *dsyB*-primer amplicon experiment (**Chapter 6**), it confirmed that, while being dominated by *Hyphomicrobiaceae,* the rest of the sequences were quite varied in DsyB sequences at Time 0 natural sediment. Indeed, this *Hyphomicrobiaceae* DsyB was the most closely related sequence to the largest cluster of clones from the Stiffkey sediment library, making it a potential species of interest in further study of DMSP production in this environment.

7.2.2 *Spartina* transect and pool water quantification show that bacteria likely play an important role in DMSP production

Spartina species such as *Spartina alternifloa* and *Spartina anglica* (the species studied from Stiffkey salt marsh) have long been considered the sole reason for the high DMSP levels detected in salt marshes (Steudler & Peterson 1984; Kocsis et al. 1998). While rough experiments performed on these plants do suggest very high endogenous DMSP concentrations, it was shown through transect of sediment that as samples were taken from sites moving away from the *Spartina* plants, DMSP levels were originally very high, with a decrease within a distance of 20 cm, after which the sediment DMSP content almost seems to stablise, maintaining a mostly constant level of DMSP production (**Chapter 4**). We propose that this could be due to the activity of bacterial/algal DMSP-synthesis taking over from the DMSP leached from the plants, as it is unlikely to have diffused that far from the plants, and would not suddenly stabilise. It should be noted that the edges of the pools were always covered in algal matts and we propose that these impose a very significant contribution to the highest DMSP levels seen in the sediment closest to the *Spartina*. For some reason these algae were not visibly prominent anywhere else except the edges. Furthermore, it should also be noted that *Spartina* like any organism producing DMSP goes to a great deal of effort to make the molecule, energetic cost, thus it is unlikely that it gives away a precious resource too easily. Thus, it is more likely that in the samples close to the *Spartina* that some root material may have been included in the sediment samples, which is contributing to the highest observed DMSP levels. It would be

very interesting in the future to look at the potential interactions between Spartina, and DMSP-producing bacteria in the rhizosphere and phylosphere of these plants.

Although there is a ~9 % abundance of eukaryotic sequences in salt marsh sediment, shown through 16S rRNA analysis to be 6% from the *Astrionellopsis* genus (**Chapter 6**), it is still thought that the DMSP content of the sediment is still a mostly bacterial domain. The levels published for *Astrionellopsis sp.* DMSP production were very varied (Keller et al. 1989; Speeckaert et al. 2018), so instead of relying on previously published data to estimate the impact that this eukaryote may have on overall DMSP levels, as, we isolated a strain of *Astrionellopsis* from the actual sediment from Stiffkey. From analysis performed on that eukaryotic isolate, which revealed a very low level of DMSP-synthesis activity, we hypothesise that it is unlikely to contribute significantly compared to bacteria. This hypothesis is supported by the fact that *dsyB* and to a lesser extent *mmtN* were far more abundant in our metagenomics analysis of the natural samples. Of course it could be that these diatoms produce DMSP via an unknown DMSP synthesis pathway with novel genes. Nevertheless theer are very high numbers of DMSP-producing bacteria in these sediments so they must play a significant role in the production of DMSP especially considering *dsyB* transcripts were detected.

When they were both sampled, Stiffkey sediment and the overlying seawater (Pool Water) were found to differ in DMSP content significantly (**Chapter 4**). DMSP content (nmol/g) of sediment was found to be >2 orders-of-magnitude higher than the DMSP (nmol/ml) of Pool Water. This was to be expected in a sample with such a high density of species, as opposed to water samples where bacteria are much more dispersed. RT-qPCR on cDNA from Stiffkey sediment and pool water also mirrored this, with the presence of *dsyB* transcripts being detected at a much higher abundance in the sediment than in the water samples, suggesting that *dsyB* activity is much more pronounced in marine sediment compared to the ocean. From analysis of the metagenomic analysis of Stiffkey and the global ocean dataset, a similar pattern emerges, with both *dsyB* and *mmtN* abundance higher in salt marsh sediment compared to the water. It is possible that with *mmtN* transcription this is even more pronounced, although this has yet to be tested. From this data we predict that marine sediments are environments of high DMSP productivity, much more than seawater which is perceived as the hub for DMSP synthesis. It will be interesting in the future to test more varied marine sediments and to study the effects of pressure, and oxygen of DMSP production. It was apparent that DMSP production does not stop in the oxic sediment zones.

7.2.3 A range of DMSP-producing isolates can be cultured from Stiffkey

From the degenerate primer design and clone library construction in **Chapter 3**, it was clear that there was a source of DMSP-producing bacteria in Stiffkey natural sediment. It was found that when plating natural sediment on agar with no selection pressure, at least 1 in 4 colonies could produce DMSP, and this number increased to 77 % of colonies picked when sediment was incubated under different conditions such as high salinity, low nitrogen and added MTHB (**Chapter 4**). The identification of some of these bacteria revealed species of *Labrenzia*, and the closely related *Stappia,* as well as *Pseudooceanicola*, all of which were known to be *dsyB*-containing species. There were also a number of isolates of genera that had not previously been shown to produce DMSP or contain *dsyB* (*Marinobacter, Novosphingobium,* and *Alteromonas*). Further characterisation and whole genome sequencing of those isolates not containing *dsyB* (through degenerate primer PCR) revealed that the three most unusual isolates did not contain *dsyB,* despite being able to produce DMSP. This was predicted to be because they either have a different isoform of DsyB that carries out the same process, despite being a different protein, or else they were using an entirely new gene, likely as part of a different production pathway. These experiments show that there is undoubtedly a more interesting, complex story to be told, and bacterial DMSP production is likely not only more widespread, but also more varied than previously thought.

7.2.4 *Novosphingobium* contains a novel DMSP-producing gene, *mmtN*

Gene discovery is an avenue of research in which the Todd lab excel, having identified all of the currently known genes involved in DMSP catabolism (*ddd),* with the exception of the eukaryotic *Alma1*, as well as the first bacterial DMSP-synthesis gene, and its eukaryotic counterpart (*dsyB/DSYB*) (Todd et al. 2009; Curson, Sullivan, et al. 2011; Todd et al. 2012; Curson et al. 2017). It was through the use of these well-established experimental procedures for gene discovery, including genomic library construction and screening that the second bacterial DMSP-synthesis gene, termed *mmtN*, was identified (**Chapter 5**).

