Distribution and diversity of isoprene degrading bacteria in the terrestrial environment

By

Lisa Gibson

A thesis submitted to the School of Environmental Sciences at the University of East Anglia in fulfilment of the requirements for the degree of Doctor of Philosophy

September 2021

©This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognize that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution

Abstract

Isoprene is the most abundantly produced biogenic volatile organic compound (BVOC) and is an important climate active gas that plays a complex role in atmospheric chemistry. Through a combination of cultivation, and cultivation-independent techniques this project set out to investigate the diversity and distribution of isoprene degrading bacteria, and to investigate the factors that influence their abundance in complex microbial communities in a range of environments.

DNA-Stable isotope probing (DNA-SIP) was combined with amplicon sequencing techniques and metagenomics, investigating the isoprene degrading community harboured by a high isoprene emitting willow (*Salix fragilis*) in the UK, and one of the highest known emitters, the oil palm tree (*Elaeis guineensis*) in Malaysia. Willow leaves harboured a diverse community of isoprene degrading bacteria and represented the first phyllosphere environment to contain communities of *Methylobacterium* and *Mycobacterium* that could metabolise isotopically-labelled isoprene during a DNA-SIP experiment. A *Mycobacterium* metagenome-assembled genome (MAG) was recovered, containing the first known instance of two complete isoprene degradation metabolic gene clusters in a single genome. Malaysian oil palm soil and phyllosphere studies showed that the soil environment harboured a novel and much more diverse community of isoprene degrading bacteria than the associated phyllosphere. Amplicon sequencing of the *isoA* gene (encoding the α-subunit of the isoprene monooxygenase) showed that genomes from phyllosphere and soil communities contained a novel diversity of *isoA* which is essential for bacterial degradation of isoprene. The abundance of isoprene degrading bacteria in a number of environments was also investigated with a combination of metagenomic gene mining and qPCR. qPCR methods were carried out with newly-designed gene probes, validated in this study, that target the *isoA* gene. Results showed that *isoA* containing bacteria are found in environments that do not emit isoprene, and make up 0.02% - 1% of the whole microbial community in a given environment.

Acknowledgements

Thanks must go first to my supervisor Colin Murrell, for giving me the opportunity to do this project, and for his time and expert advice throughout. To Andrew Crombie, for his wealth of knowledge and endless patience. You have a way of knowing when we need a vote of confidence and your gentle support has been much appreciated.

Special thanks go to Ornella Carrión. My PhD would have been far more difficult without your endless support and cheery optimism, even at times when everything seemed determined to go wrong. There's nobody else I'd rather tag team with for those graveyard SIP shifts.

A huge thanks go to the other PhD students in the ELSA lab. For their constant encouragement, companionship, empathy and understanding. To David for his sixth sense when one of us needs a little emotional support. You're a true friend and I'm incredibly lucky to have met you. To Leanne, with your endless capacity for fun, friendliness and seeing the positive in things, even when your own plate is full. You made my time in Norwich brighter with your presence. Every one of you helped to make the lab a place of kindness and support, I wouldn't have made it through without you all.

I'm incredibly lucky to have the family I do, and I could never thank them enough for getting me through just about everything up to this point. Mam, thank you for your open ears and for always being on my side. Dad, thank you for the bear hugs, constantly reminding me that you're proud of me and the assurance that those acts of love and kindness will always be waiting for me, just a plane ride or Skype call away. You've both been a rock for me over the years and I'm forever grateful.

To my sister, Sarah, who among many other things was a constant reminder that nothing good ever came from taking things (or yourself) too seriously. For making me laugh when I was at my most frustrated or sad. For being there and holding my hand when I fell to pieces, and firmly pulling me back to my feet during some of the hardest times of my life. I'm so lucky to have you. And thanks to Tessa just for existing and being so ridiculously cute that I can't help but be happy when I see her.

Where on Earth would I be without my TB wonder women, Katie and Ruth. You're both angels and have dragged me through this PhD with a combination of love, sass and cackles. Thank you for listening to *every single* woe. For your never-ending kindness, empathy and love. You're both amazing human beings and I'm better for knowing you.

Finally, thank you so much to my partner Shea. You've been there for me every single step of the way, with love and constant encouragement. Thank you for believing in me enough for the both of us. Thank you for constantly reminding me to put myself and my health first. Thank you for being a goof and allowing me to be fully myself when we're together. I'm incredibly lucky to have you in my life.

Declaration

I declare that the work presented in this thesis was conducted by me under the direct supervision of Professor J. Colin Murrell. Results obtained by, or with help from others has been acknowledged in the relevant section. None of the work presented has been previously submitted for any other degree.

Lisa Gibson

Contents

Tables

Figures

Acronyms

ANI	Average nucleotide identity
ANMS	Ammonium nitrate mineral salts
ASV	Amplicon sequence variant
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BVOC	Biogenic volatile organic compound
CoA	Coenzyme A
CT	Cycle threshold
CTAB	Cetyl trimethylammonium bromide
DGGE	Denaturing gradient gel electrophoresis
dH2O	Deionized water
DMAPP	Dimethylallyl diphosphate
DMS	Dimethylsulfide
DNA	Deoxyribonucleic acid
dw	Dry weight
EDTA	Ethylendiaminetetraacetic acid
FAO	Food and Agricultural Organization
FID	Flame ionization detector
FIS	Fast Isoprene Sensor
GC	Gas chromatography
gDNA	Genomic DNA
GMBA	2-glutathionyl-2-methyl-3-butenoate
HGMB	1-hydroxy-2-glutathionyl-2-methyl-3-butene
hr	Hour
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
IsoMO	Isoprene monooxygenase
IspS	Isoprene synthase
MAG	Metagenome assembled genome
mbp	Mega base pairs
MEP	Methylerythritol phosphate
MEP	Methylerythritol 4-phosphate
MVA	Mevalonate
OD	Optical density
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
pMMO	Particulate methane monooxygenase
ppbv	Parts per billion by volume
PQQ	Pyrroloquinoline quinone
psi	Pounds per square inch
qPCR	Quantitative polymerase chain reaction
RA	Relative abundance

RAST	Rapid Annotations Using Subsystems Technology
RDP	Ribosomal Database Project
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rpm	Revolutions per minute
rRNA	Ribosomal RNA
s	Seconds
s.d.	Standard deviation
SDIMO	Soluble diiron monooxygenase
SDS	Sodium dodecyl sulfate
SIP	Stable Isotope Probing
sMMO	Soluble methane monooxygenase
SOA	Secondary organic aerosols
SOB	Super optimized broth
SRC	Short rotation coppice
TEMED	N,N,N',N'-tetramethyl-ethane-1,2-diamine
Tris	Tris(hydroxymethyl)aminomethane
V	Volts
v/v	Volume to volume
VOC	Volatile organic compounds
w/v	Weight to volume
WGS	Whole genome sequencing
WT	Wild type

1 Introduction

1.1 Biogeochemical cycle of isoprene

Biogenic volatile organic compounds (BVOCs) play a critical role in global tropospheric chemistry, photochemical ozone creation and the oxidative capacity of the atmosphere (Fehsenfeld *et al*. 1992). The most abundant of these is isoprene (2-methyl-1,3-butadiene; C5H8), a hemiterpene and a common structural motif in isoprenoids (Lichtenthaler *et al*. 1997; Lichtenthaler 1999). The scale of isoprene emissions is equal that of all other BVOCs, such as monoterpenes, sesquiterpenes, methanol, acetone and ethene, combined. Isoprene emission is estimated by Guenther *et al.* (2012) at 550 ± 100 Tg C yr^{-1}. This figure is similar in scale to global emissions of methane, yet in comparison, we still know very little about some aspects of the global isoprene cycle, particularly its biodegradation.

1.1.1 Atmospheric chemistry of isoprene

Isoprene is a reactive diene. This high-reactivity, due to twin carbon-carbon double bonds, means that isoprene plays a complex role in atmospheric chemistry. For example, it is thought to have both warming and cooling effects on the Earth's climate. Also, isoprene's net effect on global temperature shifts is particularly susceptible to local conditions. The twin carbon-carbon double bonds leave isoprene susceptible to attack by nitrate (NO_3), hydroxyl radicals (OH) and ozone (O_3) with a reaction lifetime that can range from less than an hour to a little over a day (Steinfeld 1998; Atkinson and Arey 2003). The oxidation of isoprene is important in the modulation of global tropospheric O_3 and methane (CH_4) concentrations. However, the full scope of isoprene atmospheric chemistry is complex and new details are still being brought to light today. We know that nitrogen oxides (NO_x) are a key aspect of this process. Isoprene reacts readily with O_3 when levels of nitric oxide (NO) are low, and as a result lowers the levels of atmospheric O_3. However, in circumstances where NO levels are high, such as highly polluted urban environments, isoprene oxidation produces nitrogen dioxide (NO_2; Pacifico *et al.*, 2009) . Through photolysis, this leads to an increase of atmospheric O_3 (Monson *et al*. 2007) , a potent greenhouse gas and pollutant that impacts human health (Fowler *et al.*, 2008).

Products of isoprene oxidation can also partition in the condensed phase and lead to the formation of secondary organic aerosols (SOA; Figure 1.1; Claeys *et al.*, 2004). The yield of this process is low at >3%, however the sheer scale of isoprene emissions mean that

contribution of isoprene to the formation of SOA in non-negligible (Henze and Seinfeld 2006). SOA are an important source of cloud condensation nuclei and in creating them, isoprene contributes to a cooling effect in local climate (Carslaw *et al*. 2009).

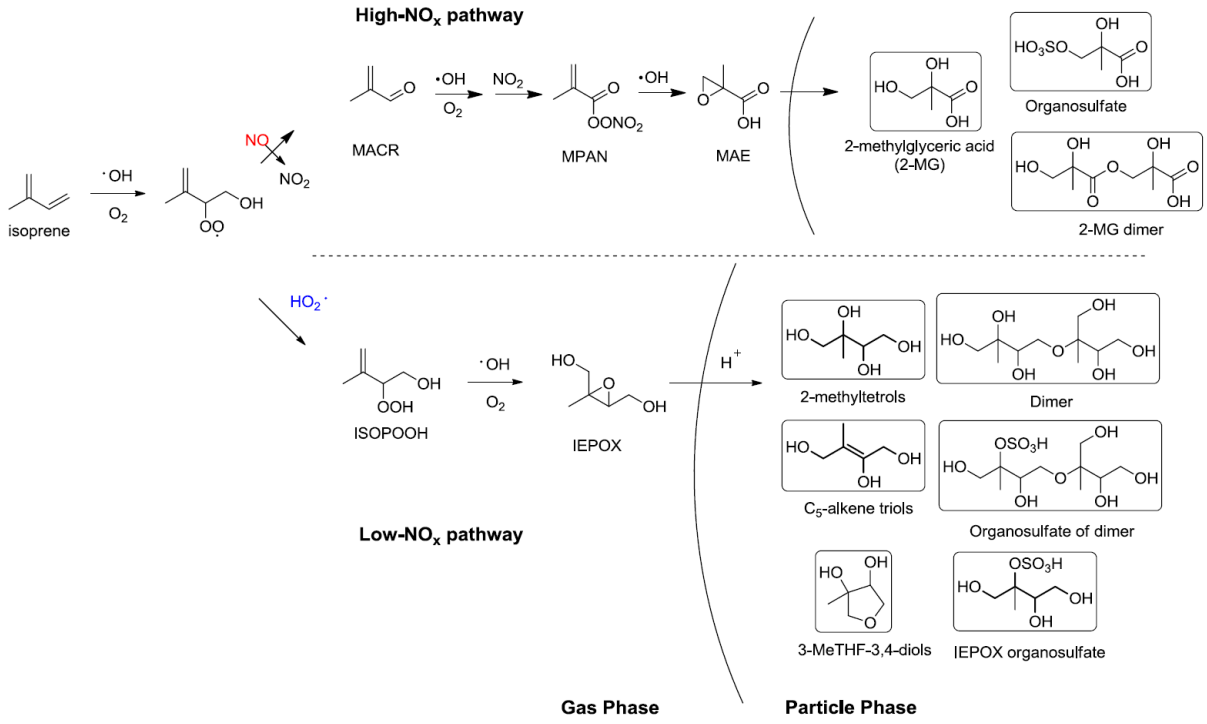

Figure 1.1. Proposed chemistry leading to formation of SOA by isoprene. Taken from Lin *et al.* (2013)

1.1.2 Anthropogenic sources of isoprene

Unlike other climate active gases of concern, only a small proportion of isoprene emissions are due to anthropogenic activity. However, there are a number of man-made sources for the compound, which is a key chemical commodity for a range of industrial products. Many elastomers used in golf balls, surgical gloves, condoms, shoes and rubber bands contain isoprene. It's most prevalent use however, is in the production of *cis*-polyisoprene (synthetic rubber), especially for tire manufacturing. Alongside these areas, isoprene is also used as an intermediate for perfumes, flavourings, vitamins and pharmaceuticals (Vickers and Sabri 2015). Originally, isoprene was synthesized chemically via the pyrolysis of natural rubber, but nowadays the bulk of isoprene production stems from fossil fuel resources, with global industrial production at over 1 million tonnes per year (Morais *et al*. 2015).

With plant emissions of isoprene at over 500 Tg yr^{-1}, it is interesting to note that industrial production currently only matches a fraction of that amount (Atkinson and Arey 2003; Morais *et al*. 2015). This means natural isoprene production is theoretically enough to meet industrial demands several hundred times over, were it effectively harnessed (Morais *et al*. 2015). Biotechnological approaches to isoprene production have been investigated and were reviewed by Ye *et al.* (2016). Pioneering work showed that isoprene could be produced by naturally occurring bacteria like *Bacillus subtilis,* but yield was quite low (Julsing *et al*. 2007). Overexpression of the biosynthetic pathways responsible for isoprene production in both bacteria and yeast has seen further success and led to the pilot-scale production of 'bioisoprene' by a small number of biotechnology companies (McAuliffe *et al.,* 2010; Vickers *et al.*, 2015). However, despite a growing interest in the development of bioisoprene, methods of production have still to reach a point of economic feasibility, with barriers such as the regulative mechanisms within host strains still to be overcome.

1.1.3 Biogenic sources of isoprene

Isoprene is unusual as a climate active gas in that the vast majority of emissions are biogenic in origin (Guenther *et al*. 2012). The sources of biogenic isoprene production are wide and varied, with emissions recorded from plants (Sharkey and Yeh, 2001; Sharkey *et al.,* 2007), bacteria (Kuzma *et al*. 1995; Ray Fall and Copley 2000), fungi (Bu'lock 1973; Bäck *et al*. 2010), animals (Thomas D. Sharkey 1996), humans (Fenske and Paulson 2011), algae (Dani *et al*. 2020; Meskhidze *et al*. 2015; D. A. Exton *et al*. 2013) and corals (Exton *et al.*, 2015; Hrebien *et al.,* 2020). The scale of emission between these groups varies by orders of magnitude, for example, isoprene emissions from the marine environment are estimated to range between 0.1 – 11.6 Tg C yr^{-1} (reviewed by Dawson *et al.* 2021) while terrestrial plants are estimated to emit approximately 535 Tg isoprene yr^{-1} (Guenther *et al*. 2012).

Table 1.1. Isoprene emission rates from a range of tree species.

Isoprene emission rates are measured in µg g (dry weight) h^{-1}.

Common Name	Species	µg g(dw) h^{-1}	Reference
Malaysian oil palm	*Elaeis guineensis*	175	Kesselmeier & Staudt, 1999
Oregon white oak	*Quercus garryana*	59.2	Hewitt & Street, 1992
Crack willow	*Salix fragilis*	37	Pio, C.A *et al*. 1993
White poplar	*Populus alba*	19.94	Hewitt & Street, 1992
Black spruce	*Picea mariana*	15	Hewitt & Street, 1992
Miscanthus	*Miscanthus x giganteus*	<0.1	Morrison *et al.*, 2016
Swamp ash	*Fraxinus caroliniana*	<0.1	Hewitt & Street, 1992
American beech	*Fagus grandifolia*	<0.1	Hewitt & Street, 1992
European fir	Abies alba	<0.1	Hewitt & Street, 1992
Silver birch	*Betula pendula*	<0.1	Hewitt & Street, 1992
Palestine oak	*Quercus calliprinos*	<0.1	Csiky & Seufert, 1999

Terrestrial trees in particular are responsible for approximately 90% of all isoprene emissions (Guenther *et al*. 2012), and the rate of emission between individual tree species varies drastically (Table 1.1). One notable example of this can be seen in the genus *Quercus,* or oak, wherein all known species of oak in North America produce isoprene, whilst many Mediterranean species produce none at all (Csiky and Seufert 1999). It is interesting to note that while the majority of tropical trees do not produce isoprene, the bulk of isoprene production from trees globally come from those few tropical species that do (Geron *et al.*, 2006; Sharkey *et al.,* 2007; Guenther *et al.*, 2012). Trees are not the only terrestrial plants that emit isoprene, emissions have also been recorded in mosses (Hanson *et al*. 1999), grasses (Hewitt *et al.,* 1990), sedges (Bai *et al*. 2006) and some commercial crops (T. D. Sharkey, Wiberley, and Donohue 2007). A relatively old but very useful comparison of recorded isoprene emission in plants collated by the Hewitt group at the University of Lancaster can be found at http://es.lancs.ac.uk/cnhgroup/iso-emissions.pdf.

1.1.4 Biological sinks of isoprene

While the atmospheric oxidation of isoprene is a rapid process, biological sinks also play a role in the biogeochemical cycle of isoprene (Figure 1.2).

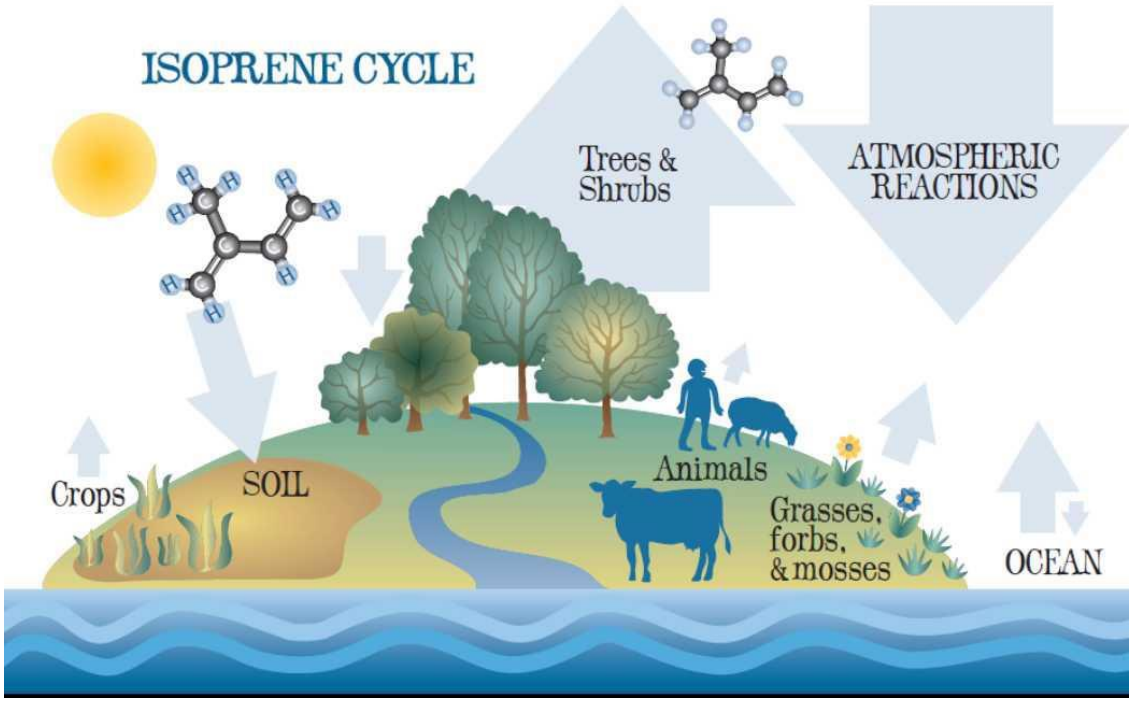

Figure 1.2. Biogeochemical cycle of isoprene. Arrows pointing up represent sources of isoprene production while downwards facing arrows represent isoprene sinks. Taken from McGenity *et al.* 2018.

The ability of microbial life within soils to act as an isoprene sink was first suggested by Cleveland and Yavitt (1997, 1998). In their study, soil samples collected from various arboreal environments around the world were incubated with 508 ppbv of isoprene in sealed containers at 25˚C in the dark. Isoprene degradation was measured via gas chromatography, and with appropriate controls accounted for, such as utilising autoclaved soils to ensure any observed reduction was due to microbial processes, results showed a consistent decrease in isoprene concentrations at a rate dependant on soil type. Since then *in situ* studies have shown rapid consumption of isoprene within temperate forest and tropical rainforest mesocosms (Pegoraro *et al.* 2006; 2005). Microbial uptake of isoprene over a range of concentrations (2-200 ppbv) was shown via continuous flow experiments utilising temperate forest soils (Gray *et al.,* 2015), these studies more closely matched the recorded average ambient concentrations of isoprene at 0.5-10 ppbv (Cleveland and Yavitt 1998). Consumption of isoprene in marine environments was first shown by Alvarez *et al.*

(2009), who examined this activity in estuarine, coastal and open marine communities. This study was also of note as it demonstrated the ability of isoprene degrading bacteria in microcosms to consume isoprene produced by marine microalgae cultures, proving localised communities could benefit directly from isoprene produced within their immediate vicinity and at concentrations relevant to environmental conditions. This new field of research informed the expanded model for the biogeochemical cycle of isoprene with inclusion of soils and marine environments as isoprene sinks as seen in Figure 1.2.

To better understand the mechanisms by which these soil microcosms act to degrade isoprene, isolation of individual degrading species was necessary. Early examples of isolated bacteria growing on isoprene as a sole carbon source came from studies by Ginkel *et al.*, (1987), Ewers *et al.,* (1990) and Cleveland and Yavitt (1997). These studies led to the isolation of isoprene-degrading bacteria from the genera *Rhodococcus, Arthrobacter, Nocardia* and *Alcaligenes.* While these isolates provided early information into the diversity of isoprene-degrading bacteria, they were not characterised in any great detail, and as such the mechanisms of degradation remained unknown.

1.1.5 Future of isoprene emissions and climate

There are a number of factors that suggest that isoprene emissions will fluctuate in the future as a result of both climate change and a shift in land-usage. A study by Feng *et al.* (2019) looked at environmental factors associated with climate change such as temperature rise, drought, increased CO_2 and increased O_3, conducting a meta-analysis to determine their impacts on isoprene emissions. They found that while a rise in global temperatures would result in a significant increase in isoprene emissions (+22%), all other environmental factors that were tested had the opposite effect. Exposure to chronic O_3 pollution in particular was shown to significantly reduce isoprene emission by -21%. Though it should be noted that the majority of the trees tested in this study were native to a temperate climate and not representative of emissions from warmer regions.

Previously, shifts in land-use were attributed mostly to deforestation for the expansion of food cropland. This had the combined impact of removing trees that produce isoprene and replacing them with food crops that do not, resulting in a 15% decrease in isoprene emissions between 1901 and 2002 (Lathière, Hewitt, and Beerling 2010). While both deforestation and the expansion of cropland are unlikely to cease in the near-future, the kinds of crops that are being planted have changed as the demand for biofuel and

alternative plant-based oils has increased. As seen in Table 1.1, the Malaysian oil palm produces 175 µg g (dw) h^{-1} of isoprene, and is also the single most cultivated crop in large swathes of South-East Asia such as Indonesia and Malaysia. Oil palm production in Malaysia has grown by a factor of 5 over the last 20 years, increasing at a rate of +7% a year in both Indonesia and Malaysia (Food and Agricultural Organization (FAO) 2015). Oil palm plantations cover >85% of total agricultural land in Malaysia (Cheng *et al*. 2019). The vast expansion of a single crop plantation that emits such high amounts of isoprene has led to serious concerns over the risk of localised increases in atmospheric isoprene levels and subsequent impacts on local air quality (Hewitt *et al*. 2009). A study by Stavrakou *et al.* (2014) stated that when oil palm plantations were explicitly taken into account with the use of the MEGAN model (Guenther *et al*. 2012), isoprene fluxes in Indonesia increased by 10%.

The combined picture creates a complex outlook for overall isoprene emissions in the future. Whether the cumulative effect of both climate change and land-usage will result in an increase or decrease in overall isoprene levels remains to be seen.

1.2 Biogenic production of isoprene in terrestrial plants

1.2.1 Isoprene biosynthesis

Isoprene is the basic subunit of terpenoids, a large and diverse class of organic compounds produced by all free-living organisms. Carotenoids, hopanoids, sterols, chlorophylls and archaeal lipids are all examples of terpenoids, which also make up many hormones and signalling molecules (Holstein and Hohl 2004; Rohmer 1999). Terpenoids are synthesized by condensations of iso-pentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). These intermediates are biosynthesized via two distinct pathways, the mevalonate pathway (MVA) or the non-mevalonate pathway, also called the methylerythritol phosphate (MEP) pathway (Figure 1.3). The MVA pathway is responsible for the biosynthesis of terpenoids in animals, fungi, archaea and some bacteria, whereas the MEP pathway is utilised by most other bacteria, and by plants (J. Yang *et al*. 2012; Rohmer 1999). The biosynthesis of isoprene in plants through the MEP pathway is catalysed by the isoprene synthase (IspS), a terpenoid synthase dependant on Mg that functions by catalysing the removal of pyrophosphate from DMAPP (Figure 1.3). IspS has been well-characterised in the *Populus x canescens* species (Köksal *et al*. 2010).

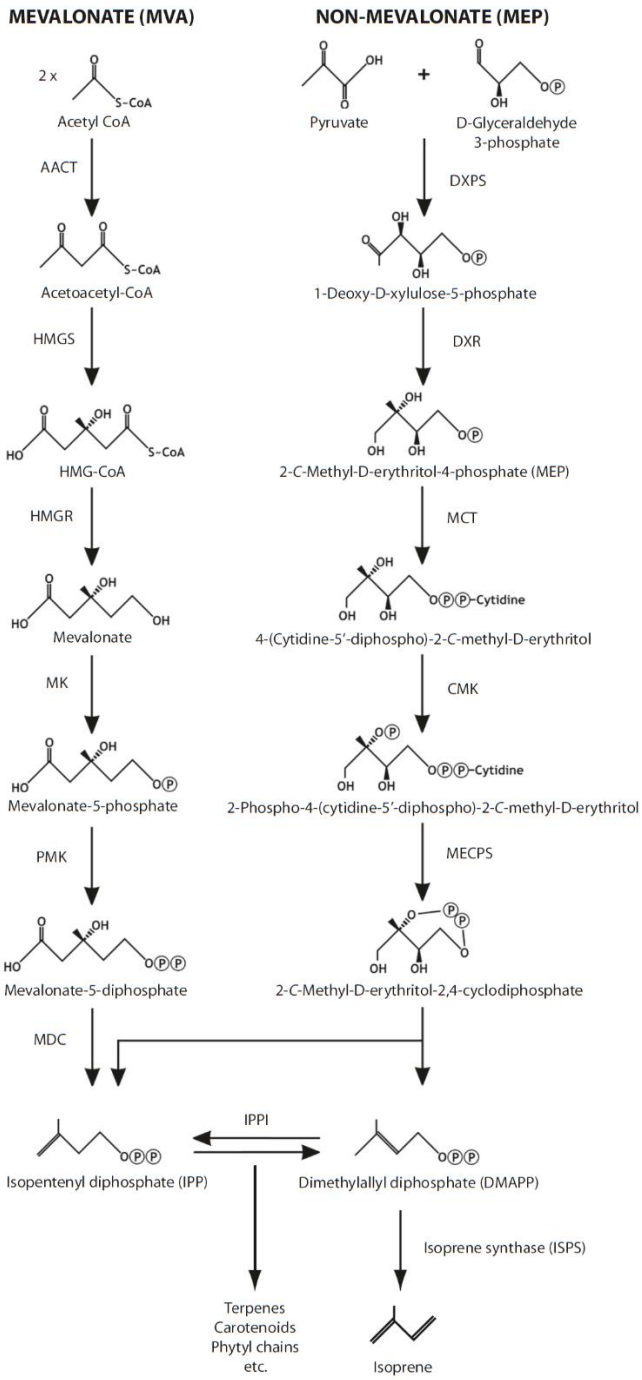

Figure 1.3. Biosynthesis of isoprene and isoprenoids via the MVA and MEP pathways. MVA pathway enzymes shown are: AACT acetoacetyl-CoA thiolase, HMGS 3-hydroxy-3-methylglutaryl-CoA synthase, HMGR 3-hydroxy-3-methylglutaryl-CoA reductase, MK mevalonate kinase, PMK phosphomevalonate kinase, MDC mevalonate 5-diphosphate decarboxylase. MEP pathway enzymes shown are: DXPS 1-deoxyxylulose 5-phosphate synthase, DXR 1-deoxyxylulose 5-phosphate reductoisomerase, MCT 2-C-methyl-D-erythritol 4-phosphate cytidyl transferase, CMK 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase, MECPS 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase. IPPI is isopentenyl diphosphate isomerase and ISPS is the isoprene synthase. Taken from McGenity *et al.*, 2018.

1.2.2 Metabolic role of isoprene biosynthesis

Synthesis of isoprene takes place in the chloroplasts of plants and is released via the stomata. Emission levels are dependant solely on biosynthesis rates however, and are not regulated by stomata control (Sharkey and Yeh, 2001). It has been shown that the carbon necessary for the production of isoprene is mostly derived from photosynthesis and can make up about 2% of a plant's photosynthetically-fixed carbon, requiring 20 ATP and 14 NADPH molecules (Fall and Monson 1992; Sharkey and Yeh 2001). Interestingly, the purpose of this marked energy investment is not yet fully understood. It has been reported that isoprene improves the resilience of plants to oxidative, thermal and biotic stress (Thomas D Sharkey and Yeh 2001; Behnke *et al*. 2007; Laothawornkitkul *et al.* 2008; Vickers *et al*. 2009), while other suggested roles for isoprene biosynthesis include predation avoidance, plant signalling and as a means to dissipate excess energy produced via photosynthesis (Magel *et al*., 2006; Loivamäki *et al*., 2008).

The hypothesis that isoprene acts as a thermoprotectant in plants was explored in a number of studies that showed that inhibiting the plant's endogenous isoprene synthesis lowered the temperature at which damage occurred to the photosynthetic apparatus, whilst reintroducing exogenous isoprene reversed the effect (Singsaas *et al*. 1997; Sharkey and Singsaas 1995). It is interesting to note however, that the addition of isoprene to plants that do not produce it, does not increase tolerance to heat stress (Logan and Monson 1999). None of the above studies clarified the mechanisms by which plants might utilise isoprene in such a way.

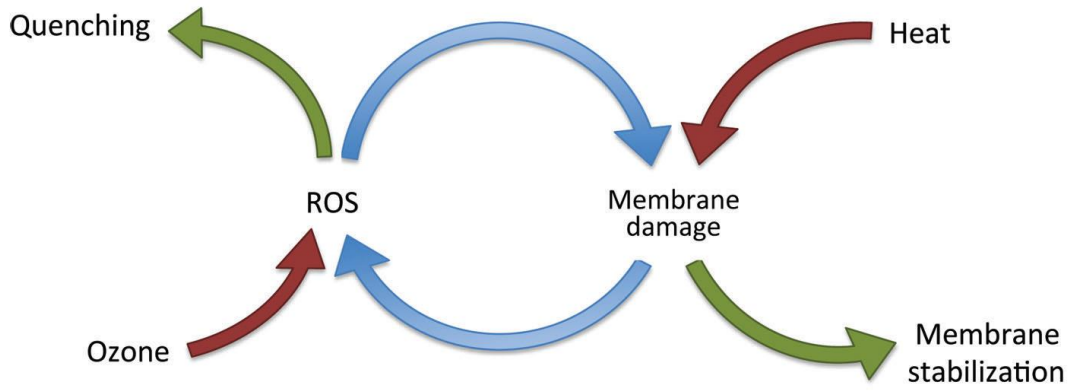

Figure 1.4. Proposed thermo- and oxidative protective role of isoprene in plants. Borrowed from Velikova *et al.* (2012).

It was originally hypothesized that isoprene acted to stabilise the photosynthetic membrane of chloroplasts by virtue of its lipophilic-hydrophilic properties. Damage caused to the photosynthetic apparatus of plants by high temperatures is due to thylakoid membrane dissolution. Isoprene was thought to negate this by intercalating into the membrane and tethering lipid bilayers together while strengthening adhesion of membrane proteins both to the bilayer and to one another. A study by Velikova *et al.* (2012) suggested that this process also acted as a feedforward cycle that involves isoprene's role as an antioxidant. In this study, isoprene was shown to quench reactive oxygen species (ROS) that can be caused by ozone stress in *Arabidopsis* plants. ROS cause damage to membrane stability which in turn release further ROS. This leads to a build-up of H_2O_2 and eventually to cell death. The study suggested that isoprene may prevent this deleterious feedforward cycle by both quenching ROS and simultaneously strengthening membrane stability as seen in Figure 1.4.

However, recent studies have disputed these older models and posit that due to the highly volatile nature of isoprene and its inability to dissolve well into cellular components, isoprene is unable to accumulate in chloroplastic membranes at a concentration high enough to provide any significant impact on membrane stability (Lantz *et al*. 2019).

It is interesting to note that while isoprene does appear to act as a thermoprotectant in isoprene emitting plants, it does not serve to protect against sustained heat stress, but specifically in circumstances of sudden sharp increases of temperature (Sharkey *et al.,* 2007). With this in mind, isoprene's function within marine algae is not well understood. Being an aquatic environment, marine habitats act as a robust buffer for rapid temperature changes. It could be that this is the reason isoprene emissions in the marine environment are lower than terrestrial levels. However, there is a definite correlation between marine isoprene emissions and temperature and light intensity. A study by Dani and Loreto (2017) suggest that isoprene's main role within marine algae is that of an antioxidant, protecting them from ROS similar to the previously mentioned Velikova *et al.* (2012) study in plants. However, the former study goes further, suggesting that dimethylsulfide (DMS) also serves this purpose, and that emissions of one or the other compound is a preferential behaviour varying between algal taxa. They speculate that due to isoprene-emitting marine algae having an evolutionary advantage over DMS-emitters in terms of withstanding the rising sea-surface temperatures caused by climate change, that marine isoprene emissions are likely to rise in tandem (Dani and Loreto 2017).

Isoprene emission in plants has shown to be both diurnal and seasonal. Emissions increase in the morning, reach their highest level during the afternoon and decline again thereafter (Wilkinson *et al*. 2006). This is supported by the circadian control of isoprene synthase upregulation (Loivamäki *et al*. 2007; Mayrhofer *et al*. 2005). In terms of seasonality, transcript levels of *IspS* were shown to start low in spring, increase rapidly with the first strong temperature increase of early summer, remain high throughout early summer but fluctuate with temperature and light variations, and then begin to decline in late summer and autumn (Mayrhofer *et al*. 2005).

Finally, a review by Lantz *et al.* (2019) suggested that isoprene may play a role in gene expression. Exposure to exogenous isoprene has been shown to illicit changes in gene expression in studies involving *Arabidopsis thaliana,* a plant that emits no isoprene itself (Harvey and Sharkey 2016). Later studies involving *Arabidopsis* transformed with the isoprene synthase of *Eucalyptus globulus* showed a similar result (Zuo *et al*. 2019). Lantz *et al.* (2019) show that the genes influenced by these changes in expression often play a role in thermotolerance, or protection from photo- and oxidative-stress. However, it is noted that in order for isoprene to illicit such changes in gene expression, plants must have a mechanism by which to detect it. At present, there are no known receptors for isoprene in plants, and direct evidence for this theory is still somewhat limited. As such, the biological role of isoprene, one of the most abundantly produced BVOCs on the planet, the biosynthesis of which warrants a substantial expenditure of energy from the plants that produce it, remains mostly a mystery.

1.3 Isoprene biodegradation

1.3.1 Metabolic pathway of isoprene degradation

Section 1.1.4 introduced the concept of bacteria as a biological sink for isoprene. Early isolation of *Rhodococcus, Arthrobacter, Nocardia* and *Alcaligenes* by Ginkel *et al.*, (1987), Ewers *et al.* (1990) and Cleveland and Yavitt (1997) opened the door to research investigating the diversity of such bacteria, but it was not until van Hylckama Vlieg *et al.* (1998) isolated a Gram-positive Actinobacteria, identified as *Rhodococcus* AD45, that characterisation of the isoprene degradation pathway really took off.

Rhodococcus AD45 was shown to oxidize isoprene to 1,2-epoxy-2-methyl-3-butene in the first step of the isoprene degradation pathway, and in a study published a year later it was

shown that this product is acted upon by a glutathione S-transferase (IsoI). IsoI carried out the detoxification of epoxyisoprene by conjugation with glutathione, forming 1-hydroxy-2-glutathionyl-2-methyl-3-butene (HGMB) (van Hylckama Vlieg *et al*. 1998; 1999; 2000). IsoI was subsequently purified and characterised along with an associated dehydrogenase IsoH (van Hylckama Vlieg *et al*. 1999). IsoH catalysed two sequential oxidation steps, yielding 2NADH and 2-glutathionyl-2-methyl-3-butenoic acid (GMBA; Figure 1.5). To date the fate of GMBA after this step remains unknown, though Crombie *et al.* (2018) suggest that the subsequent removal of the glutathione moiety and β-oxidation of the intermediates of isoprene metabolism allows the species to grow on isoprene as its sole carbon source.

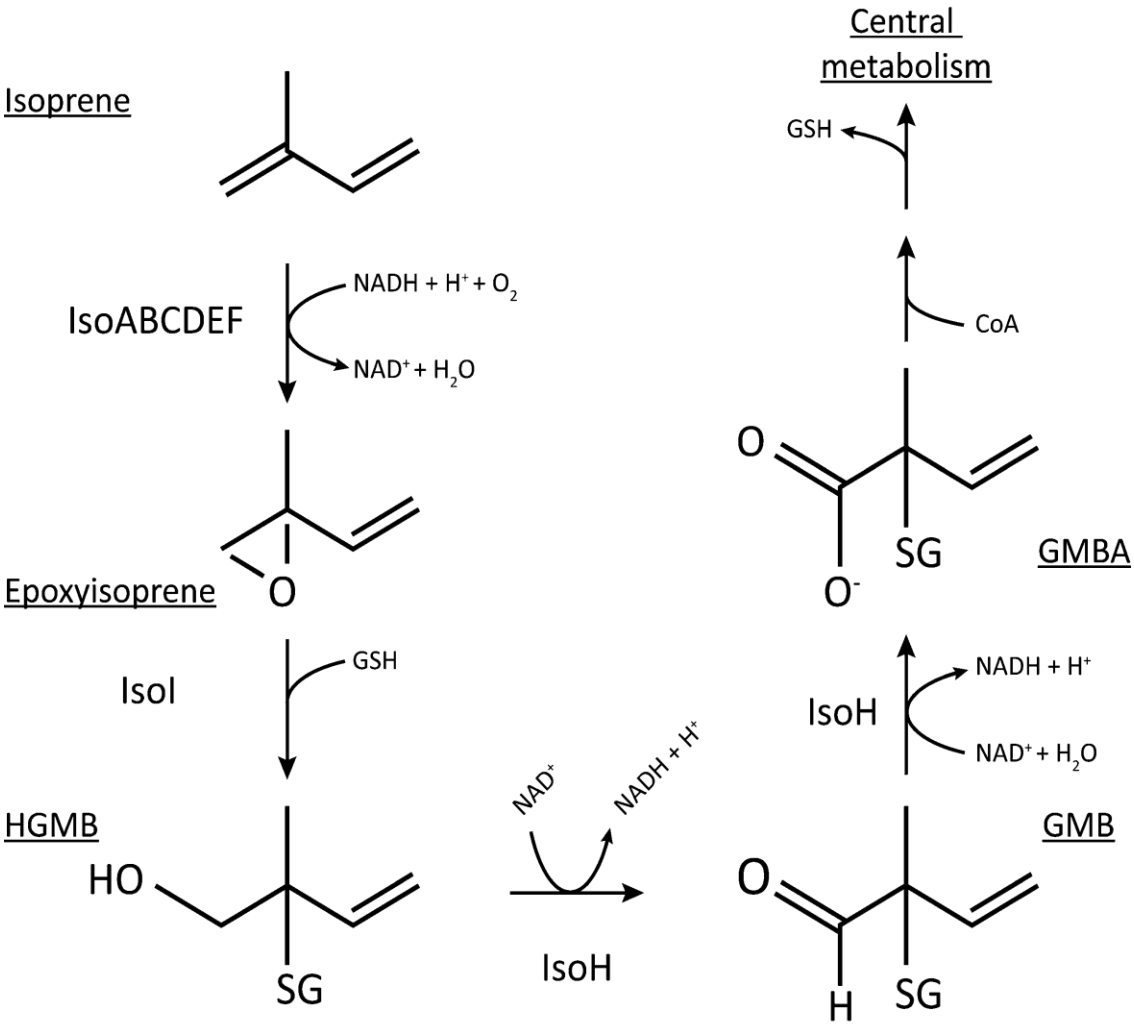

Figure 1.5. Metabolic pathway for isoprene degradation. Taken from Crombie *et al.* (2018).

With the identification of these enzymes, the design of primers to target the genes that encoded them became possible. This led to the construction of a gene library for *Rhodococcus* AD45 and the identification of a gene cluster involved in the isoprene degradation pathway (van Hylckama Vlieg *et al*. 2000). This gene cluster was predicted to encode for an isoprene monooxygenase (IsoMO) based on its sequence similarity to toluene monooxygenase from the Gram-negative species *Pseudomonas mendocina* KR1 (Whited and Gibson 1991). Later, IsoMO was shown to be a multicomponent soluble diiron monooxygenase (SDIMO) encoded for by the genes *isoABCDEF*, with all of these genes present on a 300 kbp megaplasmid (Leahy *et al.*, 2003; Holmes and Coleman, 2008; Crombie *et al.*, 2015). Crombie *et al.* (2015) confirmed that IsoMO was essential for isoprene metabolism by constructing an *isoA* deletion mutant of *Rhodococcus* AD45 and demonstrating subsequent inhibition of growth on isoprene. This study also used RNA sequencing to investigate *Rhodococcus* AD45 cells exposed to isoprene or epoxyisoprene, and found that 22 contiguous genes, including *isoABCDEFGHIJ,* were highly upregulated as a result under both conditions. This confirmed the isoprene degradation gene cluster as we know it today (Figure 1.6).

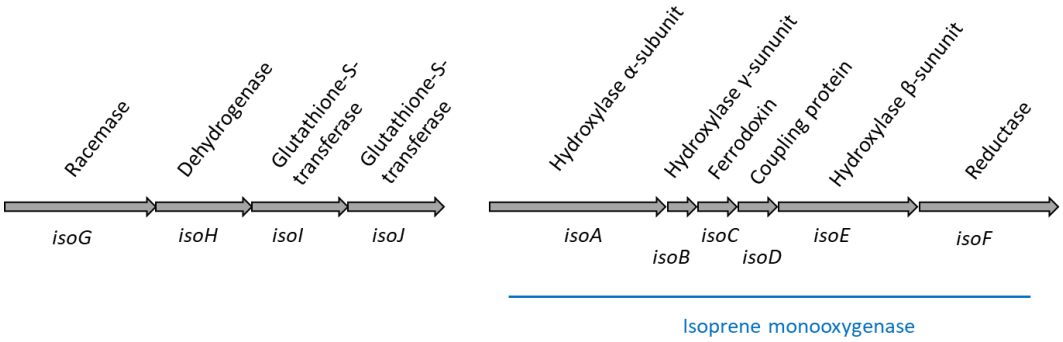

Figure 1.6. Organisation of the isoprene degradation gene cluster

As can be seen in Figure 1.6, i*soA, B* and *E* are responsible for encoding the diiron oxygenase component of the IsoMO, *isoF* encodes a flavoprotein NADH reductase, *isoC* encodes a Reiske-type ferredoxin, and together with *isoD* which encodes a coupling protein, form the multicomponent isoprene monooxygenase. Additional genes, *isoG, isoH, isoI* and *isoJ* encode a putative coenzyme A transferase, a dehydrogenase and two glutathione transferases, all of which are located upstream of the IsoMO genes (van Hylckama Vlieg *et al*. 1998; 1999; 2000).

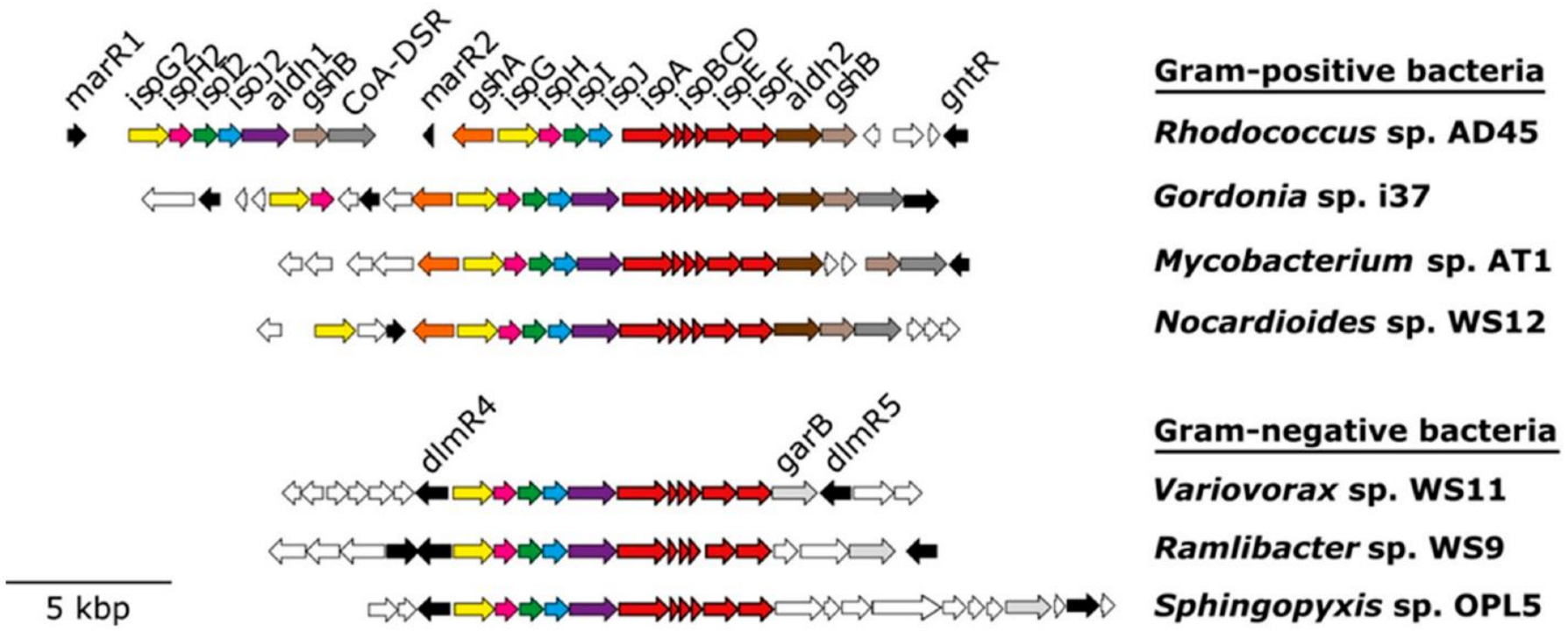

Figure 1.7. Isoprene degradation gene clusters from representative Gram-positive and Gram-negative isoprene degrading bacteria. Genes encoding the IsoMO (*isoABCDEF*) are coloured red. Regulatory genes are shown in black. Adjacent genes not known to be involved with isoprene degradation are shown in white. Borrowed from Carrion *et al.* (2020).

An interesting pattern that has emerged as more isoprene degrading isolates have been characterised, is the duplication of certain downstream genes in the isoprene degradation pathway. This trait is so far seen specifically in Gram-positive isolates wherein some or all of the genes *isoGHIJ* have non-identical copies (Crombie *et al.*, 2015; Johnston *et al.*, 2017; Gibson *et al.*, 2020 ; Figure 1.7). Work by Crombie *et al.* (2015) suggests that these duplicate genes serve to aid in the prevention of epoxide accumulation during isoprene metabolism, and that their absence in Gram-negative isoprene degrading bacteria is due to a better-adapted systems for mitigation of epoxyisoprene-induced stress. However, more Gram-negative isolates are required to determine whether the pattern between Gram-positive and negative isoprene degrading bacteria and the presence of duplicate *iso* genes holds true.

