# Molecular Ecology of Isoprene Degraders in the Terrestrial Environment

**Nasmille Liceth Larke-Mejia (née Mejía Flórez)**

**A thesis submitted to the School of Environmental Sciences in fulfilment of the requirements for the degree of Doctor in Philosophy**

**August 2018**

**University of East Anglia**

**School of Environmental Sciences**

**Norwich, UK**

# Abstract

Isoprene (2-methyl 1, 3-butadiene) is the most abundant non-methane BVOC (biogenic volatile organic compound) released into the atmosphere. Terrestrial plants are the primary producers of isoprene and release 500-750 million tonnes of isoprene per year, to protect themselves from abiotic environmental stresses such as heat and reactive oxygen species. Many studies have explored isoprene production but very little is known about consumption of isoprene by microbes. Cleveland and Yavitt in 1998 (Cleveland and Yavitt 1998), and more recently Khawand *et al.* 2016 (Khawand *et al*. 2016), demonstrated that microbes isolated from terrestrial environments are capable of using isoprene as sole carbon and energy source. By applying cultivation-dependent and cultivation-independent techniques, such as DNA Stable Isotope Probing (Dumont and Murrell 2005), my objective was to determine the distribution, diversity and activity of isoprene-degrading bacteria in the terrestrial environment. Isoprene-degrading microbes were enriched by adding 13 to 50 ppm isoprene to microcosms using topsoil from a willow tree and topsoil/leaves from an oil palm tree. DNA stable isotope probing, using $^{13}$C-labelled isoprene, assisted in revealing the diversity of active isoprene degraders by labelling organisms that incorporated the isoprene, directly or indirectly. PCR retrieval of partial 16S rRNA genes from this DNA revealed labelled members of the genera *Ramlibacter, Variovorax, Rhodococcus* and *Methylibium,* for willow soil, and *Rhodococcus*, *Gordonia, Aquabacterium, Aquincola, Methylobacterium* and members from the Sphingomonadaceae family, for the oil palm tree. Using cultivation-dependent methods I isolated seven phylogenetically different isoprene-utilizing bacteria of the genera *Rhodococcus, Nocardioides* and *Variovorax* from willow soil environment; another four phylogenetically different bacteria belonging to the genera *Gordonia*, *Sphingopyxis* and *Sphingobacterium* from the oil palm tree. Results suggest *Rhodococcus* is a cosmopolitan isoprene-degrader, present in a variety of environments, and different isoprene-degrading bacteria were found associated to willow and oil palm trees.

# Table of Contents

# List of Tables

# List of Figures

# Declaration

I declare that the work presented in this thesis was conducted by me under the direct supervision of my supervisor, J. Colin Murrell, and other members of the isoprene team (Murrell/ELSA lab). Results obtained by other people have been acknowledged in the "Acknowledgements" or "Chapter 2: Materials and Methods". Some of the results from Chapter 3 and Chapter 4 were used in a recent chapter publication (Crombie, Andrew T., Mejia-Florez, Nasmille L., McGenity, Terry J., Murrell, J. Colin. **Genetics and Ecology of Isoprene Degradation.** *Aerobic Utilization of Hydrocarbons, Oils and Lipids.* p.1-15 (2018)), a manuscript submitted to Proceedings of the National Academy of Sciences of the United States of America - PNAS (Andrew T. Crombie, Nasmille L. Larke-Mejia, Helen Emery, Robin Dawson, Jennifer Pratscher, Gordon P. Murphy, Terry J. McGenity, J. Colin Murrell. **The poplar phyllosphere harbours a wide diversity of isoprene-degrading bacteria** (2018)) and some results from Chapters 4, 5 and 6 in a manuscript submitted to Microbiome (**Functional gene probing reveals the widespread distribution, diversity and abundance of isoprene-degrading bacteria in the environment.** Ornella Carrión, Nasmille L. Larke-Mejía, Lisa Gibson, Muhammad Farhan Ul Haque, Javier Ramiro-García, Terry J. McGenity, J. Colin Murrell (2018)).

Nasmille Liceth Larke-Mejía

# Acknowledgements

I would like to thank Colciencias, the Colombian government and the Newton-Fund for opening my horizons by accepting to fund my PhD in the UK. This experience has been invaluable to me and my future as a scientist. I would like to especially thank Professor J. Colin Murrell, initially for inviting me to work in his lab, for his expert guidance, invaluable help and essential support during my studies. To my supervisory team (Professor Jonathan Todd, Dr. Jennifer Pratscher and Dr. Andrew Crombie) and Dr Terence McGenity for their valuable time, enriching discussions and difficult questions. To all the members of the Murrell lab, past and present, for making me feel part of a team, listening and helping me through the day to day scientific challenges. I would like to especially thank Dr. Andrew Crombie for his patience, responses to my many questions and very insightful positive feedback. Thank you to Ornella Carrion-Fonseca for helping me to collect and enrich the oil palm samples, my summer student Natalie Jones for her time and dedication to the isolate work during summer 2017, and Elliot Brooks/Lisa Gibson for the identification of some oil palm tree isolates.

To my family, my loving parents (Marcos Mejia Estrada and Mercedes Florez Zambrano) and my caring brothers (David Orlando Mejia Florez and Marcos Manuel Mejia Florez), thank you for always being there for me and supporting my career choices. Finally, and most especially, to my husband Stephen Peter Larke-Mejia, for always listening. Your unconditional love and care were essential every step of the way. Thank you for being my home away from home!

# Abbreviations

| | |
|---|---|
| ADP | adenosine diphosphate |
| AL | ash leaves |
| Amp | ampicillin |
| ATP | adenosine triphosphate |
| BLAST | basic local alignment search tool |
| bp | base pairs |
| BVOC | biogenic volatile organic compound |
| CI | Criegee intermediates |
| DGGE | denaturing gradient gel electrophoresis |
| $dH_2O$ | deionized water (nuclease free) |
| DMAPP | dimethylallyl diphosphate |
| DNA | deoxyribonucleic acid |
| dw | dry weight |
| e.g. | for example |
| EDTA | ethylenediaminetetraacetic acid |
| EL | eucalyptus leaves |
| g | gram |
| GC | gas chromatography |
| gdw | gram dry weight |
| GEIA | global emissions inventory activity |
| h | hour |
| HGT | horizontal gene transfer |
| i.e. | in other words |
| IPI | isopentenyl diphosphate isomerase |
| IPP | isopentenyl diphosphate |
| IEPOX | isoprene epoxydiol products |
| IGAC | international global atmospheric chemistry |
| IsMO | isoprene monooxygenase |
| IspS | isoprene synthase |
| kDa | kilo Dalton |
| l | litre |

| | |
|---|---|
| M | molar |
| MACR | methacrolein |
| MAMS | marine ammonium mineral salt |
| MEK | Myriam El Khawand |
| MEP | methyl erythritol 4-phosphate |
| mg | milligram |
| ml | millilitre |
| MM | minimal medium |
| MVA | mevalonate |
| MVK | methyl vinyl ketone |
| ng | nanogram |
| NLM | Nasmille Larke-Mejia |
| OD | optical density |
| OH | hydroxyl radical |
| OPL | oil palm leaves |
| OPLS | oil palm leaf swabs |
| OPLW | oil palm leaf washings |
| OPS | oil palm soil |
| OTU | operational taxonomic unit |
| PAGE | polyacrylamide gel electrophoresis |
| PCR | polymerase chain reaction |
| POZ | primary ozonides |
| ppmv | parts per million volume |
| ppt | parts per trillion |
| PrMO | propane monooxygenase |
| rpm | revolutions per minute |
| s.d. | standard deviation |
| SDIMO | soluble diiron centre monooxygenase |
| sec | seconds |
| SIP | stable isotope probing |
| SDS | sodium dodecyl sulfate |
| sMMO | soluble methane monooxygenase |
| SOA | secondary organic aerosol |

| | |
|---|---|
| SOB | super optimized broth |
| SRC | short rotation coppice |
| TDS | tyre dump soil |
| Tg | teragram |
| UEA | University of East Anglia |
| UV | ultraviolet |
| v/v | volume to volume |
| w/v | weight to volume |
| WS | willow soil |
| $\alpha$ | alpha |
| $\beta$ | beta |
| $\gamma$ | gamma |
| $\mu$ | micro |

# Chapter 1 Introduction

## 1.1 Background research

### 1.1.1 Isoprene

Isoprene (2-methyl-1,3-butadiene; $C_5H_8$) is the most abundant of the non-methane biogenic volatile organic compounds (BVOC), covering about one third of the annual global reactive BVOC flux (Guenther *et al*., 2012; Jenkin *et al*., 2015; Liu *et al*., 2016; Pierce *et al.,* 1998*)*. In Europe, isoprene is estimated to contribute 30% of total VOC emissions and 50% reactive BVOC emissions (Jenkin *et al*., 2015). Annual global emissions of isoprene have been estimated to produce 550 $\pm$ 100 Tg of C $y^{-1}$ under standard conditions, by extrapolating BVOC enclosure-type observations (Guenther *et al*., 2006; Guenther *et al*., 2012).

Over 90% of isoprene is produced by the biosphere and this can have an important effect on the atmospheric chemistry of the lower troposphere, influencing the Earth's climate (Crombie, Mejia-Florez, McGenity, & Murrell, 2018; Guenther *et al*., 2006; Pierce *et al*., 1998; Zenone *et al*., 2016). Containing four unsaturated carbons, two double bonds, isoprene is short-lived (minutes to hours) with atmospheric concentrations that vary over a day in orders of magnitude and over spatial scales of less than a few km (Guenther *et al*., 2006). Global changes in climate, drought, and land cover have shown BVOC production is likely to increase with warming and enhanced UV radiation (Peñuelas & Llusià, 2003; Peñuelas & Staudt, 2010)

### 1.1.2 Atmospheric chemistry, environmental relevance and contribution to the global climate

The potential impact of isoprene emission, reactive chemistry and oxidative capacity in the composition of the global atmosphere has stimulated extensive study of the atmospheric chemistry of isoprene (Jenkin *et al.*, 2015; Liu *et al.*, 2016). Isoprene reacts quickly and affects the chemical cycles of other atmospheric trace gases (**Figure 1.1**). Due to the reactivity of isoprene, the annual emissions of the compound are less accurate and some scientists have studied the distribution of isoprene oxidation products (e.g. formaldehyde and carbon monoxide) to estimate global isoprene emission rates (Guenther *et al.*, 2012).

Depending on the environmental conditions where isoprene is emitted, the products of isoprene photo-oxidation have the potential to affect (positively or negatively) the sources of atmospheric organic particulate matter and greenhouse gases. For example, in the forest ecosystem, isoprene is the main BVOC emitted and it plays an important role in the production of tropospheric $O_3$, organic nitrates, organic acids, formaldehyde, carbon monoxide, carbon dioxide and the generation of secondary organic aerosol (SOA) (Harley *et al.*, 2014; J. D. Surratt *et al.*, 2010).

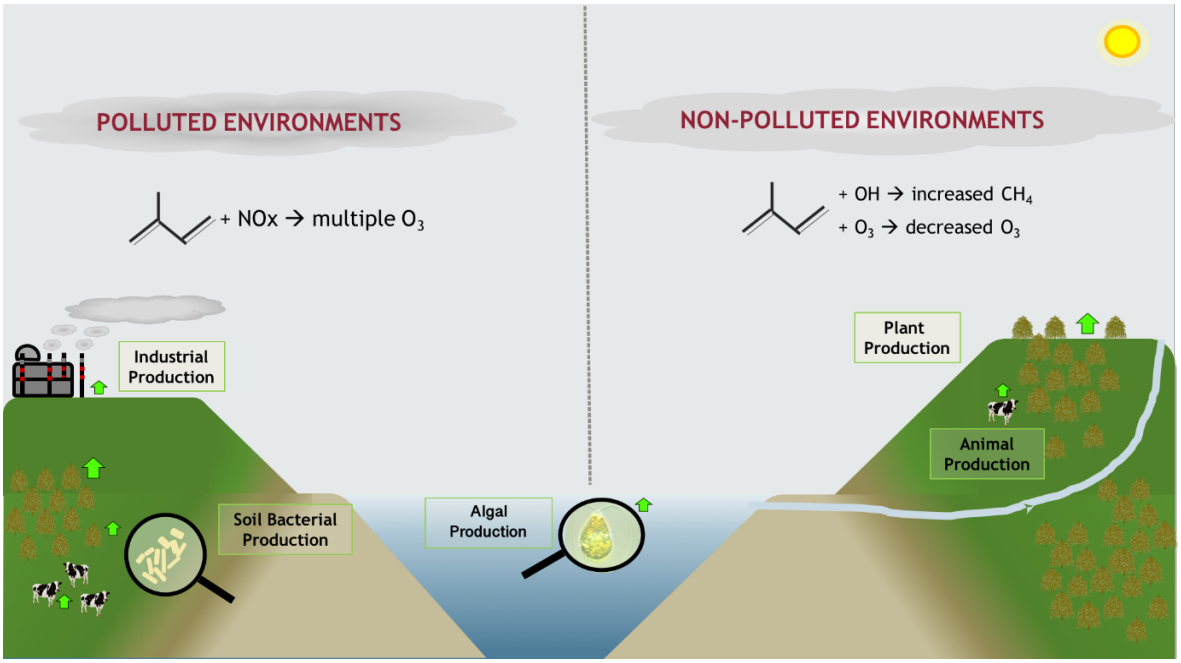



**Figure 1.1** Tropospheric fate of isoprene in polluted and unpolluted environments. The sources of isoprene emission are shown as green arrows.

Ozonolysis contributes to the removal of 10% of the emitted isoprene (**Figure 1.2;** (Nguyen *et al.*, 2016)). In the reaction, two primary ozonides (POZ) are formed from the interaction of ozone with either double bond from isoprene and can potentially form up to nine activated Criegee intermediates (CI). The CI can go through decomposition into a hydroxyl radical (OH) and a beta-oxy alkyl radical (R) to form stabilized Criegee intermediates (SCI) that react with many atmospheric species (water vapour, sulphur dioxide, formic acid, NO, $NO_2$, $O_3$, RO, alkenes, etc. (Nguyen *et al.*, 2016)). Nevertheless, the photo-oxidation of isoprene is dominated by OH radicals because its reaction with $O_3$ is relatively slow (Claeys *et al.*, 2004).

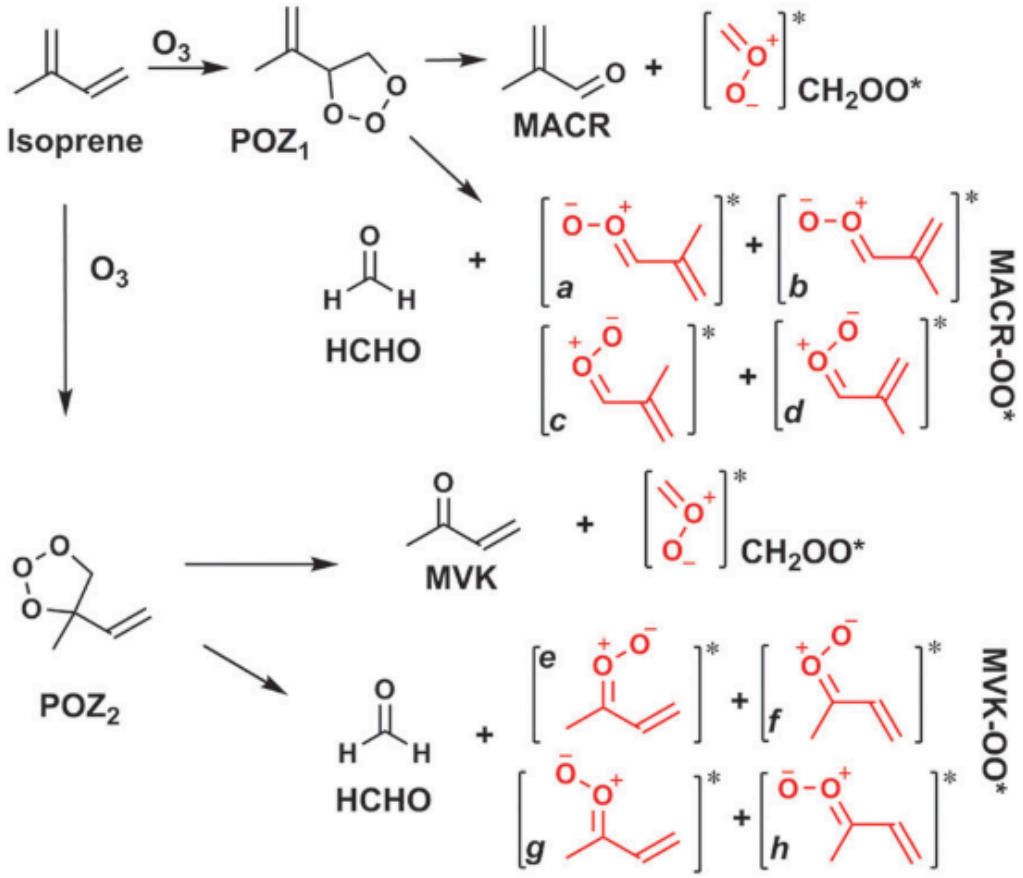



**Figure 1.2** Mechanisms of ozonolysis for isoprene. Figure taken from (Nguyen et al., 2016) *Criegee intermediates.

Field- and laboratory-based research has studied the relation of isoprene to the cycling for $HO_x$ radicals, generation of oxidised organic nitrogen species, the formation of precursors that contribute to SOA and reactions with $O_3$ (Jenkin *et al.*, 2015). Volatile methacrolein (MACR), methyl vinyl ketone (MVK), glycoaldehyde, hydroxyacetone, methylglyoxal, and formaldehyde are some of the products from the direct oxidation of isoprene (**Figure 1.3**).

Scientists have suggested the study of the oxidation products as an indirect measurement of isoprene in the environment during specific time-lapses. In recent years, the production of methyltetrols have received special attention due to their downstream products and potential precursors of SOA (Carlton *et al.*, 2009).

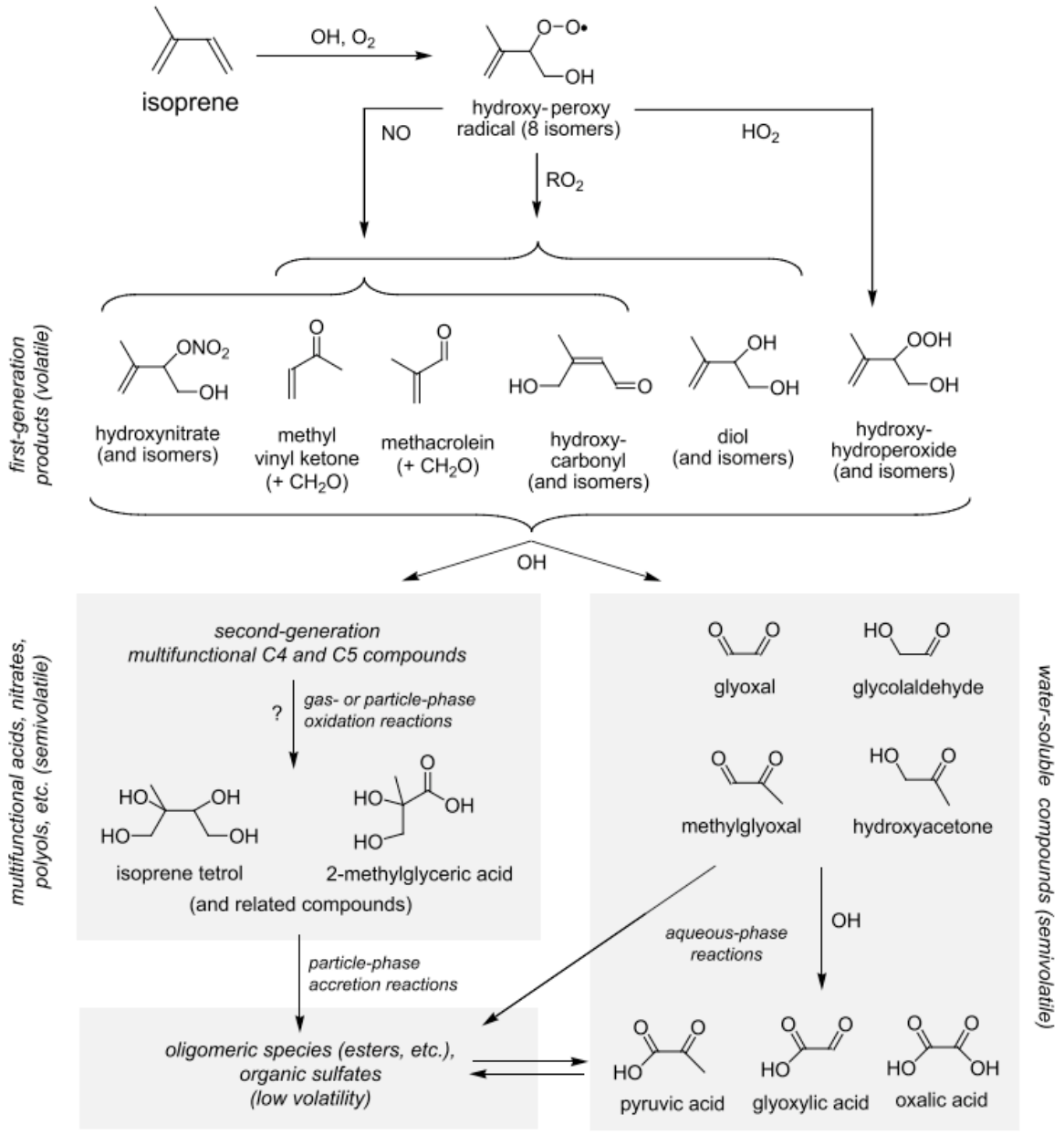



**Figure 1.3** SOA formation by isoprene oxidation pathway (Carlton et al., 2009)

The photo-oxidation of isoprene begins with the addition reaction of a hydroxyl radical (OH) across a double bond, followed by the addition of molecular oxygen ($O_2$). The reaction produces isoprene peroxyl radicals. In polluted regions, with high levels of nitric oxide (NO),

isoprene feed a cycle of NO, $NO_2$, $O_3$ which is responsible for the generation and degradation of tropospheric ozone (Calfapietra *et al.*, 2009; Kleindienst *et al.*, 2007; Pierce *et al.*., 1998). The products from the reactions are MVK and MACR which contribute to the transport of nitrogen radicals beyond the source of the isoprene through the formation of stable organic nitrogen compounds (Liu *et al.*, 2016; Zenone *et al.*, 2016).

In unpolluted environments, like the Amazonian rainforest (with concentrations of NOx that are less than 100 ppt) isoprene peroxyl radicals will react with hydroperoxyl radicals ($HO_2$ gas) (Claeys *et al.*, 2004; Surratt *et al.*, 2006). The products are an isomeric family of hydroxyl hydroperoxides, important in the production of particulate matter through the generation of isoprene epoxydiol products (IEPOX) such as 2-methyltetrols (a known source of SOA) (Ding *et al.*, 2016; Surratt *et al.*, 2006). 2-methyltetrols have low vapour pressure, are hygroscopic, and enhance the ability of aerosols to act as cloud condensation nuclei; IEPOX can produce a haze above the rainforests and other environments where the products are present (Claeys *et al.*, 2004; Li *et al.*, 2018; Surratt *et al.*, 2010). These reactions, and the buffering of atmospheric hydroxyl radicals in isoprene oxidation, have significant implications atmospheric and biosphere interactions.

In unpolluted environments with high concentrations of IEPOX, there is also an increase in solar radiation scattering, and SOA formation may be calculated in relation to three factors: organic aerosol loading, NOx level and temperature (Carlton *et al.*, 2009). Simultaneously, the diffused light increases photosynthetic activity and helps in the production of BVOCs (Gu *et al.*, 2002), which creates a cycle. In moderately polluted environments isoprene emissions may not suppress aerosol formation, which in turn would help to stabilize the atmospheric chemistry and climate over the rainforests (Kiendler-Scharr *et al.*, 2009; Taraborrelli *et al.*, 2012).

The atmospheric chemistry of the environment in which isoprene is emitted is essential in determining its reactive pathway. Although, the two pathways (ozonolysis and direct oxidation) help to maintain the atmospheric oxidation cycle and feedback from OH and $O_3$ (Liu *et al.*, 2016), Guenther algorithms only include factors such as temperature and light dependency in the emission calculations. The annual global isoprene emission is less well constrained to estimates due to its short life-time and changes in time and space

throughout the day (Guenther *et al.*, 2006). Isoprene emissions change throughout a landscape and in rural areas the emissions depend on the geography, atmospheric conditions and specific characteristics of the different source emitters (Pacifico *et al.*, 2009).

### 1.1.3 Isoprene production from natural and anthropogenic resources

#### *1.1.3.1 Sources of isoprene*

Biogenic production of isoprene by living organisms was discovered in plants by Professor Guivi Sandanze in 1957 while studying the role monoterpenes have on plant-insect interactions (Sanadze, 2004; Sharkey & Monson, 2017). While most of the work done in the last 60 years has been oriented towards understanding the mechanism of production of isoprene in plants, this hemiterpene is produced in other living organisms.

Isoprene production has been observed in Gram-positive and Gram-negative bacteria, with *Bacillus subtilis* as the most active producer during the log-phase in rich media (Ray Fall & Copley, 2000; Kuzma *et al.*, 1995; Schöller *et al.*, 1997; Wagner *et al.*, 1999). Production has also been observed in animals (Sanadze, 2004; Sharkey & Yeh, 2001); in humans it is considered an endogenous VOC, being produced at concentrations >100 ppb in resting adult human breath, as part of normal metabolic activity (mevalonate -MVA- pathway of cholesterol biosynthesis) (Deneris *et al.*1984; King *et al.*, 2010; Kushch *et al.*, 2008; Smith *et al.*, 2010; Taucher *et al.*, 1997).

Fungi like *Eurotium amstelodami* (Valencia, 1991), photosynthetic marine algae (Alvarez *et al.*2009) and cyanobacteria also emit isoprene, but the major source of biogenic isoprene is the foliage from terrestrial isoprene-emitting trees (Guenther *et al.*, 2006; Pierce *et al.*, 1998; Sanadze, 2004; Sharkey & Yeh, 2001).

### *1.1.3.2 Anthropogenic sources of isoprene*

Anthropogenic resources have also been studied as sources of isoprene, particularly in rural areas where little-to-no biogenic sources of emission are expected (Durana *et al.*, 2006; Hellén *et al.*, 2012; Hellén *et al.*, 2006; Reimann *et al.*, 2000). In urban areas, different air pollutants and compounds interact, isoprene is one of those highly active compounds. The sources of isoprene may be due to the presence of natural trees and other anthropogenic sources such as traffic and wood combustion emissions (Durana *et al.*, 2006; Hellén *et al.*, 2006; Reimann *et al.*, 2000). **Figure 1.5** shows the results when measuring the urban background using an *in-situ* gas chromatograph with a mass spectrometer in Helsinki throughout 2011 (Hellén *et al.*, 2012). Isoprene production from urban sites comes from different sources during the year, and is still produced during the winter when biogenic production is not expected. During the summer the highest rates of isoprene are observed and products can be transported into rural areas from distant sources.

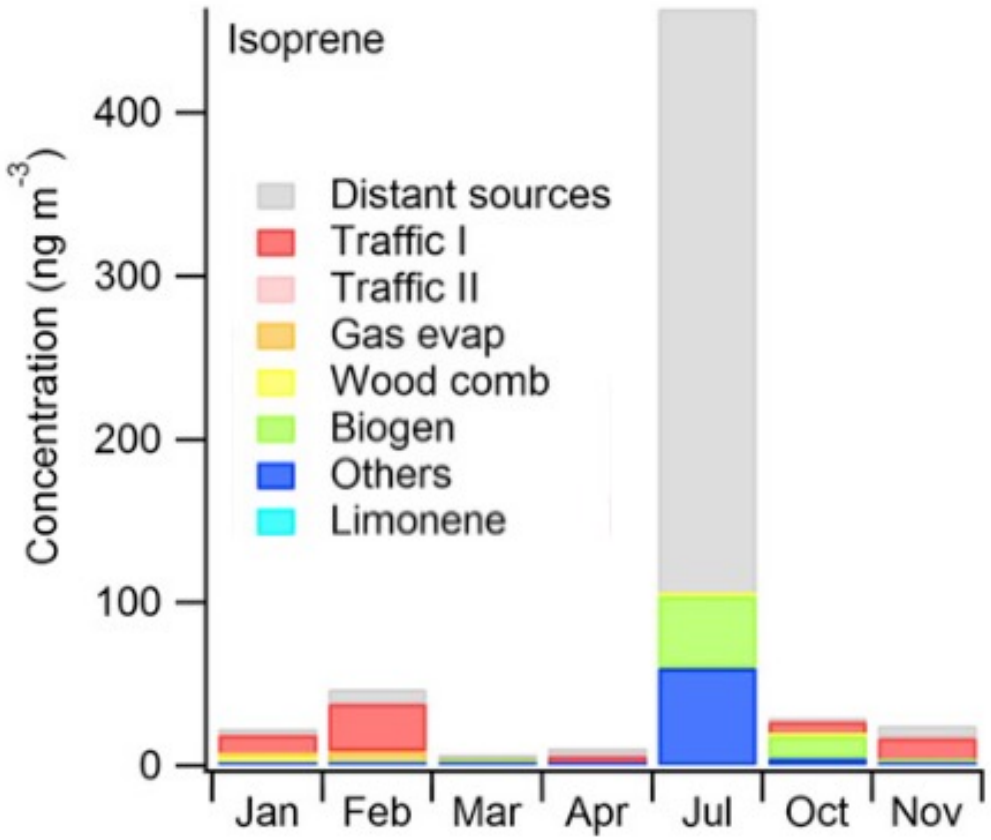



**Figure 1.4** Contribution of isoprene to the measured concentrations of urban air in Helsinki in 2011. Modified from (H. Hellén et al., 2012)

Isoprene is an hemiterpene and a key building block for isoprenoid synthesis in living organisms, including the production of carotenoids, cholesterol, chlorophyll, archaeal lipids, etc. (Crombie *et al.*, 2018). Isoprene has also been used industrially in the production of synthetic high-quality rubber and aviation fuel (Ezinkwo *et al.*, 2013; Weise *et al.*, 2013). Due to the high demand of alternative sources of renewable energy, and its potential use for the production of bio-fuels and medicine, isoprene is a compound of interest for the synthetic chemistry industry (Wang *et al.*, 2016; Whited *et al.*, 2010). Recently, its effective production though catalytic production from different raw materials has also been evaluated (Ezinkwo *et al.*, 2013; Wang *et al.*, 2016).

Some Gram- positive and Gram- negative bacteria have been found to produce isoprene (Fall & Copley, 2000). In synthetic biology, efforts to biosynthesize natural products have achieved isoprene production via the methyl erythritol 4-phosphate pathway (MEP pathway) in bacteria and cyanobacteria (Immethun *et al.*, 2013; Lindberg *et al.*, 2010; Rohmer *et al.*, 1993) and the isoprene synthase from a poplar tree has been effectively expressed in *E coli* (Miller *et al.,* 2001).

Isopentenyl diphosphate (IPP, referred to as IDP in some literature) and dimethylallyl diphosphate (DMAPP, referred to as DMADP in some literature) are the two basic building blocks and precursors for isoprenoid synthesis. The mevalonate (MVA) pathway is present in bacteria, yeast, animals, and cytosol of plants and uses acetyl coenzyme A to synthesize IPP (which is isomerized to DMAPP). The MEP pathway in bacteria, green algae, and plant plastids synthesizes IPP and DMAPP simultaneously (Zhou *et al.*, 2013). DMAPP is the only precursor for isoprene production and isoprene is the most abundant product of the MEP pathway (Zhou *et al.*, 2013). IPP is produced in the plant cytosol and it can be isomerized to DMAPP by an isopentenyl diphosphate isomerase (IPI, refer to **Figure 1.6**).

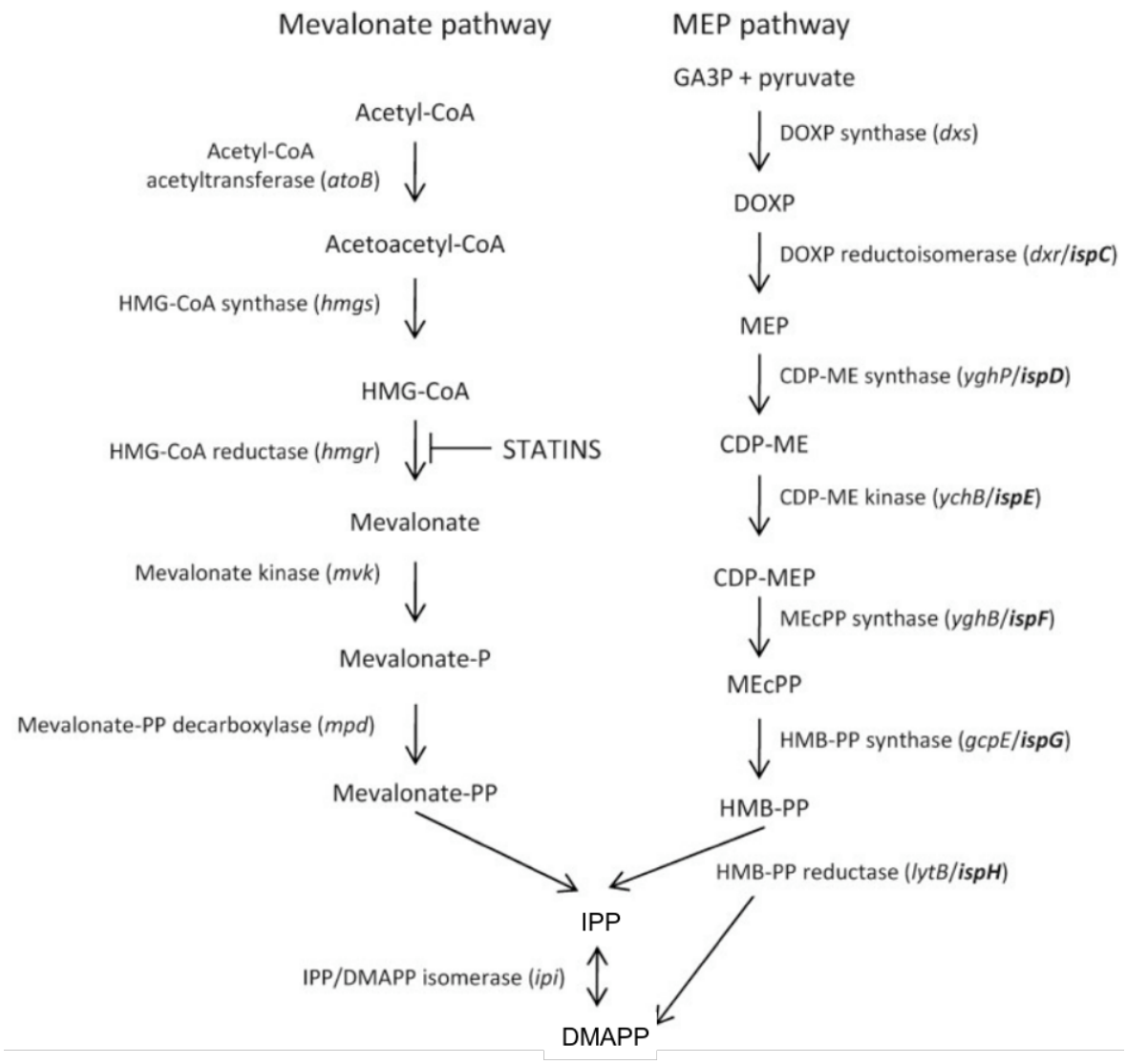



**Figure 1.5** MEP and MAV pathways for IPP and DMAPP biosynthesis. MEP genes with their historical designation and current nomenclature in bold. Image modified from Heuston *et al.*, 2012.

### *1.1.3.3 Plant bio-genic production of isoprene*

The estimation of biogenic hydrocarbon emissions relies on empirical models based on enclosure-type observations. Isoprene emission inventories and estimations are particularly challenging because not all plants emit isoprene (Fuentes *et al.*, 2000). Isoprene emission is widely spread in different taxonomic groups (mosses, ferns, gymnosperms and angiosperms) with members within each group that can emit isoprene and others that do not (Pacifico *et al.*, 2009; Pierce *et al.*, 1998; Sharkey *et al.*, 2008). It has been estimated that around 20% of the plant species on earth emit isoprene (Sanadze, 2004). Although DMAPP is usually produced via the MEP pathway, it can also be produced via the

mevalonate pathway (refer to **Figure 1.5**, (Crombie *et al.*, 2018; Rohmer *et al.*, 1993; Sharkey & Yeh, 2001)). Plants produce the hemiterpene by the action of the isoprene synthase (IspS) on DMAPP. The isoprene synthase is found in a limited number of higher plant clades (Sharkey & Yeh, 2001). Curiously the gene sequences for the isoprene synthase do not support a single origin of the trait and the way the trait is distributed in photosynthetic organisms is not clear (Hanson *et al.*1999; Sharkey & Monson, 2017).

At the landscape level, emissions differ from the individual emission potentials of each tree in a rainforest. Changes in conditions throughout the day, location of leaves in a canopy (amount of radiation) and frequent low light conditions may significantly alter the potential isoprene emission of a particular tree (Monson *et al.*, 1992; Pierce *et al.*, 1998). Klinger and colleagues divided isoprene emitting plants along a savannah rainforest in central Africa into three categories (high-, low- and non- emitters) and suggested that isoprene emissions from each organism can vary according to the successional status of the ecosystem (Klinger *et al.*, 1998). Isoprene emitters produce $\geq$0.005 μg isoprene cm $^{-2}$ h$^{-1}$ (Klinger *et al.*, 1998). Plant species, which emit isoprene, include high emitters such as *Elaeis guineensis* (oil palm tree), *Salix alba* (willow tree) and *Berlinia grandifolia*, *Afrormosia laxiflora*, and low emitters include *Mangifera indica*, *Fraxinus* spp. (Ash tree) and *Glycine max*. Some non-emitting plants include *Carica papaya* (papaya), *Coffea* sp. (coffee) and *Musa sapientum* (banana) (Hewitt & Street, 1992; Klinger *et al.*, 1998).

### *Environmental factors that influence isoprene production in plants*

Since the discovery of isoprene emission as an individual and free substance by Sanadze, confirmed later by Rasmussen in the 1970's, numerous studies have attempted to determine the environmental influence, metabolic machinery behind isoprene production and how specific factors (biotic and abiotic) can effect isoprene emission (Rasmussen, 1970; Sanadze, 2004; Sharkey & Monson, 2017). The highest rates of isoprene emission are achieved with light ('the isoprene effect") and at high radiation incidence, independent from stomatal aperture (Zenone *et al.*, 2016), and related/dependent on photosynthesis (Loreto & Sharkey, 1990; Sharkey & Yeh, 2001).

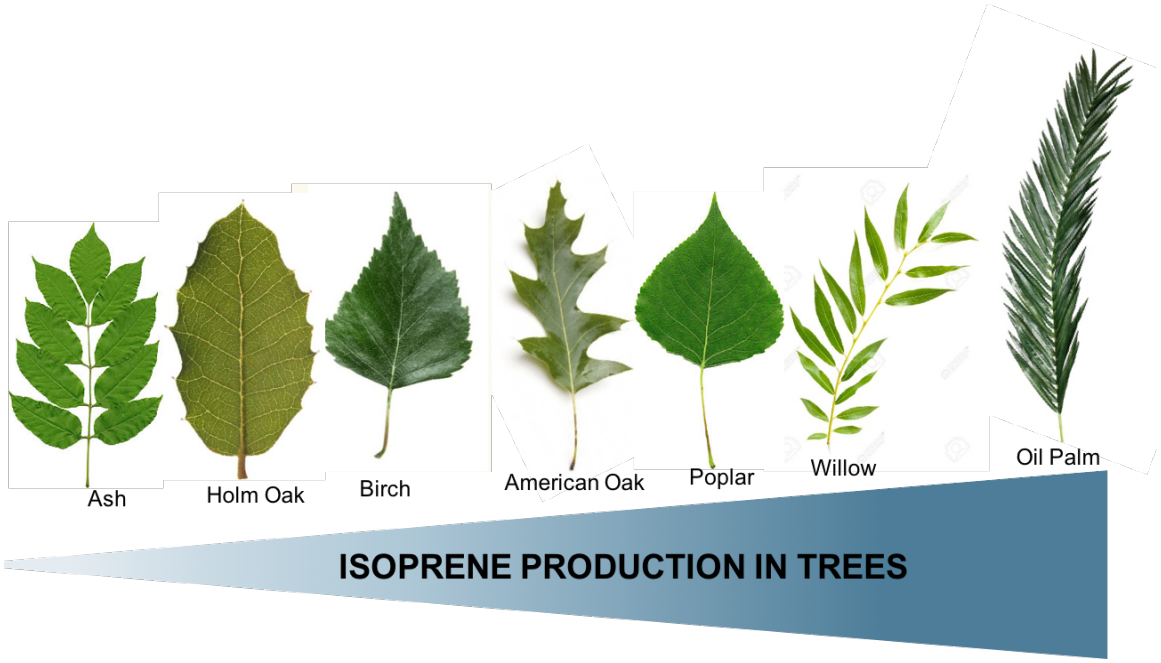



**Figure 1.6** Isoprene emitting trees and estimated range of isoprene emission based on their isoprene emission potential.

In fact, 1-2% of the photosynthetically assimilated carbon is allocated to the production of isoprene in natural environment conditions (temperate climate) (Guenther, 1995; Harrison *et al.*, 2013; Peñuelas *et al.*, 2013). Isoprene is synthesized on the thylakoid membrane of the chloroplast (Datukishvili *et al.*, 2001; J. Kuzma & Fall, 1993; Loreto & Sharkey, 1993; Wildermuth & Fall, 1996) and is released principally through the stomatal pores in leaves (Fall & Monson, 1992). The production is also dependent on $CO_2$ and $O_3$ concentrations (Fares *et al.*, 2006; Harrison *et al.*, 2013; Loreto & Velikova, 2001; Loreto *et al.*, 2001; Wang *et al.*, 2016), drought stress (Fortunati *et al.*, 2008) and even the circadian rhythm of the plant (C. N. Hewitt *et al.*, 2011). Inside the plant cell the factors that influence enzyme activity, substrate availability and gene expression also affect isoprene synthesis (Monson *et al.*, 2012). The principle environmental factors that affect emission in the short term are leaf temperature, photosynthetic photon flux density and intercellular $CO_2$ concentration, $O_2$ concentration. In the long term, the weather, soil water availability, atmospheric $CO_2$ concentration, leaf position in the canopy and developmental stage of the leaf are important (Niinemets, Arneth, *et al.*, 2010; Niinemets, Monson, *et al.*, 2010; Sharkey & Yeh, 2001).

The percentage of carbon allocated to isoprene emission from photosynthesis can increase up to 50% under stressful conditions (Loreto & Schnitzler, 2010; Pegoraro *et al.*, 2004).

Multiple studies have intended to prove if isoprene production is dependent on leaf temperature (Medori *et al.*, 2012). Sharkey and Singsaas argument is that increasing production of isoprene at temperatures over 38 °C protects the chloroplast membrane structure by making it more resistant and able to continue with photosynthesis at high temperatures (Sharkey & Singsaas, 1995; Sharkey & Yeh, 2001; Singsaas *et al.*, 1997; Zenone *et al.*, 2016) and under reactive oxygen species stress (Behnke *et al.*, 2007; Velikova *et al.*, 2011). But adding isoprene to a non-emitting plant does not increase their thermo-tolerance (Logan & Monson, 1999; Peñuelas & Llusià, 2004). Isoprene production may be an important mechanism for plants that are subjected to heat, high light radiation and drought tolerance (Bamberger *et al.*, 2017; Sharkey & Loreto, 1993), stabilizing the thylakoid membrane (Velikova *et al.*, 2012)

Studies on circadian control of isoprene emission capacity have shown that the emission capacity of a plant tends to increase in the morning, peak in the afternoon and decline in the late afternoon. This was tested with *Elaeis guineensis* in continuous light and a temperature-compensated period of 24 h (Wilkinson *et al.*, 2006) and later testing the expression of *IspS* transcript as well (Loivamäki *et al.*, 2007; Mayrhofer *et al.*, 2005). Silver and Fall identified and purified the membrane bound isoprene synthase (IspS), which has light-dependent activity in leaves of vascular plants. They confirmed the enzyme had the ability to synthesize isoprene from DMAPP (according to its availability), by eliminating a pyrophosphate (Silver & Fall, 1991). Wildermuth and Fall then discovered that the enzyme is the membrane-bound IspS and concluded isoprene synthesis is attributed to the thylakoid membrane of the chloroplasts (refer to **Figure 1.7**, (Wildermuth & Fall, 1996)).

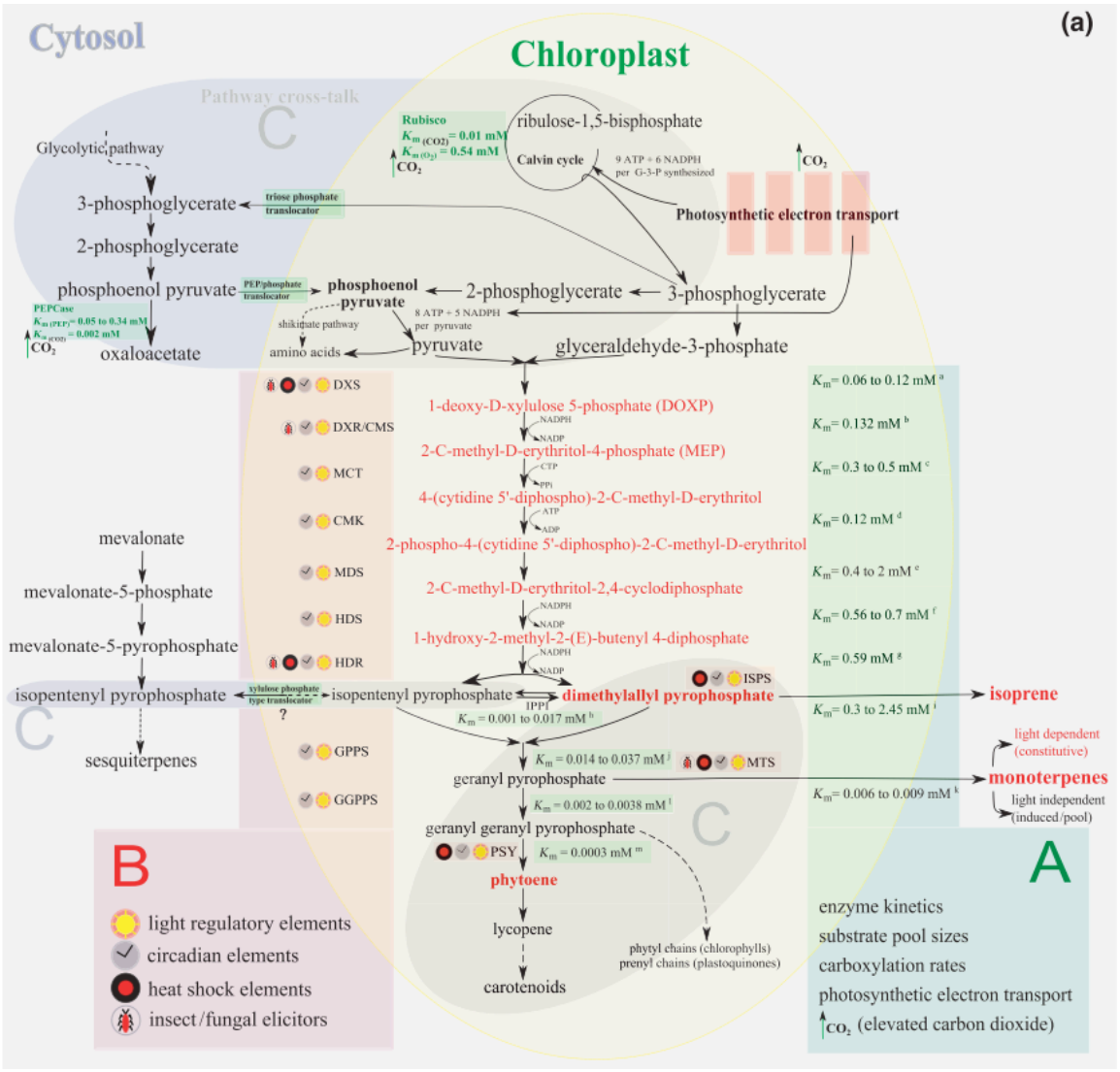



**Figure 1.7** Metabolic routes involved in plant production of isoprene and other monoterpenes (Harrison *et al.*, 2013).

### *High-emitters and climate change*

Changes in global climate and land-use will have an impact on the rates of isoprene emission, especially in areas with incidence of multiple abiotic stress factors. Heat waves and soil water deficit might be more frequent in the future and the emissions of BVOCs are expected to change with the intensity of extreme weather conditions (Coumou & Rahmstorf, 2012; Peñuelas & Staudt, 2010). Although some isoprene models have been developed, i.e. the International Global Atmospheric Chemistry (IGAC) and the Global Emissions Inventory Activity (GEIA), the simulations have included estimations for

temperate regions and leave uncertainty about isoprene emissions in the tropics (Chen *et al.*, 2018; Guenther, 1995).

Oil palm plantations around the world have vastly increased in the last decade due to the increasing demand for products such as palm oil, oleo-chemical products, palm kernel oil, combustion fuels, biomass, etc. Furthermore, isoprene emission above oil palm were on average 10 times higher than those over the rainforest (Misztal *et al.*, 2010). Indonesia, Malaysia, Nigeria, Congo, Brazil, Colombia and Ecuador are the major producers of oil palm worldwide, with the first two countries as primary producers (Corley & Tinker, 2016). Some of the products are used by consumers daily and increase the demand of oil palm products. Since oil palm is also one of the highest isoprene emitters reported to date, the study of the effects of large monocultures have been studied (refer to the studies by (Chen *et al.*, 2018; Evans & Griffiths, 2013; Koh & Wilcove, 2008; Ocampo-Peñuela *et al.*, 2018; Paterson & Lima, 2018; Schnitzler & Nouvellon, 2018; Wilkinson *et al.*, 2006) for more detail).

Isoprene emitting trees will become more abundant in the future. The European Union aims to replace 10% of transportation fuel and a proportion of power generation fuel with biomass-derived fuels by 2020 (Ashworth *et al.*, 2013). To do this there has been an increase in the cultivation of biofuel feedstock crops, afforestation, and replacement of present agricultural crops and grassland in Europe with short rotation coppice (SRC), biofuel feedstock is then used for power generation (converted into lingo-cellulosic ethanol). Planting 72 Mha of SRC species (poplar, willow or eucalyptus) will result in an increase of 11.5 Tg C $y^{-1}$ to 16.0 Tg C $y^{-1}$ (approximately 1%) of isoprene emissions and will lead to enhanced ozone formation due to high NOx in Europe, affecting mortality rates (Ashworth *et al.*, 2013). Other authors estimate the global burden of isoprene to increase 8 % by 2100 (Sharkey & Monson, 2014).

Passing from vegetation to cropland and deforestation will decrease isoprene emissions by the reforestation using SRC, which have high BVOC emissions, will exceed those of forests. (Rosenkranz *et al.*, 2015) Land use will lower isoprene levels but rising temperature will counteract by increase it. Finally, isoprene production is inversely correlated to $CO_2$ concentration and may affect emissions in the future (Way *et al.*, 2011) and be less adaptive for plants in an high $CO_2$ atmosphere. With increasing plantations of high isoprene emitters

such as oil palm, it is extremely important to understand the cycling of isoprene in the terrestrial environment associated with isoprene-emitting crops. Isoprene-degrading microorganisms play an important role, as biological sinks that are found in the natural environment.

### 1.1.4 Isoprene sinks and biodegradation

After studies have showed that soil can be a significant sink for atmospheric trace compounds such as methane, methyl bromide and monoterpenes, studies towards the contribution of soils as a sink of atmospheric isoprene began (Cleveland *et al.,* 1997). Twenty years ago, an indication of the distribution of isoprene oxidizing microorganisms in soil was shown by the capacity of various samples to take up the compound from the gas phase of soil microcosms (van Ginkel *et al.*, 1987). In fact, it was hypothesized that different soil samples from diverse origins, and soil types, have different rates of isoprene uptake (Cleveland *et al.*, 1997).

In the soil, bacteria were able to remove isoprene at trace concentrations in a short time (a couple of hours) (Cleveland & Yavitt, 1998; van Ginkel *et al.*, 1987). Scientists have estimated the abundance of isoprene-degrading microorganisms to be approximately $10^5$ cells per gram dry weight of soil (Cleveland and Yavitt, 1998). Consumption of atmospheric isoprene by soil microorganisms may be a significant component of the global isoprene budget as soils might remove as much as 5% of global emitted isoprene on an annual basis (Cleveland & Yavitt, 1998; Cleveland *et al.*, 1997). Bacterial isoprene consumption was also dependent on the initial concentration of isoprene used, temperature, oxygen availability and the soil characteristics (amount of organic matter, depth, moisture, and pH) (Cleveland & Yavitt, 1998).

In early studies *Xanthobacter autothropicus* Py2 was tested for its ability to oxidize isoprene, but did not grow on the compound (van Ginkel *et al.,* 1986). Isoprene-degrading bacteria, able to use isoprene as a source of carbon and energy, were isolated and identified as Actinobacteria from the *Nocardia*, *Arthrobac*ter and *Rhodococcus* genera and the proteobacterium *Alcaligenes* (Cleveland & Yavitt, 1998; Cleveland *et al.*, 1997; Ewers *et al.*, 1990; van Ginkel *et al.*, 1987). The presence *Alcaligenes* sp. *Klebsiella* sp. and

*Pseudomonas* sp. were detected as isoprene bio-degraders in a continuous batch bioreactor using contaminated soil from a waste rubber dumping site (Srivastva *et al.*, 2015). The potential use of *Methylobacterium* sp. PV1 as an isoprene-degrader has also been evaluated by the same group (Srivastva *et al.*, 2017); but there are no reports, to date, on the mechanisms used by these last four organisms to degrade isoprene.

In marine sediments, estuarine samples and fresh water samples, a variety of bacteria, including some *Rhodococcus* species, *Gordonia* sp. i37 and *Mycobacterium* sp. AT1, have been isolated, identified and characterized as isoprene-degraders (Alvarez *et al.*, 2009; Johnston *et al.*, 2017). Our team has recently found *Rhodococcus* species in the soils and leaves of isoprene-emitting trees (Crombie *et al.*, 2018; El Khawand *et al.*, 2016). A variety of *Rhodococcus* species have been isolated from multiple studies from soils (Ewers *et al.*, 1990; Murphy, 2015; van Hylckama Vlieg *et al.*, 1998), freshwater sediments and marine sediments associated with plants (Alvarez *et al.*, 2009; Johnston *et al.*, 2017) making it an important genus of isoprene-degraders in both terrestrial, phyllosphere and aquatic environments.

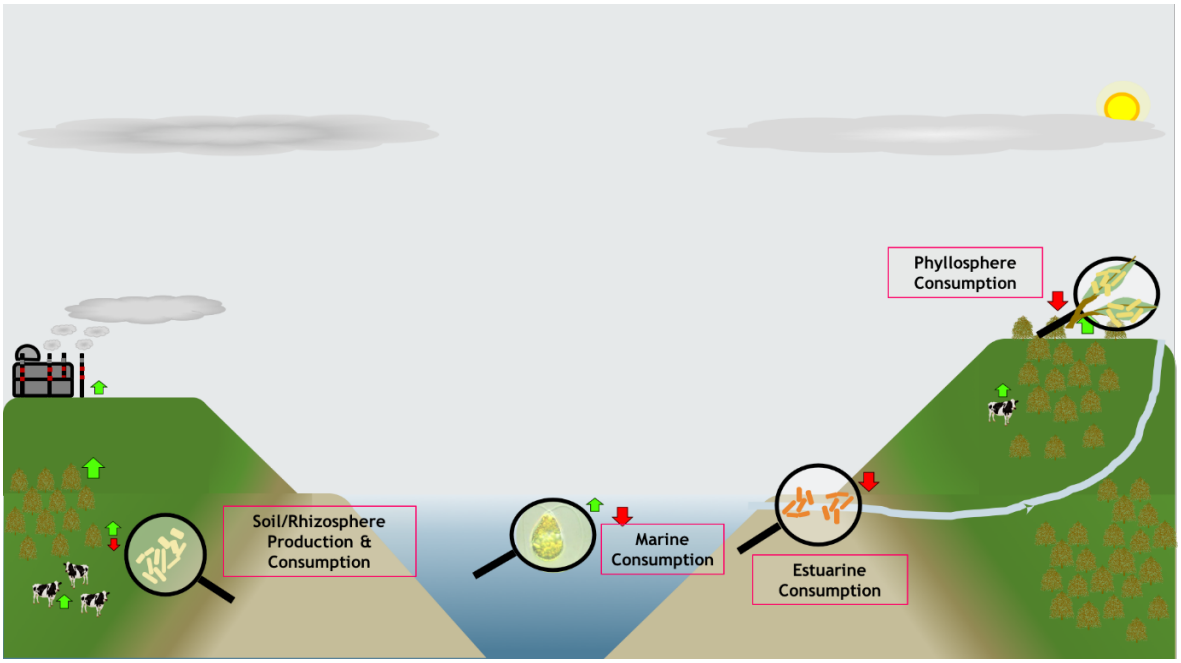



**Figure 1.8** Potential isoprene sinks in the environment. Bio-degradation of isoprene in aquatic and terrestrial samples such as soil and the phyllosphere of isoprene-emitting plants (red arrows). Isoprene sources of emission into the atmosphere shown as green arrows.

Plants are populated by microorganisms both below and above ground (Vorholt, 2012). After confirming the identification of isoprene as a compound emitted by plants (Rasmussen, 1970). Rasmussen hypothesized that the microbes in the phyllosphere and rhizosphere of tropical vegetation are a major sink for the naturally occurring organic volatiles. These volatile organics, now known as BVOCs, are a source of carbon and energy necessary for the metabolism and growth of microorganisms (Rasmussen & Hutton, 1972).

Although the plant rhizosphere has been widely studied, in recent years the phyllosphere has received more attention (Vorholt, 2012). The aerial part of plants is inhabited by qualitatively and quantitatively diverse microorganisms called epiphytes. It has been estimated that up to $10^7$ cells/$cm^2$ are present (including bacteria, fungi and eukaryotes, (Hirano & Upper, 2000)). Cultivation independent methods have shown larger and more complex communities than have been found only with culture dependent methods (Yang *et al.*, 2001).

Interactions of various epiphytic bacteria with plants suggest more than a commensalistic relationship between the phyllosphere and their host (Lindow & Brandl, 2003). Microbial epiphytes are exposed to a plants daily cycle and are influenced by the day to day changes such as the effect of direct sunlight and may also be affected by changes in plant metabolism indirectly (Vorholt, 2012). Phyllosphere populations can also change in size and proximity in the same plant species over short time scales and changing seasons (Lindow & Brandl, 2003). The physical environmental condition and availability of immigrant inoculum is important in determining the species of microbes present in the phyllosphere (Yang *et al.*, 2001).

Epiphytic bacteria depend on the composition and quality of nutrients (carbohydrates, organic acids, amino acids) which are affected by leaf age, leaf position, leaf physiological status and tissue damage (Yang *et al.*, 2001). Physical conditions like moisture emitted through stomata, the availability of carbon-containing nutrients is a major determinant of

colonization (Lindow & Brandl, 2003) and bacteria living in the phyllosphere exhibit biogeographic patterns since interspecies variability exceeds intraspecies variability (Redford *et al.*, 2010). The phyllosphere is a dynamic environment that has large changes in temperature, UV radiation, relative humidity and leaf wetness (Lindow & Brandl, 2003; Redford *et al.*, 2010; Whipps *et al.*, 2008). Precipitation and wind contribute to temporal variability in immigration and emigration of resident phyllosphere microbes (Lindow, 1996).

Plant leaves with different genotypes can release volatile organic compounds and metabolic substrates, including isoprene, $C_1$ compounds ( organic compounds containing one carbon such as: $CO_2$, methyl chloride and methanol), acetone, terpenoids, aldehydes and alcohols, into the boundary layer around leaves (Vorholt, 2012; Whipps *et al.*, 2008). Counteractively some of these compounds produced by different parts of the plant, can promote microbial development but can selectively inhibit microbial growth as well (Dingman, 2000), supporting different microbial communities.

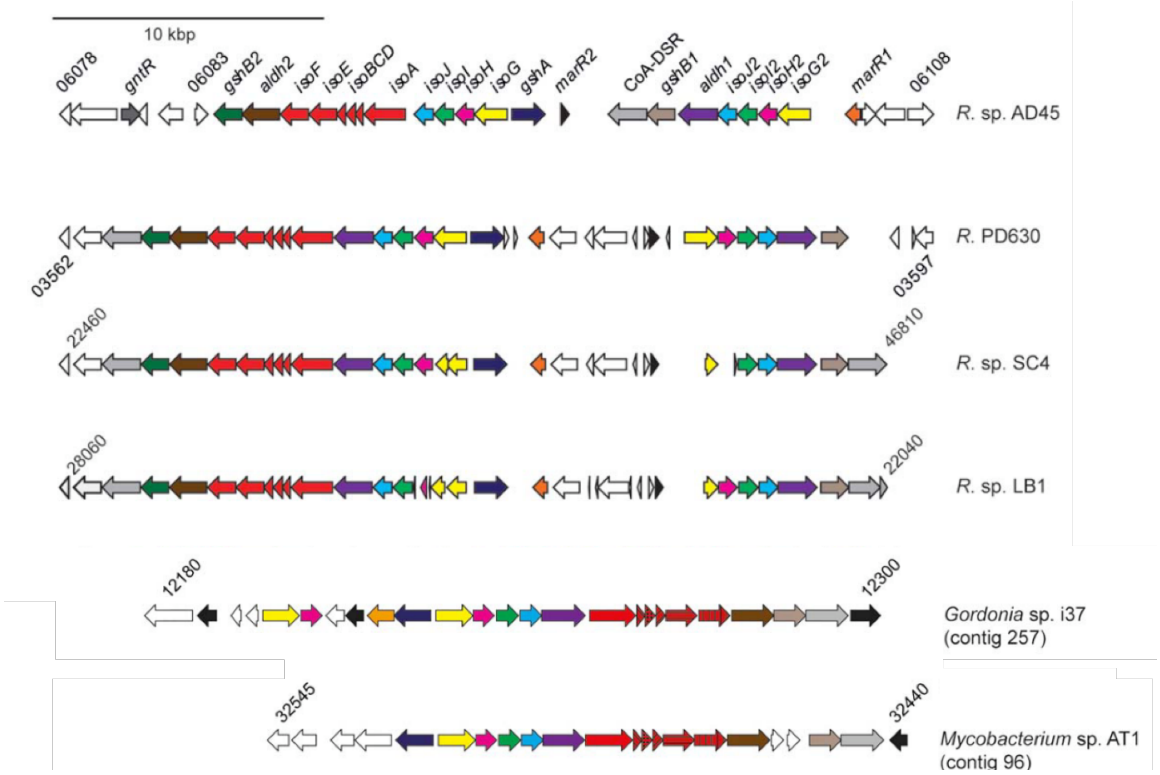
The planetary phyllosphere bacterial population may be as large as $10^{26}$ cells, covering an approximate leaf surface area of at least $10^8$ $km^2$ globally (Morris & Kinkel, 2002).The planetary phyllosphere population thought to be large enough to contribute to many processes of global importance, including effects on the behaviour of individual plants (Lindow & Brandl, 2003). Phyllosphere microbes have the potential to affect plant health (Turner *et al.*, 2013) and interact closely with compounds important in atmospheric chemistry (Vorholt, 2012).

#### *1.1.4.1 The putative isoprene degradation pathway*

A putative biochemical pathway has been characterised in the bacterium *Rhodococcus* sp. AD45, a Gram-positive strain of the phylum Actinobacteria, isolated by van Hylckama Vlieg and workers from an isoprene enrichment culture of freshwater (J. E. van Hylckama Vlieg et al., 1998). It was suggested that 1,2-epoxy-2-methyl-3-butene is the first intermediate in the isoprene biodegradation pathway (van Ginkel et al., 1987). Studies have shown that *Rhodococcus* sp. AD45 produces 3, 4-epoxy-3-methyl-1-butene as the first product in the isoprene degradation pathway. This product is metabolized by a glutathione S-transferase

(GST) encoded by *IsoI*. This is a critical step because it removes the toxic epoxide that had been shown to alkylate DNA and proteins (Johan, Hylckama, Kingma, & Janssen, 1999).

A




B

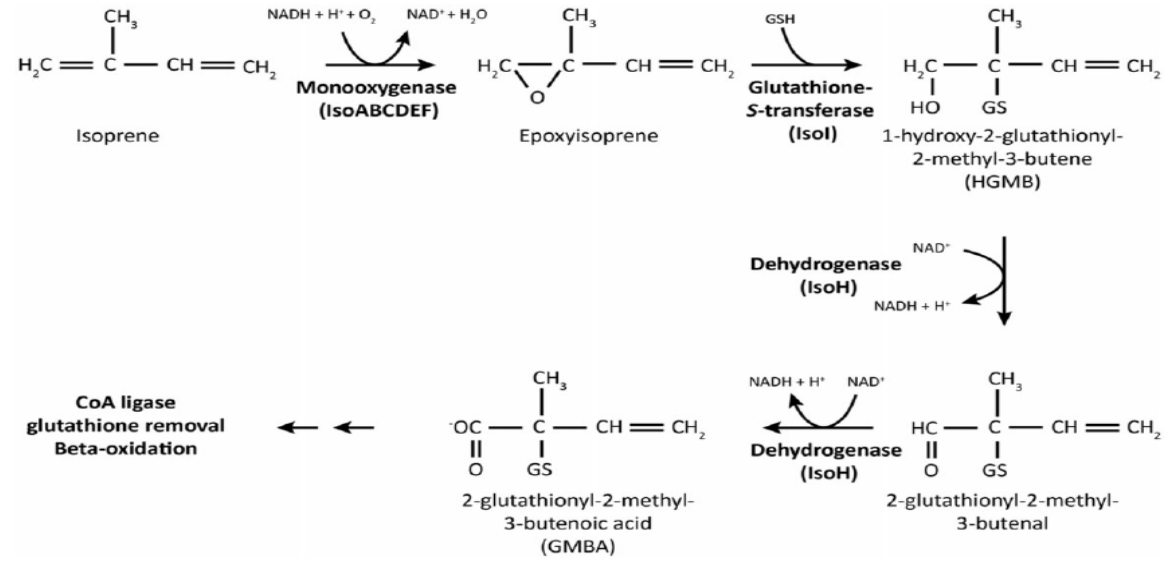



**Figure 1.9** Isoprene metabolism genes and pathway. A) Isoprene metabolic gene cluster in isoprene-degrading bacteria, including the isoprene monooxygenase genes are coloured red, and other genes are colour coded according to their corresponding predicted functions. *R*. PD630 is *Rhodococcus opacus* PD630, *R* sp. SC4 and *R*. sp. LB1 are two *Rhodococcus* species, *Gordonia* sp i37 and *Mycobacterium* sp. AT1 (El Khawand et al., 2016; Johnston et al., 2017). (B) Putative pathway of isoprene metabolism, the later steps of the pathway have not been confirmed (Crombie et al., 2015).

*Rhodococcus* sp. AD45 was used by Van Hylckama Vlieg and colleagues to construct a gene library which led to the identification of an 8.5 kb gene cluster proposed to be involved in isoprene degradation (van Hylckama Vlieg *et al.*, 2000). The sequence encodes a putative multicomponent monooxygenase that is similar to the toluene monooxygenase from *Pseudomonas mendocina* KR1 (*isoABCDEF*) and the alkene monooxygenase from *Xanthobacter autotrophicus* Py2 (Crombie *et al.*, 2015; van Hylckama Vlieg *et al.*, 2000), two glutathione S-transferases (*isoI, isoJ*), a dehydrogenase (*isoH*), and a racemase (*isoG*), shown in **Figure 1.9A**.

Crombie *et al.* showed that *Rhodococcus* sp. AD45 growth is inhibited by isoprene concentrations over 2% v/v, isoprene uptake is repressed by the presence of an alternative carbon source, and the isoprene monooxygenase is essential for isoprene metabolism (mutated strain for the hydroxylase alpha-subunit showed no growth on isoprene). (Crombie *et al.*, 2015). Transcription analysis demonstrated an increase in the transcription of 22 genes, including the isoprene metabolic gene cluster (06083 to *isoG2*), when adding isoprene or epoxyisoprene to cultures (**Figure 1.9A**).

Enrichment of terrestrial, marine and estuarine samples have shown mostly Gram- positive Actinobacteria as the active and abundant isoprene-degrading bacteria, with some Gram-negative alpha-proteobacteria in marine and estuarine samples (El Khawand et al., 2016; Johnston et al., 2017) *Rhodococcus* sp. AD45, *Rhodococcus* sp. SC1, *Rhodococcus* sp. LB1, *Rhodococcus opacus* PD630, *Gordonia* sp. i37 and *Mycobacterium* sp. AT1 have an almost identical organization of the ten genes from the isoprene metabolic gene cluster (*isoABCDEFJHIJ*, see **Figure 1.9A**), except for *aldH* gene between *isoA* and *isoJ* for *Rhodococcus* sp. AD45 (Johnston *et al.*, 2017). The isoprene monooxygenase, that catalyses the first step in the metabolic pathway (**Figure 1.9B**), has homology to enzymes of the soluble diiron centre monooxygenase (SDIMO) family (van Hylckama Vlieg *et al.*, 2000). This family of enzymes, of which the soluble methane monooxygenase (sMMO) is a well-known example (Holmes & Coleman, 2008), has been extensively studied for its role in hydrocarbon oxidation.

The SDIMO family includes the sMMO, toluene monooxygenase, phenol hydroxylase and alkene monooxygenases. They are soluble, multicomponent enzymes, which catalyse the incorporation of a dioxygen into specific organic and inorganic substances. There is a great degree of similarity between the organization of the genes that encode the functional proteins and functionality of enzymes in different groups (Leahy *et al.*, 2003). Most SDIMO proteins have 3 or 4 components that include a dimeric hydroxylase, an NADH oxidoreductase, an effector (or coupling) protein and, in some cases, a Rieske-type ferredoxin protein. Primers have been designed to evaluate SDIMO diversity in soils, sediments and enrichments (Coleman *et al.*, 2006)

Activation of molecular oxygen is achieved via iron, copper, flavin or pteridine-based cofactors in different monooxygenase classes. Monooxygenases have substrate ranges, and some are very stereo-selective, producing epoxides in high enantiomeric excess (Cheung *et al.*, 2013; Owens *et al.*, 2009). All di-iron monooxygensases characterised are believed to contain a channel of hydrophobic amino acids leading to the active site that are implicated in substrate recognition, binding and region specificity (Leahy *et al.*, 2003). sMMO methane monooxygenase plays an important role in cycling of carbon in the biosphere and has a broad substrate range ($C_1$ to $C_5$-$C_{10}$ (Van Beilen & Funhoff, 2007)). In the case of the isoprene, the isoprene-oxidizing ability is not restricted to the isoprene monooxygenase (IsoMO). *Mycobacterium* strain NBB4 ethene-inducible monooxygenase ETnABCD can oxidize isoprene producing a styrene 7,8 –oxide (Cheung *et al.*, 2013). *Xanthobacter autothropicus* Py2 can oxidize isoprene using its propane monooxygenase (PrMO, (Johnston *et al.*, 2017; van Ginkel *et al.*, 1986)). These overlapping roles suggest that in environments with a mixture of gases non-isoprene-assimilators are present expressing enzymes that can catalyse a broad range of compounds (Johnston *et al.*, 2017).

After the different monooxygenases catalyse the first step, a highly reactive epoxide is produced. The epoxide is toxic to cells and bacteria have developed complex systems for epoxide detoxification. *Mycobacterium* E20 required coenzyme-A and an unknown

cofactor for the transformation of ethylene oxide to acetyl CoA (Bont *et al.*, 1979; Shennan, 2006). *Xanthobacter autothropicus* Py2 used coenzyme M transferase (Zn dependant epoxidase) to open the peroxide ring resulting in propylene oxidation (Krum & Ensign, 2001; Shennan, 2006) For *Rhodococcus* sp. AD45 a glutathione S-transferase catalyses the opening of the epoxide ring from the 1,2-epoxy-2-methyl-3-butene resulting from isoprene oxidation (Johan *et al*., 1999) .

### 1.1.5 DNA Stable Isotope Probing (SIP)

DNA-Stable Isotope Probing (DNA-SIP) is a technique that is based on the incorporation of stable isotopes (in this case carbon) into newly synthesized DNA of microorganisms. The microorganisms are incubated with the specific isotope-labelled 'heavy' substrate ($^{13}C$), as their sole source of carbon, to later identify the DNA of microorganisms that have actively used the substrate (Dumont & Murrell, 2005; Neufeld, *et al.*, 2007; Radajewski *et al.*, 2000). The organisms that have used the heavy substrate will now have 'heavy' DNA that can be separated from the DNA of organisms ($^{12}C$-DNA) that did not use the substrate using buoyant density centrifugation. Both, heavy and light DNA fractions, are then analysed by methods including PCR amplification using 16S rRNA gene primers and analysis by denaturing gradient gel electrophoresis (DGGE) (Friedrich, 2006; Grob *et al.*., 2015) Changes between the band patterns in the gel will then confirm the efficient separation of the heavy and light DNA. Organisms can later be identified from the DNA in each fraction and it is possible to determine which organisms are the active degraders of the specific compound.

Incorporation of the $^{13}C$-label into DNA of isoprene degraders enables the study of the dynamics of isoprene-degrading bacteria present in this environment under different enrichment conditions. To study isoprene-degradation, DNA-SIP has been used to identify $^{13}C$-isoprene assimilators in the terrestrial and marine environments, and to show a greater variety of isoprene degraders (Crombie *et al.*, 2018; El Khawand *et al.*, 2016; Johnston *et al.*, 2017). In the water and sediment samples, from Colne Estuary, DNA-SIP with 0.2% v/v isoprene revealed the Actinobacteria of the genera *Mycobacterium* (fluctuating between 99% to 11-16% during their enrichment) and *Microbacterium* (from 0.4 % to 36-43% at the

end of their enrichment) as the active and abundant isoprene degraders after 12 and 15 days of enrichment. The genera *Gordonia* (relative abundance of 7%) and *Rhodococcus* (relative abundance of 31%) were found after 15 days of enrichment showing that the diversity of isoprene degraders increased with time (Johnston *et al.*., 2017).