After screening the genomic library of a DMSP-producing *Novosphingobium* and sequencing the positive fragments, one gene in particular was looked at as a candidate for DMSP production in *Novosphingobium.* This was because its amino acid sequence showed ~30 % similarity to the SAM-dependent MMT that was in *Arabidopsis thaliana* (Ranocha et al. 2000). Although this plant does not produce DMSP, it and many other angiosperms produce SMM from Met. That SAM-dependent methyltransferase is able to methylate Met, creating SMM, so this activity was looked for in *Novosphingobium* by cloning the *mmtN* gene into *E. coli* BL21, and detecting DMS production after heated

alkaline hydrolysis, where without *mmtN* there is no activity. The peak of DMS could have arisen from either SMM or DMSP (or both), although it is most likely SMM. Very recent LC-MS work has just verified that the reaction product of the pure MmtN enzyme with S-AdoMet and Met is indeed SMM (Simone Payne and Ana Bermejo Marttinez). When *mmtN* is probed against protein databases it was found to exist in ~24 strains (currently). In contrast to *dsyB*, which almost exclusively exists in alphaproteobacterial species, homologs of MmtN were found in a range of classes, including alphaproteobacteria, actinobacteria and one gammaproteobacterium. A selection of these *mmtN*-containing species were ordered so that they could be tested for DMSP production, and their *mmtN* genes were cloned and tested in the same way as *mmtN* was for *Novosphingobium*. The fact that these activity assays were functional in *E. coli* was interesting, as with *dsyB* the cloned gene had to be mobilised into the wide-range host *R. leguminosaurum* before it would function. Presumably these means that MmtN does not require any strange/uncommon co-factors that *E. coli* does not import or produce. It could also reflect the more varied identities of *mmtN*-containing strains compared to those containing *dsyB.* Following the demonstration that the *mmtN* gene can confer MMT activity to a bacteria that lacks this ability, the protein was overexpressed and purified to measure enzyme activity, through the conversion of S-AdoMet to S-AdoHyc, which can be detected by HPLC. The optimum conditions for enzyme activity and K_m values were determined, and then the MmtN protein was tested under optimal conditions with multiple compounds, but S-AdoMet was only demethylated to S-AdoHyc when L-Met was added, confirming the hypothesis from above.

Although we have not yet determined the following steps in the production pathway used by these bacteria, we know that DMSP is the final molecule produced through LC-MS confirmation. We hypothesise that Met methylation to produce SMM is the first step in the methylation DMSP production pathway **(Figure 7-1)**, used by the angiosperms that are able to produce DMSP (*Spartina, Wollastonia* and sugarcanes) (Stefels 2000). We can be confident in this because when this gene was disrupted in *T. profundimaris* WP0211, one of the *mmtN*-containing species ordered previously, DMSP was no longer produced at all, as confirmed by GC, detecting no DMS produced, and LC-MS confirming that DMSP (not just SMM) is no longer present. We also predict that out of the two routes taken by angiosperms through this pathway (either decarboxylation to DMSP-amine followed by oxidation to DMSP-aldehyde, or a transamination/ decarboxylation reaction to produce DMSP-aldehyde directly) (Dickschat et al. 2015), the latter is the pathway used by *mmtN*-containing species, as when *Novosphingobium* was incubated with intermediates from all the pathways, SMM caused an increase in DMSP production but DMSP-amine did not. Furthermore, when one scans the genomic position of *mmtN* in bacteria containing it,

unlike *dysB*, it is commonly linked to genes predicted to encode decarboxylases and aminotransferases. These are the predicted activities for the missing enzymes of the DMSP synthesis pathway and are very strong candidates for DMSP synthesis enzymes. Indeed in recent work Andrew Curson has knocked out the aminotransferase in *T. profundimaris* WP0211 and has shown that mutant no longer produces DMSP even with the addition of SMM. Thus the predictions of this thesis have been shown to be correct.

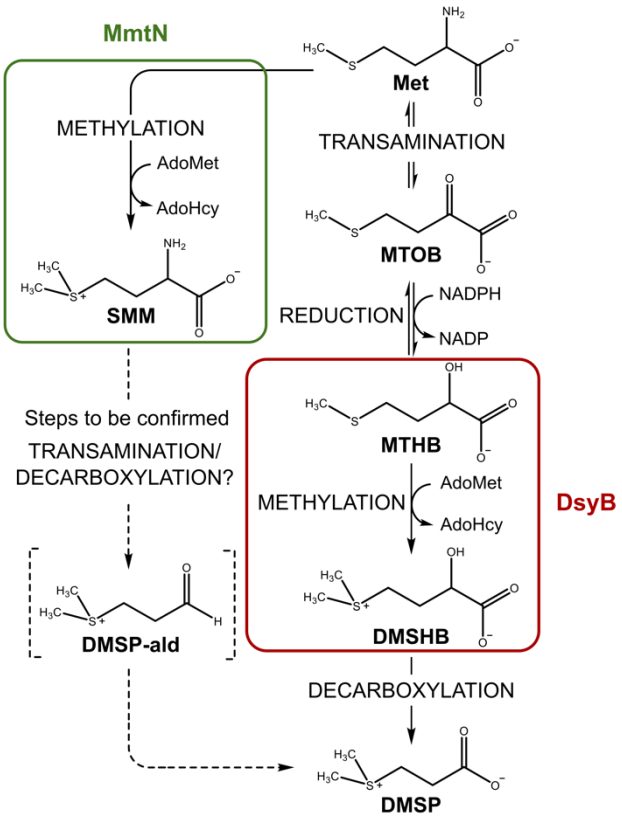

Figure 7-1: The two methods through which bacteria produce DMSP from L-Met. All intermediates from the transamination pathway have been confirmed, including the rate-limiting committed step that is catalysed by DsyB (red). Only the first step of the methylation pathway has been confirmed thus far, catalysed by MmtN (green), although the rest of the steps are theorised to include a transamination and decarboxylation reaction to produce DMSP-aldehyde before becoming DMSP.

When induction experiments were performed on *Novosphingobium* and *Thalassospira* under various different growth conditions, it was seen that DMSP production was significantly increased when in the high salinity media, with *Thalassospira* in particular functioning and producing DMSP even at levels of 70 PSU. Despite this evidence suggesting that DMSP production by is likely linked to salinity, phenotyping experiments carried out on the mutant did not show any reduction in growth compared to the W/T, even when they were grown in high salinity. Other conditions were tested, including varying

nitrogen levels, treating the cells to competition experiments and freeze-thawing them were all tested on the mutant to attempt to find a condition that produces a reduced phenotype, but none have produced a definitive result so far. This does not necessarily mean that there is not one, but so far no tested condition has produced one. One explanation could be that, as seen on LC-MS chromatographs, when DMSP is knocked out, it would appear that GBT is upregulated, perhaps in order to maintain the same level of osmoprotection.

Even though there is no phenotype as yet, the fact that *mmtN* exists is a significant discovery, as it means that the estimates of *dsyB* abundance and transcription in the environment are not descriptive of the total DSMP production, as there are many more DMSP-producing species also contributing to total DMSP levels in the environment.

7.2.5 Microcosm experiments on Stiffkey sediment dramatically increase DMSP production by the sediment.

The culture-dependent work performed on Stiffkey salt marsh is only one aspect of the story (**Chapter 4**). As a complement to the previous chapters, **Chapter 6** was almost entirely analysis of sequencing performed on community DNA extracted from Stiffkey sediment. This was either natural (Time 0) sediment, or sediment that had been used in a microcosm experiment with either a control of standard media conditions, or a combination media composition designed in **Chapter 4** to increase DMSP production. The enriched sediment showed a large amount of DMSP production compared to the control, suggesting that DMSP-producing bacteria were either increased in abundance or were highly transcribing *dsyB, mmtN* and any other potential DMSP-synthesis genes.

The sequencing performed on these samples were two sets of amplicon, one for the 16S rRNA gene to enable phylogenetic identification of the community, and one using the *dsyB* degenerate primers to create a Diversity Assay of *dsyB* in the sediment (discussed above), as well as metagenomic sequencing of all the genomes present.