1.3.2 Diversity of isoprene degrading bacteria today

Having identified soils as a sink for isoprene (see section 1.1.4), and showing the action was a microbial process (confirmed by autoclaving soils which prevented further oxidation of isoprene), further attempts to isolate bacteria responsible for isoprene degradation were carried out. Many of the earlier representatives of isoprene degrading isolates were Gram-positive Actinobacteria, such as *Gordonia, Micrococcus, Mycobacterium* and *Rhodococcus,* with *Rhodococcus* seeming to be particularly abundant in a range of environments (Ewers *et al.*, 1990; van Hylckama Vlieg *et al.*, 1998; Alvarez *et al.*, 2009; El Khawand *et al.*, 2016; Crombie *et al.*, 2017; Johnston *et al.*, 2017; Murphy, 2017; Figure 1.7).

Earlier studies focused on aquatic environments, with species of *Rhodococcus, Gordonia* and *Mycobacterium* isolated from marine sediments, estuarine and fresh water samples (Alvarez *et al.*, 2009; Johnston *et al.*, 2017). The study by Alvarez *et al.* (2009) also showed that marine sediments harboured a greater isoprene degrading potential than water samples in estuarine environments. It was noted that the rate of isoprene uptake was significantly increased in circumstances where samples were incubated with lower concentrations of isoprene. It was also noted that communities from sampling locations that experience lower atmospheric levels of isoprene showed higher rates of isoprene uptake. This suggested that higher concentrations of isoprene may have a toxic or inhibitory effect on isoprene metabolism in the natural community.

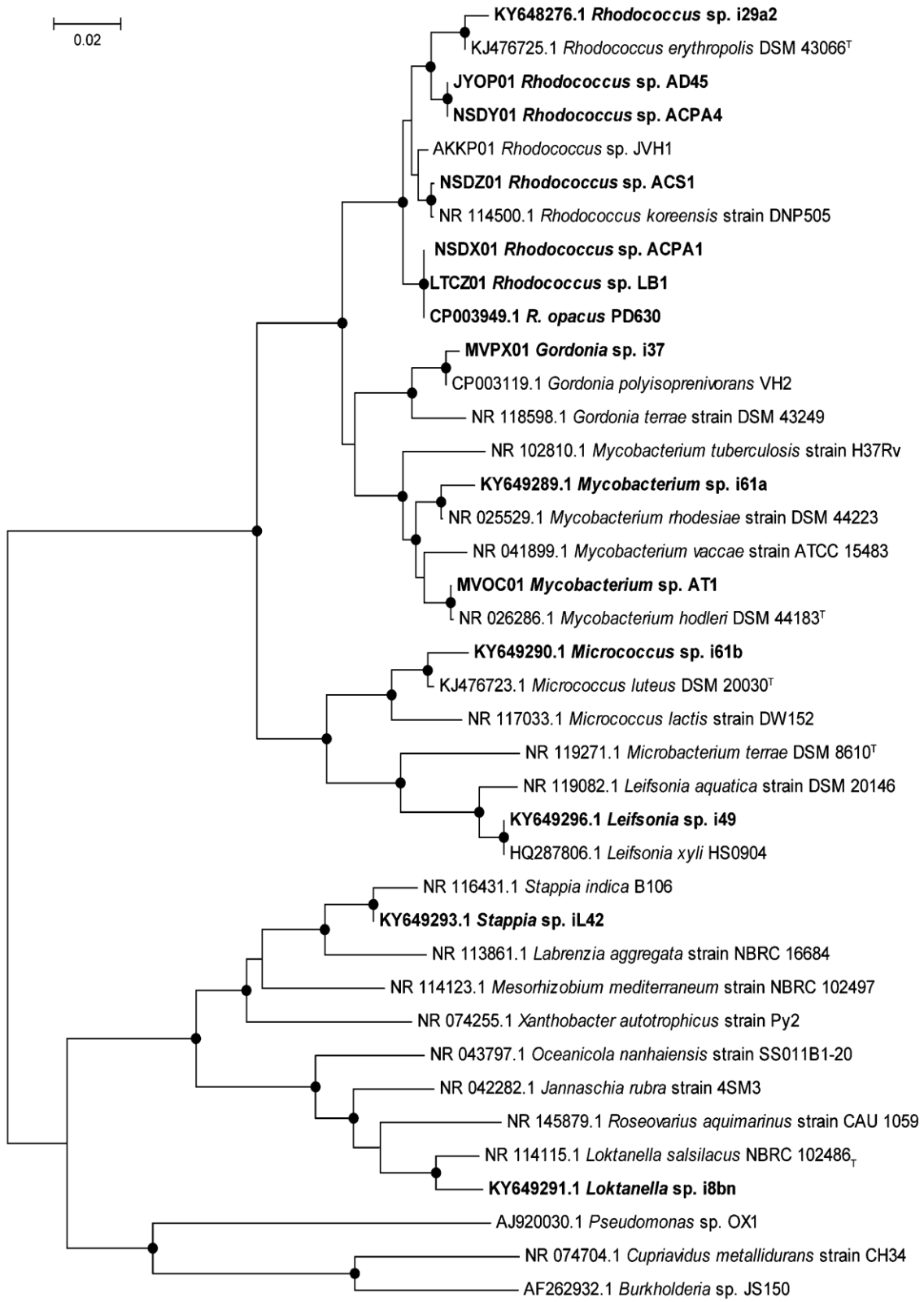

Figure 1.8. 16S rRNA gene-based phylogeny of isoprene degrading bacteria. Taken from Crombie *et al.* (2018) showing isolate strains of isoprene degrading bacteria available at time of publication (in bold) alongside other representative strains.

The effect of isoprene concentration also appeared to extend to the phylogeny of the bacteria that were isolated in these experiments. Earlier studies that were dominated by representatives of the Actinobacteria phylum (the phylogeny of which can be seen in Figure 1.8) were often characterised by relatively high concentrations of isoprene during enrichment. For example, earlier DNA-Stable Isotope Probing (DNA-SIP) experiments (explored in detail in section 1.4.2) utilised 0.5% (v/v) isoprene and consistently showed Actinobacteria to dominate the enriched communities they recovered (El Khawand *et al.*, 2016; Johnston *et al.*, 2017).

A study by Crombie *et al.* (2018) investigated this effect by carrying out two separate DNA-SIP experiments on the same samples, one involving incubation with labelled isoprene at 500 ppmv and the other with 150 ppmv. In this study, both incubations showed *Rhodococcus* was the dominant genus, followed by *Comamonadaceae,* however in the lower concentration enrichment, the abundance of *Comamonadaceae* increased from 3% relative abundance (RA) to 16% RA of the whole isoprene degrading community, with *Variovorax* being most abundant. This result suggests that incubation with higher isoprene concentrations may bias community analyses towards an overabundance of Actinobacteria as suspected (Table 1.2).

Later studies utilising the DNA-SIP method subsequently lowered the concentration of isoprene used during incubation. A study investigating the isoprene degrading bacteria harboured by soils associated with willow trees used 25 ppmv during the incubation period and identified a community profile quite distinct from previous studies (Larke-Mejia *et al.*, 2019; Table 1.2). Here, Gram-negative Proteobacteria were dominant, specifically *Burkholderiales. Ramlibacter* and *Variovorax* represented a combined 72% of the isoprene degrading community at the first timepoint, with *Rhodococcus* representing only 4.5% relative abundance (RA). Because the environment studied in this experiment is different to those previously tested with higher concentrations of isoprene, it means the species of tree these bacteria were associated with, or other abiotic factors, cannot be ruled out as the variable responsible for the distinct difference in community profile. It does suggest however, that Actinobacteria thrive mostly under higher isoprene conditions, and in situations where lower concentrations are available, a more diverse community can thrive.

Table 1.2. Isoprene degrading community identified via DNA-SIP. Active isoprene degraders are listed in order of abundance. Adapted from Carrión *et al.* (2020).

Study	Environment	Isoprene Concentration (ppm)	Active Isoprene Degraders
El Khawand *et al.*, 2016	Willow soil	5000	*Rhodococcus, Variovorax, Comamonas*
Johnston *et al.*, 2017	Estuarine water and sediment	2000	*Microbacterium, Rhodococcus, Mycobacterium, Gordonia*
Crombie *et al.*, 2018	Poplar leaves	500	*Rhodococcus, Xanthamonadaceae, Comamonadaceae*
		150	*Rhodococcus, Variovorax*
Larke-Mejía *et al.*, 2019	Willow soil	25	*Ramlibacter, Variovorax, Rhodococcus*
Larke-Mejía *et al.*, 2020	Oil palm leaves	25	*Gordonia, Sphingomonas, Aquincola*
	Oil palm soil	25	*Aquabacterium, Rhodococcus, Saccharibacter*

The variety of isoprene degrading bacterial isolates grew wider. *Pseudomonas* and *Klebsiella* isoprene degrading strains were isolated from rubber-contaminated soil (Srivastva *et al*. 2015), while isoprene degrading strains from the genera *Arthrobacter, Bacillus, Pseudomonas, Sphingobacterium, Sphingobium*, and *Pantoea* were isolated from the soils and leaves associated with tropical teak (*Tectona grandis*) and mahua trees (*Madhuca latifolia*) (Singh, Srivastava, and Dubey 2019). Following DNA-SIP experiments that indicated the presence of isoprene degrading members of the *Comamonadaceae* family, Larke-Mejia *et al.* (2019) carried out a targeted isolation experiment. The study utilised lower levels of isoprene during enrichment (25 ppmv) to avoid specifically selecting for Actinobacteria. This method was successful and resulted in the isolation of some of the first Gram-negative isoprene-degrading bacteria. Enrichments from soils and leaves associated with a willow and an oil palm tree from Kew Gardens, London, contained isoprene degrading strains of *Variovorax, Sphingopyxis* and *Ramlibacter* (Larke-Mejia *et al*. 2019). *Variovorax* sp. WS11 would go on to be investigated as a model organism to complement research done in *Rhodococcus* AD45 but within a novel, genetically pliable Gram-negative system (Dawson *et al*. 2020; Dawson 2021).

1.4 Molecular ecology methods to survey distribution and diversity of isoprene degrading bacteria

1.4.1 Functional gene probes

The range of isoprene degrading isolates described in section 1.3.2 indicates that isoprene metabolism appears to be a widespread characteristic found in many phylogenetically diverse bacteria. While enrichment cultures are a useful tool to identify bacteria with specific metabolic traits, it is time consuming as a means of widespread screening. Universal primers that target genes like the 16S rRNA gene for phylogenetic identification are essential for the characterisation of bacteria, but do not give any information as to their metabolic capabilities.

In the past, studies investigating methane metabolism successfully utilised another SDIMO, the soluble methane monooxygenase (sMMO) to screen for methane degrading bacteria in the environment. These studies designed functional gene probes targeting *mmoX*, the α-subunit of the hydroxylase of the sMMO enzyme, using them to significantly increase our knowledge of methane oxidising bacteria in the environment. (McDonald *et al.,* 1995; Dumont and Murrell, 2005a; McDonald *et al.*, 2008; Ul Haque *et al.*, 2018).

Once a range of isoprene degrading isolates had been obtained and the genetic pathway of isoprene degradation elucidated, a similar approach to investigating isoprene degrading bacteria became possible. Earlier studies were able to utilise the genetic information obtained from cultured isoprene degrading organisms available at the time, to produce functional gene probes that allowed for targeted detection of such bacteria in environmental samples. These probes were designed to detect and amplify *isoA*, encoding the α-subunit of the isoprene monooxygenase, within which the active site of the enzyme is located. As a conserved gene essential for isoprene degradation (as described in section 1.3.1), *isoA* is considered an ideal biomarker for isoprene metabolism (Crombie *et al*. 2015; El Khawand *et al*. 2016). However, these earlier *isoA* gene probes were limited by the diversity of the isoprene degrading bacteria mined to design them, meaning earlier primers were biased towards *Rhodococcus*-like sequences and those recovered from other Actinobacteria.

As with all gene probes, there is a constant need to update their design as *isoA* sequences from a wider variety of bacteria are recovered. With every addition of a phylogenetically diverse *isoA* sequence to databases, the subsequent primer design becomes more robust and capable of capturing a more accurate profile of *isoA* genes in the environment.

1.4.2 DNA-SIP

While techniques such as PCR, cloning and amplicon sequencing allow us to easily identify microbes from the environment at a molecular level, they do not readily allow for linkages between phylogeny and function at any relevant scale, especially with regards to the many unculturable bacteria known to make up the bulk of the planet's microbiome. Techniques such as DNA-SIP allow us to link the identity and function of bacteria in the environment at a broad scale, and allow us to specifically investigate groups of bacteria endowed with a shared and perhaps less common metabolic trait from within a complex community (Radajewski *et al.*, 2000; Dumont *et al.*, 2005b; Figure 1.9).

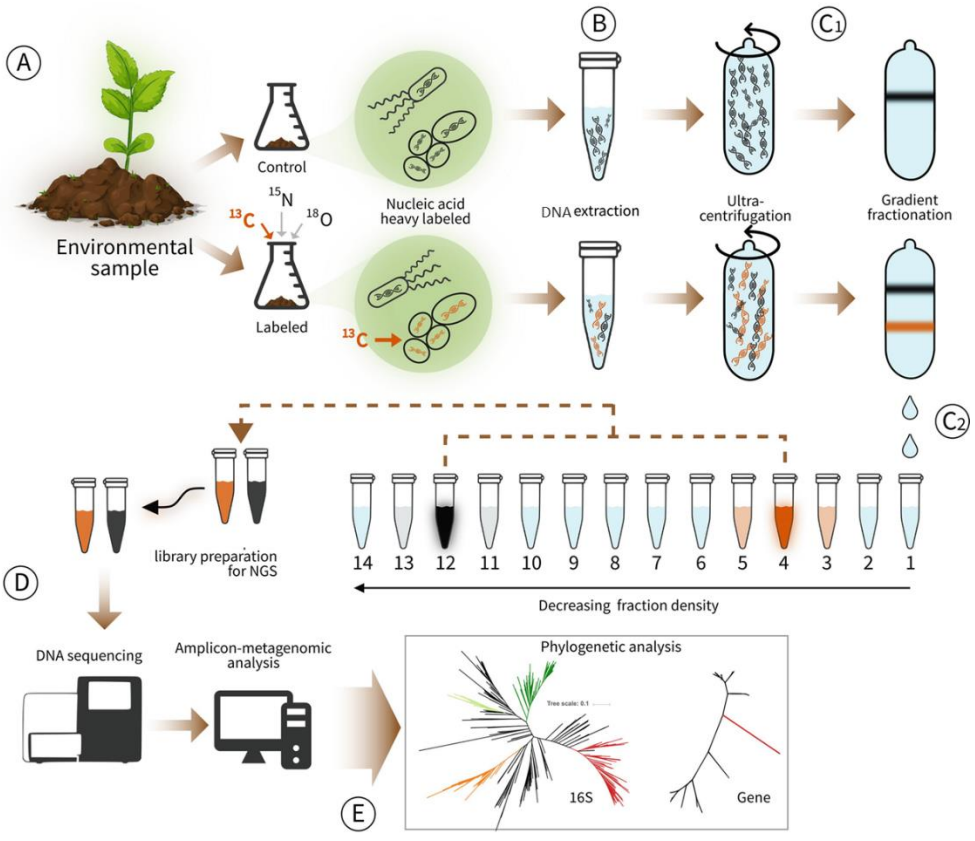

Figure 1.9. Graphical workflow of a DNA-SIP experiment. (a) Environmental samples are enriched with isotopically labelled and unlabelled control substrate. (b) Control and labelled DNA is extracted and suspended in a caesium chloride solution that undergoes isopycnic ultracentrifugation to distribute DNA into a gradient determined by buoyant density. (c_1) DNA from bacteria that have incorporated the isotopically labelled substrate is found lower in the gradient. (c_2) Samples undergo fractionation to recover DNA of varying buoyant density in individual tubes. (d) Fractions containing 'heavy' isotopically labelled DNA and 'light' unlabelled DNA are sequenced. (e) Downstream analyses such as amplicon sequencing or metagenomics can be carried out to obtain phylogenetic and functional information of the target population. Adapted from Ghori *et al.* (2019).

The principal behind this technique was demonstrated in 1958 by Meselson and Stahl who showed that stable isotope labelled ^{15}N and ^{14}N could be incorporated into DNA and subsequently separated via density-gradient centrifugation (Figure 1.9). Since then, the technique has been shown to be compatible with use of ^{2}H and ^{13}C in the labelling process. Earlier experiments utilising labelled carbon focused on methylotrophs and their incorporation of $^{13}CH_3OH$ and $^{13}CH_4$, added to soil samples or enrichment cultures. (Radajewski *et al*. 2002; Morris *et al*. 2002; Radajewski *et al*. 2000).

The specifics of DNA-SIP as a method can be found in section 2.5. The process has been a well-used and successful tool in the investigation of the diversity and distribution of isoprene degrading bacteria in the past. To date, DNA-SIP has been used by El Khawand *et al.* (2016) and Larke-Mejia *et al.* (2019) to investigate the soils associated with willow trees. Crombie *et al.* (2018) used DNA-SIP to investigate the isoprene degrading community of a poplar phyllosphere, Larke-Mejía *et al.* (2020) to investigate oil palm leaves from Kew Gardens, and Johnston *et al.* (2017) to investigate isoprene degrading bacteria from an estuarine environment. Through these studies (the results of which are summarised in Table 1.2), we have a much broader idea of the diversity of isoprene degrading bacteria than could ever be obtained through traditional isolation methods. In fact, many attempts to isolate bacteria with DNA found in the 'heavy' fraction of ^{13}C labelled samples, who were found to contain *isoA*-like sequences, have been unsuccessful. This shows the value of culture-independent methods of identification.

1.4.3 Metagenomic sequencing

One of the culture-independent methods most heavily used in tandem with DNA-SIP is that of metagenomic sequencing. Metagenomics is the study of collective genomes in an environmental community. Unlike amplicon sequencing which utilises a gene target approach, metagenomics uses a whole-genome shotgun method to fragment and sequence the entire bulk DNA of a community. As a result, the reads generated through metagenomics are not biased by primer specificity and can originate from bacteria, viruses, phages, archaea, fungi and other eukaryotes. Metagenomic data also includes plasmid DNA along with extra-chromosomal, host, chloroplast and mitochondrial elements.

While metagenomics as a concept opened up a wide and exciting vista of microbiome research, it does have its drawbacks. Specifically, the sequencing depth required to analyse resident bacteria in a studied environment at any resolution is inversely proportional to the

abundance of said bacteria. This is an issue if the purpose of the study is to investigate a particular metabolic process carried out by a relatively small group of bacteria. Increasing the depth of metagenomic sequencing to obtain enough reads from these groups can quickly become cost prohibitive (Xu 2006; Nazir 2016; Fadiji and Babalola 2020).

The combination of DNA-SIP and metagenomics sequencing is an elegant solution to this issue (reviewed in Chen and Murrell, 2010). With the use of DNA-SIP allowing the separation of DNA from bacteria that carry out isoprene degradation from those that do not, it greatly reduces the complexity of the community being sequenced and thus lowers the depth required for informative resolution (Schloss and Handelsman 2003). This decreased complexity and the resulting more granular data on individuals responsible for isoprene degradation then allows for the successful creation of metagenome assembled genomes or MAGs. With the use of MAGs, unculturable bacteria can be analysed in a similar manner to whole genome sequencing or WGS, with the potential for entire metabolic pathways to be identified and explored.

This method has been successfully used in previous studies investigating bacteria involved in the isoprene cycle. Crombie *et al.*, (2018) combined metagenomic sequencing with DNA-SIP while investigating the phyllosphere of a poplar tree. In this study they were able to recover a number of MAGs closely resembling isoprene degrading *Rhodococcus* isolates, along with a *Pseudonocardia* MAG and one designated as *Variovorax* before a physical isolate of that genera had been obtained. Larke-Mejia *et al.* (2019) used a similar process while investigating soils associated with a willow tree. Here metagenomic analysis allowed for the recovery of a number of MAGs identified as belonging to the family *Burkholderiales,* and one identified as *Sphingomonadales* harbouring genes related to isoprene biodegradation.

A drawback of the DNA-SIP – metagenomic approach is that it severely reduces the throughput potential of stand-alone metagenomic methods. DNA-SIP experiments, particularly ones based on isoprene metabolism, are both time- and labour-expensive. Also, the amount of DNA required for metagenomic sequencing is quite high, while the DNA obtained from individual fractions of a DNA-SIP experiment are typically very low.

An alternative approach would be the use of RNA-SIP, where the use of RNA instead of DNA allows for a specific insight into active communities. RNA-SIP is in general a more sensitive method due to an earlier and faster accumulation of isotope to RNA than to DNA. Benefits of RNA-SIP include faster labelling of active organisms, the ability to label active

but non-replicating bacteria and it also removes the impact of GC content on the fractionation process. However, despite the higher sensitivity of the RNA-SIP method, downstream applications are limited to those that include rRNA with implications for its usefulness in accurately interpreting the physiological role of bacteria in the studied environment (Neufeld *et al*. 2007; Jameson *et al*. 2017).

As such, to date, the use of metagenomics with DNA-SIP has been an important method in the study of the ecology of isoprene degrading bacteria and features heavily in the following research.

1.5 Aims & Objectives of this Study

The primary aim of this study was to contribute to our understanding of the role that bacteria play as biological sinks in the biogeochemical cycle of isoprene. There were three main areas of interest that the following research set out to address.

The first was to identify the bacteria involved in the biodegradation of isoprene. While this PhD project was built on a body of work carried out by previous lab members that explored this question (the results of which are detailed in Table 1.2), each experiment further expanded our understanding as to the diversity of isoprene degrading bacteria, and indicated that there was still much to be discovered with regards to their phylogenetic range.

The second area of interest was in relation to the distribution of isoprene degrading bacteria. Isoprene itself is ubiquitous and is emitted across the globe. This project aimed to contribute to our understanding as to the geographical distribution of environments that may harbour these species. A specific focus of this aim was to examine whether different locations, biomes or host species, particularly those that produce or experience high levels of isoprene, might have an impact on the presence and diversity of the associated isoprene degrading bacteria.

Finally, this project aimed to investigate the abundance of isoprene degrading bacteria in terrestrial environments associated with both isoprene-emitting and non-emitting plants. To date, all research on their ecology focused on enrichment-based methods that do not reflect the bacterial community harboured in the natural environment. If we are to truly understand the role that bacteria play in the biogeochemical cycle of isoprene, we must

first determine how common a trait isoprene metabolism is without the addition of artificial selective pressures. This project aimed to take early steps in addressing the question, to what extent do bacteria in the environment contribute to the biogeochemical cycle of isoprene under natural conditions.

2 Materials and Methods

2.1 Materials

All molecular biology- and analytical-grade components were purchased from Thermo Fisher Scientific (Waltham, MA, USA), Sigma-Aldrich (St Louis, MO, USA), Melford (Chelsworth, Ipswich, UK) or Merck (Kenilworth, NJ, USA). Gases were purchased from Sigma-Aldrich, Air Liquide UK (Birmingham, UK), BOC (Manchester, UK) or CK Special Gases Limited (Newton Unthank, UK). Custom oligonucleotides were synthesised by Thermo Fisher Scientific.

2.2 Cultivation methods

2.2.1 Growth media, buffer preparation and antibiotics

All growth media and solutions were prepared with distilled water and subsequently autoclaved at 15 psi for 15 minutes at 121˚C. Solutions or components sensitive to extreme heat were sterilised with the use of a sterile, disposable 0.2 µm syringe filter (Sartorius, Göttingen, Germany) and added when autoclaved solutions had cooled to room temperature. Solid media were prepared with the addition of 1.5% (w/v) Bacto Agar (Difco, Waltham, MA, USA). Rich solid medium was prepared with 1.8% (w/v) R2A agar (Oxoid Limited, Hampshire, UK). Ampicillin was filter sterilized and added to autoclaved, cooled growth medium to a concentration of 100 µg ml^{-1}. Liquid media utilised were prepared as follows.

Ewers medium (Dorn *et al.* 1974)

(10x) mineral salts:

$(NH_4)_2SO_4$	10 g L^{-1}
$MgSO_4.7H_2O$	2 g L^{-1}
$Ca(NO_3)_2.4H_2O$	0.725 g L^{-1}

(100x) Fe Ammonium Citrate:

Fe Ammonium Citrate	1 g L^{-1}

(1,000x) Trace Elements (SL-6) (Quayle & Pfennig 1975)

$ZnSO_4.7H_2O$	10 mg L^{-1}
$MnCl_2.4H_2O$	3 mg L^{-1}
H_3BO_3	30 mg L^{-1}
$CoCl_2.6H_2O$	20 mg L^{-1}
$CuCl_2.2H_2O$	1 mg L^{-1}
$NiCl_2.6H_2O$	2 mg L^{-1}
$Na_2MoO_4.2H_2O$	3 mg L^{-1}

1 M Phosphate Buffer (pH 6.0)

KH_2PO_4	0.5 M
Na_2HPO_4	0.5 M

(1,000x) Marine Ammonium Mineral Salts (MAMS) vitamins (Kanagawa *et al*. 1982)

Thiamine hydrochloride	10 mg L^{-1}
Nicotinic acid	20 mg L^{-1}
Pyroxidine hydrochloride	20 mg L^{-1}
p-aminobenzoic acid	10 mg L^{-1}
Riboflavin	20 mg L^{-1}
Calcium pantothenate	20 mg L^{-1}
Biotin	1 mg L^{-1}
Cyanocobalamin B12	2 mg L^{-1}
Lipoic acid	5 mg L^{-1}
Folic acid	5 mg L^{-1}

Ammonium Nitrate Mineral Salts (ANMS) (Adapted from Brenner *et.al* 1984)

(10x) ANMS salts:	
KNO_3	5 g L^{-1}
NH_4Cl	5 g L^{-1}
$MgSO_4.7H_2O$	10 g L^{-1}
$CaCl_2.2H_2O$	2 g L^{-1}
(1,000x) Fe EDTA:	
Fe EDTA	0.38 g L^{-1}
(100x) Sodium Molybdate:	
$Na_2MoO_4.2H_20$	0.26 g L^{-1}
(100x) Trace Elements	
$CuSo_4.5H_2O$	0.25 g L^{-1}
$FeSO_4.7H_20$	0.5 g L^{-1}
$ZnSo_4.7H_20$	0.4 g L^{-1}
H_3BO_3	0.015 g L^{-1}
$CaCl_2.6H_2O$	0.05 g L^{-1}
EDTA disodium salt	0.25 g L^{-1}
$MnCl_2.4H_2O$	0.02 g L^{-1}
$NiCl_2.6H_2O$	0.01 g L^{-1}
1 M Phosphate Buffer (pH 6.8)	
KH_2PO_4	0.5 M
$Na_2HPO_4.12H_20$	0.5 M

Marine Basal Medium (Baumann *et al.* 1981)

(4x) Basal media:	
Tris HCl 1M (pH 7.5)	150 mL
K_2HPO_4	87 mg
NH_4Cl	1.5 g
H_2O	375 mL
(20x) Fe EDTA	
Fe EDTA	0.5 g L-1
(1,000x) Vitamin supplement	
Biotin	20 mg L-1
Folic acid	20 mg L-1
Pyridoxine HCl	100 mg L-1
Riboflavin	50 mg L-1
Thiamine	50 mg L-1
Nicotinic acid	50 mg L-1
Pantothenic acid	50 mg L-1
Cyanocobalamin B12	1 mg L-1
P-aminobenzoic acid	50 mg L-1
Sea Salts	
Sea Salts	1 g/psu/L

2.2.2 Enrichment, isolation and maintenance

Enrichment of isoprene degrading bacteria from environmental samples was carried out in liquid medium initially, MBM was used for marine isolates and Ewers for terrestrial or freshwater isolates. Environmental enrichments were grown in sterile 125 mL serum vials sealed with butyl rubber seals and supplied with 25 ppmv gaseous isoprene. Isoprene consumption was measured by gas chromatography (GC) as described in section 2.2.3. Sub-culturing was carried out three times at 2-week intervals. Cultures were then plated to solid minimal medium and incubated in an air-tight container supplied with 25 ppmv isoprene alongside minimal media plates containing no culture as a control. Colonies that formed were individually streaked to fresh minimal medium plates and grown again under the same conditions. Biomass was gathered from plates and isolates were inoculated into fresh minimal liquid medium, supplied with 25 ppmv isoprene and incubated at 30˚C with shaking in the dark. Isoprene levels in the headspace were monitored daily via GC. Upon confirmation of consistent isoprene uptake, isolates were stored on rich R2A agar or minimal solid medium stored in anaerobic jars with isoprene for short-term storage. Isolates were identified and the presence of *isoA* genes confirmed utilising 16S rRNA gene and *isoA* gene PCR methods described in section 2.4.2

Nocardioides sp. WS12 was grown on ANMS medium with a pH of 6.8, supplemented with 1% (v/v) glucose or 125 ppmv isoprene in the headspace. Cultures were grown at 25˚C with shaking in the dark. *Nocardioides* sp. WS12 was stored on R2A agar or ANMS solid medium for short-term storage. Long-term storage at -80˚C was in the form of concentrated cultures supplemented with 10% (v/v) glycerol and snap frozen in liquid nitrogen.

Purity of isolate cultures was routinely monitored by serial dilution and subsequent plating to solid rich medium. Cultures were also observed via microscopy at 1,000x magnification to check for contamination with the use of a Zeiss Axioskop 50 microscope, 130 VA Type B (Carl Zeiss Ltd, Cambridge, UK).

2.2.3 Quantification of isoprene uptake

Isoprene uptake by isolates and environmental cultures was measured by gas chromatography. An air-tight glass syringe (Agilent, Cheadle, UK) was used to take a 50 µL gas sample from the headspace of sealed cultures. Samples were injected into to a 7820A gas chromatograph (Agilent) and measured as described by Crombie *et al.* (2015).

Measurement of isoprene uptake during the oil palm DNA-SIP experiment described in Chapter 5 was carried out with the use of a Fast Isoprene Sensor (Hills-Scientific, Boulder, CO, USA). 50 µL of headspace gas was injected into the Fast Isoprene Sensor. Data were exported to QTiPlot and the peak area produced by the photon count was used to determine isoprene concentration with the use of a standard curve of known isoprene concentrations ranging between 70 and 500 ppmv.

It should be noted that in terms of enrichment cultures, measuring the depletion of isoprene alone does not account for the possibility that certain bacteria may oxidise isoprene without utilising it for growth, or others may grow because they have the ability to utilise by-products of isoprene metabolism without the ability to metabolise isoprene itself.

2.3 Extraction of nucleic acids

2.3.1 Genomic and plasmid DNA extraction

To extract gDNA, pure cultures were grown in 50 mL liquid medium with glucose to an OD_{540} of 0.8-1. Cultures were centrifuged for 2 minutes at 11,000 x *g* and the supernatant discarded. gDNA was extracted using the Wizard Genomic DNA Purification Kit (Promega, WI, USA) according to the manufacturer's instructions.

Plasmid DNA was extracted from 5 mL overnight *Escherichia coli* cultures with the GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific) according to manufacturer's instructions.

2.3.2 High-MW DNA extraction

In circumstances when high concentration, high molecular weight gDNA was required, the following protocol based on Current Protocols in Molecular Biology Unit 2.4 (Ausubel *et al.,* 2003) was followed. Volumes given are suitable for 20 – 50 mL cell cultures.

Protocol components:

Resuspension buffer: 20 mM Tris, 2 mM EDTA, pH 8.0

RNase A: 10 mg mL^{-1}

NaCl: 5 M stock solution

CTAB/NaCl: 10 % (w/v) CTAB in 0.7 M NaCl.

Phenol:Chl:IAA

Prepared fresh on day of use:

Lysozyme: 100 mg mL^{-1}

Proteinase K: 10 mg mL^{-1}

N-laurylsarcosine: 10 % w/v

Method:

Cell pellet was suspended in 5 mL resuspension buffer. 60 µL of lysozyme was added and the solution was incubated for 60 mins at 37˚C. 375 µL of Proteinase K and 7 µL of RNase A was added. Solution was incubated again for 15 minutes at 37˚C. N-laurylsarcosine was added at a volume of 780 µL and the solution incubated for 60 minutes at 60˚C to lyse cells. 1012 µL of the 5 M stock of NaCl was added, followed by 803 µL CTAB/NaCl (to give 1% CTAB). The solution was incubated for 15 minutes at 60˚C.

An equal volume of phenol:Chl:IAA was added and the solution shaken vigorously for 5 seconds, then incubated for 10 minutes at 60˚C. The solution was shaken once at 5 minutes and once more at the end of this incubation step to separate DNA from protein and other solids at the interphase. The solution was centrifuged for 5 minutes at 8,000 x *g* and the upper aqueous phase extracted to a fresh tube. The phenol:Chl:IAA step was repeated twice more without the initial incubation step and with mixing for 5 minutes each time on a tube roller. Extraction with Chl:IAA was carried out once, with an additional second extraction where necessary if a white interface was still visible.

DNA was precipitated with the addition of 2 volumes of cold 100% ethanol. The solution was incubated for 60 minutes or overnight at 20˚C followed by centrifugation for 20 minutes at 13,000 x *g* at 4˚C. DNA was washed with cold 70% ethanol, centrifuged once more and briefly air-dried. The DNA pellet was then resuspended in 400 µL buffer (5 mM Tris pH 8.5).

2.4 Nucleic acid manipulation techniques

2.4.1 Quantification of DNA

The concentration of DNA was measured with the use of a Qubit 3.0 flurometer (Thermo Fisher Scientific) with either dsDNA Broad Range (BR) or High Sensitivity (HS) reagents depending on the expected DNA concentration range. Measurements were carried out according to the manufacturer's instructions.

2.4.2 Polymerase chain reaction (PCR) of *isoA* and 16S rRNA gene

All PCRs were carried out with the use of a DNA Engine Tetrad thermocycler (Bio-Rad, CA, USA). Reactions were prepared in 50 µL volumes using DreamTaq DNA polymerase and associated buffer (Thermo Fisher Scientific). A typical reaction included 5 µL buffer, 4 µM of each relevant primer (see Table 2.1), dNTPs (2.0 mM of each), 50 ng gDNA template or 20 ng to 1 µg environmental DNA and the remaining volume made of nuclease-free H_20. Each PCR experiment included a negative control containing no DNA template, and a positive control of genomic DNA from a known strain.

PCRs using DNA extracted directly from bacterial colonies were also carried out. In such cases, a sterile toothpick was used to pick one half of a single colony and the tip of the toothpick placed in 100 µL of sterile water in a PCR tube. Tubes were heated to 100˚C on a heat rack for 10 minutes and 1 µL of the heat-lysed culture was used as template for the PCR reaction. Colony PCRs were utilised for large-scale screening purposes.

When amplifying *isoA* the following conditions were used:

An initial step of 94˚C for 2 minutes, followed by 31 cycles of 95˚C for 15 seconds, 54˚C for 30 seconds, 72˚C for 1 minute and an extension step of 72˚C for 7 minutes. In circumstances where environmental DNA was utilised, and amplification was weak (such as the amplification of *isoA* genes from freshwater and marine samples in Chapter 3), 40 cycles were carried out.

Following amplification, PCR products were purified for downstream analyses with the Machery-Nagel NucleoSpin and PCR Clean-up Kit (Thermo Fisher Scientific) as per the manufacturer's instructions.

Table 2.1. Details of primers used in this study. GC-clamp sequences are underscored.

Name	Target	Sequence (5'-3')	Product Size (bp)	Reference
27F	Near-complete 16S rRNA gene	AGAGTTTGATCMTGGCTCAG	1400	Lane D.J. (1991)
1492R	Near-complete 16S rRNA gene	TACGGYTACCTTGTTAGGACTT		Lane D.J. (1991)
519F	V4-V5 hypervariable region of 16SrRNA gene	CAGCMGCCGCGGTAATAC	391	Stores *et al.* (2016)
907R	V4-V5 hypervariable region of 16SrRNA gene	CCGTCAATTCMTTTGAGTTT		Lane D.J. (1991)
341F-GC	V3-V5 hypervariable region of 16S rRNA gene (contains GC clamp)	CGCCCGCCGCGCGCGGCGGGCGGGGCGGGGGCACGGGGGGCCTACGGAGGCAGCAG	566	Muyzer G. (1995)
907R	V3-V5 hypervariable region of 16S rRNA gene	CCGTCAATTCMTTTRAGTTT		Muyzer G. (1995)
M13F	Plasmid inserts (various)	GTAAAACGACGGCCAG	Variable	Invitrogen
M13R	Plasmid inserts (various)	CAGGAAACAGCTATGAC		Invitrogen
isoA14F	Region of *isoA* gene encoding first iron centre of IsoA	GVGACGAYTGGTAYGACA	497	Carrion *et. al.* (2018)
isoA511R	Region of *isoA* gene encoding first iron centre of IsoA	TCGTCRAAGAARTTCTTBAC		Carrion *et. al.* (2018)

2.4.3 Preparation of *Escherichia coli* TOP10 cells and pGEM-T Easy cloning

Preparation of chemically competent *Escherichia coli* cells

Chemically competent *Escherichia coli* cells were grown in SOB medium prepared as follows. 20 g of tryptone, 5 g yeast extract and 0.5 g NaCl were added to 900 mL distilled H_2O. Then, 2.5 mM KCl was added, and 5 M NaOH was used to adjust the pH to 7.0. Total volume was brought to 1 L with distilled H_2O and the solution was autoclaved. Immediately before use, sterile 2 M $MgCl_2$ was added to make a final concentration of 10 mM.

CCMB80 buffer was made as follows. 11.8 g $CaCl_2.2H_2O$, 4 g $MnCl_2.4H_2O$, 2 g $MgCl_2.6H_2O$, 100 mL glycerol, 10 mL CH_3COOK (1 M stock, pH 7.0) were added to 800 mL distilled H_2O. The solution was sterilised via filtration and stored at 4˚C.

Glassware was treated for use by filling each container halfway with distilled water and sterilising by autoclaving.

Seed stocks were prepared by growth of *Escherichia coli* TOP10 cells on solid SOB medium at room temperature for 36 hours. Single colonies were transferred to 2 mL liquid SOB medium and incubated once more overnight at room temperature. Glycerol was added at 15% (v/v) and 1 mL was snap frozen in liquid nitrogen and stored at -80˚C.

To prepare chemically competent *Eschericia coli* TOP10 cells, 250 mL SOB medium was inoculated with 1 mL of seed stock and grown at 30˚C with shaking until the culture reached an OD_{540} of 0.3-0.4. The culture was cooled on ice for 10 minutes and cells harvested by centrifugation. The cell pellet was resuspended in 80 mL ice cold CCMB80 buffer and incubated on ice for 20 minutes before centrifugation. The cell pellet was resuspended in 10 mL ice cold CCMB80 buffer and the OD_{540} adjusted to 5-7.5. Competent cells were snap frozen in liquid nitrogen and stored at -80˚C.

Ligation of DNA

Clone libraries were created with the use of the pGEM-T Easy Vector System (Promega). The target gene was amplified via PCR from environmental DNA and the product purified as described previously. Ligations were set up at an insert:vector molar ratio of both 1:1 and 1:3. A ligation mix at the 1:1 ratio consisted of 1 µL PCR product, 1 µL pGEM-T Easy Vector (50 ng), 1 µL T4 DNA ligase, 5 µL 2x Rapid Ligation Buffer and 2 µL nuclease-free H_2O. Positive controls were set up similarly with control insert DNA in the place of PCR product. A background control with no DNA was also set up. Reactions were mixed well and stored overnight at 4˚C.

Transformation of chemically-competent *Escherichia coli* TOP10 cells

SOB medium was removed from storage in advance and allowed to warm to room temperature. Chemically competent *Escherichia coli* TOP10 cells were thawed on ice and 50 µL were combined with 50 – 100 ng of plasmid DNA or ligation mix. The mixture was put on ice for 30 minutes to 1 hour, then heat shocked for 3 minutes in a 42˚C heat bath. Immediately after this the mixture was returned to the ice for 5 – 10 seconds. 500 mL of room temperature SOC medium was added and the culture incubated at 37˚C with shaking for 1 hour. The culture was then centrifuged for 3 minutes at 7,000 x*g* and the supernatant discarded. The pellet was resuspended with what small amount of medium remained in the bottom of the tube and streaked to solid LB medium containing 50 µL ampicillin and 50 µL xgal per plate. Plates were incubated for 12 – 24 hours at 37˚C.

Plasmid Extraction from *E. coli*

Positive clones were inoculated in flasks with 5 mL LB and 5 µL ampicillin and incubated overnight at 37˚C with shaking. Once grown, 3 mL of cells were pelleted via centrifugation and plasmid DNA extracted with the use of the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) according to manufacturer's instructions.

2.4.4 Agarose gel electrophoresis

Agarose was added to 1x Tris Borate EDTA (TBE) buffer at 1% (w/v) and microwaved until the agarose had completely dissolved. Once cooled to room temperature, 0.5 µg mL^{-1} ethidium bromide was added before the gel was poured into an appropriate tank and allowed to set. A GeneRuler 1 kb DNA ladder (Thermo Fisher Scientific) was used to determine the size of sample DNA. Gels were run at 90 V for 20 – 40 minutes. Gels were visualised via a ChemiDoc XRS System (Bio-Rad).

2.4.5 Denaturing gradient gel electrophoresis

16S rRNA genes were amplified with primers 341-GC F and 907R (Valášková and Baldrian 2009). The DCode Universal Mutation Detection System (Bio-Rad) was used to carry out denaturing gradient gel electrophoresis (DGGE) following the manufacturer's instructions. PCR products were run on an 8% (w/v) polyacrylamide gel made of acrylamide-bisacrylamide 37.5:1 in x1 TAE buffer (pH 8.0) with a 30 – 70% linear denaturant gradient. 100% denaturant represented 7 M urea and 40% (w/v) formamide. Electrophoresis was

carried out at 75 V for 16 hours. Staining was carried out with 3 µL SYBR Green I Nucleic Acid Gel Stain (Thermo Fisher Scientific) and gels were imaged using a GelDoc XR system (Bio-Rad).

2.5 DNA-stable isotope probing

2.5.1 Soil sample preparation

The exact location of each set of soil samples used is detailed in each of the relevant chapters. However, as a rule, the processing of bulk soil samples associated with individual plants or locations was done using the following methods. In circumstances when bulk soil could not be transported to the ELSA lab (such as oil palm soils in Chapter 5), soil microcosms were set up with soil washings. 5 g of soils were resuspended in 50 mL of sterile distilled H_2O in 250 mL conical flasks. Flasks were shaken at 150 rpm for 30 minutes at room temperature. The resulting soil suspensions were decanted to sterile 50 mL measuring cylinders allowed to settle for 1 hour. The upper aqueous layer was transferred to a sterile flask. This method was repeated once more with the addition of a sonication step. This was to ensure the recovery of any cells that may have been tightly adhered to soil particles, but to also ensure that fragile cells that may be damaged by sonication would also make it through processing. Sonication was carried out for 5 minutes in a water bath (Mettler Electronics, CA, USA). The soil washings obtained from both methods (sonicated and un-sonicated) were combined in custom-made sterile 2 L air-tight Duran bottles with a spout that was sealed with a butyl-rubber seal to allow for easy enrichment with gas substrates. Samples were provided with 25 ppmv isoprene and incubated in the dark at 30˚C with shaking. In circumstances where soil could be used directly, the washing steps were skipped, and soil was incubated directly in sterile distilled H_2O.

2.5.2 Phyllosphere sample preparation

When dealing with leaf samples, biomass levels are expected to be low, so a concentrated washing technique was used. Leaves were placed into sterile glass bottle containing 250 mL sterile distilled H_2O. Leaf samples were sonicated for 5 minutes in a water bath (Mettler Electronics, CA, USA) and shaken at room temperature for 1 hour to dislodge cells from the leaf surface. The washings were then filtered through a 0.22 µm cellulose nitrate membrane filter (Pall, NY, USA) in order to concentrate the microbial cell mass. Filters were washed into a sterile 2 L air-tight Duran bottles using 40 mL sterile Ewers medium. Leaf samples were incubated in the presence of 25 ppmv isoprene and incubated in the dark at 30˚C with shaking.

2.5.3 DNA-SIP experiments using ^{13}C isoprene

The synthesis of ^{13}C isoprene was carried out by Gregg Whited and colleagues at Dupont, DE, USA. This ^{13}C-labelled isoprene was provided to leaf and/or soil microcosms to monitor uptake of labelled isoprene and subsequent incorporation of the isotope into the DNA of isoprene degrading bacteria. Unlabelled ^{12}C isoprene was provided to control samples. Sampling was carried out at various timepoints throughout incubation of SIP samples as detailed in each relevant chapter. Typically when enough isoprene had been incorporated to ensure sufficient labelling of the DNA in samples enriched with ^{13}C isoprene. The aim was to incubate for the minimal length of time to ensure this while also avoiding cross-feeding, which is more likely to occur the longer samples are incubated for prior to collecting material for the extraction of DNA. Isoprene uptake was monitored by GC or FIS as described in section 2.2.3.

2.5.4 Fractionation of isotopically-labelled and unlabelled DNA solutions

Fractionation was carried out as described by Neufeld *et al.* (2007). This method uses a caesium chloride gradient followed by isopycnic ultracentrifugation to separate out DNA in solution according to its density. Although density of the CsCl solution itself is incredibly difficult to accurately measure, in this case the refractive index was used as a proxy. Once the DNA was added, the solution was subjected to density gradient ultracentrifugation at 177,000 *xg* in with the use of a Beckman Vti 65.2 rotor with vacuum at 20˚C for 65 – 80 hours. No brake was used to ensure the gradient was not disturbed at the end of the ultracentrifugation run.

Once the gradient was established, the solution was fractionated to separate DNA labelled with ^{13}C isoprene from that which was not. In samples that were incubated with ^{12}C isoprene, any difference in density is due to an unusually high GC-content in the DNA, though it is always expected that this will be a rare occurrence. The contents of each tube were separated into 12 – 14 fractions via gradient fractionation with the use of a peristaltic pump (Watson Marlow Pumps, Cornwall, UK). An AR200 digital refractometer (Reichert, NY, USA) was used to determine the refractive index of each fraction. This was plotted against DNA concentration, quantified via Qubit with the use of the dsDNA HS assay kit, to identify heavy and light fractions produced by the DNA-SIP.

2.6 Amplicon sequencing analysis

2.6.1 Preparation of samples for amplicon sequencing

Amplicon sequencing was used to investigate both 16S rRNA gene and *isoA* diversity within samples described in a number of chapters. Duplicate PCRs were carried out on each sample to amplify the gene of interest and the PCR products were pooled prior to purification to avoid PCR bias. In certain circumstances (such as in Chapter 5), it was not always possible to obtain a PCR product from the heavy fraction of samples incubated with ^{12}C isoprene. This is because not enough members of the community contain a particularly high GC-content in their genome and as such little to no DNA is fractionated into the relevant refractive index range described in section 1.5.4.

In the experiments described in Chapter 3 and Chapter 5, *isoA* and 16S rRNA gene amplicon libraries were prepared and sequenced by MrDNA (Shallowater, TX, USA). PCR products provided were used in a 5 cycle PCR using the HotStarTaq Plus Master Mix Kit (Qiagen, MD, USA). After amplification, PCR products were checked on a 2% agarose gel and multiple samples pooled together in equal proportions based on their MW and DNA concentration. Pooled samples were then purified with the use of calibrated Agencourt AMPure XP beads (Beckman Coulter, CA, USA). Purified products were sequenced on an Illumina MiSeq platform following the manufacturer's instructions.

In Chapter 4, 16S rRNA gene amplicon sequencing was carried out by the Centre for Genomic Research (University of Liverpool, UK). PCR products were sequenced with the use of the Illumina MiSeq platform with 2 x 250 bp paired-end reads.