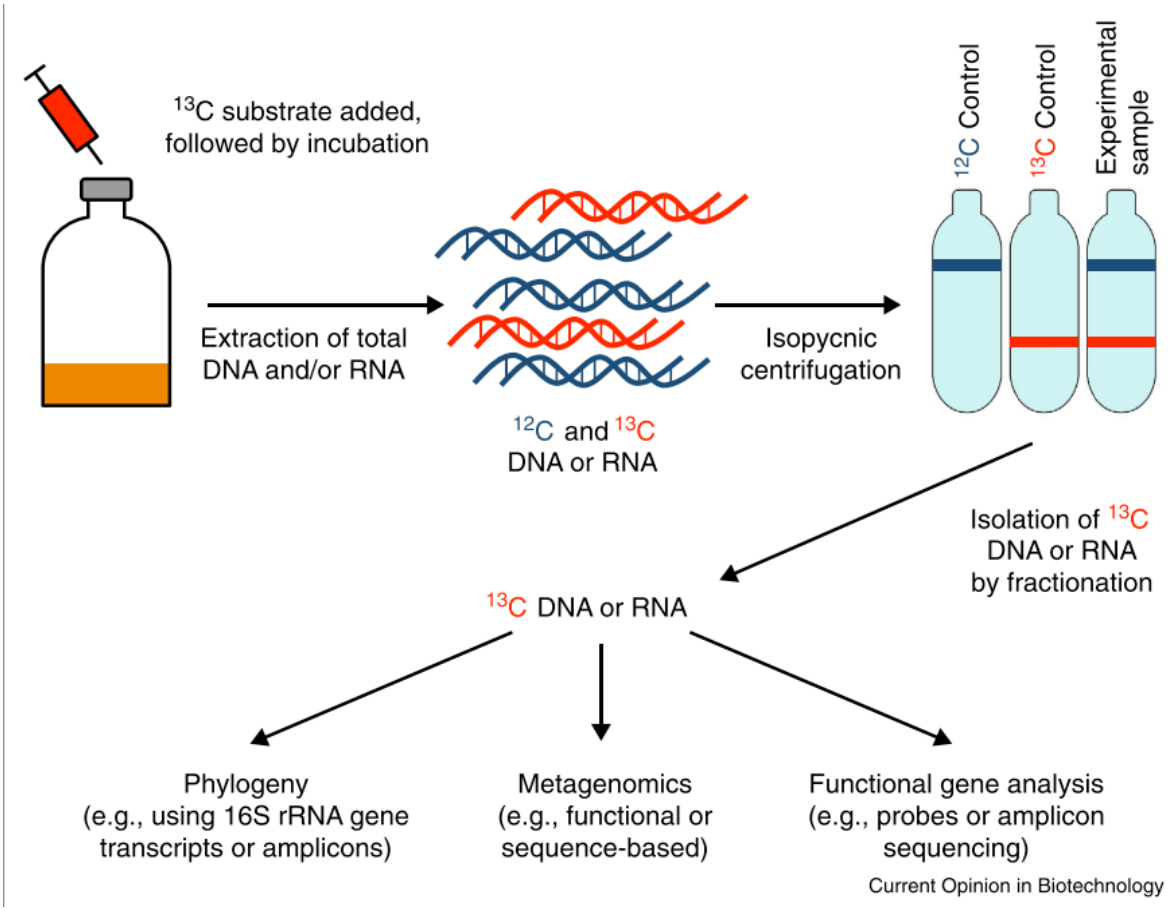



**Figure 1.10** Nucleic acid stable-isotope probing downstream sequence analysis (Coyotzi *et al.*, 2016).

Willow soil enrichment experiments with 0.5% v/v isoprene showed an increase in Actinobacteria after 15 days, dominated by *Rhodococcus* species (average of 88%, s.d. of 5% of the heavy fraction of $^{13}$C incubations). Labelled incubations also included Betaproteobacteria, *Comamonas* spp. and *Variovorax* spp. (average of 6.5%, s.d. of 1.2% of the heavy fraction of $^{13}$C incubations), suggesting the presence of isoprene-degrading bacteria from the Comamonadaceae family in the enriched soils. No evidence of isoprene metabolic genes was found in gene databases for members of the Comamonadaceae family. The authors suggested that isoprene degraders, members of this family, might be

resistant to cultivation under laboratory conditions. Cultivation-independent methods were therefore important to assess the abundance of isoprene-degraders in different environments. DNA-SIP coupled with metagenomics would also help in the search of new genes and new pathways in the use of isoprene (El Khawand et al., 2016).

## 1.2 Rationale

Global warming is a threat will affect all organisms, present and future. Various climate active volatile gases control the fluctuation of global temperature; the production of isoprene, mainly by terrestrial plants, contributes 1/3 of the global volatile organic carbon (VOC) emissions to the atmosphere, making it a very important but often overlooked gas which contributes to both cooling and warming of the Earth (Taraborrelli et al., 2012). Elucidating this gap in the biogeochemical cycle by studying isoprene degrading soil bacteria would help our understanding of the role of microbes in the biogeochemical cycling of isoprene.

## 1.3 Hypothesis

Terrestrial isoprene-degrading bacteria play an important role in moderating the flux of isoprene to the atmosphere.

## 1.4 Objectives

To understand how isoprene cycling by bacteria is regulated in the terrestrial environment.

To establish environmentally-relevant conditions for isoprene enrichment for soil and phyllosphere samples.

To use cultivation-dependent methods to isolate novel isoprene-degrading bacteria.

To evaluate the diversity and abundance of the key isoprene-degrading bacteria in willow soil using DNA-SIP (16S rRNA gene diversity and SIP-metagenomics).

To evaluate the diversity, identity and abundance of key isoprene-degrading bacteria in soil around an oil palm tree and oil palm leaves using DNA-SIP.

To compare bacterial isoprene degradation gene diversity ($\alpha$-subunit of the isoprene monooxygenase and metabolic gene cluster) from isoprene-degrading bacteria in environmental samples.

# Chapter 2 Materials and Methods

## 2.1 Materials

Media formulations, molecular biology-grade chemicals and analytical-grade chemicals were obtained from Fisher Scientific UK (BD Bacto ™ Agar, Difco ™ Agar Noble), Thermo Scientific (Oxoid R2A Agar), Formedium (Nutrient Agar), Sigma-Aldrich Corporation (Phytagel), Melford Laboratories Ltd (Agarose). Gases were obtained from Sigma-Aldrich, Air Liquide UK (Birmingham, UK), CK Special Gases Limited (Newton Unthank, UK) or BOC (Manchester, UK). Custom oligonucleotide primers were ordered from Invitrogen (**Table 2.1**).

## 2.2 Buffers, vitamin and antibiotic solutions

*Tris-acetate-EDTA (TAE) buffer 50 X* (Sambrook & Russell, 2001): Tris 242 g $l^{-1}$, glacial acetic acid 57.1 ml $l^{-1}$, 0.5 M Sodium EDTA (pH 8.0) 100 ml $l^{-1}$.

*Tris-borate-EDTA (TBE) buffer 10 X* (Sambrook & Russell, 2001): Tris base 108 g $l^{-1}$, orthoboric acid 55 g $l^{-1}$, 0.5 M Sodium EDTA (pH 8.0) 40 ml $l^{-1}$.

*Tris-EDTA (TE) buffer pH 8.0* (Sambrook & Russell, 2001)*:* Tris-HCl 10 mM, $Na_2EDTA$ 1 mM, prepared from 1 M Tris-HCl (pH 8.0) and 0.5 M $Na_2$-EDTA (pH 8.0).

*SET buffer*: EDTA 40 mM, Tris-HCl pH 9.0 50 mM, sucrose 0.75 M

*MAMS (marine ammonium mineral salt) vitamins,* modified from vitamin B mixture (Kelly & Wood, 1998)*:* 400 ml Milli-Q water, 10 mg thiamine hydrochloride, 20 mg nicotic acid, 20 mg pyridoxine hydrochloride, 10 mg *p*-aminobenzoic acid, 20 mg riboflavin, 20 mg calcium pantothenate, 1 mg biotin, 2 mg cyanocobalamin B12, 5 mg lipoic acid, 5 mg folic acid. The pH was adjusted to 4.0, diluted to 1 litre of Milli-Q water and aliquots were filter sterilized before use with 0.2 μm pore filter (Sartorius Minisart, Germany).

*Antibiotics:* Filter sterilized ampicillin stock solutions at 100 mg $ml^{-1}$ stock were kept at -20 °C. Before use, the stock solution was thawed and kept on ice. Cultivation medium was prepared, autoclaved and left to cool before adding the antibiotic at a final working solution of 100 μg $ml^{-1}$.

## 2.3 Medium preparation and cultivation of bacterial strains

All cultivation medium and solutions were prepared using de-ionised water (Milli-Q water) and sterilized by autoclaving at 15 psi for 15 minutes at 121 °C. Solid medium was prepared by adding 1.5 % (w/v) of Bacto Agar (Difco), Agar Noble (Difco) or Phytagel (Sigma) before autoclaving. Glucose, MAMS vitamins, and other temperature sensitive solutions were sterilized by passing them through a sterile 0.2 μm pore filter (Sartorius Minisart, Germany) before adding to the media. A list of the bacterial strains used during this study is shown in **Table 2.2.**

**Table 2-1.** List of primer pairs sequences used during this study. The list also includes a short description and their references.

| Primer name | Sequence (5'-3') | Description | Reference |
|---|---|---|---|
| 27F | AGAGTTTGATCMTGGCTCAG | Complete 16S rRNA gene | (Lane, 1991) |
| 1492R | TACGGYTACCTTGTTACGACTT | Complete 16S rRNA gene | (Lane, 1991) |
| 27Fmod | AGRGTTTGATCMTGGCTCAG | Partial 16S rRNA gene amplicon sequencing | (Kuske *et al.*, 2006) |
| 519R | GTNTTACNGCGGCKGCTG | Partial 16S rRNA gene amplicon sequencing | (Sotres *et al.*, 2016) |
| 341F | CCTACGGGNGGCWGCAG | Partial 16S rRNA gene amplicon sequencing | (Klindworth *et al.*, 2013) |
| 785R | GACTACHVGGGTATCTAATCC | Partial 16S rRNA gene amplicon sequencing | (Klindworth *et al.*, 2013) |
| M13F | GTAAAACGACGGCCAG | internal pGEM-T cloning | Invitrogen |
| M13R | CAGGAAACAGCTATGAC | internal pGEM-T cloning | Invitrogen |
| SP6 | TATTTAGGTGACACTATAG | internal pGEM-T sequencing | Invitrogen |
| T7 | TAATACGACTCACTATAGGG | internal pGEM-T sequencing | Invitrogen |
| isoAf | TGCATGGTCGARCAYATG | *isoA* gene degenerate primers | (El Khawand *et al.*, 2016) |
| isoAr | GRTCYTGYTCGAAGCACCACTT | *isoA* gene degenerate primers | (El Khawand *et al.*, 2016) |

| | | | |
|---|---|---|---|
| | CGCCCGCCGCGCGCGGCGGGCGGG | | |
| 341F GC | GCGGGGGCACGGGGGGCCTACGGA | | |
| | GGCAGCAG | DGGE partial 16S rRNA gene | (Muyzer *et al.*, 1993) |
| 518R | CCGTCAATTCMTTTRAGTTT | DGGE partial 16S rRNA gene | (Lane, 1991) |
| isoAf1 (11f) | GVGACGAYTGGTAYGACA | new *isoA* gene degenerate primers | this study |
| isoAr1 (1023r) | GCRTTBGGBTTCCAGAAC | new *isoA* gene degenerate primers | this study |
| isoAr2 | TCSADCATGAAYTCCTTG | new *isoA* gene degenerate primers | this study |

*Ewers minimal medium (MM) (Dorn et al., 1974)*

Cultivation and enrichment medium was prepared by using 1/10 of solution 1, 1/100 of solution 2 and 1/1000 of solution 3. The medium was then autoclaved and 3% phosphate buffer was added when cool (Ewers, *et al.*, 1990). *Solution 1 (X10 mineral salts):* 4 g $(NH_4)_2SO_4$, 0.8 g $MgSO_4 \cdot 7H_2O$; 0.29 g $Ca(NO_3)_2 \cdot 4H_2O$, complete to 400 ml using Milli-Q water and autoclave. *Solution 2 (X100 Iron ammonium citrate):* 1 g per litre Fe Ammonium Citrate and autoclaved. *Solution 3 (Trace elements SL6):* for 1 litre added 10 mg $ZnSO_4 \cdot$; 3 mg $MnCl_2 \cdot 4H_2O$; 30 mg $H_3BO_3$; 20 mg $CoCl_2 \cdot 6H_2O$; 1 mg $CuCl_2 \cdot 2H_2O$; 2 mg $NiCl_2 \cdot 6H_2O$, 3 mg $Na_2MoO_4 \cdot 2H_2O$ (Quayle and Pfennig 1974). *Phosphate buffer solution:* use at pH 7.0 for *Rhodococcus* sp. AD45 and pH 6.5 for targeted isolation of non-*Rhodococcus* isolates (describer in **section 2.11**). Solution prepared by mixing $KH_2PO_4$ 0.5 M and $Na_2HPO_4$ 0.5 M. A 1:1 mixture produced a pH of 6.8. Use at 25 ml per litre of medium (12.5 mM).

Isolates and sub-cultivation of environmental samples, in liquid medium, was done by growing on gaseous isoprene in 125 ml sterile serum vials sealed with butyl rubber seals. Stable isotope enrichment of environmental samples was done in sterile 2-litre modified glass bottles, with butyl rubber seals for addition of the gas substrate. Isoprene was added as explained in section 2.9. The vials were shaken in the dark at 150 rpm at 25-30 °C. Solid medium plates were inoculated or streaked and placed in an anaerobic jar with the gas substrate and incubated at 30 °C for 24-48 h. Additional incubation was done at room temperature.

*SOB medium(Sambrook & Russell, 2001)*

1 litre of SOB medium (pH 7.0) was prepared by adding 5 g yeast extract, 20 g tryptone, 0.5 g NaCl, and 10 ml of 250 mM KCl solution. The pH was adjusted to 7.0 with 5.0 M NaOH. Before use, 2.0 M sterile $MgCl_2$ solution was added to a final concentration of 10 mM.

*SOC medium (Sambrook & Russell, 2001)*

SOC medium was prepared from SOB medium by adding filter sterilized glucose to a final concentration of 20 mM.

*Luria Bertani (LB) medium(Sambrook & Russell, 2001)*

Tryptone 10 g, yeast extract 5 g, NaCl 10 g, and volume was adjusted to 1 litre with deionized water and autoclaved.

*Growth of Rhodococcus genus isolates*

All *Rhodococcus* isolates were grown, stored and maintained in both R2A agar (Oxoid) and/or solid Ewers minimal medium (MM) pH 6.5 supplemented with gas isoprene (details in **section 2.8**; concentration between 0.0013 % or 13 ppmv and 1% or 10000 ppmv). *Rhodococcus* sp. AD45 strain was grown on MM pH 7.0. Strains incubation time was between 24 to 48 h at 30 °C.

*Growth of non-Rhodococcus isolates*

*Variovorax, Sphingobacterium, Sphingopyxis, Gordonia* and *Nocardioides* isolates were grown, stored and maintained in MM pH 6.5 with gas isoprene (described in **section 2.8**) as their sole carbon and energy source (concentration between 0.0013 % or 13 ppmv and 1% or 10000 ppmv) . Solid MM plates were first incubated for 24 h at 30 °C and left for up to 7 to 10 days at room temperature in anaerobic chambers supplemented with gas isoprene.

*Variovorax sp. strain WS9*

The optimal growth conditions were tested at different pH (5.5 to 7), with alternative sources of carbon (10 mM glucose, yeast extract 10 mg $l^{-1}$ and 5mM succinate), and MAMS vitamins (1 to 4 µl $ml^{-1}$). The isolate was also grown with 0.1 % or 1% isoprene and 5 mM of succinate to test for constitutive consumption of isoprene.

**Table 2-2** List of bacteria strains used during this study. WS: Willow soil, WL: Willow leaf, TDS: tyre dump soil, OPL: oil palm leaf.

| Strains | Location | Reference/source |
|---|---|---|
| *Rhodococcus* sp. AD45 | Freshwater sediment (NL) | (van Hylckama Vlieg *et al.*, 1998) |
| *Rhodococcus* opacus PD630 | DMSZ (GE) | (Alvarez *et al.*, 1996) |
| *Xanthobacter autotrophicus* Py2 | Freshwater | (Small & Ensign, 1997) |
| *Mycobacterium hodleri* i29a2* | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Gordonia polyisoprenivorans* i37 | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Mycobacterium* sp. AT1 | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Rhodococcus erythropolis* i24 | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Rhodococcus wratislaviensis* i48 | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Micrococcus luteus* i61b | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Loktanella* i8bn | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Rhodococcus globerulus* i8a | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Rhodococcus globerulus* i8a1 | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Rhodococcus globerulus* i8a2 | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Leifsonia* sp. i49 | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Rhodococcus wratislaviensis* i29b | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |

| | | |
|---|---|---|
| *Mycobacterium fortuitum* i61b | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Shinella* sp. i39 | Colne Estuary (Essex, UK) | (Alvarez *et al.*, 2009) |
| *Rhodococcus* sp. SC4 | Soil (Leamington Spa, UK) | (El Khawand *et al.*, 2016) |
| *Rhodococcus* sp. LB1 | Leaf (University of Warwick, UK) | (El Khawand *et al.*, 2016) |
| *Microbacterium* sp. i39y | Colne Estuary (Essex, UK) | (Johnston *et al.*, 2017) |
| *Stappia* sp. iL42 | L4 sampling station (Plymouth, UK) | (Johnston *et al.*, 2017) |
| *Microbacterium* sp. iP1 | Penarth coast (Wales, UK) | (Johnston *et al.*, 2017) |
| *Pseudomonas stutzeri* OX1 | Activated wastewater sludge (ITA) | (Baggi *et al.*, 1987) |
| *Pseudomonas mendocina* KR1 | Algal-bacterial mat (TX, USA) | (Whited & Gibson, 1991) |
| *Ralstonia picketti* PKO1 | Sandy aquifer | (Kukor & Olsen, 1990) |
| *Ralstonia eutropha* JMP134 | Soil (CHE) | (Laemmli *et al.*, 2000) |
| *Rhodococcus* sp. Essex 74 | Soil (Essex, UK) | Terry McGenity lab collection 2016 |
| *Rhodococcus* sp. Essex 75 | Soil (Essex, UK) | Terry McGenity lab collection 2016 |
| *Rhodococcus* sp. strain WS1 | WS (Colney, UK) | This study |
| *Rhodococcus* sp. strain WS2 | WS (Colney, UK) | This study |
| *Rhodococcus* sp. strain WS3 | WS (Colney, UK) | This study |
| *Rhodococcus* sp. strain WS4 | WS SIP (Colney, UK) | This study |
| *Rhodococcus* sp. strain WS5 | WS SIP (Colney, UK) | This study |

| | | |
|---|---|---|
| *Rhodococcus* sp. strain WS6 | WS SIP (Colney, UK) | This study |
| *Rhodococcus* sp. strain WS7 | WS SIP (Colney, UK) | This study |
| *Rhodococcus* sp. strain WS8 | WS SIP (Colney, UK) | This study |
| *Variovorax* sp. strain WS9 | WS SIP (Colney, UK) | This study |
| *Rhodococcus* sp. strain WS10 | WS (Colney, UK) | This study |
| *Variovorax* sp. strain WS11 | WS (University of East Anglia, UK) | This study |
| *Nocardioides* sp. strain WS12 | WS (University of East Anglia, UK) | This study |
| *Variovorax* sp. strain WS13 | WS (University of East Anglia, UK) | This study |
| *Rhodococcus* sp. strain TD1 | TDS (Fakenham, UK) | This study |
| *Rhodococcus* sp. strain TD2 | TDS (Fakenham, UK) | This study |
| *Rhodococcus* sp. strain TD3 | TDS (Fakenham, UK) | This study |
| *Rhodococcus* sp. strain WL1 | WL (University of East Anglia, UK) | This study |
| *Rhodococcus* sp. strain OPL1 | OPL (London, UK) | This study |
| *Gordonia* sp. strain OPL2 | OPL (London, UK) | This study |
| *Sphingobacterium* sp. strain OPL3 | OP (London, UK) | This study |
| *Sphingopyxis* sp. strain OPL5 | OP (London, UK) | This study |
| *Escherichia coli* TOP10 | - | Invitrogen |

## 2.4 Bacterial purity checks, microscopy and maintenance

Purity checks consisted of serial dilutions of bacterial cultures and phase contrast microscopy. For microscopic observation, 1 ml of fresh culture was centrifuged for 10 minutes at 1000 x g to pellet cells. 900 μl of supernatant was discarded and the pellet re-suspended in the lower volume. 20-30 μl of concentrated cells was observed at 1000x magnification in phase-contrast under a Zeiss Axioskop 50 microscope, 130 VA Type B, and documented using the AxioCam camera system and Axiovison Rel 4.8 software (Carl Zeiss Ltd, Cambridge UK). To confirm the identity of the strains, a nearly complete 16S rRNA gene PCR was performed with 27f-1492r primers and the gene was sequenced.

Glycerol stocks were prepared by growing the isolates in liquid rich or minimal medium to an OD600 higher than 0.6. 1 ml of the thick culture was transferred into a cryovial along with 1 ml of 30 % (v/v) sterile glycerol and mixed by pipetting slowly. Cryovials were then drop-frozen into liquid nitrogen and stored at - 80 °C.

## *2.5* Transformation of chemically-competent *Escherichia coli*

Chemically competent *E. coli* Top10 or JM109 cells (Promega) were thawed on ice and mixed gently with 2 μl of ligation mix or plasmid DNA. After an incubation of 20 min, cells were heat shocked at 42 °C for 50 seconds and then placed back on ice for 2 min. 950 ml of SOC medium was added to the cells before incubating for 90 min at 37 °C, shaking at 150 rpm. 50 μl to 100 μl aliquots were then spread onto selective LB agar plates with ampicillin (100 μg $ml^{-1}$) and incubated at 37 °C overnight.

## *2.6* Preparation and transformation of electro-competent *Escherichia coli*

*E. coli* was grown at 37 °C overnight in 5ml of sterile LB medium. 500 ml of LB medium was then inoculated and grown to an OD600 between 0.4 and 0.5. The cell culture was cooled on ice for 15 min and then centrifuged at 4 °C for 15 min at 8,000 x g. The supernatant was removed and the cells were washed twice with cold sterile Milli-Q water, first with 500 ml

and then with 250 ml. A final wash was done with cold 10 % (v/v) glycerol. The cell pellet was re-suspended in 2.5 ml of cold 10 % (v/v) glycerol, 100 µl aliquots were placed into 1.5 ml centrifuge tubes, drop frozen in liquid nitrogen and stored at -80 °C.

Electro-competent *E. coli* were transformed by mixing, on ice, with 2 µl of plasmid DNA or ligation mix and incubating on ice for 1 min. The mixture was transferred to a chilled 0.1 cm electroporation cuvette (Plus BTX, USA) and an electric pulse was applied at 1.8 kV, 25 µF, and 200 Ω with a GenePulser Electroporation system (Bio-Rad, UK). Cells were immediately added to 1 ml of SOC medium and incubated for at 37 °C, shaking at 150 rpm, for 1 hour. 50 to 100 µl aliquots were then spread onto selective LB agar plates.

## 2.7 Environmental sampling of soil and leaves

Samples were obtained from different sites throughout the study (Table 2.3). 10 - 20 grams of soil and/or leaves were taken from three aleatory points around the tree of interest. Samples were collected in sterile bags or sterile Falcon tubes using gloves and transported back to the laboratory to be processed.

### *Soil samples*

Soil was collected 10 - 20 cm from the trunk of the selected tree. Samples were dug out from 5 - 10 cm under the surface, stored and transported immediately in sterile 50 ml Falcon tubes at room temperature back to the laboratory. All residues of grass, leaves, roots and stones were separated from the soil before storing. In the case of the tyre dump samples, collected by Dr. Jennifer Pratscher, two soil samples were taken from close to old tyres and one sample from underneath a tyre.

### *Leaf samples*

For native tree species, a few branches from the external crown (2 m above ground) of the trees were cut and stored in large sterile plastic bags. Leaves were transported immediately to the lab at room temperature and placed on sterile paper before cutting all leaflets from their petioles. Leaves were weighed and/or surface area was measured before swabbing or washing.

For non-native tree species, no leaf material was transported to the laboratory initially. Oil palm tree leaf swabs were taken directly at the Palm House in Kew Gardens in November 2016. The sampled cotton swabs were transported to the lab in sterile paper inside plastic bags. For the last sampling, in February 2017, a large section of a compound leaf (approximately 50 $cm^2$) was transported back to the lab and processed as described previously for native trees.

## 2.8 Quantifying isoprene uptake in the headspace

Concentration of isoprene in the sealed glass vials was quantified by injecting 100 µl of headspace gas into an Agilent Technologies 7890A gas chromatograph. The system was fitted with an HP-Plot/Q column (30 m, 530 µm bore, 40 µm film) at an oven temperature of 175 °C, injector 250 °C (1:5 split ratio) and flame ionization detector at 300 °C (carrier gas He, 4 ml $min^{-1}$) (Crombie *et al*. 2015). Retention time for isoprene was approximately 6.5 minutes.

### *Isoprene standards*

Samples were compared to a series of fresh isoprene standards of 10, 50, 150 and 250 ppmv in 120 ml serum vials. Vials were flushed with nitrogen ($N_2$) before adding a known volume of liquid isoprene using a glass micro-syringe.

**Table 2-3** List of environmental samples taken during this study. The list also includes the coordinates, dates at which samples were taken and a short description of the purpose of each sample.

| Sample | Location | Coordinates | Dates | Description |
|---|---|---|---|---|
| Willow tree topsoil | Colney sport fields (Colney, UK) | 52°37'17.2"N 1°13'46.3"E | April 2015 and July 2015 | Enrichment and isolation |
| Willow tree topsoil | UEA broad (Norwich, UK) | 52°37'06.6"N 1°14'09.9"E | June 2016 | Enrichment and targeted isolation |
| Tyre dump soil | Tyre dump (Fakenham, UK) | 52°50'54.6"N 0°44'49.2"E | June 2016 | Enrichment and isolation |
| Willow tree leaves | UEA broad (Norwich, UK) | 52°37'06.6"N 1°14'09.9"E | September 2015 and October 2016 | Epiphyte enrichment and isolation |
| Ash tree leaves | UEA broad (Norwich, UK) | 52°37'06.6"N 1°14'09.9"E | September 2015 and October 2016 | Epiphyte enrichment |
| Eucalyptus tree leaves | The Avenues (Norwich, UK) | 52°37'34.8"N 1°16'13.6"E | October 2016 | Epiphyte enrichment |

| | | | | |
|---|---|---|---|---|
| Oil palm tree topsoil | Kew Gardens Palm House (London, UK) | 51°28'43.6"N 0°17'33.1"W | November 2016 and February 2017 | Enrichment and isolation |
| Oil palm tree leaves | Kew Gardens Palm House (London, UK) | 51°28'43.6"N 0°17'33.1"W | November 2016 and February 2017 | Epiphyte enrichment and isolation |

## 2.9 Procedures for the enrichment of environmental samples with isoprene

Isoprene was added by injection of isoprene vapour (withdrawn from the headspace of a 2-ml vial containing a small quantity of liquid isoprene heated for 5 min in a 37 °C water bath) and incubations were set up with nominal concentrations, i.e. representing the volume of vapour added to incubations. All isoprene concentrations refer to the concentration of vapour in the headspace. Real-time concentrations in incubations were determined by gas chromatography.

### *Enrichment of soil samples with isoprene*

Incubations were done with 40 ml of minimal medium and 4 - 5 g soil in 2 litre modified flasks. Tests with diluted minimal medium (0.1X) and sterile water were also implemented to show if there was need to add additional nutrients. Isoprene concentrations used for enrichment assays started at 250 ppmv, for initial experiments, and were reduced to a minimum of 13 ppmv.

### *Enrichment of leaf-washings with isoprene*

For willow, eucalyptus and ash trees, 5 g of leaves was added to 500 ml Erlenmeyer flasks containing 100 ml minimal medium. For oil palm leaves 33.75 g of leaves (3 leaflets) were washed. Leaves were sonicated for 5 min, 50 kHz, at 4 °C (Mettler ME2, Anaheim CA, USA) and shaken for 15 min at 25 °C at 200 rpm. Cell suspensions were added to 50ml Falcon tubes and centrifuged at 5,000 x g at 22 °C for 20 min. The supernatant was separated from pellet carefully and then passed through CellTrap CT402LL001N00 filters (MEM-TEQ Ventures Ltd, Selby, UK) to collect all cells left after centrifugation. The filters were then washed with 10 ml of fresh minimal medium to eliminate all other carbon compounds washed from the leaves and that might be used as an alternative source of carbon and energy during enrichment. Filtered cells and pellet were re-suspended in 50 ml minimal medium pH 6.5 with MAMs vitamins (1µl $ml^{-1}$) and placed in a 2-litre flask with 50 ppmv of isoprene. Samples were incubated at 25 °C, in the dark, and shaken at 150 rpm and isoprene consumption was monitored by gas chromatography.

*Enrichment of leaf-swabbings with isoprene*

For eucalyptus, ash and willow leaves, an equal surface area was calculated for all types of leaves using 5 g of willow leaves as a reference i.e. 294 $cm^2$. Two sequential swabs were used to take cells from the leaves. The first swab was moistened with minimal medium and passed in a zigzag motion over all the surface of the leaf. The second swab was then passed over the surface to absorb the left over minimal medium (Hedin *et al.*, 2010). After all the leaves were sampled the cotton swabs were placed in the 50 ml Falcon tubes with 20-40 ml minimal medium. Falcon tubes were left to sit for 5 min and then vortexed for 1 min. Cotton from swabs was squeezed using a sterile 15 ml Falcon tube and discarded. Cell suspensions were then centrifuged at 5,000 x g at 22 °C for 20 min. The rest of the protocol was the same as with leaf-washings, as described earlier.

### 2.9.1 Endophytes from oil palm leaves

The protocol was obtained as a combination of Bogas *et al*., 2015; Romero, *et al*., 2014. Leaf epiphytes were removed, to ensure recovery of only endophytes from leaves, by surface-sterilizing using a stepwise washing procedure. The leaves were weighed and washed three times with 500 ml of tap water, followed by 1 min with 500 ml of 70% ethanol, 2 min with 500 ml sodium hypochlorite (4% available $Cl^-$), and three washes with 500 ml sterile distilled water. To confirm disinfection, aliquots of the last sterile-distilled water from each replicate were plated on rich R2A medium and examined for bacterial growth after 2 days incubation at 30 °C. No growth was detected.

The surface-sterile leaves were then cut into 1-2 cm pieces and ground with a sterile pestle and mortar with 10 ml of minimal medium pH 6.5 with vitamins. To separate plant material from the medium, the crushed leaves were passed through a mesh and a glass fibre filter subsequently. The recovered medium was later centrifuged at 5000 x g for 20 minutes as the leaf-washing protocol. The cell trap contents and the pellet were re-suspended in 50 ml of MM pH 6.5 with vitamins. All replicates were spiked with 25 ppmv of isoprene and isoprene consumption was monitored by gas chromatography for a few weeks.

## 2.10 Stable isotope enrichment

Fully $^{13}C$-labelled isoprene was manufactured and supplied by Gregg Whited and colleagues at Dupont, USA (El Khawand *et al.*, 2016). Samples were exposed to the fully-labelled isoprene as their sole carbon and energy substrate and non-labelled isoprene replicates were done as a control.

### *Soil and leaf-washing stable isotope incubations*

SIP incubations were carried out in 2-litre custom-made bottles sealed with a butyl rubber cap and an aluminium crimp top. For soil incubations, the flasks contained 4 to 5 g soil and 40 to 50 ml of water, respectively; leaf-washings had the washed cells suspended in 50 ml of minimal medium (pH 6.5, 1µl/ml MAMs vitamins). 25 ppmv of either $^{12}C$ or $^{13}C$-labelled isoprene was added as the sole carbon and energy source. These samples were incubated at 25°C with shaking at 150 rpm in the dark. Isoprene consumption was monitored by gas chromatography.

## 2.11 Isolation of isoprene-degrading bacteria

### *Soil and epiphyte isolates*

After successful degradation of isoprene by the different microcosms, the cultures were processed by performing serial dilutions and plating on rich agar (R2A) and minimal medium agar supplemented with isoprene vapour. Growth of bacteria was registered after 48 h at 30 °C, and up to 10 days at 21 °C. Phenotypically different isolates were selected, purified and grown to a high biomass before being tested individually for isoprene degradation in 20 ml liquid minimal medium ($OD_{600}$ of 0.1). Isolates that consumed 50 ppmv of isoprene, were continuously supplemented and sub-cultured every two weeks.

Some isoprene-degraders from oil palm leaves were obtained using leaf presses on minimal medium plates pH 6.5 with vitamins (1 µl $ml^{-1}$). Plates were transported back to the lab and placed in closed chambers supplemented with 500 ppmv isoprene in the headspace.

Colonies were transferred with a wood toothpick to minimal medium plates and rich medium plates. Monitoring and picking isolates was done every 2-3 of days for 4 weeks.

*Sub-cultivation isolation strategy*

To effectively obtain isoprene-degrading isolates from oil palm leaves, the enriched samples were supplemented with isoprene at low concentration (25 ppmv) and high concentration (250 ppmv) every two days, and transferred every 3 to 4 weeks into fresh medium (1:10). A total of 5 subcultures were done for soil samples, and 4 subcultures for leaf-washings. Selection of strains was performed as described previously.

*Targeted isolation of non-Rhodococcus isolates*

Environmental samples were enriched and transferred in liquid minimal medium pH 6.5 with 1 μl $ml^{-1}$ of MAMs vitamins with a maximum concentration of 25ppmv of isoprene in the headspace. Serial dilutions after the third subculture were plated onto minimal medium and R2A agar plates, and incubated for 24 h at 30 °C. Plates were incubated for 7 to 10 days at 21 °C and different phenotypes were selected for further analysis.

## 2.12 Analysis of proteins in isoprene-degraders

*Cell harvesting and preparation of cell extracts*

*Rhodococcus* sp. AD45, *Rhodococcus opacus* PD630 and *Variovorax* sp. strain WS9 cells were grown in 500 ml sterile flasks containing 200 ml minimal medium with 3 mM succinate or 1 % (v/v) isoprene. The flasks were incubated at 25 °C, shaking at 150 rpm, and cells were harvested at late exponential phase ($OD_{600}$: 0.8 to 1) by centrifugation for 30 min at 14,000 x g, 4 °C. After decanting the supernatant, the cells were washed with minimal medium containing no substrate and re-suspended in 2 ml of 50 mM phosphate buffer (pH 7.0) and 4 μl of 0.5 mM benzamidine. The cells were subjected to three passages through a French pressure cell (American Instrument Company, USA) at 110 MPa (on ice). Cell-free extracts were prepared by removing cell debris via centrifugation for 20 min at 10,000 x g, 4 °C.

*Protein quantification*

The concentration of proteins in the cell-free extracts was estimated using the Bio- Rad Protein Assay (Bio- Rad) whereby bovine serum albumin was used for the preparation of standards, according to the manufacturer's protocol.

### 2.12.1 Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE)

Polypeptides in the cell-free extracts were separated by SDS-PAGE using an X-Cell II Mini-Cell apparatus (Novex). A 4% (w/v) stacking gel and a 12.5 % (w/v) resolving gel. Protein samples were prepared in a total volume of 15 µl cell extract (15 µg of protein) and 5 x loading buffer. The samples were placed for 8 min in a boiling water bath, cooled on ice, and supernatant loaded into the wells. Samples were run in 5 x running buffer at 90 V through the stacking gel and 120 V through the resolving gel.

## 2.13 DNA extraction

DNA extracted from isolates and environmental samples, for routine PCR experiments, was done using the FastSpin DNA soil kit (MP Biomedicals, Santa Ana, USA) according to the manufacturer's instructions.

*Genomic DNA extraction from Rhodococcus isolates*

*Rhodococcus* strains were grown in 50 ml liquid minimal medium with 5,000 ppmv of isoprene (25 °C, shaken at 150 rpm). When cells reached an $OD_{600}$ of 0.8 to 1, the cells were spun down (6,000 x g for 20 min) and kept in the freezer. The cell pellet was re-suspended in 5 ml of resuspension buffer (20 mM Tris, 2 mM EDTA, pH 8.0) and incubated with 60 µl of lysozyme (100 mg $ml^{-1}$) for 60 min at 37 °C. After incubation, 375 µl of proteinase K (10 mg $ml^{-1}$) and 7 µl RNaseA (10 mg $ml^{-1}$) were added and incubated for another 15 min at 37 °C. 1/7 volume, approximately 780 µl, of N-lauryl sarcosine (10 %w/v) was added and incubated for 60 min at 60 °C (mixing the sample gently every 10 min). 1.012 ml of 5 M

NaCl were added to the sample along with 803 µl of CTAB/NaCl (10 % in 0.7 M NaCl) and incubated for a further 15 min at 60 °C. Equal amounts of phenol: chloroform: isoamyl alcohol (25:24:1) were added to the sample, shaken vigorously, incubated for 10 min at 60 °C, shaken again and centrifuged for 5 min at 6,000 x g. The upper phase was removed into a fresh tube and the previous step repeated twice, without incubation, and mixing by inversion. One last extraction was performed with chloroform: isoamyl alcohol. DNA was precipitated with 2 volumes of cold 100 % ethanol and incubation overnight at 4 °C. DNA was pelleted by centrifugation for 20 min at 6,000 x g 4 °C. DNA was then washed with 70 % (v/v) ethanol, centrifuged again and the DNA pellet was dried briefly. 400 µl of 10 mM Tris pH 8.0 was added to re-suspend the DNA.

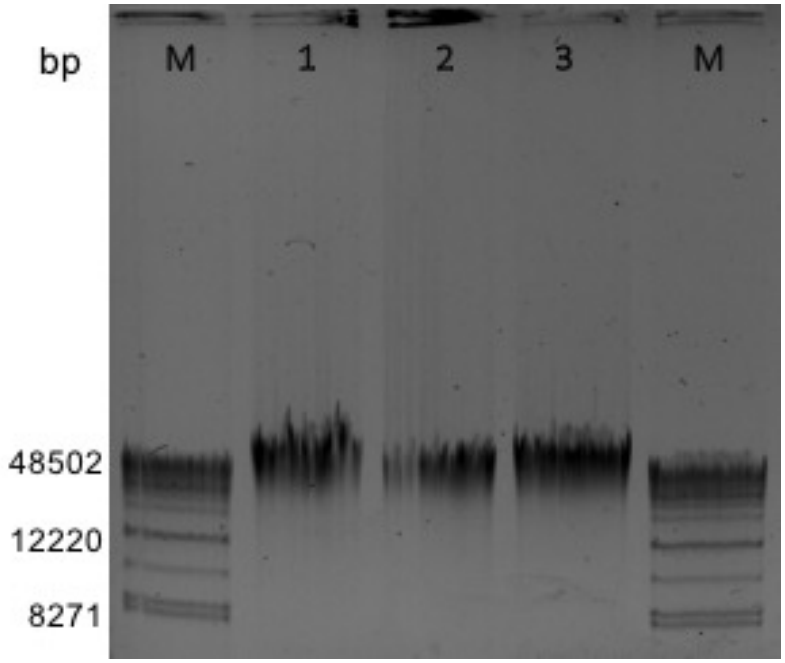



**Figure 2.1** Genomic DNA of the three *Rhodococcus* isolates (WS1, WS2 and WS3 respectively), run on a 0.6% agarose gel (75V for 100 min).M: Lambda DNA mixed marker

## 2.14 DNA stable isotope probing (DNA-SIP)

After harvesting samples, from $^{12}$C-isoprene and $^{13}$C-isoprene enrichments, at the established time-points with the stable isotope (section 2.8 and section 2.12), native and enriched samples, from each time point, were stored at -20 °C until the DNA extraction. Total DNA from all environmental samples was extracted using the FastSpin DNA soil kit following the manufacturers' instructions (**section 2.13**). For fractionation of labelled ($^{13}$C)

and unlabelled ($^{12}$C) DNA (Neufeld *et al.*, 2007). 5 μg of DNA from soil SIP enrichments, or up to 1 μg of DNA for leaf SIP enrichments, were added to gradient buffer and subsequently to caesium chloride solutions (final density of 1.725 g $ml^{-1}$). Each replicate in a polypropylene centrifuge tube (Beckman Coulter, Inc., Brea, CA, USA) was submitted to density gradient ultracentrifugation at 177,000 x $g_{av}$ in a Beckman Vti 65.2 rotor at 20 °C for 65-80 h with vacuum, maximum acceleration and without using the break on completion of the run.

DNA was then fractionated, separating labelled $^{13}$C DNA (heavy DNA) from non-labelled $^{12}$C DNA (light DNA) due to a density gradient. The contents of each tube were separated into 12-14 fractions (440-500 μl each) by gradient fractionation with a needle and a low-flow peristaltic pump (Watson-Marlow Pumps, Falmouth, Cornwall, UK). The density of each was determined with an AR200 digital refractometer (Reichert, USA) and DNA was precipitated using PEG-NaCl solution protocol (Sambrook & Russell, 2001) and re-suspended in nuclease free water. DNA from each fraction was quantified using a Qubit™ dsDNA HS Assay kit (**section 2.15.1**).

## 2.15 Processing of DNA from SIP experiments

### 2.15.1 DNA purification and quantification

DNA was purified from PCR reactions or using band-excision using the Nucleospin Gel & PCR Cleanup column kit (Macherey-Nagel, Düren, Germany) according to the manufacturers' instructions. DNA concentrations were estimated by agarose gel electrophoresis and comparison to a known quantity of 1kb ladder (Fermentas) or using a NanoDrop spectrophotometer (NanoDrop Technologies Inc., Wilmington, USA).

Fractions with lower concentrations of DNA, from DNA-SIP, were quantified using the Qubit Fluorometer and high-sensitivity DNA (dsDNA HS) protocol using instructions set by manufacturer (Thermo Fisher Scientific, UK).

### 2.15.2 Polymerase Chain Reaction (PCR)

All basic PCR reactions were carried out in 50 µl reaction volumes using a Tetrad (Bio-Rad) thermal cycler and were performed using 1 × buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs (for each), 0.4 µM forward and reverse primer, and 2.5 units of DreamTaq (Fermentas). For direct amplification of 16S rRNA gene and *isoA* gene from bacterial cultures or colonies, 0.07% w/v BSA and 5% v/v DMSO were also included in the reaction mix. PCR conditions included a step of 95 °C for 3 min, followed by 25-30 cycles of 94 °C for 1 min, annealing temperature determined according to primers utilized for 30 sec to 1 min, 72°C for 1 min. The final elongation time was 72 °C for 11 mins. When PCR was conducted directly from colonies, the initial denaturation step was 10 mins. For PCR amplification of full-length 16S rRNA genes, primers used were 27f and 1492r (**Table 2.1**).

#### *GC clamp 16S rRNA PCR for DGGE*

The PCR protocol using the 341F-GC and 518R primer pair was as follows: 95 °C for 5 min, followed by 29 cycles of 94 °C for 1 min, 60 °C for 1 min, 72°C for 1 min. The final elongation time was at 72 °C for 11 mins.

#### *isoA gene PCR*

PCR was performed with double the concentration of the *isoA* gene primer pair isoAf and isoAr (El Khawand et al., 2016). A touch down PCR protocol was 94 °C for 3 min, 19 cycles of 94 °C for 30 secs, 72 °C for 45 secs (decreased 1 °C for every cycle) and 72 °C for 1 min. 25 cycles of 94 °C for 30 secs, 54°C for 45 secs, 72 for 1 min. A final elongation step of 72 °C for 5 min was used.

### 2.15.3 Cloning of 16S rRNA gene and isoA PCR products

Freshly prepared and purified PCR products were cloned into pGEM-T Easy vector (Promega) as instructed in the manufacturers' manual. The correct insertion of the PCR product was checked using an M13f-M13r PCR product visualized in a 1% agarose gel.

### 2.15.4 Gel electrophoresis

*Agarose gel electrophoresis*

Extracted DNA and PCR gene products were examined and separated using a 0.6 - 1 % (w/v) agarose gel containing ethidium bromide (0.5 μg $ml^{-1}$). GeneRulerTM 1kb DNA (Fermentas) ladder was used as a marker for size comparison of DNA fragments. Samples were typically run for 30 min at 90 V in 1 x TBE buffer and visualized using a Bio-Rad Gel Doc XR documentation system.

*Denaturing gradient gel electrophoresis (DGGE)*

16S rRNA genes were amplified from all fractions or from pooled light and heavy fractions resulting from the SIP experiments using the 341F-GC and 518R PCR primers (Dar, et al., 2005). DGGE of bacterial 16S rRNA gene fragments was carried out using the DCode™ Universal Mutation Detection System (Bio-Rad) following the manufacturer's instructions. PCR products were loaded on a 1 mm thick vertical gel containing 8% (v/v) polyacrylamide (acrylamide-bisacrylamide, 37.5:1) in 1x Tris-acetate-EDTA (TAE) buffer (pH 8). A linear gradient of 30 to 70% denaturant (with 100% denaturant corresponding to 7 M urea and 40% (v/v) de-ionised formamide) was used for the separation of 16S rRNA gene fragments. Electrophoresis was carried out for 16 hours at 75 V. Gels were stained using 3 µl SYBR® Gold Nucleic Acid Gel Stain (Invitrogen) in 50 ml TAE buffer for 1 hour in the dark. Gels were rinsed with TAE buffer and bands were viewed under the Bio-Rad Gel Doc XR gel documentation system using the Amber Filter 5206 (Bio-Rad).

## 2.16 DNA sequencing

Short (<2 kbp) PCR products were sequenced at Eurofins Genomics (Ebersberg, Germany) with the appropriate primers. The concentration of DNA and primers was sent according to the company's specifications. Resulting sequence chromatographs were analyzed and edited manually with BioEdit software (Ibis Therapeutics, CA, USA). Overlapping sequences

were aligned and a consensus sequence obtained for either the 16S rRNA gene or the *isoA* gene.

#### *Genomic DNA sequencing*

Genomic DNA from both *Rhodococcus* WS1 and *Rhodococcus* WS3 isolates was sent for sequencing and assembly at Edinburgh Genomics (libraries were constructed and sequenced using Illumina MiSeq and assembled using SPAdes). Sequences were uploaded to RAST (Aziz *et al., 2008*) and Prokka (Seemann, 2014) to obtain annotation of the genomes. For all other isolates, bacteria biomass was sent for genome sequencing to MicrobesNG (University of Birmingham, UK) using Ilumina MiSeq and HiSeq 2500 platforms. Identity of the closest available reference genome was done using Kraken (Wood & Salzberg, 2014), and a map of the reads using BWA mem to assess the quality of the data. De novo assembly of the reads was done using SPAdes (Bankevich et al., 2012) and automated annotation was performed using Prokka (Seemann, 2014). Prokka annotations for most of the genomes were done by Dr. Andrew T. Crombie.

Search and identification of plasmid and bacteriophage sequences in the genome of *Rhodococcus* sp. strain WS4 was performed using default settings for PlasmidFinder 1.3 and PHAST (PHAge Search Tool), respectively. The programs are available at https://cge.cbs.dtu.dk/services/PlasmidFinder/ (Carattoli et al., 2014) and http://phast.wishartlab.com/index.html (Y. Zhou, Liang, Lynch, Dennis, & Wishart, 2011).

### 2.16.1 Amplicon sequencing

#### *16S rRNA gene amplicon sequencing (454)*

Colney willow soil DNA-SIP samples were sent for 16S rRNA gene amplicon sequencing at MR DNA (Molecular Research LP) in Shallowater, USA. Purified 16S rRNA PCR products (amplified using 27f-517mod primers) from native soil; time point 1: $^{12}C$ enrichment pooled replicates for light and heavy fractions, $^{13}C$ enrichment pooled replicates for light and heavy fractions; time point 2: $^{12}C$ enrichment and $^{13}C$ enrichment for light and heavy for each replicate individually.

Products were sequenced utilizing Roche 454 FLX titanium instruments and reagents and following manufacturer's guidelines (Capone *et al.*, 2011; Dowd *et al*., 2008). Sequence data were processed at MR DNA using a published pipeline. Briefly, the Q25 reads were stripped of barcodes and primers. Short sequences (<200 bp), sequences with ambiguous base calls and those with >6 bp homopolymer runs were removed. Remaining sequences were denoised using a custom pipeline, OTUs clustered at 97% sequence identity, chimeric sequences were removed using Uchime (Edgar *et al.*, 2011) and taxonomy was assigned using BLASTn against the RDPII/NCBI database (v 11.1) (Cole et al., 2014).

#### *16S rRNA gene amplicon sequencing (Ilumina)*

16S rRNA genes obtained from DNA-SIP experiments with oil palm soil and leaves were sent for amplicon sequencing to MR DNA Molecular Research, Shallowater, TX, USA. 16S rRNA cleaned PCR products using 341f-785r primers for the following samples: Native soil; $^{13}C$ labelled DNA for time point 3 of each replicate and one sample from time point one and time point 2 (replicate 2).

## 2.17 Clone library construction of 16S rRNA and *isoA* genes

Clone libraries were constructed for 16S rRNA gene and *isoA* gene using DNA extracted from soil or leaf washings isoprene enrichments. The purified PCR products were cloned into pGEMT Easy vector (Promega) and transformed into *E. coli* TOP10 cells (Invitrogen, section 2.14.3). Transformation was confirmed using M13f-M13r primers, PCR products of the correct size were confirmed, the amplifications were purified using a PCR clean-up kit, and sequenced using SP6-T7 primers (**Table 2.2**).

#### *Construction of 16S rRNA gene libraries from oil palm heavy DNA*

As part of the preliminary analysis of oil palm enrichments, the full PCR product from the $^{13}C$ heavy DNA in time-point 3 from both soil and leaves was cloned, and a small clone library was constructed. For this procedure 1 μl of each heavy fraction (4 to 6) was used as a template for the 16S rRNA PCR. The resulting PCR product was run in a 1% agarose gel and excised to eliminate other environmental DNA in the sample and purified using

Nucleospin Gel & PCR Cleanup column kit (Macherey-Nagel, Düren, Germany). The DNA fragment was cloned into pGEM-T (Promega) and positive clones were selected from an M13f-M13r PCR product run on a 1% agarose gel. The PCR products used as a template for 16S rRNA GC-clamp PCR, amplicons were purified and run in a 30 - 70% DGGE gel. Clones with different DGGE migration were sent for sequencing using SP6 and T7 primers (**Table 2.1**, see results in **Chapter 5**).

## 2.18 Metagenome analysis of DNA obtained in DNA-SIP experiments with willow tree soil

DNA samples were sent to University of Liverpool Centre for Genomic Research for metagenome sequencing. Samples consisted of an unfractionated native soil sample, and time point 1 (6 days of enrichment) and time point 2 (7 days of enrichment) heavy $^{13}C$-labelled DNA, pooled from each replicate (fragments 5 and 6). The raw Fastq files were trimmed for the presence of Illumina adapter sequences using Cutadapt version 1.2.1 (Martin, 2011). The option -O 3 was used, so the 3' end of any reads which match the adapter sequence for 3 bp, or more were trimmed. The reads were further trimmed using Sickle version 1.200 (Joshi & Fass, 2011) with a minimum window quality score of 20. Reads shorter than 10 bp after trimming were removed. If only one of a read pair passed this filter, it was included in the R0 file. Statistics were generated using fastq-stats from EAUtils (Aronesty, 2011).

### *Phylogenetic analysis*

Metaphlan (Segata *et al.*, 2012) was used to obtain the taxonomic profile and composition of microbial communities during the enrichment process of the soil samples with $^{13}C$ labelled isoprene. The program was run using the trimmed unassembled metagenome data and was compared to the clade-specific marker genes from 3,000 reference genomes by Dr. Jennifer Pratscher.

*Binning of contigs*

To enhance the number and length of contigs obtained from the metagenomes, heavy DNA metagenome sequences from 6 and 7 days of enrichment were co-assembled with Megahit (Li *et al.,* 2015). Binning with two settings, one for very sensitive and another with very specific was performed with Metabat (Eren *et al.*, 2015). A quality run on the bins to analyse the completeness, contamination and content of the bins was done with CheckM (Parks *et al.*, 2015). This process was done with the help of Dr. Jennifer Pratscher.

*isoA gene sequence extraction*

The blast+ program was used to compare a curated IsoA database to the co-assembled T1-T2 contigs using an e-value of 1e-4. All hits for each IsoA query sequence were listed and compared. Contig numbers were later extracted, cut to include only the *isoA* gene and translated to proteins using alignments (ClustalW; (Thompson *et al.*, 2002)) with the curated IsoA database. An IsoA tree was constructed as the last step using the MEGA 7 software (Kumar *et al.*, 2016).

## 2.19 Analysis of the isoprene gene cluster from novel isoprene-degrading isolates

### 2.19.1. Isoprene-degrading bacteria multiple genome alignment

The complete genomes of *Rhodococcus* sp. AD45, *Rhodococcus* sp. strain WS4, *Rhodococcus* sp. strain WS7, *Variovorax* sp. strain WS9, *Variovorax* sp. strain WS11, *Gordonia* sp. strain OPL2 and *Sphingopyxis* sp. OPL5 were included in the analysis using the MAUVE software (Darling *et al.*, 2004). The programme was run using the default settings. *Rhodococcus* sp. strain WS3 was not included due to its similarity to *Rhodococcus* sp. AD45.

### 2.19.2. Phylogenetic analysis of the isoprene cluster genes

As part of the genome comparison process, isoprene metabolic gene clusters were visualized using the Gene Graphics web application (Harrison *et al.*, 2018). To obtain the isoprene gene cluster, all genomes were first uploaded to the RAST server (Aziz *et al.*, 2008)

and observed using the SEED Viewer genome browser. The *isoA* gene (annotated as a toluene monooxygenase), was located in each genome, and a fragment of 16 kb from each genome, containing the isoprene metabolic gene cluster, was then selected around the gene of interest and exported in to the web application. The clusters were observed and edited for the 8 genomes (see **Chapter 6**).

Ten protein databases, one for each gene in the isoprene metabolic cluster (*isoABCDEFGHIJ*), were generated by extracting protein sequences from the genome annotation data of isoprene-degrading bacteria genomes. Metagenome annotation data from willow tree soil contigs that contained isoprene metabolic genes were also included in the phylogenetic analysis. For the analysis, protein sequences were aligned using ClustalW (Thompson *et al.*, 2002), and maximum Likelihood trees (Johansen & Juselius, 1990) were constructed using the James-Taylor-Thornton model, run by the MEGA 7 software (Kumar *et al.*, 2016). A bootstrap of 1k replicates was used to add robustness and reliability to the resulting trees (Felsenstein, 2011).

### 2.19.3. *isoA* gene PCR primer design

Twenty *isoA* gene reference sequences were aligned using MEGA 7 (Kumar *et al.*, 2016). Sequences corresponded to the *isoA* gene from *Gordonia polyisoprenivorans*, *Mycobacterium* sp. i61a, *Rhodococcus* sp. AD45, *Rhodococcus opacus* PD630, *Rhodococcus* isolate LB1, *Rhodococcus* isolate SC4, *Rhodococcus* isolate WS1, *Rhodococcus* isolate WS3, *Rhodococcus* isolate WS4, *Rhodococcus* isolate WS7, *Variovorax* isolate WS9, *Xanthobacter autotrophicus* Py2 alkane monooxygenase, and metagenome *isoA* sequences from previous work in the lab (8 sequences). Regions of 18 to 20 nucleotides with conserved bases between 19 sequences (not including the alkene monooxygenase) were identified. Initially five forward primers and seven reverse primers were designed. All primers were analysed individually and the oligos with the least number of degenerate bases were selected. Two primer pairs (isoAf1-isoAr1 and isoAf1-isoAr2) were tested with DNA from isolates in the laboratory. Primer set isoAf1-isoAr1 yielded one PCR product at the estimated amplicon size (1013bp). A gradient PCR was done for this primer pair and an optimum annealing temperature of 54 °C was established (see results in **Chapter 6**).

# Chapter 3 Enrichment and isolation of novel isoprene-degrading bacteria

## 3.1 Introduction

Terrestrial and marine environmental samples can oxidize isoprene due to the presence of isoprene-degrading bacteria (Alvarez *et al.*, 2009; Cleveland & Yavitt, 1998; El Khawand *et al.*, 2016; Johnston *et al.*, 2017). The first experiments that detected these microbes in soils were performed in the late 1980's and 1990's (Cleveland *et al.*, 1997; Van Ginkel *et al.*, 1987) and resulted in the isolation of bacteria of the genera *Nocardia*, *Alcaligenes* and *Rhodococcus*.

Gram-positive Actinobacteria of the genus *Rhodococcus* have been shown to play a key role in isoprene degradation (Fall and Copley, 2000; El Khawand *et al*., 2016). Also, isoprene metabolism has been well studied and characterised in *Rhodococcus* sp. AD45, isolated from fresh water sediment (Crombie *et al.*, 2015; Johan *et al.*, 1999; van Hylckama Vlieg *et al.*, 1998; van Hylckama Vlieg *et al.*, 2000). Despite the use of different isolation techniques and media formulations, mostly bacteria of this genus have been isolated from soil samples. In the search for novel isoprene-degraders, studying the main resource of natural isoprene production is essential.

Tropical trees are responsible for 80% of the total terpenoid emissions into the atmosphere i.e. 500 Tg (Guenther *et al.*, 2006; Guenther *et al.*, 2012). Plants produce isoprene inside their leaves via the methylerythritol 4-phosphate (MEP) pathway in the chloroplasts (Lichtenthaler *et al.*, 1997; Rasmussen, 1970). Since leaves are the main source of biologically-produced isoprene (Sharkey *et al.*, 2008; Sharkey & Yeh, 2001), tree foliage is a potential niche for isoprene-degrading microbes.

This chapter presents the results for enrichment and isolation of isoprene-degraders in soil and leaf samples from isoprene-emitting trees. The enrichment conditions used have aimed to simulate concentrations of isoprene closer to those present in the natural environment than those studied before, while still generating a selective pressure. These lower isoprene concentrations and longer incubation periods permitted the cultivation of a wider diversity of isoprene-degraders.

*General objectives:*

To establish the conditions for isoprene enrichment for soil and phyllosphere samples.
To use cultivation-dependent methods to isolate novel isoprene-degrading bacteria in-vitro.

*Specific objectives:*

To determine the isoprene concentration and medium formulation required to perform the enrichment of terrestrial samples.

To determine *in-vitro* isoprene degradation rates of soil and phyllosphere samples from isoprene-emitting trees.

To establish an alternative method to obtain phyllosphere samples from trees species not endemic to the UK.

To identify and characterize novel non-*Rhodococcus* isolates that use isoprene as a sole carbon source.

## 3.2 Results

### 3.2.1 Samples collected from the terrestrial environment

In order to study the isoprene metabolic potential in natural soils and the phyllosphere, various species of trees with different isoprene-emitting potentials were considered. Soil and leaf samples were collected from four different trees (see **Table 3.1**).

All samples were taken from trees present in their natural environment, except for the oil palm tree. Oil palms are tropical trees; the tree sampled in this study was from the Palm House greenhouse in Kew Gardens, London.

### 3.2.2 Soil samples used for enrichment assays

Topsoil samples were taken from 5 - 10 cm underneath the surface, close to isoprene-emitting trees. Each soil sample was obtained from two or three random points around the tree and mixed together before storing at room temperature. A list of soil samples is included in **Table 3.2**.

All samples were transported to the lab in sterile bags or sterile Falcon tubes and were processed the same day (except for tyre dump samples). Twigs, roots and rocks were removed carefully; large pieces of soil were broken down and mixed well before separating into replicates.

### 3.2.3 Leaf samples collected from isoprene-emitting trees

Leaf samples i.e. leaflets, were gathered from the lower crown of the tree, generally in the outer limits of the canopy that is directly in contact with sunlight. Broad, smooth and fully developed leaves were chosen over fragile or ill-shaped leaves. **Table 3.3** gives a list of the tree species that were sampled, and their locations.

**Table 3-1** Isoprene emission-potential of trees sampled during this study.

| Common name | Scientific name | Isoprene emission potential (μg $gdw^{-1}h^{-1}$)* |
|---|---|---|
| African oil palm tree | *Elaeis guineensis* | 172.9 |
| White willow tree | *Salix alba* L. | 37.2 |
| Eucalyptus/gum tree | *Eucalyptus* sp. | [0.96]-18 |
| European ash tree | *Fraxinus excelsior* | <0.1 |

*The symbol [ ] represents a less reliable result according to the reference (Hewitt & Street, 1992).Emission potential was calculated with a photoionization detector (PID) by enclosing approximately 20 $cm^2$ of leaf surface at with near optimum temperature and light conditions (30 °C and 1000 μmol $m^{-2}$ $s^{-1}$ PAR (photosynthetically active radiation)) for 3-5 minutes. High emitters were considered to produce >50 PID units or >0.10 μg isoprene $cm^{-2}$ $h^{-1}$ (Klinger *et al.,* 1998).

**Table 3-2** Reference names for soil samples used during enrichment and isolation experiments. Soil sample origin as well as the location is included.

| Soil sample origin | Location name | Location coordinates | Reference name |
|---|---|---|---|
| Willow tree topsoil | Colney fields, Norwich | 52°37'17.2"N 1°13'46.3"E | Colney willow soil (Colney-WS) |
| Willow tree topsoil | UEA broads, Norwich | 52°37'06.6"N 1°14'09.9"E | UEA willow soil (UEA-WS) |
| Tyre dump soil | Fakenham, Norfolk | 52°50'54.6"N 0°44'49.2"E | Tyre dump soil (TDS) |
| Oil palm tree topsoil | Kew Gardens, London | 51°28'43.6"N 0°17'33.1"W | Kew Oil palm soil (OPS) |

**Table 3-3** List of trees for phyllosphere samples and sampling sites used during this study.

| Leaf sample | Location | Reference name |
|---|---|---|
| Willow tree leaves | UEA Broad | UEA willow leaves (WL) |
| Ash tree leaves | UEA Broad | UEA ash leaves (AL) |
| Oil palm tree leaves | Palm House, Kew Gardens | Kew Oil palm leaves (OPL) |

Compound leaves were kept on their branches and transported to the lab in sterile plastic bags. In the laboratory, leaves were cut from the petioles, weighed and measured in agreement with the pre-determined enrichment procedure (details in **section 2.9**).

#### *3.2.3.1 Washing leaves to obtain microbial cells from the phyllosphere*

Phyllosphere microbes from isoprene-emitting trees were obtained by washing leaves. This technique allows the sampling of bacteria cells from the surface of leaves (details in **Chapter 2 section 2.9**). Different approaches to the processing of leaves were tested, including standardized weight and surface area, for different types of isoprene-emitting trees. Leaf weight and/or surface area were measured before the enrichment assays.

All the trees sampled had compound leaves with different phenotypic characteristics. The most distinguishing features were size, weight, texture, total surface area and 3D structure (eg. Ash leaf creases in **Figure 3.1C**). Physical features of the leaves may have influenced the number of cells recovered from leaf-washings but these were not considered as factors to be studied in this work. Some of the phenotypic characteristics of the leaves are shown in **Figure 3.1**.

*Swabbing leaves to obtain microbial cells from the phyllosphere*

Oil palm trees have the highest isoprene emission potential reported (Hewitt & Street, 1992) refer to **Table 3.1**. The leaf-swabbing procedure, generally used to evaluate persistence of harmful bacteria on hospital surface areas, was used to recover microbial cells from the phyllosphere of isoprene-emitting trees (detailed in **section 2.9**). This is an alternative method, especially when sampling non-native tree species. Tests to assess the swabbing procedure were done with isoprene-emitting trees native to the UK before evaluating the method on oil palm trees.

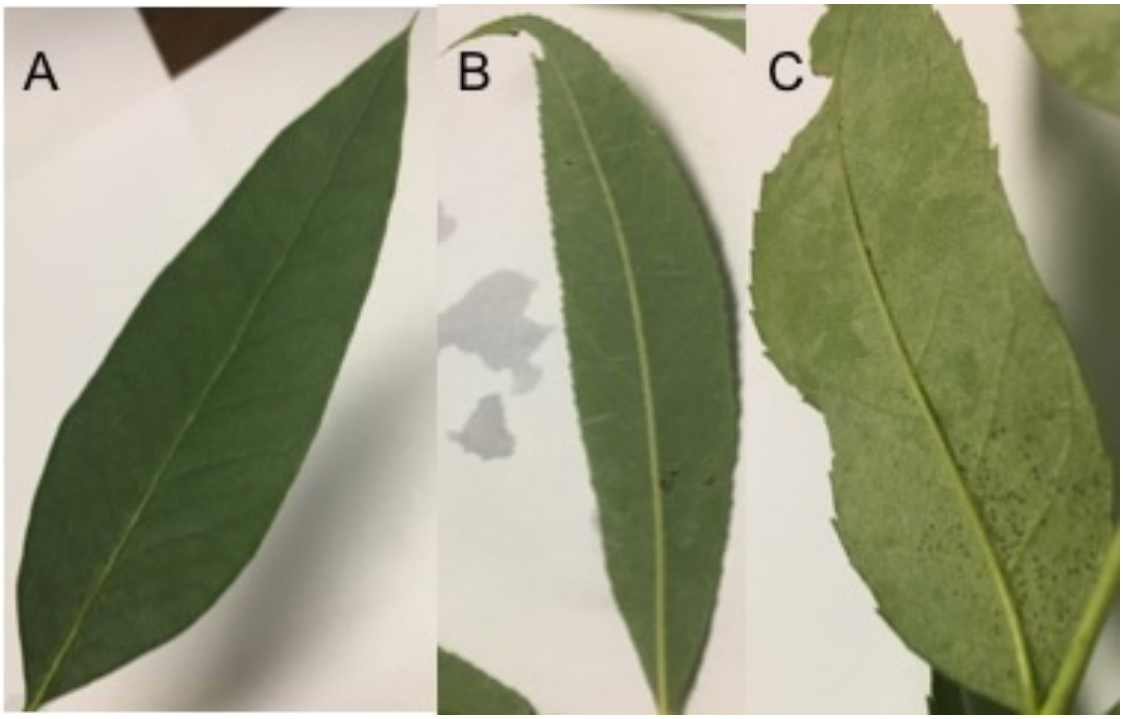



**Figure 3.1** Phenotypic characteristics of leaves from A) a eucalyptus tree, B) a willow tree and C) an ash tree. The size of the leaves shown in the figure is not the actual size of the leaf.

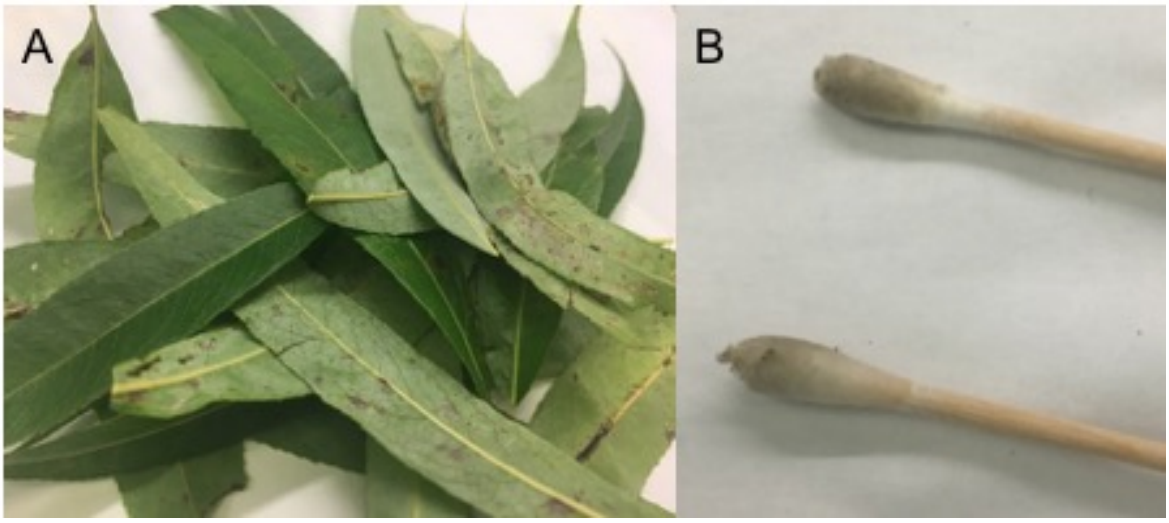



**Figure 3.2** Swabbing procedure A) preparation of willow leaves and B) two cotton swab tips dry (above) and wet (below) after phyllosphere sampling. A wet and a dry cotton swab were used for each pair of willow leaves.

The leaf-washing protocol was used as the base for the design of the leaf-swabbing methodology (refer to **section 2.9**). The first method was altered to suit the characteristics of the tree species sampled.

### 3.2.4 Evaluation of isoprene consumption from soil and leaf samples

The isoprene consumption assays aimed initially to reproduce work from previous soil enrichments, and later to recreate conditions that would mimic environmental conditions. Khawand *et al*., 2016 and Johnston *et al*., 2017 used concentrations of isoprene up to 1% (v/v) i.e. 10,000 ppmv. In the natural environment, the maximum concentrations of isoprene are much lower as mentioned in Chapter 1 (El Khawand *et al*., 2016; Johnston *et al.*, 2017). In this study, the isoprene concentrations tested ranged from 12 - 250 ppmv. Decreasing the concentration of isoprene in the headspace resulted in considerably shorter incubation times. Incubation periods that previously took a few weeks for significant assimilation of isoprene, took a total of 10 to 13 days when used with lower isoprene concentrations. The latter strategy, however, did require close and constant monitoring of isoprene uptake and multiple injections of additional isoprene throughout the assay.

#### *3.2.4.1 Consumption of isoprene in soil samples*

Different experimental conditions were studied to determine the initial concentration of isoprene and supplementary nutrients. Since isoprene is a reactive and mutagenic alkene with a strong smell, high concentrations may be toxic (Srivastva *et al.*, 2018). To achieve conditions a little closer to those in the natural environment, the concentration was lowered 100-fold to 0.005 % (v/v), i.e. nominal concentration of 50 ppmv. Diluted and full strength minimal media (MM), as sources of external nutrients, were initially evaluated.

##### 3.2.4.1.1 Willow soil enrichments (Colney-WS)

Soil isoprene enrichment cultures were set up with 5 g of soil collected from beneath a willow tree at the University of East Anglia sports fields, close to the River Yare (Colney-WS, see **Table 3.2**). Microcosm incubations were performed in sealed and sterile 2 litre airtight flasks with 50 ml Ewers minimal medium (MM) at two concentrations (1X or full, 0.1X or diluted) and an initial isoprene concentration of approximately 25 ppmv or 0.0025 %. Flasks were shaken at 150 rpm at 25 °C in the dark. Isoprene consumption was monitored using

gas chromatography (GC, see **Figure 3.3**) and isoprene was replenished when consumed by the soil microbiome.

Water and double sterilized soil were used as controls for leaking or abiotic depletion of isoprene, respectively. These negative controls were only spiked at the beginning of the assay. Data show both controls behaved similarly during the whole experiment and no abiotic removal or leakage of isoprene was detected from the vials.

Colney-WS consumption of isoprene was detected after 36 h of incubation, and the 25 ppmv was consumed in less than 48 h. After 220 h of the assay, 1X MM was spiked with isoprene 3 times, while 0.1X MM was spiked 12 times. The diluted minimal medium treatment consumed 1.2 ml of isoprene vapour during the duration of the assay. The rate of consumption of isoprene for the diluted MM increased as the enrichment progressed, requiring a more frequent spiking of isoprene. At the end of the experiment, the sample was supplemented with higher amounts of isoprene and the consumption rates were maintained.

Results with the diluted minimal medium suggested that incubations could be performed with little or no addition of nutrients. The next experiment performed compared a water slurry to the diluted minimal medium enrichment. The concentration of isoprene (approximately 25 ppmv or 0.0025 %.) was closely monitored and maintained by adding isoprene when consumption was detected. Results are shown in **Figure 3.4.**

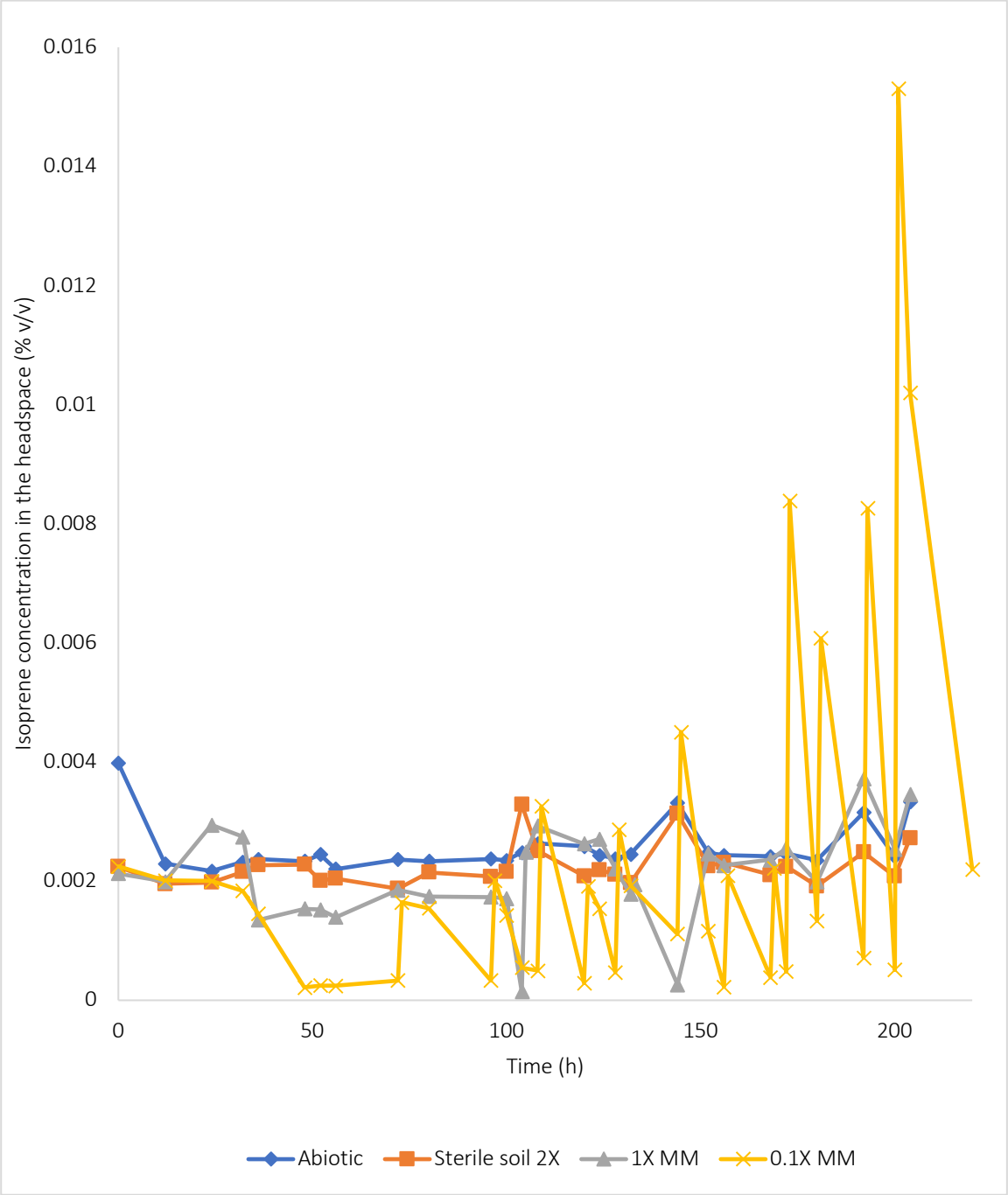



**Figure 3.3** Isoprene consumption of Colney willow soil (Colney-WS) sample with different concentrations of minimal medium (1X and 0.1X). Enrichment was performed for 220 h. No replicates for each treatment were done in this experiment. Abiotic, vials with no soil added. Sterile soil 2X, soil sterilised by autoclaving twice.