This analysis showed that there is a high number of genera predicted to include DMSP producers present in the natural Stiffkey sediments – up t0 3.5 %. These include *Streptomyces, Marinobacter* and *Roseovarius*. Of course the numbers of DMSP-producing genera increase in abundance in the enriched, as they do in the control incubation, revealing a high abundance of the genus *Alteromonas,* isolates of which have been shown to produce DMSP (**Chapter 4**). It also showed a noticeable abundance of *Novosphingobium* which was not unexpected, as when bacterial isolates were picked from sediment incubated in these conditions in **Chapter 4** a high proportion of the DMSP-producing species were *Novosphingobium*. Surprisingly, the *dsyB*-containing *Labrenzia* was actually higher in Control sediments compared to Enriched. Although these results

are promising, it is also important to note that; i) presence does not guarantee activity, and ii) some genera may contain species that do not produce DMSP, as well as ones that do.

The abundances of bacteria within the community was also analysed by metagenomics. This taxonomic data did not completely match those reported in the 16S rRNA sequencing, although this could be due to lack of coverage of the metagenomes, as well as differences in determining taxonomy. Both still show the higher abundance of *Alteromonas* in Enriched samples, alongside several *Thalassospira* species. The metagenomes were also mined for genes of interest and normalised to RecA to express them as percentage of bacterial in the community. The abundance of *dsyB* and *mmtN* increased in both Control and Enriched samples in comparison to the Time 0 sediment, with *mmtN* being slightly higher in Enriched samples, but for *dsyB* abundance the Control samples were almost twice the percentage in the Enriched. *DSYB* was only present in the Time 0 samples. The diversity of the *mmtN* sequences was roughly analysed by recording the closest-aligning *mmtN* sequence after BLASTp was performed on the reads. These showed that the Enriched samples were dominated with reads from *Novosphingobium* and *Thalassospira,* and the Control samples contained mostly hits aligning to *Labrenzia.*

To test the abundance and transcription of *dsyB* in these samples, qPCR was performed on DNA and cDNA constructed from RNA using specific primers. *mmtN* primers were also designed but were not suitable for qPCR. The abundance of *dsyB* actually appeared to be highest in the Enriched samples compared to the others, while Time 0 was higher in abundance than Cley and Pool water. The number of transcripts of *dsyB* were actually similar between all three sets of samples, especially compared to the copy numbers from DNA, where Time 0 numbers were much lower than Control and Enriched samples. However, this is not necessarily unexpected, as we have already stated that these primers need more optimisation for qPCR. The most accurate way to analyse the transcription in Stiffkey would be to perform metatranscriptomes, as this would be a less biased, all-encompassing analysis.

7.3 Recommendations for future research

This research has greatly broadened our understanding of the scope of bacterial DMSP production, fitting several pieces of the puzzle of the DMSP cycle together, but there are now other questions that need answering, and more avenues of research to pursue.

The precedents set out in this study, along with the enrichment experiment designed for the purpose of studying bacterial DMSP production, are now being put into

practice in other environments, not limited to salt marshes but also Mangrove swamps and even the Mariana Trench, which potentially has shown very similar findings as those described here.

7.3.1 Further work on salt marsh environments

Stiffkey salt marsh has been an excellent source of information on the role that bacteria play in DMSP production. There are several other experiments that perhaps could be carried out in order to complete the picture. Firstly, the coverage of the metagenomic sequencing that was performed was not as high as it could have been, so more, higher coverage metagenomic sequencing would be useful to compare to the ones described in this thesis. As previously mentioned, metatranscriptomics would also reveal much about the true transcript levels in the natural sediment in comparison to other genes, as well as perhaps on the Enriched and Control samples. Another option would be to carry out 16S rRNA amplicon and metagenomic sequencing on Stiffkey at different times of the year, to observe the change in abundance through the year. The anoxic sediment, while not producing DMSP at the same level that the oxic layer does, would still be worth analysing as it is known that they are sites in which DMSP catabolism takes place (Kiene & Visscher 1987), so DMSP may be being produced but is then degraded before being detected. At the very least, culture-dependent work could give an indication of whether or not bacterial DMSP producers exist there.

Although the *dsyB* transcript numbers seen in Cley and Yarmouth are relatively low, DMSP levels are high (**Chapter 4**) so it is possible that *mmtN* is much higher in abundance or transcription. It would therefore be a good idea to perform the same sequencing, including metatranscriptomes, on these environments as a comparison to Stiffkey.

All the work carried out on Stiffkey focussed on the lower marsh portion, but it is a very large salt marsh, and therefore likely varies greatly. It would therefore be interesting to perform similar culturing experiments and perhaps sequencing on different areas of the marsh. Indeed, the upper marshes, while being more variable depending on rainfall, can become very hypersaline after a lack of rain (Davy & Smith 1988), analysis of which could provide interesting examples of DMSP-producing species of bacteria.

To further confirm the hypothesis that bacterial contributions are important in Stiffkey compared to the eukaryotic contribution, the activity of prokaryotic and eukaryotic organisms in Stiffkey could be compared, perhaps using antibiotics to remove either set of organisms from the sediment as carried out by Carrión et al. (2017).

Several bacterial isolates from Stiffkey, namely *Alteromonas* and *Marinobacter*, were found to produce DMSP without containing either *dsyB* or the newly-identified *mmtN*. It would therefore be very interesting to fully characterise these species, create genomic libraries and screen for DMS production.

7.3.2 Further analysis of *dsyB* and *mmtN*

One of the major findings of this piece of work was the discovery of the novel DMSP-producing gene, *mmtN.* Although a disruption mutant was successfully created of *mmtN* in *T. profundimaris,* we did not observe any effect on the growth after the loss of DMSP. There are however, other conditions in which the mutant has yet to be tested, to further analyse the role that DMSP plays in the organism. These include increased levels of oxidative stress through the addition of H_2O_2 or treatment with UV light, and perhaps even higher salinity levels, as *T. profundimaris* still grows at 70 PSU conditions. The LC-MS analysis suggests that the lack of a phenotype in the *mmtN*$^-$ mutant could be due to increased production of GBT, which could be tested by finding the genes involved in GBT synthesis in that organism, and then creating a double mutant to confirm that loss of the two osmoprotectants affects growth.

Now that the gene has been identified and the protein purified there are many experiments that can be performed in continued analysis of the MmtN enzyme. X-ray crystallography would reveal the structure, and perhaps give insight to the mechanism of the MmtN and any required compounds it might need. Alongside this, the sequence upstream of the gene could be cloned into the pBluescript, and used in *lacZ* fusions to determine the promotor region of the gene, and therefore observe any conditions that increase gene expression.

As mentioned in **Figure 7-1**, although we know that the methylation of Met to SMM is the first step in the production of DMSP in bacteria, we do not yet know what the other steps are in the pathway. There is already a precedent for the pathway to take two separate routes to reach DMSP, so it could be that bacteria use a third route to achieve this. To determine the rest of the pathway intermediates LC-MS and or HPLC work should be done on *Thalassospira* wild type and mutants defective in DMSP synthesis to detect the missing intermediates, which could be either DMSP-amine or DMSP-aldehyde. This will require the use of either radiochemicals or stable isotope work. Such work is a large component of a grant Dr Todd has in review presently. One problem with this might be that methionine is not solely used in DMSP production, and therefore may be difficult to track. Another option would be to use heavy-isotope labelled SMM instead.