2.6.2 Analysis of 16S rRNA gene amplicon sequencing

The 16S rRNA gene amplicon data produced in Chapter 5 were analysed using the MrDNA analysis pipeline. Briefly, sequences were joined and short sequences >150 bp of sequences with ambiguous base calls were removed. Sequences were denoised, operational taxonomic units (OTUs) generated, and chimeras removed. OTUs were defined by 97% similarity. Final OTUs were taxonomically classified with the use of BLASTn against a specially curated database of DNA sequences extracted from The RDPII (Maidak *et al*. 1996) and NCBI databases (Altschul *et al*. 1990).

All other amplicons, both 16S rRNA gene and *isoA*, were analysed personally in-house with the use of the Bioconductor package DADA2 (Callahan *et al.*, 2016; version 1.6). Forward and reverse reads were trimmed to remove any adapter sequences and quality-filtered if the expected error rate was greater than 2 as assessed by the DADA2 quality analysis

function. Sequences were denoised using estimated error rates and the resultant sequences were dereplicated. Chimeric sequences were removed, and individual amplicon sequence variants (ASVs) were determined with the DADA2 algorithm. ASVs were taxonomically identified with the use of the RDP rRNA database (Maidak *et al.*, 1996; trainset 18) in the case of 16S rRNA gene amplicons, or a personally curated database of *isoA* sequences recovered from isoprene degrading bacterial isolate genomes in the case of *isoA* amplicon analyses. Identification of *isoA* ASVs was carried out manually with BLASTx.

2.7 Metagenome analysis

2.7.1 Sequencing of metagenomic data

Metagenomic sequencing described in Chapter 5 was carried out at MrDNA. 50 ng of each sample was used to create metagenomic libraries using the Nextera DNA Sample Preparation Kit (Illumina, CA, USA). Libraries were pooled in equimolar ratios of 1 nM and sequenced as paired end reads with the Illumina NovaSeq 6000 system.

The sequencing of metagenomic data in Chapter 6 was carried out by Centre for Genomic Research. Library preparation (insert size <500 bp) and sequencing were carried out on an Illumina HiSeq 2500 platform in high-output mode (v4) (2 x 125 bp paired-end reads). Samples were processed with the Nextera XT kit following the Nextera XT workflow with an additional purification step using Agencourt AMPure XP beads.

2.7.2 Metagenome community analysis, quality assessment, assembly and binning techniques

In both metagenomic studies, the community profiling of unassembled reads was carried out with Kraken (Wood and Salzberg, 2014; version 1.1.1). Results were fed to Bracken (Lu *et al.*, 2017; version 2.5) with a kmer of 31 to determine the relative abundance of each taxa.

Further analyses of metagenomic sequences in Chapter 5 was carried out as follows. Reads were quality filtered with the use of the iu-filter-quality-minoche script (Minoche, Dohm, and Himmelbauer 2011). Reads were then assembled using metaSPAdes (Nurk *et al.*, 2017; version 3.13) with kmers 21, 33 and 55. The quality of each of the assemblies was analysed with the use of MetaQUAST (Mikheenko *et al.*, 2016; version 4.6.3). Assembled contigs were utilised to construct metagenome-assembled genomes (MAGs) using MaxBin2 (Wu *et al.*, 2016; version 2.2.2). The completion and contamination of each MAG was assessed with CheckM (Parks *et al.*, 2015; version 1.0.18). MAGs with >75% completeness and <5%

contamination were reassembled and taxonomically identified using the MetaWRAP::reassemble_bins and MetaWRAP::classify_bins modules of the MetaWRAP pipeline (Uritskiy *et al.*, 2018; version 1.2.2). MAGs were concatenated and used to create a local database to search manually with BLAST for the presence of any homologues to known isoprene degrading proteins corresponding to IsoABCDEFGHIJ, along with associated proteins ALdH1, GshB and GarB (reviewed in Murrell *et al.*, 2020) with a cut-off value of $E < 1e\text{-}10$ in permissive searches and $E < 1e\text{-}40$ in restrictive searches. MAGs were functionally annotated with the use of PROKKA (Seemann, 2014; version 1.13.3).

The analysis of metagenomic sequence data obtained in Chapter 5 was carried out with a variety of modules from the MetaWRAP pipeline as follows. Raw pooled reads obtained from the heavy fraction of samples incubated with ^{13}C during the DNA-SIP experiment were pre-processed with the use of the metaWRAP::Read_qc module with default settings but without the bmtagger step. The MetaWRAP::Assembly module was used to assemble reads utilising MetaSPAdes. Binning was carried out with MetaWRAP::Binning using metaBAT2 (Kang *et al.*, 2019; version 2.12.1), MaxBin2 (version 2.2.6) and CONCOCT (Alneberg *et al.*, 2014; version 1.0.0). The results from each of the three assemblers were compiled to determine which MAGs were of the highest quality. These MAGs were reassembled in an effort to improve completion. Completeness, strain heterogeneity and contamination were assessed with CheckM (version 1.0.18) using the lineage-specific workflow. Taxonomy of each MAG was assigned with the use of MateWRAP::Classify_bins module. Functional annotation was carried out with PROKKA. In cases where a higher level of resolution was desired for taxonomic classification, MAGs were processed with the MiGA pipeline (Rodriguez-R *et al*. 2018).

In all cases, visual analysis and functional gene mining of annotated MAG sequences was carried out with the use of the Artemis genome browser and visualization tool (Carver *et al.*, 2012; version 16.0.0).

2.8 Genome analysis

2.8.1 Genome Sequencing

Sequencing of *Nocardioides* sp. WS12 was carried out by MicrobesNG (University of Birmingham, UK). *Nocardioides* sp. WS12 was grown in ANMS media with 10 mM glucose at 25˚C with shaking. Once the liquid culture had reached the mid-exponential growth stage, cells were collected and plated to R2A agar plates. During this period, the culture was checked for signs of contamination via microscopy. After 3 days of growth, biomass was collected and deposited into barcoded bead tubes provided by MicrobesNG who then carried out combined long-read and short-read whole genome sequencing as follows. For short-read Illumina sequencing, beads were washed with extraction buffer containing lysozyme and RNase A, then incubated for 25 minutes at 37˚C. Proteinase K and RNase A were added and the solution incubated for 5 minutes at 65˚C. gDNA was purified with an equal volume of Solid Phase Reversible Immobilisation (SPRI) beads and resuspended in EB buffer. DNA was quantified in triplicate with the Quant-It dsDNA HS assay in an Eppendorf AF2000 plate reader (Eppendorf, Hamburg, Germany). Libraries were prepared with a Nextera XT Library Prep Kit following manufacturer's instructions, with two modifications: 2 ng of DNA was used instead of 1 ng, and the PCR elongation time was increased by 30 seconds to 1 minute. Quantification and library preparation were carried out with the use of a Microlab STAR automated liquid handling system (Hamilton, NV, USA). Pooled libraries were quantified with the use of the Kapa Biosystems Library Quantification Kit for Illumina on a Roche light cycler 96 qPCR machine (Roche, Basel, Switzerland). Sequencing of libraries was carried out on the Illumina HiSeq using a 250 bp paired-end protocol.

Long-read sequencing was carried out on the same samples as follows. Liquid cultures were spun down and the pellet resuspended and stored in the cryopreservative of a Microbank tube (Pro-Lab Diagnostics UK, Wirral, UK). Approximately 2 x 10^9 *Nocardioides* sp. WS12 cells were used for HMW DNA extraction using a Nanobind CCB Big DNA Kit (Circulomics, Baltimore, MD, USA). Quantification of DNA was carried out with the Qubit dsDNA HS assay. Long-read genomic DNA libraries were prepared with an Oxford Nanopore SQK-RBK004 Kit with Native Barcoding EXP-NBD104/114 (ONT, Oxford, UK) using 400 – 500 ng of library and loaded in a FLO-MIN106 (R.9.4) flow cell in a GridION (ONT, Oxford, UK). Reads were adapter trimmed with Trimmomatic (Bolger Anthony M. *et al.*, 2014; version 0.30) with a sliding window quality cutoff of Q15.

Genome assembly with both datasets combined was carried out with Unicycler (Wick *et al.*, 2017; version 0.4.0).

2.8.2 Whole genome analysis

The quality of the combined short- and long-read *Nocardioides* sp. WS12 genome was analysed with the use of The MicroScope Microbial Genome Annotation and Analysis Platform (Vallenet *et al.*, 2020; version 3.13.5). The MicroScope platform was used to determine basic characteristics of the genome as described in Table 3.4. It was also used to annotate and query functional genes. Translated amino acid sequences that seemed likely to play a role in isoprene degradation were compared to a personally curated base of Iso proteins with tBLASTn. The MiGA pipeline was used to further investigate taxonomic affiliation, novelty and gene diversity with the use of the NCBI prokaryotic genome database.

2.9 Quantitative PCR (qPCR) methods

2.9.1 qPCR of 16S rRNA genes

Prior to use as template for quantitative PCR (qPCR), environmental DNA was utilised in a standard 16S rRNA gene PCR to ensure that amplification was possible. If a product was not obtained, DNA was cleaned with an extra run with the FastDNA SPIN kit for soils (MP Biomedicals, Solon, OH, USA) following manufacturer's instructions to remove inhibitors, which are particularly prevalent in soil samples and can prevent effective amplification.

qPCRs were carried out in 20 µL reactions with the use of a StepOnePlus machine (Applied Biosystems, Foster City, CA, USA). Sensifast SYBR Hi-ROX master mix (BioLine, London, UK) was used for all reactions. Primers 519F and 907R were used for 16S rRNA gene qPCR reactions (Mao *et al*. 2012)**Error! Reference source not found.**. Reactions contained 10 – 7 0 pg template DNA, 400 nM of each primer and 10 µL of SensiFAST SYBR Hi-ROX master mix. The qPCR reaction consisted of an initial denaturation step at 95°C for 3 min, followed by 40 cycles of 95 °C for 20 seconds, 55 °C for 20 seconds and 72 °C for 30 seconds. Data were collected at 72 °C for 15 seconds.

Efficiency was calculated with the use of a 10-fold standard curve. Standards were created by cloning the 16S rRNA gene of *Rhodococcus* AD45 into the pGEMT Easy Vector and using this as template DNA at various concentrations. The detection limit was 10^2 copies per 20 µL reaction. Specificity was measured via agarose gel electrophoresis of qPCR products and the use of melting curves in which the temperature was increased in 0.3˚C increments from 60 - 95˚C.

2.9.2 qPCR of *isoA* genes

qPCR of the *isoA* gene was carried out with the isoA14F and isoA511R primer set (Carrión *et al*. 2018). Reactions contained 1 – 18 ng of template DNA, 400 nM of each primer and 10 µL of SensiFAST SYBR Hi-ROX master mix. The qPCR reaction consisted of an initial denaturation step at 95°C for 3 minutes, followed by 40 cycles of 95°C for 20 seconds, 60°C for 20 seconds and 72°C for 30 seconds. Data were acquired at 88°C for 15 seconds to avoid quantification of primer dimers. Standards, efficiency and specificity checks were carried out as described in section 1.9.1 with the cloned *isoA* gene of *Rhodococcus* AD45 acting as a standard.

isoA copy numbers were normalised to 16S rRNA gene copy number to estimate a rough relative abundance of isoprene degrading bacteria in the sampled environment.

2.10 Mining of *isoA* genes in public metagenome datasets

Hidden Markov Model (HMM-based) surveys on both public metagenomes and those collated within the ELSA lab were carried out to search for and quantify *isoA* sequences in the environment following the method used by Curson *et al.* (2018). A collection of *isoA* sequences recovered from isoprene degrading isolates were used as training data to create an *isoA* HMM profile with the use of HMMR tools (http://hmmer.janelia.org/; version 3.1). HMM profiles for the *recA* gene were downloaded from the Functional Gene Pipeline and Repository (FunGene; Fish *et al.*, 2013). HMM searches were carried out against peptide sequences predicted from metagenome assemblies of interest with a cut-off value of $1e^{-30}$. Each IsoA sequence retrieved was manually analysed via BLASTp to ensure there were no false positives. The number of unique hits to IsoA was normalised to number of RecA sequences recovered, to the read number of the smallest metagenome and to the shortest protein length (in this case RecA).

3 Validation of novel *isoA* primers and genome analysis of *Nocardioides* sp. WS12.

3.1 Summary

This chapter details the validation of newly designed *isoA* gene primers and the characterisation of a newly isolated isoprene degrading bacterium *Nocardioides* WS12. The *isoA* gene primers were tested on DNA extracted from a number of environments via the creation of clone libraries and *isoA* gene amplicon sequencing, and were shown to amplify a wide variety of *isoA* genes with high specificity. The process of validation showed that isoprene degrading bacteria were present in terrestrial, industrial, phyllosphere, soil, freshwater and marine environments.

Nocardioides WS12 was isolated from one of the tested environments, soil associated with a willow tree. The genome of *Nocardioides* WS12 was sequenced and analysed to reveal a complete isoprene monooxygenase gene cluster along with genes associated with rubber degradation. *Nocardioides* WS12 grew on isoprene as the sole source of carbon and energy and was tested on a number of other compounds for both growth and oxidation.

This work is adapted from the original publications “Gene probing reveals the widespread distribution, diversity and abundance of isoprene-degrading bacteria in the environment” by Carrión *et al.* in Microbiome (2018), and also “Complete genome of isoprene degrading *Nocardioides* sp. WS12” by Gibson *et al.* in Microorganisms (2020).

isoA primer design and control strain validation were carried out by Dr. Ornella Carrión. Other aspects of primer validation, including enrichment of environmental samples, amplification of *isoA* genes and clone libraries were carried out jointly by Dr. Ornella Carrión and myself. *isoA* amplicon sequencing analysis and phylogenetic trees were carried out by me. Isolation of *Nocardioides* WS12 was carried out by Dr. Nasmille Larke-Mejía, all further physical and genomic characterisation of *Nocardioides* WS12 were carried out myself.

3.2 Background

Earlier studies into isoprene degrading bacteria in the environment found that Actinobacteria were the dominant phylum in both freshwater and terrestrial environments (El Khawand *et al.*, 2016; Johnston *et al.*, 2017; Crombie *et al.*, 2018). As a result, the primers developed during this period were derived from *isoA* sequences from a relatively small phylogenetic group of isoprene degraders. It was not until later studies, and the utilisation of lower isoprene concentrations in incubations and DNA-SIP experiments, that a wider diversity of isoprene degrading bacteria were recovered (Larke-Mejia *et al*. 2019; Carrión *et al*. 2020; Robin A. Dawson *et al*. 2020; Gibson, Larke-Mejía, and Colin Murrell 2020). The availability of this greater diversity of *isoA* sequences allowed for new primers to be designed with the potential to amplify a broader range of *isoA* sequences from the environment.

Studies such as these are only made possible by the physical isolation and sequencing of bona fide isoprene degrading bacteria. One example of the isolates utilised during primer design was the Actinobacterium *Nocardioides* WS12, isolated from the soil associated with a willow tree on the UEA campus (Larke-Mejia *et al*. 2019). *Nocardioides* are associated with a variety of different plant species, often in the phyllosphere environment (Yadav 2017; Franco *et al*. 2007; Bao *et al*. 2020; Larke-Mejia *et al*. 2019). However until the isolation of *Nocardioides* WS12, none had been shown to degrade isoprene.

3.3 Materials & Methods

All methods utilised for primer validation in this chapter were carried out as described in Chapter 2. This includes enrichment, clone libraries, *isoA* and 16S rRNA gene PCR and *isoA* gene amplicon analyses.

Nocardioides sp. WS12 was isolated from soil samples collected 10-20 cm from the trunk of a *Salix alba* tree and 5–10 cm below the soil surface as previously described (Larke-Mejia *et al*. 2019). Subsequent cultures were maintained in Ammonia Nitrate Mineral salts (ANMS) media (adapted from Brenner (1984)) with the addition of 5 gL^{-1} ammonium nitrate. 125 ppmv of isoprene was added and cultures were incubated at 25˚C with shaking. Growth, purity checks and substrate specificity analysis were carried out as described in Chapter 2. Once purity was confirmed, *Nocardioides* sp. WS12 cells were grown in ANMS medium with 10 mM glucose and plated to R2A (Oxoid) agar plates and grown for three days. At this point, biomass was collected and deposited into barcoded bead tubes supplied by

MicrobesNG (University of Birmingham, Birmingham, UK). Combined long-read and short-read genome sequencing was carried out by MicrobesNG as follows. Three beads were washed with extraction buffer containing lysozyme and RNase A and incubated for 25 min at 37˚C. Proteinase K and RNaseA were added and incubated for 5 min at 65˚C. Genomic DNA was purified using an equal volume of Solid Phase Reversible Immobilisation (SPRI) beads and resuspended in EB buffer. DNA was quantified in triplicate with the Quantit dsDNA High Sensitivity assay in an Ependorff AF2200 plate reader. Genomic DNA libraries were prepared using Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol with the following modifications: two nanograms of DNA instead of one were used as input, and PCR elongation time was increased to 1 min from 30 seconds. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system. Pooled libraries were quantified using the Kapa Biosystems Library Quantification Kit for Illumina on a Roche light cycler 96 qPCR machine. Libraries were sequenced on the Illumina HiSeq using a 250bp paired-end protocol. Reads were adapter trimmed using Trimmomatic 0.30 (Bolger Anthony M., Marc, and Bjoern 2014) with a sliding window quality cutoff of Q15. *De novo* assembly was performed on samples using SPAdes version 3.7 (Bankevich *et al*. 2012). All subsequent genome analyses were carried out as described in Chapter 2.

3.4 Results & Discussion

3.4.1 Validation of isoA primers

To assist in the identification of novel and diverse isoprene degrading bacteria in the environment, new gene probes that target *isoA* were designed. As mentioned, *isoA* encodes for the α-subunit of the isoprene monooxygenase and is highly conserved amongst all know isoprene degrading bacteria. This makes *isoA* an ideal marker gene for targeted functional and phylogenetic analyses of isoprene biodegradation in the environment. The *isoA* sequences recovered from 38 validated isoprene degrading bacteria, both available via Genbank or from bacteria isolated within the lab, were used to create an alignment (detailed in Table 3.1). These bacteria were members of Actinobacteria (such as *Gordonia, Mycobacterium* and *Rhodococcus*), Alphaproteobacteria (*Sphingopyxis*) and Betaproteobacteria (*Variovorax*). Sequences recovered from metagenomic datasets originating from willow soil, willow leaf and poplar leaf enrichments were also used where they shared >85% amino acid identity and 98% coverage with an *isoA* from a confirmed isoprene degrading bacteria. Similar genes encoding the α-subunit of other SDIMOs were

also included to ensure specificity of amplification. Examples include the sMMO, alkene monooxygenase and toluene monooxygenase from bacteria known to have no isoprene degrading capabilities. This alignment was used to identify conserved positions within *isoA* and primers spanning a number of these positions were designed (shown in Figure 3.2).

Of the 11 different variations tested, one pair, isoA14F and isoA511R, were chosen for their specificity and sensitivity when tested on both positive and negative control strains. Primer design was carried out by Dr. Ornella Carrión and further details of the process can be found in Carrión *et al.*(2018).

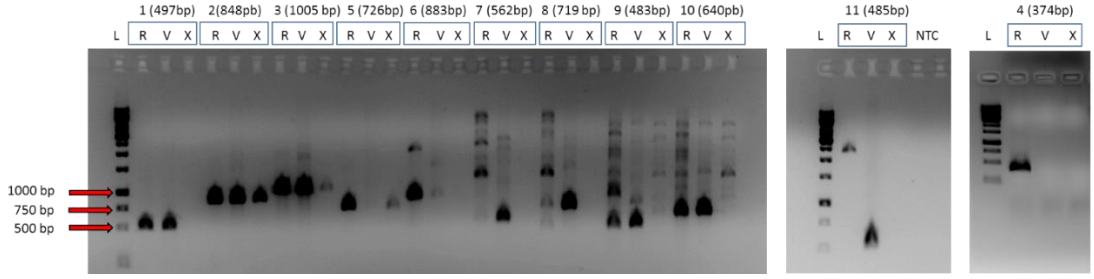

Figure 3.1. Preliminary validation of *isoA* primer combinations. Eleven *isoA* PCR primer combinations were tested with genomic DNA from the isoprene-degrading bacteria *Rhodococcus* sp. AD45 (R) and *Variovorax* sp. WS9 (V). Genomic DNA from *Xanthobacter autotrophicus* Py2 (X) was included as a negative control. Combination of primers and expected PCR amplicon size are indicated above gel images. Combination 1 represents primer pair isoA14F and isoA511R which were selected for further validation. L: GeneRulerTM 1kb Plus Ladder (ThermoFisher); NTC: non-template control.

Table 3.1. Sequences of hydroxylase α-subunits of soluble diiron monooxygenases utilised in the design of isoA primers. Taken from supplementary information provided in Carrión *et al.* (2018).

Microorganism	Enzyme	Genbank Accession Number
Gordonia sp. i37	IsoMO	KU870746.1
Gordonia sp. OPL2	IsoMO	MK176353
Leifsonia sp. i49	IsoMO	KU870737.1
Loktanella sp. i8b1	IsoMO	KU870736.1
Micrococcus sp. i61b	IsoMO	KU870739.1
Mycobacterium sp. AT1	IsoMO	KU870745.1
Mycobacterium sp. i61a	IsoMO	KU870739.1

Nocardioides sp. WS12	IsoMO	MK176348
Rhodococcus sp. ACPA1	IsoMO	NSDX01000002.1
Rhodococcus sp. ACPA4	IsoMO	NZ_NSDY01000003.1
Rhodococcus sp. ACS1	IsoMO	NZ_NSDZ01000001.1
Rhodococcus sp. ACS2	IsoMO	MK176338
Rhodococcus sp. AD45	IsoMO	AJ249207.1
Rhodococcus sp. i8a2	IsoMO	KU870743.1
Rhodococcus sp. i29a2	IsoMO	KU870744.1
Rhodococcus sp. LB1	IsoMO	LTCZ01000014.1
Rhodococcus sp. SC4	IsoMO	LSBM01000309.1
Rhodococcus sp. TD1	IsoMO	MK176350
Rhodococcus sp. TD2	IsoMO	MK176351
Rhodococcus sp. TD3	IsoMO	MK176352
Rhodococcus sp. WL1	IsoMO	MK176349
Rhodococcus sp. WS1	IsoMO	MK176339
Rhodococcus sp. WS2	IsoMO	MK176340
Rhodococcus sp. WS3	IsoMO	MK176341
Rhodococcus sp. WS4	IsoMO	MK176342
Rhodococcus sp. WS5	IsoMO	MK176343
Rhodococcus sp. WS6	IsoMO	MK176344
Rhodococcus sp. WS7	IsoMO	MK176345
Rhodococcus sp. WS10	IsoMO	MK176347
Rhodococcus sp. SK2ab	IsoMO	MK176356
Rhodococcus sp. SK5	IsoMO	MK176355
Rhodococcus erythropolis i47	IsoMO	KU870742.1
Rhodococcus opacus PD630	IsoMO	NZ_JH377098.1
Shinella sp. i39	IsoMO	KU870741.1
Sphingopyxis sp. OPL5	IsoMO	MK176354
Stappia sp. iL42	IsoMO	KU870740.1

Variovorax sp. WS9	IsoMO	MK176346
Variovorax sp. WS11	IsoMO	NZ_PXZZ01000003.1
Burkholderia cepacia G4	Toluene MO	AF349675
Gordonia sp. TY5	Propane MO	AB112920
Methylococcus capsulatus Bath	Soluble methane MO	M90050
Methylosinus trichosporium OB3b	Soluble methane MO	X55394
Mycobacterium sp. M156	Propene MO	AY455999
Mycobacterium chubuense NBB4	Ethene MO	GU174752
Mycobacterium chubuense NBB4	Propene MO	GU174753
Mycobacterium chubuense NBB4	Group 3 SDIMO	GU174751
Mycobacterium chubuense NBB4	Group 6 SDIMO	GU174750
Mycobacterium rhodosieae JS60	Ethane MO	AY243034
Nocardioides sp. JS614	Ethane MO	AY772007
Pseudomonas mendocina KR1	Toluene MO	M65106
Pseudonocardia sp. K1	Tetrohydrofuran MO	AJ296087
Rhodococcus rhodochrous B-276	Alkene MO	D37875
Thauera butanovora	Butane MO	AY093933
Xanthobacter sp. PY2	Alkene MO	AJ012090

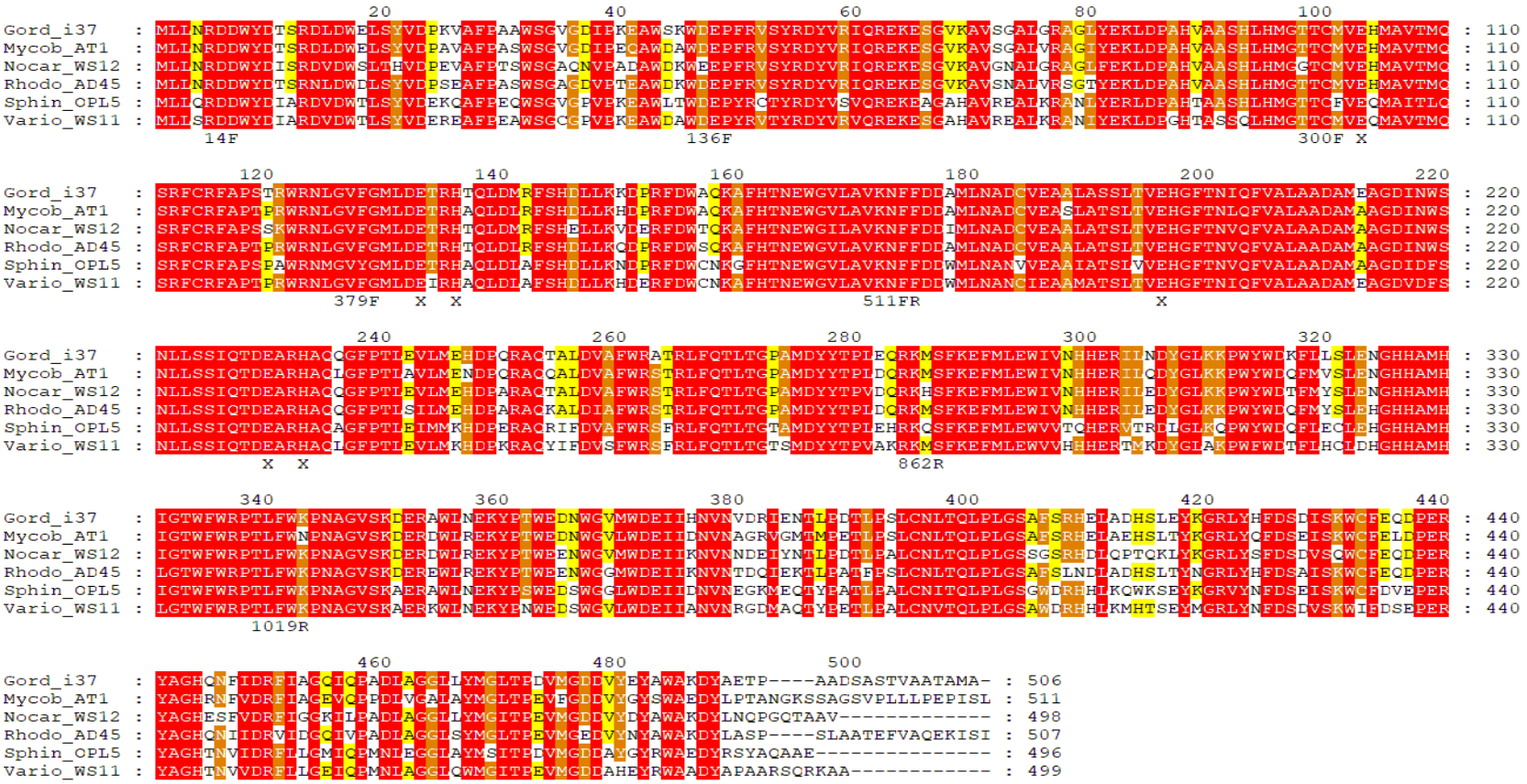

Figure 3.2. Alignment of the IsoA sequences from representative isoprene-degrading bacteria and position of the new *isoA* primers. Alignment of IsoA sequences was done using the ClustalW package included in BioEdit Sequence Alignment Editor v7.2.6 (Hall 1999). Conserved domains were analysed using GeneDoc v2.5.010 (Nicholas, K. B. 1997). Residues with identical or similar properties are highlighted in red, orange or yellow if they are conserved in all six, at least five or at least four polypeptides. The positions of iron binding ligands are marked with an "X" below. Start positions of the new isoA primers are indicated as follows: 14F: isoA14F; 136F: isoA136F; 300F: isoA300F; 379F: isoA379F; 511FR: isoA511F and isoA511R; 862R: isoA862R; 1019R3: isoA1019R. Strain names are: Gordonia sp. i37, Mycobacterium sp. AT1; Nocardioides sp. WS12; Rhodococcus sp. AD45; Sphingopyxis OPL5, and Variovorax sp. WS11. Accession numbers of these sequences are listed in Table 3.1. Taken from supplementary material supplied in Carrión *et al.,* (2018).

3.4.2 Diversity of isoA genes in environmental samples

To confirm the specificity of the new *isoA* primer set and to broaden our knowledge of the diversity and distribution of isoprene degrading bacteria, 11 different environmental sample types were enriched with isoprene (see section 2.2.2). These ranged from phyllosphere, soil, freshwater and aquatic environments (Table 3.2).

Table 3.2. Environmental samples utilised for primer validation studies. Analysis type 'C' means sample DNA was used to create clone libraries. Analysis type 'A' means DNA was used for the purpose of *isoA* gene amplicon sequencing

Material	Sampling site	Location	Analysis
Ash leaves	UEA Campus	Norwich, UK	C A
Poplar leaves	UEA Campus	Norwich, UK	C A
Willow leaves	UEA Campus	Norwich, UK	C A
Oil palm leaves B	Sepang	Selangor, Malaysia	C A
Oil palm leaves C	Kew Gardens	London, UK	C A
Oil palm soil D	Kew Gardens	London, UK	A
Willow soil	UEA Campus	Norwich, UK	C A
Tyre dump soil	Industrial park	Fakenham, UK	C A
Coastal sediment	Penarth beach	Penarth, UK	C A
Salt marsh sediment	Stiffkey salt marsh	Stiffkey, UK	C A
Freshwater sediment	UEA Lake	Norwich, UK	C A

Bulk eDNA extracted from these enrichments was utilised as template for PCR using the newly designed isoA14F and isoA511R primers. All enrichments produced a single product of the expected 497 bp size. *isoA* products from nine of the enriched environments were purified and used for the construction of *isoA* clone libraries, resulting in 69 sequenced clones in total. Each clone insert was sequenced and analysed with BLASTx (Altschul *et al*. 1990) to ensure only IsoA sequences were recovered and to compare to the closest IsoA homologue from bona fide isoprene degrading bacteria (Table 3.3).

Table 3.3. Comparison of *isoA* clone inserts to IsoA sequences from bona fide isoprene degrading bacteria.

Clone label	Closest IsoA sequence	Cover (%)	Amino Acid Identity (%)
Ash leaf clone 2	*Rhodococcus* opacus PD630	99	92.40
Ash leaf clone 7	*Rhodococcus* opacus PD630	99	100.00
Ash leaf clone 13	*Rhodococcus* opacus PD630	99	99.42
Ash leaf clone 15	*Rhodococcus* opacus PD630	99	100.00
Ash leaf clone 17	*Rhodococcus* opacus PD630	99	100.00
Ash leaf clone 32	*Rhodococcus* opacus PD630	99	100.00
Ash leaf clone 39	*Rhodococcus* opacus PD630	99	99.42
Ash leaf clone 40	*Rhodococcus* opacus PD630	99	100.00
Ash leaf clone 49	*Rhodococcus* opacus PD630	99	100.00
Oil palm leaf C clone 6	*Gordonia polyisoprenovorans* strain i37	99	94.15
Oil palm leaf C clone 11	*Gordonia polyisoprenovorans* strain i37	99	94.15
Oil palm leaf C clone 14	*Gordonia polyisoprenovorans* strain i37	99	94.15
Oil palm leaf C clone 15	*Gordonia polyisoprenovorans* strain i37	99	93.49
Oil palm leaf C clone 16	*Gordonia polyisoprenovorans* strain i37	99	93.57
Oil palm leaf C clone 17	*Gordonia polyisoprenovorans* strain i37	99	94.15
Oil palm leaf C clone 21	*Gordonia polyisoprenovorans* strain i37	99	94.15
Willow soil clone 1	*Variovorax* sp strain WS9	99	98.26
Willow soil clone 3	*Variovorax* sp strain WS9	99	98.83
Willow soil clone 5	*Variovorax* sp strain WS9	99	98.84
Willow soil clone 7	*Variovorax* sp strain WS9	99	99.41
Willow soil clone 9	*Variovorax* sp strain WS9	99	98.26
Willow soil clone 10	*Variovorax* sp strain WS9	99	99.42
Willow soil clone 11	*Variovorax* sp strain WS9	99	98.83
Willow soil clone 12	*Variovorax* sp strain WS9	99	98.84
Willow soil clone 21	*Variovorax* sp strain WS9	99	98.84
Willow soil clone 26	*Variovorax* sp strain WS9	99	98.26
Tyre dump soil clone 1	*Variovorax* sp strain WS9	99	98.83
Tyre dump soil clone 14	*Rhodococcus* opacus PD630	99	98.25
Tyre dump soil clone 22	*Ramlibacter* sp strain WS9	99	97.06
Tyre dump soil clone 28	*Ramlibacter* sp strain WS9	99	96.51
Freshwater sediment clone 1	*Rhodococcus* sp strain WS3	99	90.12
Freshwater sediment clone 3	*Rhodococcus* sp strain WS3	99	89.47
Freshwater sediment clone 4	*Rhodococcus* sp strain WS3	99	89.47
Freshwater sediment clone 5	Sphingopyxis sp strain OPL5	99	83.63
Freshwater sediment clone 6	*Rhodococcus* sp strain WS3	99	89.47
Freshwater sediment clone 8	*Rhodococcus* sp strain WS3	99	89.47
Freshwater sediment clone 9	*Rhodococcus* sp strain WS3	99	89.53
Freshwater sediment clone 11	*Rhodococcus* sp strain WS3	99	89.94
Freshwater sediment clone 14	*Rhodococcus* sp strain WS3	99	90.00
Freshwater sediment clone 16	*Rhodococcus* sp strain WS3	99	90.12
Freshwater sediment clone 18	*Rhodococcus* sp strain WS4	99	100.00

Oil palm leaf B clone 1	*Gordonia polyisoprenovorans* strain i37	99	98.83
Oil palm leaf B clone 2	*Gordonia polyisoprenovorans* strain i37	99	94.77
Oil palm leaf B clone 3	*Gordonia polyisoprenovorans* strain i37	99	94.77
Oil palm leaf B clone 4	*Gordonia polyisoprenovorans* strain i37	99	94.77
Oil palm leaf B clone 6	*Gordonia polyisoprenovorans* strain i37	99	95.93
Oil palm leaf B clone 7	*Gordonia polyisoprenovorans* strain i37	99	95.35
Salt marsh sediment A clone 1	*Rhodococcus* opacus PD630	99	100.00
Salt marsh sediment A clone 5	*Rhodococcus* opacus PD630	99	99.42
Salt marsh sediment A clone 7	*Rhodococcus* opacus PD630	99	100.00
Salt marsh sediment A clone 9	*Rhodococcus* opacus PD630	99	100.00
Salt marsh sediment A clone 10	*Rhodococcus* opacus PD630	99	99.41
Salt marsh sediment A clone 11	*Rhodococcus* opacus PD630	99	100.00
Salt marsh sediment A clone 18	*Rhodococcus* sp strain WS3	99	99.42
Salt marsh sediment A clone 22	*Rhodococcus* opacus PD630	99	100.00
Salt marsh sediment A clone 28	*Rhodococcus* opacus PD630	99	100.00
Coastal sediment clone 4	*Ramlibacter* sp strain WS9	99	100.00
Coastal sediment clone 5	*Rhodococcus* sp strain WS3	99	100.00
Coastal sediment clone 6	*Rhodococcus* sp strain WS3	99	98.84
Coastal sediment clone 9	*Rhodococcus* sp strain WS3	99	100.00
Coastal sediment clone 10	*Ramlibacter* sp strain WS9	99	98.83
Coastal sediment clone 13	*Rhodococcus* sp strain WS3	99	100.00
Coastal sediment clone 17	*Rhodococcus* sp strain WS3	99	100.00
Coastal sediment clone 18	*Rhodococcus* sp strain WS3	99	100.00
Coastal sediment clone 19	*Rhodococcus* sp strain WS3	99	98.84
Coastal sediment clone 27	*Rhodococcus* sp strain WS3	99	100.00
Coastal sediment clone 29	*Rhodococcus* sp strain WS3	99	100.00
Coastal sediment clone 30	*Rhodococcus* sp strain WS3	99	89.47
Coastal sediment clone 31	*Ramlibacter* sp strain WS9	99	98.83

All of the sequenced clone inserts were identified as containing the *isoA* gene with a shared translated amino acid identity of 83-100% with the IsoA of bona fide isoprene degrading bacteria. *isoA* inserts also shared less than 70% translated amino acid identity with the α-subunit of closely related SDIMOs such as the alkene monooxygenase from *Xanthobacter autotrophicus* PY2. These results confirmed the high specificity of the isoA14F and isoA511R primers.

Diversity of *isoA* sequences obtained from DNA extracted from within the same environments was low, with tyre dump soil samples the only environment category where *isoA* sequences were not similar at the genus level. A distinct difference between *isoA* sequences obtained from DNA extracted from differing environments was observed however. *isoA* sequences originating from Ash leaf samples all shared homology with *Rhodococcus opacus* PD630 (Furuya *et al*. 2012). *isoA* genes amplified from Salt marsh

sediment DNA were similar, although a single *isoA* sequence obtained from this environment differed in that it shared highest translated amino acid identity (99.42%) with the IsoA of *Rhodococcus* sp. strain WS3 (Larke-Mejia *et al*. 2019). Sequences from both freshwater and coastal sediment DNA were dominated by genes similar to that of the *isoA* from *Rhodococcus* sp. strain WS3, however both groups also contained *isoA* sequences with high translated amino acid identity to the IsoA of Gram-negative degraders such as *Sphingopyxis* sp. strain OPL5 and *Ramlibacter* sp. strain WS9 (Larke-Mejia *et al*. 2019). *isoA* sequences from oil palm leaf DNA all shared highest translated amino acid identity with the IsoA of *Gordonia polyisoprenovorans* sp. strain i37 (Johnston *et al.*, 2017). While *isoA* sequences from willow soil DNA all shared highest translated amino acid identity with the IsoA of *Ramlibacter* sp. strain WS9.

Having confirmed the specificity of the primer set, PCR amplicons were sent for *isoA* amplicon sequencing and analysed as described in section 2.6. Once quality-filtered, a total of 140,311 reads were retrieved with an average of 23,385 reads per environment sampled (specified in Table 3.2). Analyses with the DADA2 pipeline (Callahan *et al*. 2016) recovered 177 ASVs. ASVs were manually analysed with BLASTx and compared to the IsoA sequences of other ratified isoprene degrading bacteria. All ASVs except three shared >85% amino acid identity with an IsoA recovered from a bona fide isoprene degrading bacteria. The three ASVs that were exceptions to this shared no amino acid identity with IsoA sequences and thus were considered false positives and removed from the dataset. Alongside IsoA sequences from isolates, an IsoA sequence recovered from a metagenome assembled genome (MAG) identified as *Novosphingobium* and recovered from oil palm soil samples as described in detail in Chapter 5 was also included in the comparison database.

The most common homologue overall was that of *Rhodococcus*-like *isoA* sequences. These were found in all environments except oil palm leaves, and made up >98% of the ASVs recovered from DNA extracted from the two marine environments, coastal sediments and salt marsh sediments, one phyllosphere environment, ash leaves and one soil sample, oil palm soils Figure 3.3). *Rhodococcus*-like *isoA* sequences made up 49% RA of freshwater sediment samples, with the other 50% RA sharing homology with the *isoA* of the *Novosphingobium* MAG recovered from the oil palm soil SIP samples described in Chapter 5. The abundance of *Rhodococcus*-like *isoA* sequences is consistent with previous studies that show *Rhodococcus* to be highly abundant in temperate isoprene-rich environments (Crombie, Larke-Mejia, *et al*. 2018; El Khawand *et al*. 2016; Larke-Mejia *et al*. 2019).

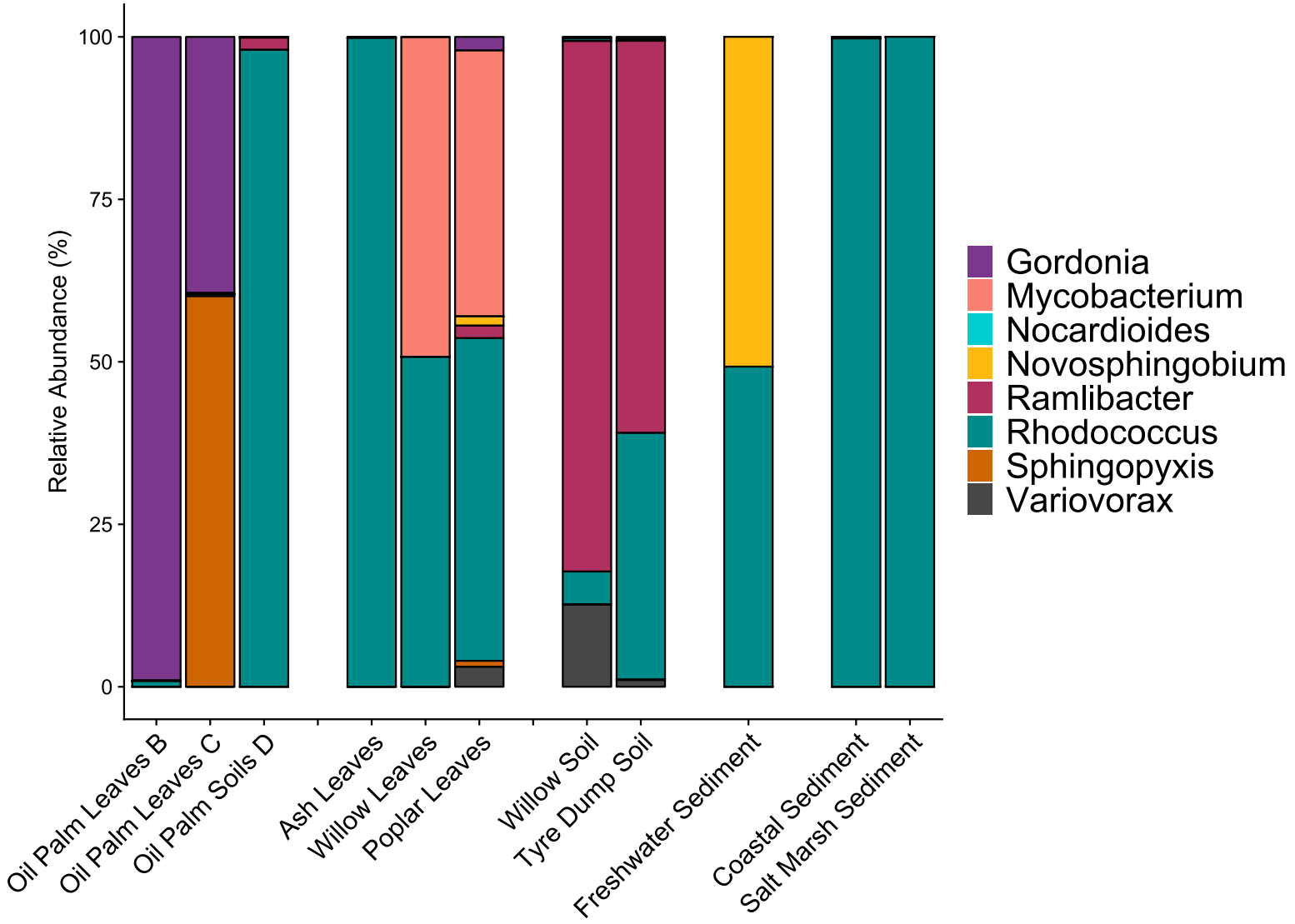

Figure 3.3. Figure 3.2. Profile of *isoA* genes retrieved from a range of isoprene-enriched environments. Analysed by *isoA* gene amplicon sequencing, figure shows relative abundance and diversity of *isoA* genes present in the bacterial community associated with each environmental sample.

Terrestrial soil samples such as willow soil and tyre dump soil samples yielded a high abundance of *Ramlibacter*-like *isoA* sequences making up 82% and 60% RA respectively. Willow soil samples also harboured *Variovorax*-like *isoA* sequences with an RA of 13% (Figure 3.3). It is of note that soil samples seem to harbour a greater abundance of Gram-negative isoprene degrading bacteria than phyllosphere environments.

Poplar leaf DNA samples showed the highest variability with ASVs similar to the *isoA* sequences of *Rhodococcus, Mycobacterium, Ramlibacter, Novosphingobium, Sphingopyxis, Variovorax* and *Gordonia,* with those similar to the *isoA* sequences of *Mycobacterium* and *Rhodococcus* being most dominant with an RA of 41% and 51% respectively. The abundance of *Mycobacterium*-like *isoA* sequences was shared with willow leaf samples, where such sequences were present at 49% RA.

isoA amplicon results from both oil palm leaf samples were distinct from the results of all other environments tested. ASVs similar to the *isoA* of *Gordonia* sp. i37 were present in high abundance in both, making up 99% RA of oil palm leaves B and 39% RA of oil palm leaves C. *Sphingopyxis*-like *isoA* sequences made up 60% RA of oil palm leaf C samples also. *Gordonia* and *Sphingopyxis* species were recently isolated from oil palm samples obtained from Kew Gardens in the UK (Larke-Mejia *et al*. 2019). Unlike all other environments tested, there is a notable absence of any *Rhodococcus*-like sequences in samples collected from the phyllosphere of oil palm trees.

The results from *isoA* amplicon sequencing and clone libraries are quite similar on a number of levels, with the exception of the abundance of *Sphingopyxis*-like *isoA* sequences amplified from oil palm leaf samples during *isoA* gene amplicon analysis. Dominant ASVs and clone library-derived *isoA* sequences can be seen alongside bona fide IsoA sequences in a phylogenetic context in Figure 3.4.

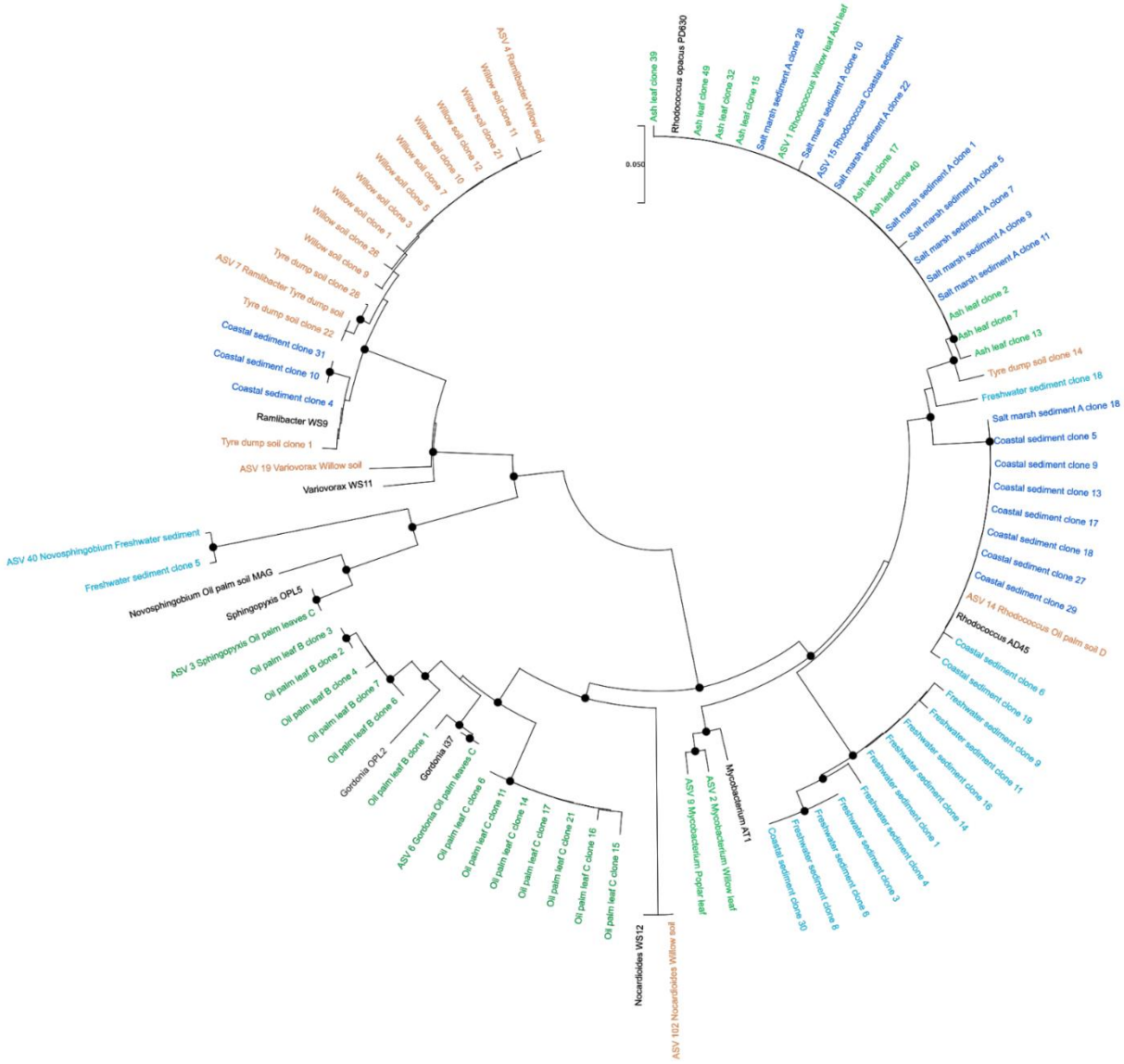

Figure 3.4. Phylogenetic tree of IsoA sequences recovered from isoprene-enriched environments. Created using the neighbour-joining method and analysed via the Jones-Taylor-Thornton model. Bootstrap values ≥50% are represented by dots at branch points with the use of 1,000 replicates. Terrestrial samples are coloured brown, freshwater samples are light blue, marine samples are dark blue and phyllosphere samples are green. Ratified isoprene degrading bacteria and IsoA sequences recovered from MAGs can be seen in black.