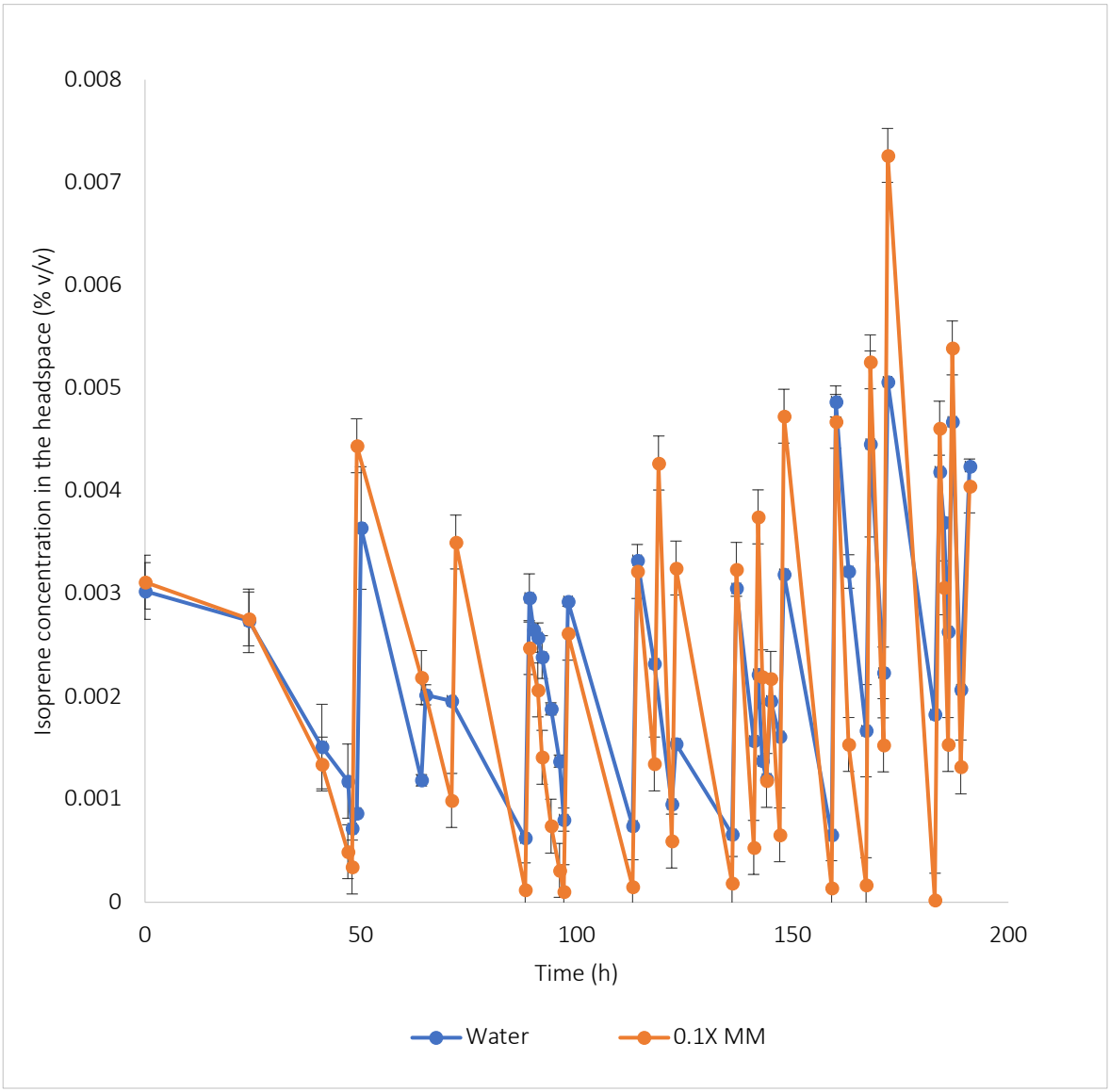



**Figure 3.4** Isoprene consumption of Colney-WS sample with 0.1X MM and a water slurry. The experiment was performed in triplicate and error bars show the standard deviation. Data not shown for the abiotic control, vials with no soil added, and sterile soil 2X, soil sterilized by autoclaving twice, since there was no change in the concentration of isoprene during the experiment.

Even though consumption rates were slightly higher in the 0.1X MM incubation, consumption with the water slurry was similar. Adding sterile water and lower concentrations of isoprene to the microcosm simulated environmental conditions while enriching isoprene-degrading bacteria. Results from the water slurry enrichment experiment (incubation time, rates and concentration of isoprene) from **Figure 3.4** were used for the Colney-WS and OPS stable isotope probing experiments in **Chapter 4** and **5**.

#### 3.2.4.1.2 Isoprene enrichments using tyre dump soil (TDS) and enrichment consumption rates for soil samples

Srivastva and colleagues published a study in which they isolated bacteria that degraded isoprene from soil in a tyre dump site (Srivastva *et al*. 2015). In the search for diverse isoprene-degraders, different soil samples were collected from a tyre dump site close to Fakenham in June 2016. Three soil samples were enriched using the same conditions as the previous water slurry experiment. Colney-WS was also sampled as a point of comparison for the enrichment process. Isoprene consumption was monitored and results are shown for the first 70 h in **Figure 3.5**.

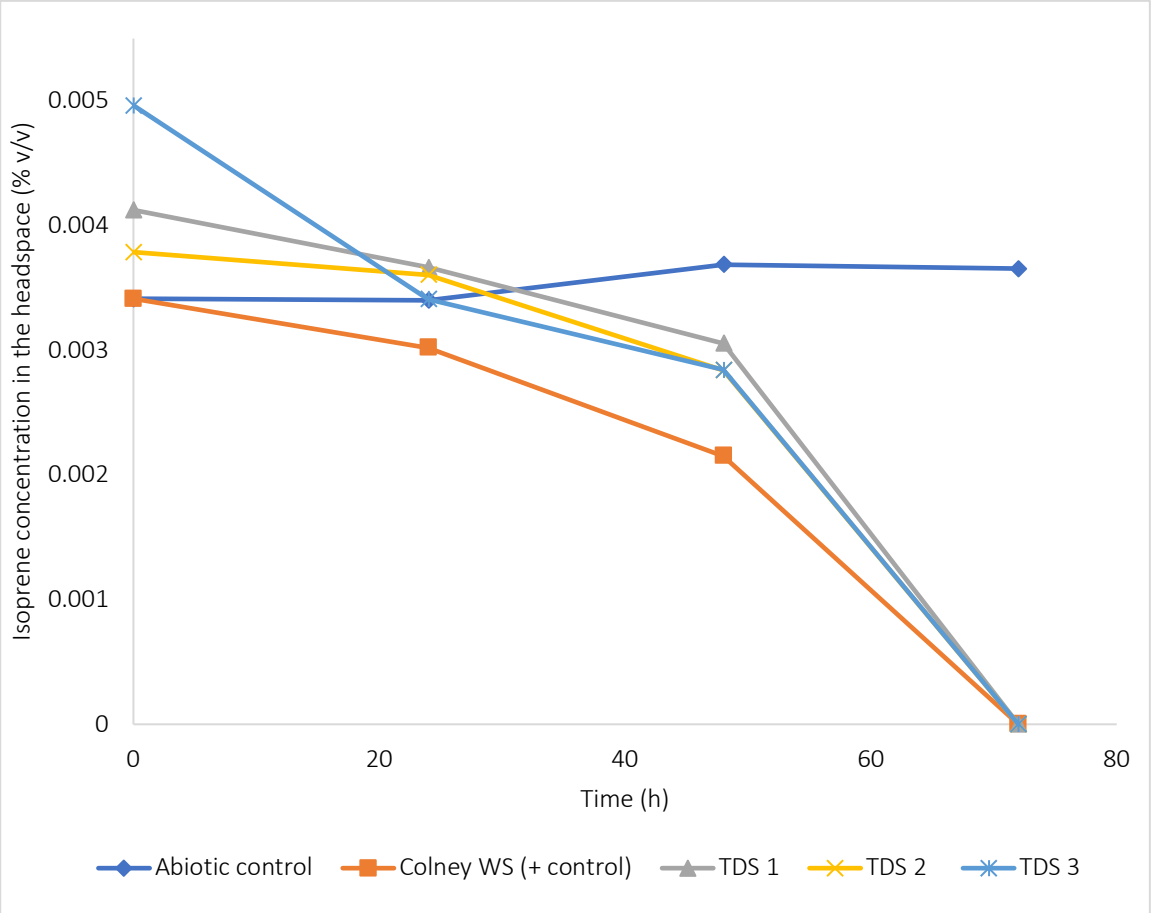



**Figure 3.5** Isoprene consumption by tyre dump soil (TDS) samples. Colney WS used as a positive control.

The TDS isoprene consumption results were comparable to Colney-WS samples. Isoprene consumption for the different enriched soil samples was compared by calculating their estimated consumption rates (**Table 3.4**).

**Table 3-4** Estimated isoprene consumption rate (%) for soil sample enrichments. The consumption rate was calculated for the initial consumption of isoprene. Description includes date the samples were taken, replication, and if the sample was stored or incubated immediately after sampling.

| Sampling site | Media | Consumption rate ($\mu mol\ h^{-1}\ g^{-1}$) | Description |
|---|---|---|---|
| Colney-WS | 0.1X MM | 0.034 | April 2015, no replicates, immediate |
| Colney-WS | 0.1XMM | 0.030 | July 2015, 3 replicates, immediate |
| Colney-WS | Water | 0.027 | July 2015, 3 replicates, immediate |
| TDS | Water | 0.016 | June 2016, biological replicates, stored |

Results from the isoprene enrichment of soil samples have shown an efficient consumption of 0.002 to 0.005% (v/v) isoprene in 2 to 3 days. The rates of isoprene consumption increased almost 2-fold from samples that were fresh. The results may be due to higher microbial activity from fresh samples close to isoprene-emitting trees. Since the fresh samples consumed the isoprene quicker, they were used for subsequent experiments. Stable isotope probing experiments (see **Chapter 4** and **5**) were processed the same day, following these guidelines, to help optimize the rate at which microbial activity was observed during the assay.

To keep consumption rates high and abundant isoprene-degraders, the successful enrichments were sub-cultured and studied for isolation of isoprene-degrading bacteria (further information in **section 3.2.5**). Sub-cultures were obtained by adding 1 ml of enrichment culture to 9 ml of fresh minimal medium in a clean sterile airtight 120 ml serum vial. The concentration of isoprene was maintained the same as in the initial enrichment for a few weeks. Sub-cultured incubations had higher isoprene consumption rates and contained less soil particles (data not shown).

#### 3.2.4.2 *Isoprene degradation of phyllosphere samples from different isoprene-emitting trees*

The first enrichment was performed to compare isoprene consumption by cells recovered from willow (WL) and ash (AL) trees at the University of East Anglia (data not shown). The results varied between trees and suggested that cells from leaves of high isoprene emitting trees have a faster isoprene consumption rate than those from low isoprene emitting trees. Even though the same weight of leaves (5 g) was used during the sample preparation, there were differences in the amount of surface area washed due to the thickness and density of the leaves.

A new phyllosphere cell incubation from WL and AL was performed taking into consideration the surface area of leaves. This experiment included phyllosphere samples from a Eucalyptus tree (EL) collected from The Avenues, close to UEA (Norwich, UK) and results for isoprene consumption are shown in **Figure 3.6**. Cells from an equal surface area (approximately 294 $cm^2$) were washed from the leaves and added to each replicate.

A larger variation in isoprene consumption between replicates was observed compared to results from soil enrichments. When the initial amount of isoprene was spiked into 2 litre airtight vials, there was approximately 5 to 10 percent of injection error recognized. Even when accounting for this error, the difference between consumption times obtained between replicates of the same type of tree (intraspecific variation) was larger than expected. The interspecific results show willow tree epiphyte cells consumed the isoprene quicker than ash tree cells. The results do however reflect the variation in between leaves of the same tree. Each replicate was therefore analysed individually to account for the possibility of differing abundance and divergent diversity of phyllosphere microbes present on each leaf.

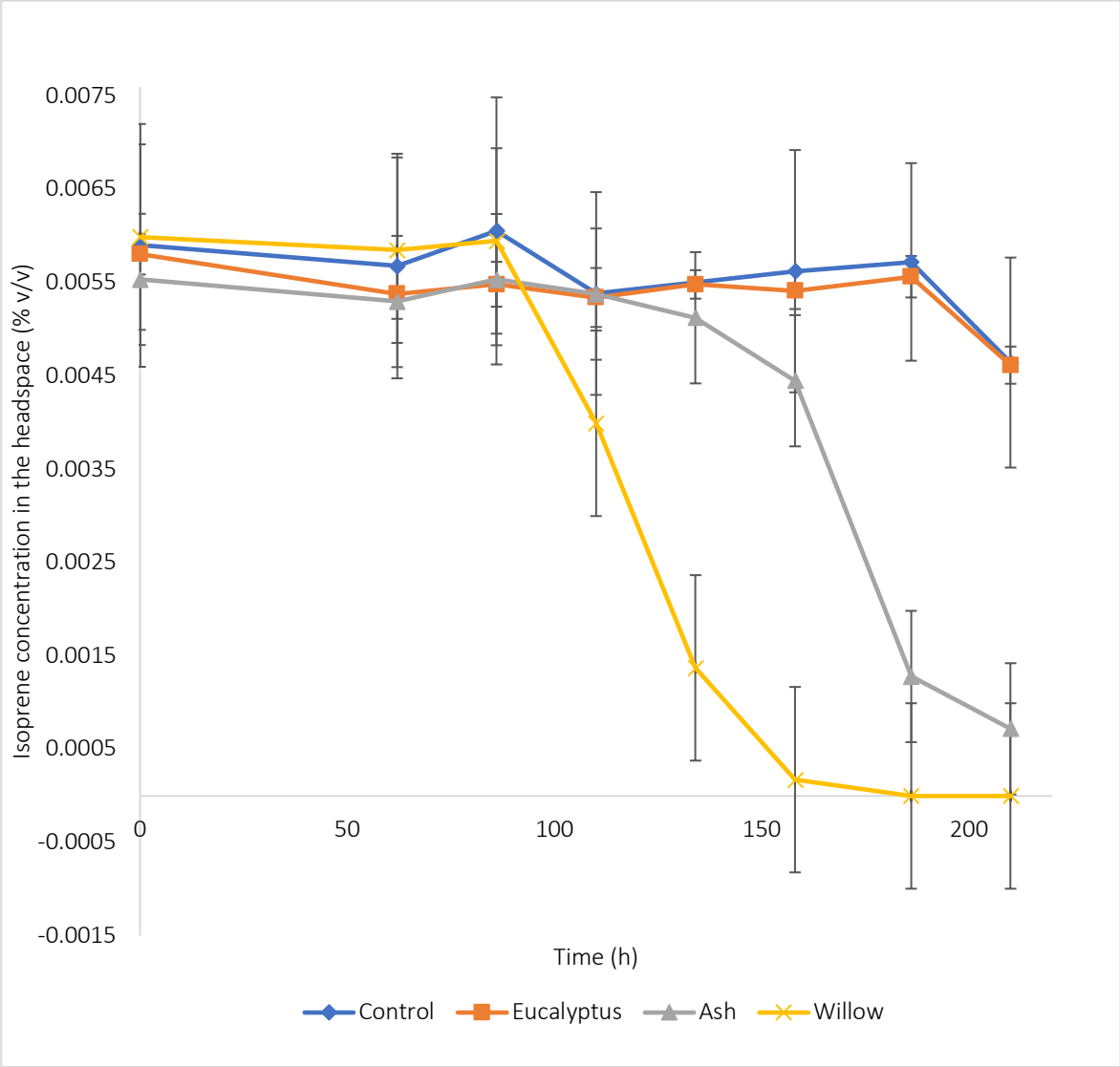



**Figure 3.6** Isoprene consumption of ash (AL), willow (WL) and eucalyptus (EL) leaf- washings. Experiment performed with 5 replicates and error bars show the standard deviation.

Ash trees are reported to have lower isoprene-emission potential compared to willow trees (**Table 3.1**). The results shown in **Figure 3.6** suggested that cells recovered from high isoprene emitting trees had higher isoprene consumption rates (shown in **Table 3.5**). This implies that each species of tree has different phyllosphere microbiomes, selected by the plant, and one of the factors that may affect the abundance and diversity of the isoprene-degraders present on the leaves is the amount of isoprene the tree naturally emits.

**Table 3-5** Estimated consumption rate (%) of isoprene for AL and WL washed phyllosphere samples. The average standard deviation of the samples was 0.00099 μmol $h^{-1}$ $g^{-1}$ for ash leaves and 0.00022 μmol $h^{-1}$ $g^{-1}$ for willow leaves.

| Sampling site | Hours to consume first spike | Consumption rate (μmol $h^{-1}$ $g^{-1}$ of leaves washed) |
|---|---|---|
| Ash tree leaves | ~230 | 0.010 |
| Willow tree leaves | ~186 | 0.018 |

Most eucalyptus tree species are reported as high emitters in the literature (C. Nicholas Hewitt & Street, 1992). The results shown in **Figure 3.6** do not show any isoprene consumption from cells recovered from the eucalyptus leaves (EL) tree sampled in this study. EL are waxy, small and light-weight which allowed for more leaves to be washed in comparison to the AL and WL. The physical characteristics of the leaves, such as the absence of creases and waxy texture, might affect the number of microbe cells that are present on the leaves before washing (**Figure 3.1A**). Despite using a larger number of EL to obtain the same surface area, the cells recovered did not consume isoprene. This result suggests that the species of Eucalyptus tree sampled did not produce isoprene and/or the physical characteristics of the leaves affected the number of phyllosphere cells recovered.

Minimal medium concentration experiments were also performed to observe the effect of nutrient supplementation on isoprene consumption in phyllosphere enrichments (shown in **Figure 3.7A**). In this experiment, the washed cells were divided into 5 replicates. A final volume of 50 ml was obtained by adding different amounts of minimal medium and water (0.2X, 0.4X, 0.6X, 0.8X and 1X). Results showed there was no consumption of isoprene from cells washed off willow leaves, which was unexpected when comparing with previous results (see **Figure 3.6**). The experimental procedure changed in this assay when the standard number of cells washed was divided into replicates. The division of the original cells washed, into 5 replicates, might have affected isoprene consumption from WL. This argument does not explain isoprene consumption by AL, previous results showed creases on the leaves that might have helped trap and subsequently wash a larger number of cells (**Figure 3.1C)**. This result is the first evidence that suggests that to obtain reliable data, a

standard and representative amount of leaves needs to be washed for each replicate experiment.

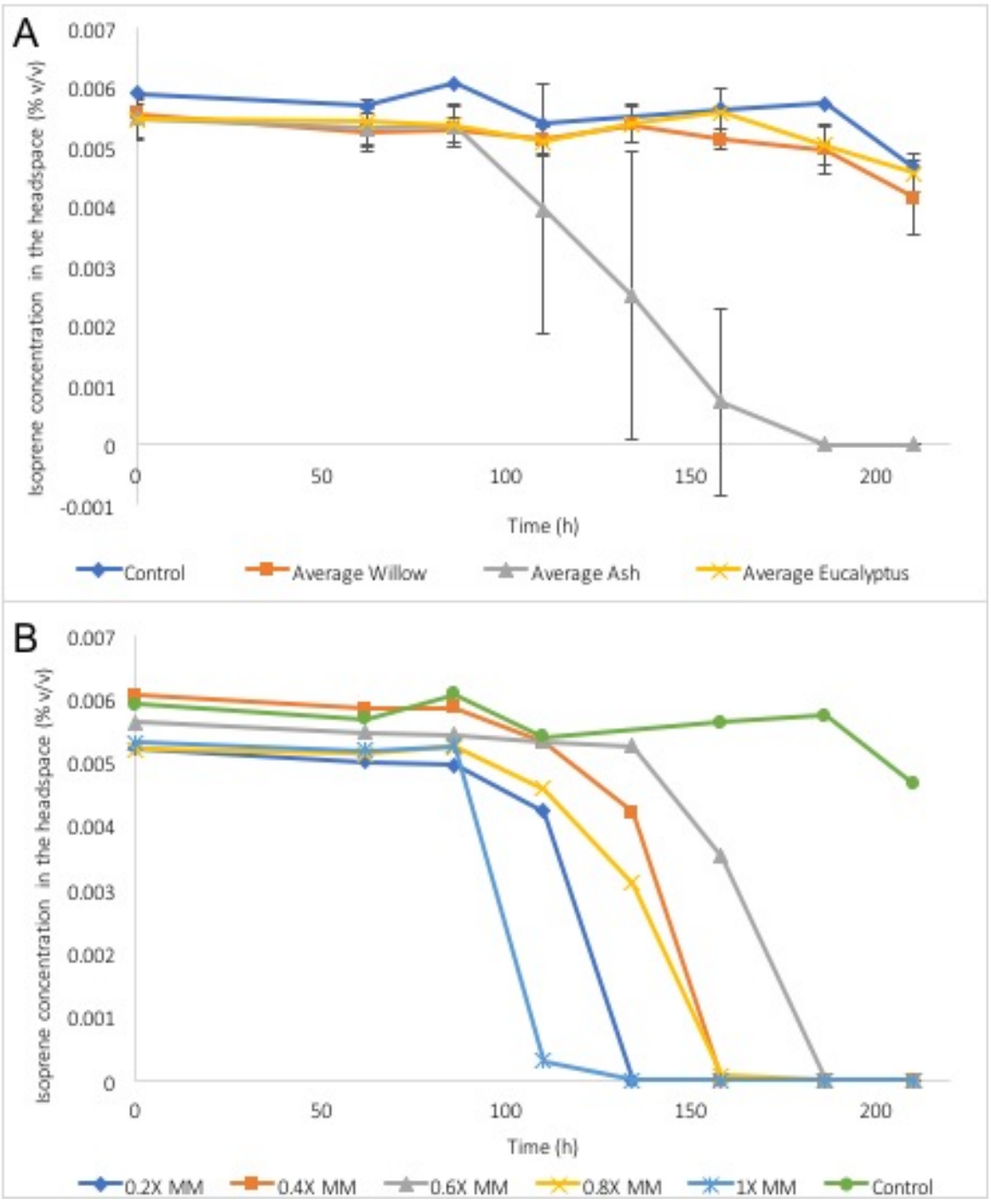



**Figure 3.7** A) Average isoprene consumption from ash (AL), willow (WL) and eucalyptus (EL) leaf-washings with different concentrations of minimal medium. Five replicates for each leaf type and error bars show the standard deviation. B) Isoprene consumption of individual replicates from ash (AL) leaf-washings with different concentrations of minimal medium.

Despite willow phyllosphere cells not using isoprene in the assay, ash and eucalyptus cell results did replicate previous data. It is important to note that the five replicates from ash phyllosphere cells consumed isoprene at different rates (see **Figure 3.7B**). Once again,

these results show the large variation between consumption rates of replicates, even when the samples come from the same original leaf washing procedure. Furthermore, there was no clear pattern between isoprene consumption and different concentrations of minimal medium. The fastest consumption was in the replicates that had full 1X MM and most diluted 0.2X MM. The results suggest the concentration of minimal medium did not affect the consumption of isoprene.

In general, the consumption of 0.002% to 0.005% (v/v) was detected after 5 to 9 days of incubation from cells washed-off from a representative number of leaves. The three dimensional (3D) characteristics i.e. creases, and texture of the leaves may be important factors that influence the number of cells that will be washed off from leaves. Some variation in the isoprene consumption of different trees may also be explained by an interspecific variation between phylloplane microbiomes. These results have shown intraspecific variations to be another source of differentiation in the phylloplane of a tree (Niinemets *et al*., 2010).

#### *3.2.4.3 Evaluation of leaf-swabbing as an alternative method to obtain phyllosphere isoprene-degraders*

Oil palm trees are important in the production of plant based fuels (biofuels) and alternative sources of energy, but they also have the highest isoprene emitting potential reported (Hewitt & Street, 1992). During this study, the locations to obtain oil palm tree phyllosphere samples were limited in the UK. Leaf-swabbing was explored as an option for obtaining phyllosphere cells without sacrificing parts of the plant and transporting specimens between cities/countries. Leaf-swabbing could also permit larger surface areas of leaves to be sampled.

**Figure 3.8** shows the results of the first leaf-swabbing experiment using 5 g of leaves from ash and willow trees, divided into 4 replicates each. There was no consumption of isoprene from cells recovered in the leaf-swabbing from 5 g of leaves. This suggested that the amount of leaves that needed to be swabbed should be greater than 1 g per replicate. As determined in **Figure 3.7A**, dividing the total amount of washed/swabbed leaves was not

recommended. Leaf swab experiments might show better results with a greater amount of leaf surface sampled.

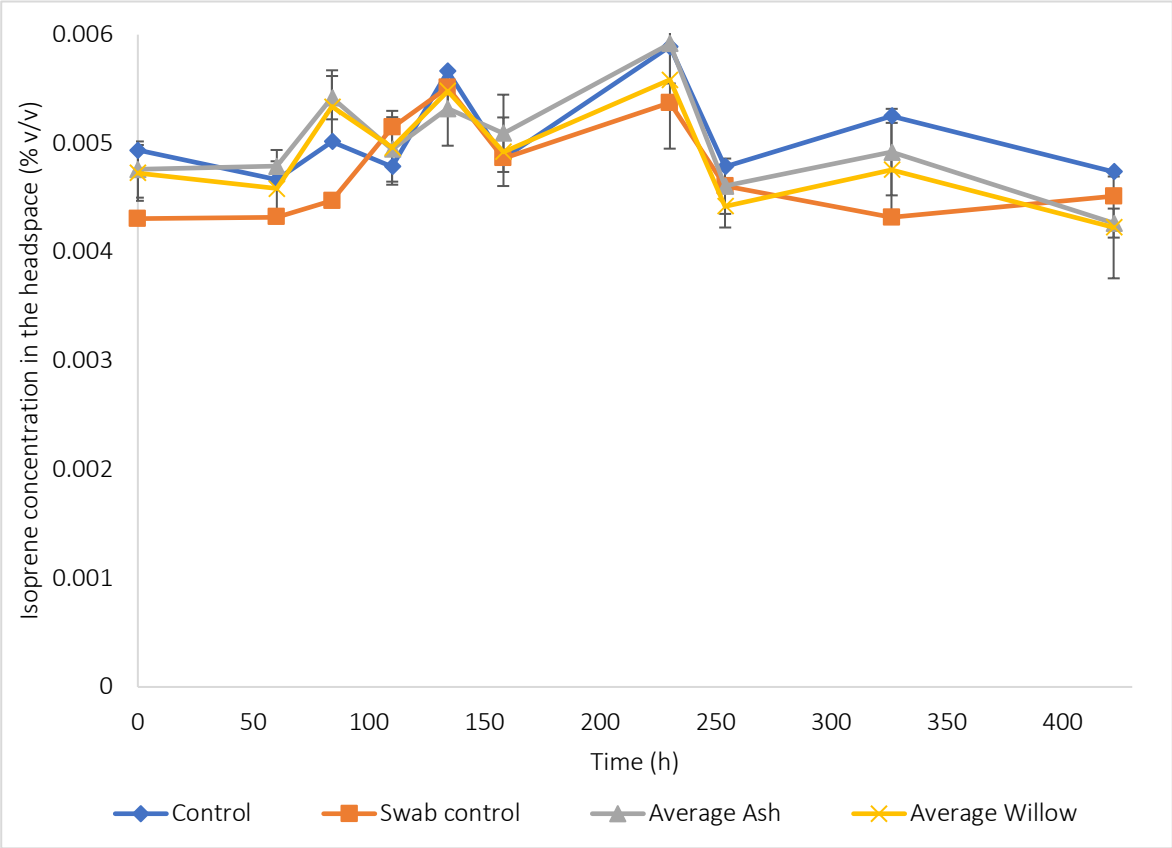



**Figure 3.8** Isoprene consumption from ash (AL), willow (WL) leaf swabs. Swab control, sterile swabs, and abiotic controls, with neither swabs nor cells, were included in the study. Four replicates for each leaf type and error bars show the standard deviation.

Despite obtaining negative results using this procedure at the first attempt, oil palm leaves were tested. Oil palm leaves are bigger and have a greater surface area. This helped in the recovery of larger amounts of phyllosphere cells from leaves. Initial experiments with OPL washing gave no isoprene consumption when working with one leaf (data not shown), which suggested that a larger amount of surface area needed to be used during the experiments.

Oil palm leaves from the Palm House were sampled with sterile cotton swabs, instead of cotton buds, in November 2016. The assay was performed by swabbing 3 leaves (of approximately 10 g each) and efficient isoprene consumption was observed (data not shown). For the next assay, the oil palm leaf swabbing (OPLS) technique was compared to

the leaf-washing experiment (OPLW) to establish if there was any difference between these two protocols when larger surface areas were considered. The experiment was done in duplicate by washing or swabbing 3 oil palm leaves for each replicate (approx. 30 g each). The results of the isoprene uptake experiments during the first 160 h are shown in **Figure 3.9.**

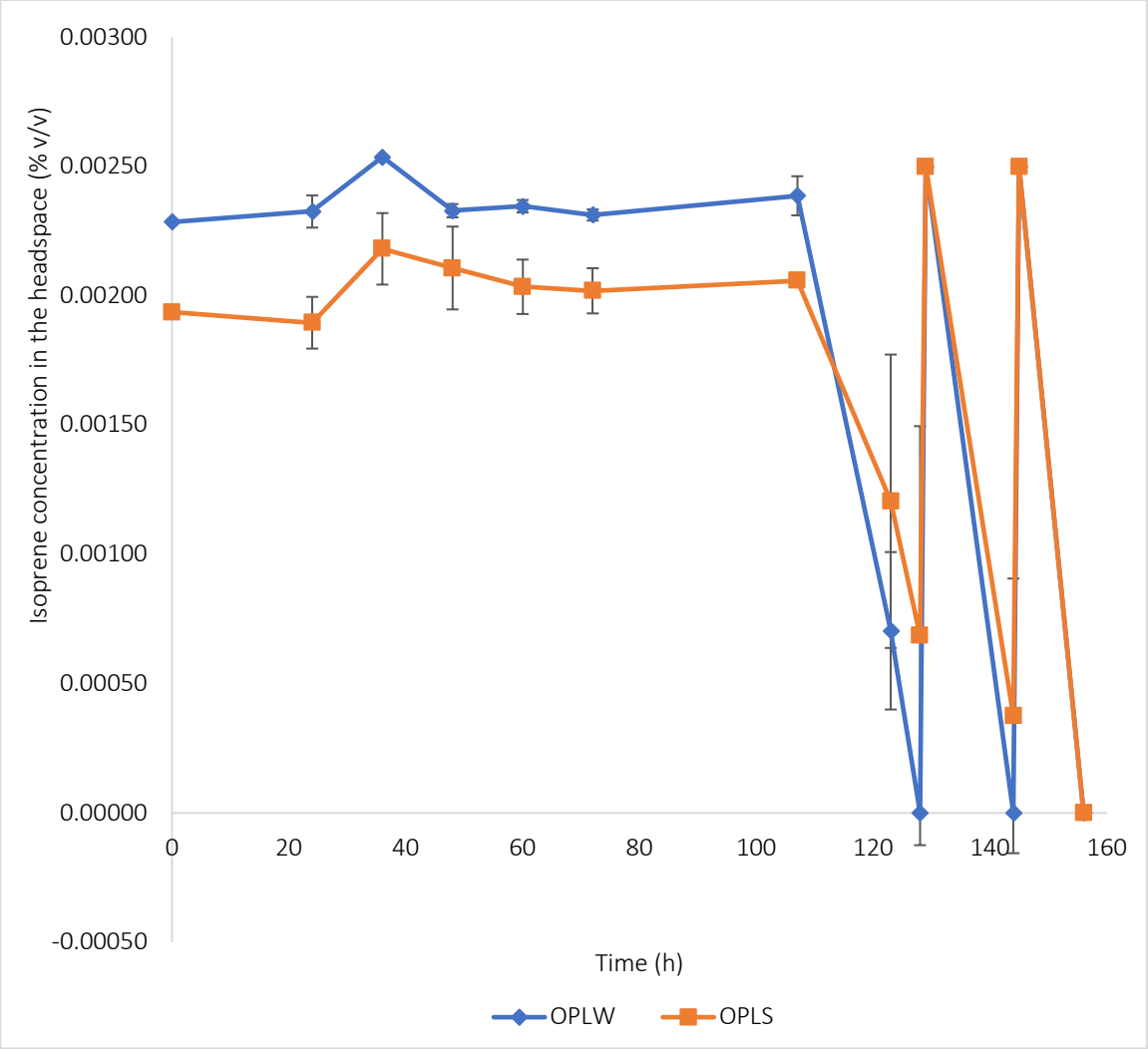



**Figure 3.9** Isoprene uptake of oil palm leaf-washings (OPLW) and oil palm leaf swabs (OPLS). Each treatment was done in duplicate and error bars show the standard deviation.

Oil palm leaf-washings and leaf-swabs yielded comparable results when larger amounts of surface area were used. OPLW and OPLS require similar times to consume the added isoprene (between 130-140 h). Leaf-swabbing can therefore be considered a good method to obtain phyllosphere microbes when sampling large leaf surface areas. In future

experiments, measuring the optical density of samples before the incubation begins may help to understand how effective each of the procedures was at removing cells from the surface of leaves.

### 3.2.5 Isolation of isoprene degrading bacteria

Soils and leaf-phyllosphere samples, that effectively degraded the isoprene, were continuously enriched with 13 to 50 ppmv isoprene for 3 to 8 weeks. Serial dilutions on minimal medium agar plates (supplemented with Noble Agar or Phytagel) incubated with approximately 1% (v/v) isoprene in the headspace and rich media plates (R2A and Nutrient Agar) were used to obtain isolates. Some bacterial colonies appeared after 24 hours while others were monitored for up to 8 weeks to obtain slower growing bacteria. To compare between cultivation-dependent and cultivation-independent data, enriched cultures from DNA stable isotope probing incubations were also monitored, sub-cultured and re-streaked in order to optimize the isolation of bacteria.

More than 250 putative isolates were obtained from Colney-WS, UEA-WS, TDS, WL and OPL on minimal and rich media. Each isolate was purified and tested for its ability to use isoprene as their sole carbon and energy source. Isoprene depletion in minimal medium liquid cultures was monitored using gas chromatography (GC). After enrichment at low isoprene concentrations, isolates were tested at higher isoprene concentrations in 120 ml sterile glass serum vials with 1% (v/v) isoprene. Bacteria that could use the isoprene and grow, evidenced by an increase in $OD_{600}$, were identified using 16S rRNA gene PCR. Purity was checked by light microscopy (1000X). When cultures were confirmed as pure (microscopy, plating, etc.; refer to **section 2.4**), glycerol stocks were made and stored at -80 °C. A list of the isolates obtained during this study is shown in **Table 3.6**.

**Table 3-6** Isolates obtained from enriched cultures. Information includes the sampling site (enrichment culture), closest blastn identity for the 16S rRNA gene, and whether the genome was sequenced. WS represents isolates isolated from willow soil, TD from tyre dump, WL from willow leaves and OPL from oil palm leaves.

| Isolate name | Sample of origin | 16S rRNA gene identified as (% identity) | Genome sequenced |
|---|---|---|---|
| *Rhodococcus* sp. strain WS1 | UEA/Colney fields | *Rhodococcus* sp. djl -6 -2 (100%) | yes |
| *Rhodococcus* sp. strain WS2 | UEA/Colney fields | *Rhodococcus* sp. AD45 (99%) | no |
| *Rhodococcus* sp. strain WS3 | UEA/Colney fields | *Rhodococcus* sp. AD45 (100%) | yes |
| *Rhodococcus* sp. strain WS4 | UEA/Colney fields | *Rhodococcus opacus* strain 1CP (99%) | yes |
| *Rhodococcus* sp. strain WS5 | UEA/Colney fields | *Rhodococcus* sp. AD45 (100%) | no |
| *Rhodococcus* sp. strain WS6 | UEA/Colney fields | *Rhodococcus* sp. AD45 (99%) | no |
| *Rhodococcus* sp. strain WS7 | UEA/Colney fields | *Rhodococcus* sp. djl -6 -2 (100%) | yes |
| *Rhodococcus* sp. strain WS8 | UEA/Colney fields | *Rhodococcus* sp. AD45 (99%) | no |
| *Variovorax* sp. strain WS9 | UEA/Colney fields | *Variovorax* sp. BZ15 (98%) | yes |

| | | | |
|---|---|---|---|
| *Rhodococcus* sp. strain WS10 | UEA/Colney fields | *Rhodococcus* sp. AD45 (98%) | no |
| *Variovorax* sp. strain WS11 | UEA broad | *Variovorax* sp. RA8 (99%) | yes |
| *Nocardioides* sp. strain WS12 | UEA broad | *Nocardioides aromaticivorans* strain H9 (97%) | yes |
| *Variovorax* sp. strain WS13 | UEA broad | *Variovorax* sp. RA8 (99%) | no |
| *Rhodococcus* sp. strain TD1 | Tyre dump | *Rhodococcus jostii* str IFO 16295 (99%) | no |
| *Rhodococcus* sp. strain TD2 | Tyre dump | *Rhodococcus koreensis* str Sedi2 (99%) | no |
| *Rhodococcus* sp. strain TD3 | Tyre dump | *Rhodococcus* sp. CAP 110 (99%) | no |
| *Rhodococcus* sp. strain WL1 | UEA broad | Rhodococcus *opacus* 04-OD7 (97%) | no |
| *Rhodococcus* sp. strain OPL1 | Kew Gardens | *Rhodococcus* sp. strain MAK1 (100%) | no |
| *Gordonia* sp. strain OPL2 | Kew Gardens | *Gordonia terrae* NRRL B-16283 (99%) | yes |
| *Sphingobacterium* sp. strain OPL3 | Kew Gardens | *Sphingobacterium* sp.VA-15b (99%) | no |
| *Sphingopyxis* sp. strain OPL5 | Kew Gardens | *Sphingopyxis* sp. QXT-31 (99%) | yes |

Exciting isolates, representing bacteria that were different to the strains present in the lab and in the published literature, were sent for genome sequencing. Two *Variovorax,* one *Nocardioides*, and *Rhodococcus* species which were different from our model *Rhodococcus* sp. AD45, were selected from the willow soil enrichments (*Variovorax* sp. strain WS9, *Variovorax* sp. strain WS11, *Nocardioides* sp. strain WS12, *Rhodococcus* sp. strain WS1, *Rhodococcus* sp. strain WS3, *Rhodococcus* sp. strain WS4 and *Rhodococcus* sp. strain WS7). Together with one *Gordonia* isolates and one *Sphingopyxis* isolates (*Gordonia* sp, strain OPL2, *Sphingopyxis* sp. strain OPL5), recovered from oil palm leaves, they represent a larger diversity of isoprene-degraders from the terrestrial environment.

#### *3.2.5.1 Targeted isolation of other Variovorax-like isolates*

The first Gram-negative isoprene-degrading bacterium obtained in this study was identified as part of the *Variovorax* genus. *Variovorax* sp. strain WS9 was the ninth isolate from the Colney-WS isoprene enrichment in July 2015 and the only one that was not a *Rhodococcus*. The isolate was initially a slow growing bacterium, therefore the optimum conditions for its growth were studied in detail.

This isolate represents an exciting bacterium for the study of isoprene degradation by Gram-negative bacteria in soils. The optimal growth of the bacteria in minimal medium was established at pH 6.5 with the addition of 1 μl per ml of MAMs vitamins (details in **section 3.2.6**). To search for other Gram-negative isolates from the Comamonadaceae family, the conditions for optimal growth of *Variovorax* sp. strain WS9 were replicated in enrichment cultures after January 2016 (see **Table 2.3**).

#### *3.2.5.2 Isolation of isoprene-degrading bacteria using a sub-cultivation strategy*

Isolation of isoprene-degrading bacteria was achieved from samples with high numbers of microbial cells. Over 250 phenotypically distinct bacteria isolates were obtained from a variety of soil and leaf samples enriched during this study.

Oil palm isolates that could use isoprene as their sole carbon and energy source, took around a year to purify and identify. To maintain favourable conditions within the Palm

House at Kew Gardens, the humidity and temperatures of their native tropical countries are replicated in the facility. The air inside the greenhouse is sprayed with water several times per hour to maintain high humidity. Leaf presses from the oil palm tree were done on minimal medium plates, pH 6.5, and due to the humid environment of the Palm House, mostly fungi grew on the plates over the first few days (data not shown).

Sub-culturing was introduced in enrichment assays of early 2015 to eliminate soil particles from enrichment cultures. Originally it helped with the observation of enriched soil samples under the microscope. A few changes in the experimental design helped the isolation of seven oil palm leaf isolates from considerably dirty samples (see details in **section 2.11**).

The samples were enriched for at least 3 to 4 weeks with low concentrations of isoprene (25 ppmv) to help slow growing bacteria (such as *Variovorax*) to flourish. 1 ml of the previous culture was added to 9 ml of sterile minimal medium after shaking well. Although this method did not guarantee that all the possible isolates were recovered, it did help to eliminate other microbes in the sample that were not using isoprene directly as a sole carbon and energy source.

A total of 4 to 5 sequential subcultures were done for each sample over 6 months. These were later plated on minimal medium with Noble Agar, with Phytagel and R2A rich medium. All plates were initially incubated for 48 h at 30 °C, then incubated for two to three weeks at 21 °C, to account for slow growing bacteria.

### 3.2.6 Novel Gram-negative terrestrial isoprene-degrading bacterium identified as *Variovorax* sp. strain WS9

This was the first isolation of a Gram-negative bacterium, member of the Comamonadaceae family, from the terrestrial environment that uses isoprene as a sole carbon and energy source.

*Variovorax* sp. strain WS9 was a slow growing bacterium. To detect growth of a single colony, on minimal or rich media plates, took at least a week. This non-pigmented isolate was observed in serial dilutions of Colney-WS isoprene enrichments from August 2015. The

tiny colonies were abundant at high dilutions on rich medium, incubated at room temperature for a month. A phase contrast microscopic image of the cells grown on isoprene is shown in **Figure 3.10**.

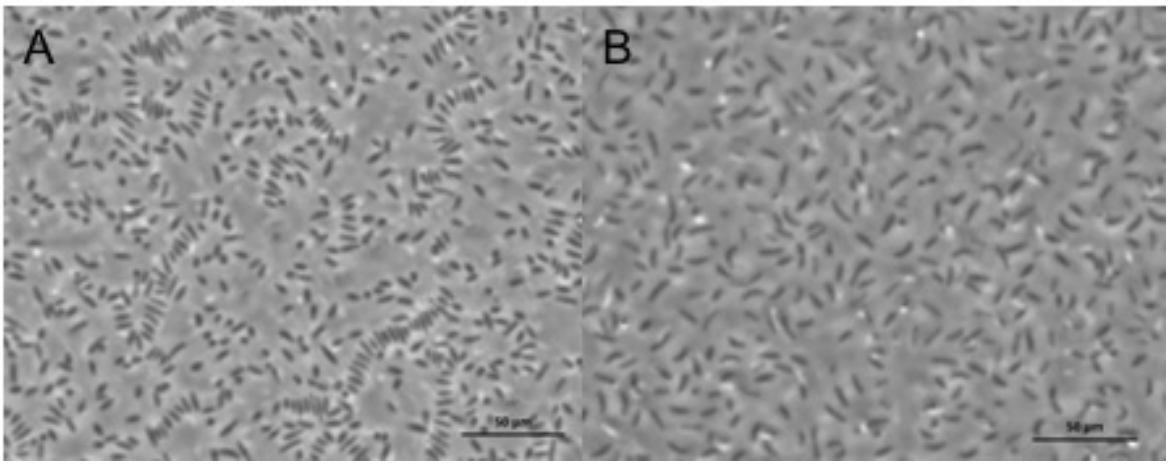



**Figure 3.10** *Variovorax* sp. WS9 bacterium phase contrast microscope images grown on (A) succinate and (B) isoprene (1000X magnification).

DNA was extracted from *Variovorax* sp. strain WS9 cells grown on isoprene using the FastSpin DNA soil kit (MP Biomedicals, details in **section 2.13**). A standard 16S rRNA gene touch-down PCR protocol (**section 2.15.2**) with 27f-1492r primers (**Table 2.1**) was performed using the isolate genomic DNA as a template. 16S rRNA gene amplicons were purified, ligated into the pGEM-T easy vector (Promega), cloned into *E. coli* competent cells, and sequenced using SP6-T7 primers (**Table 2.1**). A blastn search with the complete 16S rRNA gene (National Centre of Biotechnology Information, NCBI) results show 98% sequence identity to *Variovorax* sp. BZ15 isolated from hydrocarbon-contaminated soil (GeneBank accession number HQ588854) and 98% identity to *Caenimonas* sp. SL110 isolated from a freshwater dessert lake in northern China (Gene bank accession number KY122001, Li *et al.*, 2014).

### *3.2.6.1 Optimal growth conditions for Variovorax sp. strain WS9*

*Variovorax* sp. strain WS9 did not grow well on glucose, yeast extract did not affect its growth, and supplementing 1 % isoprene i.e. 10,000 ppmv to the headspace was not toxic (data not shown).

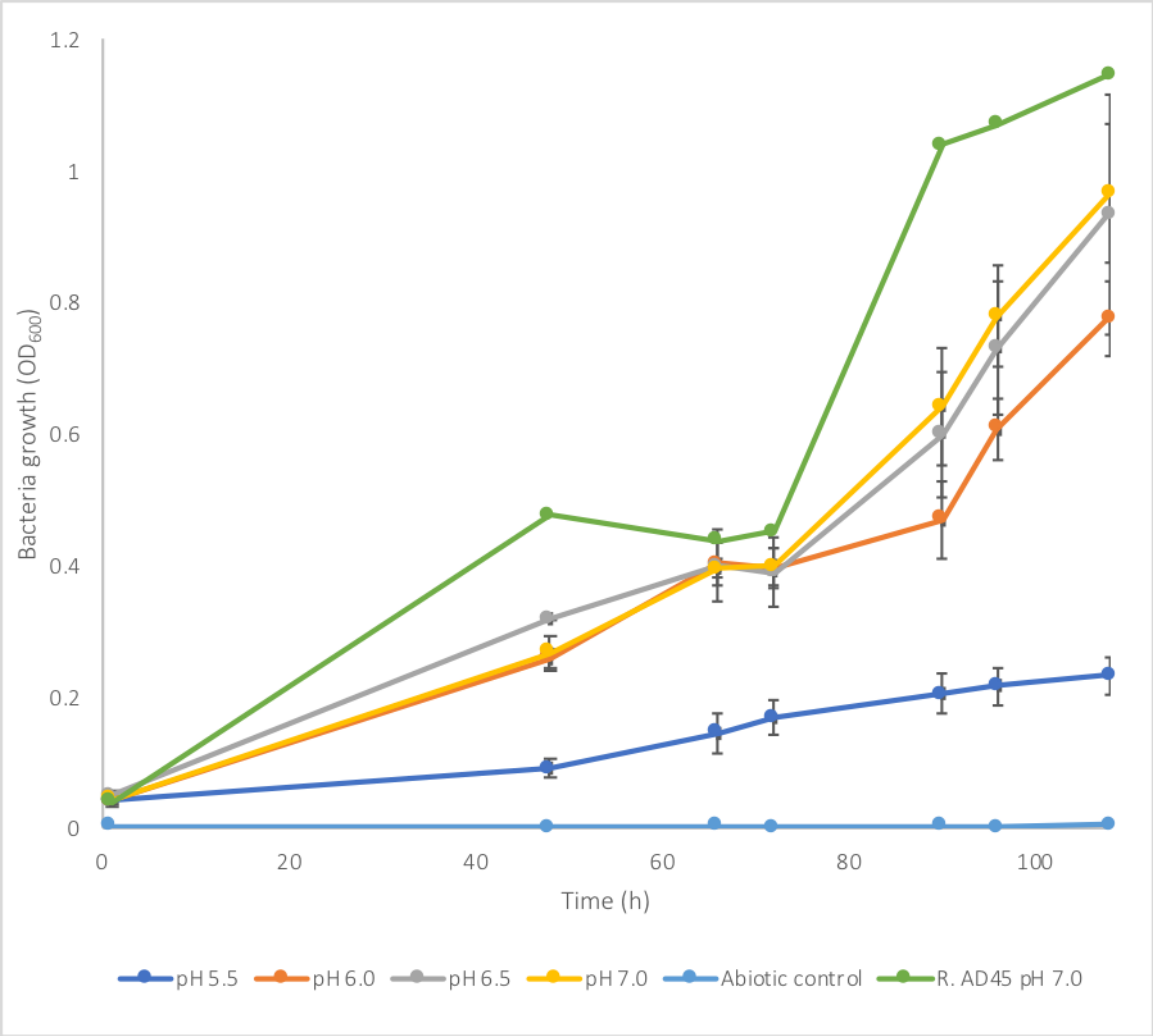



**Figure 3.11** Growth of isolate *Variovorax* strain sp. WS9 on minimal media with pH between 5.5 and 7. 1% isoprene was supplemented at the beginning of the assay and replenished after 72 hours. Media was supplemented with 1 µl per ml of MAMs vitamins. *Rhodococcus* sp. AD45 positive control was grown at pH 7.0. The experiment was performed in triplicate and error bars show the standard deviation.

**Figure 3.11** shows the optical density of *Variovorax* sp. strain WS9 grown at different pH. *Variovorax* sp. strain WS9 grew best at pH between 6.5 to 7.0, with addition of vitamins (1 µl per ml). The bacterium grew up to an $OD_{600}$ of 0.4 after consuming 1% isoprene (after 66 hours), and reached an $OD_{600}$ close to 1 when adding and additional 1% isoprene to the headspace. This information was essential for polypeptide analysis of the isoprene metabolic gene cluster.

### *3.2.6.2 Variovorax sp. strain WS9 consumes isoprene in the presence of succinate*

*Variovorax* sp. strain WS9 isoprene consumption was tested on different concentrations of isoprene (0.1% or 1% (v/v) in the headspace), or with 5 mM succinate, and with a combination of both carbon sources. In the initial experiment, there was a constant use of the isoprene even in the presence of succinate in the medium (data not shown). Start-up cultures were grown on isoprene or succinate as a carbon source and the consumption of isoprene was evaluated in triplicate (**Figure 3.12**).

The initial isoprene consumption was similar when *Variovorax* sp. strain WS9 was grown only with isoprene and with both carbon sources. This result suggested that both carbon sources were used simultaneously by *Variovorax* sp. strain WS9. Furthermore, the maximum $OD_{600}$ obtained for *Variovorax* sp. strain WS9 grown on both carbon sources (0.42-0.48) was lower than when the strain is grown only with succinate (0.57-0.58), this may be a result of the metabolic investment of *Variovorax* sp. strain WS9 when using isoprene as an additional carbon source. Finally, isoprene consumption rate at the end of the assay (40 h, isoprene below 0.2%) might have lowered due to the availability of isoprene and preferential use of succinate.

There were some limitations to the data produced in this procedure, growth curves for the use of isoprene and succinate as carbon sources need to be repeated with closer monitoring. During the experiments a faster-growing *Variovorax* isolate was recovered from a willow soil, *Variovorax* sp. strain WS11. This latter is currently being tested by Robin Dawson.

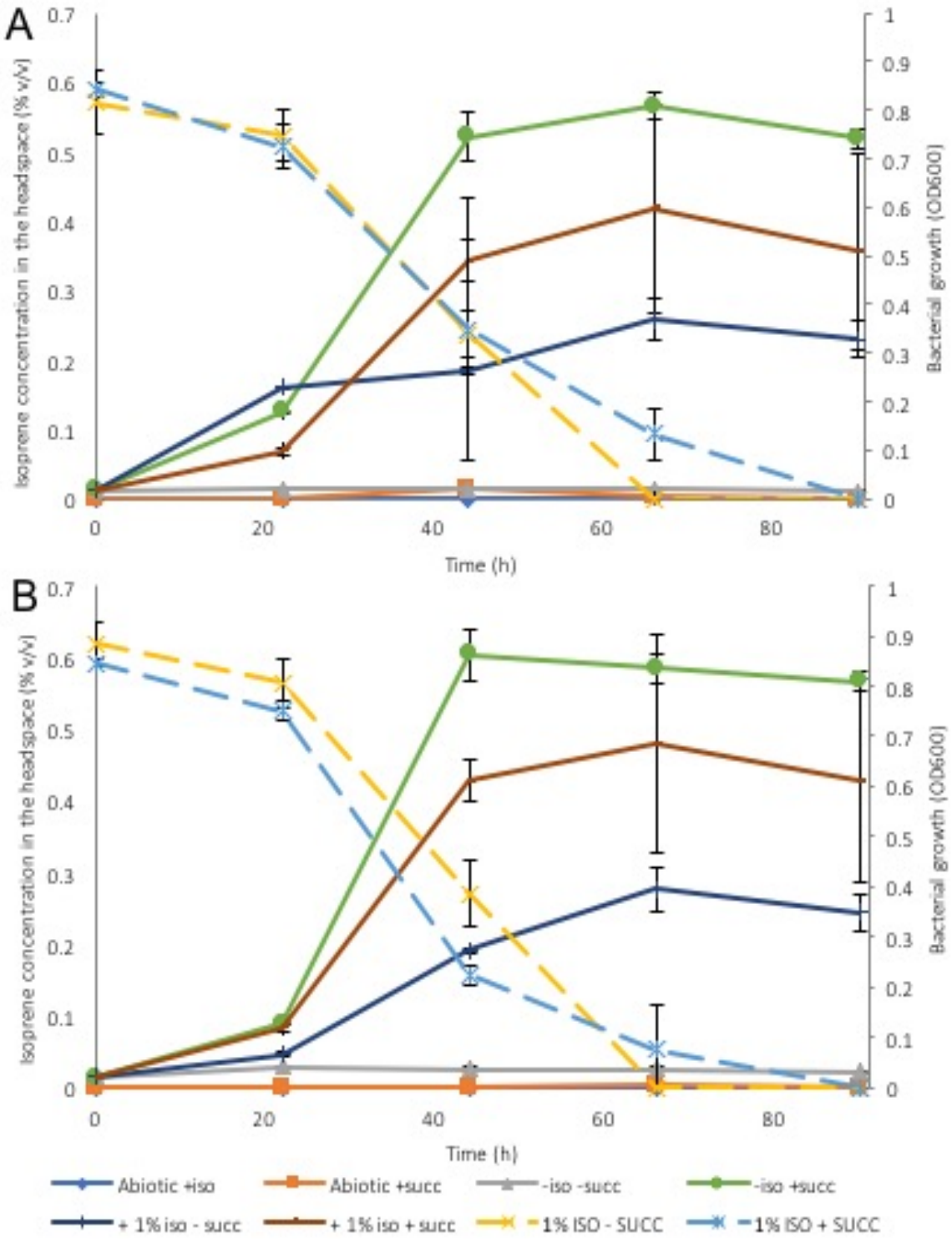



**Figure 3.12** Isoprene consumption by *Variovorax* sp. stain WS9 grown with 1% (v/v) isoprene, or succinate, and a combination of both carbon sources. Start-up cultures grown with A) isoprene or B) succinate. The experiment was performed in triplicate and error bars show the standard deviation.

### 3.2.7 Detection of isoprene metabolic genes using *isoA* gene PCR and SDS-PAGE

DNA was extracted from cells grown on isoprene using the FastSpin DNA soil kit (MP Biomedicals, **section 2.13**). An *isoA* gene PCR protocol was performed as described by Khawand *et al*., 2016 (see **section 2.15.2**) using isoAf - isoAr primer pair (**Table 2.1**) and the *Variovorax* sp. strain WS9 genomic DNA as a template.

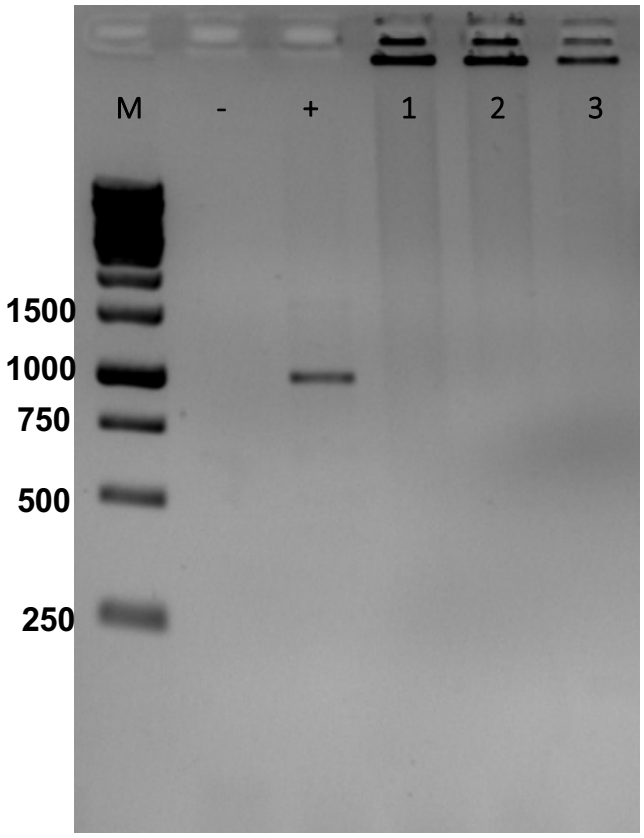



**Figure 3.13** *isoA* gene colony-PCR products for *Variovorax* sp. strain WS9. *Rhodococcus* sp. AD45 DNA was used as the positive control (+). 1, 2 and 3 are replicates. M: 1 kb DNA ladder (Fermentas).

Results for the *isoA* gene PCR showed no amplification of the gene. The *isoA* primers had shown variability of gene amplification with genomic DNA from isoprene-degrading isolates in the lab (data not shown), so another approach to observing the mechanisms used by the isolate was attempted. The presence and production of the isoprene gene cluster was studied with a protein profile of the isolate *Variovorax* sp. strain WS9 grown on isoprene, compared to the protein profiles of known isoprene degraders (*Rhodococcus* sp. AD45 and *Rhodococcus* sp. PD630, see **Figure 3.14**).

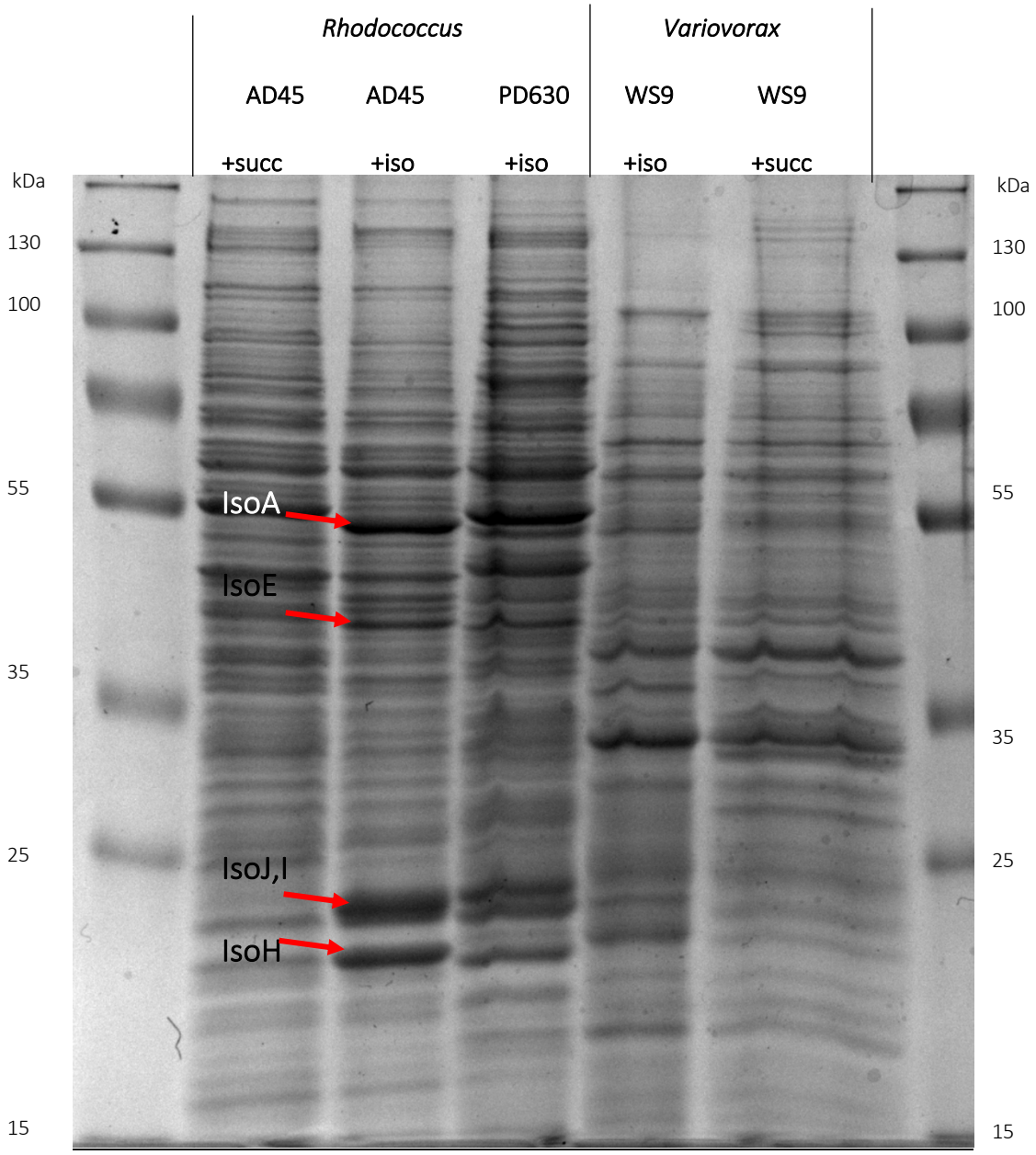



**Figure 3.14** Polypeptide profile comparison of *Variovorax* sp. strain WS9 isolate grown on isoprene compared to *Rhodococcus* sp. AD45 and *Rhodococcus* sp. PD630 (Crombie *et al*., 2015). Succinate-grown profiles for *Rhodococcus* sp. AD45 (Crombie *et al*., 2015) and *Variovorax* sp. strain WS9 protein profiles are also present as a control. Approx. 15 µg of protein added to each lane. M: Pageruler plus protein ladder

Polypeptide profiles for cell-free extracts of both *Rhodococcus* isolates grown on isoprene, show production of the polypeptides necessary for isoprene metabolism (IsoA, IsoJ, IsoI and IsoH). *Variovorax* sp. strain WS9 produced a different profile to *Rhodococcus* sp. AD45 and IsoA does not appear to be highly expressed in this isolate.

*isoA* gene PCR amplification and profiles did not show the expected results for isoprene metabolic gene products when comparing with the Gram-positive *Rhodococcus* isolates.

Genome sequencing and analysis confirmed the presence of the isoprene metabolic genes. Details on the genome and isoprene cluster of *Variovorax* sp. strain WS9 will be presented in **Chapter 6**.

### 3.2.8 The second Gram-negative terrestrial isoprene degrading bacterium identified as *Variovorax* sp. strain WS11

Targeted isolation was effective in obtaining the second isoprene-degrading isolate from the *Variovorax* genus. This isolate came from topsoil close to a willow tree in University of East Anglia (UEA-WS) from the enrichment experiment in June 2016. The experience of isolating and working with *Variovorax* sp. strain WS9 was essential for the conditions and quick selection/identification of the isolate. *Variovorax* sp. strain WS11 is also a transparent, slow growing bacterium that produces a yellow pigment on the agar after a few days. Microscope images in **Figure 3.15** show the isolate grows as a compact rod when grown on succinate and stretches into a longer rod on isoprene.

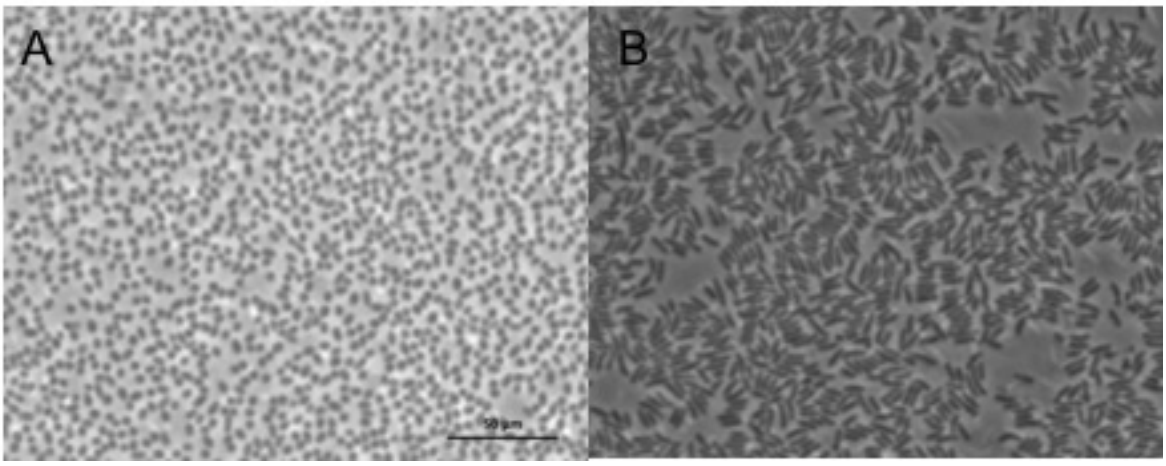



**Figure 3.15** *Variovorax* sp. strain WS11 bacterium phase contrast microscope images grown on a) succinate and b) isoprene (1000X magnification).

#### *3.2.8.1 Growth of isolate Variovorax sp. strain WS11 on isoprene*

Isolate *Variovorax* sp. strain WS11 came from a different soil and a different location to *Variovorax* sp. strain WS9, but near the same type of tree. The production of a yellow halo

around the transparent colonies suggested the two *Variovorax* isolates had phenotypic and genotypic differences. A growth curve of *Variovorax* sp. strain WS11 on isoprene as the sole carbon and energy source is shown in **Figure 3.16**. Both isolates took approximately the same time to consume 1% (v/v) isoprene and reach a similar maximum $OD_{600}$ of approximately 0.3**.**

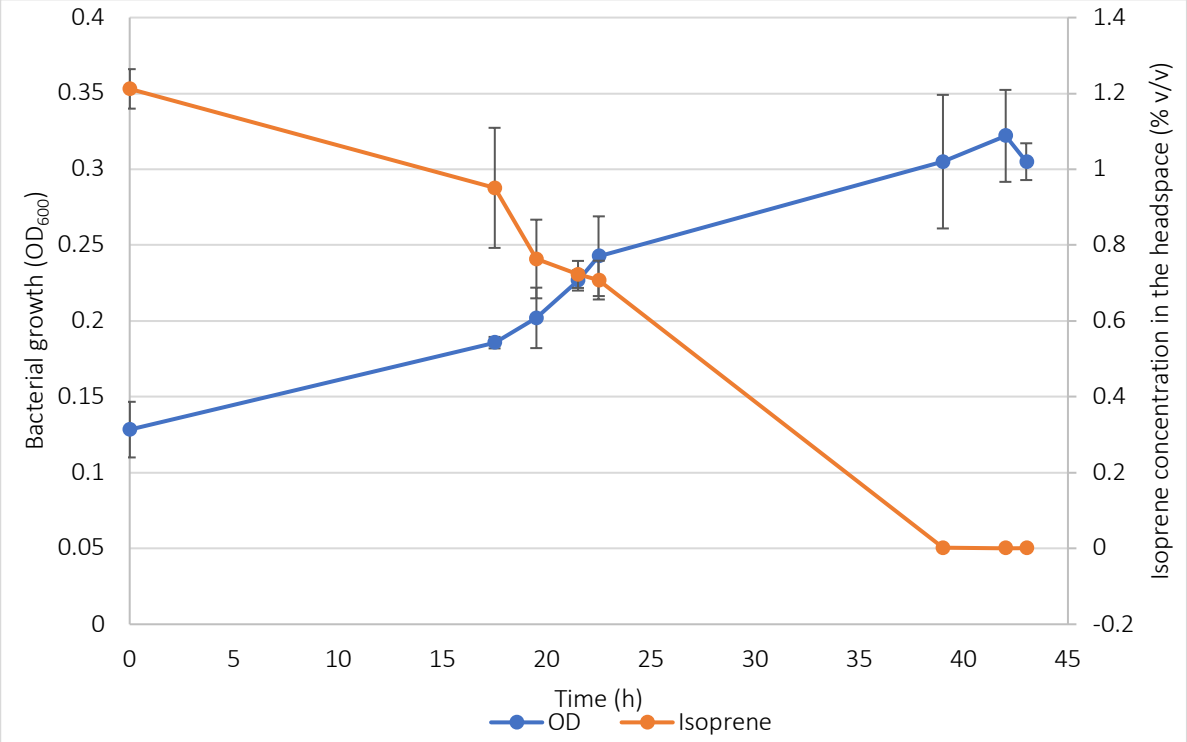



**Figure 3.16** Isoprene-grown *Variovorax* sp. strain WS11 cells grown on 1% (v/v) isoprene. Error bars show the standard variation in the three replicas. Negative controls showed no consumption of the supplemented isoprene, or growth (data not shown).

Even though the isolates seem to reach the same optical density, growth on rich media was quicker for isolate *Variovorax* sp. strain WS11. Colonies appear after 3 to 4 days of incubation compared to 7 to 10 days for *Variovorax* sp. strain WS9. Genome sequencing results and analysis of the isoprene cluster of *Variovorax* sp. strain WS11 will be presented in **Chapter 6.** Further analysis of this isoprene-degrading bacterium is being done by Rob Dawson (PhD student at the Murrell Lab).

### 3.2.9 Gram-positive terrestrial isoprene degrading bacterium identified as *Nocardioides* sp. strain WS12

Enrichment assays of the UEA-WS, aimed at targeted isolation of other *Variovorax* isolates, permitted the isolation of a non-*Rhodococcus* actinobacteria identified as *Nocardoides* sp. strain WS12. Several attempts for a complete growth curve of the isolate were performed (data not shown) and this is now being analysed by Lisa Gibson (PhD student at the Murrell Lab). Continuous monitoring and replenishment of isoprene might be the best method to study how this bacteria metabolises isoprene.

### 3.2.10 Phylogeny of new *Rhodococcus* and non-*Rhodococcus* isolates

Along with *Variovorax* sp. strain WS9, *Variovorax* sp. strain WS11 and *Nocardioides* sp. strain WS12, shotgun genome sequencing was performed for other interesting *Rhodococcus* isolates (shown in **Table 3.6**). Isolates from the *Rhodococcus* genus with a high percentage similarity to the model organism *Rhodococcus* sp. AD45 (>98%), were not studied further. 16S rRNA gene sequence alignment and neighbour-joining trees show the taxonomic organization of the willow soil isolates to closely related organisms from the *Rhodococcus* genus (**Figure 3.17**), *Variovorax* genus (**Figure 3.18**) and *Nocardioides* genus (**Figure 3.19**)

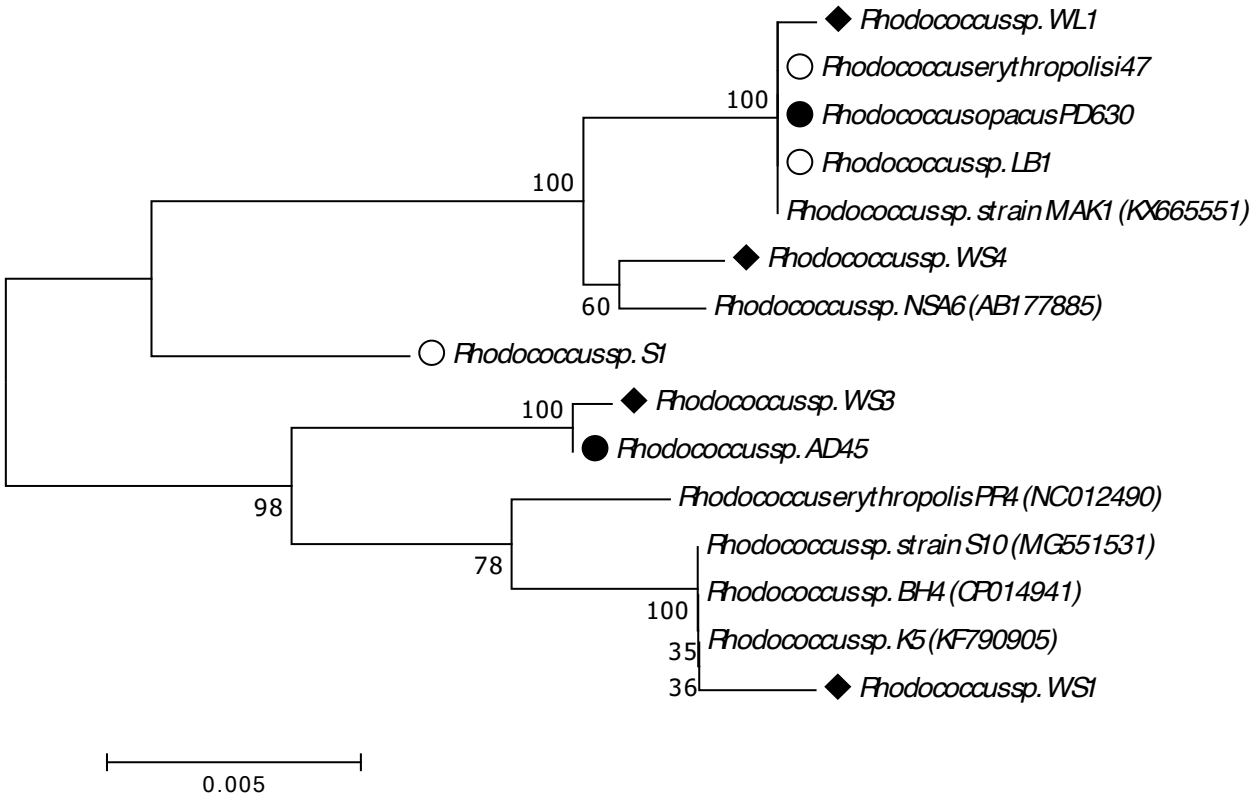



**Figure 3.17** Neighbour-joining tree with the phylogeny of isolates from the *Rhodococcus* genus (bootstrap of 1000 replicates). The tree includes 15 nucleotide sequences and 1336 positions in the final data set (gaps and missing data were eliminated). White circles, laboratory isolates; black circles, reference microorganisms; black diamonds, are the isolates obtained in this study.