7.4 Concluding Remarks

It was long thought that marine eukaryotes, specifically phytoplankton, are the most significant DMSP producers in the environment (Kiene et al. 2000), with species such as *E. huxleyi* producing it at continuously high levels with little regulation (Sunda et al. 2007). In this thesis, we show that a wide range of bacteria also possess the ability to synthesise DMSP, with many other potential bacterial species yet to be confirmed. Indeed, since the discovery of bacterial DMSP production roughly three years ago, two unrelated genes involved in this process have been identified, alongside several species that have function but do not appear to contain either of the known genes. It is clearly a most prevalent ability.

We have shown that although eukaryotic activity may be contributing strongly to DMSP levels detected in surface ocean waters, and in other eukaryote-rich environments such as algal blooms, heterotrophic bacteria are likely important DMSP producers in marine sediments like Stiffkey salt marsh, contributing noticeably to the total DMSP levels, which, per mass unit, are far more productive than overlying seawater. Indeed, experiments very recently carried out by Andy Hind on DMSP-synthesis rates using 10µM ^{13}C-Met to label processes in Stiffkey sediment has shown that the sediment is much more active than the seawater, easily detectable on the GC, in accordance with the findings of the RT-qPCR.

Despite lower abundance of *dsyB* and *mmtN* in seawater metagenomes compared to the salt marsh ones, the seawater incubation experiments performed in **Chapter 4** demonstrate that both *dsyB*-containing *Pelagibaca* and *mmtN*-containing *Novosphingobium* are able to produce detectable levels of DMSP even when incubated in almost *in situ* seawater conditions, meaning that they and likely others have the potential to produce DMSP in the seawater, as well as sediment.

It was clear from the very first experiments performed on Stiffkey salt marsh sediment, that the bulk of the DMSP production takes place in the mud. These have the highest levels of DMSP compared to the pool water and anoxic sediment, and are an excellent site in which to study bacteria. The sediment also has a higher combined abundance of *dsyB* and *mmtN* compared to ocean metagenomes (1.05 % of bacteria, compared to 0.38 % in the ocean), and higher levels of transcription of *dsyB* compared to overlying pool water.

One of the limitations to the study of DMSP cycling previous to the discovery of bacterial DMSP was that of location. The majority of work has focussed on the photic layer of the ocean and other environments, where eukaryotes grow best. This work shows that the limits of light (and to a degree, oxygen content) do not apply. Even the anoxic layer of Stiffkey sediment is an order-of-magnitude higher than the pool water.

DMSP-producing bacteria and their *dsyB* and/or *mmtN* transcripts were present in Stiffkey, Cley, Yarmouth and all tested seawater samples and *Tara* Oceans bacterioplankton datasets. We therefore hypothesise that *dsyB* and *mmtN* are far more abundant in marine surface sediment compared to ocean environments. Through this study we have demonstrated that not only are the surface marine sediments environments with high DMSP productivity, but also that heterotrophic bacteria are likely important producers in these environments.

8 REFERENCES

Aird, D. et al., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, 12(2).

Aird, D. et al., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, 12(2).

Alcolombri, U. et al., 2014. DddD Is a CoA-Transferase/Lyase Producing Dimethyl Sulfide in the Marine Environment. *Biochemistry*, 53, pp.5473–5475.

Alcolombri, U. et al., 2015. Identification of the algal dimethyl sulfide–releasing enzyme: A missing link in the marine sulfur cycle. , 348(6242), pp.1466–1469.

Andreae, M.O., 1990. Ocean-atmosphere interactions in the global biogeochemical sulfur cycle. *Marine Chemistry*, 30, pp.1–29.

Ansede, J.H., Friedman, R. & Yoch, D.C., 2001. Phylogenetic Analysis of Culturable Dimethyl Sulfide-Producing Bacteria from a Spartina-Dominated Salt Marsh and Estuarine Water. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 67(3), pp.1210–1217.

Ansede, J.H., Pellechia, P.J. & Yoch, D.C., 1999. Metabolism of acrylate to β-hydroxypropionate and its role in dimethylsulfoniopropionate lyase induction by a salt marsh sediment bacterium, Alcaligenes faecalis M3A. *Applied and Environmental Microbiology*, 65(11), pp.5075–5081.

Arakawa, T. & Timasheff, S.N., 1985. The stabilization of proteins by osmolytes. *Biophysical Journal*, 47(3), pp.411–414.

Ashelford, K.E., Weightman, A.J. & Fry, J.C., 2002. PRIMROSE : a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. *Nucleic Acids Research*, 30(15), pp.3481–3489.

Bankevich, A. et al., 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, 19(5), pp.455–477.

Bates, T.S., 1994. The cycling of sulfur in surface seawater of the northeast Pacific. *Journal of Geophysical Research*, 99(C4), pp.7835–7843.

Beszteri, B. et al., 2010. Average genome size: A potential source of bias in comparative metagenomics. *ISME Journal*, 4(8), pp.1075–1077.

Biebl, H. et al., 2007. Description of *Labrenzia alexandrii* gen. nov., sp. nov., a novel alphaproteobacterium containing bacteriochlorophyll a, and a proposal reclassification of *Stappia aggregata* as *Labrenzia aggregata* comb. nov., and of *Stappia alba* as *Labrenzia alba* comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, 57(5), pp.1095–1107.

Birben, E. et al., 2012. Oxidative Stress and Antioxidant Defense. *WAO Journal*, 5(January), pp.9–19.

Bolger, A.M., Lohse, M. & Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), pp.2114–2120.

De Bont, J. a. M., Van Dijken, J.P. & Harder, W., 1981. Dimethyl Sulphoxide and Dimethyl Sulphide as a Carbon, Sulphur and Energy Source for Growth of *Hyphomicrobium S*. *Microbiology*, 127(2), pp.315–323.

Bourgis, F. et al., 1999. S-methylmethionine plays a major role in phloem sulfur transport and is synthesized by a novel type of methyltransferase. *The Plant cell*, 11(8), pp.1485–1498.

Bratbak, G., 1996. Viral activity in relation to *Emiliania huxleyi* blooms: a mechanism of DMSP release?. *Oceanographic Literature Review*, *7*(43), p.691.

Caporaso, J.G. et al., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, 9(5), pp.335–336.

Carrión, O. et al., 2015. A novel pathway producing dimethylsulphide in bacteria is widespread in soil environments. *Nature Communications*, 6, p.6579.

Carrión, O. et al., 2017. Methanethiol-dependent dimethylsulfide production in soil environments. *ISME Journal*, 11(10), pp.2379–2390.

Challenger, F. & Simpson, M.I., 1948. Studies on Biological Methylation. Part XII. A Precursor of the Dimethyl Sulphide evolved by *Polysiphonia fastigiata*. Dimethyl- 2-carboxyethylsulphonium Hydroxide and its Salts. *Journal of the Chemical Society*, 3, pp.1591–1597.

Charlson, R.J. et al., 1987. Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature*, 326(16), pp.655–661.

Cline, J., Braman, J. & Hogrefe, H.., 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, 24(18), pp.3546–3551.

Cook, R.J. & Wagner, C., 1984. Glycine N-methyltransferase is a folate binding protein of rat liver cytosol. *Proceedings of the National Academy of Sciences. USA*, 81(12), pp.3631–4.