3.4.3 Characterisation of *Nocardioides* WS12 genome

Nocardioides sp. WS12 was isolated from the same willow soil environment that was sampled for clone library and *isoA* amplicon analyses (as described by Larke-Mejia *et al.*, 2019). *Nocardioides* sp.WS12 grew well on solid R2A medium (Oxoid) as seen in Figure 3.5. *Nocardioides* sp. WS12 was tested for growth on isoprene in a range of different media and growth temperatures and was shown to achieve peak cell density at OD_{540} 0.8 in ANMS medium adapted as described earlier at 25˚C with shaking. Under these conditions *Nocardioides* sp. WS12 exhibited a specific growth rate of 0.033 h^{-1} with a doubling time of 21.27 h when grown on 5% (v/v) isoprene in the headspace (Figure 3.5). This ability to grow in the presence of high amounts of isoprene is unusual, for comparison, growth of previous isolates such as *Rhodococcus* AD45 has been shown to be inhibited by isoprene concentrations above 2% (v/v) (Crombie *et al*. 2015). This suggests *Nocardioides* WS12 is relatively resistant to isoprene toxicity. In fact, further investigation showed no growth inhibition when cells were incubated with up to 10% (v/v) isoprene.

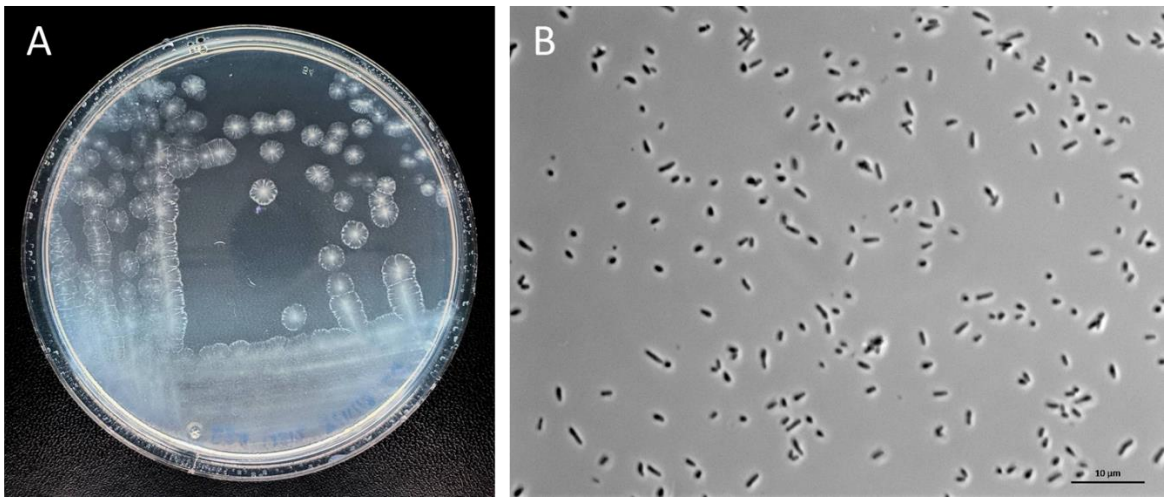

Figure 3.5. Growth of *Nocardioides* sp.WS12 on solid and liquid media. Image A) shows *Nocardioides* sp. WS12 grown on solid R2A medium. Image B) shows *Nocardioides* sp. WS12 grown on liquid ANMS medium with 5% (v/v) isoprene in the headspace, imaged at x1000 resolution.

When tested, *Nocardioides* sp. WS12 showed moderate growth (max OD_{540} >0.4) on glucose, succinate, fructose, sucrose, xylose, tri-sodium citrate and both *R*- and *L*-limonene. Lower growth was observed with maltose, glycerol and sodium pyruvate (max OD_{540} <0.4). All substrates were tested at a concentration of 10 mM.

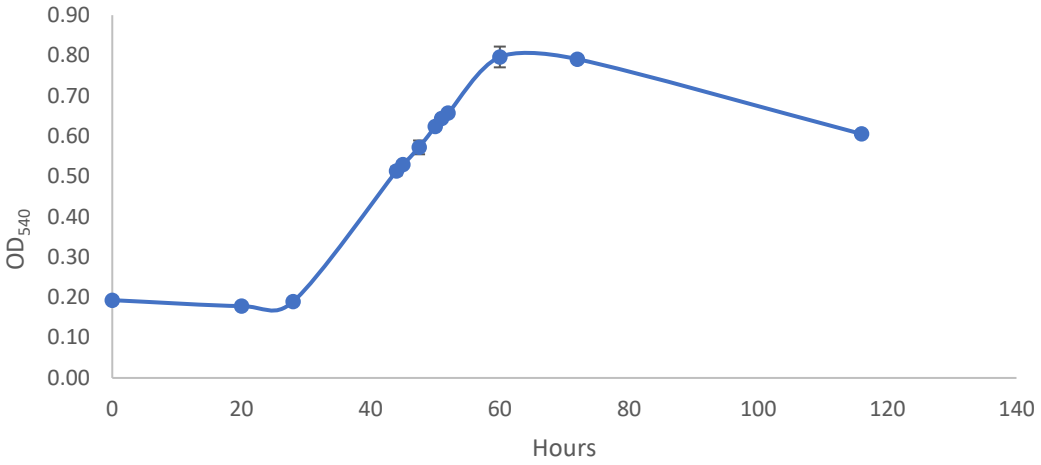

Figure 3.6. Growth of *Nocardioides* sp. WS12 on isoprene. Growth was carried out in ANMS medium supplemented with 5% (v/v) isoprene and incubated at 25˚C with shaking. Experiments were done in triplicate. Error bars are present but not visible at most data points as variance between replicates was minimal.

3.4.4 *Nocardioides* WS12 genome analyses

A combination of long and short read sequencing resulted in the retrieval of a complete, closed genome from *Nocardioides* sp.WS12. The single contig was 5.2 Mbp long, placing *Nocardioides* sp.WS12 at the larger range of genome sizes for its genus. *Nocardioides* genome sizes average at about 4 Mbp, with some, such as *Nocardioides nitrophenolicus* isolated from industrial wastewater, being significantly smaller at under 2 Mbp (Yoon *et al*. 1999). The genome contained no plasmids, and its GC-content was average for the genus at 69% (Table 3.4).

Two 16S rRNA genes were recovered from the genome of *Nocardioides* sp.WS12. Both were identified as *Nocardioides* at genus level with a nucleic acid identity of 100%. At the species level, the two 16S rRNA gene sequences diverged, with the first sharing 97.58% nucleic acid identity with *Nocardioides* sp. strain DK7869 (unpublished) and the second with *Nocardioides aromaticivorans* strain H-1 with 96.96% nucleic acid identity (Kubota *et al*. 2005). When the genome as a whole was analysed via Average Amino Identity (AAI) it shared 83.3% with *Pimelobacter simplex* (Suzuki and Komagata 1983) and 82.26% with *Nocardioides humi* (Kim *et al*. 2009). The threshold for shared identity at species level is 95% (Goris *et al*. 2007; Konstantinidis and Tiedje 2005) so these results indicate that *Nocardioides* sp. WS12 is a species not currently represented by any of the 115 *Nocardioides* genomes currently available in the NCBI prokaryotic database.

Table 3.4. Characteristics of the genome of *Nocardioides* sp.WS12

Length (bp)	5,171,066
Undetermined bases	0
GC (%)	68.66
Contigs	1
N50	5,171,066
Predicted Proteins	4,975
Ave. Length (aa)	323
Coding Density (%)	93.3
Completeness	99.23%
Contamination	0.52%
Pseudogenes	2
tRNA types	21
Total tRNAs	52

3.4.5 The complete isoprene degradation gene cluster of *Nocardioides* WS12

Analysis of the *Nocardioides* sp.WS12 genome revealed the presence of a full isoprene degradation gene cluster (*isoABCDEFGHIJ*). These genes were arranged very similarly to the isoprene degradation gene cluster of the isolate *Gordonia* sp. i37 (Johnston *et al.*, 2017;Figure 3.6). The genes encoding the isoprene monooxygenase *isoABCDEF* sit adjacent to an aldehyde dehydrogenase gene (*aldH2*), a glutathione synthase gene (*gshB*), and a CoA-disulfide reductase gene (*CoA-DSR*). The genes responsible for the subsequent steps of isoprene degradation *isoGHIJ* are upstream of the IsoMO. These genes encode the putative coenzyme A transferase, a dehydrogenase and two glutathione transferases. Similar to *Gordonia* sp. i37, *Nocardioides* sp.WS12 contains a second copy of the glutathione synthase *(gshA),* one copy of the putative transcriptional regulator *marR* and a duplicate copy of *isoG* (*isoG2*). Unlike, *Gordonia* sp. i37 however, *Nocardioides* sp.WS12 has only a single copy of *isoH*.

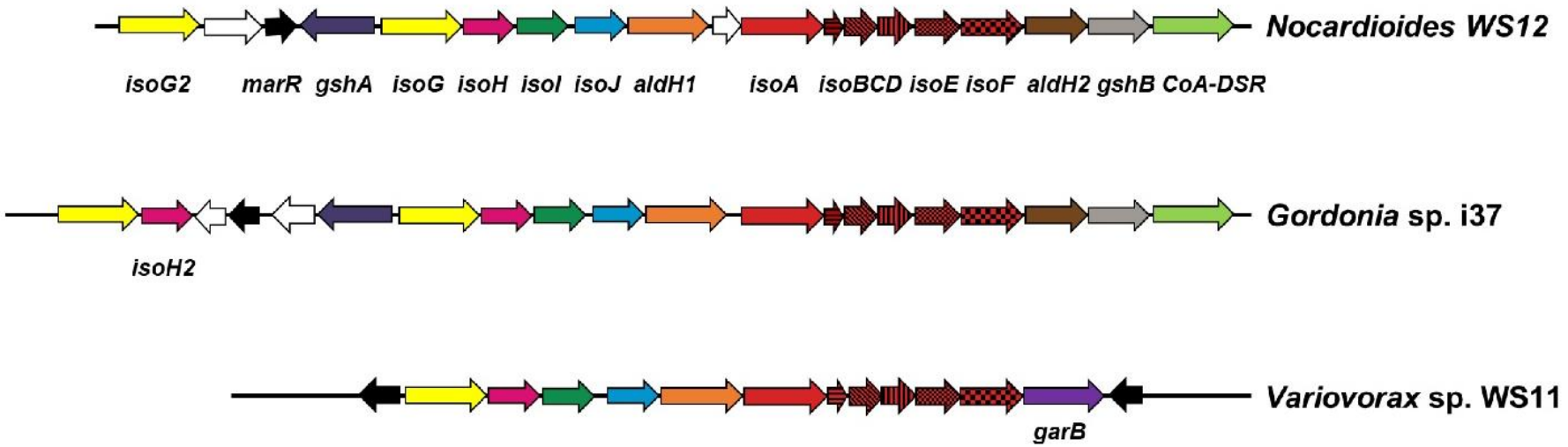

Figure 3.7. Isoprene degradation gene cluster of *Nocardioides* sp.WS12. Gram-positive Actinobacteria *Gordonia* sp.i37 (Johnston *et al.*, 2017) and Gram-negative Comamonadaceae *Variovorax* sp.WS11 (Larke-Mejia *et al*. 2019; Robin A. Dawson *et al*. 2020) are shown for comparison. Genes encoding the isoprene monooxygenase are coloured in red.

Figure 3.7 shows comparison to the Gram-negative isoprene degrader *Variovorax* sp.WS11 which contains no duplicate *iso* genes but does contain *garB,* encoding the glutathione reductase, which is not present in *Nocardioides* sp.WS12 (Dawson *et al*. 2020). The *isoABCDEFGHIJ* polypeptides found in *Nocardioides* sp.WS12 all shared >50% amino acid identity with the corresponding enzymes in other isoprene degrading Actinobacteria such as *Gordonia, Rhodococcus* and *Mycobacterium* (Table 3.5).

Table 3.5. Blastp comparison of polypeptides encoded by isoprene degradation genes in *Nocardioides* sp. WS12 to those from isoprene degrading Actinobacteria.

	Closest bona fide isoprene degrader	References	Coverage (%)	Amino acid identity (%)
IsoA	*Gordonia polyisoprenovorans* strain i37 IsoA	Alvarez *et al.*, 2009; Johnston, *et al.*, 2017	99	85.19
IsoB	*Rhodococcus opacus* strain PD630 IsoB	Alvarez *et al.*, 1996; Crombie *et al.*, 2015	100	57.95
IsoC	*Mycobacterium* sp strain AT1 IsoC	Johnston *et al.*, 2017	100	65.77
IsoD	*Rhodococcus* sp AD45 IsoD	van Hylckama Vlieg *et al.*, 1998	100	67.3
IsoE	*Rhodococcus opacus* strain PD630 IsoE	Alvarez *et al.*, 1996; Crombie *et al.*, 2015	100	62.96
IsoF	*Gordonia polyisoprenovorans* strain i37 IsoF	Alvarez *et al.*, 2009; Johnston, *et al.*, 2017	99	52.52
IsoG	*Rhodococcus opacus* strain PD630 IsoG	Alvarez *et al.*, 1996; Crombie *et al.*, 2015	100	76.56
IsoH	*Rhodococcus* sp AD45 IsoH	van Hylckama Vlieg *et al.*, 1998	100	73.45
IsoI	*Rhodococcus* sp strain WS4 IsoI	Larke-Mejia *et al.*, 2019	100	67.23
IsoJ	*Rhodococcus* sp strain WS4 IsoJ	Larke-Mejia *et al.*, 2019	100	69
AldH1	*Gordonia* sp strain OPL2 AldH1	Larke-Mejía *et al.*, 2020	98	65.3
IsoG2	*Rhodococcus* sp strain WS4 IsoG	Larke-Mejia *et al.*, 2019	96	59.64

3.4.6 Investigation of a rubber monooxygenase

Rubber (poly-cis-1,4-isoprene) is a polymer made up of multiple isoprene units and is the main source of anthropogenic isoprene production. Biodegradation of rubber is a valuable pathway within the biotechnology industry and has been observed among many members of the Actinobacteria family. Examples of this capability have been seen in *Streptomyces, Actinoplanes, Gordonia, Mycobacterium* and *Micromonospora* (Ali Shah *et al*. 2013).

The bacterial degradation of rubber is catalysed by one of three different rubber oxygenase enzymes. The first is RoxA, mainly found in Gram-negative rubber degraders such as *Xanthomonas* sp. 35Y (Jendrossek and Reinhardt 2003). The second, Lcp, is a latex-clearing protein described as such because many of the bacteria that utilise this enzyme for rubber degradation have the ability to produce clearing zones when plated onto solid media containing natural latex (Ilcu *et al*. 2017; Birke and Jendrossek 2014). The last of the three is RoxB which is loosely related to RoxA and is also found in Gram-negative degraders such as *Xanthamonoas* sp. strain 35Y, *Haliangium ochraceum, Myxococcus fulvus,* and *Corallococcus coralloides* (Jendrossek and Birke 2019).

The gene encoding Lcp (*lcp*) was found in the *Nocardioides* sp.WS12 genome. *Lcp* was located adjacent to a gene encoding a putative transcriptional regulator of the Tet-R family, a motif commonly seen in rubber degrading bacteria. Lcp is most often found in Gram-positive Actinobacteria such as *Nocardioides* (Jendrossek and Birke 2019). The Lcp found in *Nocardioides* WS12 shared 55% amino acid identity with both the Lcp from *Streptomyces* K30 and the Lcp from *Gordonia polyisoprenivorans*, the second of which is a confirmed isoprene degrading species (Linos *et al.*, 1999; Rose *et al.*, 2005; Johnston *et al.*, 2017).

qPCR analyses carried out by Dr. Ornella Carrión (described in Carrión *et al.*, 2018) investigated the abundance of *isoA* genes found in soil samples from an industrial tire dump. Results of this analysis showed a relatively abundant isoprene degrading community present in tire dump soils, with 67.7±14.4 *isoA* genes per million copies of 16S rRNA genes. While lower than the results for some of the other soil-based environments tested, this was significantly higher than the number of *isoA* harboured in all tested phyllosphere and aquatic environments. As mentioned, the most common industrial use for isoprene is in the production of synthetic rubber (*cis*-polyisoprene) for the purpose of tire manufacturing. Rubber oxygenases like Lcp and RoxA function by cleaving synthetic rubber into

oligoisoprene molecules (Bode *et al*. 2000). It is possible that isoprene degrading bacteria in these environments utilise the products of rubber degradation for growth. In the case of bacteria like *Nocardioides* WS12, perhaps the co-occurrence of an enzyme that cleaves *cis*-polyisoprene with a complete isoprene degradation pathway indicates that both of these processes can be carried out by single strain.

3.5 Conclusion

The process of developing robust functional gene probes is an ever evolving one, with each PCR primer pair only as good as the most current set of sequences representing the target gene. In this chapter a new and highly specific set of PCR primers targeting *isoA* were tested. This revealed the wide distribution of environments that harbour isoprene degrading bacteria. These *isoA* PCR primers will allow the recovery of a wider array of *isoA* sequences from the environment and increase our knowledge of the diversity and distribution of the isoprene monooxygenase enzyme in the environment and of those bacteria that harbour it. In the future, the sequences recovered with this primer pair will contribute to the design of newer gene probes that can expand this knowledge even further. However, focusing on one biomarker alone does not allow for the possibility that isoprene degradation can be carried out via another, undiscovered pathway. For this reason, other methods such a metagenomic studies and targeted isolation techniques are still needed to fully explore the breadth of isoprene degrading bacteria in the environment.

The results of *isoA* amplicon sequencing suggests that environment type, be it marine, terrestrial, phyllosphere or freshwater, is a driver for the diversity of isoprene degrading bacteria found within each location. This factor seems most pronounced, rather than species type within the same kind of biome. Further investigation into the variables that may influence the diversity of localised isoprene degrading bacteria is necessary. A greater array of environments, encompassing areas exposed to both high and low levels of isoprene emission, and including many more replicates, would be necessary to derive any strong correlation between host species, location and isoprene degrading bacterial diversity. The following chapters 4 and 5 both explore in further detail the diversity of isoprene degrading bacteria associated with plant species known to emit high amounts of isoprene. However, the current study takes the first step in illuminating the vast array of environments these bacteria can thrive in.

Nocardioides WS12 was isolated from the soil associated with a willow tree which is known to emit high levels of isoprene. Analysis of its complete isoprene degradation gene cluster adds to our current database and allows us to better identify patterns between isoprene monooxygenases originating from phylogenetically disparate bacteria. An example of one such pattern is the apparent difference in the putative regulatory genes found in the genomes of Gram-negative and Gram-positive isoprene degrading bacteria. The genome analysis described shows *Nocardioides* WS12 employs a putative MarR-type transcriptional regulator, encoded by *marR*, similar to *Rhodococcus* AD45 and *Gordonia polyisoprenivorans* i37 (Figure 3.7). However, *marR* is not present in Gram-negative isoprene degrading bacteria such as *Variovorax* WS11 and *Ramlibacter* WS9. These isolates depend on one or more putative LysR-type transcriptional regulators, encoded by *dmlR* (Dawson *et al*. 2020; Larke-Mejia *et al*. 2019). Both types of transcriptional regulator share a similar mode of action, namely, a co-inducer or co-repressor binds to the regulator to alter the specificity of DNA-binding. The co-inducer or co-repressor is typically a constituent of the pathway in question (Maddocks and Oyston 2008). While this separation of transcriptional regulator groups along phylogenetic lines is intriguing, at present we have very few representative Gram-negative isoprene degrading isolates. Confirming the presence and role of this phenomenon will not be possible until a wider variety of isoprene degrading bacteria have been isolated.

4 Cultivation independent analysis of willow leaf isoprene degrading community.

4.1 Summary

This chapter details DNA-SIP experiments to investigate the isoprene-degrading community associated with the phyllosphere of a willow (*Salix fragilis)* tree. The isoprene-degrading community was investigated via 16S rRNA gene amplicon sequencing and metagenomic analysis. Results from amplicon sequencing showed that bacteria belonging to the genera *Comamonadaceae, Acidovorax, Polaromonas, Variovorax* and *Ramlibacter* were involved in the uptake of ^{13}C isoprene. Metagenomic analysis confirmed the presence of many of the same taxa, with the addition of *Mycobacterium* and *Methylobacterium.* A *Mycobacterium* MAG was recovered and was shown to contain two complete, non-identical isoprene degradation gene clusters. A *Methylobacterium* MAG not containing any *iso* genes was recovered from the pooled heavy fractions of the ^{13}C enriched samples, indicating its ability to utilise isoprene for growth, was also examined

This work is adapted from the original publication "Isoprene-degrading bacteria associated with the phyllosphere of *Salix fragilis*, a high isoprene-emitting willow of the Northern Hemisphere" by Gibson *et al.* in Environmental Microbiome (2021). The DNA-SIP experiment including sample collection, incubation with ^{12}C and ^{13}C isoprene, PCR and extraction of nucleic acids was carried out by Dr. Andrew Crombie, all subsequent analyses were carried out myself.

4.2 Background

The emission of isoprene is not a universal trait of terrestrial plants, and even amongst those who exhibit the behaviour, the scale of the emission differs greatly species to species (Table 1.1; Loreto *et al.*, 1998; Logan *et al.*, 2000; Monson *et al.*, 2013). Unlike similar climate active gases such as methane and due in most part to its reactivity, the atmospheric bioavailability of isoprene at any one time, in any one place is low. However, isoprene concentrations in the intercellular spaces of leaves, near the stomata where emission occurs, can reach 30ppmv which is orders of magnitude above atmospheric concentrations (Fall *et al.*, 1992; Singsaas *et al.*, 1997; Brüggemann and J.-P. Schnitzler, 2002; Sun *et al.*, 2013). The logic follows that to better investigate communities that may

harbour isoprene degrading bacteria in any abundance, one must focus on environments, such as the phyllosphere, in which isoprene concentrations are higher. Willow species are common in the Northern Hemisphere and are some of the highest emitters of isoprene we know of today, emitting up to 37 µg g(dry weight) h^{-1} (Pio *et al.*, 1993). The species investigated in this study is the *Salix fragilis*, a large willow often planted by river banks and lakes for the purpose of stabilising riverside soil (Anstead and Boar 2010) .

The objective of this study was to examine the phyllosphere of a plant native to the UK know to emit high levels of isoprene and investigate whether its isoprene degrading microbial community differed to previously studied environments.

4.3 Methods & Materials

4.3.1 Sample collection & DNA-SIP

Samples were collected from the south-facing side of a *Salix fragilis* tree located in the University of East Anglia campus. Leaves were gathered from branches approximately 2.5 m above ground level. Adherent cells were removed as described in section 2.5.2*,* suspended in 50 mL of a ½ dilution of minimal medium and incubated with either unlabelled or ^{13}C labelled isoprene. 150 ppmv of isoprene vapour was injected and samples were incubated with shaking at 150 rpm at 25˚C. Isoprene consumption was monitored by gas chromatography and cells harvested when each microcosm had consumed approx. 0.5 µmol isoprene mL^{-1} which took between 13 and 53 days. Resulting cell pellets were used for nucleic acid extraction, along with material collected from unenriched (T0) samples collected prior to the enrichment period. Enriched DNA were separated via density gradient ultracentrifugation, fractionated and the density measured via refractometry to identify ^{13}C-labelled “heavy” DNA and unlabelled “light” DNA for analysis (see section 2.5.4). Of the total DNA recovered from each sample that underwent ultracentrifugation, an average of 1.2% was recovered from the heavy fractions of sample incubated with ^{12}C isoprene, whereas 20% was recovered from the heavy fraction of samples incubated with ^{13}C isoprene.

4.3.2 Community analysis

16S rRNA gene amplicon sequencing data were analysed with DADA2 as described in section 2.6 with the following specifications. Forward and reverse reads were trimmed by 33 and 37 nucleotides respectively in order to remove synthetic sequences, no max length was set. An average of 561,187 cleaned and processed reads per sample were used to collate a total of 3,369 ASVs across all 16S rRNA gene amplicon samples and replicates. ASVs were taxonomically identified with the use of the RDP rRNA database ((Maidak *et al*. 1996); version RDP trainset 18).

Metagenomic community data and the subsequent binning process were carried out as per section 2.7. MAGs were investigated for the presence of plasmid DNA with the use of the plasmidVerify script (Antipov *et al*. 2019) developed by the Centre for Algorithmic Biotechnology, Saint Petersburg State University. MAGs of interest were further analysed with the MiGA pipeline for a deeper taxonomic classification than that provided by CheckM (Rodriguez-R *et al*. 2018).

4.4 Results and Discussion

4.4.1 Bacterial community associated with *Salix fragilis* leaves

Overall results of 16S rRNA gene amplicon analysis indicated a high variability in community structure between isoprene enriched replicates. However the labelled, heavy fraction of samples incubated with ^{13}C isoprene showed clearly distinct communities from control samples indicating that the DNA-SIP had been successful in enriching and separating isoprene degrading bacteria from the rest of the population.

One set of controls in the DNA-SIP experiment were the light fractions of samples incubated with ^{12}C isoprene. Results showed that the groups enriched in these samples varied considerably between replicates (Figure 4.1). For example, *Hydrocarboniphaga* was enriched in replicate 1 with an RA of 45.35%, but was not found in other replicates. Replicates 1 and 2 showed enrichment of *Mycobacterium* with an RA of 6.94% and 13.32% respectively. *Comamonadaceae* was abundant in replicate 3 with an RA of 29.57%, and to a lesser extent in replicates 1 and 2 with an RA of 3.39% and 1.75% respectively. *Caulobacter* was also present across all replicates with an average RA of 7.91 ± 5.76%.

During DNA-SIP experiments, a certain quantity of DNA is expected to be fractionated to each of the four categories, namely the heavy and light fractions of samples incubated with

either ^{12}C or ^{13}C isoprene. In the case of the heavy fractions of samples incubated with ^{12}C isoprene however, the only variable that can result in DNA being selectively fractionated into this range is if its GC content is particularly high. For this reason, it was anticipated that a very low proportion of DNA would be recovered from these fractions (1% in this study). Outside of any differences caused by GC content, it is expected that the heavy and light fractions of samples incubated with ^{12}C isoprene would share a very similar community structure, with most DNA in the heavy fraction there as a result of an imperfect gradient with small amounts of DNA bleeding between light and heavy fractions. As can be seen in Figure 4.1, the current DNA-SIP experiment conforms to these expectations, with both heavy and light fractions of samples incubated with ^{12}C isoprene being very similar.

The light fractions of samples incubated with ^{13}C isoprene contain the community of bacteria that did not utilise isoprene for growth and reproduction and as a result their DNA was not isotopically labelled. Similarly to those samples incubated with ^{12}C isoprene, *Caulobacter* was enriched in replicates 2 and 3, with an RA of 17.58% and 15.4% respectively. *Sphingomonas* was enriched in all three replicates with an RA of 7.99 ± 5.06%, and *Tabiella* had an RA of 12.69% and 4.46% in replicates 1 and 3 respectively (Figure 4.1).

Finally, the focus of this experiment was found in the heavy fractions of those samples incubated with ^{13}C isoprene, representing those bacteria that metabolised ^{13}C isoprene during incubation and had their DNA isotopically labelled as a result. The bacterial communities enriched in these samples were distinct from control samples, indicating that the selective pressure introduced by the DNA-SIP experiment was successful in enriching an isoprene degrading bacterial community. *Comamonadaceae* was highly enriched in replicates 1 and 2, with an RA of 57.48% and 58.33% respectively. Replicate 1 showed an enrichment of *Polaromonas* with an RA of 24.78%, though this genus was not found in replicate 1 or 2. The community in replicate 3 diverged in other ways also, for example while *Comamonadaceae* was not present, *Mycobacterium* was highly enriched with an RA of 45.39%. Also, there was an abundance of *Methylobacterium,* an RA of 26.48%, which was only previously seen with a low abundance in unenriched T0 samples (Figure 4.1).

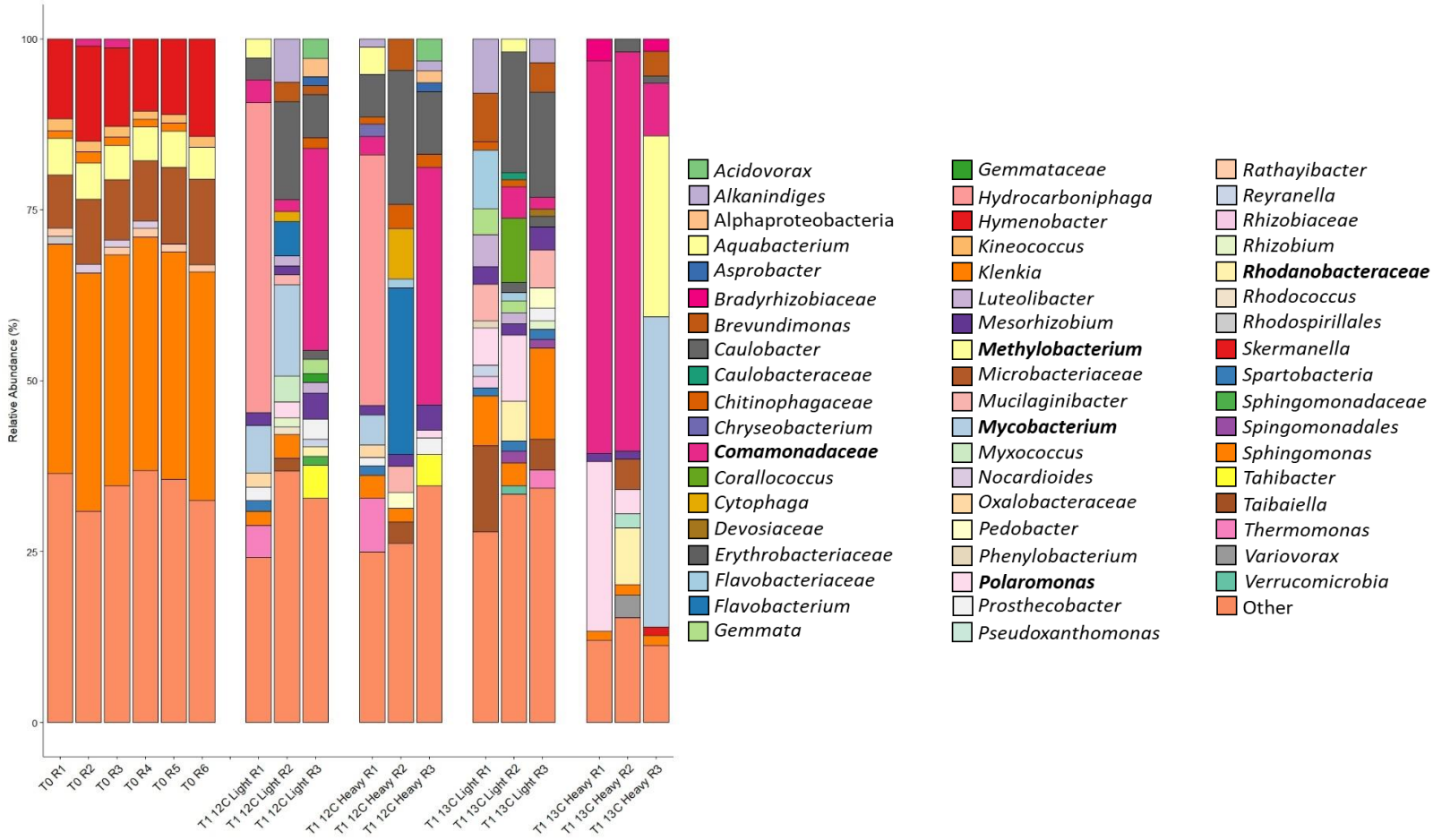

Figure 4.1. Bacterial community profile of DNA retrieved from willow leaf samples analysed by 16S rRNA gene amplicon sequencing. Samples are represented as unenriched (T0), enriched (T1), unlabelled (12C), labelled (13C), heavy DNA and light DNA fractions retrieved after DNA-SIP. R1 – 6 indicate the six replicate samples analysed. Taxa that are at less than 1% relative abundance in a sample are grouped as 'Other'. Taxa that were of >5% relative abundance in heavy fractions of ^{13}C-enriched samples are in bold.

Metagenomic analyses were utilised to further elucidate the diversity and genomic potential of those bacteria found in the heavy fraction of samples enriched with ^{13}C isoprene. Taxonomic classification was carried out with the use of Kraken (Wood and Salzberg 2014) as described in section 2.7.2 and revealed the presence of a number of genera belonging to the family *Comamonadaceae*, with *Acidovorax* (RA of 14%), *Variovorax* (RA of 10.83%), *Polaromonas* (RA of 3.8%)*, Hydrogenophaga* (RA of 3.2%)*, Ramlibacter* (RA of 2.7%) and *Rhodoferax* (RA of 2.5%) (Figure 4.2). The presence of a number of different genera of the family *Comamonadaceae* in this phyllosphere environment mirrors a previous study which focused on the bacterial community of bulk soil associated with a willow species where *Comamonadaceae* made up 21 – 30% of the relative abundance in ^{13}C incubated heavy samples (Larke-Mejia *et al*. 2019). Although the two environments exist under different abiotic pressures, this shared abundance might suggest members of the phyllosphere community are being transported to the associated soil environment, possibly though rainfall or falling leaves (Bittar *et al*. 2018).

Bradyrhizobium from the order *Rhizobiales* which was observed during 16S rRNA gene amplicon analysis (Figure 4.1) was also found during metagenomic analyses with an RA of 2.23% (Figure 4.2). *Mycolicibacterium* had been labelled and was present at an RA of 12.72%. This genus, which has recently been differentiated from the genus *Mycobacterium* (Gupta, Lo, and Son 2018), was also found enriched in the heavy fractions of samples incubated with ^{13}C isoprene after 16S rRNA gene amplicon analysis (Figure 4.1). *Mycobacterium* itself made up 4.52% of the metagenomic community. In Chapter 3 it was shown that *isoA* sequences sharing high sequence similarity with the *isoA* of *Mycobacterium* AT1 were estimated to make up half of all *isoA* sequences present in bacteria on the leaves of sampled willow leaves. In a previous study, an isoprene degrading *Mycobacterium* species was isolated from an estuarine environment also (Johnston *et al.*, 2017; Carrión *et al.*, 2018). However, this is the first instance of *Mycobacterium* being significantly enriched with isoprene in a terrestrial environment.

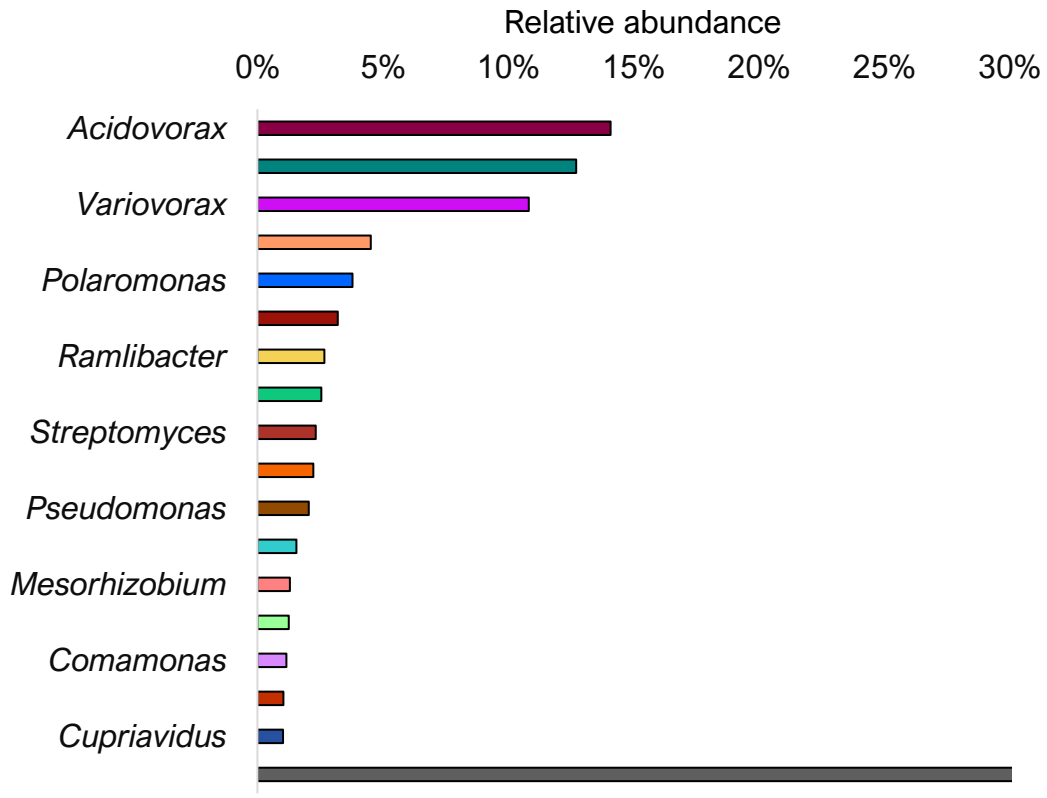

Figure 4.2. Relative abundance of bacterial taxa retrieved after analysis of the metagenome from pooled heavy fractions from ^{13}C-labelled willow leaf samples retrieved after DNA-SIP. Metagenome data were analysed and classified taxonomically using Kraken. All taxa that were at less than 1% relative abundance in the ^{13}C-labelled DNA have been grouped as 'Other'.

Methylobacterium, which was highly enriched in 16S rRNA gene amplicon analysis of one of the ^{13}C heavy replicates, also featured in the metagenome analysis with an RA of 1.5%. The appearance of a ^{13}C-labelled *Methylobacterium* here is interesting. There have been early reports of *Methylobacterium* that can potentially grow in the presence of isoprene (Murphy 2017; Srivastva *et al*. 2015), though it is not common for the genus to appear in studies examining isoprene-degrading communities in the environment.

Without an extant example of the strains that make up the *Methylobacterium* ASVs, it cannot be said with absolute certainty that they do have the metabolic capability to degrade isoprene, but their presence in the heavy fraction of ^{13}C isoprene enriched samples (while not abundant in the heavy ^{12}C controls) suggests they have indeed utilised the ^{13}C-labelled isoprene during growth. As the only representative methylotroph linked to

isoprene degradation, further evidence of *Methylobacterium* as a validated isoprene degrading bacterium would be of great interest and opens avenues for future work.

4.4.2 Analysis of an abundant *Mycobacteriaceae* MAG containing two isoprene monooxygenase gene clusters.

Metagenomic sequencing data was used to recover a number of MAGs as described in section 2.7.2. Details for each of the recovered MAGs and their identity can be seen in Table 4.1. Each of the 34 MAGs were mined for the presence of *iso* genes involved in isoprene metabolism. The most abundant of the MAGs, identified as a member of the *Mycobacteriaceae* family, was selected for further investigation due to the presence of a complete isoprene degradation gene cluster. The MAG was processed via the MiGA pipeline (Rodriguez-R *et al*. 2018) which further identified it to genus level as a member of *Mycobacterium*.

Two complete, non-identical, copies of the isoprene degradation gene cluster (*isoABCDEFGHIJ;* referred to going forward as *iso* cluster 1 and *iso* cluster 2) were found for the first time in the *Mycobacterium* MAG. Genes associated with isoprene degradation and found in the vicinity of the cluster in each isolate to date were also identified, these are *aldH1, CoA-DSR, gshB* and *marR,* encoding an aldehyde dehydrogenase, a CoA-disulfide reductase, a glutathione synthase and a putative transcriptional regulator respectively (reviewed in McGenity *et al.,* 2018 and Murrell *et al.,* 2020; described in section 1.3.1) and can be seen in Figure 4.3

Table 4.1. Statistics for the integrity and abundance of recovered MAGs

Abundance Ranking	Completeness (%)	Contamination (%)	N50	Size (Mbp)	ID
1st	99.62	1.31	426,339	7.4	*Mycobacteriaceae*
2nd	84.53	2.30	28,949	4	*Comamonadaceae*
3rd	78.65	1.48	19,712	3.4	*Comamonadaceae*
4th	71.3	1.52	21,622	5	*Comamonadaceae*
5th	92.4	1.12	70,582	4.4	*Comamonadaceae*
6th	92.75	4.48	56,291	6.1	*Comamonadaceae*
7th	89.21	2.45	47,756	4	*Comamonadaceae*
8th	88.71	2.81	33,346	4.3	*Comamonadaceae*
9th	98.86	0.17	91,899	4.8	*Methylobacteriaceae*
10th	89.89	5.65	114,304	8.7	*Myxococcales*
11th	92.9	5.46	59,280	4.8	*Comamonadaceae*
12th	94.35	1.02	64,831	3.9	*Burkholderiales*
13th	98.77	0.80	142,624	4.2	*Xanthomonadaceae*
14th	77.75	8.04	36,155	3.9	*Comamonadaceae*
15th	98.98	0.76	201,989	3.3	*Microbacteriaceae*
16th	98.77	1.31	170,797	3.8	*Sphingomonadaceae*
17th	90.67	3.08	28,589	5.7	*Burkholderiales*
18th	87.71	0.53	197,524	2.8	*Caulobacteraceae*
19th	97.15	7.72	93,325	4.1	*Caulobacteraceae*
20th	95.4	1.74	263,888	4.2	*Bradyrhizobiaceae*
21st	91.19	13.04	33,095	3.2	*Caulobacteraceae*
22nd	96.75	1.15	93,753	3.4	*Xanthomonadaceae*
23rd	96.77	2.96	58,108	8.5	*Proteobacteria*
24th	93.05	2.47	50,463	8.2	NA
25th	88.3	3.19	14,127	3.5	*Alphaproteobacteria*
26th	88.11	0.93	49,319	3.8	NA
27th	91.77	0.69	14,215	3.3	*Nocardioidaceae*
28th	91.73	3.23	29,968	3.1	*Xanthomonadaceae*
29th	91.33	1.21	15,927	5.7	*Mycobacteriaceae*
30th	82.26	9.74	5,017	4.4	*Bradyrhizobiaceae*
31st	74.49	1.32	5,029	3.2	*Xanthomonadaceae*
32nd	73.68	1.85	3,730	2.5	*Alphaproteobacteria*
33rd	80.6	0.49	6,496	4.5	*Bacteroidetes*
34th	81.82	1.21	3,958	3.3	*Sphingobacteriaceae*

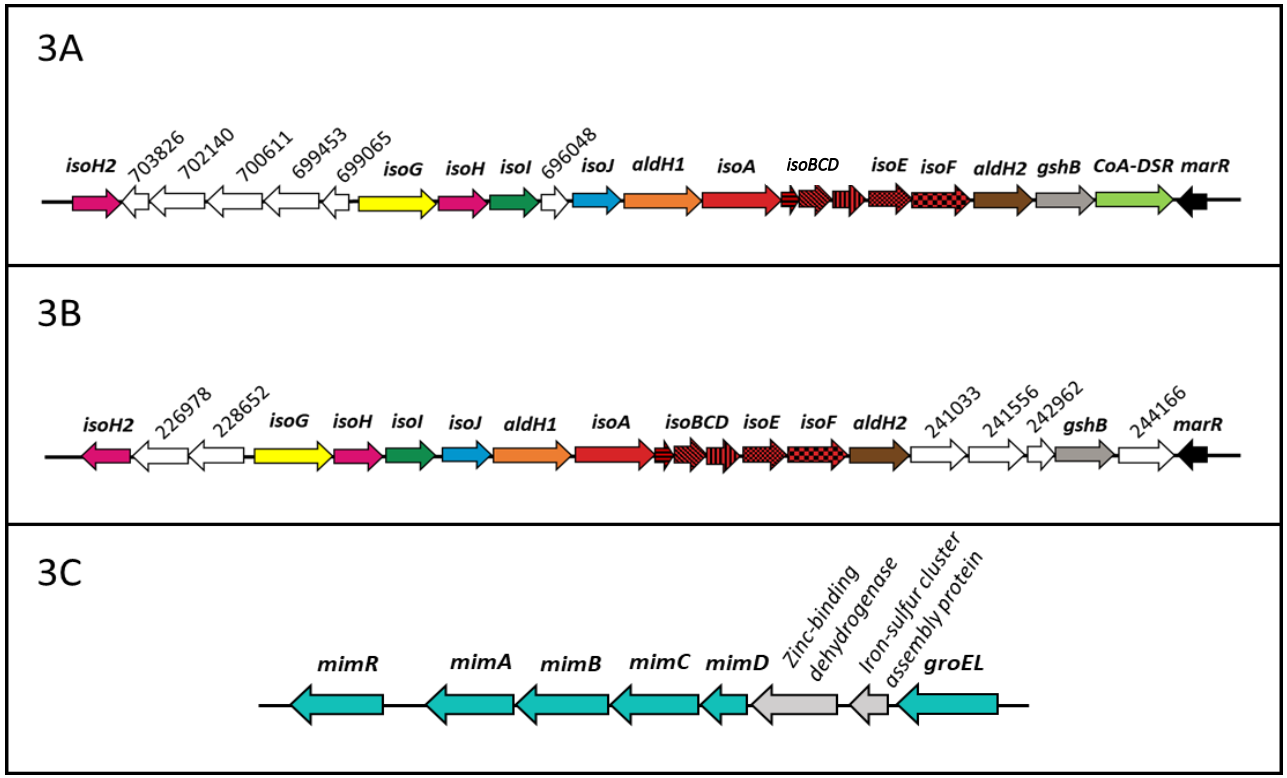

Figure 4.3. iso clusters 1 & 2 recovered from a *Mycobacterium* MAG assembled from heavy DNA retrieved after DNA-SIP. Genes encoding IsoMO (*isoABCDEF*) are coloured in red. Adjacent genes *isoGHIJ* and the duplicate gene *isoH2* encode a CoA transferase, dehydrogenase and two glutathione transferases involved in the subsequent steps of isoprene metabolism. Genes *aldH1, CoA-DSR, gshB* and *marR* encode an aldehyde dehydrogenase, a CoA-disulfide reductase, a glutathione synthase and a putative transcriptional regulator respectively. Adjacent genes that are not yet known to be involved in isoprene degradation are coloured in white. (696048 – Hypothetical protein; 699065 – Hypothetical protein; 699453 – Hypothetical protein; 700611 – Triacylglycerol lipase; 702140 - Acetyl-CoA-acetyltransferase; 703826 - AraC family transcriptional regulator; 226978 - Acetyl-CoA-acetyltransferase; 228652 – Hypothetical protein; 241033 – Hypothetical protein; 241556 – Hypothetical protein; 242962 - CaiB/BaiF family protein; 244166 - FAD-dependant oxidoreductase). Putative regulatory genes are shown in black.