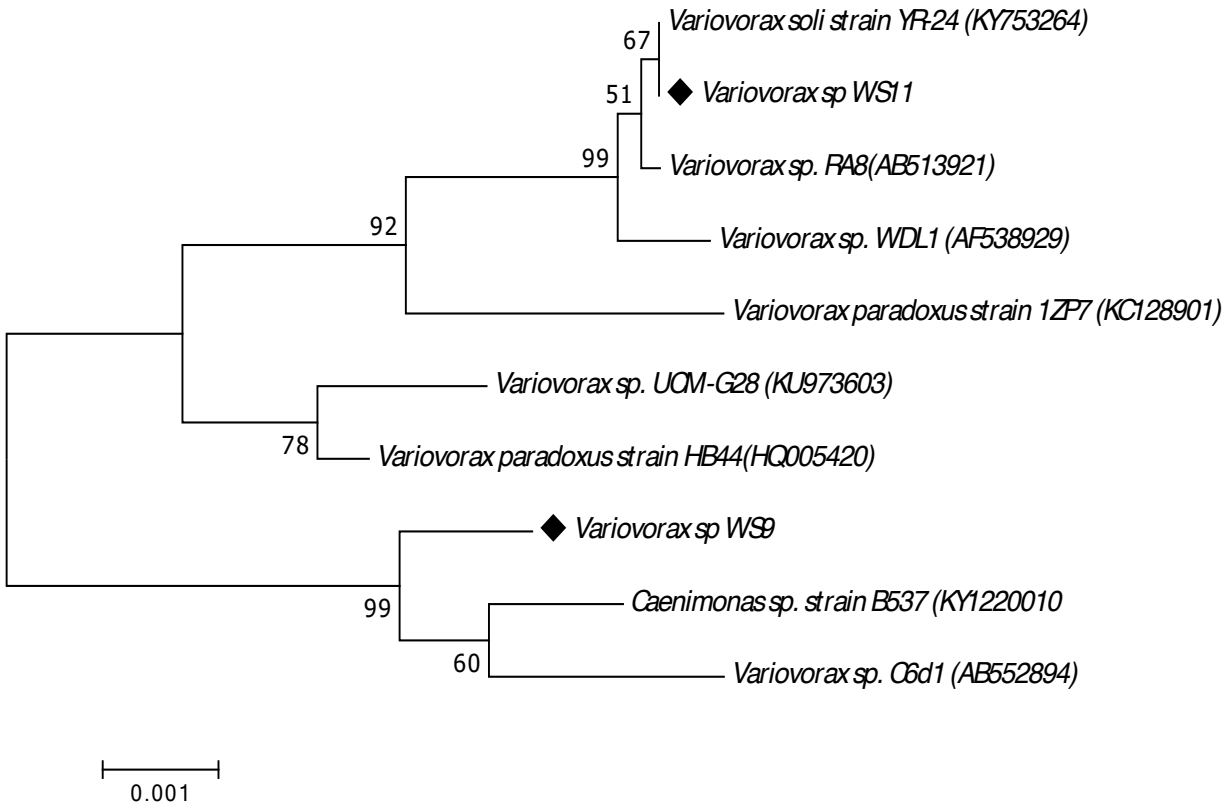



**Figure 3.18** Neighbour-joining tree with the phylogeny of isolates from the *Variovorax* genus (bootstrap of 1000 replicates). The tree includes 10 nucleotide sequences and 1370 positions in the final data set (gaps and missing data were eliminated). Black diamonds, isolates obtained in this study.

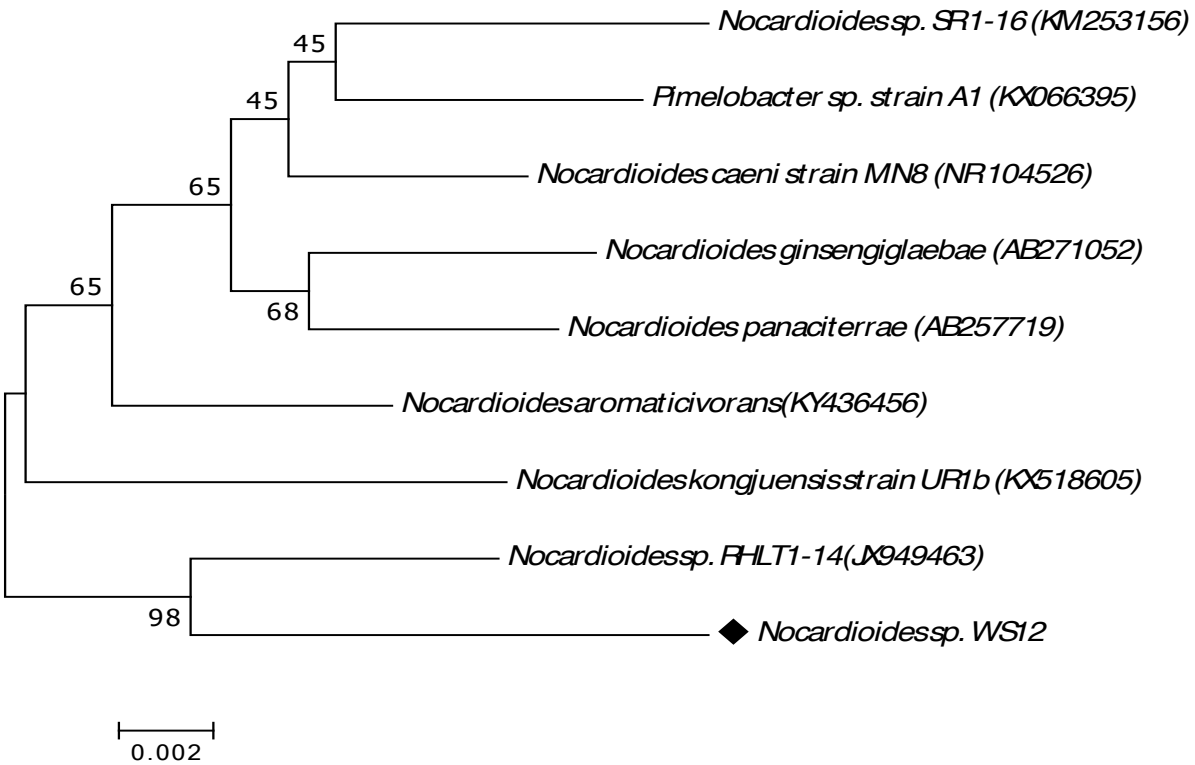



**Figure 3.19** Neighbour-joining tree with the phylogeny of isolates from the *Nocardioides* genus (bootstrap of 1000 replicates). The tree includes 9 nucleotide sequences and 1330positions in the final data set (gaps and missing data were eliminated). Black diamonds, isolate obtained in this study.

## 3.3 Discussion

Isoprene-degrading bacteria are present on leaves and in the soil under the canopy of isoprene-emitting trees. Bacterial cells are numerous on leaves, estimated to be between $10^6$ and $10^7$ cells $cm^2$ (Lindow & Brandl, 2003). The abundance and diversity of these isoprene-degraders on the phyllosphere may be directly related to the emission potential of a specific tree species, among other factors.

Cultivation-dependent approaches allowed the isolation of a larger diversity of isoprene degraders at low enrichment concentrations i.e. 13-25 ppmv. These conditions may have aided slow-growing autochthonous bacteria over other highly versatile and metabolically flexible organisms like *Rhodococcus* (Kataoka *et al.*, 1996). Although isolates of the genus *Variovorax* grew slowly on isoprene and were not inhibited at concentrations up to 1% (v/v), the isolation of the strains was difficult when the enrichment procedure from environmental samples offered high concentrations of isoprene. This may be because the high concentrations are advantageous for other isoprene degraders, such as species from the genus *Rhodococcus*.

Selection of isoprene degraders by sub-culturing at low concentrations also permitted the effective isolation of isoprene-degraders from oil palm leaves. Sequential enrichment and sub-culturing was effective at obtaining the important isolates. The gradual increase in abundance of only isoprene consumers in the inoculum was reflected in quicker uptake and depletion of isoprene in each transfer. Serial sub-cultivation of organisms has shown an increase in an organisms' production yield (Zou *et al.*, 2013), and in this case the depletion of isoprene.

Although cross-feeding was expected during the incubation process, replenishment with new isoprene every few hours might have kept isoprene-degraders in high numbers. Semi-continuous cultures would avoid frequent spiking of samples every couple of hours. The addition of fresh medium and elimination of waste to a bioreactor would mimic the sub-culturing strategy. Bioreactors can also be used to study the influence of certain factors on

microbiome dynamics and with pure bacteria (Ginige *et al.*, 2004; Madsen, 2006; Singleton *et al.*, 2005).

Bacteria living in the phyllosphere exhibit predictable biogeographic patterns (Redford *et al.,* 2010). The number of microbial cells present on leaves varies due to the environment (rainfall, location, ecosystem) and the specific phenotypic leaf characteristics of the plant species (surface area, texture, 3D structure) (Lindow & Brandl, 2003). Results suggested variation in the abundance of isoprene-degraders between leaves from the same tree, reflected as different isoprene consumption rates between replicates. It has been shown that inter-species variability may exceed intra-species variability, but the bacterial community composition driven by differences in leaf characteristics is still undetermined (Redford *et al.*, 2010). Further studies on the intra-species variation in the phyllosphere may give insights into abundance of isoprene-degraders within an individual habitat or a specific trees phyllosphere.

Although bacterial communities vary between and across individual of the species, the community structure varies more across tree species (Redford *et al.*, 2010). Inter-species microbiome variation was demonstrated with the divergent diversity of isoprene-degraders isolated from willow soil (*Rhodococcus, Nocardioides* and *Variovorax*) and oil palm trees (*Rhodococcus*, *Gordonia*, *Pseudomonas* and *Sphingopyxis*). Trees with higher isoprene emission potential may host different bacteria compared to trees with low isoprene emission potential. In either case, *Rhodococcus* isolates are present confirming it as a cosmopolitan isoprene-degrader (El Khawand *et al.*, 2016).

Recently, Khawand and colleagues observed a 6.5% relative abundance enrichment of members of the Comamonadaceae family in the heavy fractions of $^{13}$C-incubations assuming direct or indirect labelling by members of this family (El Khawand *et al.*, 2016). This study reports the first two Gram-negative isolates from the Comamonadaceae that can directly use isoprene as a sole carbon and energy source. These *Variovorax* isolates are beta-proteobacteria that have been commonly isolated from soil and river water (Willems *et al.*, 1991). Members of the genus *Variovorax* have been isolated from a diverse range of ecosystems (polluted and non-polluted, Wiliems *et al.,* 2005) as degraders of xenobiotic compounds and degraders of plant protection products (Bers *et al*., 2011). Terrestrial

isolates include *Variovorax soli* sp. nov. from a greenhouse, *Variovorax dokdonensis* sp. nov. (Yoon, 2006), *Variovorax boronicumulans* sp. nov. is a boron accumulating bacterium (Miwa *et al.*, 2008), and Artic glacier strains (Ciok *et al.*, 2016). Strains have also been isolated from other organisms such as *Variovorax* sp. strain PAMC28711 isolated from an Antarctic lichen (Han *et al.*, 2016) and *Variovorax guangxiensis* sp. nov. isolated from the banana rhizosphere (Gao *et al.*, 2015).

Finally, the *Variovorax* sp. strain WS9 isolate was grown on succinate and isoprene as carbon sources simultaneously. Isoprene depletion continued in the presence of succinate, showing no preference for either of the carbon sources. Moreover, the protein profile for isolate *Variovorax* sp. strain WS9 shows the IsoA protein does not appear to be highly produced during isoprene metabolism. These results suggest the isoprene machinery was expressed constitutively in *Variovorax* sp. strain WS9. The low production of IsoA may also explain slow consumption of isoprene and a reason why this organism is a slow-growing isoprene-degrader. Further analysis of the expression of the isoprene metabolic enzymes is necessary to conclude on how the organism is expressing isoprene metabolic genes.

# Chapter 4 Identification of active isoprene-degrading bacteria in willow soil using DNA-stable isotope probing and SIP-derived metagenomics

## 4.1 Introduction

The environment consists of complex and dynamic ecological systems. Microbiome composition, richness, and diversity are affected by changes in biotic and abiotic factors (Han *et al.*, 2007; Lakshmanan *et al.*, 2014; Widdows & Brinsley, 2002). Understanding how members within microbial communities relate to the function is a challenge in microbial ecology (Widder *et al.*, 2016). Cultivation-dependent methods are important in the genotypic and phenotypic characterisation of microbes from environmental communities, but not all microbes are easy to cultivate. To describe the diversity of microbes in a community and to accurately know the role/function of key microbes in the environment focused cultivation-independent techniques are required (Crombie *et al.*, 2018; Vieites *et al.*, 2009). Several examples of the use of cultivation-independent techniques to study the metabolism of certain organisms are found in the literature, including the study of diverse methylotrophic communities by Eyice & Schäfer, 2016; marine methylotroph studies by Grob *et al.*, 2015; the use of polycyclic aromatic hydrocarbons (PAH) by Gutierrez *et al.*, 2015; discovery of new hydrocarbon monooxygenases using cultivation-independent techniques by Li *et al.*, 2014 and marine methanol and methylamine metabolism using SIP by Neufeld *et al.*, 2007.

In this project, microbes isolated from soils and leaves from isoprene emitting trees have been grown on isoprene as their sole carbon and energy source (**Chapter 3**). The concentration of isoprene during enrichment was lowered to approximately 25 ppmv, to permit the growth of microorganisms that do not tolerate high isoprene concentrations to grow and to provide conditions closer to the natural environment. The use of DNA-stable

isotope probing can help to determine the abundance and diversity of key microbes in soils that use isoprene as a carbon source (El Khawand *et al.*, 2016).

Khawand *et al.* in 2016 demonstrated in $^{13}$C-labelled isoprene enrichments 0.5% (v/v) labelled that *Rhodococcus* was a key isoprene degrader in soil. In their study, 88% of the heavy fraction of $^{13}$C incubations corresponded to *Rhodococcus wratislaviensis*, *R. koreensis* and *R. globerulus.* Their analysis of $^{13}$C DNA from DNA-SIP enrichments also showed 6.5 % relative abundance of members of the Comamonadaceae family (El Khawand *et al.*, 2016). This chapter describes the effects of an isoprene enrichment using lower concentrations of isoprene on the microbial diversity and abundance of isoprene degraders for Colney-WS samples. DNA was extracted and fractionated, followed by community profile analysis and 16S rRNA gene amplicon sequencing of different time-points during enrichment incubations. Results were later compared to the cultivation-dependent results for the enrichment (**Chapter 3**). Finally, SIP-derived metagenomes from heavy DNA fractions and native soil was used to study the diversity of *isoA* encoding the α-subunit of the isoprene monooxygenase.

*General objective:*

To evaluate the diversity and abundance of the key isoprene-degrading bacteria in willow soil using DNA-SIP (16S rRNA gene diversity and full metagenome diversity).

*Specific objectives:*

To determine isoprene consumption rates of $^{12}$C-isoprene and $^{13}$C-labelled isoprene enrichments with Colney-willow tree soil.

To define the best conditions (time course) for effective incorporation of the $^{13}$C-label from low concentrations of isoprene in willow soil DNA-SIP experiments.

To verify incorporation of $^{13}$C-label and enrichment of isoprene-degrading bacteria using 16S rRNA gene DGGE community profiling of $^{12}$C- and $^{13}$C- DNA isolated during SIP experiments.

To determine the diversity and abundance of key isoprene-degraders in willow soil enrichments and changes over time in bacterial community composition using 16S rRNA gene amplicon sequencing.

To investigate the isoprene metabolic potential of key active microbes in Colney-WS by extracting sequences of the gene encoding the $\alpha$-subunit of the isoprene monooxygenase (*isoA*) from metagenome sequence information.

## 4.2 Results

## 4.3 Stable isotope probing (SIP) enrichments with willow tree topsoil

Soil samples adjacent to an isoprene-emitting willow tree (Colney-WS) were collected on July 1, 2015 (**Figure 4.1,** see collection details in **section 2.7**).


**Figure 4.1** Willow tree from Colney sport fields during the A) winter and B) summer months. C) Example of one sampling point in early April 2015.

Previous enrichment experiments, with willow soil (WS) collected from the same site, showed consistent consumption of isoprene (see **Chapter 3**). The results suggested the

presence of isoprene-degrading bacteria in the topsoil and resulted in the isolation of three isoprene-degrading bacteria from the genus *Rhodococcus* (**section 3.2.5**). In new experiments, fresh samples were collected and used in either $^{12}$C-isoprene or $^{13}$C-labelled isoprene enrichments.

### 4.3.1 Set up of DNA-SIP experiments

Microcosms using the Colney-WS were set-up with 5 grams of soil and 40 ml of sterile water in sterile 2 litre flasks (without any addition of nutrients). The assays were in triplicate, with either $^{12}$C-isoprene or fully-labelled $^{13}$C-isoprene (Dr Gregg Whited; DuPont Industrial Biosciences, USA refer to section 2.10) at a concentration of approximately 25 ppmv. Microcosms were incubated in the dark at 25 °C and shaking at 150 rpm (**Figure 4.2**). Leakage and abiotic depletion of isoprene were not observed in control assays.

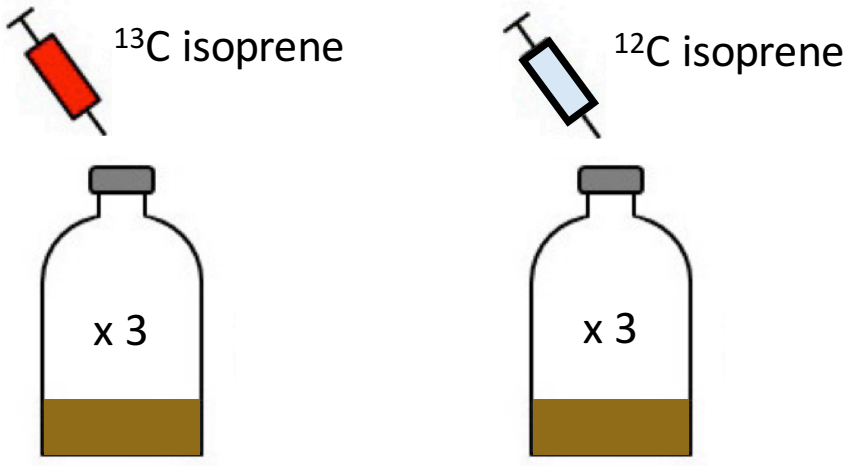



**Figure 4.2** Colney-WS SIP enrichment set-up. $^{12}$C isoprene and $^{13}$C-labelled isoprene incubations were done in triplicate.

Previous Colney-WS enrichment results were used as a guideline for the length of time needed to consume sufficient isoprene. The initial consumption rate of isoprene was established for both treatments using gas chromatography (see **Table 4.1**). The values for SIP enrichments were comparable to previous data (0.027 μmol $h^{-1}$ $g^{-1}$ for Colney-WS in July 2015, see **Table 3.4**).

**Table 4-1** Estimated isoprene consumption rate (%) for Colney-WS soil samples during the SIP experiment. The first spike of 25 ppmv isoprene was consumed in approximately 50 hours. The average standard deviation of the samples was 0.000168 μmol $h^{-1}$ $g^{-1}$ for $^{12}C$-isoprene and 0.000155 μmol $h^{-1}$ $g^{-1}$ for $^{13}C$-isoprene.

| Sampling site | Isoprene type | Consumption rate (μmol $h^{-1}$ $g^{-1}$) |
|---|---|---|
| Colney-WS | $^{12}C$ | 0.0231 |
| Colney-WS | $^{13}C$ | 0.0236 |

During the course of SIP incubations, isoprene was added to the microcosms when the concentration fell below 10 ppmv, see **Figure 4.2**. Keeping the concentration of isoprene around 25 ppmv maintained the appropriate isoprene consumption rate for degraders present in the sample. Samples were sacrificed at time points during the experiment in order provide a range of $^{13}C$-labelled carbon incorporation from isoprene into the biomass of key isoprene-degraders (Neufeld *et al*., 2007). **Table 4.2** indicates the specific incubation times at which replicate samples were sacrificed during SIP enrichments, along with the estimated amounts of $^{13}C$-Carbon incorporated into biomass. Since incorporation was sufficient after $T_1$ and $T_2$, the $T_3$ samples were not examined further.

**Table 4-2** $^{13}C$- incorporation with Colney-WS SIP incubations

| Time course | $^{13}C$ incorporation μmol $g^{-1}$ (calculated)* | Time (h) |
|---|---|---|
| Native soil | 0 | 0 |
| Time point 1 ($T_1$) | 25 (22.3) | 138 |
| Time point 2 ($T_2$) | 50 (47.26) | 172.5 |
| Time point 3 ($T_3$) | 75 (72.21) | 187 |

*T1, T2 and T3 correspond to 0.87 ml, 1.61 ml and 2.18 ml of isoprene gas spiked into the headspace, respectively. $^{13}C$arbon incorporation values were predetermined before the assay. The actual incorporation at each time point during the assay, in parenthesis, was calculated assuming 34% of $^{13}C$arbon consumed was assimilated into biomass (El Khawand *et al*., 2016). The soil sacrificed at each time-point was also included in the calculations.

### 4.3.2 Monitoring isoprene uptake throughout the Colney-WS SIP incubations

Isoprene depletion was monitored as in to previous Colney-WS enrichment experiments (**section 3.2.4.1**). Similar isoprene consumption in all replicates for both treatments permitted the same chronological spiking of fresh isoprene throughout the incubations (see **Figure 4.3**). The total incubation time for both treatments in the experiment was 187 h (approx. 75 μmol of $^{13}C$ incorporated into 1 g of biomass/soil).

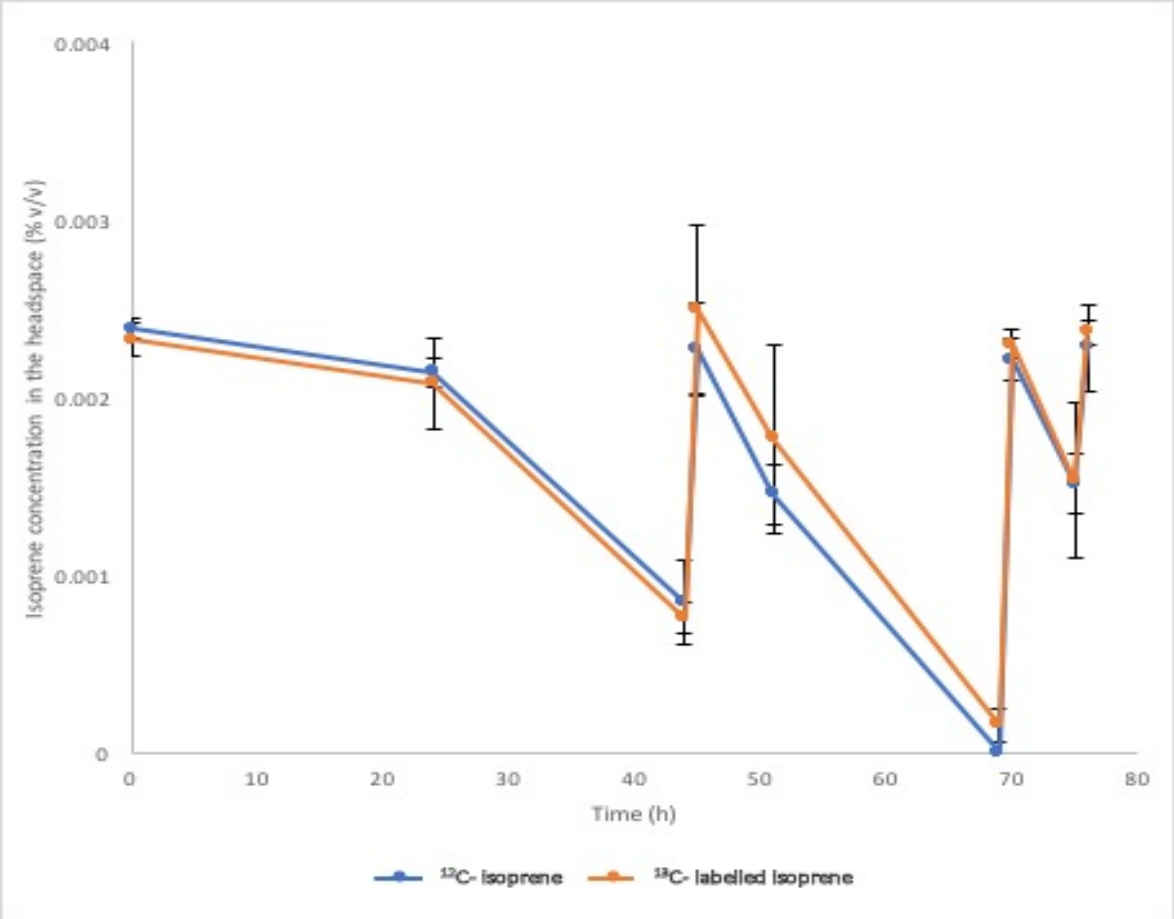



**Figure 4.3** Isoprene consumption during Colney-WS SIP incubations over the first 76 h. The total assay took 180 h. Both $^{12}C$ isoprene and $^{13}C$ labelled isoprene incubations were done in triplicate and error bars show the standard deviation.

Replenishing isoprene encouraged the growth of isoprene-consumers, resulting in higher consumption rates as the SIP-incubations progressed (data not shown). Spiking the samples frequently added small variation in the amount of isoprene supplemented to each sample (5% - 10%). Isoprene consumption rates were higher during the day, when depletion was

monitored every few hours and restocked frequently. Lower consumption rates were only observed over-night, when samples were not continuously supplemented with isoprene. For this reason, during the last 20 hours of SIP-incubations, fresh spikes of isoprene were added more frequently between the last two time points.

### 4.3.3 $^{12}C$- and $^{13}C$- DNA isolation from Colney-WS DNA-SIP experiments

**Figures 4.4** and **4.5** show the results of plotting the DNA concentration of each fraction against the density, with DNA obtained after 6 and 7 days of enrichment, respectively.

**Figures 4.4B** and **4.5B** indicate successful incorporation of the labelled $^{13}$Carbon after 6 and 7 days of enrichment. To estimate how much DNA was recovered after centrifugation and fractionation for each replicate, the total DNA recovered at each time point was calculated by adding the DNA from all fractions in each sample. 22- 62% of the DNA added to the tubes was recovered in the fractions.

The percentage of light and heavy DNA from $^{12}C$ and $^{13}C$-labelled incubations was calculated for samples taken at 6 days (25 μmol $^{13}C\ g^{-1}$ incorporated) and at 7 days (50 μmol $^{13}C\ g^{-1}$ incorporated). For the light and heavy DNA three fractions were pooled together. Fractions 4 to 6 were used for the heavy DNA and fractions 8 to 10 for the light DNA. The data recovered for both light and heavy DNA are shown in **Table 4.3**. **Figure 4.6** shows heavy DNA isolated at both time-points (6 and 7 days) with both $^{12}C$- and $^{13}C$- incubations.

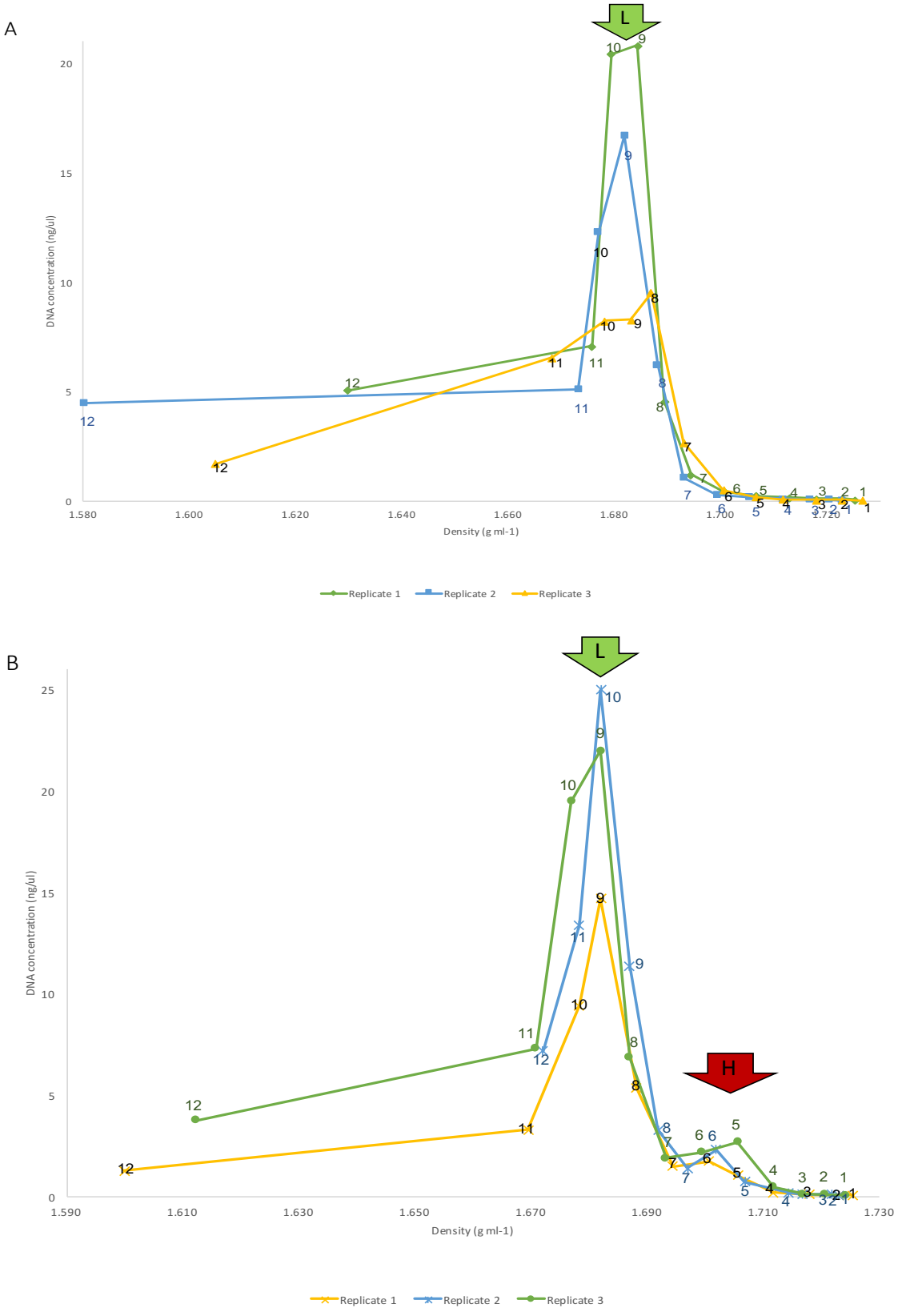



**Figure 4.4** Fractionation curves (density vs. DNA concentration) for DNA obtained with SIP enrichments with Colney-WS after 6 days of enrichment ($T_1$: 138 h) with A) $^{12}$C-isoprene and B) $^{13}$C-labelled isoprene. DNA fraction is indicated by the number on the curve. H: heavy DNA; L: light DNA.

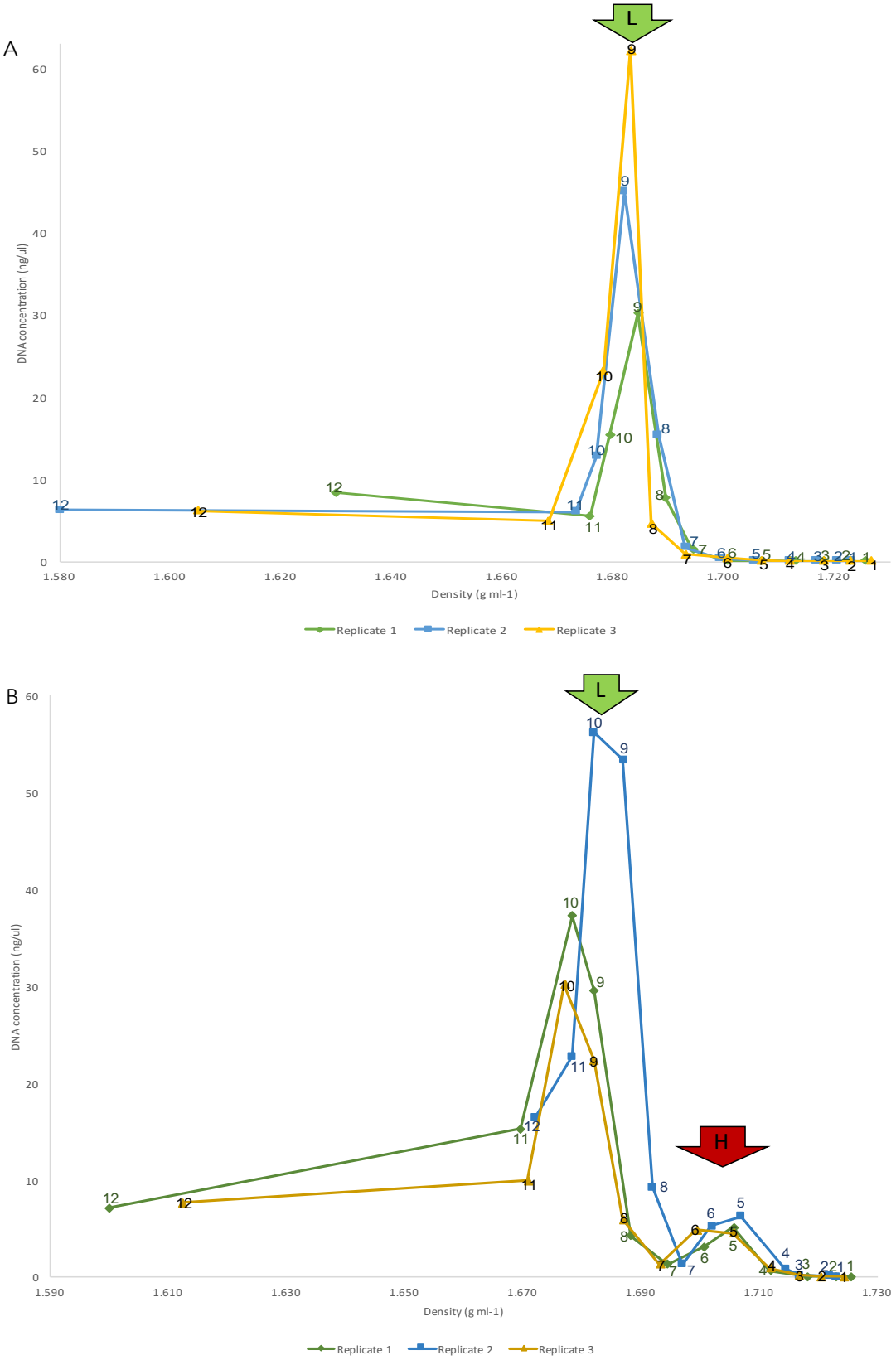



**Figure 4.5** Fractionation curves (density vs. DNA concentration) for Colney-WS DNA obtained with SIP enrichments after 7 days of enrichment ($T_2$: 172.5 h) with A) $^{12}$C-isoprene and B) $^{13}$C-labelled isoprene. DNA fraction is indicated by the number on the curve. H: heavy DNA; L: light DNA.

**Table 4-3** Percentage of DNA contained in heavy and light fractions after 6 and 7 days ($T_1$ and $T_2$, respectively) of enrichment with isoprene of Colney-WS.

| | % heavy DNA | | % light DNA | |
|---|---|---|---|---|
| **Sample** | **$T_1$ (day 6)** | **$T_2$ (day 7)** | **$T_1$ (day 6)** | **$T_2$ (day 7)** |
| $^{12}$C-Replicate 1 | 1.2 | 0.2 | 68.3 | 66.1 |
| $^{12}$C-Replicate 2 | 1.1 | 0.5 | 62.0 | 65.6 |
| $^{12}$C-Replicate 3 | 1.9 | 0.5 | 43.5 | 83.1 |
| $^{13}$C-Replicate 1 | 7.3 | 7.9 | 62.1 | 64.4 |
| $^{13}$C-Replicate 2 | 4.7 | 6.7 | 55.8 | 63.6 |
| $^{13}$C-Replicate 3 | 7.3 | 10.7 | 61.9 | 59.8 |

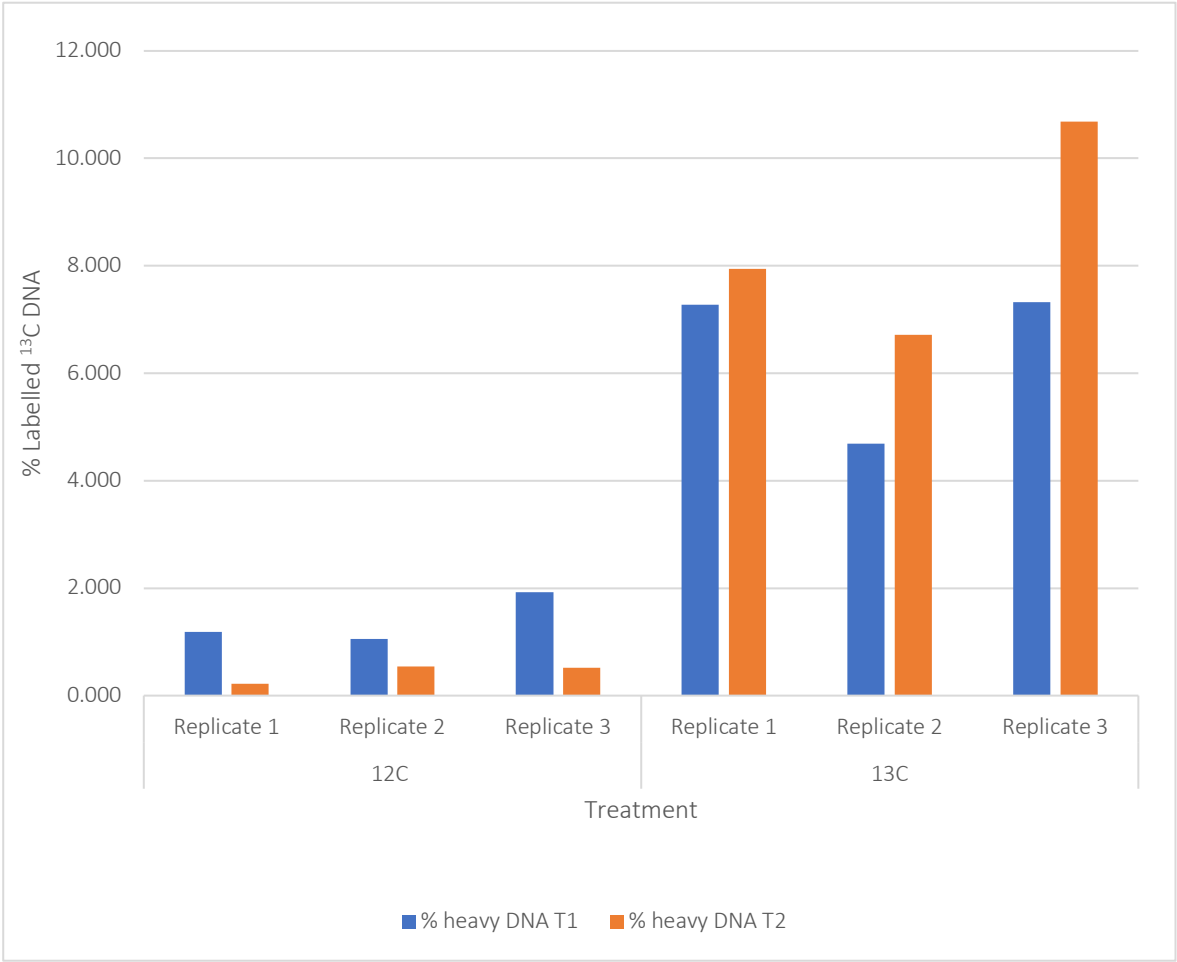



**Figure 4.6** Percentage of DNA contained in heavy-fractions from $^{12}$C-isoprene and $^{13}$C-labelled isoprene enrichments of Colney-WS after 6 ($T_1$) and 7 ($T_2$) days of enrichment.

As expected, the percentage of total heavy DNA recovered from $^{13}C$-labelled incubations increased with continuous enrichment. Incorporation of the $^{13}C$ –labelled into DNA was therefore successful in the Colney-WS microcosms.

### 4.3.4 Analysis of changes in the willow soil (WS) bacterial community profile after isoprene enrichment using Denaturant Gradient Gel Electrophoresis (DGGE)

The enriched and fractionated DNA was diluted to a concentration of 3 ng $\mu l^{-1}$ for all replicates. The DNA was then used as the template for the targeted amplification of the 16S rRNA gene using 341F-GC and 518R primer pair (see **Table 2.1**). The PCR products for each fraction were run on a 1% agarose gel to estimate the amount of DNA that would be added to the DGGE gels. **Figure 4.7** shows the PCR products obtained with DNA from incubations with $^{12}C$- and $^{13}C$- isoprene.

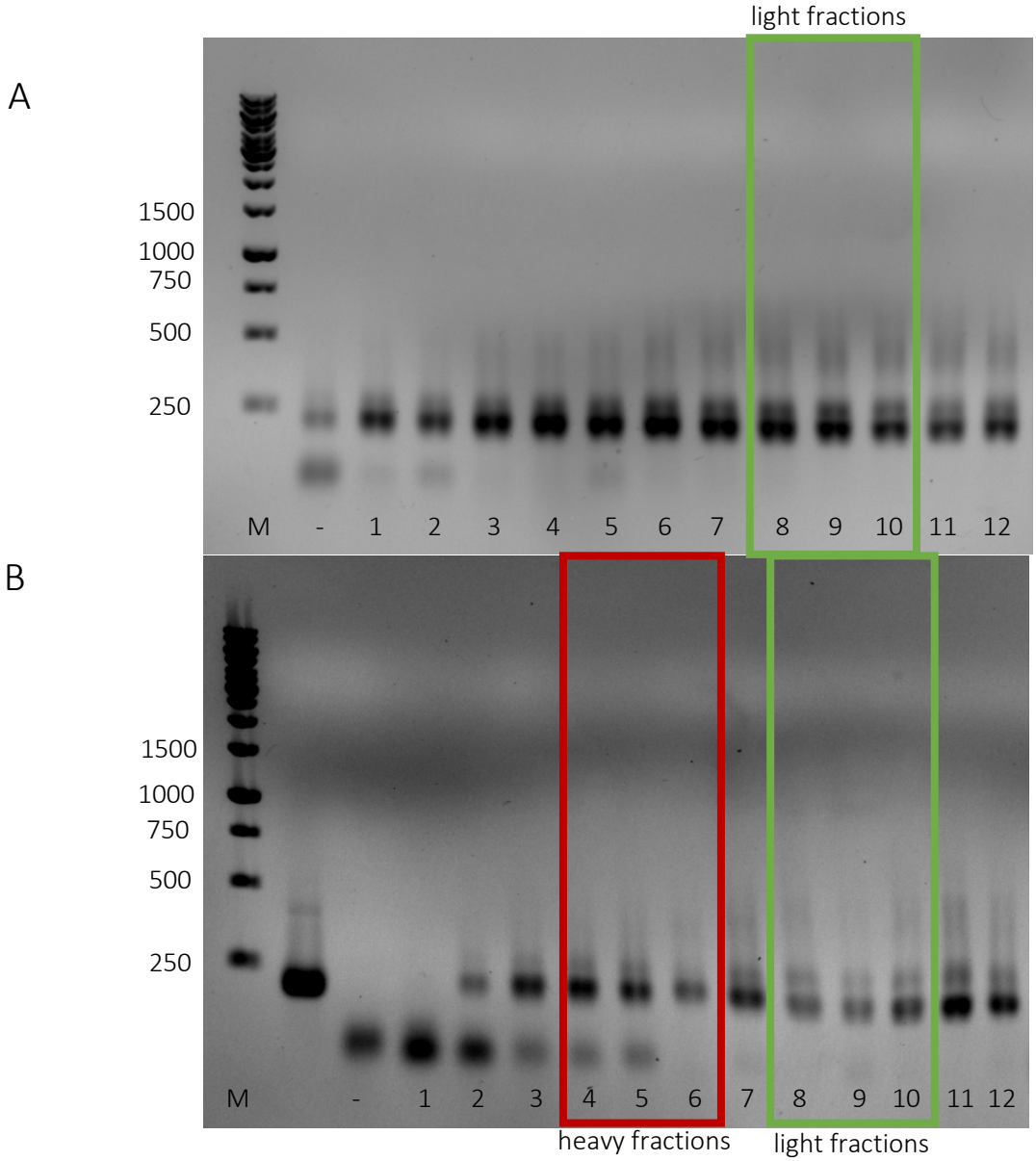



**Figure 4.7** 16S rRNA gene PCR products for DGGE analysis. 177bp product amplicons of were obtained by using 341F-GC and 518R primer pair on a 1% agarose gel for 6 days of enrichment A) $^{12}C$ enrichment replicate 1 and B) $^{13}C$ labelled enrichment replicate 1. Gel sample labels for each fraction on the bottom. M: marker

All PCR products for each replicate were run on an 8% polyacrylamide gel with a linear gradient of 30 to 70% denaturant concentration for 16 hours at 75 V in a DCode™ Universal Mutation Detection System by Bio-Rad. **Figure 4.8** shows $^{12}C$ and $^{13}C$-labelled fractions for replicate 2 after 6 days of incubation. **Figure 4.9** shows the results for replicate 1 after 7 days of enrichment.

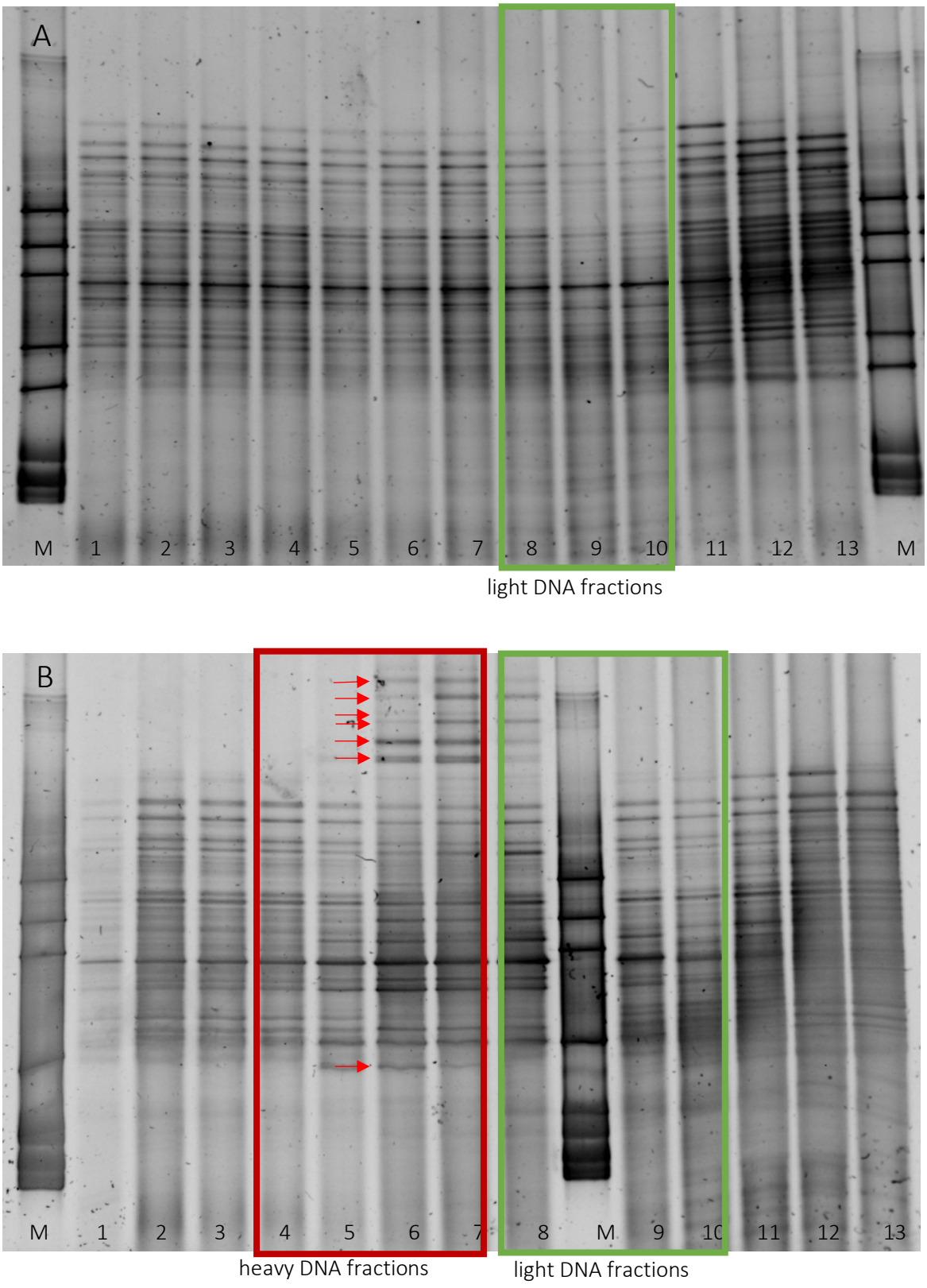



**Figure 4.8** DGGE 16S rRNA gene profile for all fractions in WS replicate 2 incubated with A) $^{12}C$ isoprene and B) $^{13}C$-labelled isoprene for 6 days. Each fraction of the DNA centrifugation from 1-12 (heavy DNA to light DNA) is shown on the bottom. Arrows indicate enriched bands. M: DGGE marker ladder.

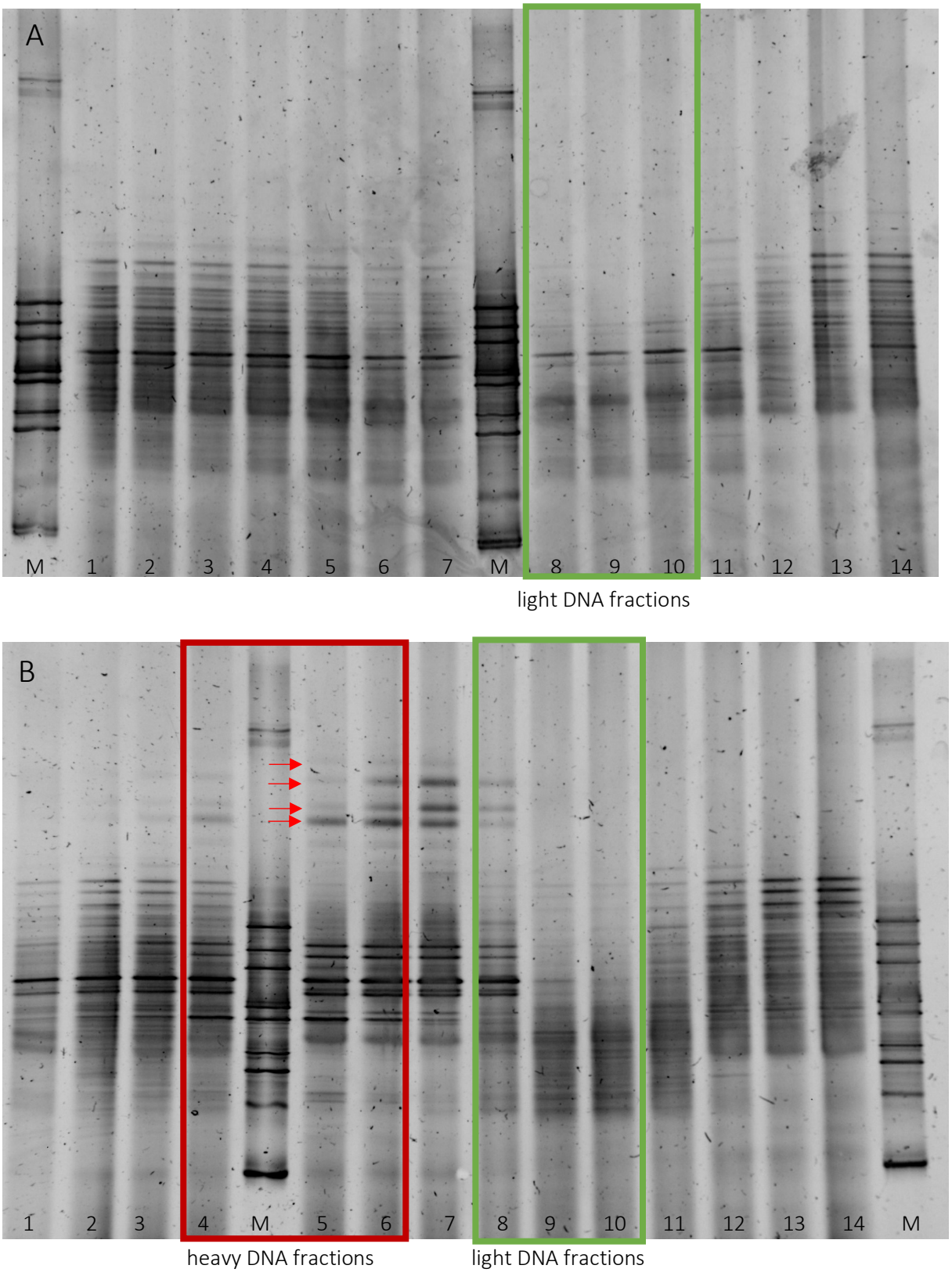



**Figure 4.9** DGGE 16S rRNA gene profile for all fractions in WS replicate 1 incubated with A) $^{12}C$ isoprene and B) $^{13}C$-labelled isoprene for 7 days. Each fraction of the DNA centrifugation from 1-12 (heavy DNA to light DNA) is shown on the bottom. Arrows indicate enriched bands. M: DGGE marker ladder.

Few differences are shown in fractions from $^{12}$C-isoprene incubations (**Figures 4.8A** and **4.9A**). **Figures 4.8B** and **4.9B** incubated with $^{13}$C-labelled DNA show a clear difference between light DNA (fractions 9-10) and heavy DNA (fractions 4-6) enrichment patterns.

**Figure 4.7B** shows at least 8 bands enriched in only the heavy DNA for replicate 2 after 6 days of incubation. Replicate 1 in **Figure 4.9B** shows more than 4 bands enriched in the heavy fractions. To compare results between time points light fractions, from each time point, were pooled together and run in a DGGE gel, alongside pooled heavy DNA fractions for each replicate for $^{13}$C labelled isoprene incubations (**Figure 4.10**).

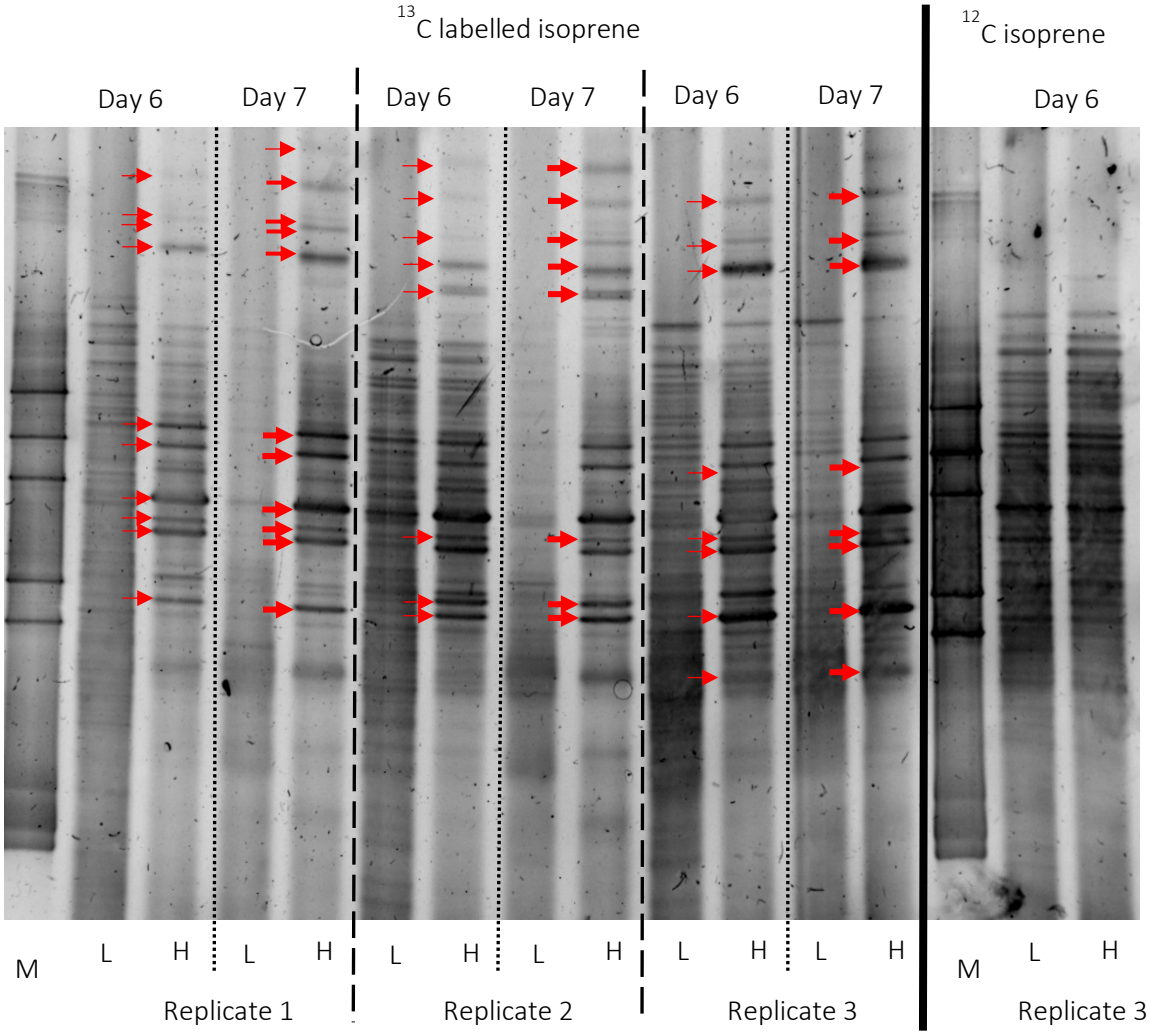



**Figure 4.10** DGGE profiles comparing 16S rRNA gene for the pooled light fractions (8 to 10) and pooled heavy fractions (3 to 5) after 6 and 7 days of incubation with $^{13}$C labelled (lanes 2 to 13) and $^{12}$C isoprene (lanes 15 and 16). Time course labels are on the top of the gel ($T_1$: 6 days; $T_2$: 7 days). Replicates, pooled heavy (H) DNA and pooled light (L) DNA are shown on the bottom. Arrows indicate enriched bands in each replicate. M: DGGE marker ladder.

### 4.3.5 16S rRNA amplicon sequencing to identify isoprene degraders in Colney-WS

Time point 1 light fractions from each replicate of incubations with labelled or unlabelled isoprene were pooled together, according to the type of isoprene used to enrich the microcosm ($T_1$Colney-WS $^{13}C$ Light DNA $T_1$Colney-WS $^{12}C$ Light DNA). The same was done for heavy fractions ($T_1$Colney-WS $^{13}C$ Heavy DNA and $T_1$Colney-WS $^{12}C$ Heavy DNA). Together with native DNA before enrichment, five DNA samples were used as templates for 16S rRNA gene PCR using the 454 sequencing primer pair (27Fmod and 519R, **Table 2.1**, data not shown). **Table 4.4** has the list of all samples sent for amplicon sequencing.

In the case of time point 2, pooled light and pooled heavy DNA from individual replicates was used as a template for 16S rRNA gene PCR. **Figure 4.11** shows an example of the PCR products for the 16S rRNA gene with the twelve DNA samples from time point 2 used as a template.

**Table 4-4** DNA-SIP samples from Colney-WS enrichment sent for 16S rRNA gene amplicon sequencing

| Sample | Time-point | Description |
|---|---|---|
| 1 | - | Native unfractionated DNA |
| 2 | $T_1$ | pooled $^{13}C$ light DNA |
| 3 | $T_1$ | pooled $^{12}C$ light DNA |
| 4 | $T_1$ | pooled $^{13}C$ heavy DNA |
| 5 | $T_1$ | pooled $^{12}C$ heavy DNA |
| 6 | $T_2$ | $^{13}C$ light DNA replicate 1 |
| 7 | $T_2$ | $^{13}C$ heavy DNA replicate 1 |
| 8 | $T_2$ | $^{12}C$ light DNA replicate 1 |
| 9 | $T_2$ | $^{12}C$ heavy DNA replicate 1 |
| 10 | $T_2$ | $^{13}C$ light DNA replicate 2 |
| 11 | $T_2$ | $^{13}C$ heavy DNA replicate 2 |
| 12 | $T_2$ | $^{12}C$ light DNA replicate 2 |
| 13 | $T_2$ | $^{12}C$ heavy DNA replicate 2 |
| 14 | $T_2$ | $^{13}C$ light DNA replicate 3 |
| 15 | $T_2$ | $^{13}C$ heavy DNA replicate 3 |
| 16 | $T_2$ | $^{12}C$ light DNA replicate 3 |
| 17 | $T_2$ | $^{12}C$ heavy DNA replicate 3 |

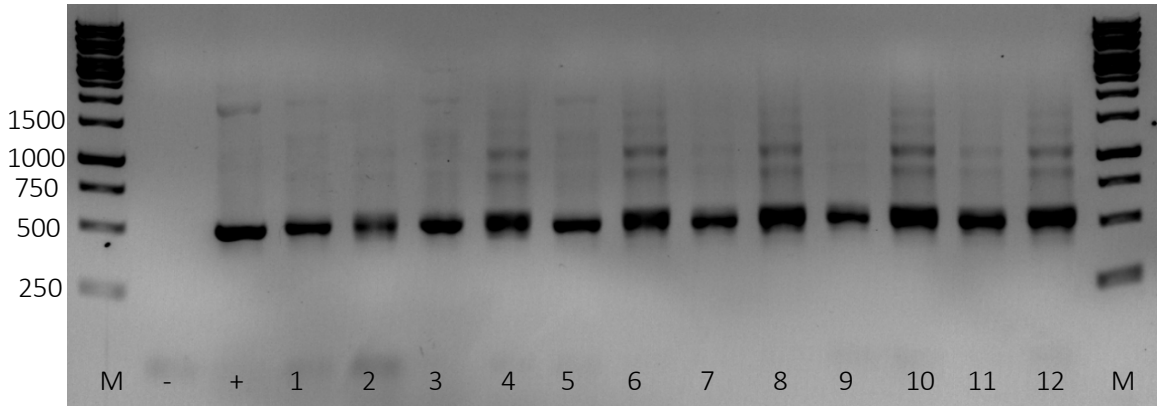



**Figure 4.11 .** 16S rRNA gene PCR products for time point 2 samples using 27Fmod and 519R primer pair. Positive control was *Rhodococcus* sp. AD45 genomic DNA. Gel sample labels for each fraction on the bottom. M: marker Samples correspond to the following:

| Number | Sample | Number | Sample |
|---|---|---|---|
| 1 | $^{13}$C heavy DNA Replicate 1 | 7 | $^{12}$C heavy DNA Replicate 1 |
| 2 | $^{13}$C light DNA Replicate 1 | 8 | $^{12}$C light DNA Replicate 1 |
| 3 | $^{13}$C heavy DNA Replicate 2 | 9 | $^{12}$C heavy DNA Replicate 2 |
| 4 | $^{13}$C light DNA Replicate 2 | 10 | $^{12}$C light DNA Replicate 2 |
| 5 | $^{13}$C heavy DNA Replicate 3 | 11 | $^{12}$C heavy DNA Replicate 3 |
| 6 | $^{13}$C light DNA Replicate 3 | 12 | $^{12}$C light DNA Replicate 3 |

Sequence data obtained were processed at MR DNA (Molecular Research LP) in Shallowater, USA. Taxonomy was assigned using BLASTn against RDPII/NCBI database. For more details see **section 2.16.1** (Cole et al., 2014). Community profiles are similar when comparing the relative abundance of bacterial communities in light and heavy DNA fractions from $^{12}$C isoprene enriched samples after 6 days of incubation. There was a shift in the community when comparing to native soil (see **Figure 4.12**). The abundance of certain bacteria such as *Flavobacterium* and *Sphingobacterium* increases compared to native soil, while other bacteria that were not detected in the native soil appear in to have an advantage in isoprene enriched microcosms (*Ramlibacter* and *Variovorax*).

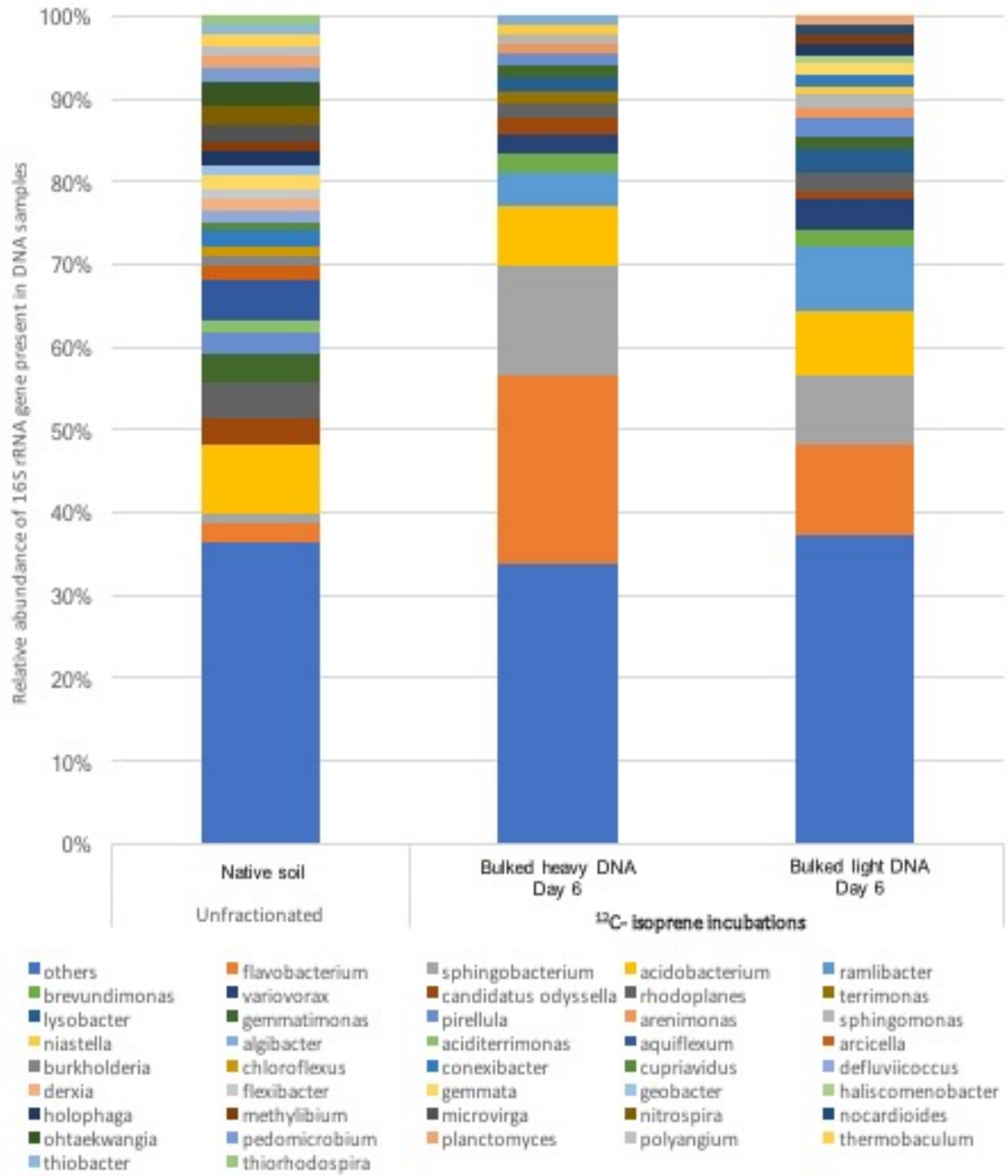



**Figure 4.12** 16S rRNA gene relative abundance of genera from Colney willow soil (WS) from native (un-fractionated) and $^{12}C$ enriched ($T_1$) pooled light and pooled heavy DNA fractions. Genera present at less than 1% relative abundance were group as "others".

The biggest shift in the abundance of the bacteria present during Colney-WS DNA-SIP was observed in the pooled heavy DNA. At the genus level (**Figure 4.13**), while bacterial profiles in the pooled light fraction from the microcosm enriched with $^{13}C$-labelled isoprene are similar to the results from $^{12}C$ isoprene enrichment, over 70% of the 16S rRNA DNA in the heavy fraction corresponds to members of the *Ramlibacter* and *Variovorax* genus. *Rhodococcus* was present in the heavy fraction as well, with a total of 4.5% abundance in the heavy DNA.

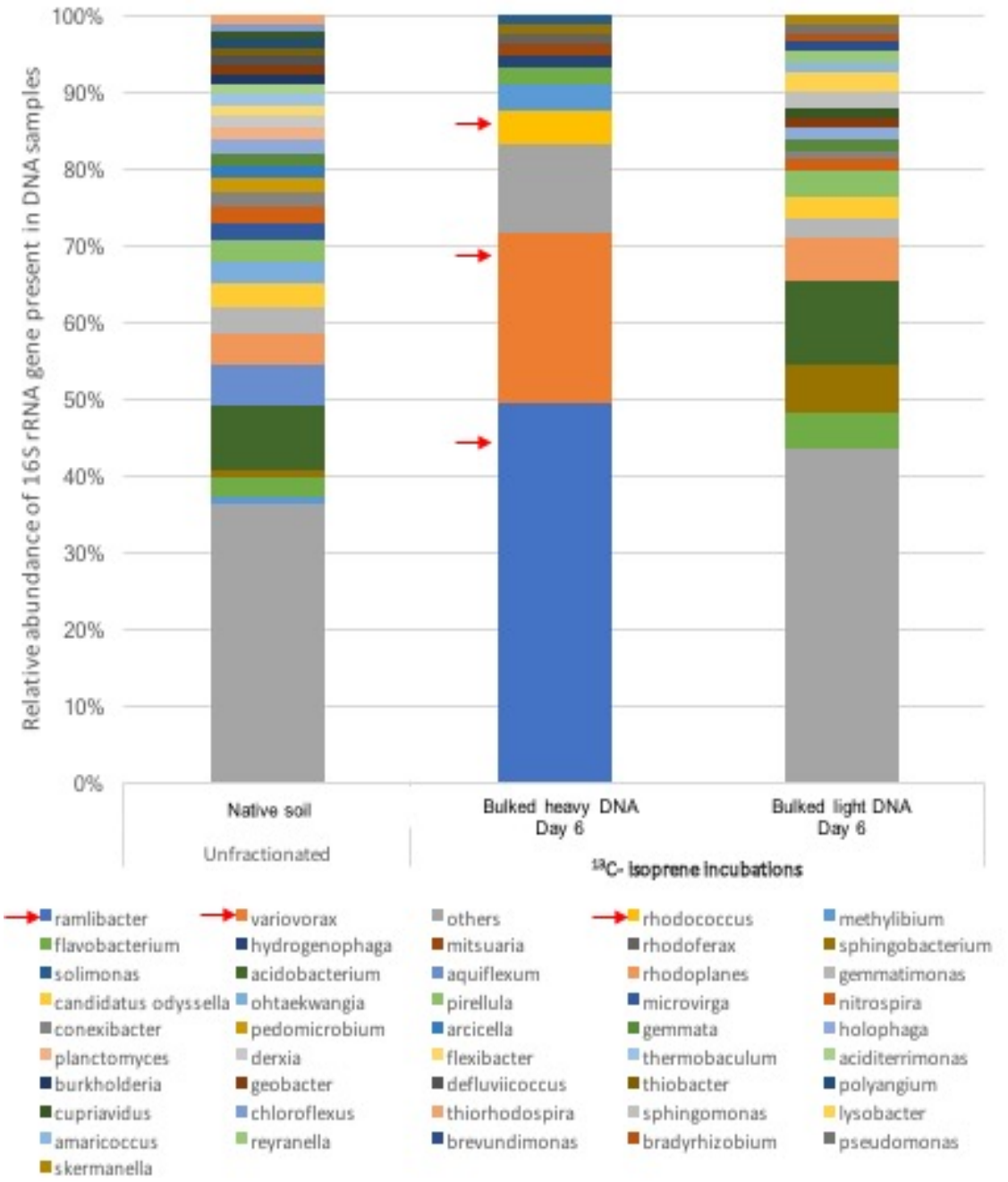



**Figure 4.13** Relative abundance of 16S rRNA gene, at the genus level, in Colney willow soil (WS) from native (un-fractionated) and $^{13}$C-labelled enriched ($T_1$) pooled light and pooled heavy DNA fractions. Genera present at less than 1% relative abundance were group as "others". Arrows indicate putative isoprene-degraders.

Since $^{12}$C-isoprene enrichment does not allow us to determine which bacteria are actively using the isoprene, figures comparing 16S rRNA gene profiles in DNA from $^{13}$C-labelled fractions were constructed. When looking at 16S rRNA gene sequences from the heavy DNA profiles, at a family level, enrichments with $^{13}$C-labelled isoprene show that 75% of the bacteria in the heavy fraction belong to the family Comamonadaceae (see **Figure 4.14**).

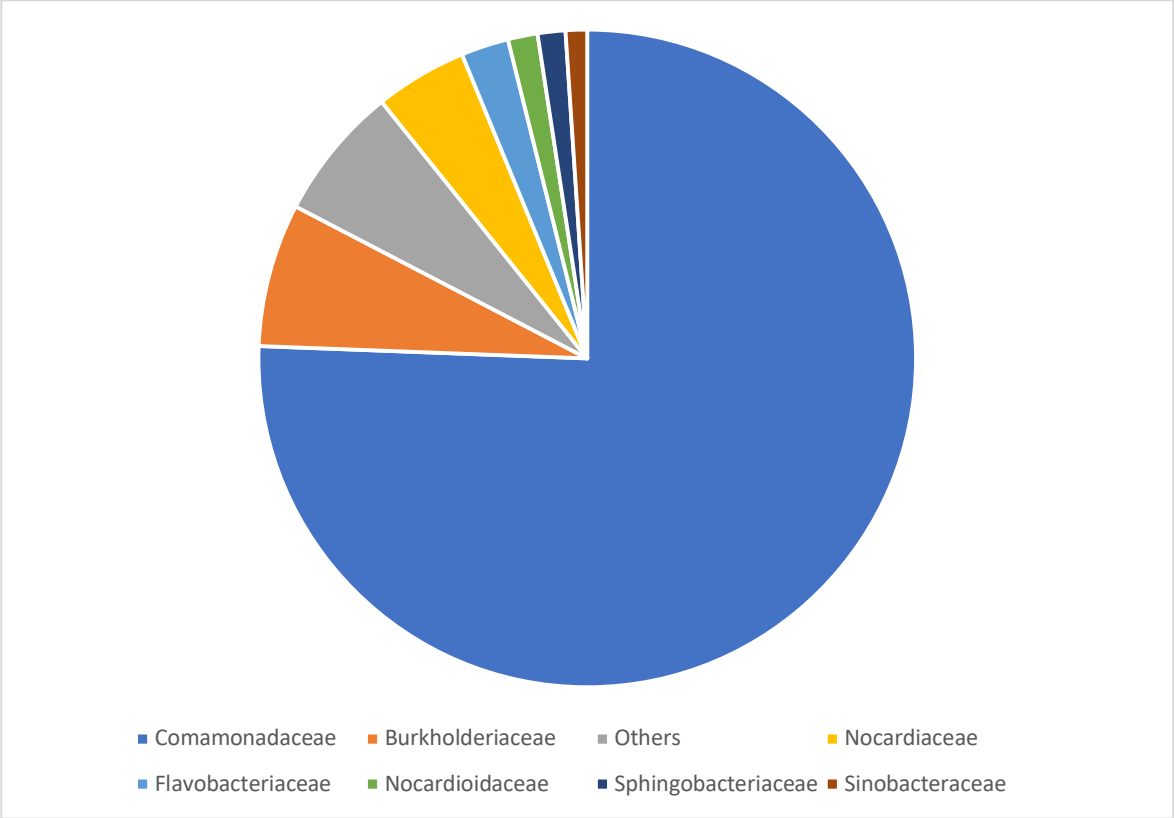



**Figure 4.14** Relative 16S rRNA gene abundance, at the family level, in heavy DNA fractions from Colney-WS DNA-SIP after 6 days of enrichment with $^{13}C$ labelled isoprene. Families present at less than 1% relative abundance were group as "others".