Cosquer, A. et al., 1999. Nanomolar Levels of Dimethylsulfoniopropionate , Dimethylsulfonioacetate, and Glycine Betaine Are Sufficient To Confer Osmoprotection to *Escherichia coli*. *Applied and Environmental Microbiology*, 65(8), pp.3304–3311.

Curson, A. et al., 2018. DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton. *Nature Microbiology*, 3, pp.430–439.

Curson, A.R.J., et al., 2011. Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes. *Nature reviews. Microbiology*, 9(12), pp.849–59.

Curson, A.R.J., Sullivan, M.J., et al., 2011. DddY, a periplasmic dimethylsulfoniopropionate lyase found in taxonomically diverse species of Proteobacteria. *The ISME journal*, 5(7), pp.1191–200.

Curson, A.R.J. et al., 2017. Dimethylsulfoniopropionate biosynthesis in marine bacteria and identification of the key gene in this process. *Nature Microbiology*, 2(February), p.17009.

Curson, A.R.J. et al., 2008. Molecular genetic analysis of a dimethylsulfoniopropionate lyase that liberates the climate-changing gas dimethylsulfide in several marine alpha-proteobacteria and *Rhodobacter sphaeroides*. *Environmental microbiology*, 10(3), pp.757–67.

Curson, A.R.J. et al., 2012. Multiple DMSP lyases in the γ-proteobacterium *Oceanimonas doudoroffii*. *Biogeochemistry*, 110, pp.109–119.

Dacey, J.W.H., King, G.M. & Wakeham, S.G., 1987. Factors controlling emission of dimethylsulfide from salt marshes. *Nature*, 330(17), pp.643–645.

Davis, K.E.R. et al., 2005. Effects of Growth Medium, Inoculum Size , and Incubation Time on Culturability and Isolation of Soil Bacteria Effects of Growth Medium, Inoculum Size , and Incubation Time on Culturability and Isolation of Soil Bacteria. *Applied and environmental microbiology*, 71(2), pp.826–834.

Davy, A.J. & Smith, H., 1988. Life-history variation and environment. In A. J. Davy, M. J. Hutchings, & A. R. Watkinson, eds. *Plant Population Ecology*. Blackwell Scientific Publications.

Debose, J.L., Lema, S.C. & Nevitt, G.A., 2008. Dimethylsulfoniopropionate as a Foraging Cue for Reef Fishes. *Science*, 319, p.1356.

DeBose, J.L. & Nevitt, G. a, 2008. The use of odors at different spatial scales: comparing birds with fish. *Journal of chemical ecology*, 34(7), pp.867–81.

Dickschat, J.S., Rabe, P. & Citron, C.A., 2015. The chemical biology of

dimethylsulfoniopropionate. *Organic & biomolecular chemistry*, 13(7), pp.1954–1968.

Dickson, D.M., Jones, R.G.W. & Davenport, J., 1980. Steady state osmotic adaptation in Ulva lactuca. *Planta*, 150(2), pp.158–165.

Dickson, D.M.J. & Kirst, G.O., 1986. The role of beta-dimethylsulphoniopropionate glycine betaine and homarine in the osmoacclimation of pPlatymonas subcordiformis. *Planta*, 167(4), pp.536–543.

Ehlers, B. et al., 1999. Detection of New DNA Polymerase Genes of Known and Potentially Novel Herpesviruses by PCR with Degenerate and Deoxyinosine-Substituted Primers. *Virus Genes*, 18(3), pp.211–220.

Eitinger, T. et al., 2011. Canonical and ECF-type ATP-binding cassette importers in prokaryotes: Diversity in modular organization and cellular functions. *FEMS Microbiology Reviews*, 35(1), pp.3–67.

Gage, D.A. et al., 1997. A new route for synthesis of dimethylsulphoniopropionate in marine algae. *Letters to Nature*, 387(June), pp.891–894.

Galí, M. et al., 2015. A remote sensing algorithm for planktonic dimethylsulfoniopropionate (DMSP) and an analysis of global patterns. *Remote Sensing of Environment*, 171(October), pp.171–184.

Giloteaux, L. et al., 2013. Characterization and transcription of arsenic respiration and resistance genes during in situ uranium bioremediation. *ISME Journal*, 7(2), pp.370–383.

Girgis, S.I. et al., 1988. Generation of DNA probes for peptides with highly degenerate codons using mixed primer PCR. *Nucleic Acids Research*, 16(21), p.10371.

González, J.M. et al., 1996. Identifying numerically abundant culturable bacteria from complex communities: An example from a lignin enrichment culture. *Applied and Environmental Microbiology*, 62(12), pp.4433–4440.

Grone, T. & Kirst, G.O., 1992. The effect of nitrogen deficiency, methionine and inhibitors of methionine metabolism on the DMSP contents of *Tetraselmis subcordiformis* (Stein). *Marine Biology*, 503, pp.497–503.

Gurevich, A. et al., 2013. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), pp.1072–1075.

Hanson, A.D. et al., 1994. Biosynthesis of 3-Dimethylsulfoniopropionate in *Wollastonia biflora* (1.) DC. *Plant Physiology*, 105, pp.103–110.

Hehemann, J.H. et al., 2014. The structure of RdDddP from *Roseobacter denitrificans*

reveals that DMSP lyases in the DddP-family are metalloenzymes. *PLoS ONE*, 9(7).

Hobbie, J.E., Daley, R.J. & Jasper, S., 1977. Use of nuclepore filter counting bacteria by fluoroscence microscopy. *Applied and Environmental Microbiologynvironmental microbiology*, 33(May), pp.1225–1228.

Holmes, A.J. et al., 1995. Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionary related. *FEMS Microbiol.Lett.*, 132Gamma-p, pp.203–208.

Howard, E.C. et al., 2008. Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environmental microbiology*, 10(9), pp.2397–410.

Howard, E.C. et al., 2006. Bacterial taxa that limit sulfur flux from the ocean. *Science*, 314, pp.649–52.

Iserte, J.A. et al., 2013. Family-Specific Degenerate Primer Design: A Tool to Design Consensus Degenerated Oligonucleotides. *Biotechnology Research International*, 2013, pp.1–9.

Ito, T. et al., 2011. Regulation of biosynthesis of dimethylsulfoniopropionate and its uptake in sterile mutant of *Ulva pertusa* (Chlorophyta). *Journal of Phycology*, 47(3), pp.517–523.

James, F., Nolte, K.D. & Hanson, A.D., 1995. Purification and Properties of S -Adenosyl-L -methionine : L - Methionine S -Methyltransferase from *Wollastonia biflora* Leaves. *The Journal of Biological chemistry*, 270(38), pp.22344–22350.

Jin, Y.O. & Mattes, T.E., 2011. Assessment and modification of degenerate qPCR primers that amplify functional genes from etheneotrophs and vinyl chloride-assimilators. *Letters in Applied Microbiology*, 53(5), pp.576–580.

Johnston, A.W., Green, R.T. & Todd, J.D., 2016. Enzymatic breakage of dimethylsulfoniopropionate—a signature molecule for life at sea. *Current Opinion in Chemical Biology*, 31(Figure 1), pp.58–65.

Johnston, A.W.B. et al., 2008. Molecular diversity of bacterial production of the climate-changing gas, dimethyl sulphide, a molecule that impinges on local and global symbioses. *Journal of experimental botany*, 59(5), pp.1059–67.