While other isoprene degrading bacteria have been shown to contain duplicate downstream genes (*isoGHIJ*) in varying combinations, such as *Rhodococcus* AD45, *Gordonia* i37 and *Nocardioides* WS12 described in Chapter 3, this is the first recorded case of any of the IsoMO genes being present in duplicate in a single genome (Crombie *et al.*, 2015; Johnston *et al.*, 2017). Due to the lack of precedent, there was concern the finding may have been due to an artefact of assembly or contamination. To investigate the possibility, the contigs both clusters were located on were mined for essential marker genes, and each marker gene located was analysed for possible duplication, which would indicate the contig may have originated from a different genome. The MAGs in question were of both good quality and length, at 995,005 bp and 363,049 bp long with coverage at over x400. Each contig contained unique marker genes consistent with the genome of a member of *Mycobacterium*, indicating neither were present as a result of contamination.

It was also considered that one of the clusters could be present on a plasmid, as is the case for the entirety of the isoprene degradation gene clusters in *Rhodococcus* AD45 and *Variovorax* WS11 (Crombie *et al*. 2015; Dawson *et al*. 2020). However, when investigated for genes that would indicate the presence of a plasmid, none were found.

Investigation of both *iso* clusters showed that each gene shared a high translated amino acid identity with the corresponding genes in the isoprene degrading *Mycobacterium* AT1 isolated by Johnston *et al.* (2017) from an estuarine environment. Though it should be noticed that *Mycobacterium* AT1 contained only one copy of the IsoMO (Table 4.2).

Table 4.2. Comparison of polypeptides recovered from the duplicate isoprene degradation gene clusters (iso cluster 1 and iso cluster 2; Figure 4.3) found in a Mycobacterium MAG to those recovered from Mycobacterium AT1 and the well-characterised *Rhodococcus* AD45. Values are given as a percentage of shared amino acid (aa) identity (ID). IsoABCDEF make up the isoprene monooxygenase IsoMO, responsible for the first step of the isoprene degradation pathway. IsoGHIJ, (a CoA transferase, dehydrogenase and two glutathione transferases) encode genes catalysing the subsequent steps of isoprene metabolism.

		Comparison to *Mycobacterium* AT1		Comparison to *Rhodococcus* AD45	
Polypeptide	**Cluster**	**Coverage**	**ID (aa%)**	**Coverage**	**ID (aa%)**
IsoA	*1*	100	91	100	86.69
	2	96	96.8	100	82.24
IsoB	*1*	98	79.57	100	69.15
	2	100	85.1	98	70.97
IsoC	*1*	95	89.8	92	75.24
	2	100	90.4	99	72.57
IsoD	*1*	98	92.2	100	81.9
	2	98	96.3	100	79.09
IsoE	*1*	100	83.6	98	75.67
	2	100	85.7	99	76.11
IsoF	*1*	97	79.5	98	63.53
	2	97	80.9	97	63.45
IsoG	*1*	99	91.8	100	85.43
	2	100	93.3	100	83.29
IsoH	*1*	100	89.8	100	78.76
	2	100	92	100	79.2
IsoI	*1*	100	93.7	100	81.51
	2	100	94.1	100	85.29
IsoJ	*1*	99	88.4	99	75.86
	2	99	89.3	99	75.86

4.4.3 Identification of a propane monooxygenase gene cluster in *Mycobacterium*

Following examination of its potential isoprene degrading abilities, the *Mycobacterium* MAG was mined for other metabolic pathways of interest. A complete propane monooxygenase gene cluster was discovered, comprised of genes *mimABCD.* The propane monooxygenase operon transcriptional regulator *mimR* and the associated chaperon *groEL* were also found (Figure 4.3; Furuya *et al.*, 2012). This enzyme is a binuclear iron monooxygenase, with genes *mimABCD* encoding an oxygenase large subunit, a reductase, an oxygenase small unit and a coupling protein respectively. MimABCD have been shown to share a high amino acid identity with PrmABCD which make up the propane monooxygenase found in *Rhodococcus* sp. strain RHA1 (Sharp *et al*. 2007) and *Gordonia* sp. strain TY-5 (Kotani *et al*. 2003). In these bacteria, the propane monooxygenase is essential for propane and acetone metabolism, and is also capable of oxidizing phenol to

hydroquinone (Furuya *et al*. 2011). *Mycobacterium* AT1 was also shown to contain a propane monooxygenase, though it did not allow growth on phenol (Johnston *et al.*, 2017).

Table 4.3. Comparison of polypeptides recovered from a propane monooxygenase gene cluster recovered from a *Mycobacterium* MAG to those recovered from *Mycobacterium* AT1 and the well-characterised *Mycobacterium smegmatis mc²2155*. Values are given as a percentage of shared amino acid (aa) identity (ID). MimABCD, an oxygenase large subunit, a reductase, an oxygenase small unit and a coupling protein respectively, make up the propane monooxygenase, with GroEL being an associated chaperonin (Kotani *et al.*, 2003; Sharp *et al.*, 2007; see Figure 4.3(C)).

	Comparison to *Mycobacterium* AT1		**Comparison to *Mycobacterium smegmatis***	
Polypeptide	**Coverage**	**ID (aa%)**	**Coverage**	**ID (aa%)**
MimR	100	97.78	99	77.87
MimA	99	98.7	100	97.23
MimB	100	98.56	100	88.79
MimC	100	99.46	98	91.3
MimD	100	96.61	94	93.69
GroEL	100	99.09	98	49.72

The MimABCDR found in the *Mycobacterium* MAG all shared high amino acid identity (>97%) with the corresponding proteins in *Mycobacterium* AT1 (Table 4.3), which was shown to grow on both propane and ethane (Johnston, 2014; Johnston *et al.*, 2017). To confirm that these were indeed separate species, amino acid identity analysis (AAI) was carried out, and showed the *Mycobacterium* MAG recovered in this study, and *Mycobacterium* AT1 had a shared amino acid identity of 92.09%. Considering a shared identity of 95% or above is considered an indication that two genomes originate from the same species, this result showed the *Mycobacterium* MAG isolated from the phyllosphere of a willow in this study was distinct to *Mycobacterium* AT1 isolated from an estuarine environment. A comparison of Mim sequences to those of the well-characterised *Mycobacterium smegmatis* strain mc^2155 can be seen in Table 4.3. *Mycobacterium smegmatis* strain mc^2155 utilises MimABCD not only to metabolise propane and acetone, but for the regioselective oxidation of phenol to hydroquinone also, similar to *Rhodococcus* sp. strain RHA1 and the *Gordonia* sp. strain TY-5 as mentioned earlier (Furuya *et al*. 2011; 2012).

4.4.4 Recovery of a *Methylobacterium* MAG, present in the isoprene-degrading community as revealed by DNA-SIP

One other MAG of interest was recovered from metagenomic sequencing data and analysed for metabolic pathways of interest. A MAG identified as a member of the *Methylobacteriaceae* family was the 9th most abundant MAG recovered, with a high completeness of 98.86% and low contamination at 0.17% (Table 4.1). *Methylobacterium* were also shown by 16S rRNA gene amplicon sequencing to be enriched in the heavy fractions of one of the samples incubated with ^{13}C isoprene, with an RA of 26.48%, in replicate 3 (Figure 4.1). Although this MAG was shown to be almost complete, no *iso* genes were found upon analysis, despite being enriched in the isoprene degrading community of the DNA-SIP experiment.

A complete *mxa* methylotrophy gene cluster was identified in the *Methylobacteriaceae* MAG, encoding a calcium-containing methanol dehydrogenase (*mxaFJGIRSACKLDEHB*). The gene *mxaW*, encoding a methanol-regulated gene of unknown function, was also identified upstream of the methylotrophy gene cluster (Springer, Auman, and Lidstrom 1998; M. Zhang and Lidstrom 2003). Six genes, *pqqABC/DE* and *pqqFG*, required for pyrroloquinoline quinone (PQQ) synthesis were also located (M. Zhang and Lidstrom 2003). Genes involved in the transcriptional regulation of the methanol oxidation system, *mxbDM* and *mxcQE,* were also present. A comparison between the genes described and those from the well-characterised *Methylobacterium extorquens* AM1 can be seen in Table 4.4 (reviewed by Ochsner *et al*. (2014)).

Table 4.4. Comparison of polypeptides associated with the oxidation of methanol to formaldehyde recovered from a *Methylobacteriaceae* MAG, with those of the well characterised *Methylobacterium extorquens* AM1. Figures are given as a percentage of shared amino acid (aa) identity (ID). Mxa polypeptides are involved in C1 metabolism in *M. extorquens* and PQQ polypeptides are involved in the synthesis of pyrolloquinoline quinone (PQQ) a cofactor of methanol dehydrogenase. The designations used are as previously described (M. Zhang and Lidstrom 2003)**.**

	Comparison to *Methylobacterium extorquens* AM1	
Polypeptide	**Coverage**	**ID (aa%)**
MxaF	100	95.37
MxaJ	100	84.19
MxaG	100	89.34
MxaI	100	95.83
MxaR	100	90.96
MxaS	97	81.29
MxaA	83	65.44
MxaC	100	82
MxaK	96	72.2
MxaL	93	80
MxaD	96	76.88
MxaE	98	68.08
MxaH	93	72.28
MxaB	93	72.28
MxaW	79	68.4
PqqA	100	96.55
PqqB	100	81.61
PqqC/D	100	78.42
PqqE	100	86.86
PqqF	98	84.55
PqqG	98	78.30
MxbD	97	70.17
MxbM	99	85.84
MxcQ	93	66.31
MxcE	90	87.09

4.5 Conclusions

The purpose of this chapter was to investigate the isoprene degrading community harboured in the phyllosphere of a tree native to the UK that is known to emit high amounts of isoprene. Much previous work in this area has focused on bulk soil environments, but with isoprene concentrations three orders of magnitude higher near the stomata, leaf surfaces provide a new and isoprene-rich environment for such analyses. A unique finding of this study was the abundance of *Mycobacterium* and *Methylobacterium* in the presence of isoprene, neither of which have previously been shown to degrade isoprene in a terrestrial environment before. These genera were not identified in the soils associated with willow trees examined in previous studies, suggesting that the isoprene degrading community supported by phyllosphere and soil environments associated with the same host can be quite distinct (Larke-Mejia *et al*. 2019). To confirm this however, the soils and phyllosphere of the same individual host need to be directly compared, a study design explored in Chapter 5.

The presence of two complete copies of the isoprene degradation gene cluster within a single genome, as seen in the *Mycobacterium* MAG recovered here, has not been previously observed in any study of this kind. Another peculiarity is the abundance of *Methylobacterium* in labelled samples and the retrieval of a corresponding MAG with an apparent lack of any genes associated with isoprene metabolism. This warrants future work that may include focused isolation to allow for oxidative and growth assays to confirm whether the genus does in fact harbour the ability to metabolise isoprene, and whether the activity is carried out by an alternative pathway.

Alternative mechanisms of isoprene metabolism have been hypothesized in the past. Two such examples include a reductive isoprene metabolism, and that of lyase-dependant isoprene metabolism. Reductive isoprene metabolism was first described by Kronen *et al.*, (2019). Their study showed evidence of an anaerobic pathway of isoprene reduction, with isoprene acting as an electron acceptor for acetogenesis with HCO_3-. The isoprene reducing community was mainly composed of *Acetobacterium* species which were shown to form 40% less acetate in the presence of isoprene, suggesting that isoprene may play a role in energy conservation in these species. However, as of yet, this pathway has not been characterised and we still do not know what genes may be involved in the process.

The second alternative mechanism of isoprene metabolism was suggested by Rohwerder *et al.* (2020). When examining the genome of isoprene degrading *Rhodococcus* sp. ACPA4

(Crombie *et al*. 2017), they identified genes located in the region of the isoprene degradation gene cluster that were involved in the metabolism of 2-hydroxyisobutyrate. It was suggested that rather than the use of glutathione, there may be another mechanism of isoprene oxidation that utilised a lyase-dependant method that follows the pathway of isobutene metabolism. Though there is no functional data to support this hypothesis as of yet, and the complete lack of *iso* genes in the *Methylobacterium* MAG suggests that if it does possess the ability to oxidise isoprene, it is not via the hypothesized reductive pathway described.

5 Cultivation independent analyses of Malaysian oil palm soils and phyllosphere.

5.1 Summary

This chapter details a DNA-SIP experiment investigating the isoprene-degrading community associated with both the associated top-soil and the phyllosphere of Malaysian oil palm trees (*Elaeis guineensis*). A combination of both 16S rRNA gene amplicon sequencing and metagenomic analysis revealed the presence of *Novosphingobium, Pelomonas, Rhodoblastus, Sphingomonas* and *Gordonia* in the isoprene degrading soil and phyllosphere communities. Metagenome data were binned and a number of MAGs containing *isoA* homologues were recovered and investigated. *isoA* gene amplicon sequencing also revealed that both environments harbour a novel diversity of the isoprene monooxygenase genes.

This chapter is adapted from the original publication "Diversity of isoprene-degrading bacteria in phyllosphere and soil communities from a high isoprene-emitting environment: a Malaysian oil palm plantation" by Carrion *et al.* in Microbiome (2020). Samples were collected by collaborators in Malaysia. The DNA-SIP experiment including sample processing, incubation, DNA extraction and PCR amplification were all carried out jointly by myself and Dr. Ornella Carrión. All bioinformatic analyses were carried out by me.

5.2 Background

In Chapter 4, a willow species native to the UK and known to be one of the highest indigenous emitters of isoprene was investigated. In terms of global emission however, trees native to the UK produce relatively low levels of isoprene. For that reason, the study described in this chapter will broaden the scale of investigation and focus on Malaysian oil palm trees (*Elaeis guineensis*). While willow trees can produce up to 37 µg g(dry weight) h^{-1} (Pio *et al.,* 1993), the Malaysian oil palm produces isoprene at over four times that rate with an estimated emission rate of 175 µg g(dry weight) h^{-1} (Kesselmeier and Staudt 1999). Oil palm is a versatile and heavily cultivated crop. In terms of its use as a food product, over 69 million tonnes of palm oil are produced annually (*OECD-FAO Agricultural Outlook* , 2017) and it makes up 30% of global vegetable oil production (Cheng *et al*. 2019). Though the plant originated in Africa, oil palm plantations are now an extremely important crop in South East Asia today and in Malaysia especially, taking up more than 85% of total

agricultural land there (Cheng *et al*. 2019). Land usage associated with oil palm increases at an estimated 6.93% a year (Ibragimov *et al.,* 2019).

The vast expansion of a single crop plantation that emits such a high amount of isoprene is of importance when considering the impact of land usage on climate change. Oil palm plantations emit 2-5 times more isoprene than the average rainforest, and isoprene emissions associated with oil palm plantations are predicted to increase with rising global temperatures (C N Hewitt *et al*. 2011; Lantz *et al*. 2019). With this in mind, oil palm plantations represent an important environment in the isoprene biogeochemical cycle and have the potential to harbour a novel community of isoprene degrading bacteria that thrive in the presence of elevated concentrations of isoprene.

5.3 Experimental Methods

5.3.1 Sample Collection & DNA-SIP

Samples were collected from an oil palm plantation in Palong (Negeri Sambilan, Malaysia). Three different locations within the plantation were chosen and samples taken from one tree within each of the locations. The current crop within the plantation was mature at 28 years old. Soil was collected from the area directly beneath the tree canopy, 50 g was gathered from a depth of 0-5 cm, following the removal of any vegetation from the soil surface. For phyllosphere studies, five healthy leaflets from the lower canopy fronds of each tree were collected. Samples were shipped to Prof. Niall McNamara's laboratory at the UK Centre for Ecology and Hydrology in Lancaster to allow for processing under the appropriate import licences.

5 g of soil from each sample was resuspended in 50 mL of sterile distilled water in a sterile 250 mL conical flask. Flasks were shaken at 150 rpm for 30 minutes at room temperature to dislodge bacterial cells from the soil into the liquid phase. Soil suspensions were then decanted into sterile 50 mL measuring cylinders and left to stand for 1 hr to allow soil particles to settle. The aqueous layer was decanted to a sterile flask. This process was carried out twice per sample, with the inclusion of a sonication step in one processing run to account for cells that may either adhere too tightly to soil to be dislodged by shaking alone, and to also allow for cells that may be damaged by sonication to make it through the processing stage in an effort not to artificially influence the diversity associated with the soil environment. The aqueous layer of both processing methods were combined and transported in sealed vials to the University of East Anglia to create soil washing

microcosms. These were set up in triplicate with 80 mL of soil washings transferred to 2 L airtight bottle and supplied with 25 ppmv of either ^{12}C- or ^{13}C-labelled isoprene. Microcosms were then incubated in the dark at 30˚C with shaking at 150 rpm.

The five leaflets collected from each of the three trees were cut into pieces approximately 10 cm long and 5 cm wide. Pieces were inserted to sterile glass bottles and 250 mL of sterile distilled water added. Samples were sonicated for 5 mins in a water bath and shaken at 150 rpm at room temperature for 1 hr to dislodge cells from the leaf surface. Leaf washings were filtered through 0.22 µM cellulose nitrate membrane filter (Pall) to concentrate microbial cells. Filters were washed with 40 mL of Ewers minimal medium (see *Materials & Methods*). Washings from filters were then transferred to 2 L airtight bottle and supplied with 25 ppmv of either ^{12}C or ^{13}C labelled isoprene. Microcosms were then incubated in the dark at 30˚C with shaking at 150 rpm.

Isoprene concentrations in all samples were monitored using a Fast Isoprene Sensor (Hills-Scientific) and replenished when isoprene dropped below 10 ppmv. A sample was taken from each sample at T0 before the addition of the initial 25 ppmv of isoprene. 10 mL samples were again taken from soil replicates once isoprene measurements indicated that replicates had assimilated 12.5 µmol isoprene C assimilated g^{-1} (5 days of incubation). The same was done for leaf replicates once they had assimilated 50 µmol isoprene C g^{-1} (10 days of incubation). Aliquots were spun down and supernatants discarded. Pellets were resuspended in 1 mL sodium phosphate solution and MT buffers supplied with the FastDNA Spin Kit for Soil (MP Biomedicals) and then transferred to Lysing matrix E tubes for subsequent to DNA extraction as described in section 2.3. All subsequent steps of the DNA-SIP proceeded as described section 2.5.

5.3.2 Bacterial Community Analysis

DGGE was carried out as described in section 2.4.5 to determine whether enrichment had been successful and how best to select samples and replicates for metagenomic sequencing. 16S rRNA gene sequencing data were obtained and analysed at MrDNA (Shallowater, TX, USA). Illumina MiSeq technology was used to produce an average of 100,757 reads per sample with an average length of 300 bp. The MrDNA analysis pipeline consisted of the following steps. Reads were joined and barcodes removed. Short sequences >150 bp and sequences with ambiguous base calls were removed. Resultant sequences were denoised, and OTUs defined with clustering at 97% similarity. Singleton

sequences and chimeras were removed. Taxonomy of OTUs were determined with the use of BLASTn against a curated database derived from RPDII and NCBI.

isoA gene amplicon sequencing and metagenomic analysis were carried out as described in section 2.6.

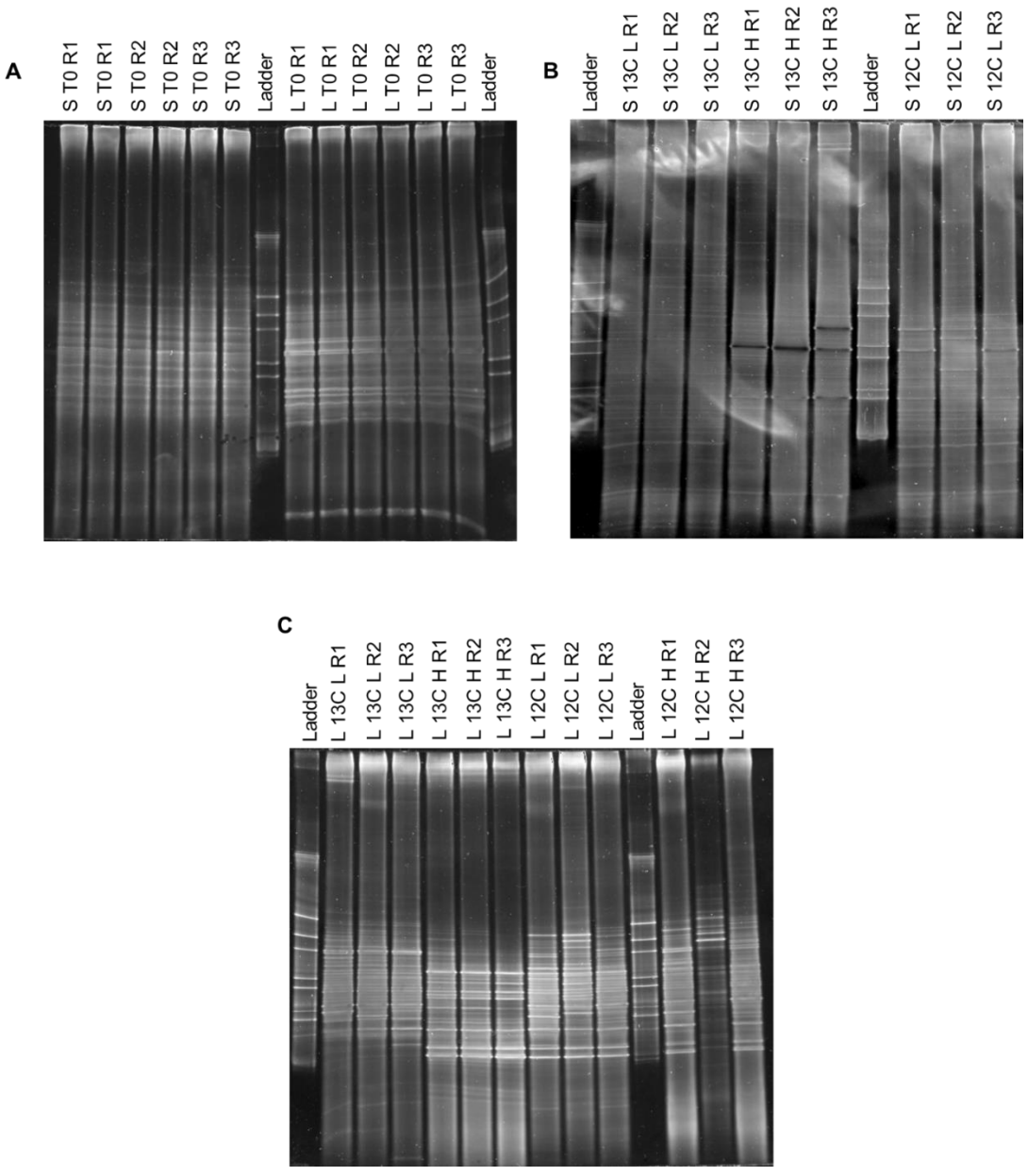

Figure 5.1. 16S rRNA gene profiles of Malaysian oil palm soil and phyllosphere samples analysed by DGGE. A: biological replicates from unenriched soil (S T0) and phyllosphere (L T0) samples, each of which was carried out in duplicate. B: biological replicates from heavy (H) and light (L) soil incubations with ^{13}C or ^{12}C isoprene. C: biological replicates from heavy (H) and light (L) phyllosphere samples enriched ^{13}C or ^{12}C isoprene.

5.4 Results & Discussion

5.4.1 Analysis of active isoprene-degrading bacteria via 16S rRNA gene amplicon sequencing

5.4.1.1 Diversity of bacteria harboured in soils associated with oil palm trees

16S rRNA gene amplicon sequencing showed a consistent bacterial community structure in all unenriched replicate soil samples. This homogeny across samples indicated that extraction and processing methods were consistent across samples and minimised the chance of artificial bias in the results (Figure 5.2). The bacterial community was dominated by phyla commonly found in soils such as Proteobacteria (40.8 ± 0.5%), Actinobacteria (13.1 ± 0.7%), Bacteroidetes (11.2 ± 1.4%) and Acidobacteria (10.8 ± 0.6%) (Janssen 2006; Roesch *et al*. 2007; Fulthorpe *et al*. 2008; Karimi *et al*. 2018). Dominant genera were *Rhodoplanes* (5.9 ± 0.1%) and *Flavobacterium* (4.0 ± 0.9%).

As mentioned in Chapter 4, the heavy fractions of samples incubated with ^{13}C-labelled isoprene represent the bacteria that have incorporated isoprene into their DNA via growth and replication and as such can be considered the isoprene degrading community. The isoprene degrading community recovered from soil samples showed some variability between replicates, however *Rhodoblastus* and *Pelomonas* were highly enriched across all replicates with an RA of 10.2-33.7% and 14.2-54.9% respectively (Figure 5.2). In replicates 2 and 3 of the isoprene degrading community *Novosphingobium* had an RA of 47.8% and 24.5% respectively. In replicate 3, *Sphingomonas* was dominant with an RA of 42.4%. These four genera showed a 19-90-fold increase in abundance when compared the ^{13}C light control. They were also abundant in ^{12}C light controls but not ^{12}C heavy, indicating that their presence in the heavy fraction of samples incubated with ^{13}C isoprene was due to incorporation of ^{13}C isoprene and not due to a high GC content (Figure 5.2).

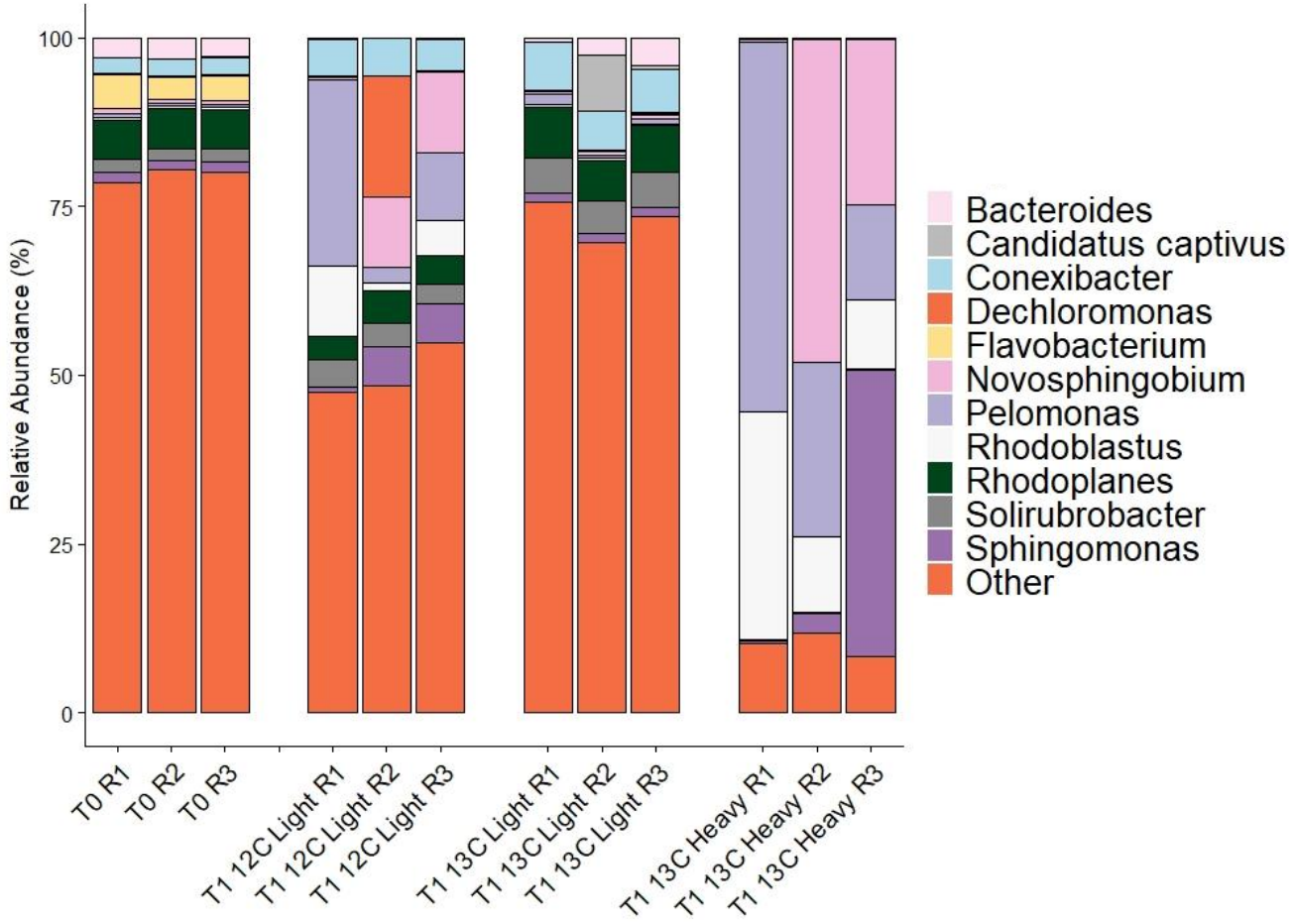

Figure 5.2. Bacterial community profile of DNA retrieved from oil palm soil samples analysed by 16S rRNA gene amplicon sequencing. Samples are represented as unenriched (T0), enriched (T1), incubated with unlabelled (12C) or labelled (13C) isoprene, with fractions separated into heavy DNA and light DNA retrieved after DNA-SIP. R1 – 3 indicate the three replicate samples analysed. Taxa that are at less than 5% relative abundance are grouped as 'Other'.

Previous investigations have also found members of *Sphingomonadaceae* and *Comamonadaceae* to be active isoprene degraders, but this is the first evidence that *Sphingomonas, Novosphingobium* and *Pelomonas* play a role in the biodegradation of isoprene (El Khawand *et al.*, 2016; Crombie *et al.*, 2018; Larke-Mejia *et al.*, 2019). Obtaining physical representatives of these strains in future work would be invaluable in confirming this greater diversity in the soil-based isoprene degrading population.

5.4.1.2 Diversity of bacteria in the phyllosphere of oil palm trees

The vast majority of unenriched phyllosphere samples were made up of Proteobacteria with an RA of 74.5% ± 0.3%. This is consistent with a number of previous community studies focused on the phyllosphere (Alvarez *et al.*, 2009; Johnston *et al.*, 2017; Larke-Mejia *et al.*, 2019). The most abundant genera were *Acinetobacter* (26.4 ± 0.7%) and *Clostridium* (22.0 ± 0.2%; Figure 5.3).

Unlike soil samples, the amount of DNA recovered from the heavy fractions of phyllosphere samples incubated with ^{13}C isoprene at earlier timepoints was not enough for the scale of phylogenetic analysis required for this study. As such, phyllosphere samples were incubated with labelled and unlabelled isoprene until T3 (10 days, when 50 µmol of isoprene C g^{-1} had been incorporated by each sample). At this point, a sufficient quantity (0.3 µg DNA) was obtained from the heavy fractions of samples incubated with ^{13}C isoprene.

It should be noted that the bacterial community profile of both the heavy and light fractions of samples incubated with ^{12}C isoprene were very similar, which is likely due to a disturbance in the CsCl gradient during processing of samples. However, the DNA recovered from the heavy fractions of these samples represents >1% of the total DNA utilised in the fractionation process, with no DNA at all being recovered from the heavy fraction of replicate 2 (Figure 5.3).

The results of 16S rRNA gene amplicon sequencing of the isoprene degrading community of phyllosphere samples were a highly consistent across replicates, as seen in the 'T1 13C Heavy' samples of Figure 5.3. *Gordonia* were particularly abundant with an RA of 51.4 ± 9.4%, followed by *Zoogloea* at 12.3 ± 2.2%. These RA values were 84.9 and 58.2-fold higher respectively in the heavy fractions of samples incubated with ^{13}C isoprene than in the light, strongly indicating that both were assimilating the ^{13}C-labelled isoprene (Figure 5.3).

The heavy enrichment of *Gordonia* in isoprene degrading fractions is consistent with previous studies that show strains of *Gordonia* are capable of utilising isoprene as a sole carbon and energy source (Johnston *et al.*, 2017; Larke-Mejia *et al.*, 2019). *Zoogloea* however, has not previously been identified in any of the environments that such studies have been carried out (El Khawand *et al.*, 2016; Johnston *et al.*, 2017; Crombie *et al.*, 2018; Larke-Mejia *et al.*, 2019). As such its presence here adds to the known diversity of isoprene degrading bacteria to date and warrants further investigation.

Although the genus *Rhizobium* was also enriched in the isoprene degrading fractions of phyllosphere samples, with an RA of 8.5 ± 2.2% across all three replicates, this only represents a 2.2-fold increase from the light fractions of samples incubated with ^{13}C isoprene. The same increase can be seen between *Rhizobium* in the heavy and light fractions of samples incubated with ^{12}C isoprene, which suggests that its presence in this fraction may be due to a higher GC content rather than the assimilation of labelled isoprene. With no previous examples of this genus utilising isoprene to support the theory, it cannot be stated that *Rhizobium* are isoprene degrading bacteria from these results alone (Figure 5.3).

Another notable result from this experiment is the distinct lack of genera that have previously been associated with isoprene degradation in the phyllosphere of trees in the UK. Examples such as *Rhodococcus, Variovorax, Sphingopyxis, Ramlibacter, Nocardioides* and *Mycobacterium* that have been present in similar studies on UK-based poplar and oil palm samples from Kew Gardens, London (Johnston *et al.*, 2017; Crombie *et al.*, 2018; Larke-Mejia *et al.*, 2019) were not present, or represented less than 1% of the labelled bacterial community in this experiment.

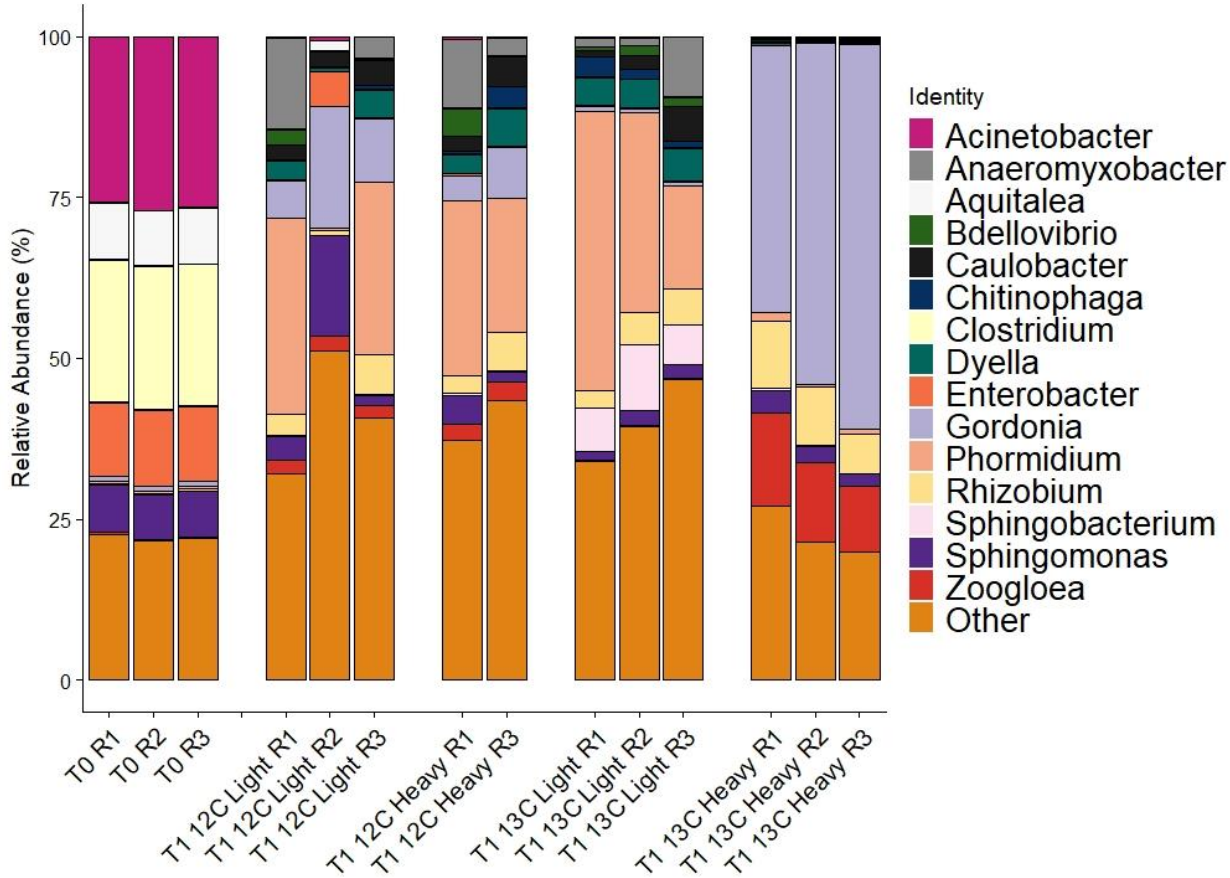

Figure 5.3. bacterial community profile of DNA retrieved from oil palm phyllosphere samples analysed by 16S rRNA gene amplicon sequencing. Samples are represented as unenriched (T0), enriched (T1), incubated with unlabelled (12C) or labelled (13C) isoprene, with fractions separated into heavy DNA and light DNA retrieved after DNA-SIP. R1 – 3 indicate the three replicate samples analysed. Taxa that are at less than 5% relative abundance are grouped as 'Other'

5.4.2 Diversity of isoprene monooxygenase genes recovered from oil palm phyllosphere and soil samples

To analyse the diversity of *isoA* genes within oil palm soil and phyllosphere samples, *isoA* amplicon sequencing was carried out and analysed via the ASV method as described in section 2.6.

Diversity of isoA genes from oil palm tree soils.

The *isoA* sequences retrieved from the heavy fractions of soil samples incubated with ^{13}C isoprene showed a higher degree of diversity than phyllosphere samples. However, all replicates were dominated by ASVs with homology closest to the *isoA* of the *Rhizobiales* MAG discussed in detail later (see section 2.7.2; Figure 5.3).The most abundant of these was ASV 2 which shared an amino acid identity of 100% with the *isoA* recovered from that MAG and had a RA of 35.5%±6.1 across the three soil replicates (Table 5.1). ASVs 5 and 4 were the next most abundant (RA 9.9%±7.1 and 7.5%±2.8 respectively) and also showed highest homology to the *isoA* the *Rhizobiales* MAG, retrieved during this SIP study (100% and 92.45% respectively). The ASV that dominated phyllosphere samples and shared a high 100% amino acid identity with *Gordonia* i37 (ASV 1) was also seen in soil samples, with an RA of 6.7%±2.8. ASV 13 (RA 7.5%±2.8) shared 100% identity with the *isoA* of *Rhodococcus* AD45, while ASV 11 shared 100% identity with the *isoA* recovered from a *Novosphingobium* MAG in this study and was present only in the heavy fraction of the ^{13}C-incubated replicate 3 with an RA of 11%.

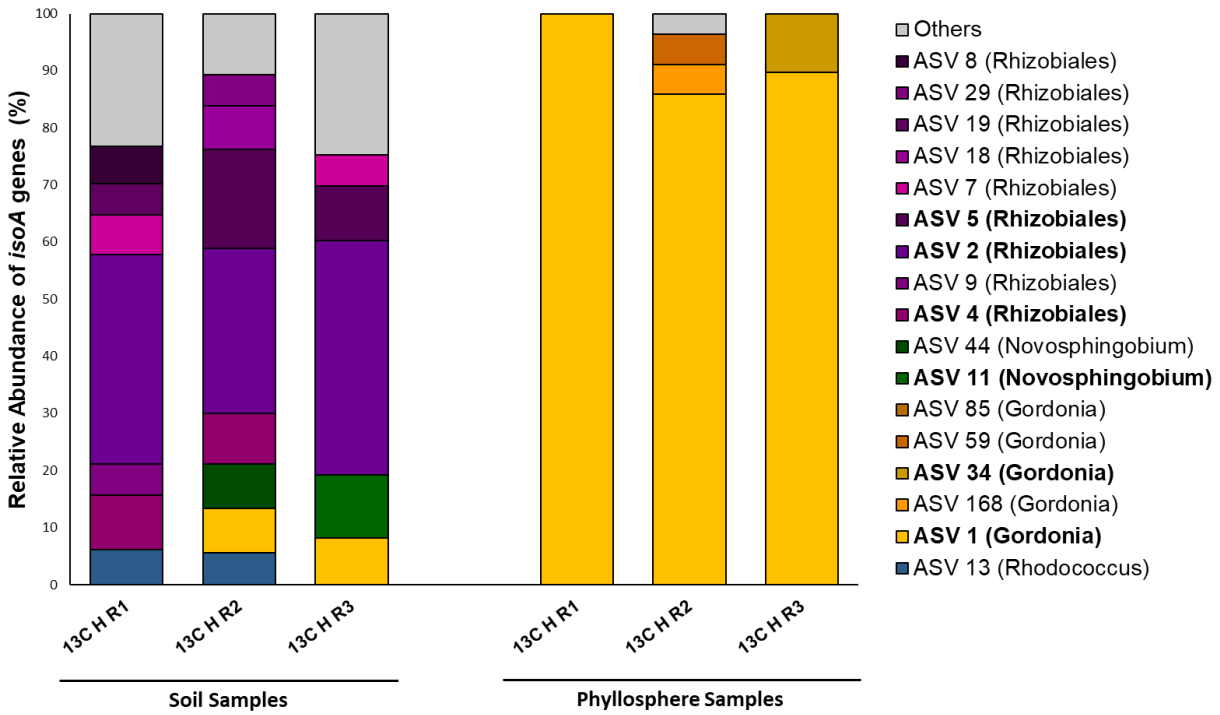

Figure 5.4. Relative abundance and diversity of *isoA* genes in the heavy fractions of oil palm soil and phyllosphere samples incubated with ^{13}C isoprene. *isoA* sequences in the heavy fractions of oil palm soil and phyllosphere samples incubated with ^{13}C isoprene were analysed by *isoA* amplicon sequencing. Only AVSs with > 5% RA in at least one replicate are represented. ASVs present at >10% in any sample are shown in bold. ASVs closely related to IsoA from the *Rhizobiales* MAG are represented in purple; ASVs with highest homology to IsoA from the *Novosphingobium* MAG are shown in green; ASVs closely related to IsoA from *Rhodococcus* are coloured in blue; ASVs with highest homology to IsoA from *Gordonia* MAG are represented in orange. ASVs with RA <5% are grouped as "Others".

Table 5.1. ASVs retrieved from isoA amplicon sequencing analysis of the heavy fractions of DNA from soil and leaf samples incubated with ^{13}C isoprene. Amino acid identity of ASVs retrieved from soil and phyllosphere samples to IsoA from ratified isoprene-degrading strains or MAGs reconstructed from metagenomes analysed in this study was determined by BLASTx (see Methods). Relative abundance (RA) of each ASV in heavy fractions of soil and phyllosphere (leaf) samples incubated with ^{13}C isoprene represent the average of three biological replicates with their respective standard deviations.

	Closest IsoA sequence	Source	Amino acid Identity(%)	Cover (%)	Soil Samples RA(%)	Leaf Samples RA(%)
ASV 1	*Gordonia i37*	Isolate	100	99	5.3±4.6	91.9±7.3
ASV 2	*Rhizobiales*	Metagenome	100	99	35.5±6.1	
ASV 4	*Rhizobiales*	Metagenome	92.5	99	6.1±5.3	
ASV 5	*Rhizobiales*	Metagenome	100	99	8.9±8.7	
ASV 7	*Rhizobiales*	Metagenome	100	99	4.2±3.7	
ASV 8	*Rhizobiales*	Metagenome	71	99	2.2±3.8	
ASV 9	*Rhizobiales*	Metagenome	98.7	99	1.8±3.1	
ASV 11	*Novosphingobium*	Metagenome	100	99	3.7±3.4	
ASV 12	*Rhizobiales*	Metagenome	98.9	99	1.4±1.3	
ASV 13	*Rhodococcus* AD45	Isolate	100	99	3.9±3.4	0.9±1.3
ASV 14	*Rhizobiales*	Metagenome	98.1	99	1.8±0.4	
ASV 15	*Rhizobiales*	Metagenome	96.2	99	2.5±2.4	
ASV 16	*Rhodococcus* AD45	Isolate	84.91	99	2.9±2.1	
ASV 17	*Rhizobiales*	Metagenome	92.45	99	0.6±1	
ASV 18	*Rhizobiales*	Metagenome	98.9	99	2.6±4.5	
ASV 19	*Rhizobiales*	Metagenome	100	99	1.8±3.1	
ASV 21	*Rhizobiales*	Metagenome	98.7	99	0.3±0.6	
ASV 23	*Rhodococcus* AD45	Isolate	99.37	99	1.2±1	
ASV 24	*Rhizobiales*	Metagenome	100	99	0.2±0.3	
ASV 26	*Rhizobiales*	Metagenome	98.1	99	0.2±0.4	
ASV 29	*Rhizobiales*	Metagenome	100	99	1.8±3.2	
ASV 30	*Rhizobiales*	Metagenome	98.7	99	0.1±0.3	
ASV 34	*Gordonia i37*	Isolate	98.1	99		3.4±5.6
ASV 37	*Rhizobiales*	Metagenome	96.2	99	0.1±0.2	
ASV 44	*Novosphingobium*	Metagenome	99.4	99	2.6±4.6	
ASV 53	*Rhizobiales*	Isolate	99.4	95	0.5±0.8	
ASV 59	*Gordonia i37*	Isolate	99.4	96		1.8±3
ASV 85	*Gordonia* i37	Isolate	99.4	96		2.7±3.8
ASV 110	*Rhizobiales*	Metagenome	99.4	100	0.1±0.2	
ASV 114	*Gordonia i37*	Isolate	100	99		0.9±1.3
ASV 122	*Rhizobiales*	Metagenome	95.6	100	0.1±0.2	
ASV 160	*Novosphingobium*	Metagenome	100	99	0.2±0.4	
ASV 168	*Gordonia i37*	Isolate	93.1	100		1.7±3.0

5.4.2.1 Diversity of isoA genes in DNA-SIP samples from the phyllosphere from oil palm trees

There was very little diversity in terms of *isoA* sequences found in the heavy fractions of phyllosphere samples incubated with ^{13}C isoprene. All three replicates were dominated by a single ASV, ASV1 which shared a 100% amino acid identity with the *isoA* from *Gordonia polyisoprenivorans* sp. i37 (Table 5.1; Johnston *et al.*, 2017). This ASV made up 100%, 85.87% and 89.75% RA of replicates 1-3 respectively (Figure 5.4). All other ASVs present in these samples also shared high homology with the *isoA* of *Gordonia* i37 though at different percentages of amino acid identity (Table 5.1).

5.4.3 Identification of active isoprene-degrading bacteria via metagenomics

The community profile and genetic potential of bacteria associated with oil palm samples were further investigated with the use of metagenomic analysis. DNA recovered from the heavy fractions of samples incubated with ^{13}C isoprene were pooled. In the case of phyllosphere samples, this meant pooling replicates 1-3, however in soil samples, DGGE experiments had indicated a difference in the taxa that were enriched in replicate 3 compared to that of replicates 1-2 (Figure 5.5). To better capture this difference, replicates 1 and 2 were pooled and replicate 3 was sequenced separately.