After 6 days of enrichment the members of the Comamonadaceae family are the active players in isoprene degradation in willow topsoil. An extra day of incubation had resulted in a doubling of the incorporation of $^{13}C$ from $^{13}C$-isoprene. In **Figure 4.15** the pooled heavy DNA fractions from 6 days of incubation were compared to three individual replicates sacrificed after 7 days of enrichment. Data show similar results, the abundance of the predominant genus (*Ramlibacter*, *Variovorax* and *Rhodococcus*) at day 6 and day 7. *Methylibium*, *Ideonella* and *Curvibacter* were only detected at an abundance of over 1% after 7 days of enrichment. All 16S rRNA gene profiles are shown in **Supplementary Figure 1**.

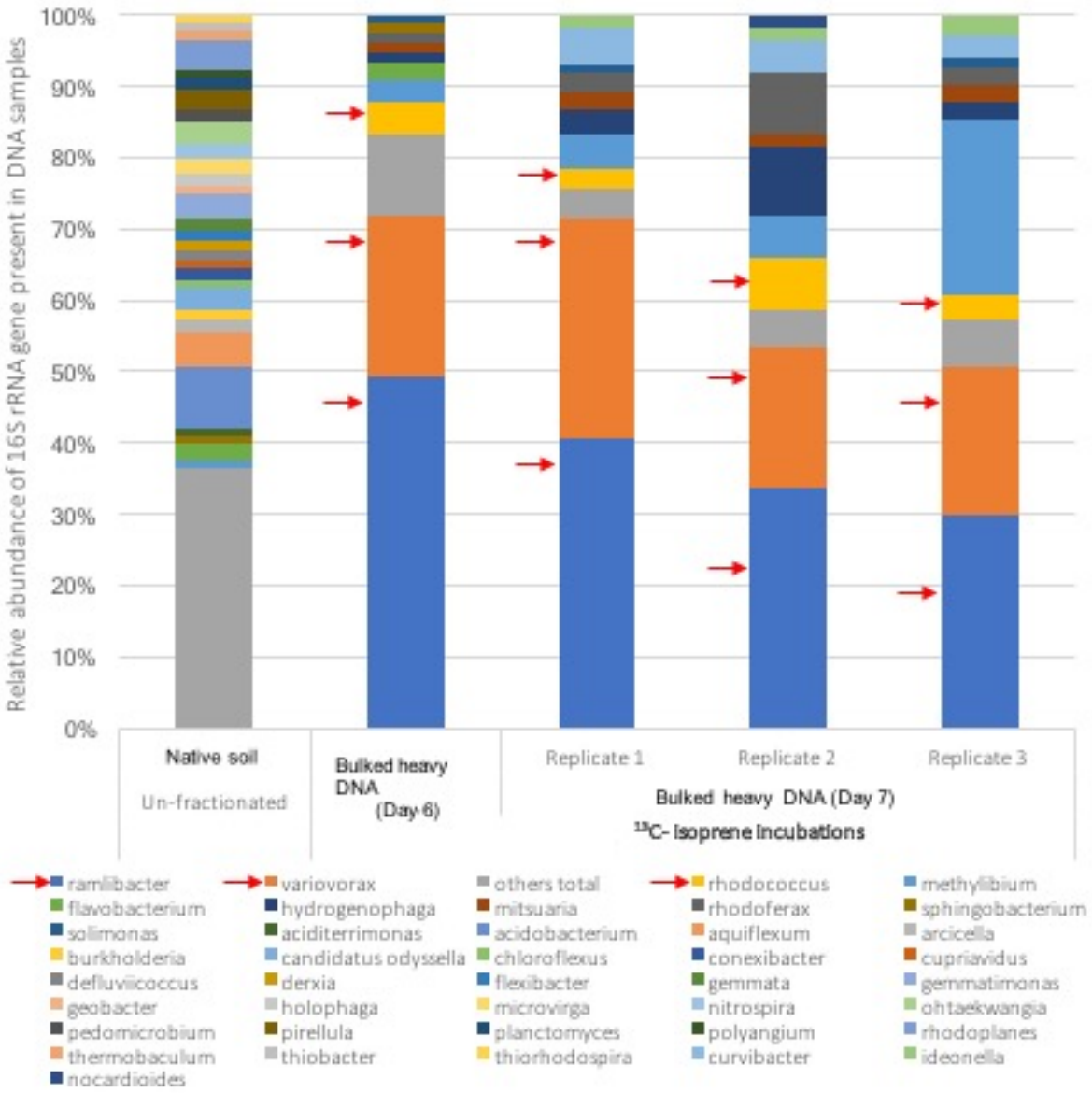



**Figure 4.15** Relative abundance of 16S rRNA, at the genus level, in Colney willow soil (WS) from native (un-fractionated) and $^{13}$C-labelled heavy DNA after 6 (pooled) and 7 days of enrichment. Genera present at less than 1% relative abundance were group as "others". Arrows indicate putative isoprene-degraders.

To observe the changes of certain bacteria throughout the enrichment the most abundant (over 1%) for native unfractionated DNA and heavy DNA fractions were compared, **Figure 4.16** shows the relative changes in the abundance of each bacterium. In the native soil, large numbers of genera were present in low numbers (referenced as "others"). Some key genera were barely detectable in the native soil but became the dominant genera after incubations with isoprene.

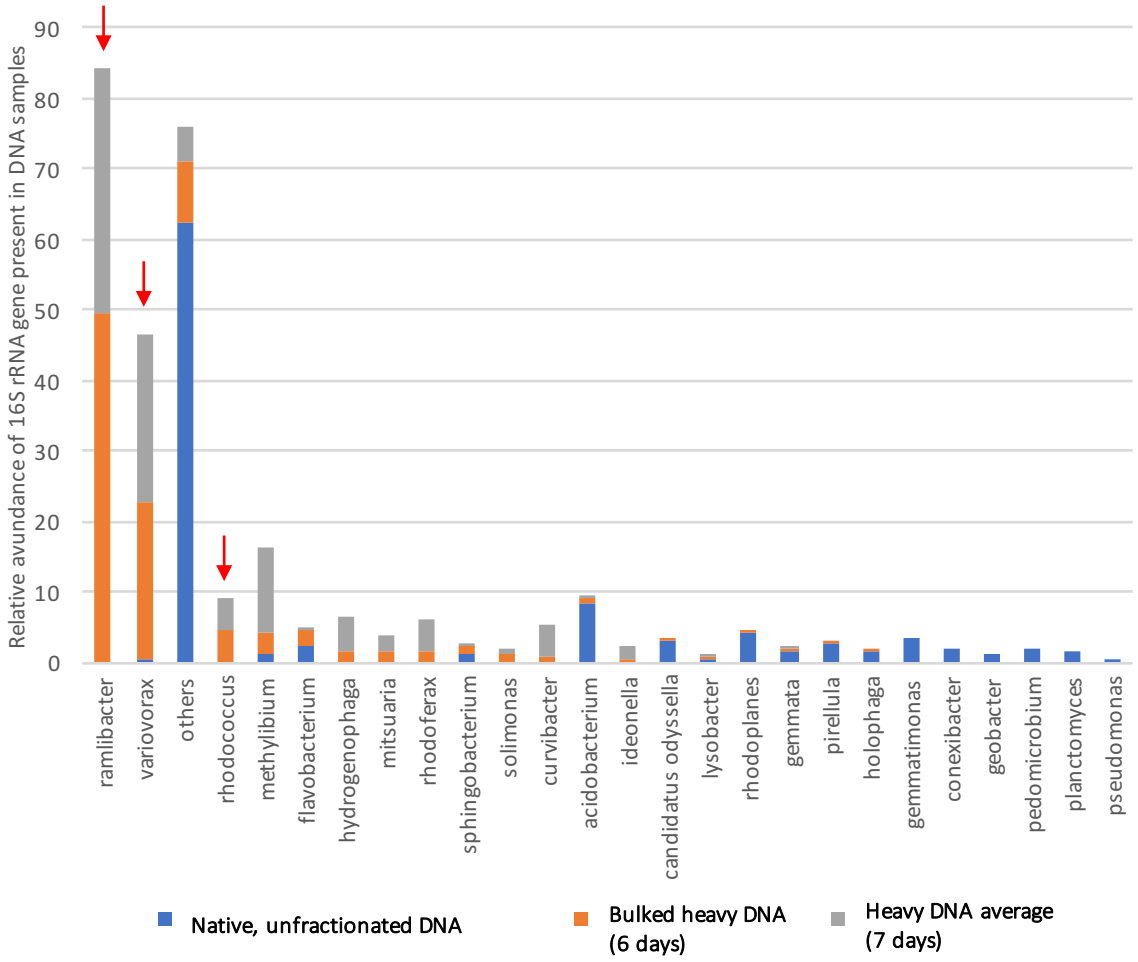



**Figure 4.16** Changes in the relative abundance of different bacteria genera throughout Colney-WS SIP experiment. $T_2$ abundance is an average between the three replicates. Genera present at less than 1% relative abundance were group as "others". Arrows indicate putative isoprene-degraders.

The enrichment process, and conditions of batch incubation, had important effects on the bacterial community present in the willow soil. The major isoprene-degrading bacteria in the Colney-WS SIP enrichment were *Ramlibacter* and *Variovorax*. Surprisingly, the well-known isoprene-degrader *Rhodococcus* was also present but at less than 5 % at relative abundance at each time-point. Other bacteria such as *Methylibium*, *Hydrogenophaga*, *Rhodobacter* and *Curvibacter* increased in abundance as enrichment progressed. *Acidobacterium*, Candidatus *Odyssella* and *Rhodoplanes*, were abundant in the native soil and their relative abundance decreased in relative abundance during enrichment. Comparison of relative abundance of 16S rRNA genes at the genus and family level in both, light and heavy fractions, $^{12}C$ and $^{13}C$ incubations after 6 days of enrichment are shown in **Supplementary Figure 2** and **Supplementary Figure 3**, respectively.

### 4.3.6 Comparison of culture-dependent and culture-independent results obtained with Colney-WS SIP experiments

Isoprene degraders were isolated from isoprene-enriched willow topsoil used during the Colney-WS SIP (isolates WS4 to WS9, see **Table 3.6**). Since 16S rRNA gene sequence analysis has elucidated the relative abundance of the key bacteria present in the sample, comparing the results obtained to cultivation-dependent data was necessary. The cultivation-independent approach verified the abundance of the *Variovorax* sp. strain WS9 isolate in the willow soil environment during the DNA-SIP. A phylogenetic tree of the most abundant OTUs (over 1%) after enrichment was constructed (**Supplementary Figure 4**). 44 OTUs, from the 16S rRNA gene amplicon sequencing data, aligned with members of the Comamonadaceae family.

A phylogenetic tree was constructed with the V1-V3 region of 16S rRNA gene (177 bp) for representative bacteria from *Ramlibacter* and *Variovorax* genus, including the *Variovorax* sp. strain WS9 isolate and its closest identity in the NCBI database (*Variovorax* sp. Cd61) (**Figure 4.17**). The size and location of the partial 16S rRNA gene was due to the size of amplicons obtained from the 16S rRNA gene amplicon sequencing analysis. The figure shows two distinct clades in which all *Ramlibacter* genus members are found in one clade. Most *Variovorax* members are found in the second clade except for *Variovorax* sp. strain WS9. *Variovorax* sp. strain WS9 16S rRNA partial gene aligned with sequences from the genus *Ramlibacter.* This result shows the V1-V3 region of the 16S rRNA gene that was amplified and sequenced was not sufficient to determine the identity of members of the Comamonadaceae present in the sample. *Variovorax* sp. strain WS9 was identified as a member of the *Ramlibacter* genus according to the GreenGenes database in the 16S rRNA amplicon sequence analysis (**Figures 4.13 to 4.16**).

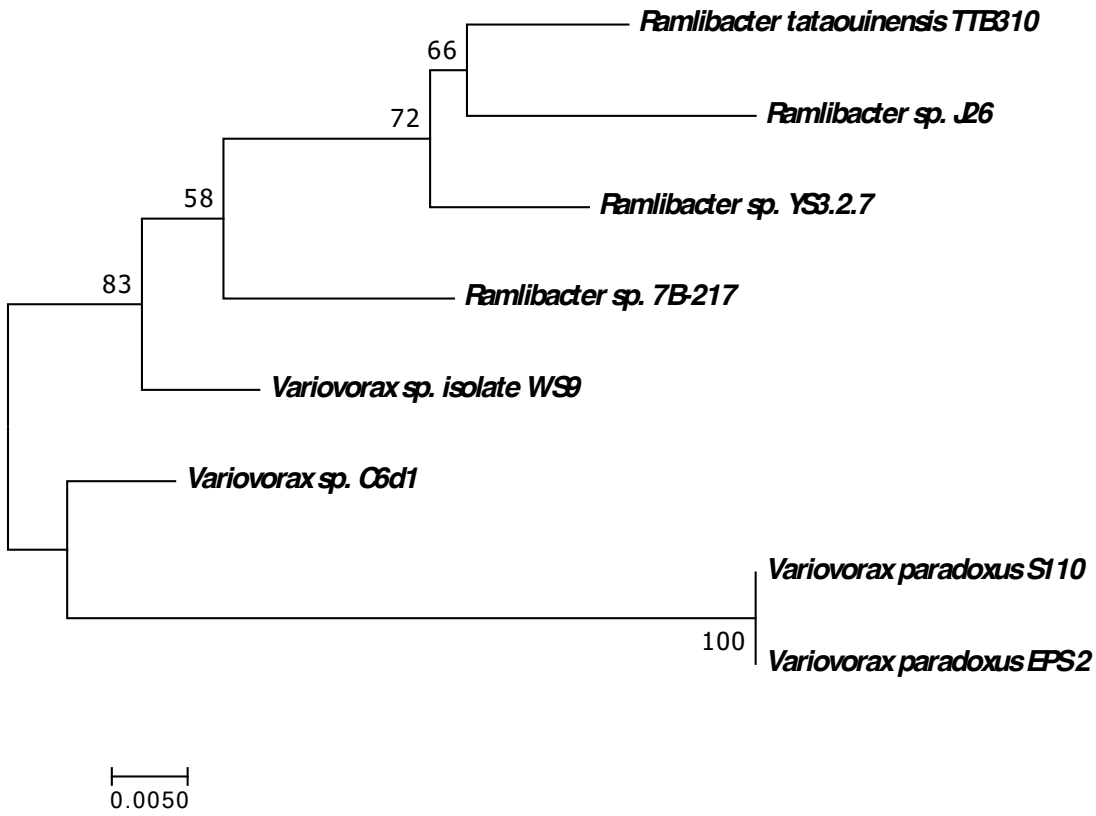



**Figure 4.17** Phylogenetic tree for a partial 16S rRNA gene (V1-V3 region, 492 bp) of representative Ramlibacter and Variovorax bacteria. The tree was generated using Maximum Likelihood and a bootstrap of 1000 replicates.

15 OTUs identified as partial 16S rRNA gene (492 bp) from *Ramlibacter* aligned with *Variovorax* sp. strain WS9. The abundance of the 15 OTUs, in the heavy DNA after 7 days of incubation, was 19% in the three replicates. 16S rRNA gene sequence data showed 76% abundance of the Comamonadaceae family after 6 days of enrichment with isoprene, the sequence data also shows that approximately 1/4 of the 16S rRNA gene sequences corresponded to the isoprene-degrader *Variovorax* sp. strain WS9.

Using phylogenetic tree clade analysis, another 18 OTUs identified as partial 16S rRNA gene from *Ramlibacter* were considered part of the *Variovorax* genus. The initial 16S rRNA gene amplicon sequencing of the V1-V3 region identifies 13 OTUs as part of the *Variovorax* genus, phylogenetic analysis has identified an additional 33 OTUs that should also be identified as part of *Variovorax* genus. These results showed that in replicate 1, after 7 days of enrichment, *Variovorax*-like OTUs account for more than 52% of the heavy fraction.

### 4.3.7 Metagenome analysis of DNA-SIP from Colney-WS

Pooled heavy fractions from each time-point ($T_1$ and $T_2$) and the unfractionated native soil were sent off for metagenome sequencing. Dr. Jennifer Pratscher performed the phylogenetic analysis, using the Metaphlan pipeline, of the unassembled metagenomes based on a wide number of marker genes are shown in **Figure 4.18**. These results, based on more than 16S rRNA gene diversity, give a holistic view of the effects of isoprene enrichment on the willow soil microcosm. Metagenome gene profiles were similar to results from 16S rRNA gene amplicon sequencing. A few closely related members of the Comamonadaceae family were abundant after incubation with isoprene (*Variovorax, Polaromonas* and *Thiomonas*). Curiously, the metagenome analysis did not show the genus *Ramlibacter*.

The data confirmed lower abundance of Actinobacteria in Colney-WS metagenomes. A specific *Rhodococcus* species (*Rhodococcus opacus*) slowly increased during the enrichment. Members of the *Nocardioides* genus were less abundant than *Rhodococcus* and were only detectable after 6 days of enrichment.

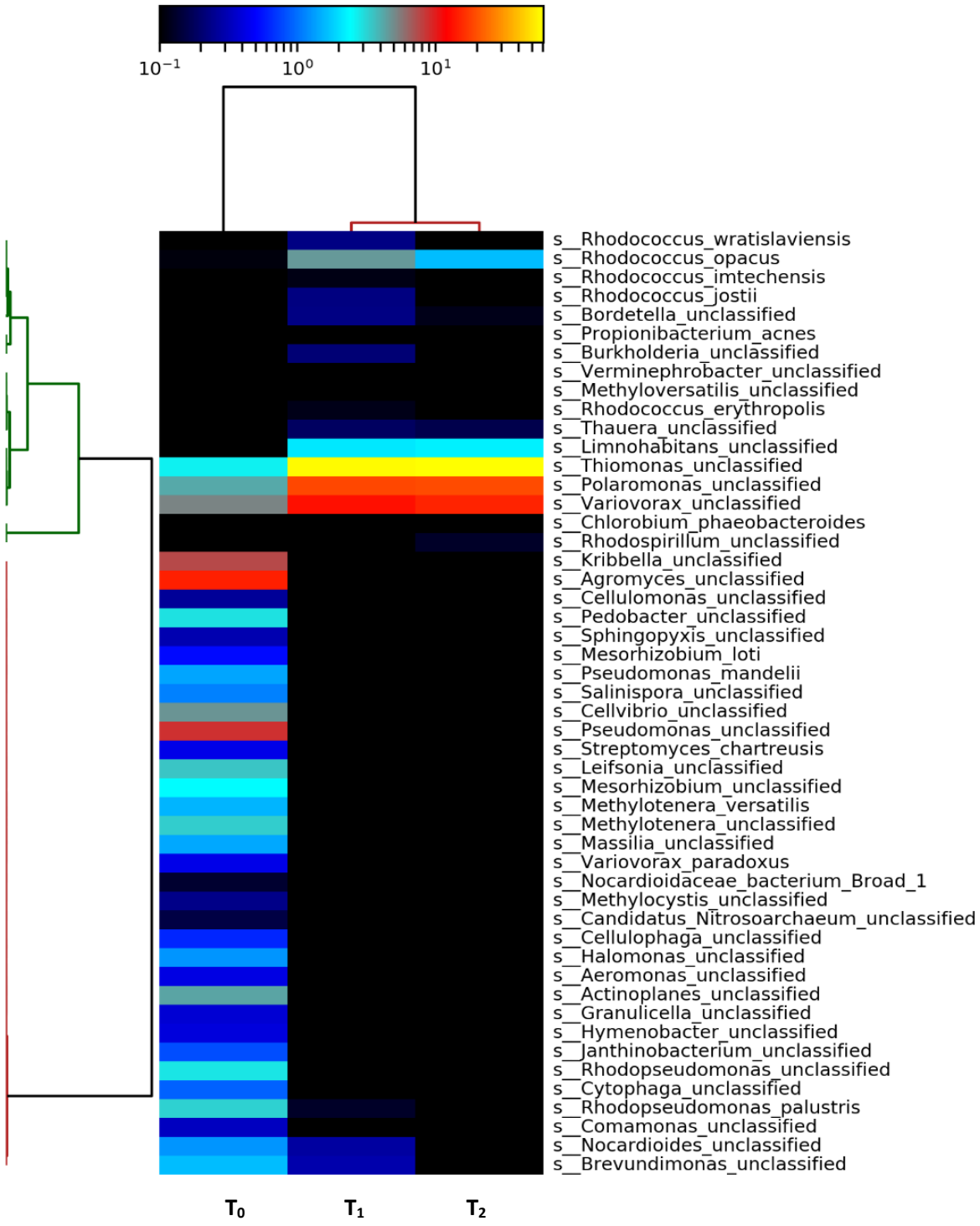



**Figure 4.18** Heat map of the Phylogenetic abundance analysis of native ($T_0$) and enriched ($T_1$ and $T_2$) metagenome raw reads using Metaphlan software (Segata et al., 2012). Analysis was performed using trimmed unassembled metagenome data and cluster analysis shows members of the Comamonadaceae family were enriched during the enrichment of willow soil. The number of raw reads for the samples was 42.6 m, 44.5 m and 52.2 m for each sample, respectively. Trimmed reads was 41.8 m, 43.7 m and 51.1 m, respectively.

Metagenome raw reads were assembled by Dr. Jennifer Pratscher at each time-point in search for functional genes of un-cultivable isoprene-degraders. Enriched time-points were

co-assembled to obtain longer contigs and complete gene sequences. Assemblies were compared to a curated IsoA database doing a tblastx search to select contigs with partial or complete IsoA gene sequences. The contigs with the best E-values and search scores were selected and the *isoA* gene sequences were extracted (details in **section 2.18**). *isoA* gene sequences from the metagenomes were translated and aligned with relevant sequences from the IsoA database to make an IsoA tree (**Figure 4.19**).

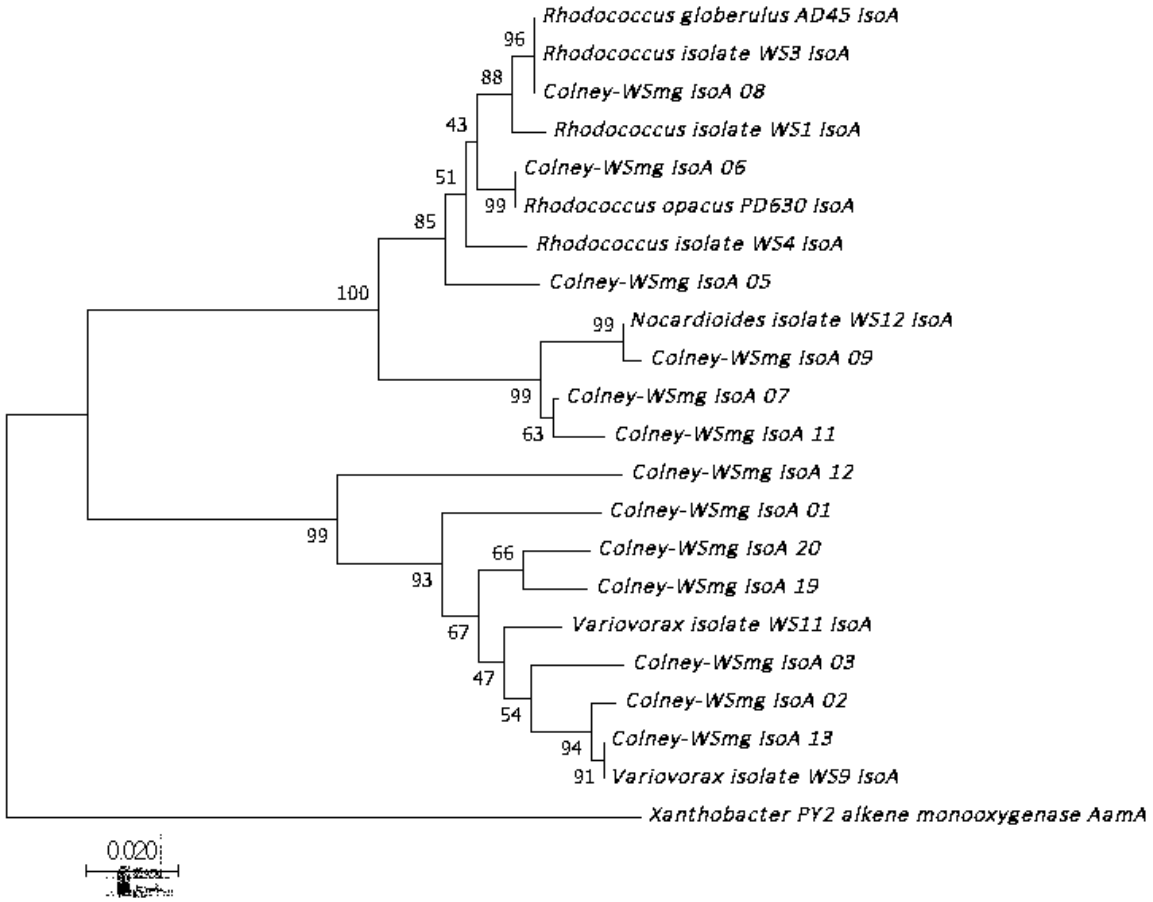



**Figure 4.19** IsoA protein tree of from isoprene-degrading bacteria and Colney-WS heavy DNA co-assembled metagenome (Colney-WSmg). *Xanthobacter autotrophicus* Py2 AamA protein sequence was used as the outgroup. The tree was constructed using Maximum Likelihood with a bootstrap of 1000 replicates. The analysis involved 22 amino acid sequences of 244 amino acid positions. All positions containing gaps and missing data were eliminated.

The phylogenetic tree for IsoA shows the representative sequences obtained from enriched heavy DNA metagenomes from Colney-WS. These sequences cluster closely to isoprene-degrading isolates obtained from the same soil sample (see **Table 3.6**) and reference bacteria that have previously been reported to degrade isoprene (*Rhodococcus* sp. AD45 and *Rhodococcus opacus* PD630). For a complete list of the sequences extracted from the metagenomes refer to **Supplementary Figure 5**.

In the terrestrial environment, Gram-negative and Gram-positive bacteria have now been identified as isoprene-degrading bacteria (**Chapter 3**). Two distinct clades, for the two group types of bacteria, are shown in **Figure 4.19**. In the Gram-Positive clade *Rhodococcus* isolates and the *Nocardioides* isolate are included. Some metagenome sequences from the heavy DNA from Colney-WS group to the Gram-positive clade (6 sequences in total). A total of three sequences aligned with *Rhodococcus* and another three with the *Nocardioides*. Only one of the three *Nocardioides*-like metagenome sequences aligned closely with the *Nocardioides* sp. strain WS12 isolate. The other two sequences are different to the *Nocardioides* isolate, which suggests more diversity of the *isoA* sequences in the Gram-positive sequences from the metagenome.

A total of 16 Colney-WS metagenome sequences aligned to the Gram-negative clade. Seven sequences aligned closely to isolate *Variovorax* sp. strain WS9 and four sequences to isolate *Variovorax* sp. strain WS11. Another 5 metagenome sequences form part of the Gram-negative clade but are not close to the available isolate IsoA sequences. The results suggest there was still a wider diversity of un-cultivable Gram-negative bacteria that can use isoprene as their carbon source in the Colney-WS SIP.

## 4.4 Discussion

The isoprene concentrations used during the isoprene enrichment of willow soil microcosms in this study are 250-fold lower than the experiments performed by Khawand *et al.,* 2016. Although the concentrations used in the experiment are still greater than in the natural environment (Wiedinmyer *et al*., 2005), these concentrations permitted the effective incorporation of $^{13}$C-labelled isoprene over a few days of incubation.

DGGE profiling at different time-points coupled with the amount of heavy DNA obtained throughout time have verified the effective incorporation of the $^{13}$C-Carbon label in certain bacteria present in the microcosm throughout time. Analysis shows qualitative and quantitative evidence of degradation of isoprene by bacteria present in the microcosm. Continuous replenishment of isoprene also increased the probabilities of slow growing microbes to use the available isoprene directly. An automated system would facilitate the monitoring process when handling low concentrations of isoprene.

Cultivation-independent data shows the members of the Comamonadaceae family as the major players in isoprene biodegradation. The results of Khawand *et al*. in 2016 show a relative abundance of around 85 % for *Rhodococcus* genus and 6.5% for Comamonadaceae family, lowering the concentration of isoprene almost inverts those values to over 75% Comamonadaceae and approximately 4.4% *Rhodococcus*. These results suggest that the concentration of isoprene supplemented to the microcosm in the soil affects which bacteria have a metabolic advantage. *Rhodococcus* bacteria may thrive in high concentrations of isoprene while the same group might be outcompeted by members of the Comamonadaceae family at lower concentrations.

16S rRNA gene amplicon sequencing results showed a few bacteria that increased in abundance after 7 days of incubation, later than what was expected for a true isoprene degrader. *Methylibium* was an example of an organism that might be a cross-feeder due to an abundance lower than 1% on day 6 of the enrichment. Isoprene degraders were considered to have increased in abundance early-on in the experiment.

SIP-metagenome data from heavy fractions reflected similar results to 16S rRNA gene amplicon sequencing. This study has found that members of the Comamonadaceae family are the predominant isoprene degraders with enrichments at lower concentrations of isoprene. While amplicon sequencing of partial 16S rRNA gene suggests the most abundant genera are *Ramlibacter* and *Variovorax*, metagenome analysis on the unassembled reads using Metaphlan suggests those to be *Variovorax*, *Polaromonas*, *Thiomonas* and *Limnohabitans*. All four genera correspond to the family Comamonadaceae; *Polaromonas* has been found to degrade hydrocarbon- and xenobiotic- compounds (Bravo *et al*., 2014; Mattes *et al*., 2008; Xie *et al*., 2011), *Thiomonas* is a chemilithoautotroph found in environments impacted by acid mine drainage and has been reported to use arsenite as an energy source (Arsène-Ploetze *et al.*, 2010; Battaglia-Brunet *et al*., 2006) and *Limnohabitans* has been found as a freshwater bacterioplankton (Kasalický *et al*., 2010, Jezbera *et al.*, 2013). Furthermore, metagenome *isoA* search has shown that most of the assembled gene sequences correspond to sequences that cluster with the Comamonadaceae/*Variovorax* isolate IsoA sequences available.

Cultivation-dependent data and cultivation-independent data obtained from the DNA-SIP have extended our knowledge of isoprene degrading bacteria and the diversity of isoprene metabolic genes in the environment. Fully-labelled isoprene enrichments at lower concentrations have shown the members of the Comamonadaceae family, namely *Variovorax*, as a major player at conditions closer to the ones found in nature. Further analysis of other isoprene metabolic genes and their abundance in nature is necessary to help understand/discriminate between Gram-positive and Gram-negative isoprene degraders in the environment. Analysis of isoprene degradation genes in the willow soil metagenome data and cultivated organisms genomes will be presented in **Chapter 6**.

# Chapter 5 Identification of active isoprene-degrading bacteria in the soil and phyllosphere of an oil palm tree using DNA-stable isotope probing

## 5.1 Introduction

500 Tg of isoprene is produced naturally by isoprene-emitting trees each year (Sharkey, 2013; Guenther *et al*., 2006). Plants are classified as high- or low- isoprene emitters and oil palm trees (*Elaeis guineensis*) have the highest isoprene emission potential of the tropical plants reported to date (Hewitt & Street, 1992). In the last few decades there has been a considerable expansion of oil palm acreage in tropical Asia, Africa and America as the demand for biofuels, palm oil, and other products from the palm have increased (Corley & Tinker, 2003).

Recent studies have evaluated the effects of large oil palm plantations on tropical biodiversity (Koh & Wilcove, 2008). Since oil palm trees are one of the highest emitters of isoprene, the study of the microbial ecology in leaves and soil is essential to understand how isoprene-degrading bacteria are distributed in the environment. Oil palm trees are naturally found in the tropics. In the UK only a few greenhouses (with the appropriate conditions) can grow the tree.

To complement the work presented in previous chapters, soil and leaf microcosms from the same tree were studied using DNA-SIP incubations. This is the first study that evaluated the presence of active isoprene-degrading bacteria from high- and low-biomass samples surrounding a single isoprene-emitting tree.

*General objective:*

To evaluate the diversity, identity and abundance of key isoprene-degrading bacteria in soil around an oil palm tree and oil palm leaves using DNA-SIP.

*Specific objectives:*

To determine isoprene consumption rates of $^{12}$C-isoprene and $^{13}$C-labelled isoprene enrichments with oil palm tree soil and leaves.

To verify incorporation of $^{13}$C-label and enrichment of isoprene-degrading bacteria using 16S rRNA gene DGGE community profiling of $^{12}$C- and $^{13}$C- DNA isolated during DNA-SIP experiments.

To determine the changes in diversity and abundance of key isoprene-degraders in soil and leaf enrichments over time by monitoring the bacterial community composition using 16S rRNA gene amplicon sequencing.

## 5.2 Results

### 5.2.1 DNA-Stable isotope probing (SIP)

Soil from around an isoprene-emitting oil palm tree (OPS), and leaves were collected in February 2017 from the Palm House greenhouse in Kew Gardens, London (**Figure 5.1**).

### 5.2.1.1 Set up of DNA-SIP experiments

*Oil palm tree soil*

Oil palm soil (OPS) microcosms consisted of 4 g of soil and 40 ml of sterile water in a sterile 2-liter flask. Enrichments with $^{12}$C-isoprene or fully-labelled $^{13}$C-isoprene (obtained from Dr Gregg Whited, DuPont Industrial Biosciences; **section 2.10**) were done in triplicate with a final concentration of approximately 25 ppmv isoprene in the headspace. All flasks were incubated in the dark at 25 °C and shaken at 150 r.p.m. Negative controls had no consumption of isoprene or growth (data not shown).


**Figure 5.1** Collection of top soil and leaves from an oil palm tree (*Elaeis guineensis*) sampled in The Palm House, Kew Gardens (November 2016). (A, C) Oil palm tree; (B) subsurface soil samples, (D) oil palm compound leaf; (C) oil palm leaflets.

*Oil palm leaf washings*

Cells recovered from the surface of the oil palm leaves (OPL) suspended in 50 ml of minimal media (MM), were placed in sterile 2-litre flasks. Each leaf washing microcosm was then sealed and enriched with $^{12}$C-isoprene or fully-labelled $^{13}$C-isoprene at a final concentration of approximately 25 ppmv in the headspace. All flasks were incubated in the dark at 25 °C and shaken at 150 r.p.m. Negative controls, with no isoprene or no biomass, had no consumption of isoprene or growth (data not shown).

*Assays for oil palm endophytes*

After retrieving epiphytic cells, leaves were washed consecutively with detergent to eliminate any cells left on the surface of the leaves. The leaves were then cut and macerated to obtain endophytic cells (details in **section 2.9.1**). $^{12}$C- isoprene was added at a final headspace concentration of 25 ppmv to each flask and monitored for a couple of weeks. No isoprene consumption was detected in that period. Two months later, isoprene concentration was tested once again and had been consumed by the bacteria recovered from inside the leaves. Due to time constraints, this experiment was terminated and not taken further. More analysis of the isoprene consumption by endophytes from isoprene-emitting trees is necessary to determine their presence, abundance and identity.

**Table 5.1** indicates the incubation times at which OPS and OPL samples were harvested during the SIP enrichments and the estimated volume of $^{13}$C- isoprene degraded.

**Table 5-1** Volume of $^{13}C$- isoprene gas consumed by oil palm soil (OPS) and leaf (OPL) incubations

| Sample | Time course | $^{13}C$ isoprene spiked into the headspace (ml) | Time (h) |
|---|---|---|---|
| OPS SIP | Native soil ($T_0$) | 0 | 0 |
| OPS SIP | Time point 1 ($T_1$) | 0.95 * | 142 |
| OPS SIP | Time point 2 ($T_2$) | 1.45* | 166 |
| OPS SIP | Time point 3 ($T_3$) | 1.95* | 190 |
| OPL SIP | Native leaf washing ($T_0$) | 0 | 0 |
| OPL SIP | Time point 1 ($T_1$) | 0.9 ˆ | 214-226 |
| OPL SIP | Time point 2 ($T_2$) | 1.5 ˆ | 274-285.5 |
| OPL SIP | Time point 3 ($T_3$) | 2.0 ˆ | 302-311.5 |

*$T_1$, $T_2$ and $T_3$ correspond to 25 (or 31.2) μmol $g^{-1}$, 50 (or 53.1) μmol $g^{-1}$, and 75 (or 85.9) μmol $g^{-1}$ of $^{13}C$- isoprene incorporated, respectively. $^{13}C$-Carbon incorporation values were predetermined before the assay based on the WS-SIP experiment (**Chapter 4**). The actual incorporation at each time point, in parenthesis, was calculated assuming that 34% of $^{13}C$arbon consumed was assimilated into biomass.

ˆ $T_1$, $T_2$ and $T_3$ correspond to targeted 25 μmol $g^{-1}$, 50 μmol $g^{-1}$, and 75 μmol $g^{-1}$ of $^{13}C$- isoprene incorporated, respectively. $^{\pm}$100% corresponds to 33.37 grams of washed leaves suspended in 50 ml of minimal medium. 10 ml of sample was harvested at each time-point.

### 5.2.1.2 Monitoring isoprene uptake throughout the OP soil and leaf SIP incubations

*Oil palm tree soil*

The oil palm soil (OPS) SIP experiment took less than 8 days to incorporate 75 µmol $g^{-1}$ $^{13}$C-carbon. Figure **5.2** shows the isoprene consumption for both $^{12}$C- isoprene and $^{13}$C- labelled isoprene incubations in the first 90 hours of the incubation.

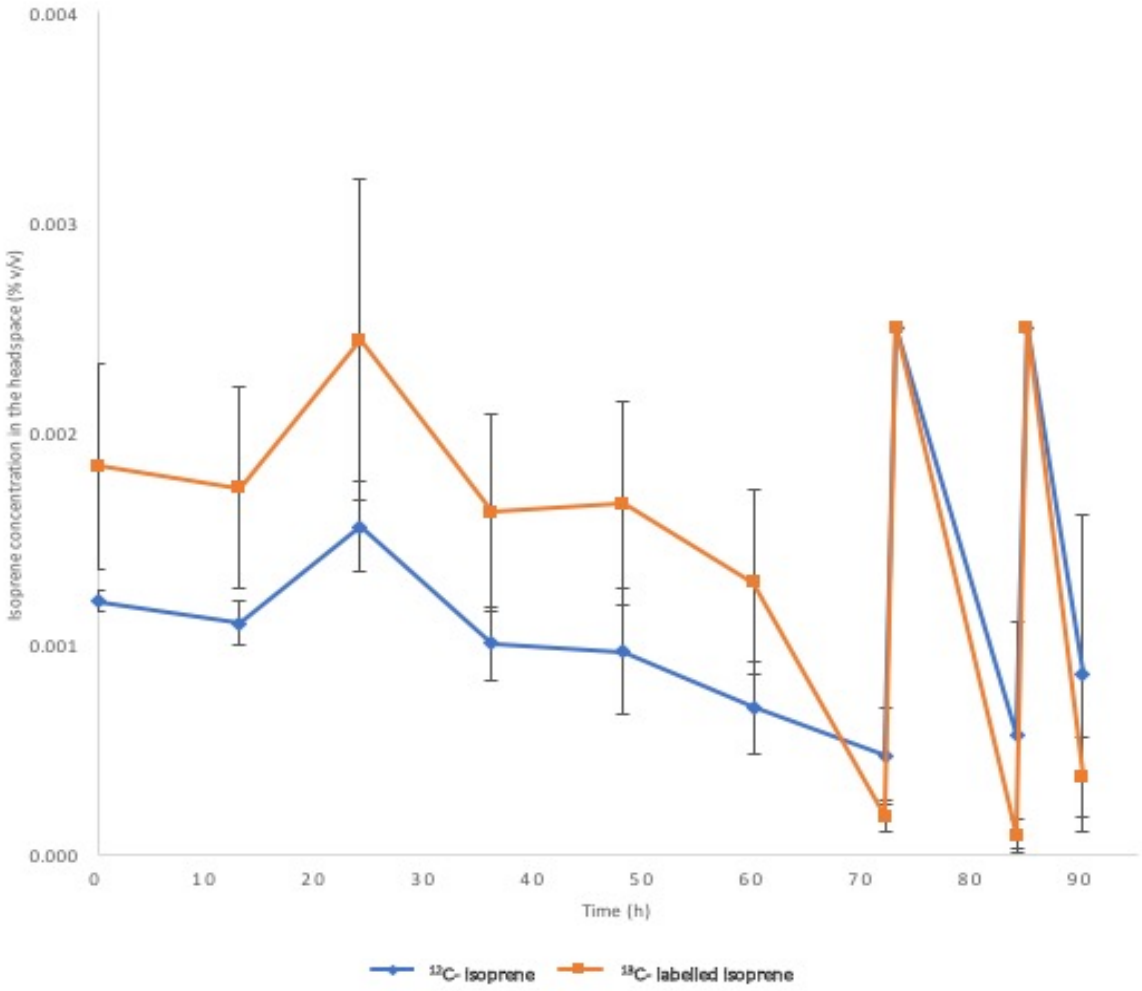



**Figure 5.2** Isoprene consumption during OPS SIP incubations over the first 90 h. Both $^{12}$C isoprene and $^{13}$C labelled isoprene incubations were done in triplicate, error bars show the standard deviation.

The incubation times, isoprene depletion and spiking pattern in the assay were similar to results obtained for the willow soil (Colney-WS) SIP experiments (**Chapter 4**). Initial isoprene consumption rates in the microcosms for $^{12}$C- and $^{13}$C-labelled isoprene are shown

in **Table 5.2**. The consumption rate for both types of isoprene was almost identical between replicates from different treatments. These results verify that there is no difference in the consumption of either type of isoprene during the soil enrichment assay.

**Table 5-2** Estimated consumption rate (%) of isoprene per hour for oil palm soil (OPS) samples during the SIP experiment. The first spike of 25ppmv isoprene was consumed in approximately 75 hours. The average standard deviation of the samples was 0.00017 μmol $h^{-1}$ $g^{-1}$ for $^{12}C$-isoprene and 0.00012 μmol $h^{-1}$ $g^{-1}$ for $^{13}C$-isoprene.

| Sampling site | Isoprene type | Consumption rate (μmol $h^{-1}$ $g^{-1}$ of soil) |
|---|---|---|
| OPS | $^{12}C$ | 0.010 |
| OPS | $^{13}C$ | 0.011 |

Depletion of the first spike of isoprene from OPS took approximately 20 hours more than Colney-WS. The biotic and abiotic characteristics of the oil palm soil might have affected the efficiency of isoprene consumption. Although the initial isoprene consumption rate of Colney-WS (0.023-0.024 μmol $h^{-1}$ $g^{-1}$) was higher than the results observed for OPS, the efficiency of isoprene consumption also increased as the soil enrichment progressed.

*Oil palm leaf washings*

Epiphytic cells from oil palm leaves (OPL-SIP) took 13 days to incorporate 75 μmol $g^{-1}$ $^{13}C$-carbon. Consumption time of isoprene from the OPL-SIP was similar to the data for OPL washings and OPL swabs presented in **Chapter 3** (i.e. 130-140h to consume the first spike). **Figure 5.3** shows the isoprene consumption for both $^{12}C$ isoprene and $^{13}C$- labelled isoprene incubations in the first 150 hours.

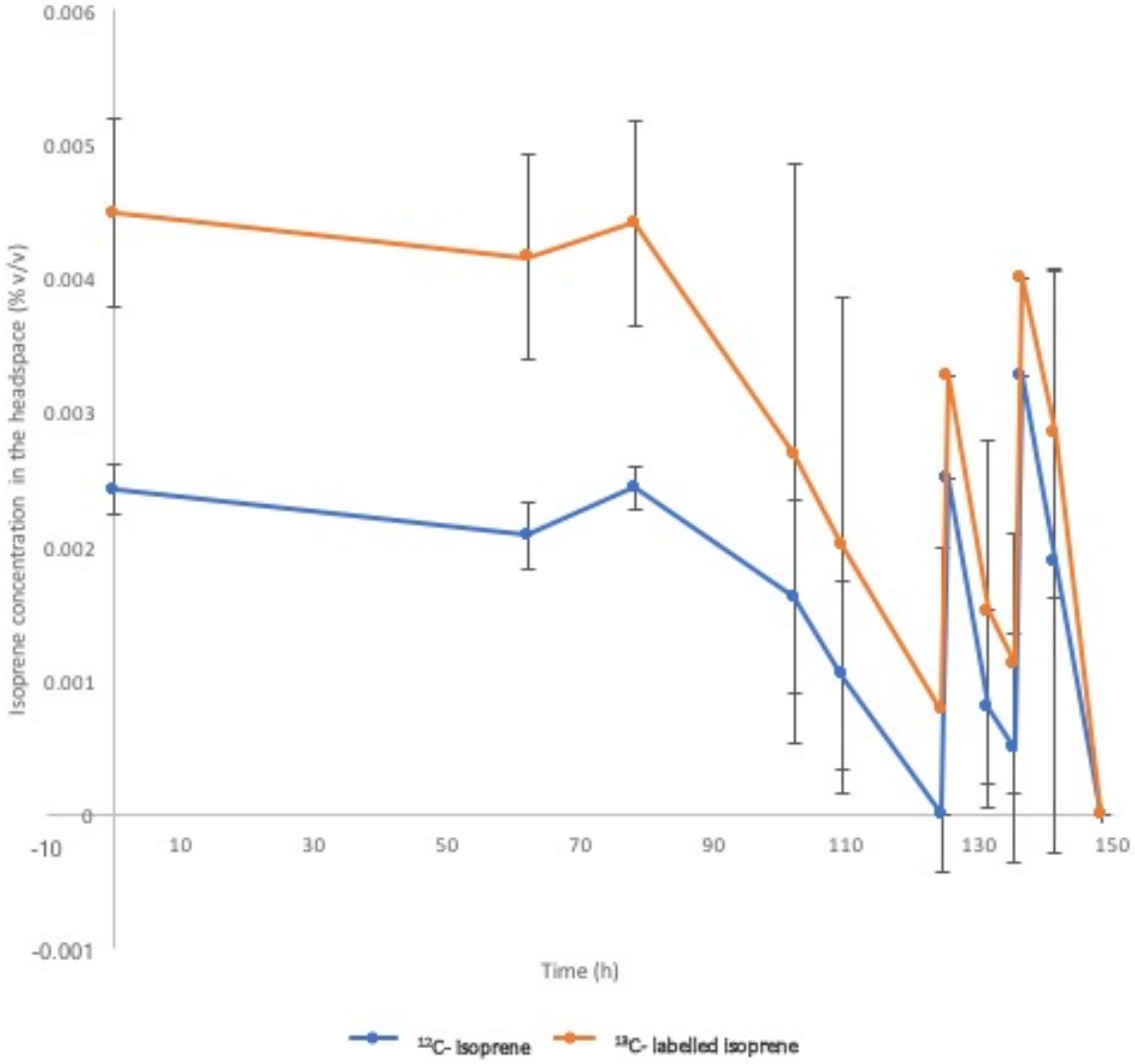



**Figure 5.3** Isoprene consumption over the first 148h of the OPL SIP incubations. Both $^{12}C$ isoprene and $^{13}C$ labelled isoprene incubations were done in triplicate, error bars show the standard deviation. The total assay took 312 h.

Replicate OPL microcosms showed variation in the consumption of isoprene. This result was observed with other leaf washing experiments (**Chapter 3**) and may reflect a difference in abundance of isoprene-degraders throughout the phytoplane of isoprene-emitting trees. Each replicate was monitored individually for consumption of isoprene to follow the level of incorporation of the $^{13}C$-label during the enrichment process. **Supplementary Table 1** shows the harvesting time for each replicate during the OPL-SIP experiment.

The initial consumption rate for OPL washings was calculated (**Table 5.4**). The consumption rates between $^{12}C$- and $^{13}C$- labelled isoprene consumption is similar in oil palm leaves. The small variation observed is therefore not due to the type of isoprene ($^{12}C$- or $^{13}C$-)

supplemented, it is probably due to variation in the abundance/types of isoprene-degraders present in each replicate.

**Table 5-3** Estimated consumption rate (%) of isoprene for oil palm leaf (OPL) samples during the SIP experiment. This is the initial rate of consumption for the first spike of 25 ppmv isoprene (consumed in approximately 130 hours). The average standard deviation of the samples was 0.00071 μmol $h^{-1}$ $g^{-1}$ for $^{12}$C-isoprene and 0.00010 μmol $h^{-1}$ $g^{-1}$ for $^{13}$C-isoprene.

| Sampling site | Isoprene type | Consumption rate (μmol $h^{-1}$ $g^{-1}$ of leaves washed) |
|---|---|---|
| OPL | $^{12}$C | 0.0044 |
| OPL | $^{13}$C | 0.0051 |

### *5.2.1.3 $^{12}$C- and $^{13}$C- DNA isolation from DNA-SIP experiments*

DNA was extracted from all soil samples replicates and from both treatments at each time point. For the high-biomass soil SIP, 5 μg of DNA from each replicate and each treatment was added individually to ultra-centrifuge tubes and subjected to CsCl isopycnic ultra-centrifugation. Up to 1 μg of DNA was added for the low biomass leaf SIP. Total DNA from each sample was fractionated, the density was calculated based on the refractive index and the DNA concentration was measured (**section 2.15.1**). **Figures 5.4** show the results when plotting the DNA concentration of each fraction against density in samples taken after 6 days of enrichment ($T_1$). **Supplementary Figure 6** and **Supplementary Figure 7** show results from samples taken after 7 and 8 days of enrichment, respectively.

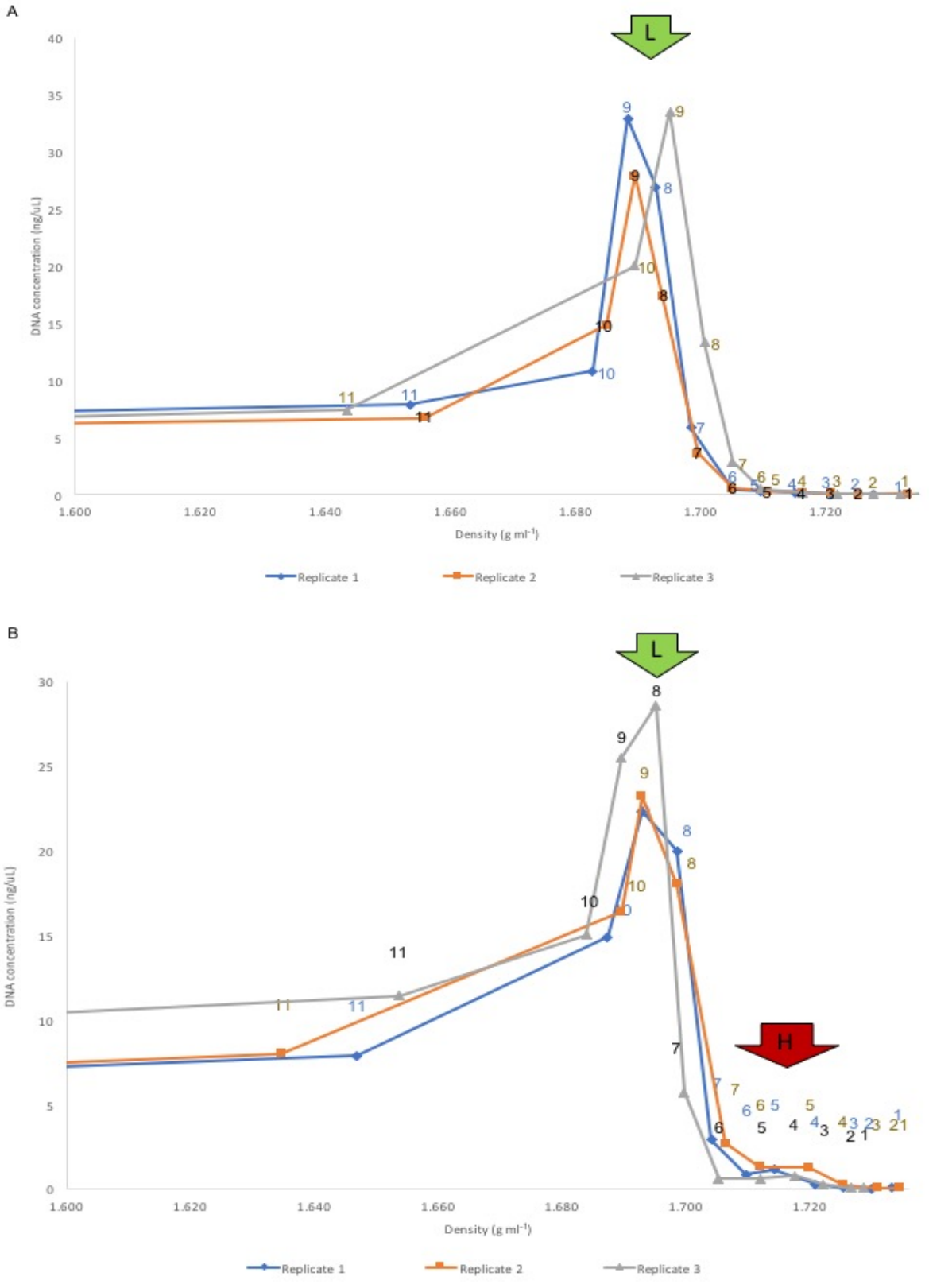



**Figure 5.4** Fractionation curves (density vs. DNA concentration) for DNA obtained with SIP enrichments with OPS after 6 days ($T_1$) of enrichment with A) $^{12}$C-isoprene and B) $^{13}$C-labelled isoprene. Fraction numbers presented on the graph.

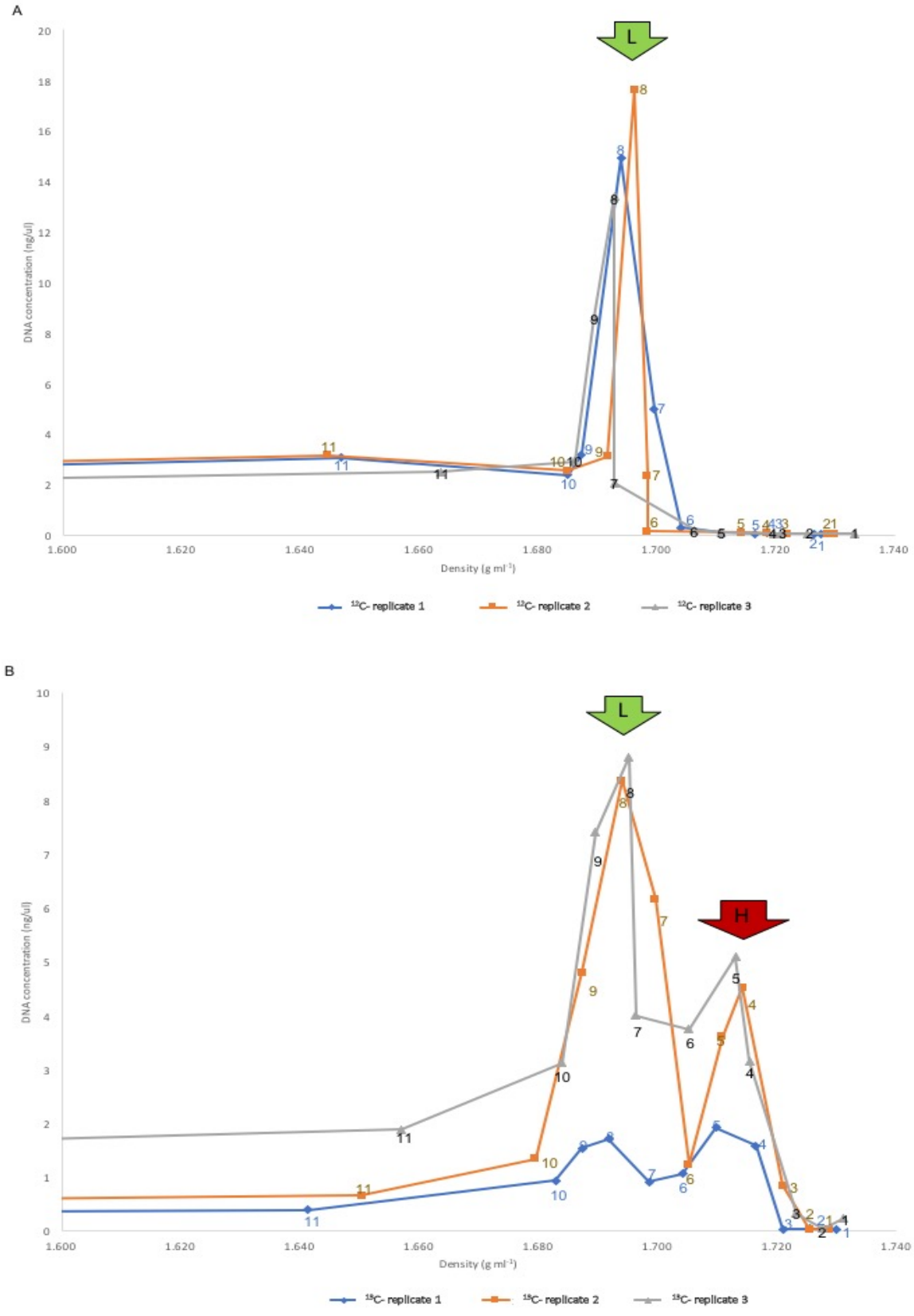



**Figure 5.5** Fractionation curves (density vs. DNA concentration) for DNA obtained with SIP enrichments with OPL washings after ~10 days ($T_1$) of enrichment with A) $^{12}$C-isoprene and B) $^{13}$C-labelled isoprene. Fraction numbers presented on the graph.

Incubation time and monitoring the incorporation of each replicate in both types of isoprene was a critical step of the enrichment process. DNA was extracted from leaf washings at $T_0$ and at three incorporation time-points (refer to **Table 5.1**). For this low-biomass SIP experiment, up to 1 μg of DNA was added to ultra-centrifuge tubes and subjected to CsCl isopycnic ultra-centrifugation to separate heavy and light DNA. Fractionation results, DNA concentrations and densities of each fraction, after 10 days of enrichment ($T_1$), are shown in **Figures 5.5** (refer to **Supplementary Figure 8** and **Supplementary Figure 9** for the last two time-points).

Incorporation of the labelled $^{13}C$- carbon into microorganisms that metabolised the $^{13}C$-labelled isoprene was efficient in both oil palm enrichments. In the case of OPS, 6 days of incubation ($T_1$), or 31.2 μmol $g^{-1}$ of incorporated $^{13}C$ labelled isoprene, was enough to produce a detectable peak of heavy DNA during SIP incubations (**Figure 5.4B**). In OPL samples, 22.3 - 33.9% of the total DNA recovered was $^{13}C$- heavy DNA after 10 days of incubation ($T_1$). The amount of OPL-SIP heavy DNA, ($^{13}C$- labelled peak, **Figure 5.5B**), was greater than that obtained in the OPS-SIP experiment.

The incorporation of $^{13}C$ label from isoprene during SIP enrichment experiments depended on the origin and characteristics of the oil palm tree sample. The first time-point was chosen after the incorporation of approximately 1 ml of isoprene gas; it took soil 6 days to incorporate set amount of isoprene, and roughly 10 days for all replicates in the leaf washing incubations. Both sets of microcosms were supplied with the same amount of $^{13}C$ labelled isoprene, yet different quantities of heavy DNA were recovered (see **Figures 5.4** and **5.5**).

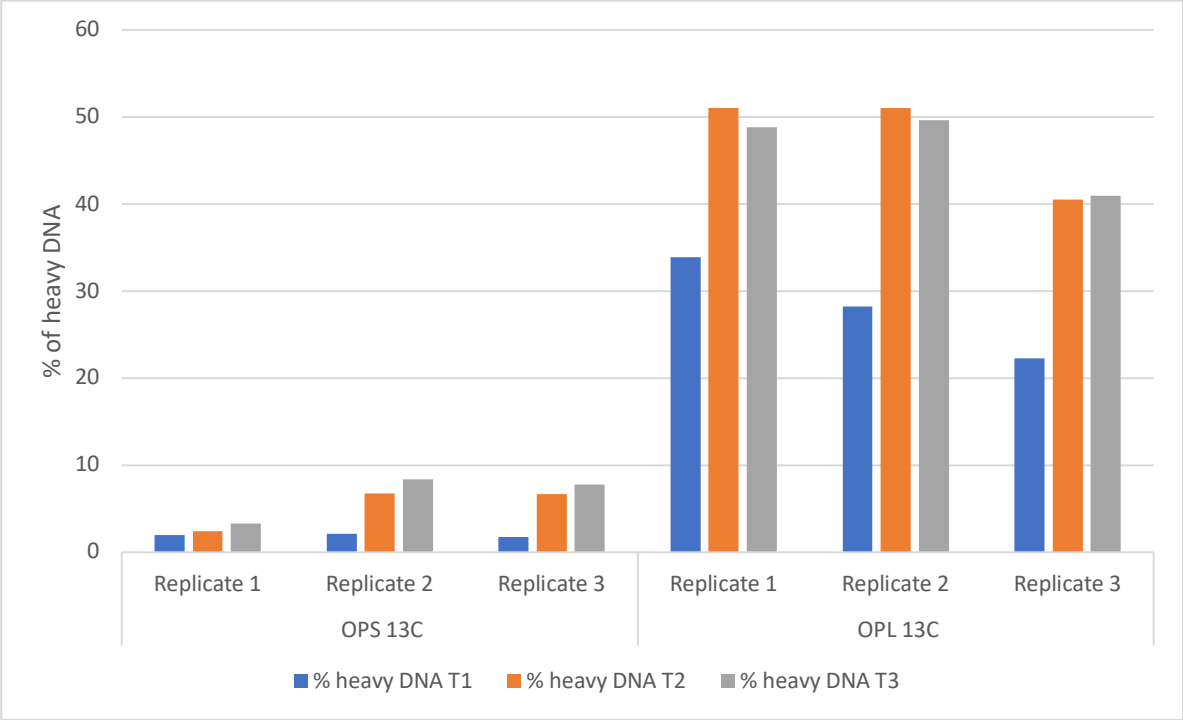



**Figure 5.6** Percentage of DNA contained in the heavy $^{13}$C-labelled fractions from $^{13}$C-labelled isoprene recovered from OPS and OPL SIP incubations at the three time-points.

As anticipated, the percentage of heavy DNA recovered from $^{12}$C-isoprene enrichments of soil and leaves was less than 1 % (refer to **Supplementary Figures 7** and **8**). **Figure 5.6** shows the percentage of heavy DNA recovered throughout both sets of SIP experiments.

### *5.2.1.4 Analysis of changes in bacteria diversity of the oil palm soil (OPS) and oil palm leaf (OPL) SIP incubations*

This was carried out using 16S rRNA gene profiling using DGGE (see **section 2.15.4**). 16S rRNA gene profiles for all fractions were run on an 8% polyacrylamide gel with a linear gradient of 30 to 70% denaturant concentration for 16 hours at 75 V in a DCode™ Universal Mutation Detection System by Bio-Rad. **Figure 5.7** shows the 16S rRNA gene profile of DNA from OPS-SIP experiments for replicate 2 after 8 days of enrichment ($T_3$). **Figure 5.8** shows the 16S rRNA gene profile of DNA OPL-SIP incubations replicate 2 after 14 days of enrichment ($T_3$).

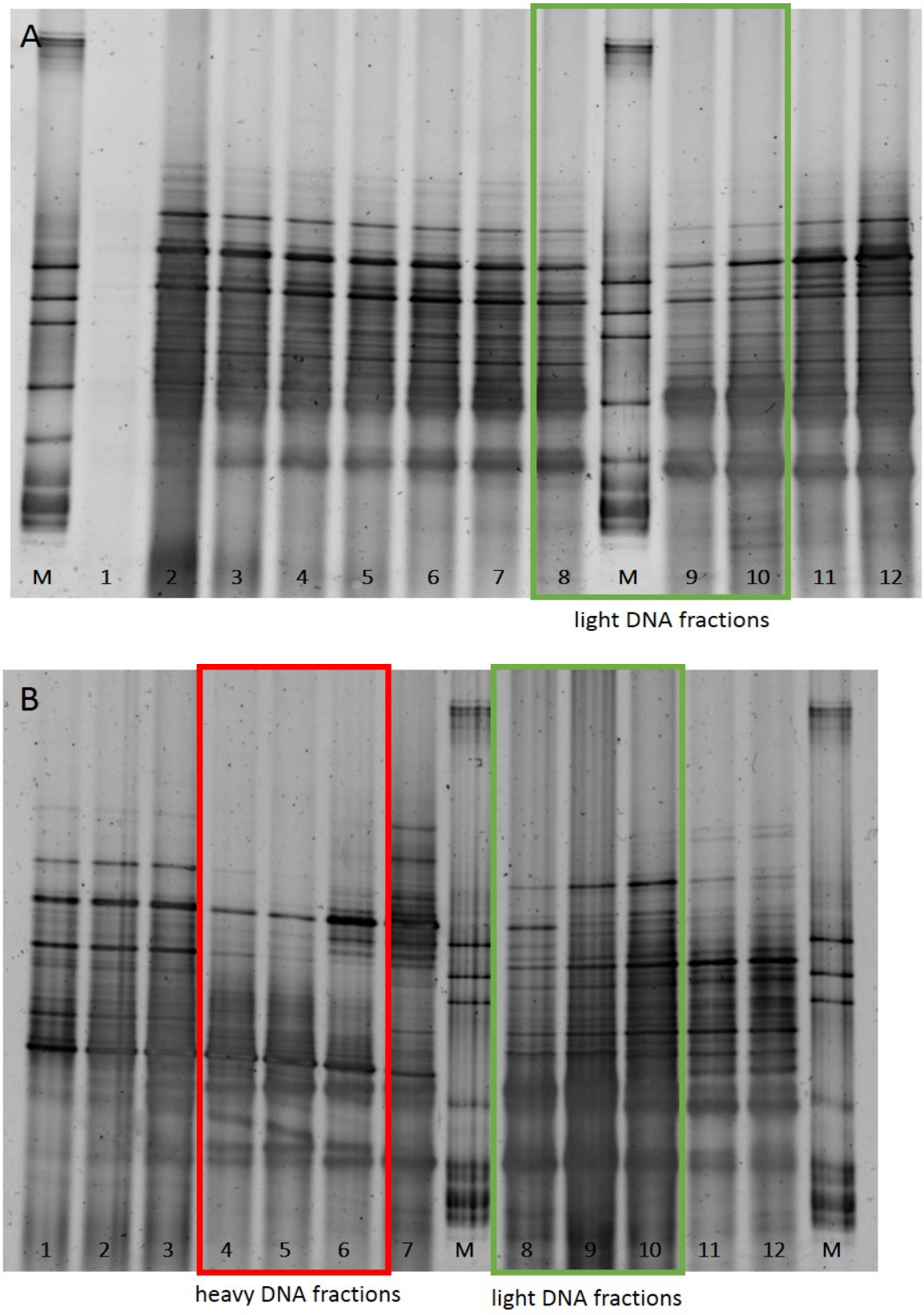



**Figure 5.7** DGGE 16S rRNA gene profile for all fractions in OPS replicate 2 incubated with A) $^{12}C$ isoprene and B) $^{13}C$-labelled isoprene for 8 days. Each fraction of the DNA centrifugation from 1-12 (heavy DNA to light DNA) is shown on the bottom. M: DGGE marker ladder

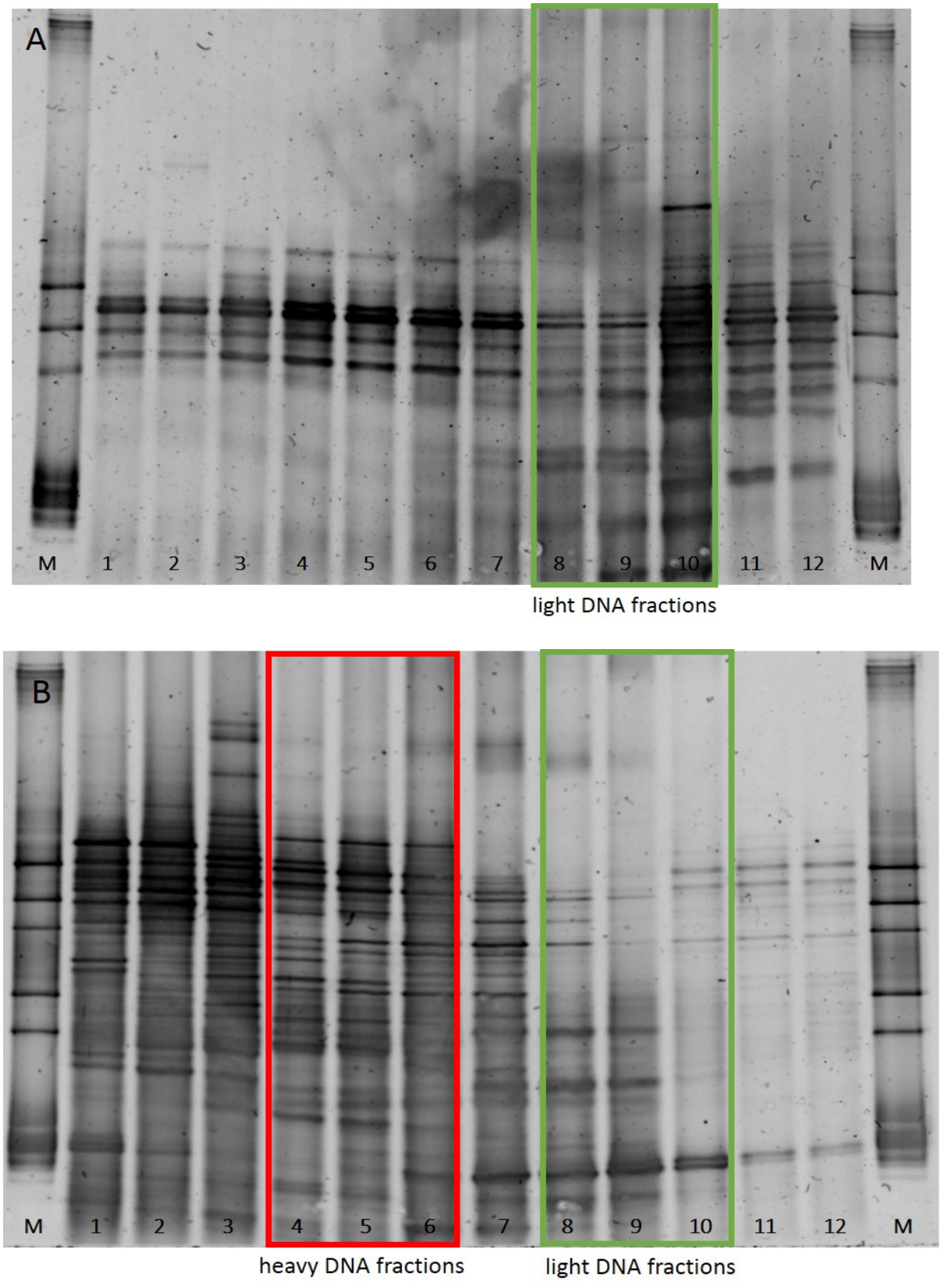



**Figure 5.8** DGGE 16S rRNA gene profile for all fractions in OPL replicate 2 incubated with A) $^{12}C$ isoprene and B) $^{13}C$-labelled isoprene for 14 days. Each fraction of the DNA centrifugation from 1-12 (heavy DNA to light DNA) is shown on the bottom. M: DGGE marker ladder

To compare the profiles for the 16S rRNA gene PCR products in heavy and light fractions, DNA samples were pooled (consistent with fractionation curves) and run side by side for $^{12}C$- and $^{13}C$- isoprene enrichments (**Figure 5.9** for OPS-SIP, **Figure 5.10** for OPL-SIP). OPS and OPL pooled light DNA and pooled heavy DNA for $^{12}C$ isoprene enrichment (replicate 2) showed similar 16S rRNA gene profiles.

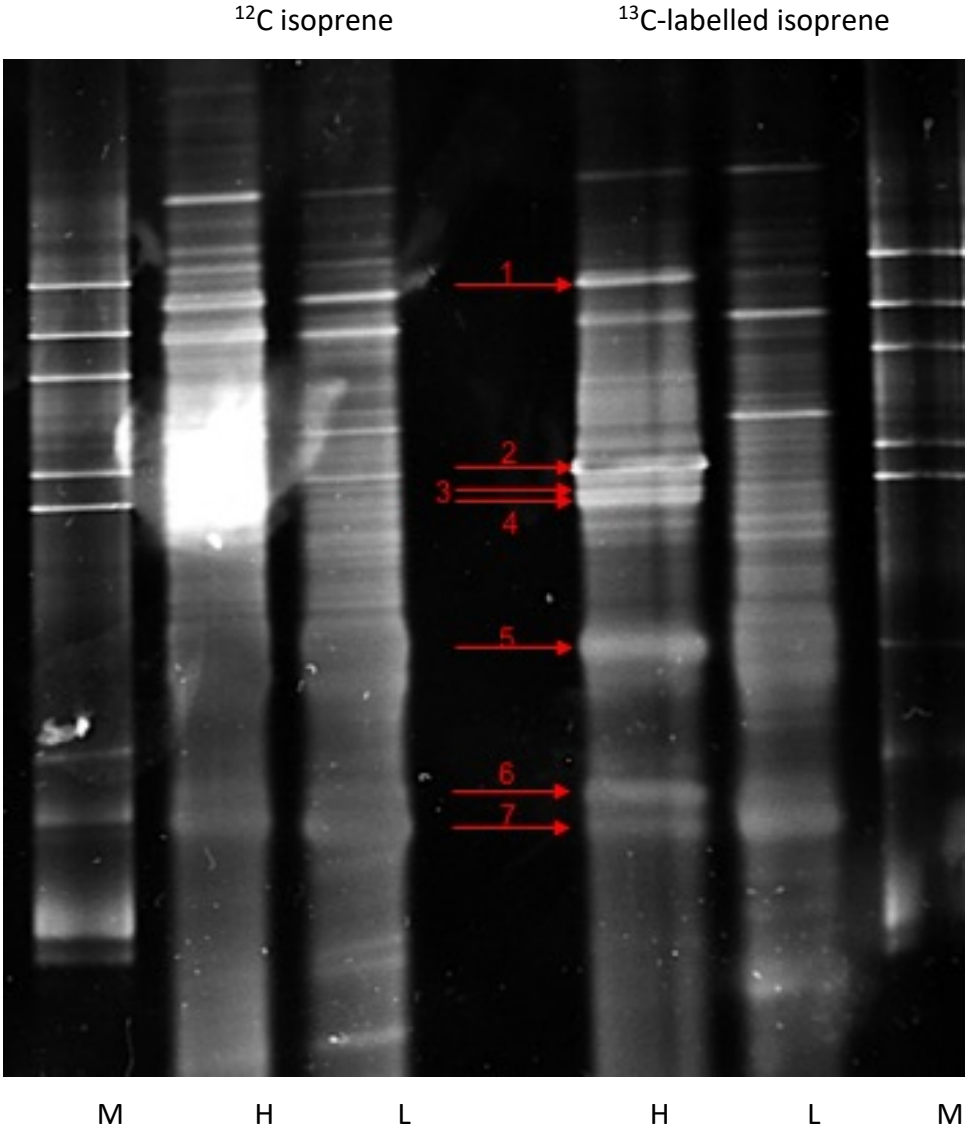



**Figure 5.9** DGGE 16S rRNA gene profile for pooled heavy (H) and light (L) fractions from oil palm soil (OPS) replicate 2 incubated with $^{12}C$ isoprene and $^{13}C$-labelled isoprene for time-point 3 ($T_3$, see **Table 5-1**). Pooled heavy (H) DNA and pooled light (L) DNA is shown on the bottom. Numbered arrows show enriched bands in the $^{13}C$-labelled treatment. M: DGGE marker ladder

Clear variations between the 16S rRNA gene profiles of the heavy and light DNA samples were observed in $^{13}C$-labelled isoprene enrichments. Twelve enriched bands were identified (red arrows) as representing potential isoprene-degraders present at the end of

the enrichment process. Efforts to identify the key active bacteria during OP-SIP incubations were important to understand bacteria dynamics during isoprene experiments (**section 5.3.5**).