Karsten, U., Kirst, G.O. & Wiencke, C., 1992. Dimethylsulphoniopropionate (DMSP) accumulation in green macioalgae from polar to temperate regions: interactive effects of light versus salinity and light versus temperature. *Polar Biology*, 12(6–7), pp.603–607.

Keller, M.D., Bellows, W.K. & Guillard, R.R.L., 1989. Dimethyl Sulfide Production in Marine

Phytoplankton. *Biogenic Sulfur in the Environment*, pp.167–182.

Kempf, B. & Bremer, E., 1998. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Archives of Microbiology*, 170(5), pp.319–330.

Kettle, A.J. et al., 1999. A global data base of sea surface dimethyl sulfide (DMS) measurements and a simple model to predict sea surface DMS as a function of latitude, longitude, and month. *Global Biogeochem. Cycles*, 13(2), pp.399–444.

Kettle, A.J. & Andreae, M.O., 2000. Flux of dimethylsulfide from the oceans : A comparison of updated data sets and flux models. *Journal of Geophysical Research*, 105, p.26,793-26,808.

Kiene, R.P. et al., 1999. Dimethylsulfoniopropionate and methanethiol are important precursors of methionine and protein-sulfur in marine bacterioplankton. *Applied and Environmental Microbiology*, 65(10), pp.4549–4558.

Kiene, R.P., Hoffmann Williams, L.P. & Walker, J.E., 1998. Seawater microorganisms have a high affinity glycine betaine uptake system which also recognizes dimethylsulfoniopropionate. *Aquatic Microbial Ecology*, 15, pp.39–51.

Kiene, R.P. & Linn, L.J., 2000. Distribution and turnover of dissolved DMSP and its relationship with bacterial production and dimethylsulfide in the Gulf of Mexico. *Limnol. Oceanogr*, 45(4), pp.849–861.

Kiene, R.P., Linn, L.J. & Bruton, J. a., 2000. New and important roles for DMSP in marine microbial communities. *Journal of Sea Research*, 43(3–4), pp.209–224.

Kiene, R.P. & Taylor, B.F., 1988. Demethylation of Dimethylsulfoniopropionate and Production of Thiols in Anoxic Marine Sediments. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 54(9), pp.2208–2212.

Kiene, R.P. & Visscher, P.T., 1987. Production and fate of methylated sulfur compounds from methionine and dimethylsulfoniopropionate in anoxic salt marsh sediments. *Applied and environmental microbiology*, 53(10), pp.2426–2434.

Kirkwood, M. et al., 2010. The dddP gene of *Roseovarius nubinhibens* encodes a novel lyase that cleaves dimethylsulfoniopropionate into acrylate plus dimethyl sulfide. *Microbiology*, 156(Pt 6), pp.1900–6.

Kirst, G.O. et al., 1991. Dimethylsulfoniopropionate (DMSP) in icealgae and its possible biological role. *Marine Chemistry*, 35(1–4), pp.381–388.

Kitaguchi, H., Uchida, a & Ishida, Y., 1999. Purification and characterization of L-methionine decarboxylase from Crypthecodinium cohnii. *Fisheries Science*, 65(4),

pp.613–617.

Ko, R., Smith, L.T. & Smith, G.M., 1994. Glycine betaine confers enhanced osmotolerance and cryotolerance on *Listeria monocytogenes*. *Journal of Bacteriology*, 176(2), pp.426–431.

Kocsis, M.G. et al., 1998. Dimethylsulfoniopropionate biosynthesis in *Spartina alterniflora*. *Plant physiology*, 117(1), pp.273–281.

Kocsis, M.G. & Hanson, a D., 2000. Biochemical evidence for two novel enzymes in the biosynthesis of 3-dimethylsulfoniopropionate in *Spartina alterniflora*. *Plant physiology*, 123(3), pp.1153–61.

Kowalewsky, S. et al., 2006. High olfactory sensitivity for dimethyl sulphide in harbour seals. *Biology letters*, 2(1), pp.106–109.

Kuwae, T. & Hosokawa, Y., 1999. Determination of abundance and biovolume of bacteria in sediments by dual staining with 4',6-diamidino-2-phenylindole and acridine orange: Relationship to dispersion treatment and sediment characteristics. *Applied and Environmental Microbiology*, 65(8), pp.3407–3412.

Lau, E. et al., 2013. The Methanol Dehydrogenase Gene, mxaF, as a Functional and Phylogenetic Marker for Proteobacterial Methanotrophs in Natural Environments. *PLoS ONE*, 8(2).

Lee, C.C. et al., 1988. Generation of cDNA Probes Directed by Amino Acid Sequence : Cloning of Urate Oxidase. *Science*, 239(4845), pp.1288–1291.

Leustek, T. & Saito, K., 1999. Sulfate Transport and Assimilation in Plants. *Plant Physiology*, 120, pp.637–643.

Levine, N.M. et al., 2012. Environmental, biochemical and genetic drivers of DMSP degradation and DMS production in the Sargasso Sea. *Environmental microbiology*, 14(5), pp.1210–23.

Malmstrom, R.R. et al., 2004. Contribution of SAR11 Bacteria to Dissolved Dimethylsulfoniopropionate and Amino Acid Uptake in the North Atlantic Ocean. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 70(7), pp.4129–4135.

Malmstrom, R.R. et al., 2005. Dimethylsulfoniopropionate (DMSP) assimilation by *Synechococcus* in the Gulf of Mexico and northwest Atlantic Ocean. *Limnology and Oceanography*, 50(6), pp.1924–1931.

McDevitt, C.A. et al., 2002. Molecular analysis of dimethyl sulphide dehydrogenase from *Rhodovulum sulfidophilum*: Its place in the dimethyl sulphoxide reductase family of microbial molybdopterin-containing enzymes. *Molecular Microbiology*, 44(6),

pp.1575–1587.

McDonald, I.R.M.C., Kenna, E.M. & Murrell, J.C., 1995. Detection of Methanotrophic Bacteria in Environmental-Samples With the PCR. *Applied and Environmental Microbiology*, 61(1), pp.116–121.

McLenon, A.L. & DiTullio, G.R., 2012. Effects of increased temperature on dimethylsulfoniopropionate (DMSP) concentration and methionine synthase activity in *Symbiodinium microadriaticum*. *Biogeochemistry*, 110(1–3), pp.17–29.

Miller, T.R. et al., 2004. Chemotaxis of Silicibacter sp. Strain TM1040 toward Dinoflagellate Products. *Society*, 70(8), pp.4692–4701.

Miller, T.R. & Belas, R., 2004. Dimethylsulfoniopropionate Metabolism by *Pfiesteria* -Associated *Roseobacter spp*. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 70(6), pp.3383–3391.

Mudd, S.H. & Datko, a H., 1990. The S-Methylmethionine Cycle in *Lemna paucicostata*. *Plant physiology*, 93(2), pp.623–630.

Mühling, M. et al., 2008. Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *ISME Journal*, 2(4), pp.379–392.

Nevitt, G.A., Velt, R.R. & Kareiva, P., 1995. Dimethyl sulphide as a foraging cue for Antarctic Procellariiform seabirds. *Letters to Nature*, pp.680–682.