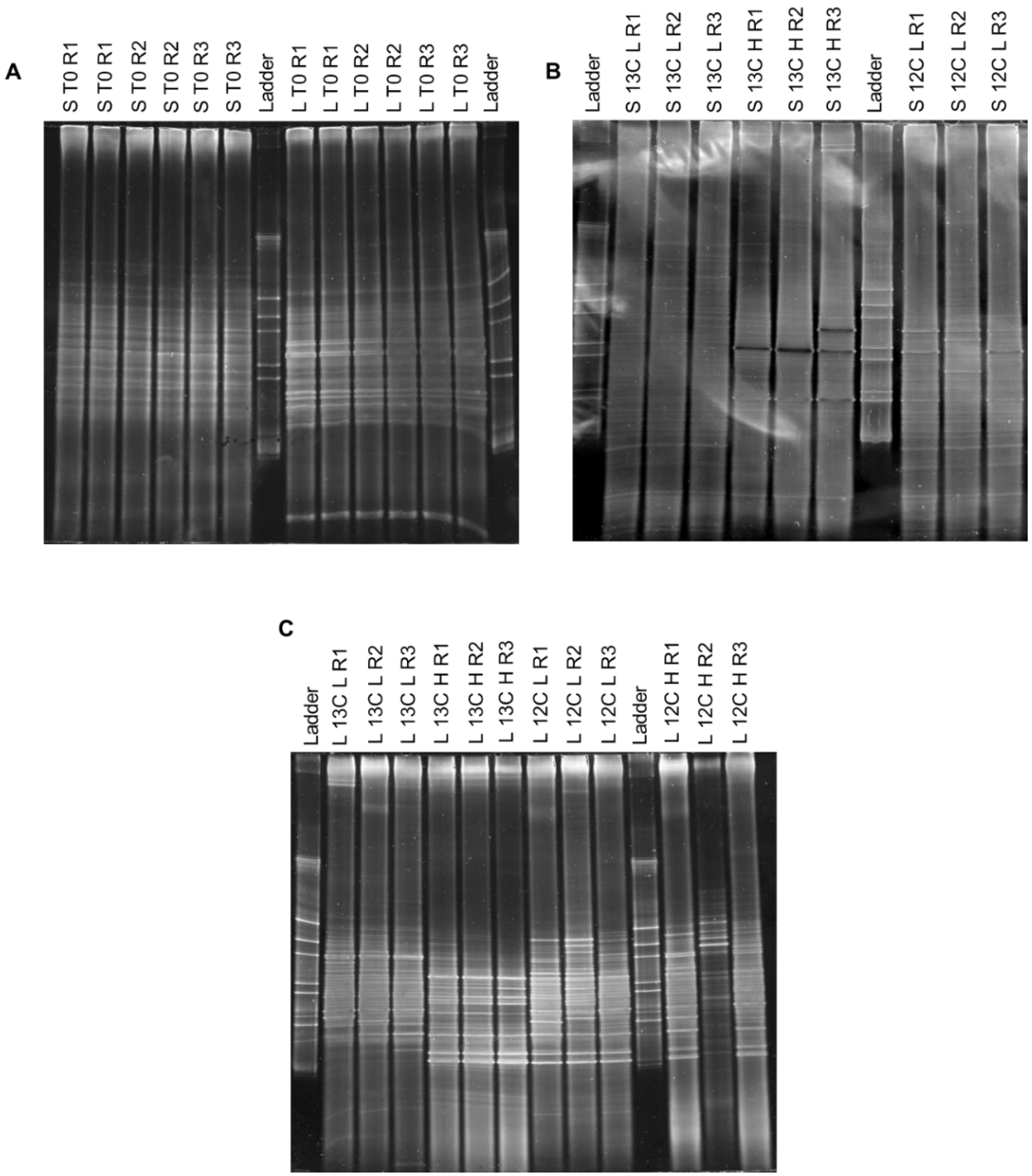

Figure 5.5. 16S rRNA gene profiles of oil palm soil and phyllosphere samples analysed by DGGE.

A: biological replicates from unenriched soil (S T0) and phyllosphere (L T0) samples, each of which was run in duplicate. B: biological replicates from heavy (H) and light (L) soil incubations with ^{13}C or ^{12}C isoprene (control). C: biological replicates from heavy (H) and light (L) leaf samples enriched ^{13}C or ^{12}C isoprene.

5.4.3.1 Metagenomic diversity of bacteria in the soil associated with oil palm trees

Metagenomic analysis of both unenriched (T0) and the enriched isoprene degrading community (T1) in soil samples showed a high diversity, with the vast majority of taxa representing less than 1% RA (represented as 'Other' in Figure 5.6). The unenriched metagenome showed an abundance of *Streptomyces* and *Bradyrhizobium* with RA of 11% and 6% respectively (Figure 5.6).

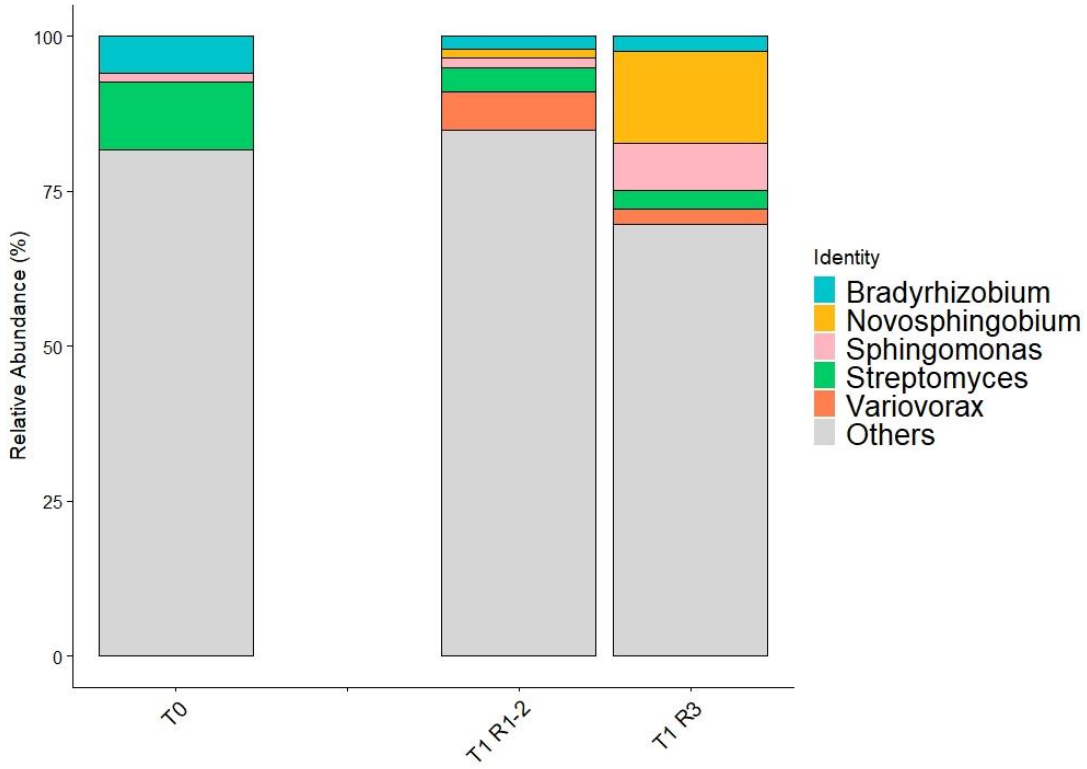

Figure 5.6. Relative abundance of bacterial taxa retrieved after analysis of the metagenome from oil palm soil samples. T0 represent unenriched samples. T1 are the heavy fractions of samples enriched with ^{13}C isoprene. Replicates 1 and 2 were combined (R1-2), replicate 3 was sequenced separately (R3). All taxa that were less than 1% RA have been grouped as 'Others'.

It was evident that the difference in the T1 isoprene degrading community structure observed via DGGE was caused by the particular enrichment of *Novosphingobium* in replicate 3, with an RA of 14.92% compared to the 1.37% found in the pooled replicates 1-2 (Figure 5.6). Replicate 3 also showed a higher enrichment of *Sphingomonas* with an RA of 7.63% in replicate 3 and 1.63% in replicates 1-2. *Variovorax* was present across all enriched samples with an RA of 6.19% in replicates 1-2 and 2.5% in replicate 3.

5.4.3.2 Metagenomic diversity of bacteria in the phyllosphere of oil palm trees

Unenriched phyllosphere samples contained a mix of different genera such as *Kosakonia* (RA of 16.5%), *Aquitalea* (RA of 16.34%), *Pantoea* (RA of 11.32%), *Enterobacter* (RA of 8%) and *Acinetobacter* (RA of 7.12%) as seen in Figure 5.7. Enriched samples however, were completely dominated by *Gordonia* with an RA of 60.34%, while no other taxa were present above 1% RA. This corroborates results from 16S rRNA gene amplicon sequencing that also showed an abundance of *Gordonia* above all other taxa within enriched phyllosphere samples (Figure 5.3). *Zooglea, Sphingomonas* and *Rhizobium* however, were not present in metagenomic community profiles.

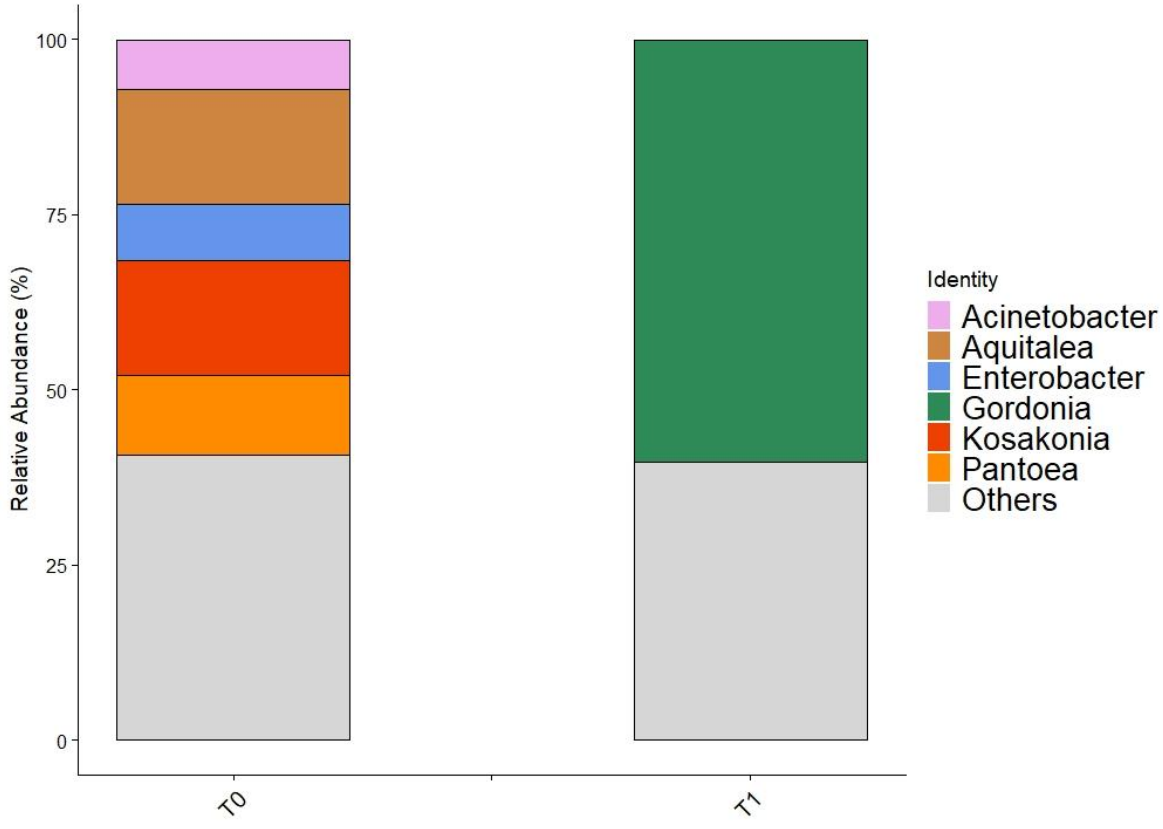

Figure 5.7. Relative abundance of bacterial taxa retrieved after analysis of the metagenome from oil palm phyllosphere samples. T0 represent unenriched samples. T1 are the heavy fractions of samples enriched with ^{13}C isoprene. All taxa that were less than 1% RA have been grouped as 'Others'

5.4.3.3 Recovery of metagenome assembled genomes (MAGs)

Metagenomic reads were assembled and utilised to recover a total of 20 MAGs from soil samples and 52 MAGs from leaf samples (Table 5.2). MAGs were investigated for the presence of genes involved in the degradation of isoprene (see section 2.7.2) and MAGs of interest chosen if the following parameters were met. MAGs must have >75% completeness, <10% contamination and contain one or more gene(s) from the isoprene monooxygenase gene cluster *isoABCDEF* or close homologue of such (*E* < 1e-40).

Table 5.2. Statistics for metagenome assemblies. Metagenomes from unenriched soil (S T0) and leaf (L T0) samples and heavy fractions from soil (S 13C H R1-2 and S 13C H R3) and leaf (L 13C H R1-3) samples incubated with ^{13}C isoprene.

	S T0	S 13C H R1-2	S 13C H R3	L T0	L 13C H R1-3
Contigs	138,886	36,166	22,852	76,131	114,878
Largest contig (bp)	75,971	703,868	1,215,493	732,750	621,426
Total size (bp)	101,440,201	64,831,442	49,671,873	152,876,094	150,393,192
GC (%)	63.9	64.9	64.6	56.5	66
N50	684	2,852	5,677	5,018	1,565
N75	568	1,078	1,322	1,171	798
L50	50,076	3,502	959	3,953	18,953
L75	91,069	13,317	6,174	21,847	54,079
Bins	4	9	7	28	24

By this metric, two MAGs from soil samples and three from phyllosphere samples were chosen for further investigation. The MAGs in question were identified as *Novosphingobium* and *Rhizobiales* from soil samples and *Gordonia, Zoogloaceae* and *Ralstonia* from phyllosphere samples. Basic assembly details for each can be seen in Table 5.3. Details of the isoprene degradation gene homologues recovered from each MAG are presented in Figure 5.8.

Table 5.3. Metagenome assembled genomes (MAGs) that contain genes encoding proteins homologous to IsoABCDEF (E<1e-40). MAGs were reconstructed from metagenomic sequencing of the heavy fractions of DNA from soil (S 13C H) and leaf (L 13C H) samples incubated with ^{13}C isoprene. MAGs completeness and contamination was assessed and taxonomically assigned using CheckM (see Methods). N50 is calculated for contigs.

MAG	Metagenome of origin	Size (Mbp)	N50	Completeness (%)	Contamination (%)	Strain heterogeneity (%)
Novosphingobium	S 13C H R3	3.7	441,005	99.5	<0.1	0
Rhizobiales	S 13C H R3	4.1	59,990	97.6	2.5	79
Gordonia polyisoprenovorans	L 13C H R1-3	6.1	194,257	99.8	0.9	40
Zoogloeaceae	L 13C H R1-3	5.2	49,423	98.7	2.1	10
Ralstonia	L 13C H R1-3	4.5	5,039	79	6.4	0

The *Novosphingobium* MAG recovered from the heavy fraction of soil samples incubated with ^{13}C isoprene contained a full isoprene gene cluster (*isoABCDEFGHIJ*) on a single contig however no accessory genes were present (Figure 5.8). The products of the isoprene degradation genes each shared an amino acid identity of >76% with the corresponding proteins in *Sphingopyxis* OPL5 (Larke-Mejia *et al.*, 2019; Table 5.3). However, the IsoA in particular shared a much higher identity (100% in some cases) with many of the ASV's recovered in the *isoA* amplicon sequencing results than to *Sphingopyxis* (Figure 5.8). This suggests that the homology to *Sphingopyxis* has more to do with the lack of representative Gram-negative isoprene degrading bacteria in the current database than to the MAG being particularly closely related to *Sphingopyxis* OPL5. The recovery of this MAG is congruent with the other methods of analysis in this study, with *Novosphingobium* also being abundant in soil samples when analysed by 16S rRNA gene amplicon sequencing (Figure 5.2) and in community analysis of metagenomic data (Figure 5.6).

Table 5.4. MAGs genes encoding polypeptides homologous to proteins involved in isoprene metabolism from ratified isoprene-degrading strains. Homology of the polypeptides encoded by genes recovered from MAGs to proteins from ratified isoprene degraders was analysed by BLASTx (see Methods) and is expressed as amino acid identity.

***Novosphingobium* MAG**

Gene	Description	Closest protein from ratified isoprene degrader	Amino acid identity (%)	Coverage (%)
isoA	Oxygenase α-subunit	IsoA from *Sphingopyxis* sp. OPL5	95.2	100
isoB	Oxygenase γ-subunit	IsoB from *Sphingopyxis* sp. OPL5	100	85.1
isoC	Reiske-type ferrodoxin	IsoC from *Sphingopyxis* sp. OPL5	100	78.6
isoD	Coupling protein	IsoD from *Sphingopyxis* sp. OPL5	100	90.6
isoE	Hydroxylase β-subunit	IsoE from *Sphingopyxis* sp. OPL5	83.7	100
isoF	Flavoprotein NADH reductase	IsoF from *Sphingopyxis* sp. OPL5	76.2	100
isoG	Racemase	IsoG from *Sphingopyxis* sp. OPL5	91.1	100
isoH	Dehydrogenase	IsoH from *Sphingopyxis* sp. OPL5	89.4	100
isoI	Glutathione-S-transferase	IsoI from *Sphingopyxis* sp. OPL5	89.8	100
isoJ	Glutathione-S-transferase	IsoJ from *Sphingopyxis* sp. OPL5	90.6	98
aldH1	Aldehyde dehydrogenase	AldH1 from *Sphingopyxis* sp. OPL5	85.3	100

***Rhizobiales* MAG**

Gene	Description	Closest protein from ratified isoprene degrader	Amino acid identity (%)	Coverage (%)
isoA	Oxygenase α-subunit	IsoA from *Sphingopyxis* sp. OPL5	84.9	100
isoB	Oxygenase γ-subunit	IsoB from *Sphingopyxis* sp. OPL5	67.4	97
isoC	Reiske-type ferrodoxin	IsoC from *Sphingopyxis* sp. OPL5	61.2	92
isoD	Coupling protein	IsoD from *Ramlibacter* sp. WS9	60	92
isoE	Hydroxylase β-subunit	IsoE from *Sphingopyxis* sp. OPL5	54.8	98
isoF	Flavoprotein NADH reductase	IsoF from *Sphingopyxis* sp. OPL5	54.9	99
isoG	Racemase	IsoG from *Sphingopyxis* sp. OPL5	72.9	96
isoH	Dehydrogenase	IsoH from *Sphingopyxis* sp. OPL5	74.3	100
isoI	Glutathione-S-transferase	IsoI from *Sphingopyxis* sp. OPL5	72.9	95
isoJ	Glutathione-S-transferase	IsoJ from *Sphingopyxis* sp. OPL5	69.9	100
aldH1	Aldehyde dehydrogenase	AldH1 from *Sphingopyxis* sp. OPL5	64.7	99

***Gordonia polyisoprenovorans* MAG**

Gene	Description	Closest protein from ratified isoprene degrader	Amino acid identity (%)	Coverage (%)
isoA	Oxygenase α-subunit	IsoA from *Gordonia polyisoprenovorans* i37	98.8	100
isoB	Oxygenase γ-subunit	IsoB from *Gordonia polyisoprenovorans* i37	82.1	100
isoC	Reiske-type ferrodoxin	IsoC from *Gordonia polyisoprenovorans* i37	99.1	100
isoD	Coupling protein	IsoD from *Ramlibacter* sp. WS9	99	95
isoE	Hydroxylase β-subunit	IsoE from *Gordonia polyisoprenovorans* i37	98.8	100
isoF	Flavoprotein NADH reductase	IsoF from *Gordonia polyisoprenovorans* i37	99.4	100
isoG	Racemase	IsoG from *Gordonia polyisoprenovorans* i37	98.3	100
isoH	Dehydrogenase	IsoH from *Gordonia polyisoprenovorans* i37	99.1	100
isoI	Glutathione-S-transferase	IsoI from *Gordonia polyisoprenovorans* i37	100	100
isoJ	Glutathione-S-transferase	IsoJ from *Gordonia polyisoprenovorans* i37	94.7	100
aldH1	Aldehyde dehydrogenase	AldH1 from *Gordonia polyisoprenovorans* i37	83	100
aldH2	Aldehyde dehydrogenase	AldH2 from *Gordonia polyisoprenovorans* i37	94.8	100
gshB	Glutathione synthetase	GshB from *Gordonia polyisoprenovorans* i37	99.1	98
coADR	CoA disulfide reductase	CoADR from *Gordonia polyisoprenovorans* i37	94	100

***Zoogloeaceae* MAG**

Gene	Description	Closest protein from ratified isoprene degrader	Amino acid identity (%)	Coverage (%)
isoA	Oxygenase α-subunit	IsoA from *Variovorax* sp. WS11	51.1	100
isoB	Oxygenase γ-subunit	IsoB from *Ramlibacter* sp. WS9	51.7	97
isoC	Reiske-type ferrodoxin	IsoC from *Ramlibacter* sp. WS9	51	93
isoD	Coupling protein	IsoD from *Gordonia* sp. OPL2	44.9	85
isoE	Hydroxylase β-subunit	IsoE from *Ramlibacter* sp. WS9	43	85
isoF	Flavoprotein NADH reductase	IsoF from *Ramlibacter* sp. WS9	39	97

***Ralstonia* MAG**

Gene	Description	Closest protein from ratified isoprene degrader	Amino acid identity (%)	Coverage (%)
isoA	Oxygenase α-subunit	IsoA from *Variovorax* sp. WS11	48.3	100
isoB	Oxygenase γ-subunit	IsoB from *Ramlibacter* sp. WS9	52.5	83
isoC	Reiske-type ferrodoxin	IsoC from *Ramlibacter* sp. WS9	47.6	92
isoD	Coupling protein	IsoD from *Rhodococcus opacus* PD630	44.2	81
isoE	Hydroxylase β-subunit	IsoE from *Nocardioides* sp. WS12	45.1	94
isoF	Flavoprotein NADH reductase	IsoF from *Rhodococcus opacus* PD630	39.1	100

The second MAG recovered from soil samples was identified as a member of the order *Rhizobiales*, though further identification was not successful despite the MAGs high completeness (97.63%) and low contamination (2.48%; Table 5.3). This is likely due to the high strain heterogeneity (78.95%) which indicates what contamination is present and stems from closely related bacteria, making robust high-resolution identification difficult.

This *Rhizobiales* MAG also contained a full isoprene gene cluster (*isoABCDEFGHIJ*), and the products of these genes also shared the highest amino acid identity (54% and above; Table 5.4) with *Sphingopyxis* OPL5. Two of the ASVs recovered in *isoA* amplicon sequencing shared 100% homology with the IsoA recovered from the *Rhizobiales* MAG (Figure 5.8). 16S rRNA gene amplicon sequencing of ^{13}C-incubated soil samples showed an enrichment of *Rhodoblastus*, a member of the *Rhizobiales* family (Figure 5.2).

The first of three MAGs recovered from phyllosphere samples was identified to the species level as *Gordonia polyisoprenivorans* and contained a full isoprene monooxygenase gene cluster on one contig, with the downstream genes *isoGHIJ* found on another (Figure 5.8).

IsoI was only partially recovered at the end of the contig containing the downstream genes. The *Gordonia* MAG also contained accessory genes *aldH2, garB* and *gshB*, often located in the region of the isoprene degradation pathway. These genes encode for an aldehyde dehydrogenase, a CoA-disulfide reductase and a glutathione synthetase respectively (Crombie *et al*. 2015). The products of genes directly related to isoprene degradation and all accessory genes shared an amino acid identity >98% with the corresponding proteins recovered from the genome of *Gordonia polyisoprenovorans* strain i37 (Johnston *et al.*, 2017; Table 5.4), except IsoB and AldH1 which shared 82.11% and 83% respectively.

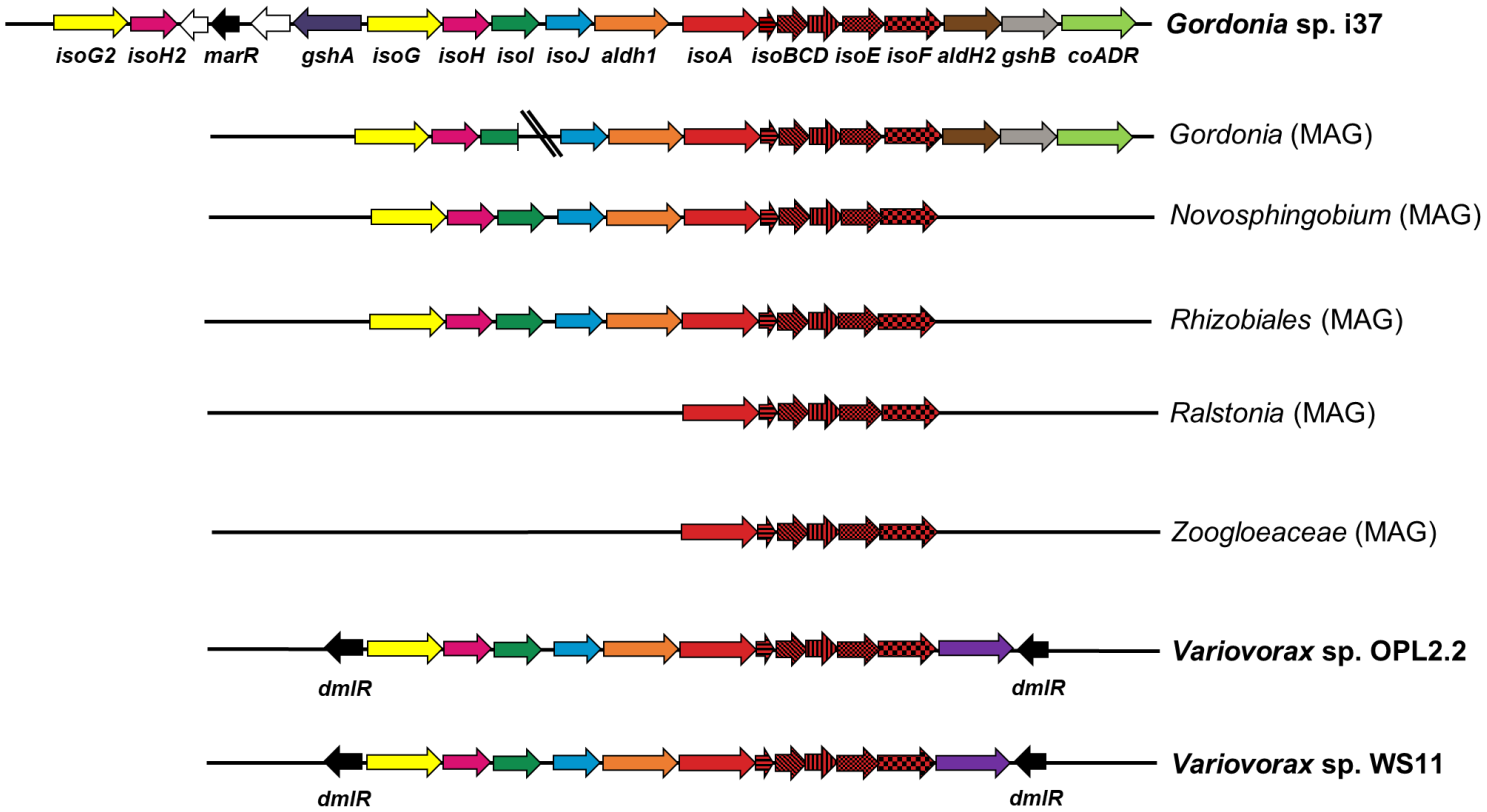

Figure 5.8. Isoprene metabolic gene clusters from representative isoprene-degrading strains (in bold) and metagenome assembled genomes (MAGs). Genes encoding IsoMO (*isoABCDEF*) are coloured in red. Adjacent genes not involved in isoprene degradation are coloured in white. Regulatory genes are shown in black. "\\" represents a discontinuity between two DNA contigs. Variovorax sp. OPL2.2 was isolated in this study from oil palm leaf enrichments by Dr. Ornella Carrión.

The remaining MAGs recovered from phyllosphere samples, *Zoogloaceae* and *Ralstonia* contained genes homologous to the isoprene monooxygenase *isoABCDEF* but not the downstream *isoGHIJ* genes (Figure 5.8). The shared identity with isoprene monooxygenase genes from bona fide isoprene degraders was lower than seen in previous MAGs also. Some of the gene products in the *Zoogloaceae* MAGs cluster shared an amino acid identity with toluene monooxygenase recovered from *Pseudomonas mendocina* KR1 (Whited and Gibson 1991) and others with isoprene monooxygenase proteins recovered from *Variovorax* WS11 (Robin A. Dawson *et al*. 2020). However the amino acid identity ranged from 38%-71% making classification unreliable (Table 5.4). A member of the family *Zoogloeaceae, Zooglea* was also found in 16S rRNA gene amplicon results for phyllosphere samples (Figure 5.3).

The *Ralstonia* MAG was similar, in that some of the recovered genes shared homology with proteins of the toluene monooxygenase of *Pseudomanas mendocina* KR1, and others alternated between homology with proteins recovered from known isoprene degrading bacteria such as *Variovorax* WS11*, Ramilibacter* WS9, *Rhodococcus opacus* PD630 and *Nocardioides* WS12 (Crombie *et al.*, 2017; Larke-Mejia *et al.*, 2019; Dawson *et al.*, 2020; Gibson *et al.*, 2020; Table 5.4). *Ralstonia* was previously included in the family *Pseudomonas* which may play a role in the shared homology between the two.

It is possible that the lack of known downstream genes and the non-specific characterisation of these clusters could indicate the presence of the novel pathway for the degradation of isoprene. In particular, the fact that these MAGs were recovered from samples that had been provided with isoprene as the sole carbon and energy source, and had obviously incorporated ^{13}C isoprene into their DNA to be recovered from heavy fractions after CsCl density gradient ultracentrifugation, implies they can assimilate isoprene by some means. However, without the isolation of a representative bacteria that grows on isoprene, or the heterologous expression of their homologous genes in an *isoA-F* mutant, it cannot be stated with confidence that the *Zoogloaceae* and *Ralstonia* MAGs represent the presence of novel isoprene degrading bacteria.

5.4.4 Comparison of bacterial diversity in the oil palm phyllosphere and associated soils

While the importance of plant-microbe interactions is a well-studied area with many implications for biotechnology, this is one of a small number of studies that compares the microbial communities in the phyllosphere and soils of the same species (Knief *et al*. 2012; Bodenhausen, Horton, and Bergelson 2013; Wei *et al*. 2018; Q. Zhang *et al*. 2019; R. Yang, Liu, and Ye 2017; Azevedo-Silva *et al*. 2021).

The results from 16S rRNA gene amplicon sequencing showed the natural, unenriched community structures of oil palm soils and phyllosphere were quite distinct (Figure 5.2; Figure 5.3). In general, the combined results of this study indicated that the isoprene degrading community harboured by soils associated with oil palm is more diverse than that found in the oil palm phyllosphere environment. This is particularly notable for metagenome and *isoA* amplicon results, where *Gordonia* and homologues of its genes dominate all other taxa in their abundance in phyllosphere samples, though 16S rRNA results do show a somewhat greater diversity. This abundance of *Gordonia* and *Gordonia*-like sequences in oil palm phyllosphere samples is consistent with *isoA* amplicon results described in Chapter 3.

The vast majority of *isoA* sequences recovered from soils were homologous to *isoA* from Gram-negative groups such as *Rhizobiales* and *Novosphingobium*, though shared amino acid identity ranged from 71-100% which suggests some of these genes are derived from other bacteria than those previously isolated. At the time of study, there were very few *isoA* sequences available from Gram-negative species, meaning those available were the only candidates for comparison, which may explain their high representation in soil samples. The higher diversity of *isoA* sequences in soils is unsurprising as research shows that soils are some of the most complex environments to be studied to date (Delmont *et al*. 2011).

The dominance of Gram-positive related sequences in the phyllosphere suggests these Gram-positive species thrive under higher concentrations of isoprene, which is consistent with research (described in section 1.3.2) which showed that Gram-positive Actinobacteria such as *Rhodococcus* and *Gordonia* were recovered with experiments utilising higher concentrations of isoprene during incubation, whereas Gram-negative species like those found in soil samples, such as *Variovorax* and *Sphingopyxis,* were only recovered when

lower concentrations were utilised (Crombie *et al.*, 2015, 2018; Johnston *et al.*, 2017; Larke-Mejia *et al.*, 2019).

Metagenomic data shared some strong similarities to 16S rRNA gene amplicon data. They showed a homogenous community structure in phyllosphere samples completely dominated by *Gordonia*, which was highly abundant in 16S rRNA amplicon data. Analysis of soil samples revealed an abundance of *Novosphingobium*, as found with 16S rRNA gene amplicon data and likely acting as the source of the recovered *Novosphingobium* MAG and the ASVs homologous to its *isoA*. *Novosphingobium* has been shown to have plant-growth promoting properties and is often found in the rhizosphere, so it is perhaps unsurprising to find it in each of the three soil datasets (Vives-Peris, Gómez-Cadenas, and Pérez-Clemente 2018; Rangjaroen *et al*. 2017). *Novosphingobium* species are also known for their ability to degrade a range of recalcitrant compounds such as polychlorophenol and high-molecular-mass polycyclic aromatic hydrocarbons (Tiirola *et al.*, 2002; Sohn *et al.*, 2004; Kertesz and Kawasaki, 2010; Lyu *et al.*, 2014).

While it has been shown that the isoprene degrading community harboured by oil palm soils are more diverse than their phyllosphere counterpart, whether they were also more abundant in this environment was also investigated. Unenriched oil palm soil and phyllosphere metagenomic datasets were analysed for the presence and relative abundance of *isoA* genes. This study will be explained in detail in Chapter 6, but results showed that the number of *isoA* sequences harboured by oil palm soil samples were five times higher than was found in phyllosphere samples. This represents 1% of 16S rRNA genes in soil samples, compared to 0.2% in the phyllosphere. As such, this study suggests that in oil palm at least, isoprene degrading bacteria are both more diverse and more abundant in associated soils than in the phyllosphere. This is interesting when one considers the greater availability of isoprene in the canopy than at ground-level (Rasmussen and Khalil 1988) and indicates that soils could be a more important sink for isoprene than previously thought.

5.5 Conclusion

Oil palm plantations are, and will continue to be, a monumental part of the economy in places like Malaysia and Indonesia. Palm oil production is also a rapidly growing industry in parts of South America such as Colombia, Ecuador and Guatemala. These plantations take up huge swathes of agricultural land that spread on an annual basis to supply palm oil for the food industry, the production of domestic products and for use as biofuel (Cheng *et al*. 2019; Ibragimov *et al.,* 2019; Carter *et al*. 2007). While such a monoculture has wide-ranging impacts on deforestation and biodiversity, the fact that oil palms produce such high amounts of isoprene has implications for future air quality too (Hewitt *et al*. 2009). A better understanding of the biological isoprene sinks associated with these environments is an important part of understanding how these biogeochemical fluxes may change over time with the impacts of increased land-usage and climate change.

As with most ecological studies of this kind, limitations include the small number of sampled locations and the impact that soil-type, weather and other biotic and abiotic factors may have on the isoprene degrading population harboured in the sampled soils and leaves. A truly robust investigation of the association between oil palms and their isoprene degrading bacteria would require samples from the same species of tree but from a wide range of different environments, particularly the inclusion of wild oil palms rather than heavily cultivated crops like those used in this study. However, the benefit of using single crop plantations is that there should be no cross-mixing from bacteria associated with different nearby tree species.

Future work to consolidate these findings should include targeted isolation techniques to recover physical isolates and confirm their isoprene degrading capabilities. Of particular interest would be representative isolates of the MAGs identified as *Ralstonia* and *Zoogloaceae*, both of which are novel in terms of isoprene degrading bacteria, and both MAGs seem to be missing the downstream genes associated with the degradation pathway despite growing on isoprene as the sole source of carbon and energy. As suggested earlier, an alternative to this labour-intense method would be the homologous expression of their *isoA-F* homologues in an *isoA-F* mutant to determine whether the ability to degrade isoprene could be restored, as was shown by Crombie *et al.* (2015) for *isoA-F* from *Rhodococcus* AD45.

6 Isoprene biodegradation survey associated with a variety of terrestrial plants

6.1 Summary

This chapter details efforts to lay the groundwork for a large-scale survey to locate and quantify the proportion of bacteria in unenriched natural communities that have the genetic potential to degrade isoprene. The first section is lab-based and uses qPCR techniques in conjunction with the primers validated in Chapter 3 to amplify and quantify *isoA* genes in soil and phyllosphere samples collected from isoprene-emitting willow (*Salix viminalis*), poplar (*Populus nigra*) and non-emitting *Miscanthus* (*Miscanthus x giganteus*) plants. Samples were collected from three plants of each species, from each individual plant, three samples each were taken from the soil and phyllosphere environment. The results of this survey showed high intra-species variability in the abundance of *isoA* gene copies in both phyllosphere and soil sample. Between species, the pattern of *isoA* abundance was the same in both phyllosphere and soil samples, with bacteria in the poplar environments harbouring the highest number of *isoA* sequences when normalized to copies of 16S rRNA genes, followed by those in willow-associated environments, and finally those associated with *Miscanthus*. These results do not support the hypothesis that the scale of isoprene emission per plant species would be the strongest variable in terms of proportion of resident bacteria carrying an *isoA* gene, as *Salix viminalis* produces more isoprene than *Populus nigra,* and *Miscanthus* species produce no isoprene at all.

The abundance of *isoA* genes in the environment was also investigated via metagenomic analyses. Both publicly available natural, unenriched metagenomic datasets and also metagenomic datasets from previous studies carried out from within the lab were collated, processed in a reproducible manner and then investigated for the number of *isoA* sequences per 16S rRNA genes found in each metagenome.

6.2 Introduction

While most studies have focused on the diversity of isoprene degrading bacteria in the environment, not much is known about the abundance of these bacteria under unenriched conditions. Although global emissions of isoprene are substantial, it's volatile nature means isoprene does not persist in the atmosphere. This means the bioavailability of isoprene is relatively low compared to other climate active gases such as methane or CO_2. It is for this reason that almost all diversity studies related to isoprene focus on enriched samples. However, to accurately assess the scale of bacterial degradation of isoprene, and understand the role of isoprene metabolism in terms of its place in the biogeochemical cycle, we must know how common isoprene biodegradation is under true environmental conditions.

It was described in Chapter 3, that the gene encoding the α-subunit of the isoprene monooxygenase is highly conserved and acts as an excellent biomarker for isoprene degradation in the environment. As such, in this chapter, the presence of *isoA* in a bacterium's genome is used as an indicator of the ability to degrade isoprene. Both qPCR and metagenomic quantification techniques are utilised across a range of environments. For qPCR assays, 16S rRNA gene copy number was used as a normalisation factor. While single copy house-keeping genes such as recA or *rpoB* would be more suitable standards for normalisation, as it is known that 16S rRNA genes can be present anywhere from 1-15 times in a single genome (Klappenbach *et al*. 2001), *recA* and *rpoB* genes do not have well-validated universal primers and cannot be used confidently across complex environmental systems. This variability means that the 16S rRNA gene normalisation method cannot function as an absolute abundance measure, though it can give an approximate and comparative idea of the abundance of *isoA* across a variety of environment types analysed in the same manner. In the case of metagenome analyses however, the availability of universal primers was not an issue and as such, *recA* was used to normalise *isoA* numbers to give the most accurate estimation of abundance possible.

The benefits of qPCR are many and obvious. It is a well-know and familiar work-flow, relatively inexpensive compared to next-generation sequencing (NGS) and is a widely-used and validated method for a number of use-cases, from ecological studies to medical diagnostics. In terms of ecological surveys, while qPCR assays themselves take little time, collecting and processing samples can be more difficult, depending on the breadth and scale of the survey. For this reason, supplementing such research with readily available

metagenomic datasets can be of great benefit. At the time these data were analysed (July 2021) the metagenomic database MG-RAST contained 469,279 publicly available metagenomes, TerrestrialMetagenomeDb had 20,206 and the Sequence Read Archive (SRA) had 458,433 (Meyer *et al*. 2008; Corrêa *et al*. 2020; Leinonen *et al*. 2011). Utilising publicly available datasets removes the need to physically visit areas of interest in order to gather sample material, greatly expanding the potential scope of any survey and contributes to the benefits of open access research as a concept.

This chapter provides an introductory, proof of concept series of experiments to examine the feasibility and utility of quantifying isoprene degrading bacteria in a range of different environments with both lab-based and bioinformatics-based methods. The goal of this study was to address the scale and abundance of isoprene biodegradation in natural, complex microbial ecosystems. A tentative hypothesis for the qPCR experiments described was that isoprene emission potential would be a driving factor in the relative abundance of isoprene degrading bacteria associated with a given plant species. In this case, the species tested with the highest emission potential was the willow *Salix viminalis* at 80-130 µg g^{-1}(dw) h^{-1}, followed by the poplar *Populus nigra* at 29-76 µg g^{-1}(dw) h^{-1} and finally the *Miscanthus x giganteus* which does not contain an isoprene synthase gene and as such emits no isoprene (Rasmussen, 1978; Hewitt and Street, 1992; Pio, C.A. *et al.,* 1993; Hu *et al.*, 2018).

6.3 Materials & Methods

Sample collection for qPCR analyses was carried out at a commercial farm in Lincolnshire, NE England (53º18’55”N; 0º34’40”W). Each species of willow SRC (*Salix viminalis*), poplar (*Populus nigra*) and *Miscanthus* (*Miscanthus x giganteus*) was grown in a single crop plantation. Images of each sampling location can be seen in Figure 6.1.

When sampling poplar and willow, three trees of each were chosen, representing biological replicates 1-3. From each of those trees, three leaf and three soil samples were taken, with leaves collected from opposite sides of each tree, and soil samples collected from the top 10 cm of soil next to the trunk once debris and leaf matter had been removed. *Miscanthus* fronds and associated soil were collected from six varying locations within the *Miscanthus* growth area. 2 g of leaf material and 500 mg of soil was processed for each sample.

All subsequent materials & methods for work described in this chapter were carried out as described in Chapter 2. This includes leaf washing, nucleic acid extraction and extra purification (necessary for soil samples only), cloning and qPCR. Metagenomic analyses were carried out as described in section 2.10.

Figure 6.1. Sampling sites for qPCR-based *isoA* survey. (a) *Populus nigra* plantation.(b) *Miscanthus x giganteus* crop. (c) *Salix viminalis* crop. Sampling for all three species was carried out on a commercial farm in Lincolnshire, NE England (53º18'55"N; 0º34'40"W).

6.4 Results & Discussion

6.4.1 qPCR analysis of *isoA* abundance associated with a range of plant species.

The first stage of the survey focused on qPCR analyses of samples collected from two isoprene emitting species, *Salix viminalis* and *Populus nigra* and one non-emitting species, *Miscanthus x giganteus.* Both leaf and associated soil samples were tested. Results showed that overall, soil samples harboured far more *isoA* containing bacteria than the corresponding phyllosphere samples, a 6.9-fold difference was seen in poplar samples between soils and leaves, 4.1-fold difference in willow samples and a 4.4-fold difference in *Miscanthus* samples (Figure 6.2). This corresponds with further work carried out by Dr. Ornella Carrión during the validation of the *isoA* primers described in Chapter 3 and published in Carrión *et al.* (2018). The study in question showed that the number of *isoA* genes harboured by bacteria in soil environments was an order of magnitude higher than was found in the phyllosphere. The experiments described here took a more targeted approach with more replicates for statistical robustness and found the same to be true, though the difference between the two environments was smaller in scale.

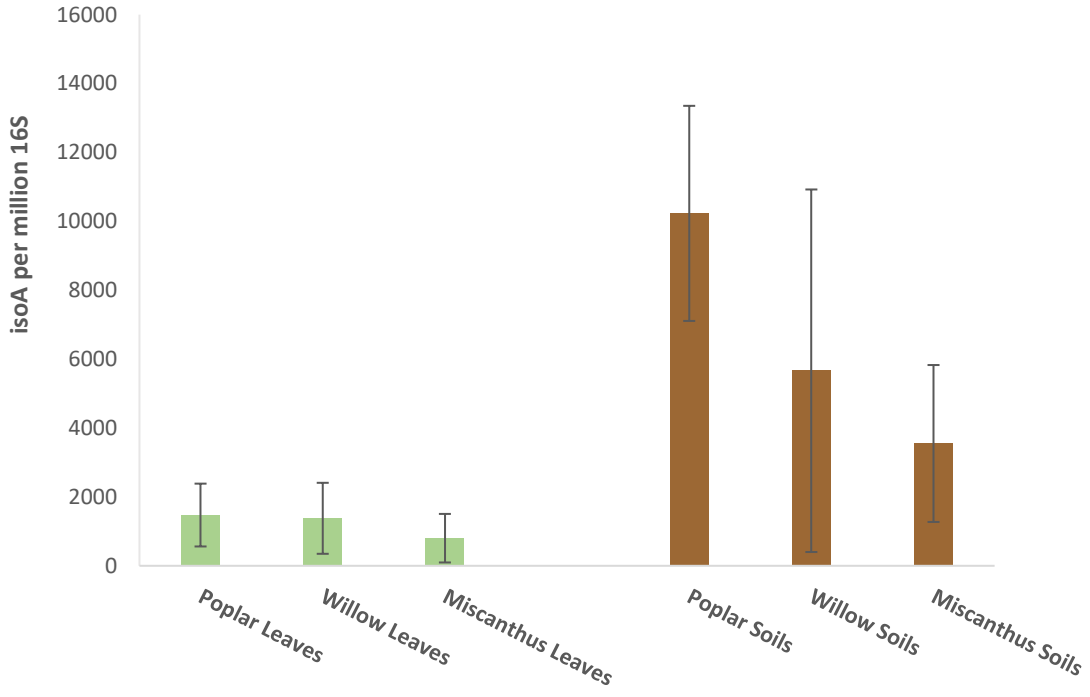

Figure 6.2. Relative abundance of isoA-containing bacteria associated with Salix, Populus and Miscanthus soils and phyllosphere. Results are displayed as *isoA* copy number per million 16S rRNA gene copy numbers per species environment type. Averages of replicates and pseudo-replicates are shown, error bars represent standard deviations.

On first glance (Figure 6.2), there appears to be a pattern in both environment types, with *Miscanthus* samples harbouring a smaller proportion of isoprene degrading bacteria than the two emitting species. The soils associated with willow species appear to harbour a smaller proportion of *isoA* genes than poplar soil samples. This pattern was also seen in phyllosphere samples, though to a much smaller degree.

A few things stand out as counter to the original hypothesis that isoprene emission potential would be a driving factor in the relative abundance of isoprene degrading bacteria associated with a given plant species. Firstly, it is interesting to observe that *Miscanthus x giganteus* species harbour any bacteria containing an *isoA* at all, given the plant produces no isoprene and was located in a distinct area of the farm to any other isoprene emitting tree species (Figure 6.1). Although the lower numbers of *isoA* sequences found associated with *Miscanthus* compared to the high emitting trees tested still lend to emission playing a role in relative abundance of associated isoprene degrading bacteria.

Secondly, we know that *Salix viminalis* has a higher isoprene emission potential than *Populus nigra* (C. Nicholas Hewitt and Street 1992; Hu *et al*. 2018). The fact that soils associated with the willow seem to harbour a smaller proportion of bacteria with an *isoA* than the poplar soils suggest other factors are at play.

The most notable aspect of the data as seen in Figure 6.2 however, is the high intra-sample variability. To get a more granular view of this, individual data points from phyllosphere samples (Figure 6.3) and soil samples (Figure 6.4) are shown.

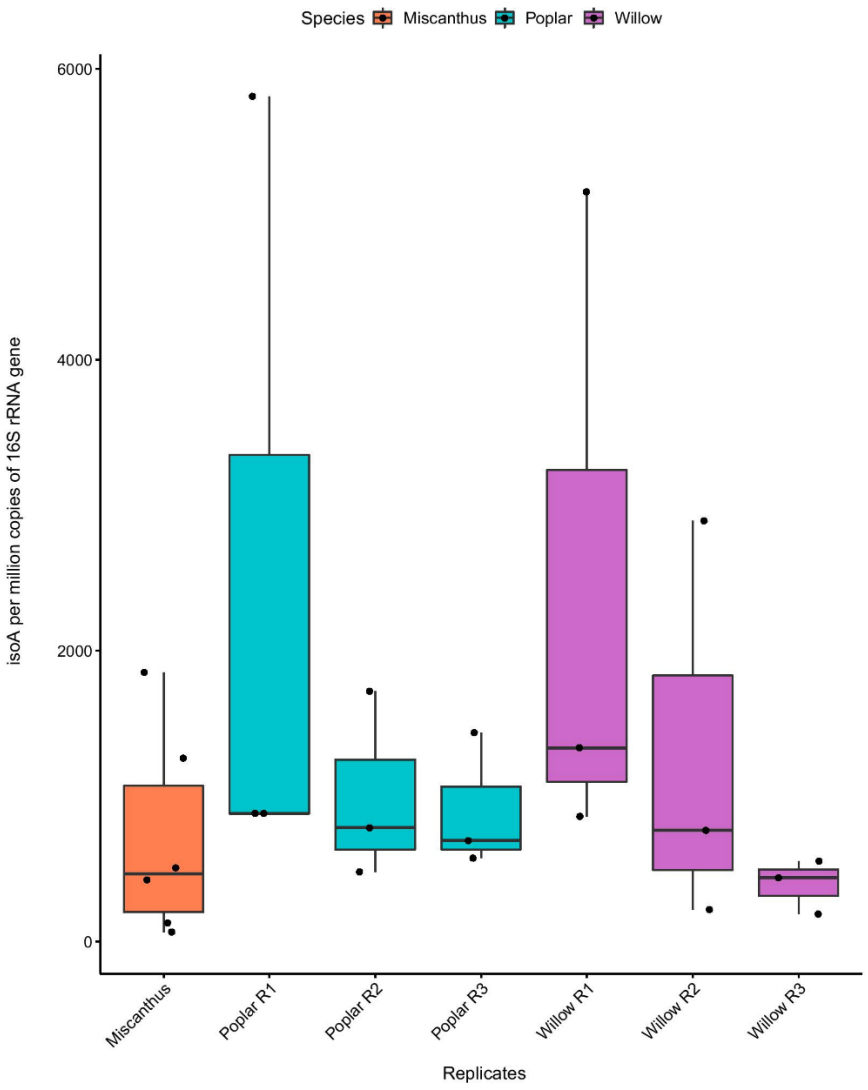

Figure 6.3. Relative abundance of *isoA*-containing bacteria in the phyllosphere of *Salix, Populus* and *Miscanthus* species.