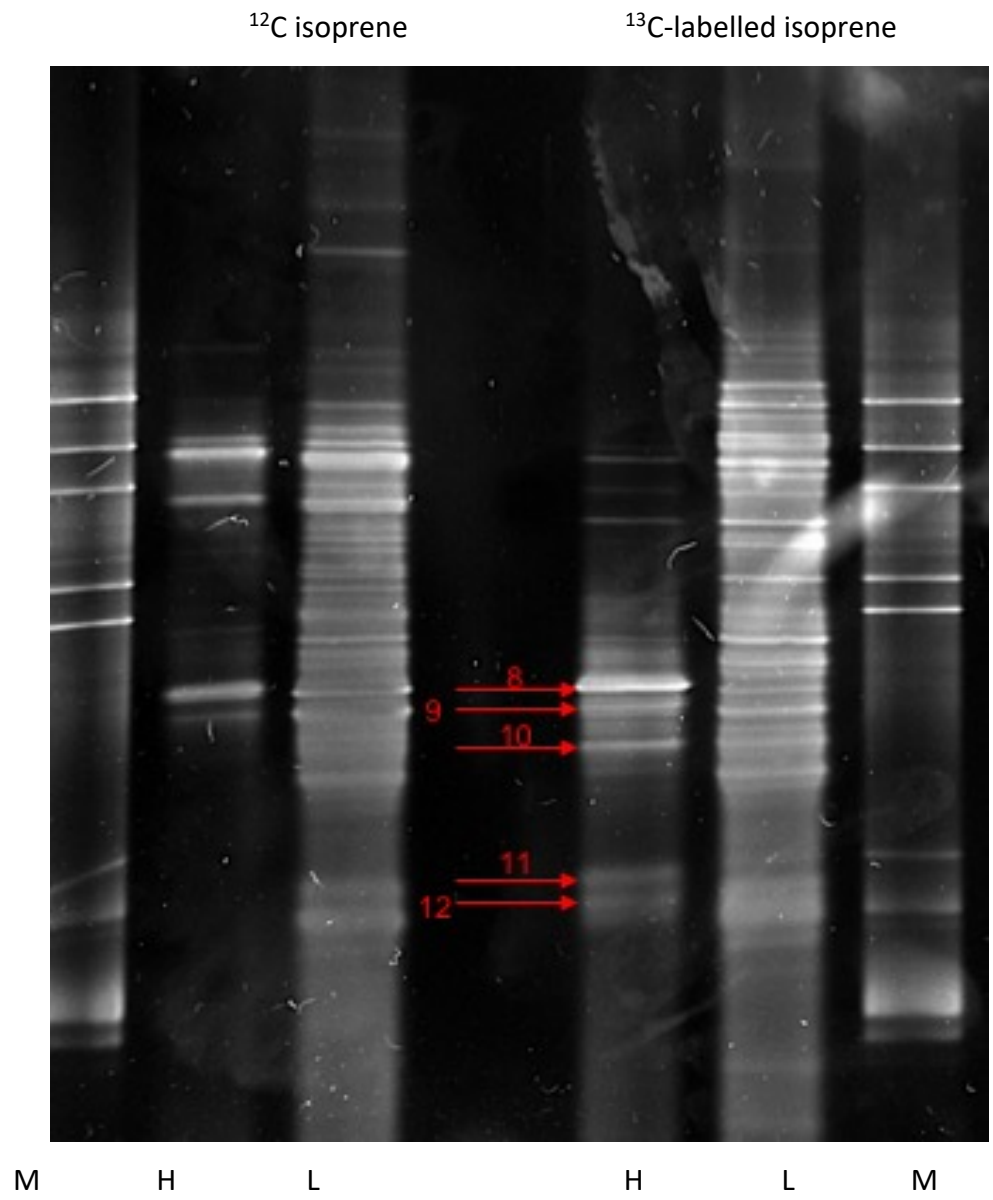



**Figure 5.10** DGGE 16S rRNA gene profile for pooled heavy (H) and light (L) fractions from oil palm leaf washings (OPL) replicate 2 incubated with $^{12}$C isoprene and $^{13}$C-labelled isoprene isoprene for time-point 3 ($T_3$, see **Table 5-1**). Pooled heavy (H) DNA and pooled light (L) DNA is shown on the bottom. Numbered arrows show enriched bands in the $^{13}$C-labelled treatment. M: DGGE marker ladder

#### *5.2.1.5 16S rRNA gene clone libraries from pooled heavy DNA fractions arising from Oil palm soil (OPS) and oil palm leaf (OPL) SIP experiments*

A 16S rRNA gene clone library from OPS-SIP and OPL-SIP pooled heavy DNA (approximately 2.0 ml isoprene incorporated, replicate 2, fraction 5) was constructed to identify the enriched isoprene degraders. Full length 16S rRNA gene PCR products (27f and 1492r primer pair, **Table 2.1**) from the pooled heavy DNA were cloned into pGEM-T Easy vector (Promega). Clones were selected by using a nested PCR (amplified with 341F-GC and 518R primer pair, **Table 2.1**) and observing migrations pattern on a DGGE gel. Key 16S rRNA genes identified in this way were sequenced to elucidate the identity of enriched bacteria during the SIP incubations (data not shown, details in **section 2.17**).

Enriched bands from **Figures 5.9** and **Figure 5.10** were examined as putative sequences from isoprene-degrading bacteria. **Table 5.4** presents the Blastn identity of 16S rRNA genes representative of bacteria which appeared to be enriched during the soil and leaf SIP incubations.

**Table 5-4** Blastn identity of full 16S rRNA genes (27f to 1492r) representative of bacteria which appeared to be enriched in the heavy DNA from oil palm soil (OPS) and oil palm leaf (OPL) washings at the end of the SIP enrichment process.

| DGGE band number | Sample origin | Blastn result* |
|---|---|---|
| 1 | Oil palm soil | n.i. |
| 2 | Oil palm soil | *Aquincola* sp.ID0723 (98%, 0.0) |
| 3 | Oil palm soil | n.i. |
| 4 | Oil palm soil | uncultured Sphingomonadaceae bacterium clone T302H2 (99%, 0.0) |
| 5 | Oil palm soil | *Aquincola* sp.ID0723 (98%, 0.0) |
| 6 | Oil palm soil | uncultured Sphingomonadaceae bacterium clone T302H2 (99%, 0.0) |
| 7 | Oil palm soil | n.i. |
| 8 | Oil palm leaves | *Aquincola tertiaricarbonis* strain L10 |
| 9 | Oil palm leaves | *Methylobacterium populi* strain ICGV-1 (99%, 0.0) |
| 10 | Oil palm leaves | *Deinicoccus* sp. JJ521 (99%, 0.0) |
| 11 | Oil palm leaves | *Gordonia polyisoprenivorans* strain W8130 (99%, 0.0) |
| 12 | Oil palm leaves | *Methylobacterium populi* strain ICGV-1 (99%, 0.0) |

*DGGE band number corresponds to arrow numbers in **Figure 5.9** and **Figure 5.10**. Full 16S rRNA gene PCR consensus sequence (1300 bp) was used for database comparison. Sequence identity and E-value in parenthesis. n.i.: not identified.

#### *5.2.1.6 16S rRNA gene amplicon sequencing results for oil palm soil (OPS) and oil palm leaf (OPL) washings SIP experiments*

Native DNA ($T_0$), pooled heavy DNA from replicate 2 (at each time point), and heavy DNA from replicate 1 and 3 for time-point 3 from both SIP experiments was analysed by 16S rRNA gene amplicon sequencing (**Table 5.5**, **section 2.16.1**).

**Table 5-5** Analysis of DNA-SIP samples from OPS and OPL enrichment by 16S rRNA gene amplicon sequencing

| Sample | Time-point | Description |
|---|---|---|
| 1 | $T_0$ | Unfractionated DNA |
| 2 | $T_1$ | OPS $^{13}$C heavy DNA replicate 2 |
| 3 | $T_2$ | OPS $^{13}$C heavy DNA replicate 2 |
| 4 | $T_3$ | OPS $^{13}$C heavy DNA replicate 2 |
| 5 | $T_3$ | OPS $^{13}$C heavy DNA replicate 1 |
| 6 | $T_3$ | OPS $^{13}$C heavy DNA replicate 3 |
| 7 | $T_1$ | OPL $^{13}$C heavy DNA replicate 2 |
| 8 | $T_2$ | OPL $^{13}$C heavy DNA replicate 2 |
| 9 | $T_3$ | OPL $^{13}$C heavy DNA replicate 2 |
| 10 | $T_3$ | OPL $^{13}$C heavy DNA replicate 1 |
| 11 | $T_3$ | OPL $^{13}$C heavy DNA replicate 3 |

*Oil Palm Soil -SIP*

The bacteria community profiles in samples before the SIP enrichments ($T_0$) and throughout the enrichment experiment ($T_1$, $T_2$ and $T_3$) for replicate 2 are shown at the genus level (**Figure 5.11**). The bacteria present in the sample changed significantly through the SIP enrichments with isoprene. This was confirmed with the appearance of bacteria that were not detected in the native soil ($T_0$). The $^{13}$C-labelled bacteria (*Rhodococcus*, *Aquabacterium*, *Gordonia* and *Aquincola*) were present and abundant throughout the SIP enrichment. Some bacteria were observed only after 7 days of enrichment

(*Saccharibacter*), and others do not persist during the enrichment (*Flavisolibacter* and *Flavobacterium*). These results suggest that *Rhodococcus*, *Aquabacterium*, *Gordonia* and *Aquincola* are the putative isoprene degraders in oil palm soil recovered at Kew Gardens.

**Figure 5.11** shows the changes in the abundance of 16S rRNA genes from key genera in replicate 2 during the SIP incubations. The most abundant genera during the OPS-SIP were *Aquabacterium* and *Rhodococcus.* At the species level, 16S rRNA genes from *Aquabacterium* spp and *Saccharibacter* spp. increased in abundance during isoprene-SIP enrichments, while the abundance of 16S rRNA genes from *Rhodococcus wratislaviensis*, *Rhodococcus koreensis* and *Rhodococcus opacus* decreased. 16S rRNA genes from two species of *Gordonia* were also detected (at low abundance) throughout the SIP incubation.

### *Oil Palm Leaf -SIP*

The bacterial community profile in the $T_0$ oil palm leaf washings was substantially different from that of $T_0$ oil palm soil. Plastid 16S rRNA gene sequences from eukaryotic organisms were eliminated from native ($T_0$) leaf washings data. 16S rRNA gene sequences from *Sphingomonas*, *Methylobacterium*, *Bacillus*, *Flavobacterium*, *Sphingobacterium*, *Plectonema* and *Thermomonas* were the most abundant sequences present prior to enrichment (**Figure 5.12**).

The relative abundance of genera labelled in replicate 2 was followed throughout the incubation with $^{13}$C-isoprene. Enrichment of verified some well-known isoprene-degraders i.e. *Gordonia, Rhodococcus*, and other bacteria that might be able to oxidize isoprene or its downstream metabolic products. In oil palm leaves *Methylobacterium* and *Sphingomonas* were present in the native ($T_0$) soil (1-2%) and later increased in abundance during the enrichment process, becoming labelled with $^{13}$Carbon. The key genera in the oil palm leaf enrichment for replicate 2 were *Aquincola*, *Sphingomonas*, *Methylobacterium*, *Aquabacterium*, *Gordonia* and *Rhodococcus*.

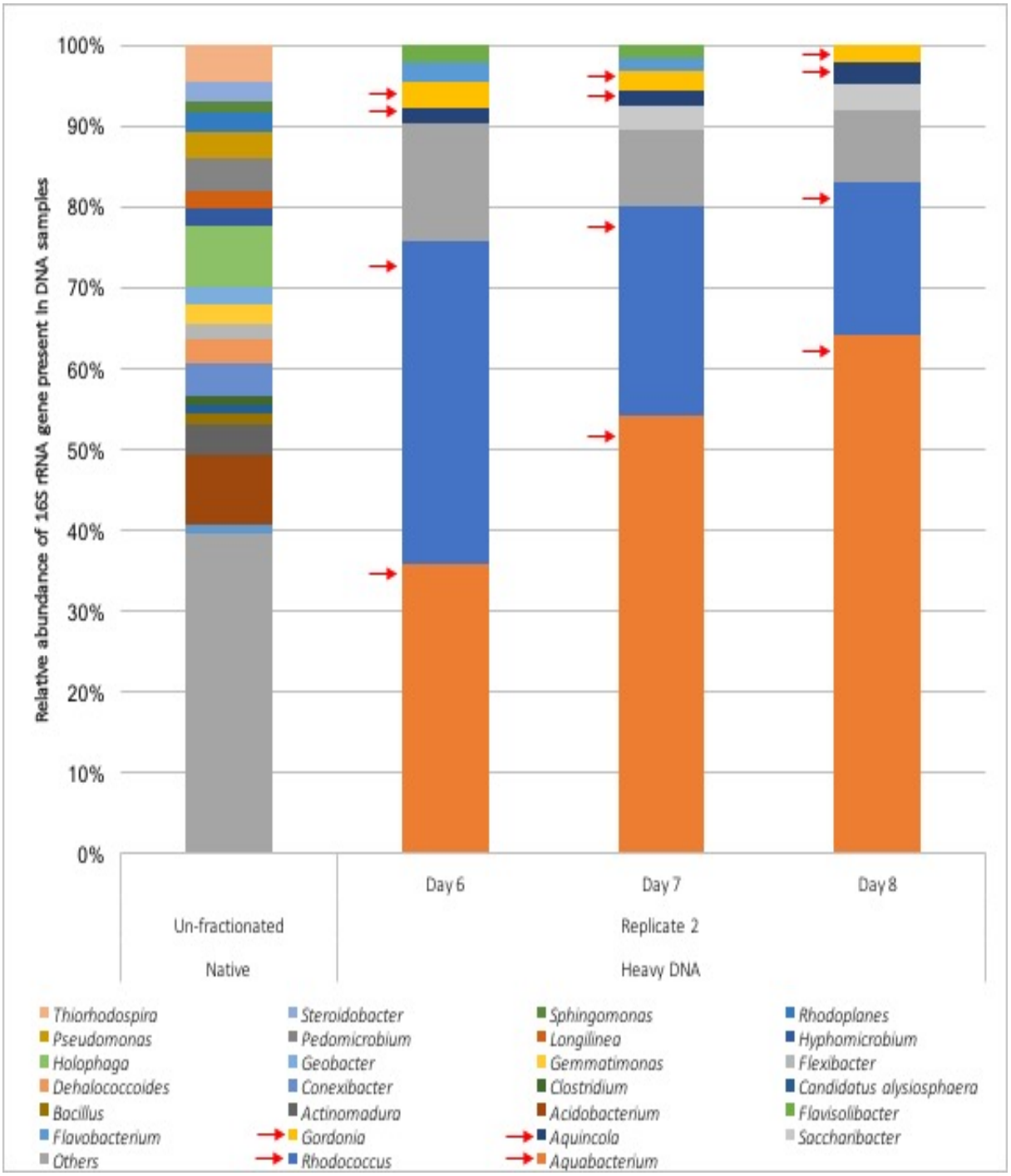



**Figure 5.11** 16S rRNA gene relative abundance of genera present from oil palm soil (OPS) from $T_0$ (un-fractionated) and enriched $^{13}$C-isoprene pooled heavy DNA fractions (replicate 2, at the three time-points). Genera present at less than 1% relative abundance were group as "others". Arrows indicate putative isoprene-degraders.

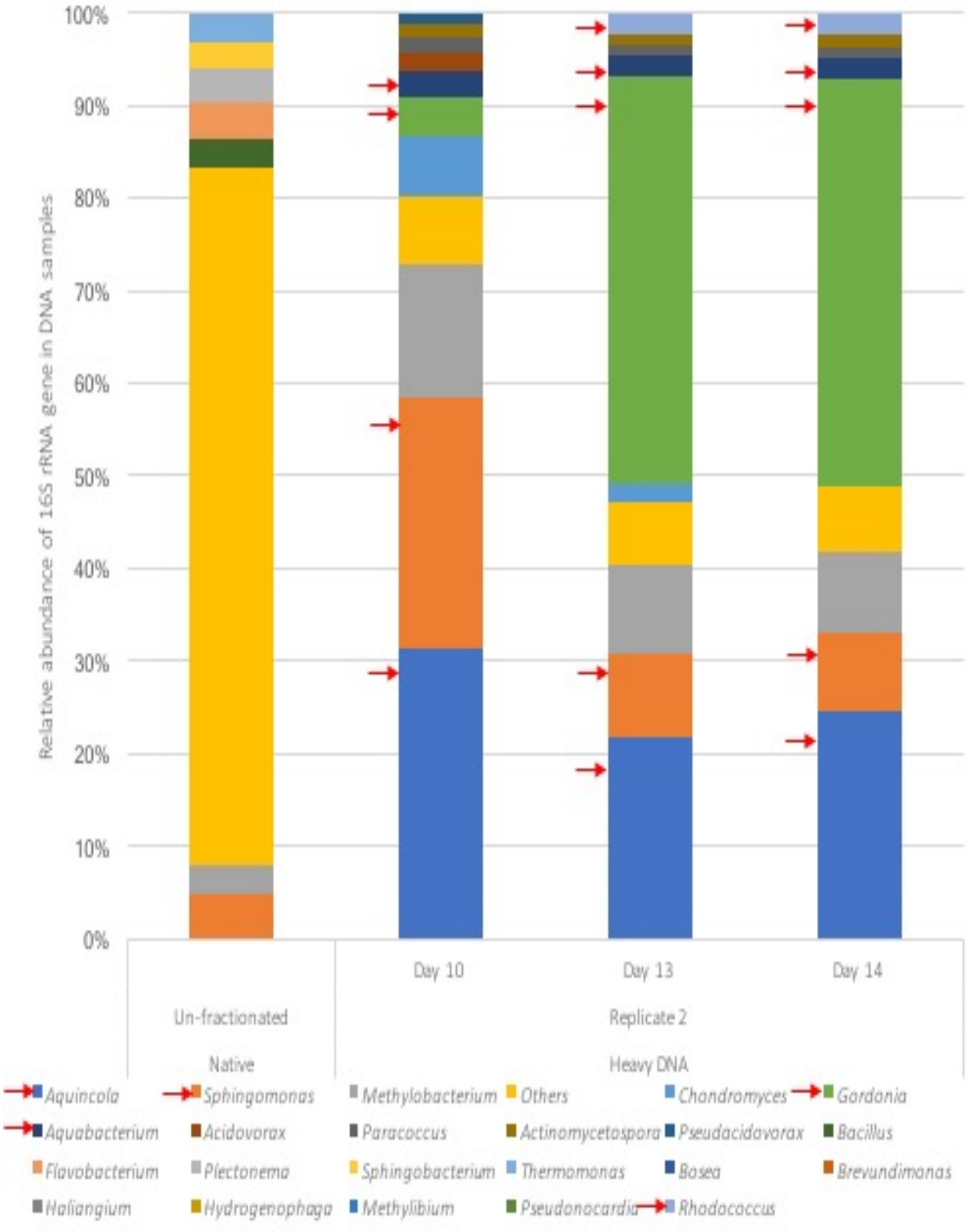



**Figure 5.12** 16S rRNA gene relative abundance of genera present from oil palm leaf (OPL) washings from native (un-fractionated) and enriched $^{13}$C-isoprene pooled heavy DNA fractions (replicate 2, at the three time-points). Genera present at less than 1% relative abundance were group as "others". Arrows indicate putative isoprene-degraders.

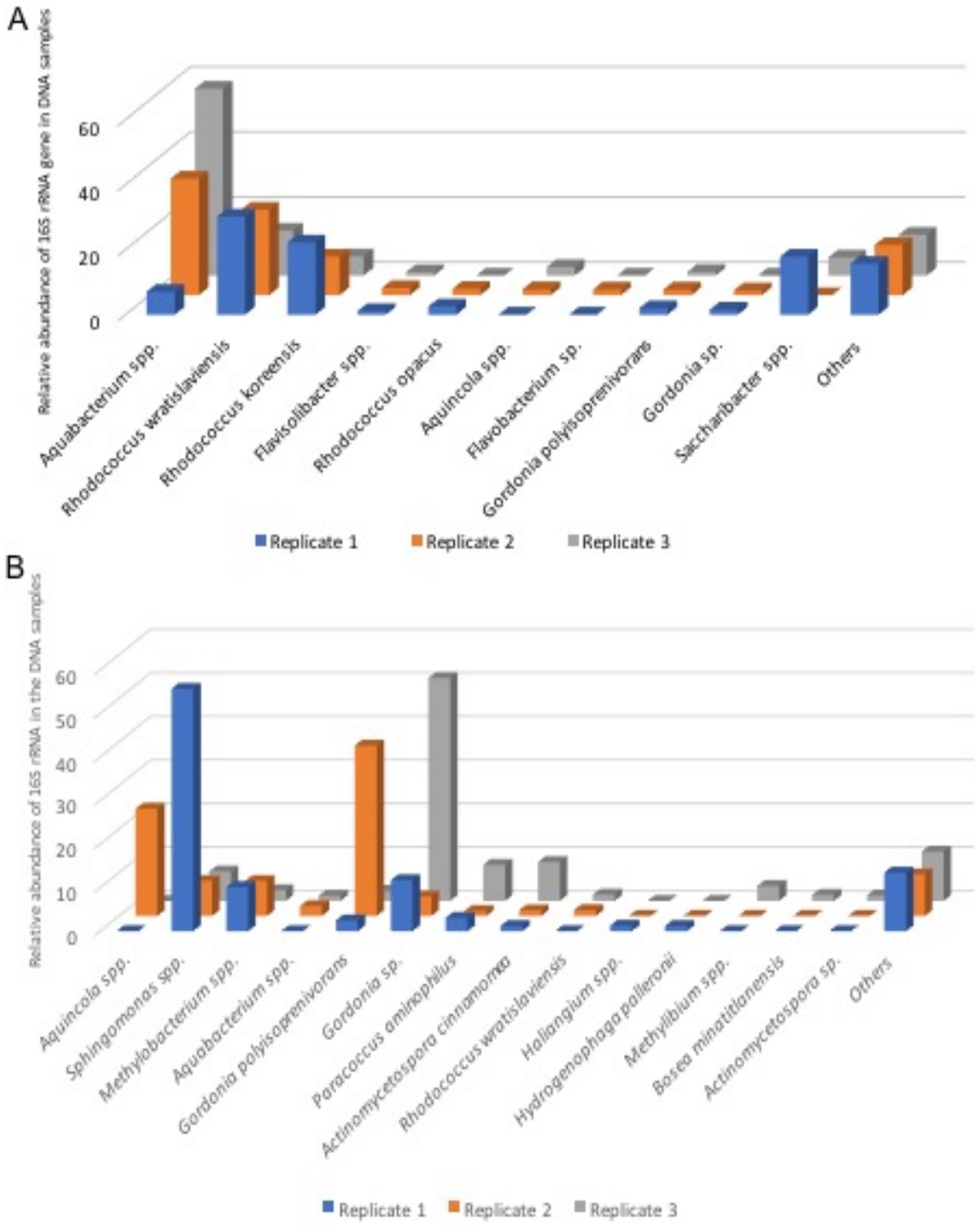



**Figure 5.13** 16S rRNA gene relative abundance of top species from (A) oil palm soil (OPS) and (B) oil palm leaf (OPL) washings from $^{13}$C-isoprene enriched heavy DNA after 14 days of enrichment. Species present at less than 1% relative abundance were group as "others"

**Figure 5.13** compares the abundance of the top species from the three replicates at the end of each OP-SIP experiments. These results show how the enrichment process of active isoprene-degraders differs between similar samples of the same origin. In **Figure 5.13A** for OPS the most abundant bacteria were *Aquabacterium*, *Rhodococcus wrastiviansis*,

*Rhodococcus koreensis*, and *Saccharibacter*. The abundance of each species changes between replicates but the pattern of suggested isoprene degraders from oil palm soil is maintained. This is observed by the repetition of species names between replicates. The presence of other genera in low abundance such as *Gordonia*, *Aquincola* and *Flavisolibacter* between replicates might suggest their potential as isoprene-degraders or cross-feeders.

Previous leaf washing enrichments suggested variation in the diversity of isoprene-degrading bacteria on the leaves of isoprene-emitting trees, reflected and the varying consumption rates between samples (**Chapter 3**). **Figure 5.13B** shows *Aquincola* spp. and *Gordonia polyisoprenivorans* were important in replicate 2, *Sphingomonas* spp. and *Gordonia* sp. for replicate 1, and *Gordonia* spp. with *Sphingomonas* spp., *Paracoccus aminophilus* and *Actinomycetospora cinnamomea* were enriched in replicate 3. These results confirm a poorer replication in communities recovered from leaf washings compared to soil samples. Furthermore, a very interesting result is the presence of *Aquincola*, *Sphingomonas* and *Aquabacterium* in the leaf enrichment, those genera were also enriched in oil palm soil samples.

## 5.3 Discussion

Isoprene consumption by the oil palm soil and leaf microcosms reflected similar data patterns to those observed in previous soil assays, the WS-SIP experiment and different leaf washing enrichments (refer to **Chapter 3** and **Chapter 4**). Replicate microcosms from OPS consumed 2.0 ml of isoprene gas in 8 days of enrichment, with low variation in consumption time and rates between samples (**Chapter 4**). OPL washings consumed the first spike of isoprene in approx. 130 h of incubation and differences in the consumption of isoprene between replicates was evident (**Chapter 3**). Previous enrichment assays were essential in the estimation of some factors that affected the OP-SIP experiments. Data proved the value of the design and execution of isoprene enrichment assays with $^{13}$C-labelled isoprene.

In OPS-SIP, the amount of isoprene supplemented, the estimated incorporation of $^{13}$C label, and the percentage of heavy DNA recovered from OPS-SIP at each harvesting point was

also comparable to WS-SIP data. However, the amount of heavy DNA incorporated into biomass with OPS samples was lower than data observed for WS (refer to fractionation curves in **Chapter 4**). Characteristics of the greenhouse soil such as the present isoprene-degrading bacteria present, the use of fertilizers, the surrounding vegetation, and the abundant persistence fungi in samples due to continuous irrigation, etc., may have been important factors in the incorporation into DNA. Soil enrichment experiments from different isoprene-emitting trees can be used as models for other isoprene-emitting trees, but this study has seen that each one is subjected to various particularities that might affect isoprene consumption and incorporation. In this chapter, the recovery of samples from a greenhouse environment gave the oil palm holistic characteristics of its own.

OPL samples (washings and swabs) from the greenhouse had been tested previously for isoprene consumption. Individual monitoring of replicates was the first lesson from leaf washing incubations (**Chapter 3**). The OPL-SIP isolation of heavy and light DNA demonstrated the efficient incorporation of $^{13}$C-carbon from labelled isoprene. The three time-points also contributed enough heavy DNA for this study. An additional time-point, with incorporation of 12-13 $\mu$mol $^{13}$C $g^{-1}$, would be desirable to study the first bacteria that incorporate the label for future work.

This chapter compares the results for a high-biomass SIP (soil) and low-biomass SIP (leaf washings) from an OP tree. OP soil and leaf SIP samples were harvested when both treatments consumed roughly the same volume of $^{13}$C labelled isoprene. For the first time-point, OPS (0.1 g soil $ml^{-1}$) needed 6 days to consume 0.9 ml of isoprene, while OPL (0.675g washed leaves $ml^{-1}$) required 10 days. The characteristics of the leaves, i.e. prevalence of fungi, may have significantly altered the natural conditions of the enrichment results, nonetheless show that soil is quicker at using the isoprene even after sampling larger leaf surface.

Isoprene consumption rate and time for each replicate from OPL-SIP varied during the enrichment process. The amount of leaves washed for the enrichment was 6.75 times higher than the amount of leaves washed in AL, WL and EL experiments (**Chapter 3, Figure 3.6**). Increasing the amount of leaves washed might have been important for OP experiments due to the origin and state of the samples but did not mitigate the variation

in bacteria present between OP samples. Increasing the amount of OP leaf biomass washed did, appropriately, provide a clearer representation of the diversity of the epiphytic bacterial isoprene degraders present on the oil palm tree phytoplane (**Figure 5.13B**).

Isoprene-degraders presence and dynamics were different on each OP leaf sample; results show differences throughout the phyllosphere microcosm (predicted with OP, AL and WL in **Chapter 3**). The size of samples, the canopy location, and tree location (weather conditions) may play an important role in these dynamics throughout the year (Lindow & Brandl, 2003). SIP experiments with high- or low-biomass samples enlisted isoprene degraders as present and active members of the phyllosphere and soil dynamics.

Analysis of 16S rRNA gene amplicon sequencing identified *Rhodococcus*, *Gordonia, Aquabacterium* and *Aquincola* were present in both soil and leaf enrichments. The first two genera are confirmed as a genuine isoprene-degrader due to the presence of the isoprene metabolic cluster (Crombie *et al*., 2018; Johnston *et al*., 2017; Khawand *et al*., 2016). The genera *Aquabacterium* and *Aquincola* are beta-proteobacteria and unclassified members of the Burkholderiales order have not been studied for their role in isoprene metabolism. Analysis of genomes available for organisms of these genera is necessary to conclude if they are active isoprene-degraders or cross-feeders.

Finally, in the OPL-SIP, analysis of a partial 16S rRNA gene identified the genera *Sphingomonas* and *Methylobacterium* were present in the native ($T_0$) leaf washings and were labelled during OPL-SIP. No other evidence was found for *Methylobacterium* as an isoprene-degrader. However, cultivation-dependent strategies allowed the isolation of three confirmed isoprene-degrading bacteria members of the Sphingomonadaceae family from OPL enrichments identified as *Sphingobacterium* sp. strain OPL3 and *Sphingopyxis* sp. OPL5 (**Chapter 3**). Cultivation-dependent methods have allowed the confirmation of members of the Sphingomonadaceae family as verified Gram-negative isoprene-degrading bacteria. This strategy was used to verify DNA-SIP data for isoprene-degrading bacteria of the Comamonadaceae family: *Variovorax* sp. strain WS9 isolated from Colney-WS (refer to **Chapter 4**).

Oil palm DNA-SIP experiments, coupled with other cultivation-dependent and cultivation-independent strategies, have shown to be valuable to studies of the microbial ecology of isoprene-degrading bacteria. Certain genera of isoprene-degraders were found throughout the phyllosphere and surrounding top-soil. The presence and abundance of certain isoprene-degrading bacteria may be species-specific, as seen for *Variovorax* with Colney-WS and *Sphingopyxis* and *Gordonia* for OP, while other genera as *Rhodococcus* might be found in a wide range of environments (soils and leaves). Further studies using samples from a naturally occurring oil palm tree, are necessary to better understand isoprene-degradation oil palms in their natural environment.

# Chapter 6 Use of the functional gene *isoA* to investigate the diversity of isoprene degraders

## 6.1 Introduction

The discovery of the first Gram-negative isoprene-degrading bacterium in the terrestrial environment has already expanded our knowledge of isoprene metabolism (**Chapter 3**). The willow soil DNA-SIP experiment demonstrated that bacteria of the Comamonadaceae family were $^{13}$C-labelled and therefore active utilizers of isoprene (**Chapter 4**). Furthermore, screening for *isoA* genes in the metagenome data from these DNA-SIP experiments also suggested the presence of the complete isoprene metabolic gene cluster in these bacteria.

Since the discovery of *Variovorax* sp. strain WS9 and *Variovorax* sp. strain WS11 (described in **Chapter 3)**, other novel isoprene degrading bacteria have been recovered from different environmental samples (refer to **Table 2.2**). The next step in the analysis of the selected novel isoprene-degrading bacteria was whole genome sequencing. Analysis of the isoprene metabolic gene clusters between different organisms was necessary to understand more about the evolution of this metabolic gene cluster. Finally, the design of new degenerate primers for *isoA,* the gene encoding the $\alpha$-subunit of the isoprene monooxygenase, was necessary to improve the screening capability for isoprene metabolic genes in the environment.

*General objective*

To compare isoprene degradation gene diversity (*isoA* gene encoding the $\alpha$-subunit of the isoprene monooxygenase and the isoprene metabolic gene cluster) in isoprene-degrading bacteria and in metagenomes.

*Specific objectives*

To compare the general genome features, metagenome data and specifically isoprene metabolic gene clusters for the novel isoprene-degrading bacteria isolated in this study.

To carry out phylogenetic comparisons for each polypeptide encoded by the isoprene metabolic gene cluster.

To design novel *isoA* gene primers to better assess the diversity of the $\alpha$-subunit of the isoprene monooxygenase in nature.

## 6.1 Results

### 6.2.1 Novel isoprene-degrading bacteria

$^{12}C$- and $^{13}C$- labelled isoprene DNA-SIP incubations, using 13-25ppmv isoprene, confirmed the presence and enrichment of active isoprene-degrading bacteria in the terrestrial environment (**Chapter 4** and **Chapter 5**). 21 different isoprene-degrading bacteria strains were isolated from a variety of enriched terrestrial samples (see **Table 6.1**). 76% of the isolates were Gram-positive, and were dominated by the genera *Rhodococcus* (14), and including *Nocardioides* (1) and *Gordonia* (1). The other isolates were Gram-negative bacteria identified as from the genera *Variovorax* (3), *Sphingopyxis* (1), *Sphingobacterium* (1).

*Variovorax* sp. strain WS9 was the first novel isolate recovered from willow tree soil in 2015. Targeted isolation, using the optimal growth conditions used for the previous *Variovorax* strain, was implemented. This resulted in the isolation of *Variovorax* sp. WS11 and *Nocardioides* sp. strain WS12 from another willow tree soil in 2016 (refer to **Table 2.3**

**and Table 3.6**). Finally, the enrichment of oil palm soil and leaves using targeted isolation and sub-cultivation isolation strategies (details in **section 2.11**), yielded *Gordonia*, *Sphingobacterium* and *Sphingopyxis* isolates in 2017.

Due to the novelty of some genera (no reports for evidence of them being isoprene-degraders) and data suggesting some interesting metabolic features of others, 11 strains were sent for whole genome sequencing. In early 2016, the genome sequence from *Variovorax* sp. strain WS9 was analysed and the data utilized in the design of a new pair of degenerate PCR primers targeting the *isoA* gene (see **section 6.1.2**, some data included in Carrion-Fonseca *et al*., 2018 unpublished).

### 6.1.2 Genome sequencing of novel isoprene-degrading bacteria

Novel taxa, and isolates which were phenotypically distinct from other isoprene-degrading bacterial representatives present in the lab, were sent for genome sequencing (details in **section 2.16**). **Table 6.2** shows the general information from genomes of novel isoprene-degrading isolates. Two willow soil metagenome bins (with high completeness and low contamination) were also included in the analysis due to the presence of the isoprene metabolic gene cluster in the RAST annotation data (Aziz et al., 2008). Briefly, six bins were obtained initially from the contig co-assembly and binning of DNA corresponding to the heavy fraction from 6 and 7 days of enrichment with isoprene (**Chapter 4**, details in **section 2.18**). All the bins were uploaded to the RAST server and screened for isoprene metabolic genes; only two bins showed evidence of one or more isoprene metabolic genes (bin 19 and bin 20). **Table 6.2** also shows the 16S rRNA gene identification using the complete 16S rRNA gene retrieved from each bin/genome used in this cluster analysis. RAST Seed Viewer tools confirmed *Rhodococcus* sp. strain WS1 and *Rhodococcus* sp. strain WS7 were the same isolate, recovered from the same willow tree soil, at different sampling times (Aziz et al., 2008).

**Table 6-1** List of isoprene-degrading bacteria isolated in this study.

| Isolate name | Date isolated | Sample origin | Isoprene-emitting tree[±] | Tree location (UK) | Genome sequenced |
|---|---|---|---|---|---|
| *Rhodococcus* sp. strain WS1 | Apr-2015 | Soil | *Salix alba L.* | Colney fields, Norwich | yes |
| *Rhodococcus* sp. strain WS2 | Apr-2015 | Soil | *Salix alba L.* | Colney fields, Norwich | no |
| *Rhodococcus* sp. strain WS3 | Apr-2015 | Soil | *Salix alba L.* | Colney fields, Norwich | yes |
| *Rhodococcus* sp. strain WS4 | Jul-2015 | Soil | *Salix alba L.* | Colney fields, Norwich | yes |
| *Rhodococcus* sp. strain WS5 | Jul-2015 | Soil | *Salix alba L.* | Colney fields, Norwich | no |
| *Rhodococcus* sp. strain WS6 | Jul-2015 | Soil | *Salix alba L.* | Colney fields, Norwich | no |
| *Rhodococcus* sp. strain WS7 | Jul-2015 | Soil | *Salix alba L.* | Colney fields, Norwich | yes |
| *Rhodococcus* sp. strain WS8 | Jul-2015 | Soil | *Salix alba L.* | Colney fields, Norwich | no |
| *Variovorax* sp. strain WS9 | Jul-2015 | Soil | *Salix alba L.* | Colney fields, Norwich | yes |
| *Rhodococcus* sp. strain WS10 | Jul-2015 | Soil | *Salix alba L.* | Colney fields, Norwich | no |
| *Variovorax* sp. strain WS11 | Jun-2016 | Soil | *Salix alba L.* | UEA broads, Norwich | yes |

| | | | | | |
|---|---|---|---|---|---|
| *Nocardioides* sp. strain WS12 | Jun-2016 | Soil | *Salix alba L.* | UEA broads, Norwich | yes |
| *Variovorax* sp. strain WS13 | Jun-2016 | Soil | *Salix alba L.* | UEA broads, Norwich | no |
| *Rhodococcus* sp. strain TD1 | Jun-2016 | Soil | -* | Fakenham, Norfolk | no |
| *Rhodococcus* sp. strain TD2 | Jun-2016 | Soil | -* | Fakenham, Norfolk | no |
| *Rhodococcus* sp. strain TD3 | Jun-2016 | Soil | -* | Fakenham, Norfolk | no |
| *Rhodococcus* sp. strain WL1 | Jun-2016 | Phyllosphere | *Salix alba L.* | UEA broads, Norwich | no |
| *Rhodococcus* sp. strain OPL1 | Feb-2017 | Phyllosphere | *Elaeis guineensis* | Kew Gardens, London | no |
| *Gordonia* sp. strain OPL2 | Feb-2017 | Phyllosphere | *Elaeis guineensis* | Kew Gardens, London | yes |
| *Sphingobacterium* sp. strain OPL3 | Feb-2017 | Phyllosphere | *Elaeis guineensis* | Kew Gardens, London | no |
| *Sphingopyxis* sp. strain OPL5 | Feb-2017 | Phyllosphere | *Elaeis guineensis* | Kew Gardens, London | yes |

[±] The same *Salix alba* L and *Elaeis guineensis* isoprene emitting trees were sampled throughout this study.
*There was no evidence of plants or shrubs close to samples taken from the tyre dump site.

The genome size and mol % G+C content of the isolated strains were compared to the genome data of their closest 16S rRNA gene identity and a reference organism for the genus (using the NCBI database). For example, *Variovorax* sp. strain WS9 (genome size of 7.0 Mbp; 65.3 % GC content) and *Variovorax* sp. strain WS11 (8.6 Mbp; 67.2 %) had a genome size similar to *Variovorax* sp. HW608 (7.7 Mbp; 66.5%; GeneBank assembly accession: GCA_900090195), *Variovorax paradoxus* EPS (6.5 Mbp; 66.5 %, (Jamieson *et al.*, 2009)), and *Variovorax paradoxus* S110 (6.7Mbp; 67.5 %, (Han *et al*., 2011)). *Sphingopyxis* sp. OPL5 (4.7 Mbp; 65.8 %) was compared to *Sphingopyxis* QXT-31 (4.29 Mbp; 66.5 %; GeneBank assembly accession: GCA_001984035) and styrene-degrading *Sphingopyxis fribergensis* (5.2 Mbp; 63.8 %; GeneBank assembly accession: GCA_000803645). For the *Rhodococcus* isolates, the reference genome used was *Rhodococcus* sp. AD45 (6.8 Mbp; 61.7 %). The genome sizes, for organisms of the genus, were as expected, except for *Rhodococcus* sp. strain WS4 (12.7 Mbp; 66.4 %). The only 16S rRNA gene sequence found in the *Rhodococcus* sp. strain WS4 genome annotation had a significant identity to *Rhodococcus opacus* 1CP (8.6 Mbp; 67 %, (Eulberg *et al.*, 1998)) and *Rhodococcus jostii* RHA1 (9.7 Mbp; 67 %). The genome size of the latter, corresponds to one chromosome and three plasmids -pRHL1, pRHL2 and pRHL3-(McLeod *et al.*, 2006). The genome size for *Rhodococcus* sp. strain WS4 was still surprisingly larger than its closest relative and to *Rhodococcus* sp. AD45. Genome size and the number of coding sequences suggested that the *Rhodoccoccus* sp. strain WS4 genome sequence was contaminated, and that it contained plasmids or bacteriophages.

A plasmid and bacteriophage search for the genome sequence of *Rhodococcus* sp. strain WS4 was performed using the PlasmidFinder and PHAST webservers. The plasmid database gave no evidence for the presence of plasmids in the genome (Carattoli et al., 2014), while the PHAST webserver identified one contig to containing prophage-like coding sequences (presented in **Supplementary Table 6** and **Supplementary Figure 12** (Y. Zhou et al., 2011)). Moreover, the RAST Seed Viewer identified two isoprene metabolic gene clusters which confirmed contamination of the *Rhodococcus* sp. strain WS4 genome. DNA from *Rhodococcus* sp. strain WS4 used for genome sequencing was most likely contaminated with a second isoprene-degrading strain of the *Rhodococcus* genus.

**Table 6-2** Genome information for isoprene-degrading bacteria. Data include representative isoprene-degrading bacteria, isolates from this study and isoprene cluster-containing metagenome bins from the willow soil DNA-SIP. Data for genome size, mol %G+C content and number of coding sequences was obtained using the RAST Server (Aziz et al., 2008).

| Strain | Isolate/bin identified as (identity)* | Size (Mbp) | % GC content | Number of coding sequences |
|---|---|---|---|---|
| *Rhodococcus* sp. AD45 | - | 6.8 | 61.7 | 6,252 |
| *Rhodococcus opacus* PD630 | - | 9.3 | 66.8 | 8,448 |
| *Gordonia polyisoprenivorans* i37 | - | 6.2 | 66.8 | 5,587 |
| *Rhodococcus* sp. strain WS1 | *Rhodococcus* sp. Djl-6-2 (100%)<br>*Rhodococcus* sp. strain S10 (100%) | 6.6 | 62.3 | 6,276 |
| *Rhodococcus* sp. strain WS3 | *Nocardia globerula* strain DSM 44596 (100%) | 6.9 | 61.7 | 6,600 |
| *Rhodococcus* sp. strain WS4 | *Rhodococcus opacus* strain 1CP (99%)<br>*Rhodococcus jostii* RHA1 (99%) | 12.7± | 66.4 | 12,726 |
| *Variovorax* sp. strain WS9 | *Caenimonas* sp. SL 110 (98%)<br>*Variovorax* sp BZ15 (98%) | 7.0 | 65.3 | 6,614 |
| *Variovorax* sp. strain WS11 | *Variovorax* sp. RA8 (99%)<br>*Variovorax* sp. HW608 (99%) | 8.6 | 67.2 | 8,130 |

| | | | | |
|---|---|---|---|---|
| *Nocardioides* sp. strain WS12 | *Nocardioides aromaticivorans* strain H9 (97%)<br>*Pimelobacter simplex* strain VKM Ac-2033D (97%) | 5.2 | 68.6 | 4,959 |
| *Gordonia* sp. strain OPL2 | *Gordonia terrae* strain NRRL B-16283 (99%)<br>*Gordonia terrae* strain Bu15_45 (99%) | 5.8 | 67.3 | 5,372 |
| *Sphingopyxis* sp. strain OPL5 | *Sphingopyxis* sp. QXT-31 (99%)<br>*Sphingopyxis macrogoltabida* strain EY-1 (99%) | 4.7 | 65.8 | 4,450 |
| WS.mg.bin19 | *Rhodococcus erythropolis*[±] | 6.7 | 61.9 | 6,473 |
| WS.mg.bin20 | Sphingomonadales order[±] | 4.4 | 62.5 | 4,510 |

*Full 16S rRNA gene sequence was extracted from genomes and aligned using blastn (without environmental samples/uncultured hits) against NCBI database.
[±]WS.mg bins had no 16S rRNA gene, Check M identity is displayed.

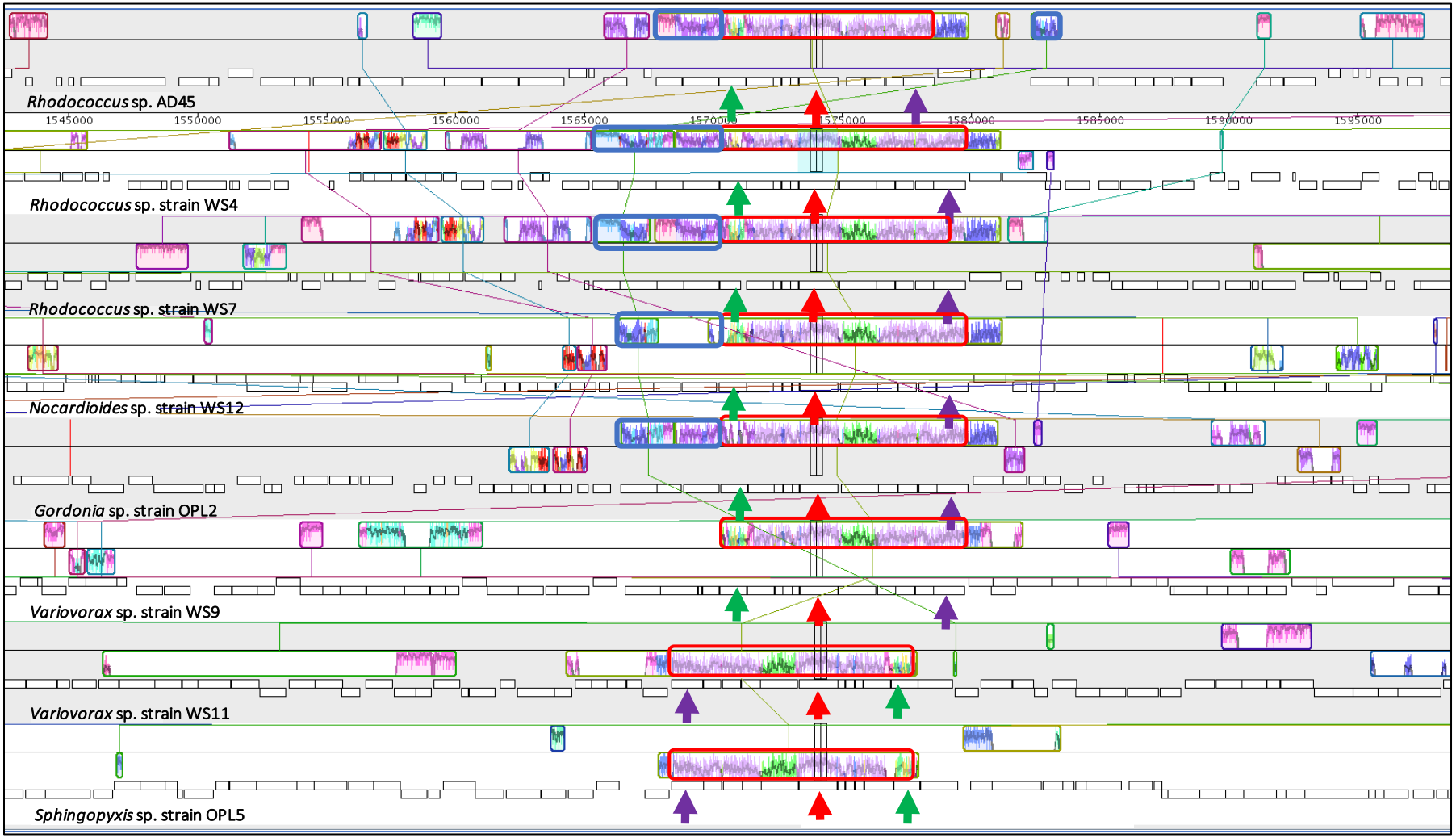



**Figure 6.1** Multiple genome alignment of novel isoprene-degrading bacteria to locate conserved regions close to the isoprene metabolic gene cluster using Mauve programme (Darling et al., 2004). *Rhodococcus* AD45 was used as the reference for the isoprene gene cluster (top). The *isoA* gene (red arrows), *isoF* gene (purple arrows) and *isoG* genes (green arrows) are shown. When including an *E. coli* strain and *V. paradoxus* EPS, as negative controls, there was no positive alignment to isoprene metabolic genes in those genomes. This image includes an alignment region of approximately 60 Kbp.

#### *6.1.1.1 Isoprene metabolic gene clusters in novel isoprene-degrading bacteria*

The first step in the comparison of isoprene metabolic genes of the novel isoprene-degrading bacteria was to locate the gene cluster in the sequenced genomes. The multiple genome alignment tool Mauve (Darling et al., 2004) was used to compare 8 genomes and to find the *isoA* gene, annotated as *tmoA* in the computer-derived genome annotations due to its sequence similarity with the *tmoA* gene encoding the large subunit of the oxygenase component of the toluene monooxygenase from *Pseudomonas mendocina* KR1 (Yen *et al.,* 1991; Crombie *et al*., 2015). The conserved region was determined through the alignment algorithm used by Mauve. The algorithm first finds local alignments by matching regions that exist in all genomes, called homologous regions or anchors; the program then partitions anchors to determine the boundaries of the co-linear blocks and determines the full region of homology. In this analysis, the conserved region and neighbouring genes were found to contain the canonical isoprene metabolic gene cluster for the novel isoprene-degrading bacteria, as compared to the model organism *Rhodococcus* sp. AD45 (**Figure 6.1**). This comparison of genome sequences permitted the analysis of large scale evolutionary events (Darling et al., 2004), and a broader analysis of neighbour homologous sequences compared to the analysis of individual genes involved in isoprene degradation.

As mentioned, the conserved region was identified in all the isoprene-degrader genomes examined. In **Figure 6.1** the region of interest was confirmed by the presence of ten key genes and an additional gene, *aldH*_1 (encoding an aldehyde dehydrogenase, shown with the green similarity plot), not included inside the cluster for *Rhodococcus* sp. strain AD45. The anchor, or homologous block of genes (red block), contains the key genes of the isoprene metabolic gene cluster. Details of the polypeptides encoded by the ten key genes, and their function, are shown in **Table 6.3** (Johan et al., 1999; J. E. van Hylckama Vlieg et al., 1998; J. E. T. Van Hylckama Vlieg, Leemhuis, Lutje Spelberg, & Janssen, 2000).

**Table 6-3** Functional genes of the isoprene metabolic gene cluster.

| Gene | Number of polypeptides* | Function |
|---|---|---|
| *isoG* | 405 | Putative coenzyme A transferase |
| *isoH* | 226 | Dehydrogenase |
| *isoI* | 238 | Glutathione-S-transferase |
| *isoJ* | 233 | Glutathione-S-transferase |
| *isoA* | 514 | Hydroxylase $\alpha$-subunit |
| *isoB* | 94 | Hydroxylase $\gamma$-subunit |
| *isoC* | 114 | Ferredoxin |
| *isoD* | 110 | Coupling protein |
| *isoE* | 342 | Hydroxylase $\beta$-subunit |
| *isoF* | 345 | Reductase |

*Polypeptide sequence size was taken from the reference isoprene-degrading bacterium *Rhodococcus* sp. AD45 (Crombie et al., 2015).

In addition to the main isoprene gene cluster, the alignment visualization shows other neighbour genes that are conserved in all Gram-positive isolates (blue blocks in **Figure 6.1** for *Rhodococcus, Nocardioides* and *Gordonia*). The genes encode another aldehyde dehydrogenase, a glutathione synthase and NADH oxidase. Interestingly, those three genes are not found together in the genome of *Rhodococcus* sp. AD45.

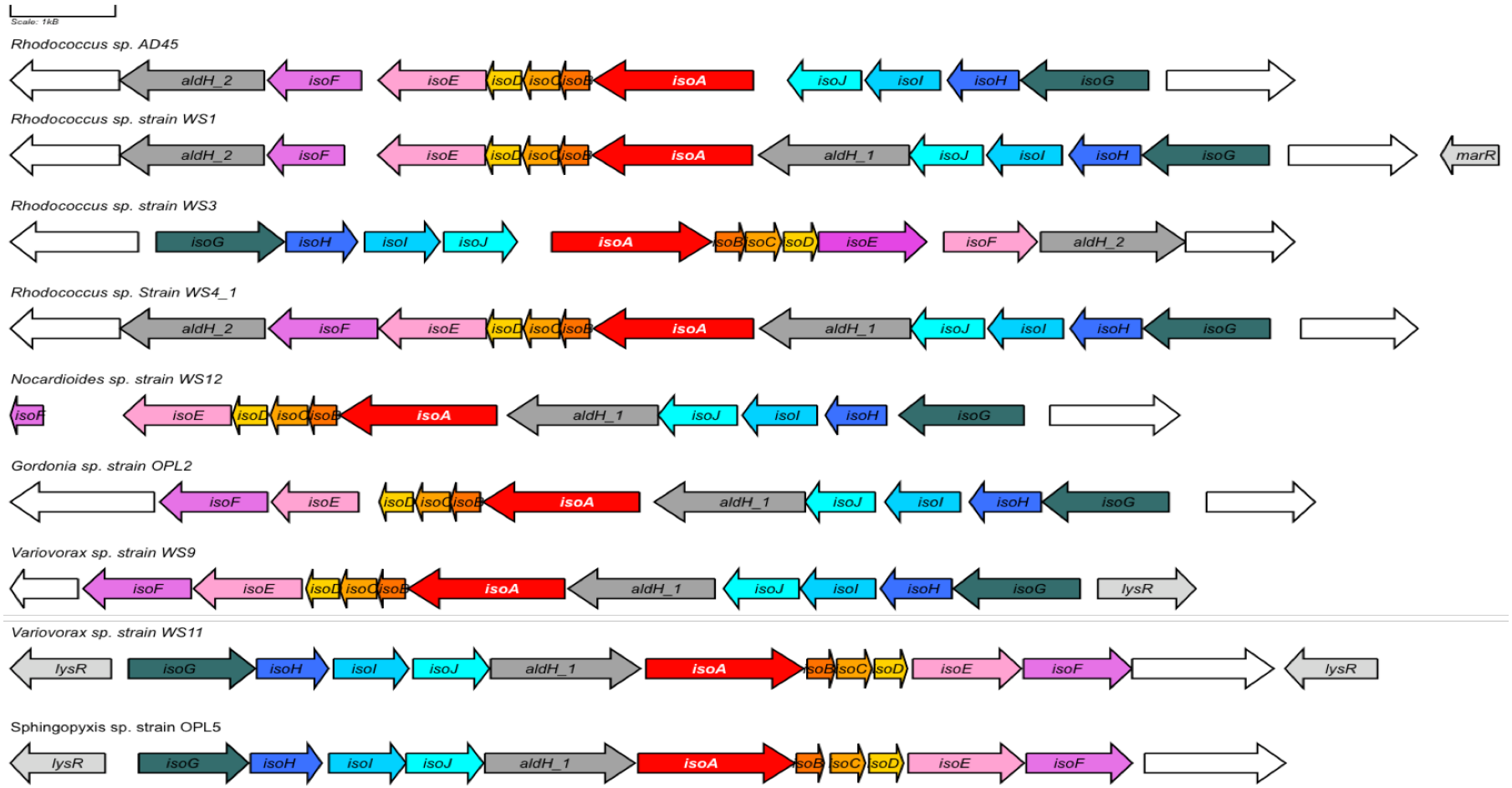



**Figure 6.2** Isoprene metabolic gene clusters of isoprene-degrading bacteria. The model isoprene-degrading bacterium *Rhodococcus* sp. AD45 is placed at the top. The *isoA* gene, that encodes for the $\alpha$-subunit of the monooxygenase, is shown in red.

**Table 6-4** One-on-one polypeptide sequence comparison between isoprene-degrading bacteria. For IsoA, the TmoA sequence from *Pseudomonas mendocina* KR1 was included (top), and for IsoG, a putative coenzyme A transferase was compared (below). Refer to **Table 6.3** for identity of isoprene degrading genes.

| 100 | 99-90 | 89-80 | 79-70 | 69-60 | 59-50 | 49-35 | 34-0 |
|---|---|---|---|---|---|---|---|

| **IsoA** | *P m* KR1 | *R* AD45 | *R o* PD630 | *G p* i37 | *M* AT1 | *R* WS4 | *R* WS7 | *N* WS12 | *G* OPL2 | *V* WS9 | *V* WS11 | *S* OPL5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *P m* KR1 | 100 | 48 | 48 | 50 | 49 | 49 | 48 | 47 | 48 | 47 | 48 | 48 |
| *R* AD45 | 48 | 100 | 91 | 86 | 83 | 90 | 91 | 84 | 87 | 71 | 73 | 72 |
| *R o* PD630 | 48 | 91 | 100 | 91 | 88 | 96 | 93 | 88 | 90 | 74 | 74 | 75 |
| *G p* i37 | 50 | 86 | 91 | 100 | 86 | 91 | 88 | 85 | 93 | 75 | 75 | 75 |
| *M* AT1 | 49 | 83 | 88 | 86 | 100 | 88 | 86 | 82 | 86 | 73 | 74 | 73 |
| *R* WS4 | 49 | 90 | 96 | 91 | 88 | 100 | 93 | 86 | 90 | 74 | 75 | 75 |
| *R* WS7 | 48 | 91 | 93 | 88 | 86 | 93 | 100 | 86 | 89 | 73 | 73 | 74 |
| *N* WS12 | 47 | 84 | 88 | 85 | 82 | 86 | 86 | 100 | 85 | 72 | 73 | 71 |
| *G* OPL2 | 48 | 87 | 90 | 93 | 86 | 90 | 89 | 85 | 100 | 72 | 73 | 72 |
| *V* WS9 | 47 | 71 | 74 | 75 | 73 | 74 | 73 | 72 | 72 | 100 | 92 | 81 |
| *V* WS11 | 48 | 73 | 74 | 75 | 74 | 75 | 73 | 73 | 73 | 92 | 100 | 82 |
| *S* OPL5 | 48 | 72 | 75 | 75 | 73 | 75 | 74 | 71 | 73 | 81 | 82 | 100 |

| **IsoG** | *R* AD45 | *R o* PD630 | *G p* i37 | *M* AT1 | *R* WS4 | *R* WS7 | *N* WS12 | *G* OPL2 | *V* WS9 | *V* WS11 | *S* OPL5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *R* AD45 | 100 | 91 | 79 | 84 | 91 | 90 | 75 | 78 | 61 | 60 | 61 |
| *R o* PD630 | 91 | 100 | 83 | 88 | 96 | 96 | 76 | 81 | 60 | 57 | 59 |
| *G p* i37 | 79 | 83 | 100 | 81 | 84 | 81 | 74 | 90 | 57 | 56 | 56 |
| *M* AT1 | 84 | 88 | 81 | 100 | 87 | 87 | 74 | 80 | 60 | 56 | 58 |
| *R* WS4 | 91 | 96 | 84 | 87 | 100 | 98 | 75 | 82 | 60 | 57 | 60 |
| *R* WS7 | 90 | 96 | 81 | 87 | 98 | 100 | 75 | 81 | 59 | 57 | 59 |
| *N* WS12 | 75 | 76 | 74 | 74 | 75 | 75 | 100 | 73 | 60 | 57 | 59 |
| *G* OPL2 | 78 | 81 | 90 | 80 | 82 | 81 | 73 | 100 | 57 | 55 | 55 |
| *V* WS9 | 61 | 60 | 57 | 60 | 60 | 59 | 60 | 57 | 100 | 85 | 67 |
| *V* WS11 | 60 | 57 | 56 | 56 | 57 | 57 | 57 | 55 | 85 | 100 | 65 |
| *S* OPL5 | 61 | 59 | 56 | 58 | 60 | 59 | 59 | 55 | 67 | 65 | 100 |

| **IsoH** | *R* AD45 | *R o* PD630 | *G p* i37 | *M* AT1 | *R* WS4 | *R* WS7 | *N* WS12 | *G* OPL2 | *V* WS9 | *V* WS11 | *S* OPL5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *R* AD45 | 100 | 86 | 79 | 78 | 88 | 88 | 73 | 81 | 62 | 59 | 61 |
| *R o* PD630 | 86 | 100 | 83 | 83 | 93 | 91 | 71 | 83 | 62 | 60 | 60 |
| *G p* i37 | 79 | 83 | 100 | 82 | 81 | 80 | 71 | 85 | 61 | 59 | 60 |
| *M* AT1 | 78 | 83 | 82 | 100 | 80 | 79 | 69 | 83 | 62 | 61 | 63 |
| *R* WS4 | 88 | 93 | 81 | 80 | 100 | 93 | 71 | 84 | 62 | 60 | 63 |
| *R* WS7 | 88 | 91 | 80 | 79 | 93 | 100 | 70 | 81 | 61 | 58 | 62 |
| *N* WS12 | 73 | 71 | 71 | 69 | 71 | 70 | 100 | 70 | 65 | 64 | 61 |
| *G* OPL2 | 81 | 83 | 85 | 83 | 84 | 81 | 70 | 100 | 60 | 59 | 60 |
| *V* WS9 | 62 | 62 | 61 | 62 | 62 | 61 | 65 | 60 | 100 | 82 | 60 |
| *V* WS11 | 59 | 60 | 59 | 61 | 60 | 58 | 64 | 59 | 82 | 100 | 61 |
| *S* OPL5 | 61 | 60 | 60 | 63 | 63 | 62 | 61 | 60 | 60 | 61 | 100 |

After the genome comparisons, the arrangement of core isoprene metabolic gene clusters for each strain was examined (**Figure 6.2**). This confirmed the synteny in the organization of the genes (Junier & Rivoire, 2013). Additionally, protein databases for each gene in the metabolic gene cluster were generated by first identifying each gene belonging to the isoprene metabolic cluster (identified as shown in **Supplementary Table 8** for *Variovorax* sp. strain WS9), extracting the encoded protein sequences from the genome annotations and placing them in individual databases for each functional protein. Polypeptide sizes, derived from each gene, are shown in **Supplementary Table 9**. Protein sequence alignments, between homologous genes, were generated using the blastp suite in the BLAST website and ClustalW using MEGA7 software (Kumar *et al.*, 2016; Thompson *et al.*, 2002). Blastp results for IsoA and IsoG/IsoH are shown in **Table 6.4.**

Finally, a phylogenetic analysis of protein sequences for the genes encoded in the isoprene metabolic gene cluster was performed using MEGA7 software (Kumar *et al.*, 2016). The trees used a maximum likelihood test with 1000 replicates to increase the reliability in the results (Douady *et al.*, 2003; Felsenstein, 2011). All polypeptide trees included protein sequences from confirmed isoprene-degrading reference bacteria (7-9 sequences) and new isolates from this study (8 sequences). In order to increase the robustness of the phylogenetic analysis to the basic 15-17 sequences from isolates, protein sequences recovered from the willow soil metagenome were also included in the analysis. Adding the metagenome sequences also allowed us to study the origin and to correlate with the abundance of the sequences during the willow tree soil DNA-SIP experiment (sequences obtained in **Chapter 4**).

The phylogenetic trees for IsoA and IsoG are presented in **Figure 6.3** and **Figure 6.4**, respectively. All other phylogenetic trees can be found in **Supplementary Figures 13** to **20**. IsoABCDEF phylogenetic trees included the sequences for TmoABCDEF from *Pseudomonas mendocina* KR1 as an outgroup (Yen *et al*., 1991). There was no reference organisms, to use as an outgroup for IsoGHIJ phylogenetic analysis.

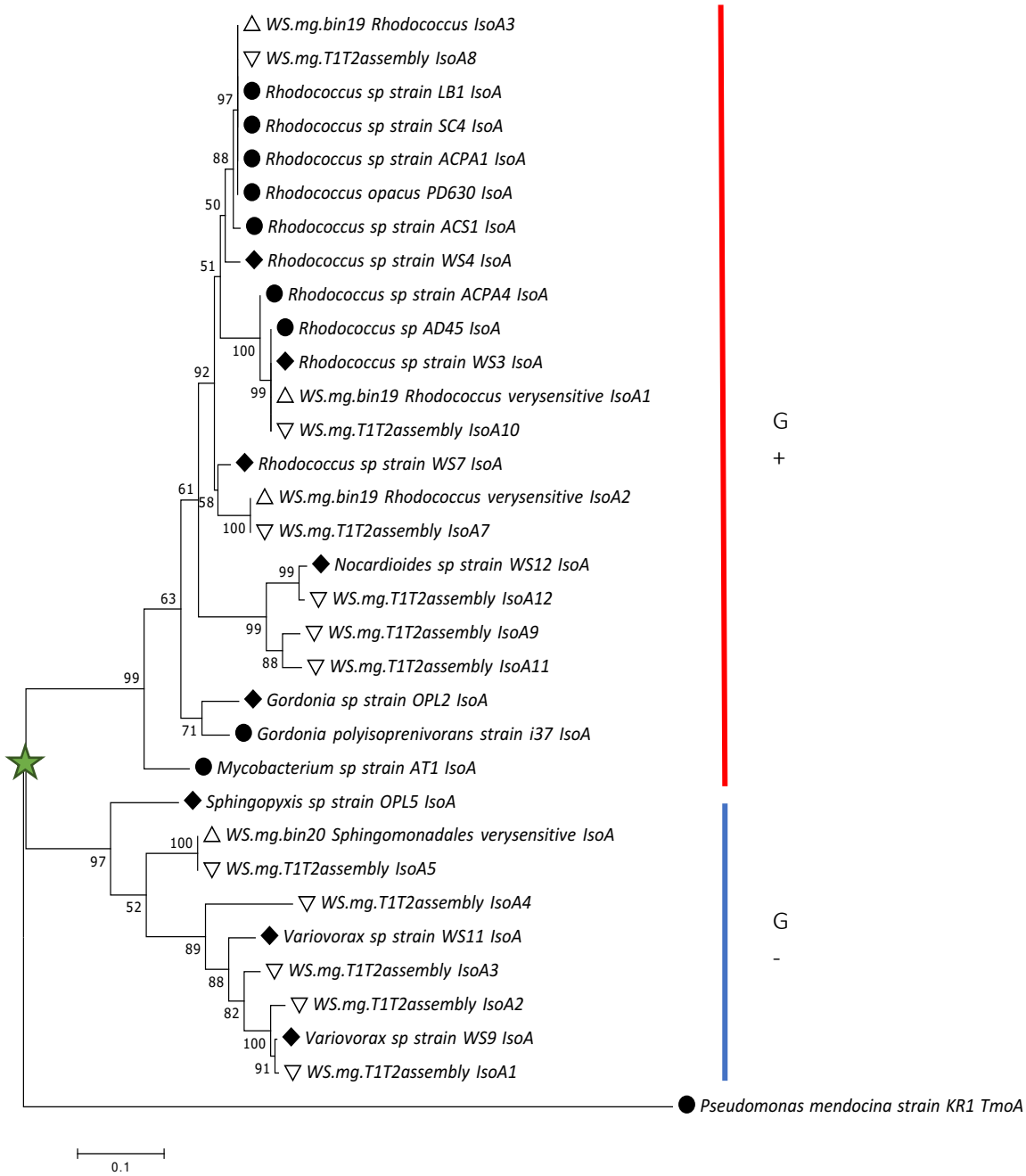



**Figure 6.3** Phylogenetic analysis using IsoA sequences. Trees were constructed with maximum likelihood method and a bootstrap of 1000 replicates. The analysis was carried out with 33 different IsoA sequences. All gaps and missing data were eliminated, a total of 485 amino acid residues were included in the final dataset. Gram-positive sequences (red group) and Gram-negative sequences (blue group) and the tree root (green star) that separates both groups are shown. Reference sequences (filled circle), novel isolate sequences (filled diamond), and metagenome sequences (empty triangles). The four WS.mg (willow soil metagenome) bin IsoJ sequences were obtained using the *very sensitive* settings using Metabat programme (details in section 2.18).

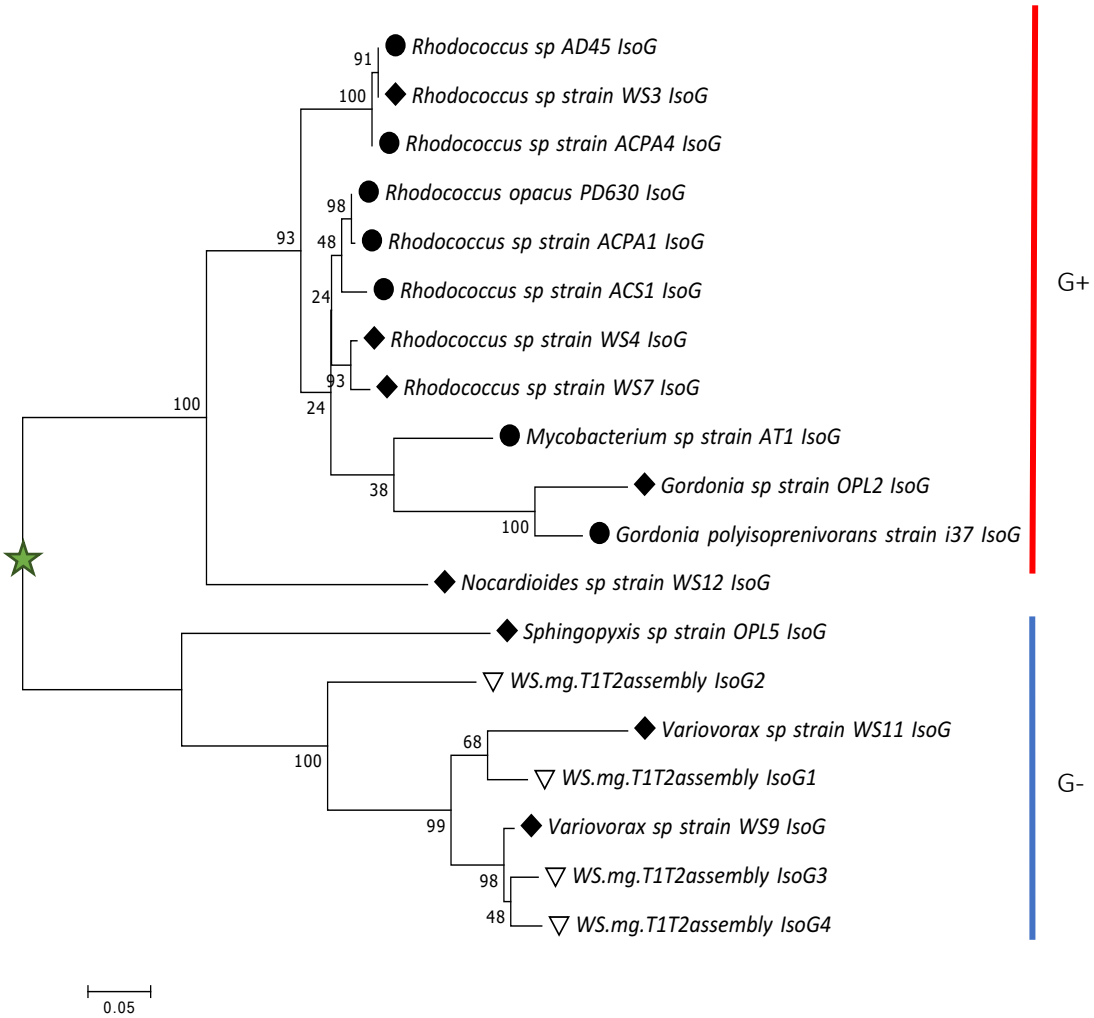
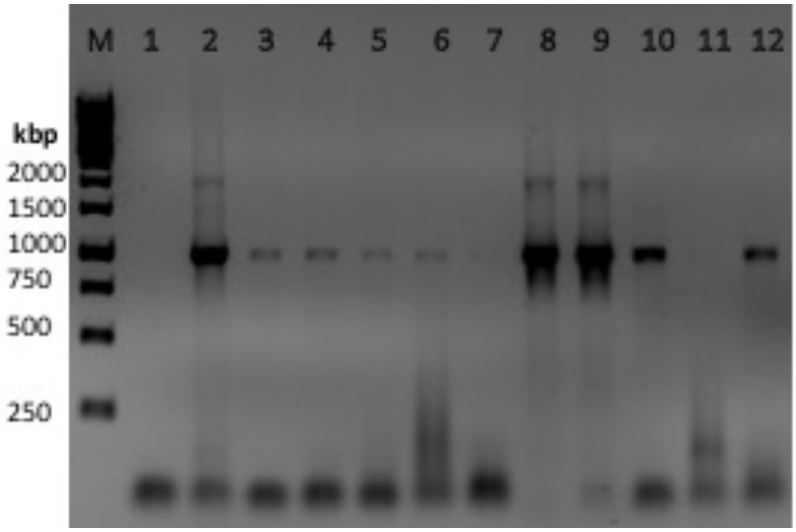



**Figure 6.4** Phylogenetic analysis using IsoG sequences. Trees were constructed with maximum likelihood method and a bootstrap of 1000 replicates. The analysis was carried out with 19 different IsoG sequences. All gaps and missing data were eliminated, a total of 396 amino acid residues were included in the final dataset. Gram-positive sequences (red group) and Gram-negative sequences (blue group) and the tree root (green star) that separates both groups are shown. Reference sequences (filled circle), novel isolate sequences (filled diamond), and metagenome sequences (empty triangles).