Oduro, H., Van Alstyne, K.L. & Farquhar, J., 2012. Sulfur isotope variability of oceanic DMSP generation and its contributions to marine biogenic sulfur emissions. *Proceedings of the National Academy of Sciences of the United States of America*, 109(23), pp.9012–6.

Otte, M.L. et al., 2004. Dimethylsulphoniopropionate (DMSP) and related compounds in higher plants. *Journal of Experimental Botany*, 55(404), pp.1919–1925.

Pinheiro, L.B. et al., 2012. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Analytical Chemistry*, 84(2), pp.1003–1011.

Quinn, P.K. & Bates, T.S., 2011. The case against climate regulation via oceanic phytoplankton sulphur emissions. *Nature*, 480(7375), pp.51–6.

Raina, J.-B. et al., 2013. DMSP biosynthesis by an animal and its role in coral thermal stress response. *Nature*, 502(7473), pp.677–80.

Rainey, F.A. et al., 1996. *Clostridium paradoxum* DSM 7308(T) contains multiple 16S rRNA genes with heterogeneous intervening sequences. *Microbiology*, 142(8),

pp.2087–2095.

Ranocha, P. et al., 2000. Characterization and Functional Expression of cDNAs Encoding Methionine-sensitive and -insensitive Homocysteine S-Methyltransferases from *Arabidopsis*. *Journal of Biological Chemistry*, 275(21), pp.15962–15968.

Reisch, C.R., Stoudemayer, M.J., et al., 2011. Novel pathway for assimilation of dimethylsulphoniopropionate widespread in marine bacteria. *Nature*, 473(7346), pp.208–11.

Reisch, C.R., Moran, M.A. & Whitman, W.B., 2011. Bacterial Catabolism of Dimethylsulfoniopropionate (DMSP). *Frontiers in microbiology*, 2(August), p.172.

Rhodes, D. et al., 1997. S-Methylmethionine Conversion to Dimethylsulfoniopropionate: Evidence for an Unusual Transamination Reaction. *Plant physiology*, 115(1 997), pp.1541–1548.

Rouillon, a., Surdin-kerjan, Y. & Thomas, D., 1999. Transport of Sulfonium Compounds. *The Journal of Biological Chemistry*, 274(40), pp.28096–28105.

Saleh-Lakha, S. et al., 2005. Microbial gene expression in soil: Methods, applications and challenges. *Journal of Microbiological Methods*, 63(1), pp.1–19.

Schäfer, H., Myronova, N. & Boden, R., 2010. Microbial degradation of dimethylsulphide and related C1-sulphur compounds: organisms and pathways controlling fluxes of sulphur in the biosphere. *Journal of experimental botany*, 61(2), pp.315–34.

Schloss, P.D., Gevers, D. & Westcott, S.L., 2011. Reducing the effects of PCR amplification and sequencing Artifacts on 16s rRNA-based studies. *PLoS ONE*, 6(12).

Schloss, P.D. & Handelsman, J., 2003. Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology*, 14(3), pp.303–310.

Segata, N. et al., 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods*, 9(8), pp.811–814.

Seymour, J.R. et al., 2010. Chemoattraction to dimethylsulfoniopropionate throughout the marine microbial food web. *Science (New York, N.Y.)*, 329(5989), pp.342–5.

Shaw, G.E., 1987. Aerosols as climate regulators: a climate-biosphere linkage? *Atmospheric Environment*, 21, pp.1983–1984.

Sheets, E.B. and Rhodes, D., 1996. Determination of DMSP and other onium compounds in *Tetraselmis subcordiformis* by plasma desorption mass spectrometry. In *Biological and environmental chemistry of DMSP and related sulfonium compounds* (pp. 55-63). Springer, Boston, MA.

Sievert, S.M., Kiene, R.P. & Schulz-vogt, H.N., 2007. The Sulfur Cycle. *Oceanography*, 20(2), pp.117–123.

Simó, R. et al., 2009. Annual DMSP contribution to S and C fluxes through phytoplankton and bacterioplankton in a NW Mediterranean coastal site. *Aquatic Microbial Ecology*, 57, pp.43–55.

Skinner, F.A., Jones, P.C.T. & Mollison, J.E., 1952. A Comparison of a Direct- and a Plate-counting Technique for the Quantitative Estimation of Soil Micro-organisms. *Journal of General Microbiology*, 6, pp.261–271.

Slezak, D.M., Puskaric, S. & Herndl, G.J., 1994. Potential role of acrylic acid in bacterioplankton communities in the sea. *Marine Ecology Progress Series*, 105(1–2), pp.191–198.

de Souza, M.P. & Yoch, D.C., 1995. Purification and Characterization of Dimethylsulfoniopropionate Lyase from an *Alcaligenes* -Like Dimethyl Sulfide-Producing Marine Isolate. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 61(1), pp.21–26.

de Souza M.P. and Yoch D.C. 1996. N-Terminal Amino Acid Sequences and Comparison of DMSP Lyases from*Pseudomonas Doudoroffii* and Alcagenes Strain M3A. In: Kiene R.P., Visscher P.T., Keller M.D., Kirst G.O. (eds) *Biological and Environmental Chemistry of DMSP and Related Sulfonium Compounds*. Springer, Boston, MA

Speeckaert, G. et al., 2018. Annual cycle of dimethylsulfoniopropionate (DMSP) and dimethylsulfoxide (DMSO) related to phytoplankton succession in the Southern North Sea. *Science of the Total Environment*, 622–623(December 2017), pp.362–372.

Staskawicz, B. et al., 1987. Molecular Characterization of Cloned Avirulence Genes From Race-0 and Race-1 of *Pseudomonas syringae* pv. <i>glycinea</ i>. *Journal of Bacteriology*, 169(12), pp.5789–5794.

Stefels, J. et al., 2007. Environmental constraints on the production and removal of the climatically active gas dimethylsulphide (DMS) and implications for ecosystem modelling. *Biogeochemistry*, 83(1–3), pp.245–275.

Stefels, J., 2000. Physiological aspects of the production and conversion of DMSP in marine algae and higher plants. *Journal of Sea Research*, 43, pp.183–197.

Stefels, J. & van Boekel, W.H.., 1993. Production of DMS from dissolved DMSP in axenic cultures of the marine phytoplankton species Phaeocystis sp. *Marine Ecology Progress Series*, 97, pp.11–18.

Steinke, M., Stefels, J. & Stamhuis, E., 2006. Dimethyl sulfide triggers search behavior in copepods. *Limnology and Oceanography*, 51(4), pp.1925–1930.

Steudler, P.A. & Peterson, B.J., 1984. Contribution of gaseous sulphur from salt marshes to the global sulphur cycle. *Nature*, 311(5985), pp.455–457.

Steven, B. et al., 2007. Characterization of the microbial diversity in a permafrost sample from the Canadian high Arctic using culture-dependent and culture-independent methods. *FEMS Microbiology Ecology*, 59(2), pp.513–523.

Strom, S. et al., 2003. Chemical defense in the microplankton I: Feeding and growth rates of heterotrophic protists on the DMS-producing phytoplankter Emiliania huxleyi. *Limnology and Oceanography*, 48(1), pp.217–229.

Summers, P.S. et al., 1998. Identification and Stereospecificity of the First Three Enzymes of 3-Dimethylsulfoniopropionate Biosynthesis in a Chlorophyte Alga 1. In *Plant Physiology*. pp. 369–378.