With each individual data point taken into consideration, the apparent patterns suggested earlier become less convincing. In phyllosphere samples, qPCR results showed willow leaves harboured an average of 1,379±1,030 *isoA* sequences per million copies of 16S rRNA genes, indicating an average of ~0.14%. Poplar phyllosphere samples contained an average of 1,473±911 *isoA* sequences per million 16S rRNA genes, equating to approximately ~0.15%. The non-emitting *Miscanthus* phyllosphere contained an average of 801±704 *isoA* sequences per million 16S rRNA genes, representing ~0.07%.

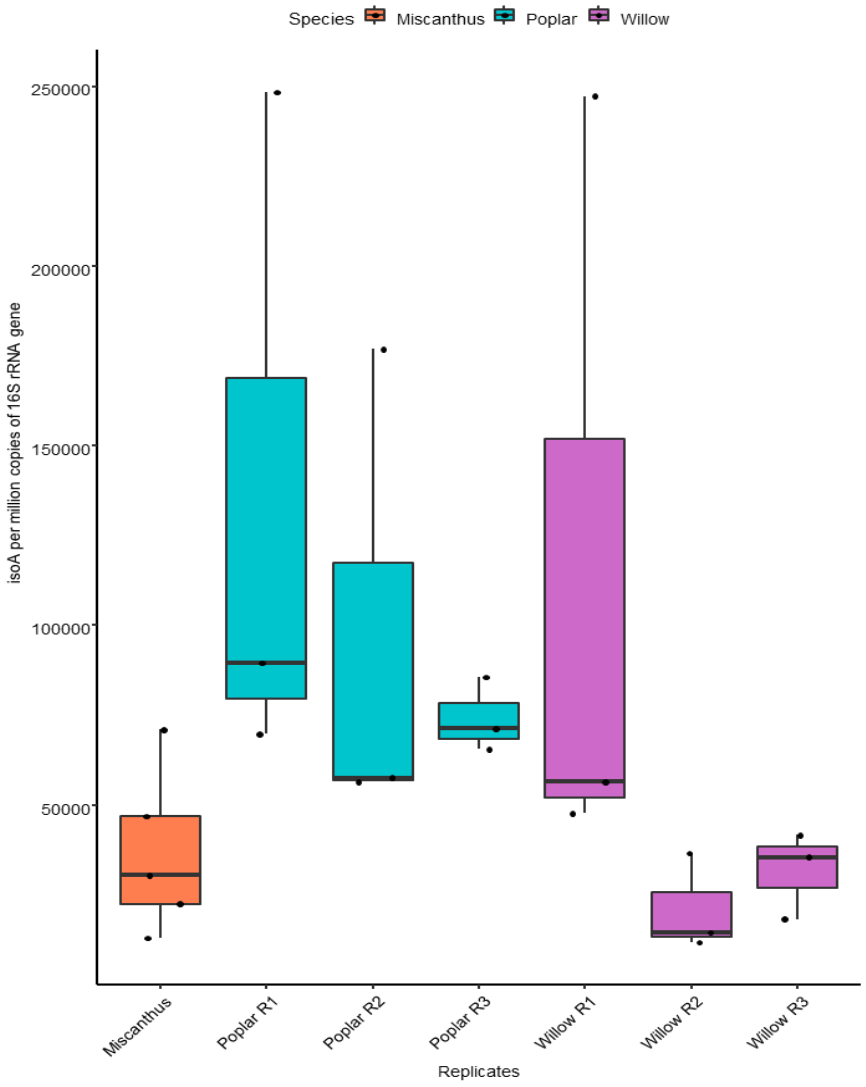

Figure 6.4. . Relative abundance of isoA-containing bacteria in the soils associated with *Salix*, *Populus* and *Miscanthus* species.

In associated soil samples, willow samples harboured an average of 5,661±5,262 *isoA* sequences per million copies of 16S rRNA genes, indicating an average of ~0.57%. Poplar samples contained an average of 10,231±3,122 *isoA* sequences per million 16S rRNA genes, equating to approximately ~1.02%. The non-emitting *Miscanthus* phyllosphere contained an average of 3,551±2,277 *isoA* sequences per million 16S rRNA genes, representing ~0.37%.

However, as can be seen by the wide spread of interquartile ranges in each individual replicate, there is a high level of overlap between each species, in both soil and

phyllosphere samples. When analysed by ANOVA neither environment-type showed a significant difference in *isoA* abundance between any of the three species. *isoA* copy number comparison from phyllosphere sample communities between all species gave a global p-value of 0.59. A comparison of *isoA* copy numbers recovered from soil samples communities gave a global p-value of 0.15.

For this reason, qPCR analyses suggests that while biome may be a strong driving factor in the abundance of isoprene degrading bacteria, emission potential of specific associated plants appears not to be.

6.4.2 Mining for *isoA* in unenriched global metagenomic datasets

Another tool utilised to investigate the abundance of bacteria in the environment with the potential to degrade isoprene was the mining of metagenomic datasets for *isoA* sequences using the techniques described by Curson *et al.* (2018). This involved the creation of Hidden Markov Models, using HMMR tools (Eddy 2011) to identify all possible *isoA* or *recA* sequences in a given metagenome, and then manually curating them to ensure each hit was positive and unique. Details of each metagenome analysed, and associated references, can be found in Table 6.1.

Publicly available metagenomes retrieved from samples from plant species that were closely related to environments that had been tested by qPCR were prioritised to investigate whether metagenomic results would support those results obtained via qPCR. Environments sampled were bulk soils associated with a willow tree and Malaysian oil palm, rhizosphere soils associated with a poplar and with *Miscanthus*, phyllosphere samples from a poplar and oil palm tree and permafrost soil from the Canadian High Arctic.

Table 6.1. Details of metagenomes mined for *isoA* genes. Emissions are measured in μg g^{-1}(dry weight leaves) h^{-1}, except where marked with (*) in which case emissions were measured in ng g^{-1} (dry weight of soil) h^{-1}. Estimated % abundance of *isoA* was normalised to number of *recA* sequences, to metagenome read length, to shortest gene length.

Sample	Species	Biome	Location	Emission	Reference	Estimated % of bacteria with *isoA*
Poplar leaves	*Populus alba*	Phyllosphere	Norfolk, UK	19.94	Crombie *et al.,* 2018	0.02
Poplar soil	*Populus deltoides*	Root rhizosphere	Oregon, USA	32.68-37	Blair *et al.,* 2018	0.90
Willow soil	*Salix fragilis*	Bulk soil	Norfolk, UK	37	Larke-Mejía *et al.,* 2019	0.65
***Miscanthus* soil**	*Miscanthus x giganteus*	Root rhizosphere	Michigan, USA	0	Guo *et al.,* 2015	0.83
Oil Palm Leaves	*Elaeis guineensis*	Phyllosphere	Palong, Malaysia	175	Carrion *et al.,* 2020	0.16
Oil Palm Soil	*Elaeis guineensis*	Bulk soil	Palong, Malaysia	175	Carrion *et al.,* 2020	1.00
C.A. Permafrost	NA	Permafrost	Canadian High Arctic	1.7 ± 0.6*	Chauhan *et al,* 2014	0.64

As found with qPCR results, a link between biome type and abundance of bacteria harbouring an *isoA* gene was seen via metagenomic screening, wherein soil environments, be they bulk or rhizosphere, contained far more isoprene degrading bacteria (containing *isoA*) than phyllosphere environments (Figure 6.5).

Interestingly, while qPCR results suggested that plant species was not a strong factor in predicting abundance of associated isoprene degrading bacteria, some aspects of the pattern between species in qPCR results was also supported by metagenomic analyses. Specifically, the high abundance of isoprene degrading bacteria associated with poplar soils compared to willow soils, counter to the opposing emission factors of both species. In this case however, the non-emitting *Miscanthus* also harboured more *isoA*-containing bacteria than willow. Soils associated with the extremely high-emitting oil palm trees harboured the most *isoA* sequences at a full 1% of the total community, though an abundance of only 0.16% was seen in the phyllosphere (Figure 6.5).

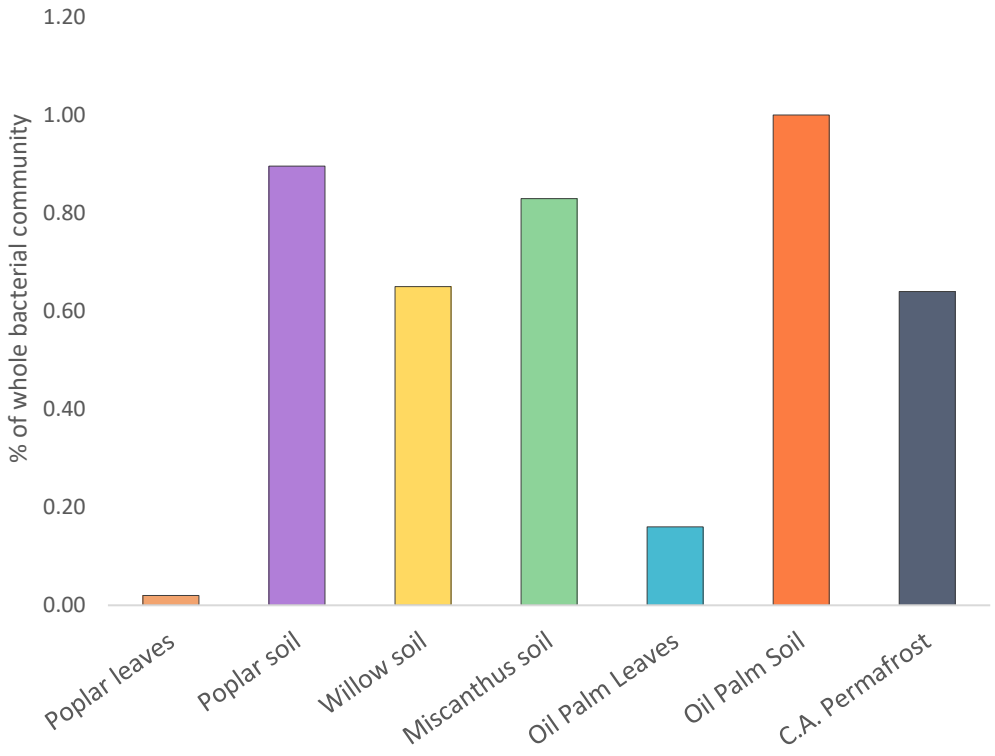

Figure 6.5. Percentage of bacteria found in metagenomic datasets harbouring an *isoA* sequence. Abundance is measured as *isoA* gene number as a percentage of *recA* gene number per dataset, normalised to read number and gene length.

One of the environments examined, specifically to demonstrate the potential to widen the scope of such surveys via public metagenome datasets, was permafrost soil from the Canadian High Arctic (Table 6.1). Interestingly, though no data for isoprene emissions from that exact area the metagenomic data originated from were available, data from warming permafrost in Finnish Lapland have shown isoprene emissions to be extremely low at 1.7±0.6 ng g^{-1} (dry weight of soil) h^{-1} (Li *et al*. 2020). However, the study that the metagenome sample originated from does not specify the temperature at the time of sampling, or if there was any plant matter in the vicinity. In a study by Lindwall *et al.* (2016), isoprene measurements were taken during growing season in a fen in low arctic Greenland, and found that in control plots, isoprene emissions were also low at 1.1±0.17 µg m^{-2} h^{-1}, though this increased by an order of magnitude when a warming treatment that raised the canopy air temperature by 1.9˚C was carried out.

Despite the low level of isoprene emissions expected in such an environment, results showed that the number of bacteria harbouring an *isoA* in the Canadian High Arctic permafrost was 0.64% of the whole community (Table 6.1). This is almost identical to results from soils associated with a willow tree (0.65% of the whole community), though *Salix fragilis* has been shown to emit over thirty times the amount of isoprene.

6.4.3 Gene mining for diverse *isoA*-like sequences

In section 5.4.3 focusing on oil palm samples, two of the recovered MAGs, identified as belong to *Zoogloeaceae* and *Ralstonia*, contained *isoA*-like sequences with translated-amino acid identities of 51.1% and 48.3% respectively to the IsoAs of the closest ratified isoprene degrading isolates (as shown in Table 5.4). These MAGs contained genes with homology to the *isoA-F* genes of the isoprene monooxygenase, but not to the downstream genes *isoG-J.* While these MAGs and their associated *isoA*-like sequences cannot be designated as verified isoprene degrading bacteria without further investigation (discussed in section 5.4.3), they could potentially represent a divergent pathway for isoprene metabolism in the environment. With this in mind, the metagenomic screening process (described in section 2.10) was run a second time on a select number of metagenomic datasets, namely both oil palm-associated metagenomes (soils and leaves), the poplar phyllosphere metagenome and the willow bulk soil metagenome.

When the *isoA*-like sequences from the *Zoogloaceaea* and *Ralstonia* MAGs were included in the HMM used to screen these metagenomes, the proportion of *isoA* sequences

recovered from each was doubled. The proportion of *isoA* sequences per *recA* sequences in the oil palm soil community was 1.99%, in the oil palm phyllosphere community was 0.23%, 0.05% in the poplar phyllosphere community and 1.29% in the willow soil community. If the function of these genes could be verified, either by isolation of representative strains or heterologous expression of their *iso* cluster as previously described by Crombie *et al.* (2018), it would suggest our current database of *isoA* sequences accounts for less than half of the diversity actually present in the environment.

6.4.4 Limitations of survey methods

There are obvious limitations of this kind of study that prevent direct comparison of *isoA* abundance between environments. As shown above, much of the associated literature for publicly available metagenomes give little information on abiotic factors that may significantly affect isoprene emissions from the plant or environment being investigated. As discussed in section 1.2.2 factors such as season, temperature, age of the plant, recent weather and stressors such as drought or atmospheric chemistry could all play a role in regulating isoprene emission levels. How important these factors are to the question of *isoA* abundance begs the question how quick the turnover of isoprene degrading bacteria is in a given environment, and whether their abundance is significantly impacted by short term stimuli.

Another factor that could impact the abundance of isoprene degrading bacteria is the vicinity of other isoprene emitting plant species to the one being tested. It is not always possible to confirm whether the plant in question is isolated enough that other species would not share a significant proportion of their associated microbiome, particularly in soil-based environments.

On a more technical front, the quality of the metagenome data itself can also play a disruptive role in apparent diversity and gene abundance. It is not uncommon for older metagenomic datasets to suffer from a shallow sequencing depth, which risks aspects of the community diversity or the true scale of gene abundance being biased or masked. In the same vein, qPCR analyses suffer from their own technical drawbacks. Specifically the reliance on *isoA* primers risks the failure of more diverse *isoA* sequences being amplified if their host organisms or a close relative is not well represented in our current database.

Another consideration for studies like this is that while gene copy number is a useful metric for abundance of a given pathway in an environment, it does not truly address how important that metabolic process is in the overall community. To accurately investigate the scale of isoprene metabolism in the environment, wide-scale metatranscriptomic studies would be required to detail how active the *isoA* genes are in a given sample, compared with genes of other metabolic pathways of interest. Though even this method can be biased, as transcription is complex and does not always lead to gene activity due to post-translational effects. With this in mind, a combination of metagenomic, metatranscriptomic and metaproteomic sequencing data would give the most detailed view of isoprene metabolism in a natural environment.

6.5 Conclusion

The combination of qPCR and metagenomic methods to investigate the abundance of *isoA* genes in the natural environment has revealed that isoprene metabolism is a surprisingly common trait amongst bacterial in the environments tested. Results contradicted the initial hypothesis that the isoprene emission potential of a given plant species would dictate the scale of isoprene degrading bacteria associated with it's leaves and soils. In fact, the variability of isoprene degrading bacteria distribution associated with a single tree suggests the driver for such things may be far more localised. Factors such as sunlight, wind or rain exposure may play a more significant role in effecting the presence, absence or abundance of isoprene degraders on a given leaf or area of soil. To examine the impacts of these abiotic factors, a survey of a much broader scale, under much more controlled conditions would be necessary (as described later in Chapter 7: Conclusions and Future Considerations).

The only conclusive factor revealed by analysis to date is that soil environments as a whole appear to be a preferential environment for isoprene degrading bacteria. Again, considering the higher concentration of isoprene found in the canopy region compared to ground level (R Fall and Monson 1992; Brüggemann and Schnitzler 2002b), suggests local isoprene concentration is not a significant driver of abundance for isoprene degrading bacteria. The higher abundance in soils in that case is perhaps unsurprising, isoprene degradation has been shown to be an inducible process (Crombie *et al*. 2015; Robin A. Dawson *et al*. 2020; Murrell, McGenity, and Crombie 2020) and soil environments, known to harbour the most diverse bacterial communities of any biome, are rich in alternative

potential carbon sources that could sustain a population of bacteria otherwise capable of degrading isoprene. For this reason especially, a multi-omic approach would be invaluable in teasing apart the action of the isoprene monooxygenase in a complex community under natural conditions.

7 Conclusion and Future Considerations

7.1 Aims of this study

Isoprene is an abundant and complex climate-active gas, emitted in massive amounts to our atmosphere by a wide range of biological sources, particularly terrestrial plants. Isoprene is thought to play a protective role against thermal and oxidative stress and is responsible for 1-2% of the total photosynthetically-sourced energy output in the plants that emit it (Fall and Monson 1992; Sharkey and Yeh 2001). Though isoprene is the most abundantly produced global BVOC, many aspects of the isoprene biogeochemical cycle are still to be fully elucidated. Specifically, the biodegradation of isoprene is considerably underexplored compared to other VOCs emitted to a similar scale. The purpose of this study was to contribute to our understanding of the ecology of isoprene degrading bacteria and help to fill in the following gaps in our knowledge related to the global isoprene cycle.

Question 1. Identity of isoprene degrading bacteria.

One aim of this project was to build on our existing knowledge as to what specific clades of bacteria are involved in the biodegradation of isoprene and examine how diverse this group of organisms is in the environment. Earlier research suggested that isoprene metabolism was found mostly in Gram-positive Actinobacteria, with more recent studies indicating a small number of Gram-negative bacteria may also contribute to the process (Reviewed by McGenity *et al.* 2018 and Carrión *et al.*, 2020). One of the key aspects of this study was to broaden this database of known or putative isoprene degrading bacteria.

Question 2. Distribution of isoprene degrading bacteria.

Isoprene biosynthesis is a widespread metabolic process, shared by plants, animals, algae, humans, bacteria and more. This project aimed to contribute to the known locations that harbour isoprene degrading bacteria and address whether different host species or biomes, particularly those producing large amounts of isoprene, impact the diversity of the isoprene degrading bacteria associated with them.

Question 3. Abundance of isoprene degrading bacteria.

One aspect of the ecology of isoprene degradation that has only been lightly explored in the past is how abundant the process is in the natural environment. Most previous work has focused on locating and identifying these bacteria, but in doing so relied on enrichment-based assays that cannot be correlated to the natural community. The final question this project set out to investigate was how abundant isoprene degrading bacteria

are in-situ, and what factors might impact their abundance. In doing so, this project aimed to begin answering the question – to what extent do microbes in the environment contribute to the global isoprene cycle.

7.2 Major findings of this thesis

7.2.1 Novel *isoA* gene probes reveal the presence and diversity of *isoA* genes in a range of environments.

The *isoA* gene encoding the α-subunit of the isoprene monooxygenase, an SDIMO responsible for catalysing the initial oxidation of isoprene to 1,2-epoxyisoprene, was identified as an ideal marker gene for isoprene biodegradation in the environment. Novel gene probes were designed to target the *isoA* via alignments of *isoA* genes recovered from the most up-to-date database of verified isoprene degrading isolates available. In Chapter 3, the primers isoA14F and isoA511R were validated and tested to examine the presence and diversity of *isoA* genes in a range of different environments.

The presence of *isoA*-containing bacteria was confirmed in each of the location- and biome-types tested. This included such environments as ash leaves, poplar leaves, willow leaves, leaves from a natively grown Malaysian oil palm, from an oil palm grown in Kew Gardens, London (UK), soils associated with a willow tree, soil from a tyre dump, coastal sediment, salt marsh sediment and freshwater sediment. The use of primers isoA14F and isoA511R to create clone libraries resulted in 100% true positives for *isoA* and gave an indication of the variety of *isoA* genes and how they are distributed across differing environment types. For example, while many of the recovered *isoA* genes shared high homology with the *isoA* sequences recovered from traditional isoprene degraders such as *Rhodococcus* and *Gordonia*, this was not always the case. Results from sampling locations such as the soil associated with a willow tree, and most of those recovered from tyre dump soil samples, were most closely related to *isoA* sequences from Gram-negative Comamonadaceae. Considering willow is a high emitter of isoprene, this indicates that the prevalence of Gram-positive isoprene degraders may not only be a result of higher isoprene concentration as discussed in section 1.3.2, but could be location- or host-dependent also.

isoA amplicon sequencing confirmed this distribution of *isoA* genes that were homologous to those recovered from Gram-negative degraders, with one notable addition in the results from oil palm leaf samples taken from Kew Gardens. Here, *isoA* amplicon sequencing results showed that 60% of the recovered sequences were homologous to the *isoA* from *Sphingopyxis,* another Gram-negative genus, from the Sphingomonadaceae family.

7.2.2 DNA-SIP experiments reveal a number of novel isoprene degrading bacteria and reveal distinct communities associated with related phyllosphere and soil environments.

DNA-SIP experiments utilising C^{13}-labelled isoprene were carried out in the phyllosphere environment of a willow tree, and in the phyllosphere and associated soils of Malaysian oil palm trees from Palong, Malaysia. Through a combination of 16S rRNA gene sequencing and metagenomics, a number of genera not previously seen in isoprene-related studies were identified as having incorporated isotopically labelled isoprene into their DNA via growth and DNA replication. Of particular note in the willow SIP experiment described in Chapter 4 was the *Methylobacterium* with an RA of 26.48% in the heavy fraction of one of the replicates incubated with C^{13}isoprene, and this was seen again in the metagenomic sequencing results. A high-quality MAG identified as *Methylobacterium* was recovered, though no known isoprene metabolism genes were identified, suggesting either the presence of an unknown pathway, missing genomic regions in the MAG, or significant cross-feeding at work. The isolation of a *Methylobacterium* isolate from an isoprene enrichment would be of great interest to confirm whether this genus is indeed a novel isoprene degrader and first confirmed representative of a methylotroph capable of isoprene metabolism. Previous work by Murphy (2017) initially reported growth of a *Methylobacterium* strain using low concentrations of isoprene, however further examination of the strain failed to replicate successful growth with isoprene as the only source of carbon. This phenomenon has occurred more than once, with a *Methylobacterium* isolate showing initial signs of isoprene metabolism that were difficult to replicate at a later date. It is possible that there are strains of *Methylobacterium* that can degrade isoprene, but that the trait is an unstable one. Perhaps these earlier isolates, and indeed the *Methylobacterium* MAG described in Chapter 5, once contained an isoprene degradation gene cluster, but the isoprene degradation trait was plasmid-mediated, unstable, and was eventually lost.

A number of novel isoprene degrading bacteria were identified when investigating Malaysian oil palm soils and leaves via DNA-SIP described in Chapter 5. *Novosphingobium*, *Pelomonas*, *Rhodoblastus* and *Zoogloea* were all identified for the first time as having the ability to metabolise isoprene. Results of this study also made it clear that though the isoprene in this environment originates from the same source, the isoprene degrading community in the phyllosphere and soils associated with oil palms were completely distinct. This may be a consequence of the different availability of isoprene at the leaf

surface level where it is much higher, and the plantation floor level where isoprene levels would be significantly lower. Other abiotic factors related to the biomes themselves, such as sun, wind and rain exposure or soil carbon and nitrogen content may also play a significant role in shaping the isoprene degrading community, leading to the next significant finding.

7.2.3 Host isoprene emission potential may not be a driving factor influencing the abundance of associated isoprene degrading bacteria.

In Chapter 6, a wide range of environments were investigated to examine how abundant isoprene degrading bacteria are in natural ecosystems, to gain a better understanding at to the scale of their contribution to the global isoprene cycle. The hypothesis of this set of experiments was that isoprene emission potential would be a driving factor in the relative abundance of isoprene degrading bacteria associated with a given plant species. The survey utilised both qPCR and metagenomic methods. Results showed that not only are isoprene degrading communities in soil environments more diverse as shown in the DNA-SIP experiments, but they are also consistently more abundant. Of the soil samples tested by qPCR, *isoA* copy numbers made up 0.37% - 1.02% of the 16S rRNA gene copy numbers, with phyllosphere samples containing 0.07% - 0.15%. The difference between the number of isoprene degraders harboured by soils compared to phyllosphere samples was even more pronounced in the metagenomic data, with differences between communities associated with the same host species differing by an order of magnitude.

The first of the main findings in this chapter was that *isoA* copy number assayed by qPCR varied drastically between samples taken from the same tree. The second finding was that there was no significant difference between the proportion of *isoA* copy numbers from DNA recovered from the leaves and soils of a non-emitting *Miscanthus* species, and two high-isoprene emitting species, willow and poplar. This lack of a correlation between emission potential and abundance of *isoA* genes in the associated community was mirrored in metagenomic results, with *Miscanthus*-sourced communities harbouring a higher proportion of *isoA* genes than willow, despite the fact *Miscanthus x giganteus* contains no isoprene synthase. These two findings combined suggests that the host species, and thus its isoprene emission potential, might be a minor factor in determining the abundance of isoprene degrading bacteria, and could be completely irrelevant altogether.

On closer look at the individual data points from each tree in Figure 6.3 and Figure 6.4 (represented by R1-3 of poplar and willow samples), it is interesting that in most cases, only one datapoint is responsible for dragging the upper quartile range far beyond that of the other two points collected from each tree. Sampling of each tree was done in a triangulate manner. A potential hypothesis here is that the single datapoint in the upper quartile range of each tree may represent that side of the tree that received most sun exposure. While a targeted and more detailed study would be required to confirm this effect, as it is, the data suggest that other abiotic factors are likely having more of an influence on the abundance of isoprene degrading bacteria than the species of plant they are associated with.

7.2.4 Sequences recovered from MAGs reveal novel arrangement of isoprene degradation gene cluster and suggest *isoA* sequences may be more diverse than current isolate-driven data suggests.

The DNA-SIP experiment involving willow leaves in Chapter 65 involved the recovery of MAGs from metagenomic sequencing data. One MAG of interest was identified as a *Mycobacterium*. While *Mycobacterium* have been identified in previous isoprene-focused DNA-SIP experiments (Johnston *et al.*, 2017), this was the first case of *Mycobacterium* shown to metabolise isoprene in a terrestrial environment. The MAG in question was shown to contain two full, non-identical isoprene degradation gene clusters, including the IsoMO (*isoABCDEF*), the downstream genes (*isoGHIJ*) and the associated aldehyde dehydrogenase *aldH2*, and each also containing a putative transcriptional regulator found in all Gram-positive isoprene degrading isolates to date. While duplicate downstream genes have been seen in other isoprene degrading bacteria (Crombie *et al.*, 2015; Johnston *et al.*, 2017; Gibson *et al.*, 2020), this is the first case of any genes encoding the IsoMO being present in duplicate in a single genome. While every attempt was taken to ensure this was not a result of miss-assembly or contamination, it is always possible that is still the case. A physical isolate and subsequent whole genome sequencing would be necessary to confirm the existence of a duplicate isoprene degradation gene cluster. However, if correct, it could be the result of an interesting quirk of horizontal gene transfer, and subsequent expression analyses would be of great interest.

Another notable finding from MAGs recovered from the phyllosphere environment of Malaysian oil palm trees in Chapter 5, was the presence of gene clusters similar to, but of relatively low homology to the genes encoding the isoprene monooxygenase. The two

MAGs that harboured these clusters were identified as *Zoogloaceae* and *Ralstonia*, and their *isoA*-like genes shared a translated amino acid identity of 51.1% and 48.3% respectively to the IsoA of *Variovorax* WS11. The lack of genes similar to *isoGHIJ* is interesting, and considering we are still unsure as to the exact role these downstream genes play in isoprene degradation, could suggest another pathway for isoprene metabolism after the action of the IsoMO. The presence of such a similar cluster in MAGs recovered from heavy DNA of cultures incubated with C^{13}isoprene certainly warrants further investigation.

7.3 Recommendations for future investigations

7.3.1 Validation of gene clusters recovered from *Zoogloaceae* and *Ralstonia*

The novel gene clusters mentioned in section 7.2.4 cannot be confirmed to play a role in isoprene degradation from metagenomic data alone. To fully investigate their potential role in isoprene metabolism, one of two methods should be followed. The first is a targeted isolation approach similar to that carried out by Larke-Mejia *et al.* (2019). Namely, returning to the sample site utilised for the oil palm DNA-SIP experiment described in Chapter 5, collecting leaf samples and using leaf washings to prepare isoprene enrichment cultures. Following the isolation method described in section 2.2.2, attempts should be made to isolate representatives of these isoprene degrading *Zoogloaceae* and *Ralstonia* for whole genome sequencing and further investigation. However, this method is time consuming, expensive when considering the sample site is based in Malaysia, and success is far from assured as we do not know the precise cultivation requirements of these bacteria.

Another route is the heterologous expression of the gene clusters. Considering both are Gram-negative, *Variovorax* WS11 would be a good candidate host for such an experiment (Dawson 2021). The gene cluster inserts could be synthetically produced, or primers designed with sequence data recovered from the metagenomes the MAGs originated from could be utilised to amplify the clusters from the DNA sequenced to produce the metagenomic datasets in the original study. If isoprene degradation could be restored by insertion of these novel gene clusters to a *Variovorax* WS11 mutant with its own IsoMO knocked out, the validity of these bacteria as novel isoprene degraders could be verified. At this point, further analysis to determine the downstream process in bacteria that do not contain *isoGHIJ* would be of great interest.

7.3.2 A future study design for a focused survey to determine the factors influencing abundance of isoprene degrading bacteria in the natural environment.

The experiments described in Chapter 6 have made it obvious that the distribution and abundance of isoprene degrading bacteria in the environment are not solely dependent on their host plant species. It would appear that factors such as sun, rain and wind exposure, or soil carbon content could be more important driving factors for isoprene degradation in the environment. Abiotic factors such as these can lead to stressors such as desiccation and nutrient limitation, impacting the profile of any microbial community. For this reason, a long-term experiment that artificially manipulates abiotic factors such as these and subsequently quantifies the resident isoprene degrading bacteria is recommended.

Although the second half of the experiments in Chapter 6 utilised publicly available metagenomes in an effort to widen the geographical scope of the survey, this method is less viable when specific and changeable abiotic factors become the main focus of the study. The metadata associated with publicly available metagenomes varies wildly in quality and depth, and one cannot assume that a valuable information available for one dataset will be available for direct comparison in the next. For this reason, it is suggested that the next survey maintains the metagenomic approach but instead relies on manually collated metagenome data to ensure consistent access to variable data.

The reason a metagenome approach is recommended over methods such as qPCR is that it does not rely on the sensitivity of a primer set that may be limited by the data available at the time of their creation. The use of such primers is indeed useful in a variety of use-cases, but risk a low assessment of abundance in cases where diverse *isoA* sequences that have not yet been encountered are present in environments of interest. In terms of a metagenomic study that utilises HMM-profiles, while results are still dependent on the sequences used to create the HMM-profile, the E-value can be adjusted to allow for more diverse matches that can then be manually analysed. Metagenome analyses also do not depend on the use of universal primers for house-keeping genes to carry out normalisation, for example single-copy genes such as *recA* and *rpoB* can be utilised, giving a more accurate estimate of true abundance.

For the initial manipulation experiment, a plantation of an isoprene-emitting species that is relatively small in size would be preferable. Species like the Spanish Reed *Arundo donax* L. have been shown to emit large amounts of isoprene and would be a good candidate

(Hewitt *et al.,* 1990). Ideally the crop would be grown separately to other species so that any impact on the associated microbiome can be discounted. A number of plots should be designated to provide a suitable number of replicates for each individual treatment to ensure robust statistical analyses are possible after data have been collected. Treatment factors should include (at the least) warming, precipitation and wind exposure.

In terms of warming, infrared heaters should be suspended above each plot to simulate an increase in ambient air temperature, with soil temperature monitored in a regular and automated manner, ensuring the temperature difference between experimental and control plots remains within 2˚C of each other. Extra control plots should contain an inactive suspended heater to account for any impact of shade on community composition.

To examine the impacts of precipitation, increased and decreased precipitation plots should be adjacent, but walled off with sheets of solid material to prevent water run-off between plots. Plastic open-topped funnelling should cover 50% of the decreased precipitation plots, with the excess funnelled into the increased precipitation plots.

The effects of wind exposure should be measured with the use of fans to the side of increased exposure plots, and transparent wind-proof barriers around the sides of decreased wind exposure plots.

Soil and frond samples should be collected up to a year after the experiment was set up, and metagenomic sequencing carried out on the DNA extracted. These metagenomic datasets should be mined for *isoA* genes and the abundance compared between each of the various treatments in a statistically robust manner. The availability of such a sample set could be utilised in enrichment studies to investigate changes in diversity of isoprene degrading bacteria also. It could be the case that the identity of associated isoprene degrading bacteria may also be tied to abiotic factors rather than simply the associated host species. A factor that would not be overly apparent in most DNA-DIP experiments to date.

While such a study is much smaller in scale in terms of geographical scope and species investigated than a broad scale metagenomic survey of publicly available metagenomes, it is one of the few ways to categorically determine specific factors influencing the abundance of isoprene degrading bacteria in a given environment. Such a study design would have the added benefit of providing robust data as to the reaction of isoprene

degrading communities to stressors associated with climate change, helping us to better understand their future role in the biogeochemical cycle of isoprene in a warming climate.

A strong caveat here is that it should be acknowledged this study design exists in an ideal world where access to funds and experimental materials is not an issue. With this in mind, the inclusion of both of transcriptomics and/or proteomics would be highly informative and provide a whole other level of informative data. While metagenomic studies are extremely useful in identifying and even quantifying a target group in a given microbial community, it gives no indication as to specific activity. Transcriptomic or proteomic analyses on the other hand could provide invaluable information as to which of the bacterial groups present is the most active in terms of the degradation of localised isoprene. Under the study design described here, it would also show whether that structure of activity in the community remains the same under different abiotic stressors.

7.4 Gaps remaining in the field

Upon completion of this PhD project, there are a number of interesting questions still to be answered with regards to the molecular ecology of isoprene degrading bacteria.

The known phylogenetic breadth of isoprene degrading bacteria grows more diverse with every environment investigated and there are surely many other taxonomic clades involved in the process we have yet to encounter. Contributions from fungi and archaea have yet to be investigated to any extent and could represent an exciting new chapter in the field should they be shown to possess the ability to degrade isoprene, potentially harbouring completely new pathways for its metabolism.

Another area of interest relates to apparent difference in diversity found in samples incubated with high and low levels of isoprene. This suggests a potential niche differentiation between low- and high-affinity isoprene degrading bacteria. Such differentiation has previously been recorded in methanotrophic bacteria (Dunfield *et al*. 1999; A J Holmes *et al*. 1999; Henckel *et al*. 2000). Investigation of low-affinity isoprene degrading microbes that are potentially being missed by current cultivation techniques would contribute greatly to our understanding of the role of microbes in the isoprene cycle.

Finally, the activity of isoprene degrading bacteria in the environment has been little explored. Utilising metatranscriptomic and metaproteomic approaches to investigate how active microbial metabolism of isoprene is *in-situ* is an area that warrants significant further

investigation. These studies could shed further light on how the process is regulated under environmental conditions and add to our understanding of their significance in the biogeochemical cycle of isoprene.

8 References

Ali Shah, Aamer, Fariha Hasan, Ziaullah Shah, Nida Kanwal, and Samia Zeb. 2013. "Biodegradation of Natural and Synthetic Rubbers: A Review." *International Biodeterioration and Biodegradation*. Elsevier.

Alneberg, Johannes, Brynjar Smári Bjarnason, Ino de Bruijn, Melanie Schirmer, Joshua Quick, Umer Z Ijaz, Leo Lahti, Nicholas J Loman, Anders F Andersson, and Christopher Quince. 2014. "Binning Metagenomic Contigs by Coverage and Composition." *Nature Methods* 11 (11): 1144–46.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10.

Alvarez, Hector M., Frank Mayer, Dirk Fabritius, and A. Steinbüchel. 1996. "Formation of Intracytoplasmic Lipid Inclusions by *Rhodococcus Opacus* Strain PD630." *Archives of Microbiology 1996 165:6* 165 (6): 377–86.

Alvarez, Laura Acuña, Daniel A. Exton, Kenneth N. Timmis, David J. Suggett, and Terry J. McGenity. 2009. "Characterization of Marine Isoprene-Degrading Communities." *Environmental Microbiology* 11 (12): 3280–91.

Anstead, Lenka, and Rosalind R. Boar. 2010. "Willow Spiling: Review of Streambank Stabilisation Projects in the UK." *Freshwater Reviews* 3 (1): 33–47.

Antipov, Dmitry, Mikhail Raiko, Alla Lapidus, and Pavel A. Pevzner. 2019. "Plasmid Detection and Assembly in Genomic and Metagenomic Data Sets." *Genome Research* 29 (6): 961–68.

Atkinson, Roger, and Janet Arey. 2003. "Gas-Phase Tropospheric Chemistry of Biogenic Volatile Organic Compounds: A Review." *Atmospheric Environment* 37 (January): 197–219.

Azevedo-Silva, D., J. A. Rasmussen, M. Carneiro, M. T.P. Gilbert, and H. Azevedo. 2021. "Feasibility of Applying Shotgun Metagenomic Analyses to Grapevine Leaf, Rhizosphere and Soil Microbiome Characterisation." *Australian Journal of Grape and Wine Research* 27 (4): 519–26.

Bäck, Jaana, Hermanni Aaltonen, Heidi Hellén, Maija K. Kajos, Johanna Patokoski, Risto Taipale, Jukka Pumpanen, and Jussi Heinonsalo. 2010. "Variable Emissions of Microbial Volatile Organic Compounds (MVOCs) from Root-Associated Fungi Isolated from Scots Pine." *Atmospheric Environment* 44 (30): 3651–59.

Bai, Jianhui, Bradly Baker, Baosheng Liang, James Greenberg, and Alex Guenther. 2006. "Isoprene and Monoterpene Emissions from an Inner Mongolia Grassland." *Atmospheric Environment* 40 (30): 5753–58.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, *et al*. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology* 19 (5): 455–77.

Bao, Lijun, Likun Gu, Bo Sun, Wenyang Cai, Shiwei Zhang, Guoqiang Zhuang, Zhihui Bai, and Xuliang Zhuang. 2020. "Seasonal Variation of Epiphytic Bacteria in the Phyllosphere of *Gingko Biloba, Pinus Bungeana* and *Sabina Chinensis*." *FEMS Microbiology Ecology* 96 (3).

Behnke, Katja, Barbara Ehlting, Markus Teuber, Martina Bauerfeind, Sandrine Louis, Robert Hänsch, Andrea Polle, Jörg Bohlmann, and Jörg-Peter Schnitzler. 2007. “Transgenic, Non-Isoprene Emitting Poplars Don’t like It Hot.” *The Plant Journal* 51 (3): 485–99.

Birke, Jakob, and Dieter Jendrossek. 2014. “Rubber Oxygenase and Latex Clearing Protein Cleave Rubber to Different Products and Use Different Cleavage Mechanisms.” *Applied and Environmental Microbiology* 80 (16): 5012–20.

Bittar, Thais B., Preston Pound, Ansley Whitetree, L. Dean Moore, and John T. Van Stan. 2018. “Estimation of Throughfall and Stemflow Bacterial Flux in a Subtropical Oak-Cedar Forest.” *Geophysical Research Letters* 45 (3): 1410–18.

Bode, H. B., A. Zeeck, K. Pluckhahn, and D. Jendrossek. 2000. “Physiological and Chemical Investigations into Microbial Degradation of Synthetic Poly(Cis-1,4-Isoprene).” *Applied and Environmental Microbiology* 66 (9): 3680–85.

Bodenhausen, Natacha, Matthew W. Horton, and Joy Bergelson. 2013. “Bacterial Communities Associated with the Leaves and the Roots of *Arabidopsis Thaliana*.” *PLoS ONE* 8 (2): e56329.

Bolger Anthony M., Lohse Marc, and Usadel Bjoern. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30: 2114. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4103590/.

Brenner, Don J. 1984. “Family I *Enterobacteriaceae*.” In *Bergey’s Manual of Systemic Bacteriology*, edited by Noel R. Krieg and John G. Holt, 1st., 408–20. Baltimore, Maryland, USA: Wiliams and Wilkins.

Brüggemann, N., and J.-P. Schnitzler. 2002a. “Comparison of Isoprene Emission, Intercellular Isoprene Concentration and Photosynthetic Performance in Water-Limited Oak (*Quercus Pubescens* Willd. and *Quercus Robur* L.) Saplings.” *Plant Biology* 4 (4): 456–63.

Brüggemann, N., and Jörg P. Schnitzler. 2002b. “Comparison of Isoprene Emission, Intercellular Isoprene Concentration and Photosynthetic Performance in Water-Limited Oak (Quercus Pubescens Willd. and Quercus Robur L.) Saplings.” *Plant Biology* 4 (4): 456–63.

Bu’lock, John D. 1973. “Comparative and Functional Aspects of the Isoprene Pathway in Fungi.” *Pure and Applied Chemistry* 34 (3–4): 435–62.

Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. 2016. “DADA2: High-Resolution Sample Inference from Illumina Amplicon Data.” *Nature Methods* 13 (7): 581–83.

Carrión Ornella, Lisa Gibson, Dafydd M.O. Elias, Niall P. McNamara, Theo A. Van Alen, Huub J.M. Op Den Camp, Christina Vimala Supramaniam, Terry J. McGenity, and J. Colin Murrell. 2020. “Diversity of Isoprene-Degrading Bacteria in Phyllosphere and Soil Communities from a High Isoprene-Emitting Environment: A Malaysian Oil Palm Plantation.” *Microbiome* 8 (1).

Carrión, Ornella, Nasmille L. Larke-Mejía, Lisa Gibson, Muhammad Farhan Ul Haque, Javier Ramiro-García, Terry J. McGenity, and J. Colin Murrell. 2018. “Gene Probing Reveals the Widespread Distribution, Diversity and Abundance of Isoprene-Degrading Bacteria in the Environment.” *Microbiome* 6 (1): 219.

Carrión, Ornella, Terry J. McGenity, and J. Colin Murrell. 2020. “Molecular Ecology of Isoprene-Degrading Bacteria.” *Microorganisms* 8 (7): 967.

Carslaw, K S, O Boucher, D V Spracklen, G W Mann, J G L Rae, S Woodward, and M Kulmala. 2009. “Atmospheric Aerosols in the Earth System Atmospheric Aerosols in the Earth System: A Review of Interactions and Feedbacks Atmospheric Aerosols in the Earth System.” *Chem. Phys. Discuss* 9: 11087–183.

Carter, Claire, Willa Finley, James Fry, David Jackson, and Lynn Willis. 2007. “Palm Oil Markets and Future Supply.” *European Journal of Lipid Science and Technology* 109 (4): 307–14.

Carver, T., S. R. Harris, M. Berriman, J. Parkhill, and J. A. McQuillan. 2012. “Artemis: An Integrated Platform for Visualization and Analysis of High-Throughput Sequence-Based Experimental Data.” *Bioinformatics* 28 (4): 464–69.

Chen, Yin, and J. Colin Murrell. 2010. “When Metagenomics Meets Stable-Isotope Probing: Progress and Perspectives.” *Trends in Microbiology* 18 (4): 157–63.

Cheng, Yuqi, Le Yu, Yidi Xu, Hui Lu, Arthur P. Cracknell, Kasturi Kanniah, and Peng Gong. 2019. “Mapping Oil Palm Plantation Expansion in Malaysia over the Past Decade (2007–2016) Using ALOS-1/2 PALSAR-1/2 Data.” *International Journal of Remote Sensing* 40 (19): 7389–7408.

Claeys, Magda, Bim Graham, Gyorgy Vas, Wu Wang, Reinhilde Vermeylen, Vlada Pashynska, Jan Cafmeyer, *et al*. 2004. “Formation of Secondary Organic Aerosols Through Photooxidation of Isoprene.” *Science* 303 (5661): 1173–76.

Cleveland, C C, and J B Yavitt. 1998. “Microbial Consumption of Atmospheric Isoprene in a Temperate Forest Soil.” *Applied and Environmental Microbiology* 64 (1): 172–77. http://www.ncbi.nlm.nih.gov/pubmed/16349477.

Cleveland, Cory C., and Joseph B. Yavitt. 1997. “Consumption of Atmospheric Isoprene in Soil.” *Geophysical Research Letters* 24 (19): 2379–82.

Corrêa, Felipe Borim, João Pedro Saraiva, Peter F Stadler, and Ulisses Nunes da Rocha. 2020. “TerrestrialMetagenomeDB: A Public Repository of Curated and Standardized Metadata for Terrestrial Metagenomes.” *Nucleic Acids Research* 48 (D1): D626–32.

Crombie, Andrew T., Myriam El Khawand, Virgil A. Rhodius, Kevin A. Fengler, Michael C. Miller, Gregg M. Whited, Terry J. Mcgenity, and J. Colin Murrell. 2015. “Regulation of Plasmid-Encoded Isoprene Metabolism in *Rhodococcus*, a Representative of an Important Link in the Global Isoprene Cycle.” *Environmental Microbiology* 17 (9): 3314–29.

Crombie, Andrew T., Nasmille L. Larke-Mejia, Helen Emery, Robin Dawson, Jennifer Pratscher, Gordon P. Murphy, Terry J. McGenity, and J. Colin Murrell. 2018. “Poplar Phyllosphere Harbors Disparate Isoprenedegrading Bacteria” 115 (51): 13081–86. http://www.ncbi.nlm.nih.gov/pubmed/30498029.

Crombie, Andrew T., Nasmille L. Mejia-Florez, Terry J. McGenity, and J. Colin Murrell. 2018. “Genetics and Ecology of Isoprene Degradation.” In *Aerobic Utilization of Hydrocarbons, Oils and Lipids*, 1–15. Cham: Springer International Publishing.

Crombie, Andrew T, Helen Emery, Terry J McGenity, and J Colin Murrell. 2017. “Draft Genome Sequences of Three Terrestrial Isoprene-Degrading *Rhodococcus* Strains.” *Genome Announcements* 5 (45): e01256-17.

Csiky, O, and G Seufert. 1999. “Terpenoid Emissions of Mediterranean Oaks and Their Relation to Taxonomy.” *Ecological Applications* 9 (4): 1138–46.

Curson, Andrew R.J. J., Beth T. Williams, Benjamin J. Pinchbeck, Leanne P. Sims, Ana Bermejo Martínez, Peter Paolo L. Rivera, Deepak Kumaresan, *et al*. 2018. "DSYB Catalyses the Key Step of Dimethylsulfoniopropionate Biosynthesis in Many Phytoplankton" 3 (4): 430–39.