The ten phylogenetic trees, derived from the ten isoprene metabolic genes analysed, show a clear division between Gram- positive or Gram- negative sequences. Also, most sequences obtained from environmental samples can be clearly assigned to a specific genus. This now permits a quick way to establish which organism the *iso* gene sequences belong to in further environmental screening experiments for isoprene metabolic gene clusters. These analyses present no evidence of recent horizontal gene transfer between organisms that contain the isoprene metabolic gene cluster.

### 6.1.2 Diversity of the *isoA* gene and the design of a new degenerate PCR primer pair

Screening for the presence/absence of isoprene-related genes in isolates and environmental samples has been done with an *isoA* gene PCR primer pair designed by Myriam El Khawand and colleagues (MEK-*isoA* primers) (El Khawand *et al.*, 2016). PCR amplification experiments using the MEK-*isoA* primers have showed variable amplification intensity for Gram-positive isolates (**Figure 6.5**) and no PCR product was obtained with genomic DNA from *Variovorax* sp. strain WS9 (refer to **Figure 3.13**). The results obtained using these MEK-*isoA* primers suggested the absence of the *isoA* gene and the isoprene metabolic gene cluster in *Variovorax* sp. strain WS9 (**Chapter 3**).




| | | | |
|---|---|---|---|
| 1 | Negative control | 7 | *Rhodococcus* sp. strain WS5 |
| 2 | Positive control (R. AD45) | 8 | *Rhodococcus* sp. strain WS1 |
| 3 | *Rhodococcus* sp. strain WS6 | 9 | *Rhodococcus* sp. strain WS7 |
| 4 | *Rhodococcus* sp. strain WS2 | 10 | *Rhodococcus* sp. strain WS8 |
| 5 | *Rhodococcus* sp. strain WS3 | 11 | *Rhodococcus* sp. strain WS10 |
| 6 | *Rhodococcus* sp. strain WS4 | 12 | *Rhodococcus* sp. strain WL1 |

**Figure 6.5** *isoA* gene PCR products obtained using genomic DNA from willow soil isolates using the MEK-*isoA* primers.

**Table 6-5** Gene sequences used for the design of the NLM-*isoA* primers.

| Isolate | Gene | Reference |
|---|---|---|
| *Xanthobacter autotrophicus* strain Py2 | *aamA** | (Zhou *et al.* 1999) |
| *Rhodococcus* sp. AD45 | *isoA* | (van Hylckama Vlieg *et al.*, 1998) |
| *Rhodococcus opacus* PD630 | *isoA* | (Crombie *et al.*, 2015) |
| *Gordonia polyisoprenivorans* i37 | *isoA* | (Alvarez *et al*., 2009) |
| *Mycobacterium* sp. AT1 | *isoA* | (Alvarez *et al*., 2009) |
| *Rhodococcus* sp. LB1 | *isoA* | (El Khawand *et al*., 2016) |
| *Rhodococcus* sp. SC4 | *isoA* | (El Khawand *et al*., 2016) |
| *Rhodococcus* sp. strain WS1 | *isoA* | This study |
| *Rhodococcus* sp. strain WS3 | *isoA* | This study |
| *Rhodococcus* sp. strain WS4 | *isoA* | This study |
| *Rhodococcus* sp. strain WS7 | *isoA* | This study |
| *Variovorax* sp. strain WS9 | *isoA* | This study |
| Metagenome sequences (8 sequences) | *isoA* | Leaf washing experiment done by Andrew T. Crombie in 2015 |

*the gene encodes for the $\alpha$-subunit of the alkene monooxygenase.


**Figure 6.6** Location of MEK-*isoA* primers (grey arrows) and NLM-*isoA* primers (blue arrows). The *isoA* gene is approx. 1515 bp. Figure includes two iron-binding sites (red ovals) and γ-binding site (green oval).

**Table 6-6** General parameters for the NLM-*isoA* degenerate PCR primer pair

| **Parameter*** | **NLM-*isoA*f1 (GVGACGAYTGGTAYGACA)** | **NLM-*isoA*r1 (GCRTTBGGBTTCCAGAAC)** | **Parameters recommended‡** |
|---|---|---|---|
| Length (bp) | 18 | 18 | 18-28 |
| GC content (%) | 44-61 | 44-61 | 50-60 |
| Melting Temperature Tm (°C) | 45.8-52.6 | 45.8-52.6 | 55-80 |
| Self-complementarity | None | None | None |
| Potential hairpin formation | None | No strict matches | None |

*All primer parameters were acquired using: http://biotools.nubic.northwestern.edu/OligoCalc.html

‡Recommended parameters (Innis, 1990)

The presence of the isoprene metabolic gene cluster, and the *isoA* gene, was later confirmed in the genome of *Variovorax* sp. strain WS9. After genome annotation, the *isoA* gene was identified due to its sequence similarity to *tmoA* as shown by Crombie et al., 2015 (48% identity). Genome annotation results for the isoprene metabolic gene cluster in *Variovorax* sp. strain WS9 are shown in **Supplementary Table 8**. The presence of the isoprene metabolic gene cluster in a Gram-negative terrestrial isolate demonstrated the existence of a wider diversity of the isoprene metabolic genes and the need to redesign the *isoA* PCR primer set.

A new pair of degenerate *isoA* gene PCR primers (NLM-*isoA* primers) were designed manually in 2016 using 20 curated *isoA* gene sequences (**Table 6.5**). The primer pair yielded one product that amplified between nucleotide position 11 and 1023 of the *isoA* gene, with a 1012 bp product (primer parameters in **Table 6.6**). **Figure 6.6** compares the location of the old and new primer pair locations on the *isoA* gene. The targets for the new PCR primer pair permit the amplification of both iron-binding sites of the *isoA* gene

### *6.1.2.1 Optimization of a PCR protocol for isoA gene amplification*

The NLM-*isoA* primer pair was evaluated with DNA from isolates confirmed to contain *isoA*. Different annealing temperatures were tested during the PCR protocol (50 °C - 60 °C) with DNA from *Rhodococcus* sp. AD45 (as the positive control), *Rhodococcus* sp. strain WS1 and *Variovorax* sp. strain WS9 (data not shown). The best amplification intensity, with non-specific amplification, was obtained with an annealing temperature of 54 °C. The complete, optimized PCR protocol is presented in **Table 6.7** and components in **Table 6.8**.

**Table 6-7** PCR protocol for the *isoA* gene amplification using the new degenerate NLM-*isoA* primers.

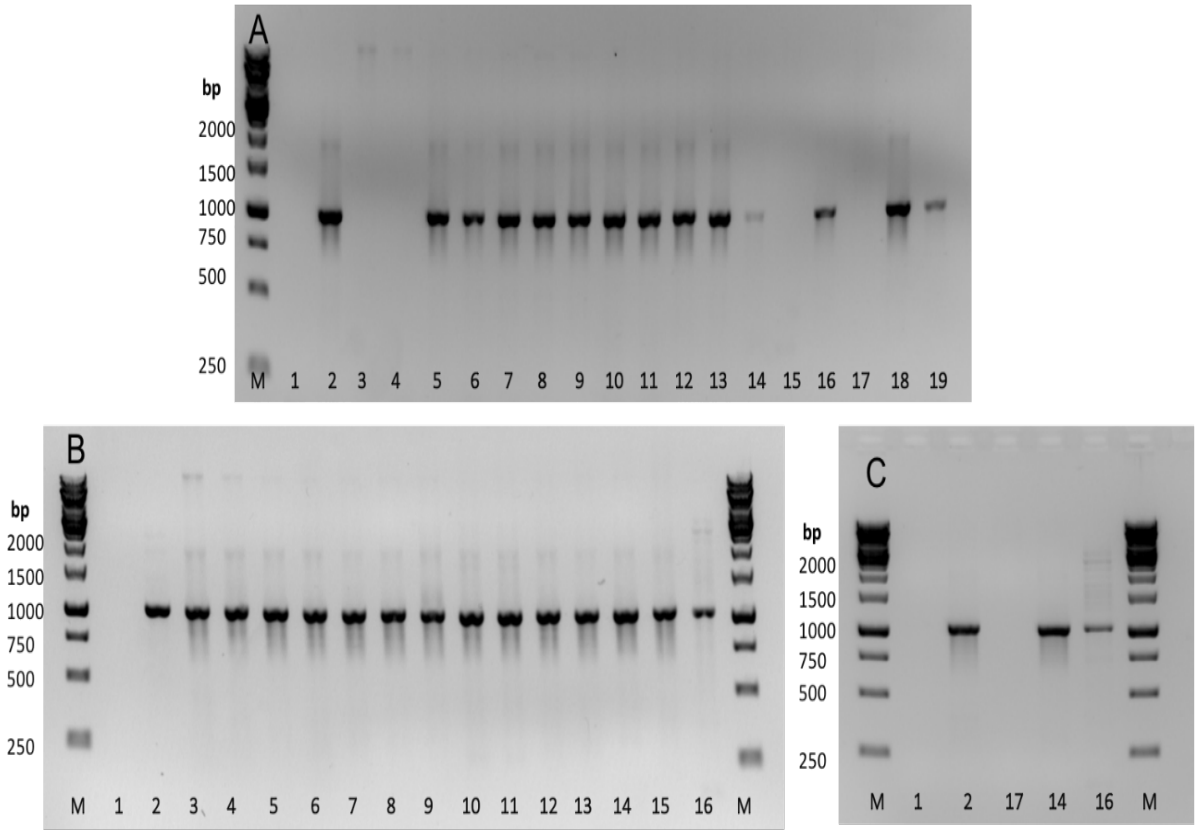
| Cycle | Step | Description | Temperature | Time |
|---|---|---|---|---|
| 1 | 1 | Denature | 94 °C | 2 min |
| 2* | 1 | Denature | 95 °C | 15 secs |
| | 2 | Anneal | 54 °C | 30 secs |
| | 3 | Elongate | 72 °C | 1 min |
| 3 | 1 | Elongate | 72 °C | 7 min |

*This cycle was repeated 30 times.

**Table 6-8** Components used during the NLM *isoA* primer PCR.

| Reactant/solution | Final concentration |
|---|---|
| $dH_2O$ | - |
| Buffer (Fermentas) | 1X |
| NMF isoAf1 primer | 0.4 μM |
| NMF isoAr1 primer | 0.4 μM |
| dNTPs | 0.2mM each |
| DNA Taq Polymerase (Fermentas) | 2.5 units |

NLM-*isoA* primers were also compared to the MEK-*isoA* primers by amplifying the *isoA* gene using genomic DNA from isoprene-degrading isolates and with isoprene enriched and unenriched environmental samples (**Figure 6.7BC**). The PCR amplicon product size is similar for both primer sets (1kb). MEK-*isoA* primers pair yielded a positive PCR product with all the Gram-positive isolates i.e. *Rhodococcus* and *Gordonia*, and the enriched environmental sample. NLM-*isoA* primers amplified the *isoA* gene for all isoprene-degraders tested (Gram-positive and Gram-negative) and for the enriched samples. The native soil sample produced no visible product with either primer pair. All PCR products were sent for sequencing and confirmed the amplification of the *isoA* gene. Results demonstrated that a wider diversity of *isoA* genes could be targeted with the new *isoA* gene PCR primer pair. Due to time constraints, the primer pair was not tested on the isolates recovered in 2017 i.e. *Gordonia, Sphingobacterium* and *Sphingopyxis.*




| | | | |
|---|---|---|---|
| 1 | Negative control | 11 | R*hodococcus* sp. strain Essex75 |
| 2 | Positive control | 12 | *Gordonia polyisoprenivorans* i37 |
| 3 | *Variovorax* sp. strain WS11 | 13 | *Rhodococcus erythropolis* i47 |
| 4 | *Variovorax* sp. strain WS13 | 14 | Willow leaf washing enrichment |
| 5 | *Rhodococcus* sp. strain WL1 | 15 | *Variovorax* sp. strain WS9 |
| 6 | *Rhodococcus* sp. strain WS10 | 16 | Enriched willow soil |
| 7 | *Rhodococcus* sp. strain TD1 | 17 | Willow soil |
| 8 | *Rhodococcus* sp. strain TD2 | 18 | *Rhodococcus* sp. strain WS1 |
| 9 | *Rhodococcus* sp. strain TD3 | 19 | *Rhodococcus* sp. strain WS3 |
| 10 | *Rhodococcus* sp. strain Essex74 | | |

**Figure 6.7** *isoA* gene PCR products from isoprene-degraders and environmental samples. A) PCR products with primer set designed by MEK, B and C) PCR products with primer set designed by NLM; M: marker 1kb ladder. Positive controls in A is *Rhodococcus* sp. AD45, in B is *Variovorax* sp. WS9

## 6.2 Discussion

Whole genome sequencing has been essential in the genotypic understanding of the novel isoprene-degrading bacteria strains recovered from environmental samples. The screening of the *isoA* gene and polypeptide profile analysis with *Variovorax* sp. strain WS9 was inconclusive in revealing the presence or absence of the gene cluster (**Chapter 3**). However, the genome sequence data demonstrated the presence of the complete isoprene

metabolic gene cluster in Gram-negative terrestrial strains. Sequence data of those gene clusters also permitted the identification of isoprene metabolic genes in the metagenome data from the willow soil DNA-SIP; the work helped to recover two bins and a few contigs that contained the partial isoprene metabolic cluster (**Chapter 4**).

Protein sequences encoded by the isoprene metabolic gene clusters, from cultivation-dependent and cultivation-independent data, were then aligned to generate phylogenetic trees. The *isoA* gene has been previously analysed phylogenetically to show two groups, terrestrial and marine/estuarine, with diverse phylogenetic groups including Gram-positive and Gram-negative strains (El Khawand et al., 2016). My study has also shown that terrestrial strains are separated into two groups as well but separated by Gram staining, not environment. Gram-positive and Gram-negative isoprene degraders are separated in all 10 phylogenetic trees (**Figure 6.3, Figure 6.4** and **Supplementary Figures 13** to **20**) for proteins encoded in the isoprene metabolic gene cluster. Even more relevant to the cultivation-independent studies is the fact that all phylogenetic trees have also separated the Gram-negative and Gram-positive genera into confined clusters. This phylogenetic organization suggest the evolution of the metabolic cluster within the specific genera and no recent horizontal gene transfer events (Boto *et al*., 2010; Gogarten & Townsend, 2005).

In these, the early years of research of isoprene degradation, the analysis and comparison of functional gene information with nucleic acid work is essential in the analysis of the microbial ecology of isoprene-degraders. The construction of databases, as "living documents", will aid in the recognition of isoprene degraders and provide further independence from cultivation-based methods. From the ten conserved isoprene genes analysed in this study, the *isoA* encoding $\alpha$-subunit of the isoprene monooxygenase was shown to be the most conserved gene in the isoprene metabolic gene cluster. The *isoA* genes also have enough internal diversity in their sequences to be able to examine the bacterial genera of origin. *isoA* is therefore a good functional gene marker to study natural isoprene-degrading populations.

The NLM-*isoA* primer pair was redesigned as part of this study to also amplify members of the Gram- negative Comamonadaceae family. The primer pair should be redesigned once again to include all the willow soil metagenome *isoA* sequences and *isoA* sequence

information for the latest isolates, *Sphingopyxis* sp. strain OPL5 and *Gordonia* sp. strain OPL2. The 1 kb product from the NLM-*isoA* primer pair has enough information to determine the bacterial genus of origin and insight into the quality of isoprene-degradation in a specific sample. To determine the quantity and abundance of the cluster, a new primer pair is being designed by Ornella Carrion Fonseca. The latter will also be used for amplicon sequencing analysis of *isoA* as a functional gene marker.

Another functional gene marker to study the microbial ecology of isoprene metabolism would be the *isoG, isoH, isoI* or *isoJ* gene. Since IsoI and IsoJ are not as conserved as IsoG and IsoH (see **Supplementary Table 6.2**), a probe for *isoH* or *isoI* would be a great option when coupled with screening for *isoA*. The results for IsoH and IsoI sequence comparison have been shown in **Table 6.4** and **Supplementary Table 7**, and although they are less conserved than *isoA* gene, they are both important and relatively unique genes for the isoprene metabolic pathway compared to other alkene metabolic pathways.

# Chapter 7 Final Discussion and Future Work

*Environments with isoprene degraders*

This study has expanded our knowledge of isoprene degradation in the terrestrial environment, specifically on and around isoprene-emitting trees. Environments and organisms associated with isoprene-production/emission, such as isoprene-emitting trees, are important in the cycling of isoprene in the environment and may be considered a hotspot for isoprene-degraders. In the natural environment, isoprene can be oxidized by isoprene assimilating (isoprene degraders) and non-assimilating microbes due to the presence of monooxygenases of specific (IsoMO) or broad-range in compound catalysis mechanisms (PrMO or sMMO, Johnston *et al.*, 2017). While non-assimilating microbes can only oxidize isoprene to epoxyisoprene (co-oxidation) in the first step in the catalysis (for example *Xanthobacter* Py2 using an alkene monooxygenase (van Ginkel et al., 1986)), isoprene-degraders use the epoxyisoprene produced in this first oxidation step as a source of carbon and energy.

To help us understand the diversity of isoprene-degraders around isoprene-emitting trees, leaves and soil from different isoprene-emitters were sampled. Enrichment experiments from leaf samples showed higher variation in the isoprene consumption rates, even between replicates of the same tree. The variation can be partially attributed to intra-species variation but there was also evidence of predation during incubation. Some replicates that began with high consumption rates reduced or stopped consuming isoprene during the procedure. This problem was solved partially by continuously supplementing isoprene to keep the communities growing and helped to overcome the loss of part of the community.

Data from cultivation-dependent and cultivation-independent experiments have helped to identify, understand the distribution and assess the abundance of isoprene degraders in the terrestrial environment. Lower concentrations of isoprene supplemented to enrichment assays yielded a greater diversity of isoprene-degrading bacteria.

Cultivation-dependent: This study has shown a wider diversity of isoprene degraders with isolates from the genera *Variovorax*, *Nocardioides* and *Rhodococcus* associated with willow trees and *Gordonia*, *Sphingopyxis*, and *Rhodococcus* with oil palm trees. The presence of different strains of *Rhodococcus* in various soil and leaf environments, along with previous studies (El Khawand *et al*., 2016; Jens Ewers *et al*., 1990, van Hylckama Vlieg *et al.*, 1998), may be consistent with consideration of this genus as a cosmopolitan isoprene-degrader in the terrestrial environment. The *Rhodoccocus* genus also seems to have a metabolic advantage at high isoprene concentrations. In the case of *Gordonia*, originally found in isoprene enrichments from marine and estuarine environments (L. A. Alvarez et al., 2009; Johnston et al., 2017), this is the first report of the genus as an isoprene-degrading epiphyte in the phyllosphere.

Cultivation-independent: High throughput sequencing has proven to be an effective tool for investigating bacteria in different research areas (Edwards & Holt, 2013). After enriching with 13-25 ppmv $^{13}$C-isoprene, DNA-SIP partial 16S rRNA gene amplicon sequencing diversity analysis in the willow soil showed over 75 % abundance of Comamonadaceae family in the heavy DNA (20 % confirmed to correspond to *Variovorax* sp. strain WS9) and less than 5 % from the *Rhodococcus* genus. DNA diversity analysis with heavy DNA from DNA-SIP experiments with oil palm soil showed over 30 % of 16S rRNA gene sequences were from *Aquabacterium*, 20-30 % *Rhodococcus* and <5 % for *Gordonia* and for the *Aquincola* genus. Finally, the same 16S rRNA gene analysis of oil palm leaf epiphytes has over 40 % Gordonia, 10-20 % *Sphingomonas* and <5 % for *Rhodococcus* and for *Aquincola*. Cultivation-independent data verified the presence of the isolated genera from each environment.

Willow soil SIP-metagenomics was an informative method to evaluate key isoprene degraders from different environments. The results verified heavy fraction 16S rRNA gene amplicon sequencing data and bins were also extracted along with incomplete isoprene metabolic gene clusters. Interesting, even with the high abundance of members of the Comamonadaceae family present during the enrichment and the extraction of *isoA* sequences along with a few incomplete clusters, there were no extractable bins for *Variovorax* or members of the Comamonadaceae family with isoprene metabolic genes. These results suggest a limitation to the binning method used during the analysis of the willow soil SIP-metagenomes. In this case, only the isolate genomes gave enough evidence of the presence of *Variovorax* as an isoprene-degrading isolate in this study. Recent results in our lab, from Dr. Andrew Crombie, have shown the recovery of a *Variovorax*-like bin from Poplar leaf SIP enrichments (Crombie, *et al*., 2018 submitted to PNAS).

A few studies have used cultivation-dependent and cultivation-independent methods to prove the presence of specific organisms in environmental samples (Gutierrez et al., 2013; Tang et al., 2018). The use of DNA-SIP has confirmed the presence and abundance of these key isoprene-degrading isolates in the environments from which they were recovered. In this study, cultivation-independent methods have helped to partially identify the key active microorganisms and cultivation-dependent methods have verified our findings.

*Future work:*

-Characterization of novel isolates (*Variovorax*, *Nocardioides* and *Sphingopyxis*) and search databases for isoprene genes or broad range monooxygenases in the *Aquincola, Thiomonas, Polaromonas, Limnohabitans* and the *Aquabacterium* genera.

-As mentioned previously, *Rhodococcus* was found in low abundance after enrichment with 13-25 ppmv isoprene. Comparison of cultivation-independent work from enriching samples with high- and low- concentrations of isoprene, by adding known amounts of *Variovorax* and *Rhodococcus*, might be useful to study the dynamics of both genera and their behaviour *in-vitro*.

-Soil and leaf sampling and experimentation with oil palm samples in the natural environment to compare to enrichment results from the greenhouse oil palm tree. These

results might confirm the presence of specific genera on oil palm trees or distinguish between the specific greenhouse from Kew Gardens.

-Microorganisms from the genera *Methylibium* (> 20 %) and *Methylobacterium* (> 10 %) were enriched and labelled throughout the willow soil and oil palm leaf SIP experiments, respectively. *Methylobacterium* has been studied as an isoprene degrader by Srivastva *et al*., in 2017 although there are no nucleotide-based studies that demonstrate that the bacterium is a isoprene-degrader. More work to study the presence of broad range methane/alkene monooxygenases (sMMO, PrMO) might explain the co-oxidation of isoprene by other organisms.

### *Sampling of isoprene-degraders in the environment*

One of the biggest challenges during this study was to obtain novel isolates from different environmentally relevant samples for isoprene degradation. Soil and leaf samples associated to isoprene-emitting trees have delivered a wide variety of isoprene-degraders by using different methodologies and isolation strategies. First, isoprene was consumed by the microcosm without the need of additional nutrients in soils and isoprene consumption was also observed in leaf washing enrichments, although it was not as efficient. The use of no additional nutrients would be a great advantage in environmentally-relevant studies and would also avoid the immediate selection of certain isoprene degraders during experimentation. The diversity of isolates obtained from the study may be skewed by the formulation of minimal medium used in the study, which might give an advantage to members of the *Rhodococcus* genus (enhanced with the use of high-isoprene concentrations). The use of different formulations of minimal medium can help in the isolation process but should not be used during the enrichment process.

Second, some samples may be difficult to access and/or transport, such as large leaves and or branches from trees. In some cases, such as ours with the oil palm trees, there was limitation in the number of trees present in the city/country where the work was done. The swabbing of leaves was therefore used as an alternative method, designed using a clinical

environment sampling methodology. Although the results are comparable when sampling large surfaces, the leaf washing methodology is still the best option to obtain epiphytes from leaves.

Finally, the sub-cultivation strategy was used to select and effectively isolate isoprene-degrading bacteria from complex microcosms. The oil palm tree samples came from a greenhouse environment. The Palm House, in Kew Gardens, includes a great diversity of tropical trees which are subjected to different conditions compared to the natural environment in which they would be found. Other varying factors may include the use of fertilizers, different types of soil, migration of microbes between plants and increased temperature/humidity. The combination of all these conditions made isolation of isoprene-degraders complex. A sub-cultivation methodology was therefore designed, coupled with the use of different media formulations, to be an effective way of isolation or targeted isolation of a wider diversity of isoprene degraders.

*Future work:*

-A thorough ecological study of the potential and abundance of isoprene-degradation in a broad number of different environments, using marker genes like *isoA*, and associated to isoprene-emitting trees.

-With the surprising observation of an overlap between organisms found in the soil and leaves of the oil palm tree, as seen in *Arabidopsis* (Bai *et al*., 2015), it would be interesting to continue to compare the diversity of isoprene degraders between these two environments from the same isoprene-emitting tree.

-*Variovorax* is a known endophyte (Han *et al*., 2011) and now is a known isoprene degrading bacterium. Isoprene-degraders may be present in the xylem and/or phloem of plants, being transported from the leaves to the roots or vice-versa, respectively. The study of isoprene-degrading endophytes is an area that might be interesting to explore and might explain the overlap/presence of certain bacteria in the leaves and soil of specific isoprene-emmiting trees.

-Continue to study and verify that different isoprene degraders are found associated with different isoprene-emitting trees (inter-species variation).

*Understanding the genetics and functionality of isoprene degraders in the environment*

When studying the isoprene metabolic gene cluster diversity, the analysis of the genome of *Variovorax* sp. strain WS9 revealed the presence of the isoprene degradation gene cluster that has 43-72% identity to *Rhodococcus sp.* AD45 metabolic genes. This result reflects greater variability of the isoprene metabolic gene cluster in the environment. The protein databases generated in this study have shown clear clustering of the sequences according to their genus and their Gram-stain, which suggests the traits have been present in the microorganisms for many years. The multiple genome analysis demonstrated that 10 genes are shared between all isoprene degraders. These genes form the core isoprene metabolic gene cluster, *isoGHIJABCDEF*.

The *isoA* gene, that encodes the $\alpha$-subunit of the IsoMO, is found in the middle of the cluster. In the metagenome analysis, most of the core isoprene-related gene sequences recovered were from this gene. This might be due to the location of the gene in the cluster and how conserved the gene is compared to the others. Our *isoA* database had more sequences, compared to other genes, which might have helped in the search of similar sequences throughout the assembly. Additionally, a new pair of *isoA* PCR primers were designed because the previous PCR primer pair did not amplify *Variovorax sp.* WS9 and *Variovorax sp.* WS11 *isoA* genes. An additional primer pair, for a unique gene in the isoprene metabolic pathway, is suggested to study the abundance of Gram-positive sequences (that include a duplication of IsoGHIJ) and as a new functional marker that will separate isoprene metabolism from other SDIMOs (specifically alkene monooxygenases). *isoG* or *isoH* have been suggested due to the percentage of sequence conservation and sequence length between isolates (refer to **Chapter 6**).

*Future work:*

-Analysis of genes up- and down-stream of the isoprene metabolic gene cluster. Specifically, glutathione synthesis genes and *aldH* to study the evolutionary mechanism(s) by which the cluster has appeared and how it differs from other alkene metabolic gene clusters.

-Some isoprene degraders have two copies of the *isoGHIJ* cluster (Crombie et al., 2015, 2018). It would be interesting to understand the genetic characteristics/evolution of the duplication and if is only present in Gram-positive isoprene degraders, as suggested in this study. The duplication of the *isoGHIJ* cluster may be studied in Gram-positive bacteria as an alternative method for efficiently detoxifying the cell from the accumulation of epoxyisoprene during isoprene metabolism.

# References


Alimentos, I. de A. y T. de. (1991). Spoilage of a bakery product (Sobao pasiego) by isoprene-producing molds. *Revista de Agroquimica y Tecnologia de Alimentos (Spain)*. Retrieved from http://agris.fao.org/agris-search/search.do?recordID=ES9200277

Alvarez, H. M., Mayer, F., Fabritius, D., & Steinbüchel, A. (1996). Formation of intracytoplasmic lipid inclusions by *Rhodococcus opacus* strain PD630. *Archives of Microbiology*, *165*(6), 377–386. https://doi.org/10.1007/s002030050341

Alvarez, L. A., Exton, D. A., Timmis, K. N., Suggett, D. J., & McGenity, T. J. (2009). Characterization of marine isoprene-degrading communities. *Environmental Microbiology*, *11*(12), 3280–3291. https://doi.org/10.1111/j.1462-2920.2009.02069.x

Aronesty, E. (2011). ea–utils: Command–line tools for processing biological sequencing data - Google Académico. Retrieved from http://code.google.com/p/ea–utils

Arsène-Ploetze, F., Koechler, S., Marchal, M., Coppée, J.-Y., & Chandler, M. (2010). Structure, Function, and Evolution of the *Thiomonas* spp. *Genome. PLoS Genet*, *6*(2), 1000859. https://doi.org/10.1371/journal.pgen.1000859

Ashworth, K., Wild, O., & Hewitt, C. N. (2013). Impacts of biofuel cultivation on mortality and crop yields. *Nature Climate Change*, *3*(5), 492–496. https://doi.org/10.1038/nclimate1788

Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., … Zagnitko, O. (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*, *9*, 1–15. https://doi.org/10.1186/1471-2164-9-75

Baggi, G., Barbieri, P., Galli, E., & Tollari, S. (1987). Isolation of a Pseudomonas stutzeri strain that degrades o-xylene. *Applied and Environmental Microbiology*, *53*(9), 2129–2132.

Bai, Y., Müller, D. B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., … Schulze-Lefert, P. (2015). Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature*, *528*(7582), 364–369. https://doi.org/10.1038/nature16192

Bamberger, I., Ruehr, N. K., Schmitt, M., Gast, A., Wohlfahrt, G., & Arneth, A. (2017). Isoprene emission and photosynthesis during heat waves and drought in black locust. *Biogeosciences Discussions*, (February), 1–26. https://doi.org/10.5194/bg-2017-32

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477.

https://doi.org/10.1089/cmb.2012.0021

Battaglia-Brunet, F., Joulian, C., Garrido, F., Dictor, M.-C., Morin, D., Coupland, K., … Baranger, P. (n.d.). Oxidation of arsenite by *Thiomonas* strains and characterization of Thiomonas arsenivorans sp. nov. https://doi.org/10.1007/s10482-005-9013-2

Behnke, K., Ehlting, B., Teuber, M., Bauerfeind, M., Louis, S., Hänsch, R., … Schnitzler, J. P. (2007). Transgenic, non-isoprene emitting poplars don't like it hot. *Plant Journal*, *51*(3), 485–499. https://doi.org/10.1111/j.1365-313X.2007.03157.x

Bogas, A. C., Ferreira, A. J., Araújo, W. L., Astolfi-Filho, S., Kitajima, E. W., Lacava, P. T., & Azevedo, J. L. (2015). Endophytic bacterial diversity in the phyllosphere of Amazon Paullinia cupana associated with asymptomatic and symptomatic anthracnose. *SpringerPlus*, *4*(1). https://doi.org/10.1186/s40064-015-1037-0

Bont, J. A. M., Attwood, M. M., Primrose, S. B., & Harder, W. (1979). Epoxidation of short chain alkenes in *Mycobacterium* E20: The involvement of a specific mono-oxygenase. *FEMS Microbiology Letters*, *6*(3), 183–188. https://doi.org/10.1111/j.1574-6968.1979.tb04306.x

Boto, L. (2010). Horizontal gene transfer in evolution: Facts and challenges. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1683), 819–827. https://doi.org/10.1098/rspb.2009.1679

Bravo, D., Braissant, O., Cailleau, G., Verrecchia, E., & Junier, P. (2014). Isolation and characterization of oxalotrophic bacteria from tropical soils. *Archives of Microbiology*, *197*(1), 65–77. https://doi.org/10.1007/s00203-014-1055-2

Calfapietra, C., Fares, S., & Loreto, F. (2009). Volatile organic compounds from Italian vegetation and their interaction with ozone. *Environmental Pollution*, *157*(5), 1478–1486. https://doi.org/10.1016/j.envpol.2008.09.048

Capone, K. A., Dowd, S. E., Stamatas, G. N., & Nikolovski, J. (2011). Diversity of the human skin microbiome early in life. *Journal of Investigative Dermatology*, *131*(10), 2026–2032. https://doi.org/10.1038/jid.2011.168

Carattoli, A., Zankari, E., Garciá-Fernández, A., Larsen, M. V., Lund, O., Villa, L., … Hasman, H. (2014). In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrobial Agents and Chemotherapy*, *58*(7), 3895–3903. https://doi.org/10.1128/AAC.02412-14

Carlton, A. G., Wiedinmyer, C., & Kroll, J. H. (2009). A review of secondary organic aerosol (SOA) formation from isoprene. *Atmospheric Chemistry and Physics*, *9*(14), 4987–

5005. https://doi.org/10.5194/acp-9-4987-2009

Chen, W. H., Guenther, A. B., Wang, X. M., Chen, Y. H., Gu, D. S., Chang, M., … Zhang, Y. Q. (2018). Regional to global biogenic isoprene emission responses to changes in vegetation from 2000 to 2015. *Journal of Geophysical Research: Atmospheres*, *123*(7), 3757–3771. https://doi.org/10.1002/2017JD027934

Cheung, S., McCarl, V., Holmes, A. J., Coleman, N. V., & Rutledge, P. J. (2013). Substrate range and enantioselectivity of epoxidation reactions mediated by the ethene-oxidising *Mycobacterium* strain NBB4. *Applied Microbiology and Biotechnology*, *97*(3), 1131–1140. https://doi.org/10.1007/s00253-012-3975-6

Ciok, A., Dziewit, L., Grzesiak, J., Budzik, K., Gorniak, D., Zdanowski, M. K., & Bartosik, D. (2016). Identification of miniature plasmids in psychrophilic Arctic bacteria of the genus *Variovorax*. *FEMS Microbiology Ecology*, *92*(4), 1–9. https://doi.org/10.1093/femsec/fiw043

Claeys, M., Graham, B., Vas, G., Wang, W., Vermeylen, R., Pashynska, V., … Maenhaut, W. (2004). Formation of secondary organic aerosols through photooxidation of isoprene. *Science*, *303*(5661), 1173–1176. https://doi.org/10.1126/science.1092805

Cleveland, C. C., Yavitt, B., & Yavitt, J. B. (1997). Consumption of atmospheric isoprene. *Geophysical Research Letters*, *24*(19), 2379–2382. https://doi.org/10.1029/97GL02451

Cleveland, C. C., & Yavitt, J. B. (1998). Microbial consumption of atmospheric isoprene in a temperate forest soil. *Appl Environ Microbiol*, *64*(1), 172–177.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., … Tiedje, J. M. (2014). Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, *42*(D1), 633–642. https://doi.org/10.1093/nar/gkt1244

Coleman, N. V., Bui, N. B., & Holmes, A. J. (2006). Soluble di-iron monooxygenase gene diversity in soils, sediments and ethene enrichments. *Environmental Microbiology*, *8*(7), 1228–1239. https://doi.org/10.1111/j.1462-2920.2006.01015.x

Corley, R. H. V., & Tinker, P. B. (2016). *The oil palm*. (R. H. V. Corley & P. B. Tinker, Eds.) (Fifth Edit). Chichester, West Sussex, UK: John Wiley & Sons, Ltd.

Coumou, D., & Rahmstorf, S. (2012). A decade of weather extremes. *Nature Climate Change*, *2*(7), 491–496. https://doi.org/10.1038/nclimate1452

Coyotzi, S., Pratscher, J., Murrell, J. C., & Neufeld, J. D. (2016). Targeted metagenomics of active microbial populations with stable-isotope probing. *Current Opinion in*

*Biotechnology*, *41*, 1–8. https://doi.org/10.1016/j.copbio.2016.02.017

Crombie, A. T., Khawand, M. El, Rhodius, V. A., Fengler, K. A., Miller, M. C., Whited, G. M., … Murrell, J. C. (2015). Regulation of plasmid-encoded isoprene metabolism in *Rhodococcus*, a representative of an important link in the global isoprene cycle. *Environmental Microbiology*, *17*(9), 3314–3329. https://doi.org/10.1111/1462-2920.12793

Crombie, A. T., Mejia-Florez, N. L., McGenity, T. J., & Murrell, J. C. (2018). Genetics and Ecology of Isoprene Degradation. *Aerobic Utilization of Hydrocarbons, Oils and Lipids*, 1–15. https://doi.org/10.1007/978-3-319-39782-5_27-1

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve : Multiple alignment of conserved genomic sequence with rearrangements, 1394–1403. https://doi.org/10.1101/gr.2289704

Datukishvili, N. T., Tarkhnishvili, G. M., Mikeladze, D. G., Beridze, T. G., & Sanadze, G. A. (2001). Isolation and purification of protein responsible for the conversion of dimethylallylpyrophosphate from poplar leaves into isoprene. *Russian Journal of Plant Physiology*, *48*(2), 222–225. https://doi.org/10.1023/A:1009008318764

Deneris, E. S., Stein, R. A., & Mead, J. F. (1984). Invitro biosynthesis of isoprene from mevalonate utilizing a rat liver cytosolic fraction. *Biochemical and Biophysical Research Communications*, *123*(2), 691–696. https://doi.org/10.1016/0006-291X(84)90284-5

Ding, X., He, Q. F., Shen, R. Q., Yu, Q. Q., Zhang, Y. Q., Xin, J. Y., … Wang, X. M. (2016). Spatial and seasonal variations of isoprene secondary organic aerosol in China: Significant impact of biomass burning during winter. *Scientific Reports*, *6*(1), 20411. https://doi.org/10.1038/srep20411

Dingman, D. W. (2000). Growth of *Escherichia coli* O157:H7 in bruised apple (*Malus domestica*) tissue as influenced by cultivar, date of harvest, and source. *Applied and Environmental Microbiology*, *66*(3), 1077–1083. https://doi.org/10.1128/AEM.66.3.1077-1083.2000

Dorn, E., Reineke, W., & Knackmuss, H.-J. (1974). Isolation and characterization of a 3-chlorobenzoate degrading pseudomonad. *Archives of Microbiology*, *99*(1), 61–70.

Douady, C. J., Delsuc, F., Boucher, Y., Doolittle, W. F., & Douzery, E. J. P. (2003). Comparison of bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Molecular Biology and Evolution*, *20*(2), 248–254.

https://doi.org/10.1093/molbev/msg042

Dowd, S. E., Callaway, T. R., Wolcott, R. D., Sun, Y., McKeehan, T., Hagevoort, R. G., & Edrington, T. S. (2008). Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiology*, *8*, 1–8. https://doi.org/10.1186/1471-2180-8-125

Dumont, M. G., & Murrell, J. C. (2005). Stable isotope probing - Linking microbial identify to function. *Nature Reviews Microbiology*, *3*(6), 499–504. https://doi.org/10.1038/nrmicro1162

Durana, N., Navazo, M., Gómez, M. C., Alonso, L., García, J. A., Ilardia, J. L., … Iza, J. (2006). Long term hourly measurement of 62 non-methane hydrocarbons in an urban area: Main results and contribution of non-traffic sources. *Atmospheric Environment*, *40*(16), 2860–2872. https://doi.org/10.1016/j.atmosenv.2006.01.005

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*(16), 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Edwards, D. J., & Holt, K. E. (2013). Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microb Inform Exp*, *3*(1), 2. https://doi.org/10.1186/2042-5783-3-2

El Khawand, M., Crombie, A. T., Johnston, A., Vavlline, D. V., McAuliffe, J. C., Latone, J. A., … Murrell, J. C. (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), 2743–2753. https://doi.org/10.1111/1462-2920.13345

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*, *3*, e1319. https://doi.org/10.7717/peerj.1319

Eulberg, D., Lakner, S., Golovleva, L. A., & Schlömann, M. (1998). Characterization of a protocatechuate catabolic gene cluster from *Rhodococcus opacus* 1CP: Evidence for a merged enzyme with 4- carboxymuconolactone-decarboxylating and 3-oxoadipate enol-lactone- hydrolyzing activity. *Journal of Bacteriology*, *180*(5), 1072–1081. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9495744

Evans, R., & Griffiths, G. (2013). Palm oil, land rights and ecosystem services in Gbarpolu County, Liberia. *Walker Institute for Climate System Research, University of Reading*,

(3). Retrieved from http://medcontent.metapress.com/index/A65RM03P4874243N.pdf

Ewers, J., Freier-Schröder, D., Knackmuss, H. J., Frier-Schroeder, D., Knackmuss, & H J. (1990). Selection of trichloroethene (TCE) degrading bacteria that resist inactivation by TCE. *Archives of Microbiology*, *154*(4), 410–413. https://doi.org/10.1007/BF00276540

Ewers, J., Frier-Schroeder, D., Knackmuss, & H J. (1990). Selection of trichloroethene (TCE) degrading bacteria that resist inactivation by TCE, *154*, 410–413.

Ezinkwo, G. O., Tretjakov, V. F., Talyshinky, R. M., Ilolov, A. M., & Mutombo, T. A. (2013). Overview of the catalytic production of isoprene from different raw materials; prospects of isoprene production from bio-ethanol. *Catalysis for Sustainable Energy*, *1*, 100–111. https://doi.org/10.2478/cse-2013-0006

Fall, R., & Copley, S. D. (2000). Bacterial sources and sinks of isoprene, a reactive atmospheric hydrocarbon. *Environmental Microbiology*, *2*(2), 123–130. https://doi.org/10.1046/j.1462-2920.2000.00095.x

Fall, R., & Monson, R. K. (1992). Isoprene emission rate and intercellular isoprene concentration as influenced by stomatal distribution and conductance. *Plant Physiology*, *100*(2), 987–992. https://doi.org/10.1104/pp.100.2.987

Fares, S., Barta, C., Brilli, F., Centritto, M., Ederli, L., Ferranti, F., … Loreto, F. (2006). Impact of high ozone on isoprene emission, photosynthesis and histology of developing *Populus alba* leaves directly or indirectly exposed to the pollutant. *Physiologia Plantarum*, *128*(3), 456–465. https://doi.org/10.1111/j.1399-3054.2006.00750.x

Felsenstein, J. (2011). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, *39*(4), 783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Fortunati, A., Barta, C., Brilli, F., Centritto, M., Zimmer, I., Schnitzler, J. P., & Loreto, F. (2008). Isoprene emission is not temperature-dependent during and after severe drought-stress: A physiological and biochemical analysis. *Plant Journal*, *55*(4), 687–697. https://doi.org/10.1111/j.1365-313X.2008.03538.x

Friedrich, M. W. (2006, February). Stable-isotope probing of DNA: Insights into the function of uncultivated microorganisms from isotopically labeled metagenomes. *Current Opinion in Biotechnology*. https://doi.org/10.1016/j.copbio.2005.12.003

Fuentes, J. D., Gu, L., Lerdau, M., Atkinson, R., Baldocchi, D., Bottenheim, J. W., … Stockwell, W. Biogenic hydrocarbons in the atmospheric boundary layer: A review, 81 Bulletin of

the American Meteorological Society § (2000). American Meteorological Society. https://doi.org/10.1175/1520-0477(2000)081<1537:BHITAB>2.3.CO;2

Gao, J. lian, Yuan, M., Wang, X. ming, Qiu, T. lei, Li, J. wei, Liu, H. can, … Sun, J. guang. (2015). *Variovorax guangxiensis* sp. nov., an aerobic, 1-aminocyclopropane-1-carboxylate deaminase producing bacterium isolated from banana rhizosphere. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *107*(1), 65–72. https://doi.org/10.1007/s10482-014-0304-3

Ginige, M. P., Hugenholtz, P., Daims, H., Wagner, M., Keller, J., & Blackall, L. L. (2004). Use of Stable-Isotope Probing, Full-Cycle rRNA Analysis, and Fluorescence In Situ Hybridization-Microautoradiography to Study a Methanol-Fed Denitrifying Microbial Community. *Applied and Environmental Microbiology*, *70*(1), 588–596. https://doi.org/10.1128/AEM.70.1.588-596.2004

Gogarten, J. P., & Townsend, J. P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology*, *3*(9), 679–687. https://doi.org/10.1038/nrmicro1204

Grob, C., Taubert, M., Howat, A. M., Burns, O. J., Dixon, J. L., Richnow, H. H., … Murrell, J. C. (2015). Combining metagenomics with metaproteomics and stable isotope probing reveals metabolic pathways used by a naturally occurring marine methylotroph. *Environmental Microbiology*, *17*(10), 4007–4018. https://doi.org/10.1111/1462-2920.12935

Gu, L., Baldocchi, D., Verma, S. B., Black, T. A., Vesala, T., Falge, E. M., & Dowty, P. R. (2002). Advantages of diffuse radiation for terrestrial ecosystem productivity. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *107*(D6), ACL-2.

Guenther, A. (1995). A global model of natural volatile organic compound emissions. *Journal of Geophysical Research*, *100*(D5), 8873–8892. https://doi.org/10.1029/94JD02950

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., & Wang, X. (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), 1471–1492. https://doi.org/10.5194/gmd-5-1471-2012

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., & C., G. (2006). Estimates of global terrestrial isoprene emissions using MEGAN. *Atmospheric Chemistry and*

*Physics Discussions*, *6*(1), 107–173. https://doi.org/10.5194/acpd-6-107-2006

Gutierrez, T., Singleton, D. R., Berry, D., Yang, T., Aitken, M. D., & Teske, A. (2013). Hydrocarbon-degrading bacteria enriched by the Deepwater Horizon oil spill identified by cultivation and DNA-SIP. *ISME Journal*, *7*(11), 2091–2104. https://doi.org/10.1038/ismej.2013.98

Han, J. I., Choi, H. K., Lee, S. Y. S. W., Orwin, P. M., Kim, J., LaRoe, S. L., … Han, C. (2011). Complete genome sequence of the metabolically versatile plant growth-promoting endophyte *Variovorax paradoxus* S110. *Journal of Bacteriology*, *193*(5), 1183–1190. https://doi.org/10.1128/JB.00925-10

Han, S. R., Lee, J. H., Kang, S., Park, H., & Oh, T. J. (2016). Complete genome sequence of opine-utilizing *Variovorax* sp. strain PAMC28711 isolated from an Antarctic lichen. *Journal of Biotechnology*, *225*, 46–47. https://doi.org/10.1016/j.jbiotec.2016.03.042

HAN, X. mei, WANG, R. qing, LIU, J., WANG, M. cheng, ZHOU, J., & GUO, W. hua. (2007). Effects of vegetation type on soil microbial community structure and catabolic diversity assessed by polyphasic methods in North China. *Journal of Environmental Sciences*, *19*(10), 1228–1234. https://doi.org/10.1016/S1001-0742(07)60200-9

Hanson, D. T., Swanson, S., Graham, L. E., & Sharkey, T. D. (1999). Evolutionary significance of isoprene emission from mosses. *American Journal of Botany*, *86*(5), 634–639. https://doi.org/10.2307/2656571

Harley, P., Eller, A., Guenther, A., & Monson, R. K. (2014). Observations and models of emissions of volatile terpenoid compounds from needles of ponderosa pine trees growing in situ: Control by light, temperature and stomatal conductance. *Oecologia*, *176*(1), 35–55. https://doi.org/10.1007/s00442-014-3008-5

Harrison, K. J., Crécy-Lagard, V. de, & Zallot, R. (2018). Gene Graphics: A genomic neighborhood data visualization web application. *Bioinformatics*, *34*(8), 1406–1408. https://doi.org/10.1093/bioinformatics/btx793

Harrison, S. P., Morfopoulos, C., Dani, K. G. S., Prentice, I. C., Arneth, A., Atwell, B. J., … Wright, I. J. (2013). Volatile isoprenoid emissions from plastid to planet. *New Phytologist*, *197*(1), 49–57. https://doi.org/10.1111/nph.12021

Hedin, G., Rynbäck, J., & Loré, B. (2010). New technique to take samples from environmental surfaces using flocked nylon swabs. *Journal of Hospital Infection*, *75*(4), 314–317. https://doi.org/10.1016/j.jhin.2010.02.027

Hellén, H., Hakola, H., Pirjola, L., Laurila, T., & Pystynen, K. H. (2006). Ambient air

concentrations, source profiles, and source apportionment of 71 different C2-C10 volatile organic compounds in urban and residential areas of Finland. *Environmental Science and Technology*, *40*(1), 103–108. https://doi.org/10.1021/es051659d

Hellén, H., Tykkä, T., & Hakola, H. (2012). Importance of monoterpenes and isoprene in urban air in northern Europe. *Atmospheric Environment*, *59*, 59–66. https://doi.org/10.1016/j.atmosenv.2012.04.049

Heuston, S., Begley, M., Gahan, C. G. M., & Hill, C. (2012). Isoprenoid biosynthesis in bacterial pathogens. *Microbiology*, *158*(Pt_6), 1389–1401. https://doi.org/10.1099/mic.0.051599-0

Hewitt, C. N., Ashworth, K., Boynard, A., Guenther, A., Langford, B., MacKenzie, A. R., … Wild, O. (2011). Ground-level ozone influenced by circadian control of isoprene emissions. *Nature Geoscience*, *4*(10), 671–674. https://doi.org/10.1038/ngeo1271

Hewitt, C. N., & Street, R. A. (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), 3069–3077. https://doi.org/10.1016/0960-1686(92)90463-U

Hirano, S. S., & Upper, C. D. (2000). Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae* a pathogen, ice Nucleus, and epiphyte. *Microbiology and Molecular Biology Reviews*, *64*(3), 624–653. https://doi.org/10.1128/MMBR.64.3.624-653.2000

Holmes, A. J., & Coleman, N. V. (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. https://doi.org/10.1007/s10482-008-9227-1

Immethun, C. M., Hoynes-O'Connor, A. G., Balassy, A., & Moon, T. S. (2013). Microbial production of isoprenoids enabled by synthetic biology. *Frontiers in Microbiology*, *4*(APR), 1–8. https://doi.org/10.3389/fmicb.2013.00075

Innis, M. (1990). In Innis, MA, Gelfand, DH, Sninsky, JJ and White, TJ (eds), PCR Protocols. Retrieved from https://scholar.google.co.uk/scholar?cluster=9016418962445793868&hl=en&as_sdt=2005&sciodt=0,5#d=gs_cit&p=&u=%2Fscholar%3Fq%3Dinfo%3ATHqtekbBIH0J%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26scfhb%3D1%26hl%3Den

Jamieson, W. D., Pehl, M. J., Gregory, G. A., & Orwin, P. M. (2009). Coordinated surface activities in *Variovorax paradoxus* EPS. *BMC Microbiology*, *9*(1), 124. https://doi.org/10.1186/1471-2180-9-124

Jenkin, M. E., Young, J. C., & Rickard, A. R. (2015). The MCM v3.3.1 degradation scheme for isoprene. *Atmospheric Chemistry and Physics*, *15*(20), 11433–11459. https://doi.org/10.5194/acp-15-11433-2015

Jezbera, J., Jezberová, J., Kasalický´vkasalický´kasalický´v, S., ˇ Imek, K., & Hahn, M. W. (2013). Patterns of *Limnohabitans* microdiversity across a large set of freshwater Habitats as Revealed by Reverse Line Blot Hybridization. *PLoS ONE*, *8*(3), 58527. https://doi.org/10.1371/journal.pone.0058527

Johan, E. T., Hylckama, V. V., Kingma, J., & Janssen, D. B. (1999). Purification of a glutathione S -transferase and a glutathione conjugate-specific dehydrogenase involved in isoprene metabolism in *Rhodococcus* sp . strain. *J Bacteriol*, *181*(7), 2094–2101.

Johansen, S., & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration- with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, *52*(2), 169–210. https://doi.org/10.1111/j.1468-0084.1990.mp52002003.x

Johnston, A., Crombie, A. T., El Khawand, M., Sims, L., Whited, G. M., McGenity, T. J., & Colin Murrell, J. (2017). Identification and characterisation of isoprene-degrading bacteria in an estuarine environment. *Environmental Microbiology*, *19*(9), 3526–3537. https://doi.org/10.1111/1462-2920.13842

Joshi, N. ., & Fass, J. . (2011). Sickle: A sliding-window, adaptive, quality-based trimming tools for FastQ files. Retrieved from https://github.com/najoshi/sickle

Junier, I., & Rivoire, O. (2013). Synteny in bacterial genomes: inference, organization and evolution, (c). Retrieved from http://arxiv.org/abs/1307.4291

Kasalický, V., Jezbera, J., Šimek, K., & Hahn, M. W. (2010). *Limnohabitans planktonicus* sp. nov. and Limnohabitans parvus sp. nov., planktonic betaproteobacteria isolated from a freshwater reservoir, and emended description of the genus *Limnohabitans*. *International Journal of Systematic and Evolutionary Microbiology*, *60*(12), 2710–2714. https://doi.org/10.1099/ijs.0.018952-0

Kataoka, N., Tokiwa, Y., Tanaka, Y., Takeda, K., & Suzuki, T. (1996). Enrichment culture and isolation of slow growing bacteria. *Applied Microbiology and Biotechnology*, *45*, 771–777. https://doi.org/10.1007/s002530050761

Kelly, D. P., & Wood, A. P. (1998). *Microbes of the sulfur cycle*. (R. S. Burlage, R. Atlas, D. Stahl, G. Geesey, & G. Sayler, Eds.), *Techniques in Microbial Ecology* (Oxford Uni).

Kiendler-Scharr, A., Wildt, J., Maso, M. D., Hohaus, T., Kleist, E., Mentel, T. F., … Wahner, A. (2009). New particle formation in forests inhibited by isoprene emissions. *Nature*, *461*(7262), 381–384. https://doi.org/10.1038/nature08292

King, J., Koc, H., Unterkofler, K., Mochalski, P., Kupferthaler, A., Teschl, G., … Amann, A. (2010). Physiological modeling of isoprene dynamics in exhaled breath. *Journal of Theoretical Biology*, *267*(4), 626–637. https://doi.org/10.1016/j.jtbi.2010.09.028

Kleindienst, T. E., Lewandowski, M., Offenberg, J. H., Jaoui, M., & Edney, E. O. (2007). Ozone-isoprene reaction: Re-examination of the formation of secondary organic aerosol. *Geophysical Research Letters*, *34*(1), L01805. https://doi.org/10.1029/2006GL027485

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, *41*(1), 1–11. https://doi.org/10.1093/nar/gks808

Klinger, L. F., Greenburg, J., Guenther, A., Tyndall, G., Zimmerman, P., M'Bangui, M., … Kenfack, D. (1998). Patterns in volatile organic compound emissions along a savanna-rainforest gradient in central Africa. *Journal of Geophysical Research: Atmospheres*, *103*(D1), 1443–1454. https://doi.org/10.1029/97JD02928

Koh, L. P., & Wilcove, D. S. (2008). Is oil palm agriculture really destroying tropical biodiversity? *Conservation Letters*, *1*(2), 60–64. https://doi.org/10.1111/j.1755-263X.2008.00011.x

Krum, J. G., & Ensign, S. A. (2001). Evidence that a linear megaplasmid encodes enzymes of aliphatic alkene and epoxide metabolism and coenzyme M (2-mercaptoethanesulfonate) biosynthesis in *Xanthobacter* strain Py2, *183*(7), 2172–2177. https://doi.org/10.1128/JB.183.7.2172

Kukor, J. J., & Olsen, R. H. (1990). Molecular cloning, characterization, and regulation of a *Pseudomonas pickettii* PKO1 gene encoding phenol hydroxylase and expression of the gene in Pseudomonas aeruginosa PAO1c. *Journal of Bacteriology*, *172*(8), 4624–4630. https://doi.org/10.1128/jb.172.8.4624-4630.1990

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–

1874. https://doi.org/10.1093/molbev/msw054

Kushch, I., Arendacká, B., Štolc, S., Mochalski, P., Filipiak, W., Schwarz, K., … Amann, A. (2008). Breath isoprene – aspects of normal physiology related to age, gender and cholesterol profile as determined in a proton transfer reaction mass spectrometry study. *Clinical Chemistry and Laboratory Medicine*, *46*(7), 1011–1018. https://doi.org/10.1515/CCLM.2008.181

Kuske, C. R., Barns, S. M., Grow, C. C., Merrill, L., & Dunbar, J. (2006). Environmental survey for four pathogenic bacteria and closely related species using phylogenetic and functional genes. *J Forensic Sci*, *51*(3), 548–558. https://doi.org/JFO131 [pii]\r10.1111/j.1556-4029.2006.00131.x

Kuzma, J., & Fall, R. (1993). Leaf isoprene emission rate is dependent on leaf development and the level of isoprene synthase. *Plant Physiology*, *101*(2), 435–440. https://doi.org/10.1104/PP.101.2.435

Kuzma, J., Nemecek-Marshall, M., Pollock, W. H., & Fall, R. (1995). Bacteria produce the volatile hydrocarbon isoprene. *Current Microbiology*, *30*(2), 97–103. https://doi.org/10.1007/BF00294190

Laemmli, C. M., Leveau, J. H. J., Alexander, J. B., & Meer, J. R. Van Der. (2000). Characterization of a second tfd gene cluster for chlorophenol and chlorocatechol metabolism on plasmid pJP4 in *Ralstonia eutropha* JMP134 (pJP4), *134*(15), 4165–4172. https://doi.org/10.1128/JB.182.15.4165-4172.2000.Updated

Lakshmanan, V., Selvaraj, G., & Bais, H. P. (2014). Functional soil microbiome: Belowground solutions to an aboveground problem. *Plant Physiology*, *166*(2), 689–700. https://doi.org/10.1104/pp.114.245811

Lane, D. J. (1991). *16S/23S rRNA sequencing. In: Nucleic acid techniques in bacterial systematics.* (E. Stackebrandt & M. Goodfellow, Eds.). New York: John Wiley & Sons, Inc. https://doi.org/10.4135/9781446279281.n7

Leahy, J. G., Batchelor, P. J., & Morcomb, S. M. (2003). Evolution of the soluble diiron monooxygenases. *FEMS Microbiology Reviews*, *27*(4), 449–479. https://doi.org/10.1016/S0168-6445(03)00023-8

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, J. J., Wang, G., Wu, C., Cao, C., Ren, Y., Wang, J., … Zhu, T. (2018). Characterization of isoprene-derived secondary organic aerosols at a rural site in North China Plain with implications for anthropogenic pollution effects. *Scientific Reports*, *8*(1), 535. https://doi.org/10.1038/s41598-017-18983-7

Li, X., Feng, F., & Zeng, Y. (2014). Genome of betaproteobacterium *Caenimonas* sp. strain SL110 contains a coenzyme F420 biosynthesis gene clusters. *Journal of Microbiology and Biotechnology*, *24*(11), 1490–1494. https://doi.org/10.4014/jmb.1405.05039

Lichtenthaler, H. K., Schwender, J., Disch, A., & Rohmer, M. (1997). Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway. *FEBS Letters*, *400*(3), 271–274. https://doi.org/10.1016/S0014-5793(96)01404-4

Lindberg, P., Park, S., & Melis, A. (2010). Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism. *Metabolic Engineering*, *12*(1), 70–79. https://doi.org/10.1016/J.YMBEN.2009.10.001

Lindow, S. E. (1996). Role of immigration and other processes in determining epiphytic bacterial populations. In *Aerial Plant Surface Microbiology* (pp. 155–168). Boston, MA: Springer US. https://doi.org/10.1007/978-0-585-34164-4_10

Lindow, S. E., & Brandl, M. T. (2003). Microbiology of the phyllosphere. *Applied and Environmental Microbiology*, *69*(4), 1875–1883. https://doi.org/10.1128/AEM.69.4.1875-1883.2003

Liu, Y., Brito, J., Dorris, M. R., Rivera-Rios, J. C., Seco, R., Bates, K. H., … Martin, S. T. (2016). Isoprene photochemistry over the Amazon rainforest. *Proceedings of the National Academy of Sciences*, *113*(22), 6125–6130. https://doi.org/10.1073/pnas.1524136113

Logan, B. A., & Monson, R. K. (1999). Thermotolerance of leaf discs from four isoprene-emitting species is not enhanced by exposure to exogenous isoprene. *Plant Physiology*, *120*(July), 821–826. https://doi.org/10.1104/pp.120.3.821

Loivamäki, M., Louis, S., Cinege, G., Zimmer, I., Fischbach, R. J., & Schnitzler, J.-P. (2007). Circadian rhythms of isoprene biosynthesis in grey poplar leaves. *Plant Physiology*, *143*(1), 540–551. https://doi.org/10.1104/pp.106.092759

Loreto, F., Mannozzi, M., Maris, C., Nascetti, P., Ferranti, F., & Pasqualini, S. (2001). Ozone quenching properties of isoprene and its antioxidant role in leaves. *Plant Physiology*, *126*(3), 993–1000. https://doi.org/10.1104/pp.126.3.993

Loreto, F., & Schnitzler, J.-P. (2010). Abiotic stresses and induced BVOCs. *Trends in Plant*

*Science*, *15*(3), 154–166. https://doi.org/10.1016/J.TPLANTS.2009.12.006

Loreto, F., & Sharkey, T. D. (1990). A gas-exchange study of photosynthesis and isoprene emission in *Quercus rubra* L. *Planta*, *182*(4), 523–531. https://doi.org/10.1007/BF02341027

Loreto, F., & Sharkey, T. D. (1993). On the relationship between isoprene emission and photosynthetic metabolites under different environmental conditions. *Planta*, *189*(3), 420–424. https://doi.org/10.1007/BF00194440

Loreto, F., & Velikova, V. (2001). Isoprene Produced by Leaves Protects the Photosynthetic Apparatus against Ozone Damage, Quenches Ozone Products, and Reduces Lipid Peroxidation of Cellular Membranes. *Plant Physiology*, *127*(4), 1781–1787. https://doi.org/10.1104/pp.010497

Madsen, E. L. (2006). The use of stable isotope probing techniques in bioreactor and field studies on bioremediation. *Current Opinion in Biotechnology*, *17*(1), 92–97. https://doi.org/10.1016/j.copbio.2005.12.004

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Mattes, T. E., Alexander, A. K., Richardson, P. M., Munk, A. C., Han, C. S., Stothard, P., & Coleman, N. V. (2008). The Genome of *Polaromonas* sp. Strain JS666: Insights into the evolution of a hydrocarbon-and xenobiotic-degrading bacterium, and features of relevance to biotechnology †. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, *74*(20), 6405–6416. https://doi.org/10.1128/AEM.00197-08

Mayrhofer, S., Teuber, M., Zimmer, I., Louis, S., Fischbach, R. J., & Schnitzler, J.-P. (2005). Diurnal and seasonal variation of isoprene biosynthesis-related genes in grey poplar leaves. *Plant Physiology*, *139*(1), 474–484. https://doi.org/10.1104/pp.105.066373

McLeod, M. P., Warren, R. L., Hsiao, W. W. L., Araki, N., Myhre, M., Fernandes, C., … Eltis, L. D. (2006). The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proceedings of the National Academy of Sciences*, *103*(42), 15582–15587. https://doi.org/10.1073/pnas.0607048103

Medori, M., Michelini, L., Nogues, I., Loreto, F., & Calfapietra, C. (2012). The impact of root temperature on photosynthesis and isoprene emission in three different plant species. *The Scientific World Journal*, *2012*, 1–10. https://doi.org/10.1100/2012/525827

Miller, B., Oschinski, C., & Zimmer, W. (2001). First isolation of an isoprene synthase gene

from poplar and successful expression of the gene in *Escherichia coli*. *Planta*, *213*(3), 483–487. https://doi.org/10.1007/s004250100557

Misztal, P.K., Owen, S.M., Guenther, A.B., Rasmussen, R. (2010). Large estragole fluxes from oil palms in Borneo. *Atmospheric Chemistry and Physics*. https://doi.org/10.1177/0964663912467814

Miwa, H., Ahmed, I., Yoon, J., Yokota, A., & Fujiwara, T. (2008). *Variovorax boronicumulans* sp. nov., a boron-accumulating bacterium isolated from soil. *International Journal of Systematic and Evolutionary Microbiology*, *58*(1), 286–289. https://doi.org/10.1099/ijs.0.65315-0

Monson, R. K., Grote, R., Niinemets, Ü., & Schnitzler, J. P. (2012). Modeling the isoprene emission rate from leaves. *New Phytologist*, *195*(3), 541–559. https://doi.org/10.1111/j.1469-8137.2012.04204.x

Monson, R. K., Jaeger, C. H., Adams, W. W., Driggers, E. M., Silver, G. M., & Fall, R. (1992). Relationships among isoprene emission rate, photosynthesis, and isoprene synthase activity as influenced by temperature. *Plant Physiology*, *98*(3), 1175–1180. https://doi.org/10.1104/pp.98.3.1175

Morris, C. E., & Kinkel, L. L. (2002). Fifty years of phyllosphere microbiology: significant contributions to research in related fields. *Phyllosphere Microbiology*, 365–375. Retrieved from https://ci.nii.ac.jp/naid/10019340794/

Murphy, G. P. (University of E. (2015). *Isoprene degradation in the terrestrial environment*. University of Essex.

Muyzer, G., De Waal, E. C., & Uitterlinden, A. G. (1993). Profiling of complex microbial populations by DGGE analysis of PCR amplified genes coding for 16S rRNA. *Appl Environ Microbiol*, *59*, 695–700.

Neufeld, J. D., Vohra, J., Dumont, M. G., Lueders, T., Manefield, M., Friedrich, M. W., & Murrell, C. J. (2007). DNA stable-isotope probing. *Nature Protocols*, *2*(4), 860–866. https://doi.org/10.1038/nprot.2007.109

Nguyen, T. B., Tyndall, G. S., Crounse, J. D., Teng, A. P., Bates, K. H., Schwantes, R. H., … Wennberg, P. O. (2016). Atmospheric fates of Criegee intermediates in the ozonolysis of isoprene. *Physical Chemistry Chemical Physics*, *18*(15), 10241–10254. https://doi.org/10.1039/c6cp00053c

Niinemets, Ü., Arneth, A., Kuhn, U., Monson, R. K., Peñuelas, J., & Staudt, M. (2010). The emission factor of volatile isoprenoids: Stress, acclimation, and developmental

responses. *Biogeosciences*, *7*(7), 2203–2223. https://doi.org/10.5194/bg-7-2203-2010

Niinemets, Ü., Monson, R. K., Arneth, A., Ciccioli, P., Kesselmeier, J., Kuhn, U., … Staudt, M. (2010). The leaf-level emission factor of volatile isoprenoids: Caveats, model algorithms, response shapes and scaling. *Biogeosciences*, *7*(6), 1809–1832. https://doi.org/10.5194/bg-7-1809-2010

Ocampo-Peñuela, N., Garcia-Ulloa, J., Ghazoul, J., & Etter, A. (2018). Quantifying impacts of oil palm expansion on Colombia's threatened biodiversity. *Biological Conservation*, *224*, 117–121. https://doi.org/10.1016/J.BIOCON.2018.05.024

Owens, C. R., Karceski, J. K., & Mattes, T. E. (2009). Gaseous alkene biotransformation and enantioselective epoxyalkane formation by *Nocardioides* sp. strain JS614. *Applied Microbiology and Biotechnology*, *84*(4), 685–692. https://doi.org/10.1007/s00253-009-2019-3

Pacifico, F., Harrison, S. P., Jones, C. D., & Sitch, S. (2009). Isoprene emissions and climate. *Atmospheric Environment*, *43*(39), 6121–6135. https://doi.org/10.1016/j.atmosenv.2009.09.002

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from. *Cold Spring Harbor Laboratory Press Method*, *1*, 1–31. https://doi.org/10.1101/gr.186072.114

Paterson, R. R. M., & Lima, N. (2018, January). Climate change affecting oil palm agronomy, and oil palm cultivation increasing climate change, require amelioration. *Ecology and Evolution*. Wiley-Blackwell. https://doi.org/10.1002/ece3.3610

Pegoraro, E., Rey, A., Greenberg, J., Harley, P., Grace, J., Malhi, Y., & Guenther, A. (2004). Effect of drought on isoprene emission rates from leaves of *Quercus virginiana* Mill. *Atmospheric Environment*, *38*(36), 6149–6156. https://doi.org/10.1016/J.ATMOSENV.2004.07.028

Peñuelas, J., & Llusià, J. (2003). BVOCs: Plant defense against climate warming? *Trends in Plant Science*, *8*(3), 105–109. https://doi.org/10.1016/S1360-1385(03)00008-6

Peñuelas, J., & Llusià, J. (2004). Plant VOC emissions: Making use of the unavoidable. *Trends in Ecology and Evolution*, *19*(8), 402–404. https://doi.org/10.1016/j.tree.2004.06.002

Peñuelas, J., Marino, G., Llusia, J., Morfopoulos, C., Farré-Armengol, G., & Filella, I. (2013). Photochemical reflectance index as an indirect estimator of foliar isoprenoid emissions at the ecosystem level. *Nature Communications*, *4*(1), 2604.

https://doi.org/10.1038/ncomms3604

Peñuelas, J., & Staudt, M. (2010). BVOCs and global change. *Trends in Plant Science*, *15*(3), 133–144. https://doi.org/10.1016/j.tplants.2009.12.005

Pierce, T., Geron, C., Bender, L., Dennis, R., Tonnesen, G., & Guenther, A. (1998). Influence of increased isoprene emissions on regional ozone modeling. *Journal of Geophysical Research Atmospheres*, *103*(D19), 25611–25629. https://doi.org/10.1029/98JD01804

Radajewski, S., Ineson, P., Parekh, N. R., & Murrell, J. C. (2000). Stable isotope probing as a tool in microbial ecology. *Nature*, *403*(February), 646–649. Retrieved from http://dx.doi.org/10.1038/35001054

Rasmussen, R. A. (1970). Isoprene: identified as a forest-type emission to the atmosphere. *Environmental Science and Technology*, *4*(8), 667–671. https://doi.org/10.1021/es60043a008

Rasmussen, R. A., & Hutton, R. S. (1972). Utilization of atmospheric organic volatiles as an energy source by microorganisms in the tropics. *Chemosphere*, *1*(1), 47–50. https://doi.org/10.1016/0045-6535(72)90012-4

Redford, A. J., Bowers, R. M., Knight, R., Linhart, Y., & Fierer, N. (2010). The ecology of the phyllosphere: Geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environmental Microbiology*, *12*(11), 2885–2893. https://doi.org/10.1111/j.1462-2920.2010.02258.x

Reimann, S., Calanca, P., & Hofer, P. (2000). The anthropogenic contribution to isoprene concentrations in a rural atmosphere. *Atmospheric Environment*, *34*(1), 109–115. https://doi.org/10.1016/S1352-2310(99)00285-X

Rohmer, M., Knani, M., Simonin, P., Sutter, B., & Sahm, H. (1993). Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *The Biochemical Journal*, *295 ( Pt 2*(2), 517–524. https://doi.org/10.1042/bj2950517

Romero, F. M., Marina, M., & Pieckenstain, F. L. (2014). The communities of tomato (Solanum lycopersicum L.) leaf endophytic bacteria, analyzed by 16S-ribosomal RNA gene pyrosequencing. *FEMS Microbiology Letters*, *351*(2), 187–194. https://doi.org/10.1111/1574-6968.12377

Rosenkranz, M., Pugh, T. A. M., Schnitzler, J.-P., & Arneth, A. (2015). Effect of land-use change and management on biogenic volatile organic compound emissions - selecting climate-smart cultivars. *Plant, Cell & Environment*, *38*(9), 1896–1912. https://doi.org/10.1111/pce.12453

Sambrook, J., & Russell, D. (2001). *Molecular Cloning: A laboratory manual*. https://doi.org/10.3109/1061186X.2013.832767

Sanadze, G. A. (2004). Biogenic isoprene (a review). *Russian Journal of Plant Physiology*. https://doi.org/10.1023/B:RUPP.0000047821.63354.a4

Schnitzler, J.-P., & Nouvellon, Y. (2018). Vegetation - atmosphere interactions – the crucial role of tropospheric ozone. Retrieved from http://agritrop.cirad.fr/586929/

Schöller, C., Molin, S., & Wilkins, K. (1997). Volatile metabolites from some gram-negative bacteria. *Chemosphere*, *35*(7), 1487–1495. https://doi.org/10.1016/S0045-6535(97)00209-9

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., & Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade- specific marker genes. *Nat. Methods*, *9*(8). https://doi.org/10.1038/Nmeth.2066

Sharkey, T. D., & Loreto, F. (1993). Water stress, temperature, and light effects on the capacity for isoprene emission and photosynthesis of kudzu leaves. *Oecologia*, *95*(3), 328–333. https://doi.org/10.1007/BF00320984

Sharkey, T. D., & Monson, R. K. (2014). The future of isoprene emission from leaves, canopies and landscapes. *Plant, Cell and Environment*, *37*(8), 1727–1740. https://doi.org/10.1111/pce.12289

Sharkey, T. D., & Monson, R. K. (2017). Isoprene research – 60 years later, the biology is still enigmatic. *Plant Cell and Environment*, *40*(9), 1671–1678. https://doi.org/10.1111/pce.12930

Sharkey, T. D., & Singsaas, E. L. (1995). Why plants emit isoprene. *Nature*. https://doi.org/10.1038/374769a0

Sharkey, T. D., Wiberley, A. E., & Donohue, A. R. (2008). Isoprene emission from plants: Why and how. *Annals of Botany*, *101*(1), 5–18. https://doi.org/10.1093/aob/mcm240

Sharkey, T. D., & Yeh, S. (2001). Isoprene emission from plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, *52*(1), 407–436. https://doi.org/10.1146/annurev.arplant.52.1.407

Shennan, J. L. (2006, March). Utilisation of C2-C4 gaseous hydrocarbons and isoprene by microorganisms. *Journal of Chemical Technology and Biotechnology*. https://doi.org/10.1002/jctb.1388

Silver, G. M., & Fall, R. (1991). Enzymatic synthesis of isoprene from dimethylallyl

diphosphate in aspen leaf extracts. *Plant Physiology*, *97*(4), 1588–1591. https://doi.org/10.1104/pp.97.4.1588

Singleton, D. R., Powell, S. N., Sangaiah, R., Gold, A., Ball, L. M., Aitken, M. D., … Aitken, M. D. (2005). Stable-Isotope Probing of bacteria capable of degrading salicylate, naphthalene, or phenanthrene in a bioreactor treating contaminated soil, *71*(3), 1202–1209. https://doi.org/10.1128/AEM.71.3.1202

Singsaas, L., Lerdau, M., Winter, K., & Sharkey, T. D. (1997). lsoprene lncreases Thermotolerance of 1994 ). Field-based measurements were made with a field-. *Plant Physiology*, (1997), 1413–1420.