Sun, D.L. et al., 2013. Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Applied and Environmental Microbiology*, 79(19), pp.5962–5969.

Sun, J. et al., 2016. The abundant marine bacterium *Pelagibacter* simultaneously catabolizes dimethylsulfoniopropionate to the gases dimethyl sulfide and methanethiol. *Nature Microbiology*, 1(8).

Sun, L. et al., 2012. Diversity of DMSP transport in marine bacteria, revealed by genetic analyses. *Biogeochemistry*, 110, pp.121–130.

Sunagawa, S. et al., 2015. Structure and function of the global ocean microbiome. , 348(6237), pp.1–10.

Sunda, W. et al., 2002. An antioxidant function for DMSP and DMS in marine algae. *Nature*, 418(6895), pp.317–20.

Sunda, W.G. et al., 2007. The effect of nitrogen limitation on cellular DMSP and DMS release in marine phytoplankton: Climate feedback implications. *Aquatic Sciences*, 69(3), pp.341–351.

Tan, D. et al., 2013. Crystal Structure of DmdD, a Crotonase Superfamily Enzyme That Catalyzes the Hydration and Hydrolysis of Methylthioacryloyl-CoA. *PLoS ONE*, 8(5), pp.1–9.

Todd, J.D. et al., 2011. DddQ, a novel, cupin-containing, dimethylsulfoniopropionate lyase in marine roseobacters and in uncultured marine bacteria. *Environmental microbiology*, 13(2), pp.427–38.

Todd, J.D. et al., 2012. DddW, a third DMSP lyase in a model Roseobacter marine bacterium, Ruegeria pomeroyi DSS-3. *The ISME journal*, 6(1), pp.223–6.

Todd, J.D. et al., 2010. Molecular dissection of bacterial acrylate catabolism--unexpected links with dimethylsulfoniopropionate catabolism and dimethyl sulfide production. *Environmental microbiology*, 12(2), pp.327–43.

Todd, J.D. et al., 2007. Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science (New York, N.Y.)*, 315(5812), pp.666–9.

Todd, J.D. et al., 2009. The dddP gene, encoding a novel enzyme that converts dimethylsulfoniopropionate into dimethyl sulfide, is widespread in ocean metagenomes and marine bacteria and also occurs in some Ascomycete fungi. *Environmental microbiology*, 11(6), pp.1376–85.

Torsvik, V. & Øvreås, L., 2002. Microbial diversity and function in soil: from genes to ecosystems. *Current Opinion in Microbiology*, 5, pp.240–245.

Trevena, A.J. et al., 2003. Profiles of dimethylsulphoniopropionate (DMSP), algal pigments, nutrients, and salinity in the fast ice of Prydz Bay, Antarctica. *Journal of Geophysical Research*, 108(C5), p.3145.

Tripp, H.J. et al., 2008. SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature*, 452(7188), pp.741–4.

Trossat, C., Nolte, K.D. & Hanson, A.D., 1996. Evidence That the Pathway of Dimethylsulfoniopropionate Biosynthesis Begins in the Cytosol and Ends in the Chloroplast. *Plant physiology*, 111(4), pp.965–973.

Trossat, C., Rathinasabapathi, B. & Hanson, a. D., 1997. Transgenically Expressed Betaine Aldehyde Dehydrogenase Efficiently Catalyzes Oxidation of Dimethylsulfoniopropionaldehyde and [omega]-Aminoaldehydes. *Plant physiology*, 113(4), pp.1457–1461.

Vallina, S.M. & Simó, R., 2007. Strong relationship between DMS and the solar radiation dose over the global surface ocean. *Science (New York, N.Y.)*, 315(5811), pp.506–8.

Varaljay, V. a et al., 2015. Single-taxon field measurements of bacterial gene regulation controlling DMSP fate. *The ISME Journal*, 10.1038/is, pp.1–10.

Varaljay, V.A. et al., 2010. Deep sequencing of a dimethylsulfoniopropionate-degrading gene (dmdA) by using PCR primer pairs designed on the basis of marine metagenomic data. *Applied and Environmental Microbiology*, 76(2), pp.609–617.

Vila-Costa, M. et al., 2006. Dimethylsulfoniopropionate uptake by marine phytoplankton.

Science (New York, N.Y.), 314(5799), pp.652–4..

Vila-Costa, M. et al., 2010. Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. *The ISME journal*, 4(11), pp.1410–20.

Visscher, P.T. & Taylor, B.F., 1993. A new mechanism for the aerobic catabolism of dimetyl sulfide. *Appl. Environ. Microbiol.*, 59(11), pp.3784–3789.

Walters, W. et al., 2015. Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems*, 1(1), pp.e0009-15.

Werpy, T. & Petersen, G., 2004. Top Value Added Chemicals from Biomass. *National Renewable Energy Laboratory, Richland, WA (US)*, PNNL-14808, p.76.

Wolfe, G. V. & Steinke, M., 1996. Grazing-activated production of dimethyl sulfide (DMS) by two clones of *Emiliania huxleyi*. *Limnology and Oceanography*, 41(6), pp.1151–1160.

Wolfe, G.V., 1996. Accumulation of dissolved DMSP by marine bacteria and its degradation via bacterivory. In *Biological and environmental chemistry of DMSP and related sulfonium compounds* (pp. 277-291). Springer, Boston, MA.

Wolfe, G. V, Steinke, M. & Kirst, G.O., 1997. Grazing-activated chemical defence in a unicellular marine alga. *Nature (London)*, 387(June), pp.894–897.

Yergeau, E. et al., 2010. The functional potential of high Arctic permafrost revealed by metagenomic sequencing, qPCR and microarray analyses. *ISME Journal*, 4(9), pp.1206–1214.

Yoch, D.C., 2002. Dimethylsulfoniopropionate : Its Sources , Role in the Marine Food Web, and Biological Degradation to Dimethylsulfide. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 68(12), pp.5804–5815.

Yoch, D.C., Ansede, J.H. & Rabinowitz, K.S., 1997. Evidence for Intracellular and Extracellular Dimethylsulfoniopropionate (DMSP) Lyases and DMSP Uptake Sites in Two Species of Marine Bacteria. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 63(8), pp.3182–3188.

Yost, D.M., Jones, R.J. & Mitchelmore, C.L., 2010. Alterations in dimethylsulfoniopropionate (DMSP) levels in the coral Montastraea franksi in response to copper exposure. *Aquatic Toxicology*, 98(4), pp.367–373.

Zakrzewski, M. et al., 2016. Calypso: A user-friendly web-server for mining and visualizing microbiome-environment interactions. *Bioinformatics*, 33(5), pp.782–783.

Zhuang, G. et al., 2011. Production of DMS and DMSP in different physiological stages

and salinity conditions in two marine algae. *Chinese Journal of Oceanology and Limnology*, 29(2), pp.369–377.

Ziegler, C., Bremer, E. & Krämer, R., 2010. The BCCT family of carriers: From physiology to crystal structure. *Molecular Microbiology*, 78(1), pp.13–34.

Zimmer-Faust, R.K., De Souza, M.P. & Yoch, D.C., 1996. Bacterial chemotaxis and its potential role in marine dimethylsulfide production and biogeochemical sulfur cycling. *Limnology and Oceanography*, 41(6), pp.1330–1334.