Dani, K. G. Srikanta, Giuseppe Torzillo, Marco Michelozzi, Rita Baraldi, and Francesco Loreto. 2020. "Isoprene Emission in Darkness by a Facultative Heterotrophic Green Alga." *Frontiers in Plant Science* 11 (November): 598786.

Dani, K.G. Srikanta, and Francesco Loreto. 2017. "Trade-Off Between Dimethyl Sulfide and Isoprene Emissions from Marine Phytoplankton." *Trends in Plant Science* 22 (5): 361–72.

Dawson, R. A., A. T. Crombie, P. Pichon, M. Steinke, T. J. McGenity, and J. C. Murrell. 2021. "The Microbiology of Isoprene Cycling in Aquatic Ecosystems." *Aquatic Microbial Ecology* 87 (July): 79–98.

Dawson, Robin. 2021. "Molecular and Biochemical Characterisation of Isoprene Metabolism by Variovorax Sp. WS11." University of East Anglia.

Dawson, Robin A., Nasmille L. Larke-Mejía, Andrew T. Crombie, Muhammad Farhan Ul Haque, and J. Colin Murrell. 2020. "Isoprene Oxidation by the Gram-Negative Model Bacterium *Variovorax* Sp. WS11." *Microorganisms* 8 (3): 349.

Delmont, Tom O., Patrick Robe, Sébastien Cecillon, Ian M. Clark, Florentin Constancias, Pascal Simonet, Penny R. Hirsch, and Timothy M. Vogel. 2011. "Accessing the Soil Metagenome for Studies of Microbial Diversity." *Applied and Environmental Microbiology* 77 (4): 1315–24.

Dumont, Marc G., and J. Colin Murrell. 2005a. "Community-Level Analysis: Key Genes of Aerobic Methane Oxidation." *Methods in Enzymology* 397 (January): 413–27.

Meyer, F, Paarmann, D, D'Souza, M, Olson, R, Glass, EM, Kubal, M, Paczian, T, Rodriguez, A, Stevens, R, Wilke, A, Wilkening, J Edwards. 2005b "Stable Isotope Probing — Linking Microbial Identity to Function." *Nature Reviews Microbiology 3:6* 3 (6): 499–504.

Dunfield, Peter F., Werner Liesack, Thilo Henckel, Roger Knowles, and Ralf Conrad. 1999. "High-Affinity Methane Oxidation by a Soil Enrichment Culture Containing a Type II Methanotroph." *Applied and Environmental Microbiology* 65 (3): 1009–14.

Eddy, Sean R. 2011. "Accelerated Profile HMM Searches." *PLOS Computational Biology* 7 (10): e1002195.

Ewers, Jens, Doris Freier-Schröder, and Hans-Joachim Knackmuss. 1990. "Selection of Trichloroethene (TCE) Degrading Bacteria That Resist Inactivation by TCE." *Archives of Microbiology* 154 (4): 410–13.

Exton, D. A., D. J. Suggett, T. J. McGenity, and M. Steinke. 2013. "Chlorophyll-Normalized Isoprene Production in Laboratory Cultures of Marine Microalgae and Implications for Global Models." *Limnology and Oceanography* 58 (4): 1301–11.

Exton, Dan A., Terry J. McGenity, Michael Steinke, David J. Smith, and David J. Suggett. 2015. "Uncovering the Volatile Nature of Tropical Coastal Marine Ecosystems in a Changing World." *Global Change Biology* 21 (4): 1383–94.

Fadiji, Ayomide Emmanuel, and Olubukola Oluranti Babalola. 2020. "Metagenomics Methods for the Study of Plant-Associated Microbial Communities: A Review." *Journal of Microbiological Methods* 170 (March): 105860.

Fall, R, and R K Monson. 1992. “Isoprene Emission Rate and Intercellular Isoprene Concentration as Influenced by Stomatal Distribution and Conductance.” *Plant Physiology* 100 (2): 987–92.

Fall, Ray, and Shelley D. Copley. 2000. “Bacterial Sources and Sinks of Isoprene, a Reactive Atmospheric Hydrocarbon.” *Environmental Microbiology* 2 (2): 123–30.

Farhan Ul Haque, Muhammad, Andrew T. Crombie, Scott A. Ensminger, Calin Baciu, and J. Colin Murrell. 2018. “Facultative Methanotrophs Are Abundant at Terrestrial Natural Gas Seeps.” *Microbiome 2018 6:1* 6 (1): 1–14.

Fehsenfeld, Fred, Jack Calvert, Ray Fall, Paul Goldan, Alex B. Guenther, C. Nicholas Hewitt, Brian Lamb, *et al*. 1992. “Emissions of Volatile Organic Compounds from Vegetation and the Implications for Atmospheric Chemistry.” *Global Biogeochemical Cycles* 6 (4): 389–430.

Feng, Zhaozhong, Xiangyang Yuan, Silvano Fares, Francesco Loreto, Pin Li, Yasutomo Hoshika, and Elena Paoletti. 2019. “Isoprene Is More Affected by Climate Drivers than Monoterpenes: A Meta-Analytic Review on Plant Isoprenoid Emissions.” *Plant, Cell & Environment* 42 (6): 1939–49.

Fenske, Jill D., and Suzanne E. Paulson. 2011. “Human Breath Emissions of VOCs.” *Journal of the Air & Waste Management Association* 49.5 (1999): 594-98.

Fish, Jordan A., Benli Chai, Qiong Wang, Yanni Sun, C. Titus Brown, James M. Tiedje, and James R. Cole. 2013. “FunGene: The Functional Gene Pipeline and Repository.” *Frontiers in Microbiology* 4 (OCT): 291.

Food and Agricultural Organization (FAO). 2015. “Global Forest Resources Assessment.” Rome, Italy.

Fowler, David *et al*. 2008. “Ground-Level Ozone in the 21st Century: Future Trends, Impacts and Policy Implications.” *Royal Society Science Policy Report*, no. 15: 134.

Franco, Christopher, Philip Michelsen, Nigel Percy, Vanessa Conn, Erna Listiana, Simon Moll, Rosemary Loria, and Justin Coombs. 2007. “Actinobacterial Endophytes for Improved Crop Performance.” *Australasian Plant Pathology* 36.6 (2007): 524-531.

Frederick M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, and Kevin Struhl. 2003. *Current Protocols in Molecular Biology*. JohnWiley & Sons Inc.

Fulthorpe, Roberta R., Luiz F.W. Roesch, Alberto Riva, and Eric W. Triplett. 2008. “Distantly Sampled Soils Carry Few Species in Common.” *ISME Journal* 2 (9): 901–10.

Furuya, Toshiki, Mika Hayashi, Hisashi Semba, and Kuniki Kino. 2012. “The Mycobacterial Binuclear Iron Monooxygenases Require a Specific Chaperonin-like Protein for Functional Expression in a Heterologous Host.” *FEBS Journal* 280 (3): (2013): 817-26.

Furuya, Toshiki, Satomi Hirose, Hisashi Osanai, Hisashi Semba, and Kuniki Kino. 2011. “Identification of the Monooxygenase Gene Clusters Responsible for the Regioselective Oxidation of Phenol to Hydroquinone in Mycobacteria.” *Applied and Environmental Microbiology* 77 (4): 1214–20.

Geron, Chris, Sue Owen, Alex Guenther, Jim Greenberg, Rei Rasmussen, Jian Hui Bai, Qing-Jun Li, and Brad Baker. 2006. “Volatile Organic Compounds from Vegetation in Southern Yunnan Province, China: Emission Rates and Some Potential Regional Implications.” *Atmospheric Environment* 40 (10): 1759–73.

Ghori, Noor-Ul-Huda, Benjamin Moreira-Grez, Paton Vuong, Ian Waite, Tim Morald, Michael Wise, and Andrew S. Whiteley. 2019. “RNA Stable Isotope Probing (RNA-SIP).” *Methods in Molecular Biology* 2046: 31–44.

Gibson, L., Nasmille L. Larke-Mejía, and J. Colin Murrell. 2020. “Complete Genome of Isoprene Degrading *Nocardioides* Sp. Ws12.” *Microorganisms* 8 (6): 1–8.

Ginkel, C.G., E. Jong, J.W.R. Tilanus, and J.A.M. Bont. 1987. “Microbial Oxidation of Isoprene, a Biogenic Foliage Volatile and of 1,3-Butadiene, an Anthropogenic Gas.” *FEMS Microbiology Letters* 45 (5): 275–79.

Ginkel, C G van, H G Welten, and J A de Bont. 1987. “Oxidation of Gaseous and Volatile Hydrocarbons by Selected Alkene-Utilizing Bacteria.” *Applied and Environmental Microbiology* 53 (12): 2903–7.

Goris, Johan, Konstantinos T. Konstantinidis, Joel A. Klappenbach, Tom Coenye, Peter Vandamme, and James M. Tiedje. 2007. “DNA–DNA Hybridization Values and Their Relationship to Whole-Genome Sequence Similarities.” *International Journal of Systematic and Evolutionary Microbiology* 57 (1): 81–91.

Gray, Christopher M., Detlev Helmig, and Noah Fierer. 2015. “Bacteria and Fungi Associated with Isoprene Consumption in Soil.” *Elementa: Science of the Anthropocene* 3 (0): 000053.

Guenther, A. B., X. Jiang, C. L. Heald, T. Sakulyanontvittaya, T. Duhl, L. K. Emmons, and X. Wang. 2012. “The Model of Emissions of Gases and Aerosols from Nature Version 2.1 (MEGAN2.1): An Extended and Updated Framework for Modeling Biogenic Emissions.” *Geoscientific Model Development* 5 (6): 1471–92.

Gupta, Radhey S., Brian Lo, and Jeen Son. 2018. “Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus Mycobacterium into an Emended Genus Mycobacterium and Four Novel Genera.” *Frontiers in Microbiology* 9 (FEB): 67.

Hall, T.A. 1999. “BIOEDIT: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/ NT.” *Nucleic Acids Symposium Series* 41: 95–98.

Hanson, David T., Sarah Swanson, Linda E. Graham, and Thomas D. Sharkey. 1999. “Evolutionary Significance of Isoprene Emission from Mosses.” *American Journal of Botany* 86 (5): 634–39.

Harvey, Christopher M., and Thomas D. Sharkey. 2016. “Exogenous Isoprene Modulates Gene Expression in Unstressed Arabidopsis Thaliana Plants.” *Plant, Cell & Environment* 39 (6): 1251–63.

Henckel, Thilo, Udo Jäckel, Sylvia Schnell, and Ralf Conrad. 2000. “Molecular Analyses of Novel Methanotrophic Communities in Forest Soil That Oxidize Atmospheric Methane.” *Applied and Environmental Microbiology* 66 (5): 1801–8.

Henze, Daven K., and John H. Seinfeld. 2006. “Global Secondary Organic Aerosol from Isoprene Oxidation.” *Geophysical Research Letters* 33 (9): L09812.

Hewitt, C. N., A. R. MacKenzie, P. Di Carlo, C. F. Di Marco, J. R. Dorsey, M. Evans, D. Fowler, *et al*. 2009. “Nitrogen Management Is Essential to Prevent Tropical Oil Palm Plantations from Causing Ground-Level Ozone Pollution.” *Proceedings of the National Academy of Sciences* 106 (44): 18447–51.

Hewitt, C. Nicholas, Russell K. Monson, and Ray Fall. 1990. “Isoprene Emissions from the Grass Arundo Donax L. Are Not Linked to Photorespiration.” *Plant Science* 66 (2): 139–44.

Hewitt, C. Nicholas, and Rachel A. Street. 1992. “A Qualitative Assessment of the Emission of Non-Methane Hydrocarbon Compounds from the Biosphere to the Atmosphere in the U.K.: Present Knowledge and Uncertainties.” *Atmospheric Environment Part A, General Topics* 26 (17): 3069–77.

Hewitt, C N, K Ashworth, A Boynard, A. Guenther, B. Langford, A. R. MacKenzie, P. K. Misztal, *et al*. 2011. “Ground-Level Ozone Influenced by Circadian Control of Isoprene Emissions.” *Nature Geoscience* 4 (10): 671–74.

Holmes, A J, P Roslev, I R McDonald, N Iversen, K Henriksen, and J C Murrell. 1999. “Characterization of Methanotrophic Bacterial Populations in Soils Showing Atmospheric Methane Uptake.” *Applied and Environmental Microbiology* 65 (8): 3312–18.

Holmes, Andrew J., and Nicholas V. Coleman. 2008. “Evolutionary Ecology and Multidisciplinary Approaches to Prospecting for Monooxygenases as Biocatalysts.” *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* 94 (1): 75–84.

Holstein, Sarah A., and Raymond J. Hohl. 2004. “Isoprenoids: Remarkable Diversity of Form and Function.” *Lipids 2004 39:4* 39 (4): 293–309.

Hrebien, Victoria, Elisabeth Deschaseaux, and Bradley D. Eyre. 2020. “Isoprene Flux from Permeable Carbonate Sediments on the Great Barrier Reef.” *Marine Chemistry* 225 (September): 103856.

Hu, Bin, Ann Mareike Jarosch, Martin Gauder, Simone Graeff-Hönninger, Jörg Peter Schnitzler, Rüdiger Grote, Heinz Rennenberg, and Jürgen Kreuzwieser. 2018. “VOC Emissions and Carbon Balance of Two Bioenergy Plantations in Response to Nitrogen Fertilization: A Comparison of *Miscanthus* and *Salix*.” *Environmental Pollution* 237 (June): 205–17.

Hylckama Vlieg, J E van, J Kingma, W Kruizinga, and D B Janssen. 1999. “Purification of a Glutathione S-Transferase and a Glutathione Conjugate-Specific Dehydrogenase Involved in Isoprene Metabolism in *Rhodococcus* Sp. Strain AD45.” *Journal of Bacteriology* 181 (7): 2094–2101.

Hylckama Vlieg, J E van, J Kingma, A J van den Wijngaard, and D B Janssen. 1998. “A Glutathione S-Transferase with Activity towards Cis-1, 2-Dichloroepoxyethane Is Involved in Isoprene Utilization by Rhodococcus Sp. Strain AD45.” *Applied and Environmental Microbiology* 64 (8): 2800–2805.

Hylckama Vlieg, J E van, H Leemhuis, J H Spelberg, and D B Janssen. 2000. “Characterization of the Gene Cluster Involved in Isoprene Metabolism in *Rhodococcus* Sp. Strain AD45.” *Journal of Bacteriology* 182 (7): 1956–63.

Ibragimov, Abdulla, Shaufique Fahmi Sidique, and Yeong Sheng Tey. 2019. “Productivity for Sustainable Growth in Malaysian Oil Palm Production: A System Dynamics Modeling Approach.” *Journal of Cleaner Production* 213 (March): 1051–62.

Ilcu, Lorena, Wolf Röther, Jakob Birke, Anton Brausemann, Oliver Einsle, and Dieter Jendrossek. 2017. “Structural and Functional Analysis of Latex Clearing Protein (Lcp) Provides Insight into the Enzymatic Cleavage of Rubber.” *Scientific Reports 2017 7:1* 7 (1): 1–11.

Jameson, Eleanor, Martin Taubert, Sara Coyotzi, Yin Chen, Özge Eyice, Hendrik Schäfer, J. Colin Murrell, Josh D. Neufeld, and Marc G. Dumont. 2017. "DNA-, RNA-, and Protein-Based Stable-Isotope Probing for High-Throughput Biomarker Analysis of Active Microorganisms." *Methods in Molecular Biology (Clifton, N.J.)* 1539: 57–74.

Janssen, Peter H. 2006. "Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S RRNA and 16S RRNA Genes." *Applied and Environmental Microbiology*.

Jendrossek, Dieter, and Jakob Birke. 2019. "Rubber Oxygenases." *Applied Microbiology and Biotechnology*. Springer Verlag.

Jendrossek, Dieter, and Simone Reinhardt. 2003. "Sequence Analysis of a Gene Product Synthesized by *Xanthomonas* Sp. during Growth on Natural Rubber Latex." *FEMS Microbiology Letters* 224 (1): 61–65.

Johnston, Antonia. 2014. "Molecular Ecology of Marine Isoprene Degradation." University of East Anglia.

Johnston, Antonia, Andrew T. Crombie, Myriam El Khawand, Leanne Sims, Gregg M. Whited, Terry J. McGenity, J Colin Murrell, *et al*. 2017. "Identification and Characterisation of Isoprene-Degrading Bacteria in an Estuarine Environment." *Environmental Microbiology* 19 (9): 3526–37.

Johnston, Antonia, Andrew T Crombie, Myriam El Khawand, Leanne Sims, Gregg M Whited, Terry J McGenity, and J Colin Murrell. 2017. "Identification and Characterisation of Isoprene-Degrading Bacteria in an Estuarine Environment." *Environmental Microbiology* 19 (9): 3526–37.

Julsing, Mattijs K., Michael Rijpkema, Herman J. Woerdenbag, Wim J. Quax, and Oliver Kayser. 2007. "Functional Analysis of Genes Involved in the Biosynthesis of Isoprene in *Bacillus Subtilis*." *Applied Microbiology and Biotechnology 2007 75:6* 75 (6): 1377–84.

Kang, Dongwan D., Feng Li, Edward Kirton, Ashleigh Thomas, Rob Egan, Hong An, and Zhong Wang. 2019. "MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies." *PeerJ* 2019 (7): e7359.

Karimi, Battle, Sébastien Terrat, Samuel Dequiedt, Nicolas P.A. Saby, Walid Horrigue, Mélanie Lelièvre, Virginie Nowak, *et al*. 2018. "Biogeography of Soil Bacteria and Archaea across France." *Science Advances* 4 (7): eaat1808.

Kertesz*, M. A., and A. Kawasaki. 2010. "Hydrocarbon-Degrading *Sphingomonads: Sphingomonas, Sphingobium, Novosphingobium, and Sphingopyxis*." In *Handbook of Hydrocarbon and Lipid Microbiology*, 1693–1705. Springer Berlin Heidelberg.

Kesselmeier, J., and M. Staudt. 1999. "Biogenic Volatile Organic Compounds (VOC): An Overview on Emission, Physiology and Ecology." *Journal of Atmospheric Chemistry* 33 (1): 23–88.

Khawand, Myriam El, Andrew T. Crombie, Antonia Johnston, Dmitrii V. Vavlline, Joseph C. McAuliffe, Jacob A. Latone, Yuliya A. Primak, *et al*. 2016. "Isolation of Isoprene Degrading Bacteria from Soils, Development of IsoA Gene Probes and Identification of the Active Isoprene-Degrading Soil Community Using DNA-Stable Isotope Probing." *Environmental Microbiology* 18 (8): 2743–53.

Kim, Myung Kyum, Sathiyaraj Srinivasan, Min Ju Park, Gayathri Sathiyaraj, Yeon Ju Kim, and Deok Chun Yang. 2009. "*Nocardioides Humi* Sp. Nov., a β-Glucosidase-Producing Bacterium Isolated from Soil of a Ginseng Field." *International Journal of Systematic and Evolutionary Microbiology* 59 (11): 2724–28.

Klappenbach, Joel A., Paul R. Saxman, James R. Cole, and Thomas M. Schmidt. 2001. "Rrndb: The Ribosomal RNA Operon Copy Number Database." *Nucleic Acids Research* 29 (1): 181–84.

Knief, Claudia, Nathanaël Delmotte, Samuel Chaffron, Manuel Stark, Gerd Innerebner, Reiner Wassmann, Christian Von Mering, and Julia A. Vorholt. 2012. "Metaproteogenomic Analysis of Microbial Communities in the Phyllosphere and Rhizosphere of Rice." *ISME Journal* 6 (7): 1378–90.

Köksal, Mustafa, Ina Zimmer, Jörg Peter Schnitzler, and David W. Christianson. 2010. "Structure of Isoprene Synthase Illuminates the Chemical Mechanism of Teragram Atmospheric Carbon Emission." *Journal of Molecular Biology* 402 (2): 363–73.

Konstantinidis, Konstantinos T., and James M. Tiedje. 2005. "Genomic Insights That Advance the Species Definition for Prokaryotes." *Proceedings of the National Academy of Sciences* 102 (7): 2567–72.

Kotani, Tetsuya, Tazuko Yamamoto, Hiroya Yurimoto, Yasuyoshi Sakai, and Nobuo Kato. 2003. "Propane Monooxygenase and NAD+-Dependent Secondary Alcohol Dehydrogenase in Propane Metabolism by *Gordonia* Sp. Strain TY-5." *Journal of Bacteriology* 185 (24): 7120–28.

Kronen, Miriam, Matthew Lee, Zackary L. Jones, and Michael J. Manefield. 2019. "Reductive Metabolism of the Important Atmospheric Gas Isoprene by Homoacetogens." *The ISME Journal 2019 13:5* 13 (5): 1168–82.

Kubota, Masahiro, Kazuyoshi Kawahara, Kachiko Sekiya, Tetsuya Uchida, Yasuko Hattori, Hiroyuki Futamata, and Akira Hiraishi. 2005. "*Nocardioides Aromaticivorans* Sp. Nov., a Dibenzofuran-Degrading Bacterium Isolated from Dioxin-Polluted Environments." *Systematic and Applied Microbiology* 28 (2): 165–74.

Kuzma, Jennifer, Michele Nemecek-Marshall, Walter H. Pollock, and Ray Fall. 1995. "Bacteria Produce the Volatile Hydrocarbon Isoprene." *Current Microbiology* 30 (2): 97–103.

Lantz, Alexandra T., Joshua Allman, Sarathi M. Weraduwage, and Thomas D. Sharkey. 2019. "Isoprene: New Insights into the Control of Emission and Mediation of Stress Tolerance by Gene Expression." *Plant Cell and Environment*. Blackwell Publishing Ltd.

Laothawornkitkul, Jullada, Nigel D. Paul, Claudia E. Vickers, Malcolm Possell, Jane E. Taylor, Philip M. Mullineaux, and C. Nicholas Hewitt. 2008. "Isoprene Emissions Influence Herbivore Feeding Decisions." *Plant, Cell and Environment* 31 (10): 1410–15.

Larke-Mejía, Nasmille L, Ornella Carrión, Andrew T Crombie, Terry J McGenity, and J Colin Murrell. 2020. "*Sphingopyxis* Sp. Strain OPL5, an Isoprene-Degrading Bacterium from the *Sphingomonadaceae* Family Isolated from Oil Palm Leaves." *Microorganisms* 8 (10): 1–16.

Larke-Mejia, Nasmille Liceth, Andrew Crombie, Jennifer Pratscher, Terry John McGenity, and Colin Murrell. 2019. "Novel Isoprene-Degrading Proteobacteria from Soil and Leaves Identified by Cultivation and Metagenomics Analysis of Stable Isotope Probing Experiments." *Frontiers in Microbiology* 10: 2700.

Lathière, J., C. N. Hewitt, and D. J. Beerling. 2010. “Sensitivity of Isoprene Emissions from the Terrestrial Biosphere to 20th Century Changes in Atmospheric CO_2 Concentration, Climate, and Land Use.” *Global Biogeochemical Cycles* 24 (2010).

Leahy, Joseph G., Patricia J. Batchelor, and Suzanne M. Morcomb. 2003. “Evolution of the Soluble Diiron Monooxygenases.” *FEMS Microbiology Reviews* 27 (4): 449–79.

Leinonen, Rasko, Hideaki Sugawara, Martin Shumway, and on behalf of the International Nucleotide Sequence Database Collaboration. 2011. “The Sequence Read Archive.” *Nucleic Acids Research* 39 (suppl_1): D19–21.

Li, Haiyan, Minna Väliranta, Mari Mäki, Lukas Kohl, A Britta K Sannel, Jukka Pumpanen, Markku Koskinen, Jaana Bäck, and Federico Bianchi. 2020. “Overlooked Organic Vapor Emissions from Thawing Arctic Permafrost.” *Environmental Research Letters* 15 (10): 104097.

Lichtenthaler, Hartmut K. 1999. “The 1-Deoxy-D-Xylulose-5-Phosphate Pathway of Isoprenoid Biosynthesis in Plants.” *Annual Review of Plant Biology* 50 (November): 47–65.

Lichtenthaler, Hartmut K., Jörg Schwender, Andrea Disch, and Michel Rohmer. 1997. “Biosynthesis of Isoprenoids in Higher Plant Chloroplasts Proceeds via a Mevalonate-Independent Pathway.” *FEBS Letters* 400 (3): 271–74.

Lin, Y. H., E. M. Knipping, E. S. Edgerton, S. L. Shaw, and J. D. Surratt. 2013. “Investigating the Influences of SO_2 and NH_3 Levels on Isoprene-Derived Secondary Organic Aerosol Formation Using Conditional Sampling Approaches.” *Atmospheric Chemistry and Physics* 13 (16): 8457–70.

Lindwall, Frida, Sophie Sylvest Svendsen, Cecilie Skov Nielsen, Anders Michelsen, and Riikka Rinnan. 2016. “Warming Increases Isoprene Emissions from an Arctic Fen.” *Science of The Total Environment* 553 (May): 297–304.

Linos, Alexandros, Alexander Steinbüchel, Cathrin Spröer, and Reiner M. Kroppenstedt. 1999. “*Gordonia Polyisoprenivorans* Sp. Nov., a Rubber-Degrading Actinomycete Isolated from an Automobile Tyre.” *International Journal of Systematic Bacteriology* 49 (4): 1785–91.

Logan, Barry A., and Russell K. Monson. 1999. “Thermotolerance of Leaf Discs from Four Isoprene-Emitting Species Is Not Enhanced by Exposure to Exogenous Isoprene.” *Plant Physiology* 120 (3): 821–26.

Logan, Barry A., Russell K. Monson, and Mark J. Potosnak. 2000. “Biochemistry and Physiology of Foliar Isoprene Production.” *Trends in Plant Science*.

Loivamäki, Maaria, Sandrine Louis, Gyöngyi Cinege, Ina Zimmer, Robert J. Fischbach, and Jörg-Peter Schnitzler. 2007. “Circadian Rhythms of Isoprene Biosynthesis in Grey Poplar Leaves.” *Plant Physiology* 143 (1): 540–51.

Loreto, Francesco, Paolo Ciccioli, Enzo Brancaleoni, Riccardo Valentini, Manuela De Lillis, Olav Csiky, and Guenther Seufert. 1998. “A Hypothesis on the Evolution of Isoprenoid Emission by Oaks Based on the Correlation between Emission Type and *Quercus* Taxonomy.” *Oecologia* 115 (3): 302–5.

Lu, Jennifer, Florian P Breitwieser, Peter Thielen, and Steven L Salzberg. 2017. “Bracken: Estimating Species Abundance in Metagenomics Data.” *PeerJ Computer Science* 2017 (1).

Lyu, Yihua, Wei Zheng, Tianling Zheng, and Yun Tian. 2014. “Biodegradation of Polycyclic Aromatic Hydrocarbons by *Novosphingobium Pentaromativorans* US6-1.” *PLoS ONE* 9 (7): 101438.

Maddocks, Sarah E., and Petra C. F. Oyston. 2008. “Structure and Function of the LysR-Type Transcriptional Regulator (LTTR) Family Proteins.” *Microbiology* 154 (12): 3609–23.

Maidak, Bonnie L., Gary J. Olsen, Niels Larsen, Ross Overbeek, Michael J. McCaughey, and Carl R. Woese. 1996. “The Ribosomal Database Project (RDP).” *Nucleic Acids Research*. Oxford Academic.

Mao, Dan Ping, Quan Zhou, Chong Yu Chen, and Zhe Xue Quan. 2012. “Coverage Evaluation of Universal Bacterial Primers Using the Metagenomic Datasets.” *BMC Microbiology* 12 (1): 1–8.

Mayrhofer, Sabine, Markus Teuber, Ina Zimmer, Sandrine Louis, Robert J Fischbach, and Jörg Peter Schnitzler. 2005. “Diurnal and Seasonal Variation of Isoprene Biosynthesis-Related Genes in Grey Poplar Leaves.” *Plant Physiology*.

McAuliffe, J. C., LaDuca, R. J., Ben-Shoshan, E. A., & Sanford, K. J. 2010. “Development of a Gas-Phase Bioprocess for Isoprenemonomer Production Using Metabolic Pathway Engineering.” *Macromol Rapid Commun* 34: 837.

McDonald, I. R., E. M. Kenna, and J. C. Murrell. 1995. “Detection of Methanotrophic Bacteria in Environmental Samples with the PCR.” *Applied and Environmental Microbiology* 61 (1): 116–21.

McDonald, Ian R, Levente Bodrossy, Yin Chen, and J Colin Murrell. 2008. “Molecular Ecology Techniques for the Study of Aerobic Methanotrophs.” *Applied and Environmental Microbiology* 74 (5): 1305–15.

McGenity, Terry J., Andrew T. Crombie, and J. Colin Murrell. 2018. “Microbial Cycling of Isoprene, the Most Abundantly Produced Biological Volatile Organic Compound on Earth.” *The ISME Journal* 12 (4).

Meselson, Matthew, and Franklin W. Stahl. 1958. “The Replication of DNA in *Escherichia Coli*.” *Proceedings of the National Academy of Sciences* 44 (7): 671–82.

Meskhidze, N., A. Sabolis, R. Reed, and D. Kamykowski. 2015. “Quantifying Environmental Stress-Induced Emissions of Algal Isoprene and Monoterpenes Using Laboratory Measurements.” *Biogeosciences* 12 (3): 637–51.

Meyer, F, D Paarmann, M D’Souza, R Olson, EM Glass, M Kubal, T Paczian, *et al*. 2008. “The Metagenomics RAST Server – a Public Resource for the Automatic Phylogenetic and Functional Analysis of Metagenomes.” *BMC Bioinformatics 2008 9:1* 9 (1): 1–8.

Mikheenko, Alla, Vladislav Saveliev, and Alexey Gurevich. 2016. “MetaQUAST: Evaluation of Metagenome Assemblies.” *Bioinformatics* 32 (7): 1088–90.

Minoche, André E., Juliane C. Dohm, and Heinz Himmelbauer. 2011. “Evaluation of Genomic High-Throughput Sequencing Data Generated on Illumina HiSeq and Genome Analyzer Systems.” *Genome Biology* 12 (11).

Monson, Russell K., Ryan T. Jones, Todd N. Rosenstiel, and Jörg Peter Schnitzler. 2013. “Why Only Some Plants Emit Isoprene.” *Plant, Cell and Environment* 36 (3): 503–16.

Monson, Russell K, Nicole Trahan, Todd N Rosenstiel, Patrick Veres, David Moore, Michael Wilkinson, Richard J Norby, *et al*. 2007. “Isoprene Emission from Terrestrial Ecosystems in Response to Global Change: Minding the Gap between Models and Observations.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365 (1856): 1677–95.

Morais, Ana R.C., Sylwia Dworakowska, Alberto Reis, Luisa Gouveia, Cristina T. Matos, Dariusz Bogdał, and Rafał Bogel-ŁUkasik. 2015. “Chemical and Biological-Based Isoprene Production: Green Metrics.” *Catalysis Today* 239 (January): 38–43.

Morris, Samantha A, Stefan Radajewski, Toby W Willison, and J Colin Murrell. 2002. “Identification of the Functionally Active Methanotroph Population in a Peat Soil Microcosm by Stable-Isotope Probing.” *Applied and Environmental Microbiology* 68 (3): 1446–53.

Murphy, Gordon P. 2017. “Isoprene Degradation in the Terrestrial Environment.” University of Essex.

Murrell, J. Colin, Terry J. McGenity, and Andrew T. Crombie. 2020. “Microbial Metabolism of Isoprene: A Much-Neglected Climate-Active Gas.” *Microbiology (United Kingdom)*. Microbiology Society.

Nazir, Asiya. 2016. “Microbial Degradation of Poly Aromatic Hydrocarbons View Project Review on Metagenomics and Its Applications.” *Imperial Journal of Interdisciplinary Research IJIR* 2 (3).

Neufeld, Josh D., Marc G. Dumont, Jyotsna Vohra, and J. Colin Murrell. 2007. “Methodological Considerations for the Use of Stable Isotope Probing in Microbial Ecology.” *Microbial Ecology* 53 (3): 435–42.

Neufeld, Josh D, Jyotsna Vohra, Marc G Dumont, Tillmann Lueders, Mike Manefield, Michael W Friedrich, and J Colin Murrell. 2007. “DNA Stable-Isotope Probing.” *Nature Protocols* 2 (4): 860–66.

Nicholas, K. B., and H. B. Nicholas Jr. 1997. “GeneDoc: A Tool for Editing and Annotating Multiple Sequence Alignments.” *Pittsburgh Supercomputing Center’s National Resource for Biomedical Supercomputing* ver. 2.7.0.

Nurk, Sergey, Dmitry Meleshko, Anton Korobeynikov, and Pavel A. Pevzner. 2017. “MetaSPAdes: A New Versatile Metagenomic Assembler.” *Genome Research* 27 (5): 824–34.

OECD-FAO Agricultural Outlook 2017-2026. 2017. OECD-FAO Agricultural Outlook. OECD.

Pacifico, F., S.P. Harrison, C.D. Jones, and S. Sitch. 2009. “Isoprene Emissions and Climate.” *Atmospheric Environment* 43 (39): 6121–35.

Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. 2015. “CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes.” *Genome Research* 25 (7): 1043–55.

Pegoraro, Emiliano, Leif Abrell, Joost Van Haren, Greg Barron-Gafford, Katherine Ann Grieve, Yadvinder Malhi, Ramesh Murthy, and Guanghui Lin. 2005. “The Effect of Elevated Atmospheric CO2 and Drought on Sources and Sinks of Isoprene in a Temperate and Tropical Rainforest Mesocosm.” *Global Change Biology* 11 (8): 1234–46.

Pegoraro, Emiliano, Ana Rey, Leif Abrell, Joost Van Haren, and Guanghui Lin. 2006. “Drought Effect on Isoprene Production and Consumption in Biosphere 2 Tropical Rainforest.” *Global Change Biology* 12 (3): 456–69.

Pio, C.A., Nunes, T.V. and Brito, S. 1993. “Volatile Hydrocarbon Emissions from Common and Native Species of Vegetation in Portugal.” In *Proceedings of the Joint Workshop of CEC/BIATEX of EUROTRAC, General Assessment of Biogenic Emissions and Deposition of Nitrogen Compounds, Sulfur Compounds and Oxidants in Europe*, edited by S. Slanina, J., Angeletti, G., and Beilke, 291–98. Air Pollution Research Report.

Radajewski, Stefan, Philip Ineson, Nisha R. Parekh, and J. Colin Murrell. 2000. “Stable-Isotope Probing as a Tool in Microbial Ecology.” *Nature* 403 (6770): 646–49.

Radajewski, Stefan, James I. Prosser, David B. Nedwell, Philip Ineson, Samantha A. Morris, David S. Reay, Gordon Webster, and J. Colin Murrell. 2002. “Identification of Active Methylotroph Populations in an Acidic Forest Soil by Stable-Isotope Probing C.” *Microbiology* 148 (8): 2331–42.

Rangjaroen, Chakrapong, Rungroch Sungthong, Benjavan Rerkasem, Neung Teaumroong, Rujirek Noisangiam, and Saisamorn Lumyong. 2017. “Untapped Endophytic Colonization and Plant Growth-Promoting Potential of the Genus Novosphingobium to Optimize Rice Cultivation.” *Microbes and Environments* 32 (1): 84–87.

Rasmussen, R. A., and M. A. K. Khalil. 1988. “Isoprene over the Amazon Basin.” *Journal of Geophysical Research: Atmospheres* 93 (D2): 1417–21.

Rasmussen, R.A. 1978. “Isoprene Plant Species List. Special Report of Air Pollution Research Section.” Pullman.

Rodriguez-R, Luis M, Santosh Gunturu, William T Harvey, Ramon Rosselló-Mora, James M Tiedje, James R Cole, and Konstantinos T Konstantinidis. 2018. “The Microbial Genomes Atlas (MiGA) Webserver: Taxonomic and Gene Diversity Analysis of Archaea and Bacteria at the Whole Genome Level.” *Nucleic Acids Research* 46 (W1): W282–88.

Roesch, Luiz F.W., Roberta R. Fulthorpe, Alberto Riva, George Casella, Alison K.M. Hadwin, Angela D. Kent, Samira H. Daroub, Flavio A.O. Camargo, William G. Farmerie, and Eric W. Triplett. 2007. “Pyrosequencing Enumerates and Contrasts Soil Microbial Diversity.” *ISME Journal* 1 (4): 283–90.

Rohmer, M. 1999. “The Discovery of a Mevalonate-Independent Pathway for Isoprenoid Biosynthesis in Bacteria, Algae and Higher Plants.” *Natural Product Reports* 16 (5): 565–74.

Rohwerder, Thore, Maria-Teresa Rohde, Nico Jehmlich, and Jessica Purswani. 2020. “Actinobacterial Degradation of 2-Hydroxyisobutyric Acid Proceeds via Acetone and Formyl-CoA by Employing a Thiamine-Dependent Lyase Reaction.” *Frontiers in Microbiology* 0 (April): 691.

Rose, Karsten, Klaus B. Tenberge, and Alexander Steinbüchel. 2005. “Identification and Characterization of Genes from Streptomyces Sp. Strain K30 Responsible for Clear Zone Formation on Natural Rubber Latex and Poly(Cis-1,4-Isoprene) Rubber Degradation.” *Biomacromolecules* 6 (1): 180–88.

Schloss, Patrick D., and Jo Handelsman. 2003. “Biotechnological Prospects from Metagenomics.” *Current Opinion in Biotechnology* 14 (3): 303–10.

Seemann, T. 2014. “Prokka: Rapid Prokaryotic Genome Annotation.” *Bioinformatics* 30 (14): 2068–69.

Sharkey, T. D., A. E. Wiberley, and A. R. Donohue. 2007. “Isoprene Emission from Plants: Why and How.” *Annals of Botany* 101 (1): 5–18.

Sharkey, Thomas D. 1996. “Isoprene Synthesis by Plants and Animals.” *Endeavour* 20 (2): 74–78.

Sharkey, Thomas D., and Eric L. Singsaas. 1995. “Why Plants Emit Isoprene.” *Nature 1995 374:6525* 374 (6525): 769–769.

Sharkey, Thomas D, and Sansun Yeh. 2001. “Isoprene Emission from Plants.” *Annual Review of Plant Physiology and Plant Molecular Biology* 52 (1): 407–36.

Sharp, Jonathan O., Christopher M. Sales, Justin C. LeBlanc, Jie Liu, Thomas K. Wood, Lindsay D. Eltis, William W. Mohn, and Lisa Alvarez-Cohen. 2007. “An Inducible Propane Monooxygenase Is Responsible for N- Nitrosodimethylamine Degradation by Rhodococcus Sp. Strain RHA1.” *Applied and Environmental Microbiology* 73 (21): 6930–38.

Singh, Abhishek, Navnita Srivastava, and Suresh Kumar Dubey. 2019. “Molecular Characterization and Kinetics of Isoprene Degrading Bacteria.” *Bioresource Technology* 278 (April): 51–56.

Singsaas, E. L., M. Lerdau, K. Winter, and T. D. Sharkey. 1997. “Isoprene Increases Thermotolerance of Isoprene-Emitting Species.” *Plant Physiology* 115 (4): 1413–20.

Singsaas, Erie L, Manuel Lerdau2, Katherine Winter, and Thomas D Sharkey. 1997. “Lsoprene Lncreases Thermotolerance of Isoprene-Emitting Species’.” *Plant Physiol*. Vol. 115. https://plantphysiol.org.

Sohn, Jae Hak, Kae Kyoung Kwon, Ji Hyun Kang, Hong Bae Jung, and Sang Jin Kim. 2004. “Novosphingobium Pentaromativorans Sp. Nov., a High-Molecular-Mass Polycyclic Aromatic Hydrocarbon-Degrading Bacterium Isolated from Estuarine Sediment.” *International Journal of Systematic and Evolutionary Microbiology* 54 (5): 1483–87.

Springer, Amy L, Ann J Auman, and Mary E Lidstrom. 1998. “Sequence and Characterization of MxaB , a Response Regulator Involved in Regulation of Methanol Oxidation, and of MxaW , a Methanol-Regulated Gene in *Methylobacterium Extorquens* AM1.” *FEMS Microbiology Letters* 160 (1): 119–24.

Srivastva, Navnita, Awadhesh Kumar Shukla, Ram Sharan Singh, Siddh Nath Upadhyay, and Suresh Kumar Dubey. 2015. “Characterization of Bacterial Isolates from Rubber Dump Site and Their Use in Biodegradation of Isoprene in Batch and Continuous Bioreactors.” *Bioresource Technology* 188 (July): 84–91.

Stavrakou, T, J.-F Müller, M Bauwens, I De Smedt, M Van Roozendael, A Guenther, M Wild, and X Xia. 2014. “Atmospheric Chemistry and Physics Isoprene Emissions over Asia 1979-2012: Impact of Climate and Land-Use Changes.” *Atmos. Chem. Phys* 14: 4587–4605.

Steinfeld, Jeffrey I. 1998. “Atmospheric Chemistry and Physics: From Air Pollution to Climate Change.” *Environment: Science and Policy for Sustainable Development* 40 (7): 26–26.

Sun, Zhihong, Katja Hüve, Vivian Vislap, and Ülo Niinemets. 2013. “Elevated [CO2] Magnifies Isoprene Emissions under Heat and Improves Thermal Resistance in Hybrid Aspen.” *Journal of Experimental Botany* 64 (18): 5509–23.

Suzuki, Ken Ichiro, and Kazuo Komagata. 1983. “Pimelobacter Gen. Nov., a New Genus of Coryneform Bacteria with LL-Diaminopimelic Acid in the Cell Wall.” *The Journal of General and Applied Microbiology* 29 (1): 59–71.

Tiirola, M. A., M. K. Männistö, J. A. Puhakka, and M. S. Kulomaa. 2002. “Isolation and Characterization of *Novosphingobium* Sp. Strain MT1, a Dominant Polychlorophenol-Degrading Strain in a Groundwater Bioremediation System.” *Applied and Environmental Microbiology* 68 (1): 173–80.

Uritskiy, Gherman V., Jocelyne Diruggiero, and James Taylor. 2018. “MetaWRAP - A Flexible Pipeline for Genome-Resolved Metagenomic Data Analysis” *Microbiome* 6 (1).

Valášková, V., and P. Baldrian. 2009. “Denaturing Gradient Gel Electrophoresis as a Fingerprinting Method for the Analysis of Soil Microbial Communities.” *Plant, Soil and Environment* 55 (2009) (No. 10): 413–23.

Vallenet, David, Alexandra Calteau, Mathieu Dubois, Paul Amours, Adelme Bazin, Mylène Beuvin, Laura Burlot, *et al*. 2020. “MicroScope: An Integrated Platform for the Annotation and Exploration of Microbial Gene Functions through Genomic, Pangenomic and Metabolic Comparative Analysis.” *Nucleic Acids Research* 48 (D1): D579–89.

Velikova, Violeta, Thomas D. Sharkey, and Francesco Loreto. 2012. “Stabilization of Thylakoid Membranes in Isoprene-Emitting Plants Reduces Formation of Reactive Oxygen Species.” *Plant Signaling & Behavior* 7 (1): 139–41.

Vickers, Claudia E., Mareike Bongers, Sarah F. Bydder, Panagiotis Chrysanthopoulos, and Mark P. Hodson. 2015. “Protocols for the Production and Analysis of Isoprenoids in Bacteria and Yeast,” 23–52.

Vickers, Claudia E., Malcolm Possell, Cristian I. Cojocariu, Violeta B. Velikova, Jullada Laothawornkitkul, Annette Ryan, Philip M. Mullineaux, and C. Nicholas Hewitt. 2009. “Isoprene Synthesis Protects Transgenic Tobacco Plants from Oxidative Stress.” *Plant, Cell and Environment* 32 (5): 520–31.

Vickers, Claudia E., and Suriana Sabri. 2015. “Isoprene.” *Advances in Biochemical Engineering/Biotechnology* 148: 289–317.

Vives-Peris, Vicente, Aurelio Gómez-Cadenas, and Rosa María Pérez-Clemente. 2018. “Salt Stress Alleviation in Citrus Plants by Plant Growth-Promoting Rhizobacteria *Pseudomonas Putida* and *Novosphingobium* Sp.” *Plant Cell Reports* 37 (11): 1557–69.

Wei, Yu Jie, Yun Wu, Yin Zhuo Yan, Wan Zou, Jie Xue, Wen Rui Ma, Wei Wang, Ge Tian, and Li Ye Wang. 2018. “High-Throughput Sequencing of Microbial Community Diversity in Soil, Grapes, Leaves, Grape Juice and Wine of Grapevine from China.” *PLoS ONE* 13 (3): e0193097.

Whited, G. M., and D. T. Gibson. 1991. “Toluene-4-Monooxygenase, a Three-Component Enzyme System That Catalyzes the Oxidation of Toluene to p-Cresol in Pseudomonas Mendocina KR1.” *Journal of Bacteriology* 173 (9): 3010–16.

Wick, Ryan R., Louise M. Judd, Claire L. Gorrie, and Kathryn E. Holt. 2017. “Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads.” *PLOS Computational Biology* 13 (6): e1005595.

Wilkinson, Michael J., Susan M. Owen, Malcolm Possell, James Hartwell, Peter Gould, Anthony Hall, Claudia Vickers, and C. Nicholas Hewitt. 2006. “Circadian Control of Isoprene Emissions from Oil Palm (*Elaeis Guineensis*).” *Plant Journal* 47 (6): 960–68.

Wood, Derrick E., and Steven L. Salzberg. 2014. “Kraken: Ultrafast Metagenomic Sequence Classification Using Exact Alignments.” *Genome Biology* 15 (3): R46.

Wu, Yu Wei, Blake A. Simmons, and Steven W. Singer. 2016. “MaxBin 2.0: An Automated Binning Algorithm to Recover Genomes from Multiple Metagenomic Datasets.” *Bioinformatics* 32 (4): 605–7.

Xu, Jianping. 2006. “Microbial Ecology in the Age of Genomics and Metagenomics: Concepts, Tools, and Recent Advances.” *Molecular Ecology*. John Wiley & Sons, Ltd.

Yadav, Ajar Nath. 2017. “Plant Microbiomes and Its Beneficial Multifunctional Plant Growth Promoting Attributes.” *International Journal of Environmental Sciences & Natural Resources* 3 (1).

Yang, Jianming, Guang Zhao, Yuanzhang Sun, Yanning Zheng, Xinglin Jiang, Wei Liu, and Mo Xian. 2012. “Bio-Isoprene Production Using Exogenous MVA Pathway and Isoprene Synthase in *Escherichia Coli*.” *Bioresource Technology* 104 (January): 642–47.

Yang, Ruixian, Ping Liu, and Wenyu Ye. 2017. “Illumina-Based Analysis of Endophytic Bacterial Diversity of Tree Peony (Paeonia Sect. Moutan) Roots and Leaves.” *Brazilian Journal of Microbiology* 48 (4): 695–705.

Ye, Lidan, Xiaomei Lv, and Hongwei Yu. 2016. “Engineering Microbes for Isoprene Production.” *Metabolic Engineering* 38 (November): 125–38.

Yoon, Jung Hoon, Young Gyun Cho, Sung Lee Taik, Ken Ichiro Suzuki, Takashi Nakase, and Yong Ha Park. 1999. “*Nocardioides Nitrophenolicus* Sp. Nov., a p-Nitrophenol-Degrading Bacterium.” *International Journal of Systematic Bacteriology* 49 (2): 675–80.

Zhang, Meng, and Mary E. Lidstrom. 2003. “Promoters and Transcripts for Genes Involved in Methanol Oxidation in *Methylobacterium Extorquens* AM1.” *Microbiology*. Microbiology Society.

Zhang, Qian, Jacquelinne J. Acuña, Nitza G. Inostroza, María Luz Mora, Sergio Radic, Michael J. Sadowsky, and Milko A. Jorquera. 2019. “Endophytic Bacterial Communities Associated with Roots and Leaves of Plants Growing in Chilean Extreme Environments.” *Scientific Reports* 9 (1): 1–12.

Zuo, Zhaojiang, Sarathi M. Weraduwage, Alexandra T. Lantz, Lydia M. Sanchez, Sean E. Weise, Jie Wang, Kevin L. Childs, and Thomas D. Sharkey. 2019. “Isoprene Acts as a Signaling Molecule in Gene Networks Important for Stress Responses and Plant Growth.” *Plant Physiology* 180 (1): 124–52.