Small, F., & Ensign, S. (1997). Alkene monooxygenase from *Xanthobacter* strain Py2. *Journal of Biological Chemistry*, *272*(40), 24913–24920. Retrieved from http://www.jbc.org/content/272/40/24913.short

Smith, D., Španěl, P., Enderby, B., Lenney, W., Turner, C., & Davies, S. J. (2010). Isoprene levels in the exhaled breath of 200 healthy pupils within the age range 7–18 years studied using SIFT-MS. *Journal of Breath Research*, *4*(1), 017101. https://doi.org/10.1088/1752-7155/4/1/017101

Sotres, A., Tey, L., Bonmatí, A., & Viñas, M. (2016). Microbial community dynamics in continuous microbial fuel cells fed with synthetic wastewater and pig slurry. *Bioelectrochemistry*, *111*, 70–82. https://doi.org/10.1016/j.bioelechem.2016.04.007

Srivastva, N., Shukla, A. K., Singh, R. S., Upadhyay, S. N., & Dubey, S. K. (2015). Characterization of bacterial isolates from rubber dump site and their use in biodegradation of isoprene in batch and continuous bioreactors. *Bioresource Technology*, *188*, 84–91. https://doi.org/10.1016/j.biortech.2015.01.014

Srivastva, N., Singh, A., Bhardwaj, Y., & Dubey, S. K. (2018). Biotechnological potential for degradation of isoprene: a review. *Critical Reviews in Biotechnology*, *38*(4), 587–599. https://doi.org/10.1080/07388551.2017.1379467

Srivastva, N., Vishwakarma, P., Bhardwaj, Y., Singh, A., Manjunath, K., & Dubey, S. K. (2017). Kinetic and molecular analyses reveal isoprene degradation potential of *Methylobacterium* sp. *Bioresource Technology*, *242*, 87–91. https://doi.org/10.1016/j.biortech.2017.02.002

Surratt, J. D., Chan, A. W. H., Eddingsaas, N. C., Chan, M., Loza, C. L., Kwan, A. J., … Seinfeld, J. H. (2010). Reactive intermediates revealed in secondary organic aerosol formation from isoprene. *Proceedings of the National Academy of Sciences*, *107*(15), 6640–6645.

https://doi.org/10.1073/pnas.0911114107

Surratt, J. D., Murphy, S. M., Kroll, J. H., Ng, N. L., Hildebrandt, L., Sorooshian, A., … Seinfeld, J. H. (2006). Chemical composition of secondary organic aerosol formed from the photooxidation of isoprene. *Journal of Physical Chemistry A*, *110*(31), 9665–9690. https://doi.org/10.1021/jp061734m

Tang, K., Zhang, Y., Lin, D., Han, Y., Chen, C. T. A., Wang, D., … Jiao, N. (2018). Cultivation-independent and cultivation-dependent analysis of microbes in the shallow-sea hydrothermal system off Kueishantao Island, Taiwan: Unmasking heterotrophic bacterial diversity and functional capacity. *Frontiers in Microbiology*, *9*, 1–15. https://doi.org/10.3389/fmicb.2018.00279

Taraborrelli, D., Lawrence, M. G., Crowley, J. N., Dillon, T. J., Gromov, S., Groß, C. B. M., … Lelieveld, J. (2012). Hydroxyl radical buffered by isoprene oxidation over tropical forests. *Nature Geoscience*, *5*(3), 190–193. https://doi.org/10.1038/ngeo1405

Taucher, J., Hansel, A., Jordan, A., Fall, R., Futrell, J. H., & Lindinger, W. (1997). Detection of isoprene in expired air from human subjects using proton-transfer-reaction mass spectrometry. *Rapid Communications in Mass Spectrometry*, *11*(11), 1230–1234. https://doi.org/10.1002/(SICI)1097-0231(199707)11:11<1230::AID-RCM3>3.0.CO;2-Z

Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics*. https://doi.org/10.1002/0471250953.bi0203s00

Turner, T. R., James, E. K., & Poole, P. S. (2013). The plant microbiome. *Genome Biology*, *14*(6), 209. https://doi.org/10.1186/gb-2013-14-6-209

Van Beilen, J. B., & Funhoff, E. G. (2007). Alkane hydroxylases involved in microbial alkane degradation. *Applied Microbiology and Biotechnology*, *74*(1), 13–21. https://doi.org/10.1007/s00253-006-0748-0

Van Ginkel, C. G., De Jong, E., Tilanus, J. W. R., & De Bont, J. A. M. (1987). Microbial oxidation of isoprene, a biogenic foliage volatile and of 1,3-butadiene, an anthropogenic gas. *FEMS Microbiology Letters*, *45*(5), 275–279. https://doi.org/10.1016/0378-1097(87)90004-8

van Ginkel, C. G., Welten, H. G., & de Bont, J. a. (1987). Oxidation of gaseous and volatile hydrocarbons by selected alkene-utilizing bacteria. *Applied and Environmental Microbiology*, *53*(12), 2903–2907.

van Ginkel, C. G., Welten, H. G. J., & de Bont, J. A. M. (1986). Epoxidation of alkenes by

alkene-grown Xanthobacter spp. *Applied Microbiology and Biotechnology*, *24*(4), 334–337. https://doi.org/10.1007/BF00257059

van Hylckama Vlieg, J. E., Kingma, J., van den Wijngaard, a J., & Janssen, D. B. (1998). A glutathione S-transferase with activity towards cis-1, 2-dichloroepoxyethane is involved in isoprene utilization by *Rhodococcus* sp. strain AD45. *Applied and Environmental Microbiology*, *64*(8), 2800–2805.

van Hylckama Vlieg, J. E. T., Leemhuis, H., Jeffrey, H., Spelberg, L., Janssen, D. B., Ad, S., … Janssen, D. B. (2000). Characterization of the gene cluster involved in isoprene metabolism in *Rhodococcus* sp . Strain AD45. *Journal of Bacteriology*, *187*(7), 1956–1963. https://doi.org/10.1128/JB.182.7.1956-1963.2000.Updated

Van Hylckama Vlieg, J. E. T., Leemhuis, H., Lutje Spelberg, J. H., & Janssen, D. B. (2000). Characterization of the gene cluster involved in isoprene metabolism in *Rhodococcus* sp. strain AD45. *Journal of Bacteriology*, *182*(7), 1956–1963. https://doi.org/10.1128/JB.182.7.1956-1963.2000

Velikova, V., Sharkey, T. D., & Loreto, F. (2012). Stabilization of thylakoid membranes in isoprene-emitting plants reduces formation of reactive oxygen species. *Plant Signaling & Behavior*, *7*(1), 139–141. https://doi.org/10.4161/psb.7.1.18521

Velikova, V., Varkonyi, Z., Szabo, M., Maslenkova, L., Nogues, I., Kovacs, L., … Loreto, F. (2011). Increased thermostability of thylakoid membranes in isoprene-emitting leaves probed with three biophysical techniques. *Plant Physiology*, *157*(2), 905–916. https://doi.org/10.1104/pp.111.182519

Vorholt, J. A. (2012). Microbial life in the phyllosphere. *Nature Reviews Microbiology*, *10*(12), 828–840. https://doi.org/10.1038/nrmicro2910

Wagner, W. P., Nemecek-marshall, M., & Fall, R. A. Y. (1999). Three distinct phases of isoprene formation during growth and sporulation of *Bacillus subtilis*. *J Bacteriol*, *181*(15), 4700–4703.

Wang, B., Shugart, H. H., Shuman, J. K., & Lerdau, M. T. (2016). Forests and ozone: Productivity, carbon storage, and feedbacks. *Scientific Reports*, *6*(1), 22133. https://doi.org/10.1038/srep22133

Wang, S., Li, R., Yi, X., Fang, T., Yang, J., & Bae, H. J. (2016). Isoprene production on enzymatic hydrolysate of peanut hull using different pretreatment methods. *BioMed Research International*, *2016*(Usda 2015). https://doi.org/10.1155/2016/4342892

Way, D. A., Schnitzler, J. P., Monson, R. K., & Jackson, R. B. (2011). Enhanced isoprene-

related tolerance of heat- and light-stressed photosynthesis at low, but not high, CO2 concentrations. *Oecologia*, *166*(1), 273–282. https://doi.org/10.1007/s00442-011-1947-7

Weise, S. E., Li, Z., Sutter, A. E., Corrion, A., Banerjee, A., & Sharkey, T. D. (2013). Measuring dimethylallyl diphosphate available for isoprene synthesis. *Analytical Biochemistry*, *435*(1), 27–34. https://doi.org/10.1016/j.ab.2012.11.031

Whipps, J. M., Hand, P., Pink, D., & Bending, G. D. (2008, December). Phyllosphere microbiology with special reference to diversity and plant genotype. *Journal of Applied Microbiology*. Wiley/Blackwell (10.1111). https://doi.org/10.1111/j.1365-2672.2008.03906.x

Whited, G. M., Feher, F. J., Benko, D. A., Cervin, M. A., Chotani, G. K., McAuliffe, J. C., … Sanford, K. J. (2010). Technology update: Development of a gas-phase bioprocess for isoprene-monomer production using metabolic pathway engineering. *Industrial Biotechnology*, *6*(3), 152–163. https://doi.org/10.1089/ind.2010.6.152

Whited, G. M., & Gibson, D. T. (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), 3010–3016. https://doi.org/10.1128/jb.173.9.3010-3016.1991

Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., … Soyer, O. S. (2016). Challenges in microbial ecology: Building predictive understanding of community function and dynamics. *ISME Journal*, *10*(11), 2557–2568. https://doi.org/10.1038/ismej.2016.45

Widdows, J., & Brinsley, M. (2002). Impact of biotic and abiotic processes on sediment dynamics and the consequences to the structure and functioning of the intertidal zone. *Journal of Sea Research*, *48*(2), 143–156. https://doi.org/10.1016/S1385-1101(02)00148-X

Wiedinmyer, C., Greenberg, J., Guenther, A., Hopkins, B., Baker, K., Geron, C., … Janssen, M. (2005). Ozarks Isoprene Experiment (OZIE): Measurements and modeling of the "isoprene volcano." *Journal of Geophysical Research D: Atmospheres*, *110*(18), 1–17. https://doi.org/10.1029/2005JD005800

Wildermuth, M. C., & Fall, R. (1996). Light-dependent isoprene emission (characterization of a thylakoid-bound isoprene synthase in *Salix discolor* chloroplasts). *Plant Physiology*, *112*(1), 171–182. https://doi.org/10.1104/pp.112.1.171

Wilkinson, M. J., Owen, S. M., Possell, M., Hartwell, J., Gould, P., Hall, A., … Nicholas Hewitt, C. (2006). Circadian control of isoprene emissions from oil palm (*Elaeis guineensis*). *Plant Journal*, *47*(6), 960–968. https://doi.org/10.1111/j.1365-313X.2006.02847.x

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3). https://doi.org/10.1186/gb-2014-15-3-r46

Xie, S., Sun, W., Luo, C., & Cupples, A. M. (2011). Novel aerobic benzene degrading microorganisms identified in three soils by stable isotope probing. *Biodegradation*, *22*(1), 71–81. https://doi.org/10.1007/s10532-010-9377-5

Yang, C. H., Crowley, D. E., Borneman, J., & Keen, N. T. (2001). Microbial phyllosphere populations are more complex than previously realized. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(7), 3889–3894. https://doi.org/10.1073/pnas.051633898

Yen, K. M., Karl, M. R., Blatt, L. M., Simon, M. J., Winter, R. B., Fausset, P. R., … Chen, K. K. (1991). Cloning and characterization of a *Pseudomonas mendocina* KR1 gene cluster encoding toluene-4-monooxygenase. *Journal of Bacteriology*, *173*(17), 5315–5327. https://doi.org/10.1128/jb.173.17.5315-5327.1991

Yoon, J.-H. (2006). *Variovorax dokdonensis* sp. nov., isolated from soil. *International Journal of Systematic and Evolutionary Microbiology*, *56*(4), 811–814. https://doi.org/10.1099/ijs.0.64070-0

Zenone, T., Hendriks, C., Brilli, F., Fransen, E., Gioli, B., Portillo-Estrada, M., … Ceulemans, R. (2016). Interaction between isoprene and ozone fluxes in a poplar plantation and its impact on air quality at the European level. *Scientific Reports*, *6*(1), 32676. https://doi.org/10.1038/srep32676

Zhou, C., Li, Z., Wiberley-Bradford, A. E., Weise, S. E., & Sharkey, T. D. (2013). Isopentenyl diphosphate and dimethylallyl diphosphate/isopentenyl diphosphate ratio measured with recombinant isopentenyl diphosphate isomerase and isoprene synthase. *Analytical Biochemistry*, *440*(2), 130–136. https://doi.org/10.1016/j.ab.2013.05.028

Zhou, N. Y., Jenkins, A., Chan Kwo Chion, C. K. N., & Leak, D. J. (1999). The alkene monooxygenase from *Xanthobacter* strain Py2 is closely related to aromatic monooxygenases and catalyzes aromatic monohydroxylation of benzene, toluene, and phenol. *Applied and Environmental Microbiology*, *65*(4), 1589–1595.

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., & Wishart, D. S. (2011). PHAST: A Fast Phage

Search Tool. *Nucleic Acids Research*, *39*(SUPPL. 2), W347–W352. https://doi.org/10.1093/nar/gkr485

Zou, Y., Hu, M., Lv, Y., Wang, Y., Song, H., & Yuan, Y. J. (2013). Enhancement of 2-keto-gulonic acid yield by serial subcultivation of co-cultures of *Bacillus cereus* and *Ketogulonigenium vulgare*. *Bioresource Technology*, *132*, 370–373. https://doi.org/10.1016/j.biortech.2012.10.151

# Supplementary Tables

**Supplementary Table 1**. Time course for OPL-SIP in hours.

| | $^{12}$C Replicate 1 | $^{12}$C Replicate 2 | $^{12}$C Replicate 3 | $^{13}$C Replicate 1 | $^{13}$C Replicate 2 | $^{13}$C Replicate 3 |
|---|---|---|---|---|---|---|
| $T_1$ | 214 | 214 | 214 | 226 | 214 | 214 |
| $T_2$ | 285 | 274 | 274 | 285.5 | 281 | 274 |
| $T_3$ | 311.5 | 302 | 302 | 307 | 307 | 302 |

**Supplementary Table 2** Percentage of heavy and light DNA recovered after 6, 7 and 8 days ($T_1$ $T_2$ and $T_3$, respectively) of enrichment with isoprene ($^{12}C$ and $^{13}C$-labelled) of OPS samples.

| | % heavy DNA* | | | % light DNA* | | |
|---|---|---|---|---|---|---|
| **OPS Sample** | **$T_1$** | **$T_2$** | **$T_3$** | **$T_1$** | **$T_2$** | **$T_3$** |
| $^{12}C$-Replicate 1 | 0.5 | 0.7 | 0.4 | 86.1 | 90.2 | 88.3 |
| $^{12}C$-Replicate 2 | 0.6 | 0.8 | 0.5 | 84.7 | 81.6 | 84.4 |
| $^{12}C$-Replicate 3 | 0.2 | 0.6 | 0.2 | 87.8 | 88.4 | 86.9 |
| $^{13}C$-Replicate 1 | 2.0 | 2.4 | 3.3 | 83.6 | 83.7 | 83.6 |
| $^{13}C$-Replicate 2 | 2.1 | 6.7 | 8.4 | 84.1 | 76.5 | 74.2 |
| $^{13}C$-Replicate 3 | 1.8 | 6.6 | 7.8 | 81.1 | 89.2 | 80.6 |

**Supplementary Table 3** Percentage of heavy and light DNA recovered after 10, 13 and 14 days ($T_1$ $T_2$ and $T_3$, respectively) of enrichment with isoprene ($^{12}C$ and $^{13}C$-labelled) of OPL samples.

| | % heavy DNA | | | % light DNA | | |
|---|---|---|---|---|---|---|
| **OPL Sample** | **$T_1$** | **$T_2$** | **$T_3$** | **$T_1$** | **$T_2$** | **$T_3$** |
| $^{12}C$-Replicate 1 | 0.69 | 0.44 | 0.28 | 85.73 | 85.79 | 88.58 |
| $^{12}C$-Replicate 2 | 0.70 | 0.23 | 0.46 | 84.90 | 89.37 | 87.36 |
| $^{12}C$-Replicate 3 | 0.61 | 0.16 | 0.55 | 87.02 | 85.97 | 89.30 |
| $^{13}C$-Replicate 1 | 33.91 | 51.04 | 48.83 | 31.48 | 25.28 | 37.50 |
| $^{13}C$-Replicate 2 | 28.22 | 51.06 | 49.65 | 65.11 | 36.02 | 38.93 |
| $^{13}C$-Replicate 3 | 22.28 | 40.54 | 40.99 | 60.77 | 45.88 | 44.12 |

**Supplementary Table 4** Total DNA recovered after DNA fractionation from enriched OPS samples.

| | Total DNA (μg) | | |
|---|---|---|---|
| **Sample** | **$T_1$ (6 days)** | **$T_2$ (7 days)** | **$T_3$ (8 days)** |
| $^{12}$C-Replicate 1 | 2.7 | 2.7 | 2.6 |
| $^{12}$C-Replicate 2 | 2.3 | 3.0 | 2.4 |
| $^{12}$C-Replicate 3 | 2.4 | 3.6 | 2.3 |
| $^{13}$C-Replicate 1 | 2.2 | 3.4 | 2.8 |
| $^{13}$C-Replicate 2 | 2.2 | 3.1 | 2.9 |
| $^{13}$C-Replicate 3 | 2.8 | 3.2 | 2.0 |

**Supplementary Table 5** Total DNA recovered after DNA fractionation from enriched OPL samples.

| | Total DNA (ng) | | |
|---|---|---|---|
| **Sample** | **$T_1$ (10 days)** | **$T_2$ (13 days)** | **$T_3$ (14 days)** |
| $^{12}$C-Replicate 1 | 890.6 | 440.9 | 613.0 |
| $^{12}$C-Replicate 2 | 904.2 | 394.0 | 815.3 |
| $^{12}$C-Replicate 3 | 929.8 | 412.1 | 974.9 |
| $^{13}$C-Replicate 1 | 308.7 | 364.6 | 622.5 |
| $^{13}$C-Replicate 2 | 952.0 | 448.9 | 695.5 |
| $^{13}$C-Replicate 3 | 1151.3 | 331.0 | 718.0 |

**Supplementary Table 6A** Prophage sequences found in the *Rhodococcus* sp. strain WS4 isolate DNA.

| Isolate | Region (contig) | Region Length (Kbp) | Phage completeness | Score * | Number of coding sequences | Region position | Possible phage | % GC |
|---|---|---|---|---|---|---|---|---|
| *R* sp.WS4 | Node 8 | 14 | Incomplete | 20 | 12 | 62816-76829 | PHAGE_Gordon_Nyceirae_NC_031004 | 67.5 |

*****score given by the keywords in the coding regions (transposase, integrase, tail, fiber, etc), number of phage related proteins constitute >70% of proteins in the region (Y. Zhou et al., 2011).

**Supplementary Table 6B** Blast hits for coding sequences found in the prophage region (*Rhodococcus* sp. strain WS4). Phage-like sequences in yellow.

| Coding region | Coding region position | Blast hit | E-value |
|---|---|---|---|
| 1 | 62816-66700 | PHAGE_Entero_933W_NC_000924:putative tail fiber protein; PP_0068; phage (gi9632525) | 3e-05 |
| 2 | 66748-67329 | PHAGE_Sinorh_phiM9_NC_028676: DNA-directed RNA polymerase specialized sigma subunit; PP_0069; phage(gi966199376) | 2e-08 |
| 3 | 67408-68148 | hypothetical protein RHA1_ro03646 [Rhodococcus jostii RHA1]. gi\|111020636\|ref\|YP_703607.1\|;PP_00070 | 3e-113 |
| 4 | 68152-68724 | TetR family transcriptional regulator [Rhodococcus jostii RHA1]. gi\|111020636\|ref\|YP_703608.1\|;PP_00071 | 3e-93 |
| 5 | 68814-69785 | PHAGE_Synech_ACG_2014f_NC_026927: hypothetical protein; PP00072; phage(gi100233) | 6e-06 |
| 6 | 69823-71010 | NADH-dependent flavin oxidoreductase [Rhodococcus jostii RHA1]. gi\|111020638\|ref\|YP_703610.1\|;PP_00073 | 0.0 |
| 7 | 71122-72105 | PHAGE_Bacill_G_NC023719: gp344;PP_00074; phage(gi593777800) | 5e-67 |
| 8 | 72150-72473 | PHAGE_Dinoro_DFL12phi1_NC_024367: putative host-like protein; PP_00075; phage(gi658311036) | 7e-11 |
| 9 | 72638-73837 | PHAGE_Bacill_AR9_NC_031039: hypothetical protein; PP_00076; phage(gi100271) | 1e-13 |
| 10 | 73872-74570 | hypothetical protein RHA1_ro03653 [Rhodococcus jostii RHA1]. gi\|111020642\|ref\|YP_703614.1\|; PP_00077 | 7e-129 |

| | | | |
|---|---|---|---|
| 11 | 74728-75819 | PHAGE_Rhizob_RR1_B_NC_021557: chromosome partitioning protein parB; PP_00078; phage(gi514231140) | 1e-08 |
| 12 | 75816-76829 | PHAGE_Natria_PhiCh1_NC_004084: putative plasmid partitioning protein Soj; PP_00079; phage(gi22091150) | 6e-25 |

**Supplementary Table 7** One-on-one protein sequence comparison between isoprene-degrading bacteria (IsoBCDEFHIJ and AldH_1).

| 100 | 99-90 | 89-80 | 79-70 | 69-60 | 59-50 | 49-35 | 34-0 |
|---|---|---|---|---|---|---|---|

| IsoB | *P m* KR1 | *R* AD45 | *G p* i37 | *R* WS7 | *N* WS12 | *G* OPL2 | *V* WS11 | *S* OPL5 |
|---|---|---|---|---|---|---|---|---|
| *P m* KR1 | 100 | 33 | 49 | 36 | 30 | 52 | 38 | 33 |
| *R* AD45 | 33 | 100 | 57 | 77 | 56 | 66 | 54 | 44 |
| *G p* i37 | 49 | 57 | 100 | 58 | 56 | 83 | 45 | 48 |
| *R* WS7 | 36 | 77 | 58 | 100 | 50 | 60 | 56 | 43 |
| *N* WS12 | 30 | 56 | 56 | 50 | 100 | 57 | 39 | 40 |
| *G* OPL2 | 52 | 66 | 83 | 60 | 57 | 100 | 46 | 45 |
| *V* WS11 | 38 | 54 | 45 | 56 | 39 | 46 | 100 | 52 |
| *S* OPL5 | 33 | 44 | 48 | 43 | 40 | 45 | 52 | 100 |

| IsoC | *P m* KR1 | *R* AD45 | *G p* i37 | *R* WS7 | *N* WS12 | *G* OPL2 | *V* WS11 | *S* OPL5 |
|---|---|---|---|---|---|---|---|---|
| *P m* KR1 | 100 | 45 | 46 | 42 | 43 | 47 | 40 | 50 |
| *R* AD45 | 45 | 100 | 78 | 82 | 62 | 81 | 49 | 54 |
| *G p* i37 | 46 | 78 | 100 | 84 | 78 | 88 | 56 | 53 |
| *R* WS7 | 42 | 82 | 84 | 100 | 64 | 89 | 50 | 54 |
| *N* WS12 | 43 | 62 | 78 | 64 | 100 | 67 | 52 | 56 |
| *G* OPL2 | 47 | 81 | 88 | 89 | 67 | 100 | 57 | 57 |
| *V* WS11 | 40 | 49 | 56 | 50 | 52 | 57 | 100 | 50 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *S* OPL5 | 50 | 54 | 53 | 54 | 56 | 57 | 50 | 100 |

| IsoD | *P m* KR1 | *R* AD45 | *G p* i37 | *R* WS7 | *N* WS12 | *G* OPL2 | *V* WS11 | *S* OPL5 |
|---|---|---|---|---|---|---|---|---|
| *P m* KR1 | 100 | 44 | 42 | 41 | 38 | 40 | 39 | 38 |
| *R* AD45 | 44 | 100 | 74 | 94 | 65 | 73 | 58 | 50 |
| *G p* i37 | 42 | 74 | 100 | 74 | 66 | 82 | 56 | 50 |
| *R* WS7 | 41 | 94 | 74 | 100 | 67 | 71 | 57 | 50 |
| *N* WS12 | 38 | 65 | 66 | 67 | 100 | 69 | 56 | 58 |
| *G* OPL2 | 40 | 73 | 82 | 71 | 69 | 100 | 61 | 52 |
| *V* WS11 | 39 | 58 | 56 | 57 | 56 | 61 | 100 | 56 |
| *S* OPL5 | 38 | 50 | 50 | 50 | 58 | 52 | 56 | 100 |

| IsoE | *P m* KR1 | *R* AD45 | *G p* i37 | *R* WS7 | *N* WS12 | *G* OPL2 | *V* WS11 | *S* OPL5 |
|---|---|---|---|---|---|---|---|---|
| *P m* KR1 | 100 | 37 | 39 | 35 | 38 | 34 | 39 | 33 |
| *R* AD45 | 37 | 100 | 71 | 83 | 61 | 68 | 53 | 53 |
| *G p* i37 | 39 | 71 | 100 | 71 | 60 | 79 | 53 | 51 |
| *R* WS7 | 35 | 83 | 71 | 100 | 62 | 69 | 53 | 53 |
| *N* WS12 | 38 | 61 | 60 | 62 | 100 | 62 | 53 | 53 |
| *G* OPL2 | 34 | 68 | 79 | 69 | 62 | 100 | 53 | 51 |
| *V* WS11 | 39 | 53 | 53 | 53 | 53 | 53 | 100 | 57 |
| *S* OPL5 | 33 | 53 | 51 | 53 | 53 | 51 | 57 | 100 |

| IsoF | *P m* KR1 | *R* AD45 | *G p* i37 | *R* WS7 | *N* WS12 | *G* OPL2 | *V* WS11 | *S* OPL5 |
|---|---|---|---|---|---|---|---|---|
| *P m* KR1 | 100 | 30 | 30 | 25 | 29 | 30 | 34 | 35 |
| *R* AD45 | 30 | 100 | 60 | 76 | 49 | 61 | 41 | 44 |
| *G p* i37 | 30 | 60 | 100 | 59 | 53 | 69 | 45 | 45 |
| *R* WS7 | 25 | 76 | 59 | 100 | 47 | 61 | 38 | 40 |
| *N* WS12 | 29 | 49 | 53 | 47 | 100 | 51 | 39 | 40 |
| *G* OPL2 | 30 | 61 | 69 | 61 | 51 | 100 | 42 | 45 |
| *V* WS11 | 34 | 41 | 45 | 38 | 39 | 42 | 100 | 44 |
| *S* OPL5 | 35 | 44 | 45 | 40 | 40 | 45 | 44 | 100 |

| IsoI | *R* AD45 | *G p* i37 | *R* WS7 | *N* WS12 | *G* OPL2 | *V* WS11 | *S* OPL5 |
|---|---|---|---|---|---|---|---|
| *R* AD45 | 100 | 80 | 86 | 66 | 79 | 48 | 45 |
| *G p* i37 | 80 | 100 | 86 | 62 | 95 | 49 | 47 |
| *R* WS7 | 86 | 86 | 100 | 65 | 85 | 52 | 47 |
| *N* WS12 | 66 | 62 | 65 | 100 | 62 | 48 | 46 |
| *G* OPL2 | 79 | 95 | 85 | 62 | 100 | 50 | 48 |
| *V* WS11 | 48 | 49 | 52 | 48 | 50 | 100 | 53 |
| *S* OPL5 | 45 | 47 | 47 | 46 | 48 | 53 | 100 |

| IsoJ | *R* AD45 | *G p* i37 | *R* WS7 | *N* WS12 | *G* OPL2 | *V* WS11 | *S* OPL5 |
|---|---|---|---|---|---|---|---|
| *R* AD45 | 100 | 71 | 77 | 66 | 68 | 55 | 58 |
| *G p* i37 | 71 | 100 | 74 | 65 | 81 | 53 | 60 |
| *R* WS7 | 77 | 74 | 100 | 68 | 68 | 57 | 61 |
| *N* WS12 | 66 | 65 | 68 | 100 | 64 | 55 | 56 |
| *G* OPL2 | 68 | 81 | 68 | 64 | 100 | 56 | 60 |
| *V* WS11 | 55 | 53 | 57 | 55 | 56 | 100 | 52 |
| *S* OPL5 | 58 | 60 | 61 | 56 | 60 | 52 | 100 |

| AldH_1 | *R* WS7 | *N* WS12 | *G* OPL2 | *V* WS11 | *S* OPL5 |
|---|---|---|---|---|---|
| *R* WS7 | 100 | 63 | 74 | 52 | 50 |
| *N* WS12 | 63 | 100 | 64 | 53 | 52 |
| *G* OPL2 | 74 | 64 | 100 | 51 | 49 |
| *V* WS11 | 52 | 53 | 51 | 100 | 56 |
| *S* OPL5 | 50 | 52 | 49 | 56 | 100 |

**Supplementary Table 8** *Variovorax* sp. strain WS9 genome sequence analysis of the isoprene cluster compared to *Rhodococcus* sp. AD45. *Variovorax* sp. strain WS9 annotation was done using PROKKA.

| **Rhodococcus sp. AD45** | | **NCBI** | ***Variovorax* sp. strain WS9 PROKKA annotation** | | |
|---|---|---|---|---|---|
| **Gene** | **Description** | **% tbalstn identity** | **Gene** | **#** | **Description** |
| *gshB* | glutathione synthase | - | | 00612 | hypothetical protein |
| *aldH* | aldehyde dehydrogenase | - | *ompW* | 00613 | Outer membrane protein W precursor |
| *isoF* | isoprene MO reductase | 44% (JVH1*) | *carAd* | 00614 | Ferredoxin--NAD(P)(+) reductase |
| *isoE* | isoprene MO β subunit | 53% | *tmoE* | 00615 | Toluene-4-monooxygenase system protein E |
| *IsoD* | isoprene MO coupling protein | 55% | *tmoD* | 00616 | Toluene-4-monooxygenase system protein D |
| *isoC* | isoprene MO ferredoxin | 55% | *tmoC* | 00617 | Toluene-4-monooxygenase system ferredoxin subunit |
| *isoB* | isoprene MO γ subunit | 50% | *tmoB* | 00618 | Toluene-4-monooxygenase system protein B |
| *isoA* | isoprene MO α subunit | 71% | *tmoA* | 00619 | Toluene-4-monooxygenase system protein A |
| *aldH* | aldehyde dehydrogenase | 54% | *feaB_1* | 00620 | Phenylacetaldehyde dehydrogenase |
| *isoJ* | glutathione S-transferase | 55% | *yfcG_1* | 00621 | Disulfide-bond oxidoreductase |
| *isoI* | glutathione S-transferase | 50% | | 00622 | hypothetical protein |
| *isoH* | dehydrogenase | 62% | | 00623 | C-factor |
| *isoG* | CoA transferase | 61% | | 00624 | formyl-coenzyme A transferase |

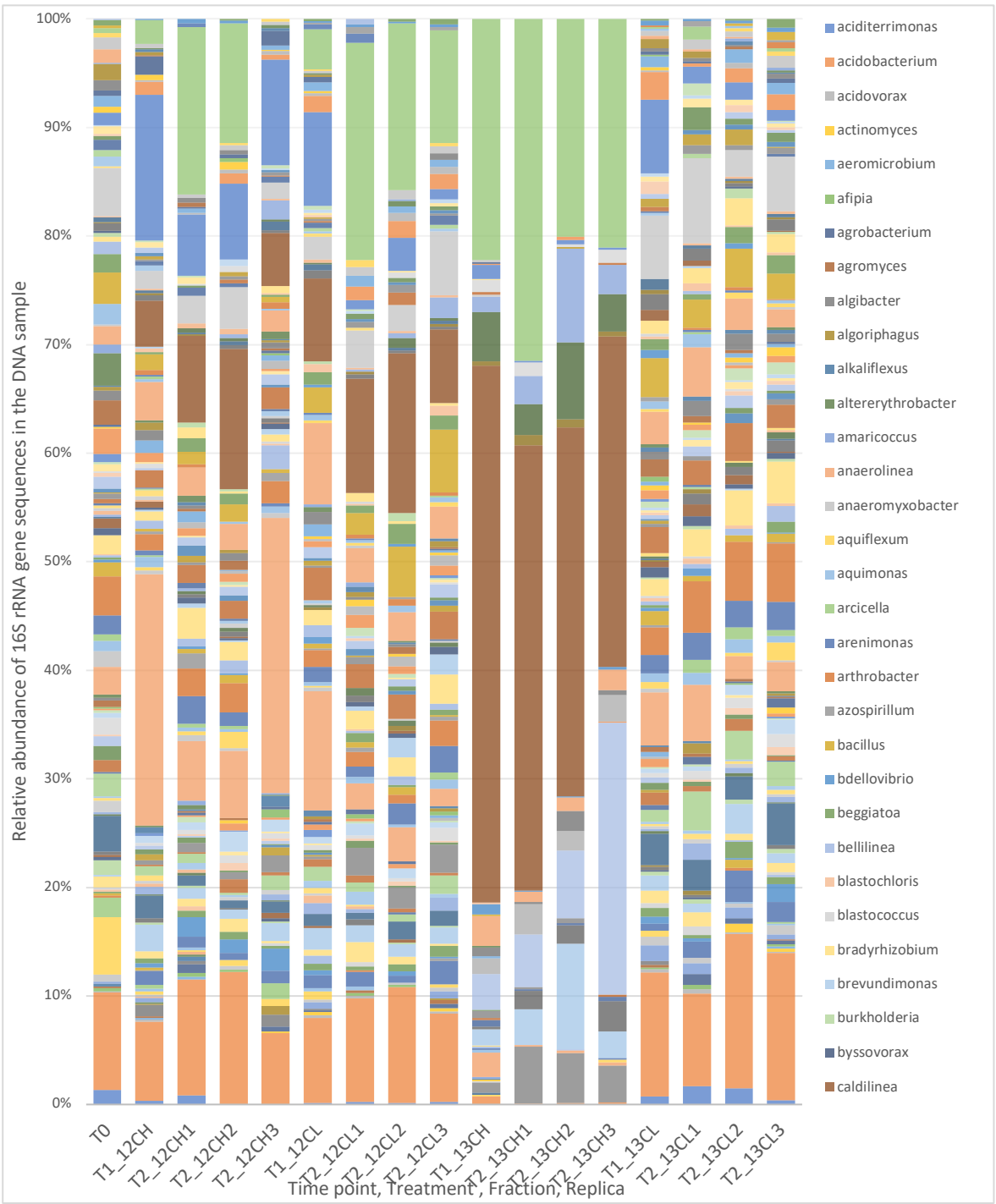
| | | | | | |
|---|---|---|---|---|---|
| *gshA* | glutamate cysteine ligase | - | *dmlR_2* | 00625 | HTH-type transcriptional regulator DmlR |
| *marR* | transcriptional regulator | - | *gcvA_3* | 00626 | Glycine cleavage system transcriptional activator |
| *marR* | transcriptional regulator | - | | 00627 | hypothetical protein |

**isoF* had a 44% identity to the *tmoF* gene in *Rhodococcus* sp. JHV1. All other results are compared to *Rhodococcus* sp. AD45

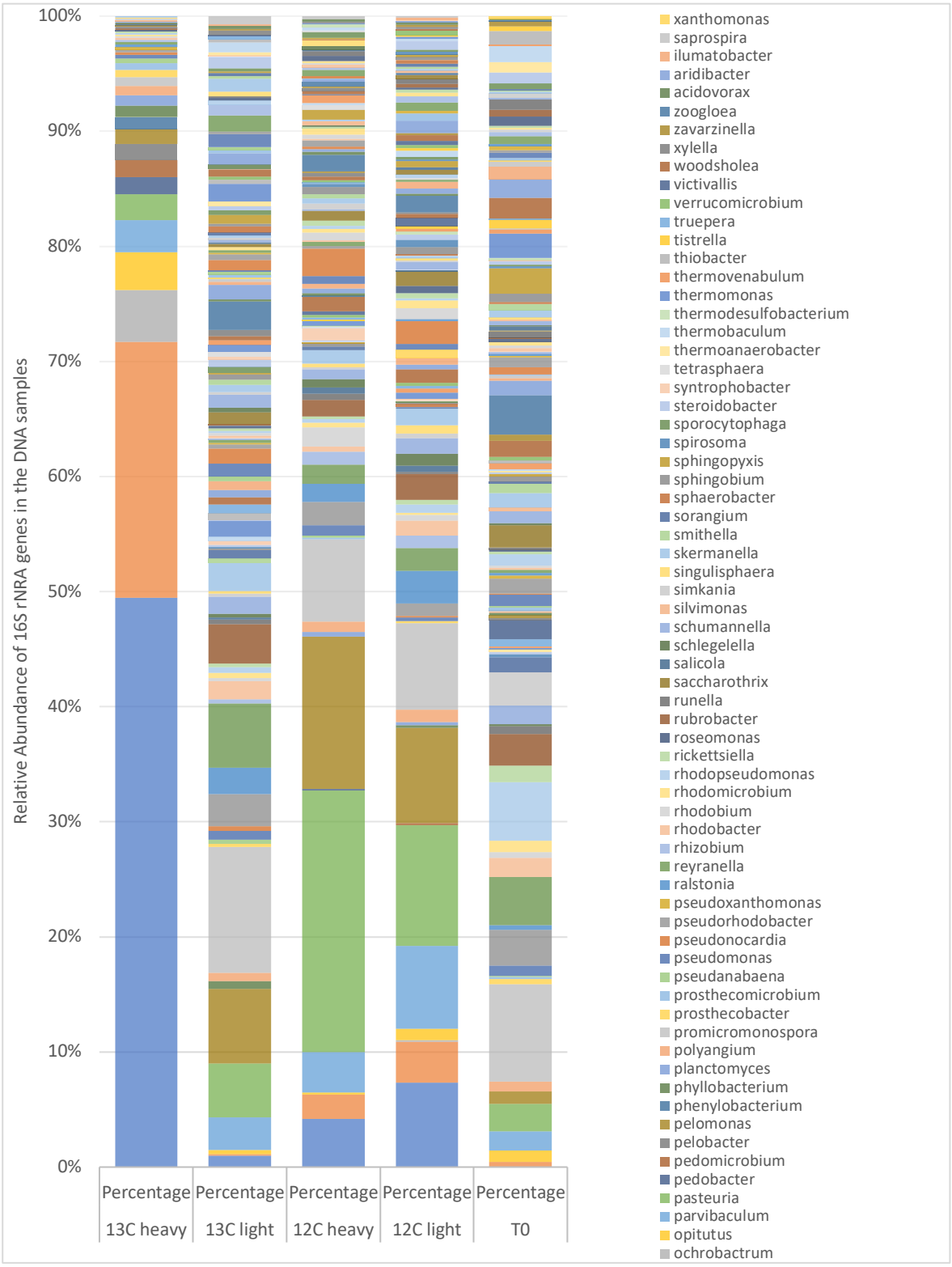
**Supplementary Table 9** Number of amino acids in each polypeptide sequence from the isoprene metabolic gene cluster. Sequences were extracted from genome annotation files from MicrobesNG. Used for protein to protein comparison.

| **Isoprene degrading bacterium** | **IsoA** | **IsoB** | **IsoC** | **IsoD** | **IsoE** | **IsoF** | **IsoG** | **IsoH** | **IsoI** | **IsoJ** | **AldH_1** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pseudomonas mendocina KR1 | 500 | 84 | 112 | 103 | 327 | 326 | - | - | - | - | - |
| Rhodococcus sp. AD45 | 514 | 94 | 114 | 110 | 342 | 345 | 405 | 226 | 238 | 233 | - |
| Rhodococcus opacus PD630 | 507 | 94 | 114 | 110 | 340 | 332 | 405 | 226 | 238 | 233 | 478 |
| Gordonia polyisoprenivorans i37 | 506 | 95 | 121 | 101 | 338 | 341 | 406 | 226 | 239 | 236 | - |
| *Mycobacterium* sp. strain AT1 | 511 | 94 | 114 | 108 | 342 | 340 | 401 | 226 | 238 | 234 | 452 |
| *Rhodococcus* sp. strain WS1 | 507 | 97 | 114 | 110 | 340 | 269 | 405 | 226 | 238 | 233 | 477 |
| *Rhodococcus* sp. strain WS4 | 503 | 94 | 114 | 110 | 340 | 345 | 405 | 226 | 238 | 233 | - |
| *Variovorax* sp. strain WS9 | 497 | 88 | 119 | 104 | 343 | 342 | 406 | 226 | 238 | 239 | 475 |
| *Variovorax* sp. strain WS11 | 500 | 88 | 111 | 104 | 344 | 344 | 402 | 227 | 238 | 242 | 475 |
| *Nocardioides* sp. strain WS12 | 498 | 89 | 118 | 110 | 340 | 355 | 404 | 226 | 237 | 250 | 477 |
| *Gordonia* sp. strain OPL2 | 496 | 95 | 109 | 106 | 340 | 342 | 405 | 227 | 239 | 236 | 478 |
| *Sphingopyxis* sp. strain OPL5 | 497 | 87 | 112 | 106 | 367 | 336 | 405 | 226 | 244 | | |

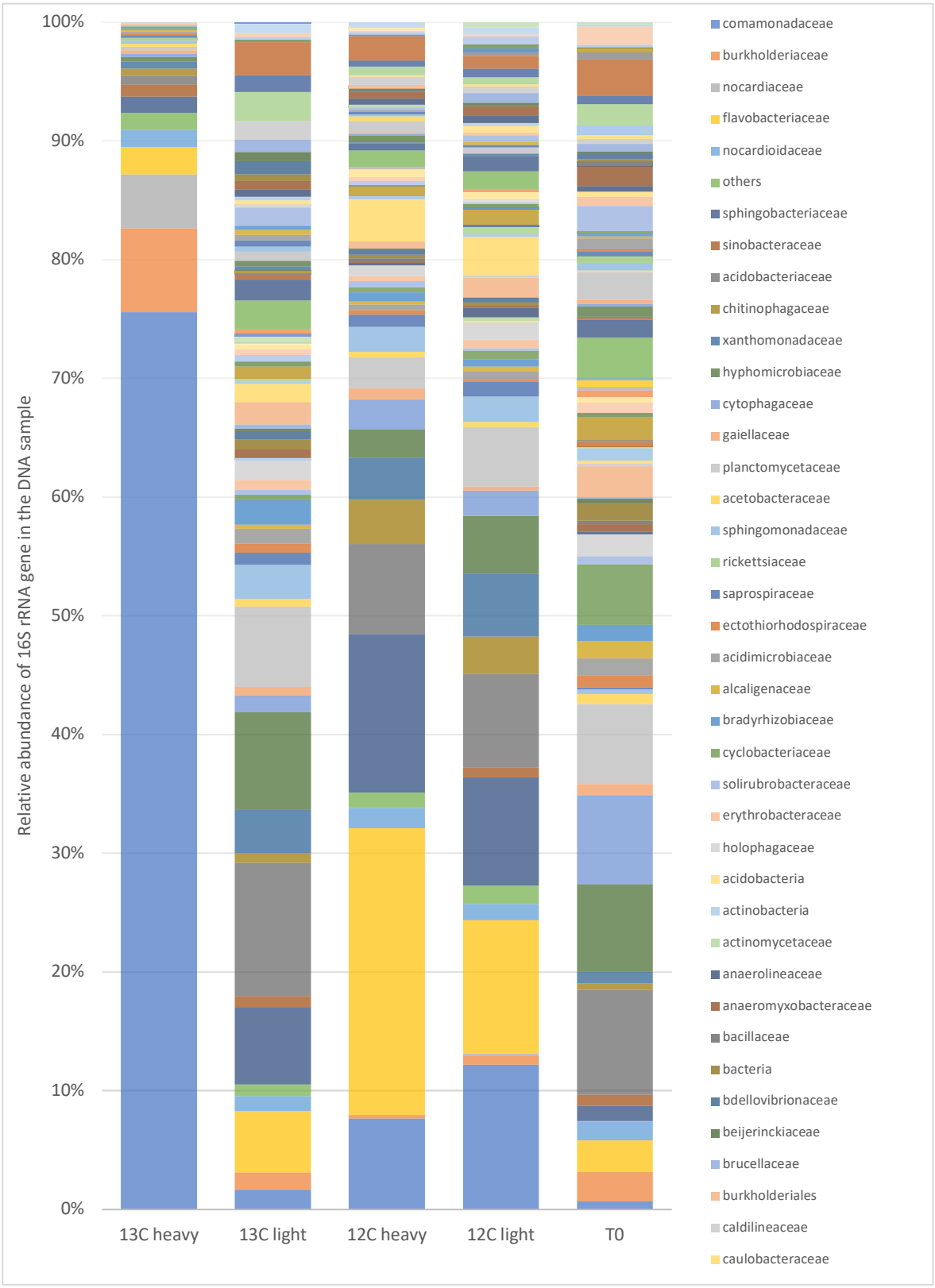
# Supplementary Figures



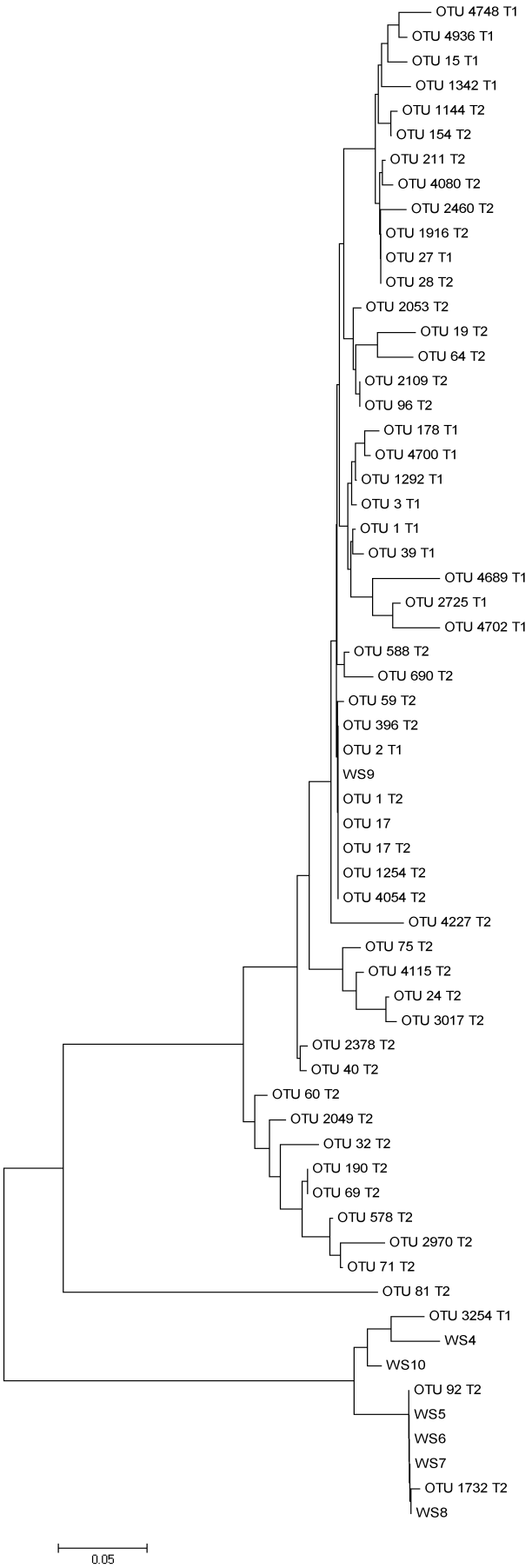

**Supplementary Figure 1.** Relative abundance of amplicon sequencing data for 16S rRNA gene in $T_0$, $T_1$ and $T_2$ samples for native and $^{12}C$ and $^{13}C$ enriched samples. Samples are ordered as follows: First $T_0$, followed by all $^{12}C$ incubations (heavy fractions $T_1$ first and replicates for $T_2$, light fractions $T_1$ and replicates for $T_2$), finally all $^{13}C$ incubations (heavy fractions $T_1$ first and replicates for $T_2$, light fractions $T_1$ and replicates for $T_2$).



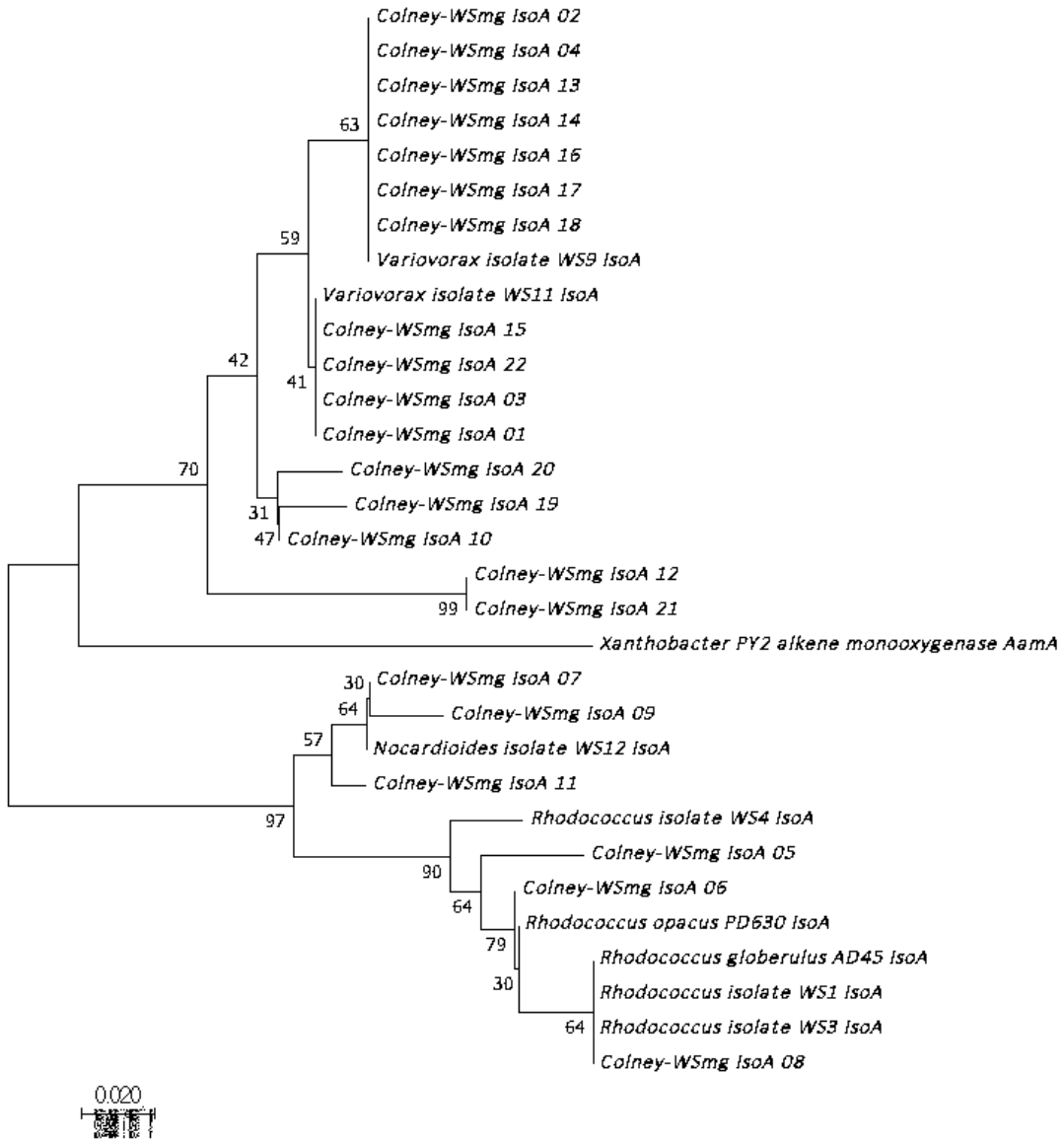

**Supplementary Figure 2.** Relative abundance of amplicon sequencing data at the genus level for 16S rRNA gene in native and $T_1$ samples ($^{12}C$ and $^{13}C$).



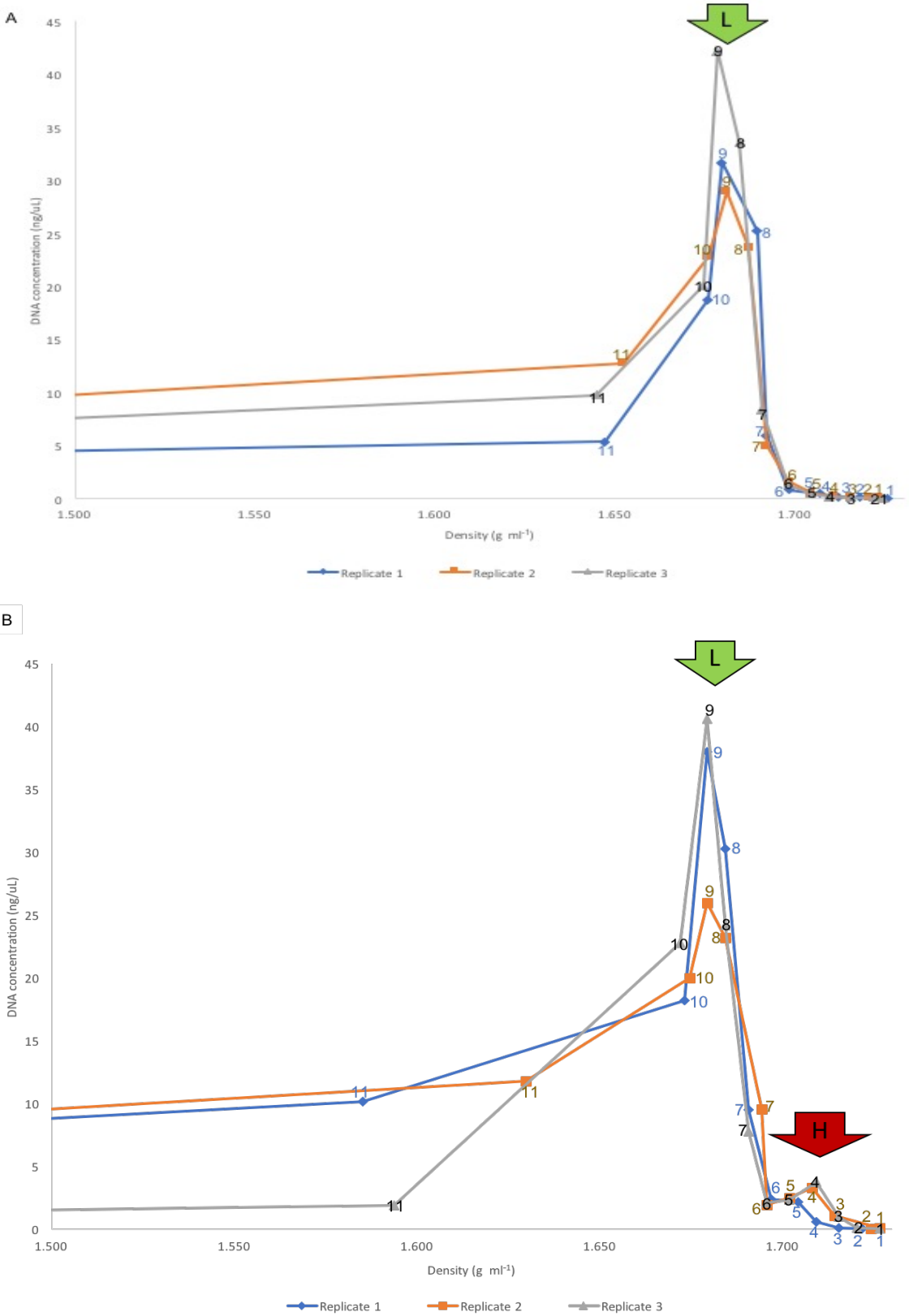

**Supplementary Figure 3.** Relative abundance of amplicon sequencing data at the family level for 16S rRNA gene in native and $T_1$ samples ($^{12}C$ and $^{13}C$).



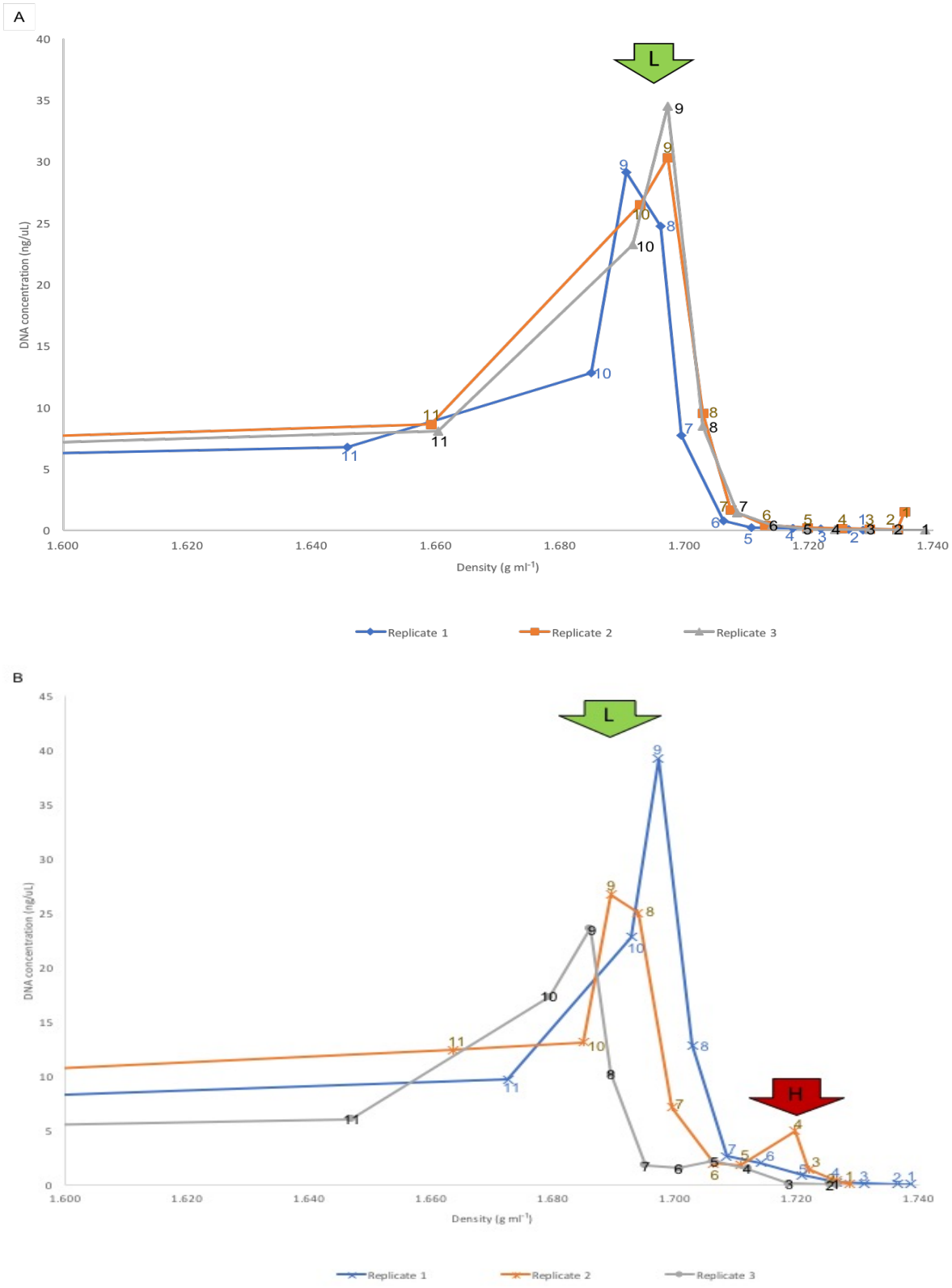

**Supplementary Figure 4**. Phylogenetic tree (V1-V3 16S rRNA gene) of most abundant (>1%) OTUs from heavy fraction of $^{13}C$ isoprene enrichments.



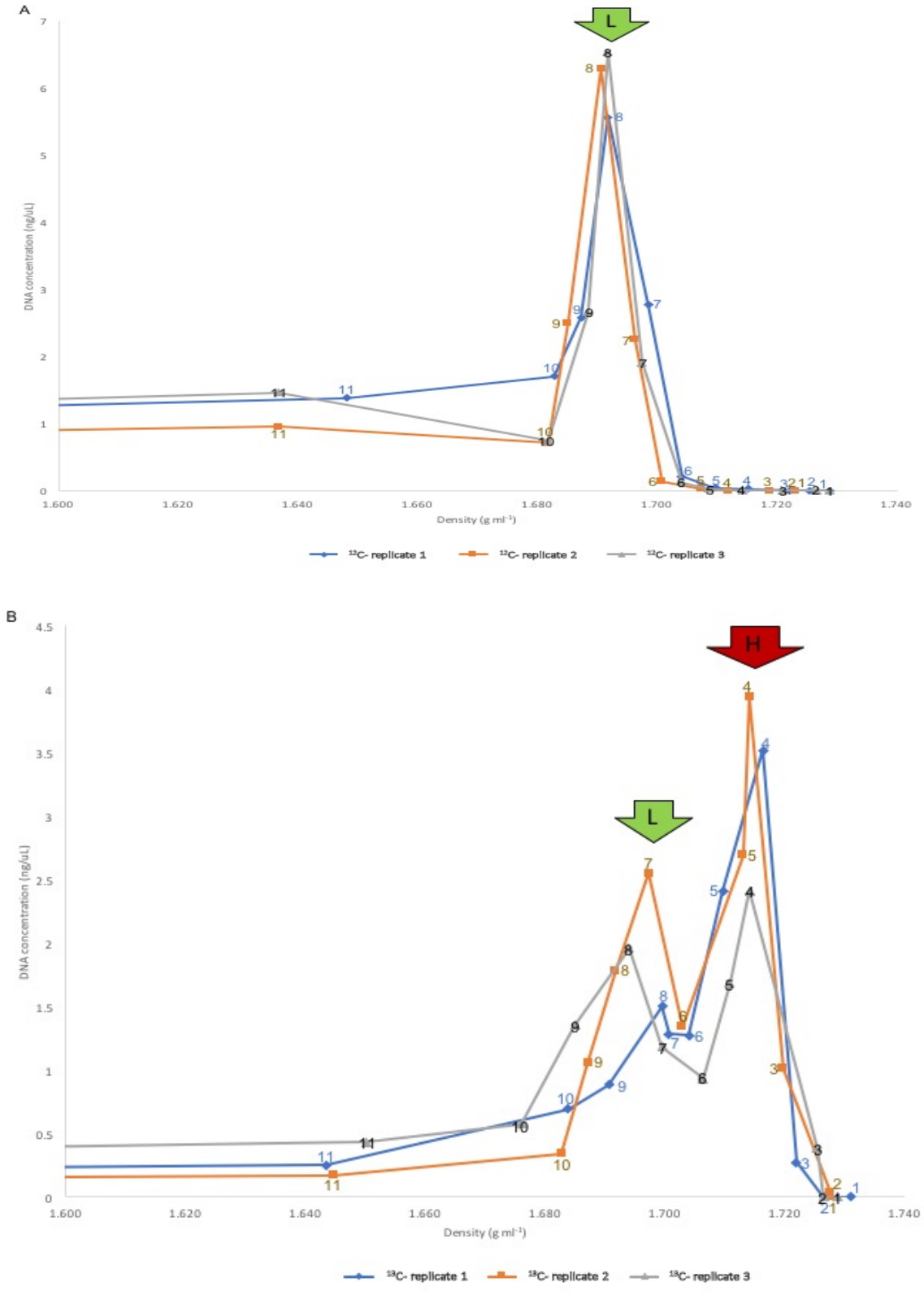

**Supplementary Figure 5**. Phylogenetic tree with all recovered metagenome IsoA sequences.



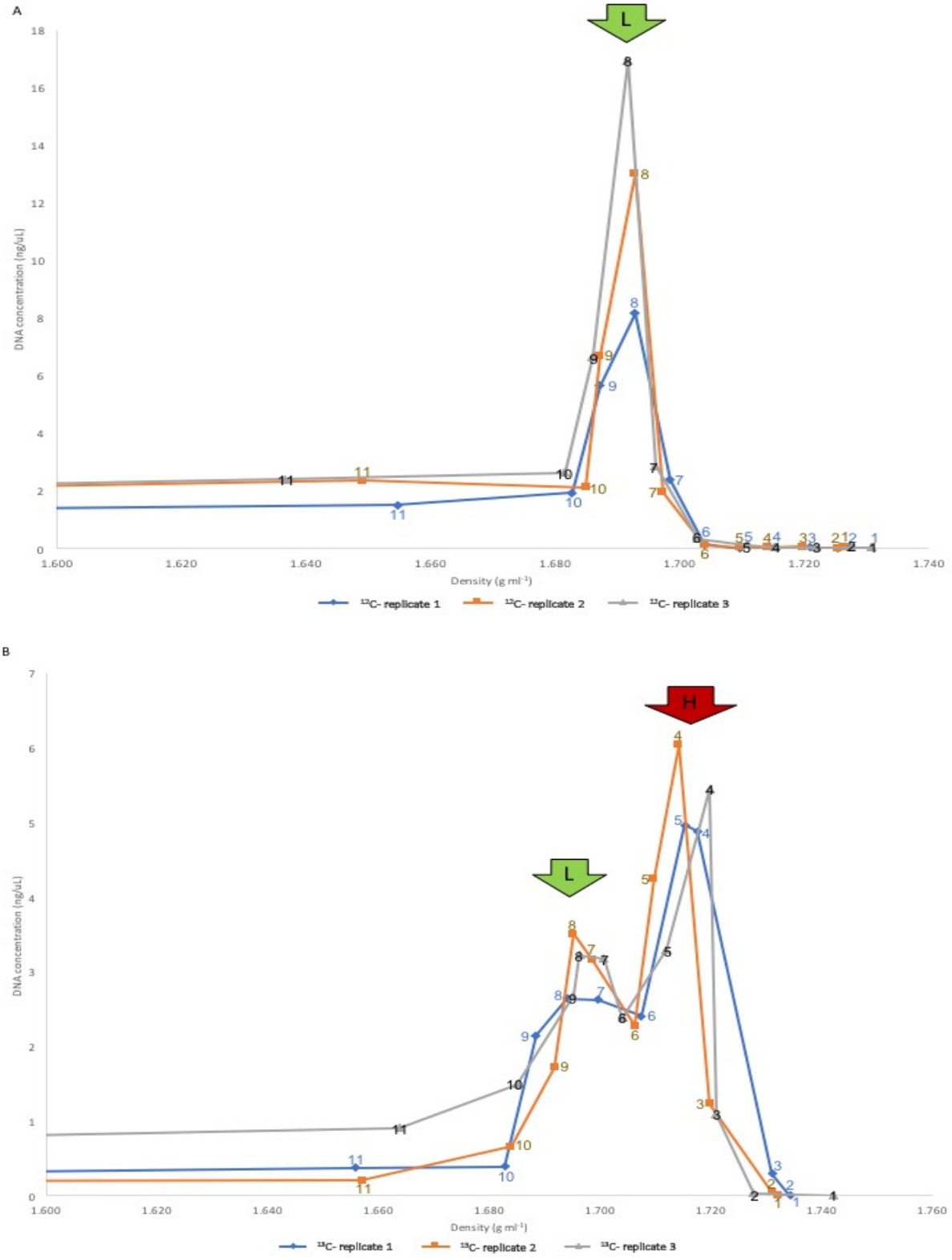

**Supplementary Figure 6** Fractionation curves (density vs. DNA concentration) for DNA obtained with SIP enrichments with OPS after 7 days of enrichment with A) $^{12}$C-isoprene and B) $^{13}$C-labelled isoprene. Fraction numbers presented on the graph.



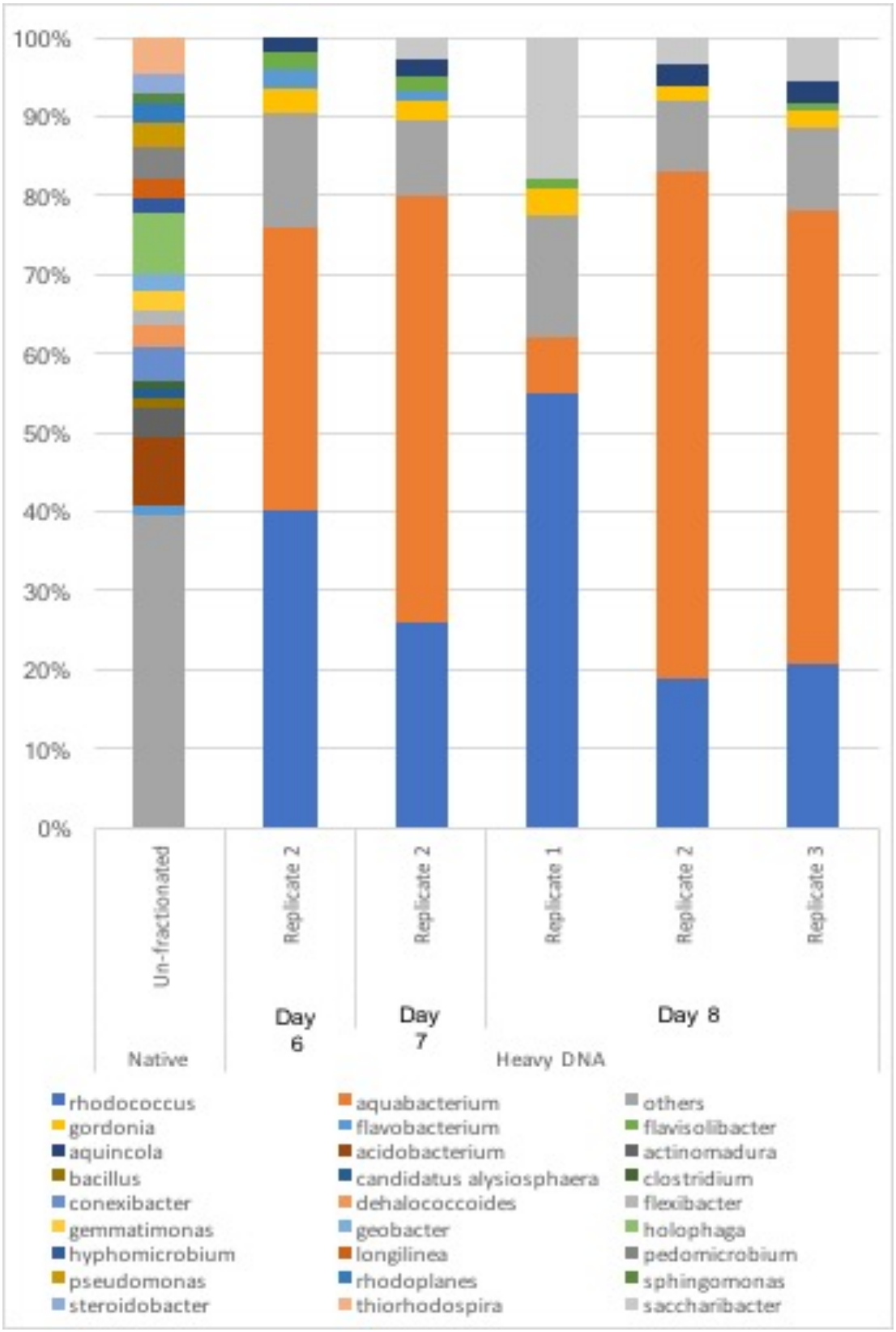

**Supplementary Figure 7** Fractionation curves (density vs. DNA concentration) for DNA obtained with SIP enrichments with OPS after 8 days of enrichment with A) $^{12}$C-isoprene and B) $^{13}$C-labelled isoprene. Fraction numbers presented on the graph.



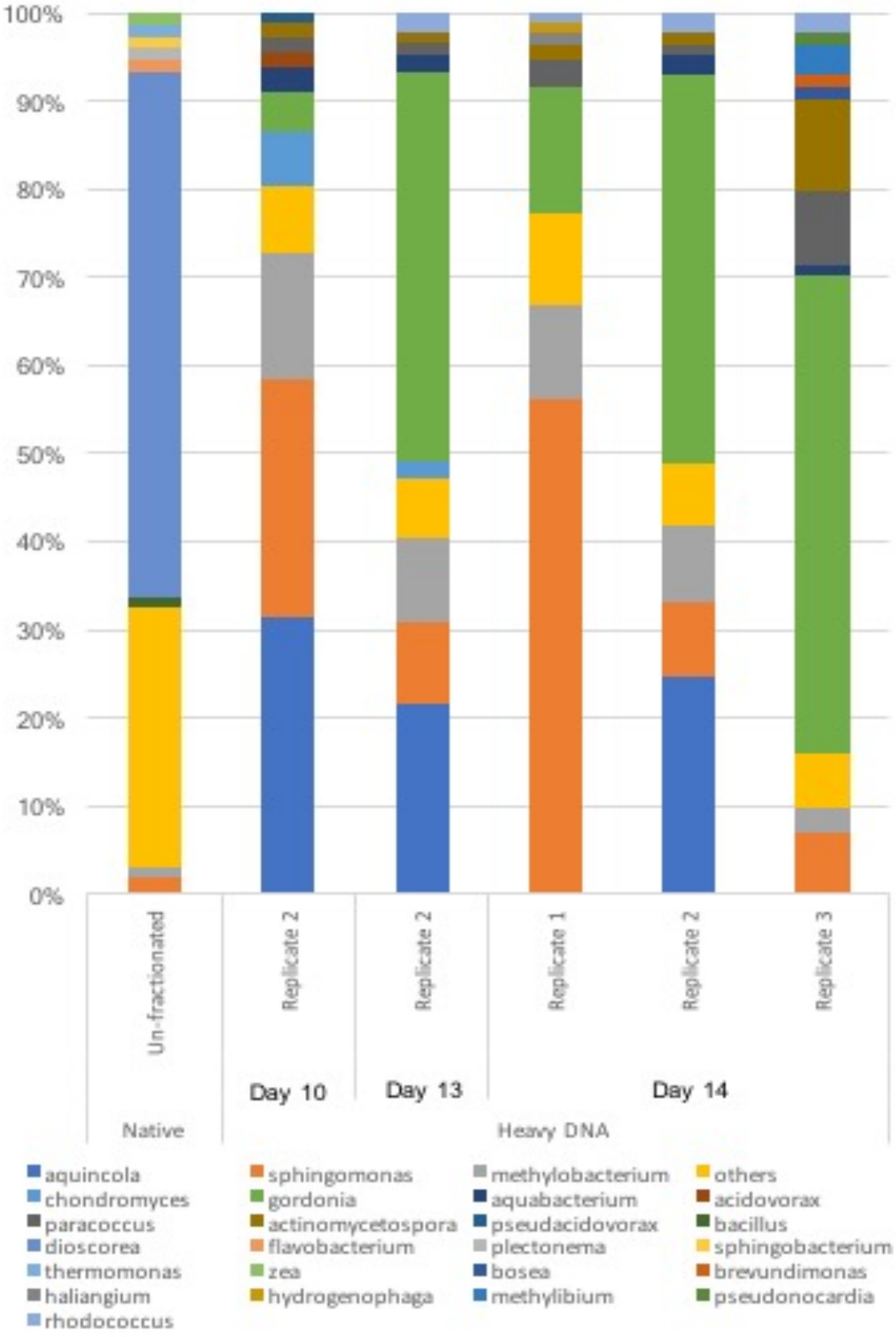

**Supplementary Figure 8** Fractionation curves (density vs. DNA concentration) for DNA obtained with SIP enrichments with OPL washings after 13 days of enrichment with A) $^{12}$C-isoprene and B) $^{13}$C-labelled isoprene. Fraction numbers presented on the graph.



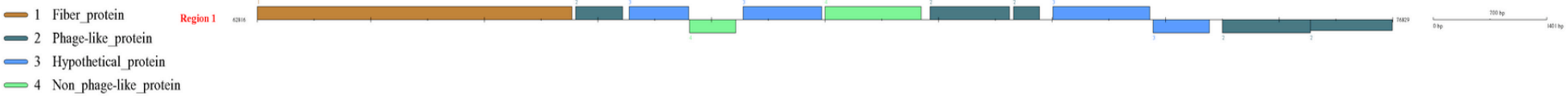

**Supplementary Figure 9** Fractionation curves (density vs. DNA concentration) for DNA obtained with SIP enrichments with OPL washings after 14 days of enrichment with A) $^{12}$C-isoprene and B) $^{13}$C-labelled isoprene. Fraction numbers presented on the graph.



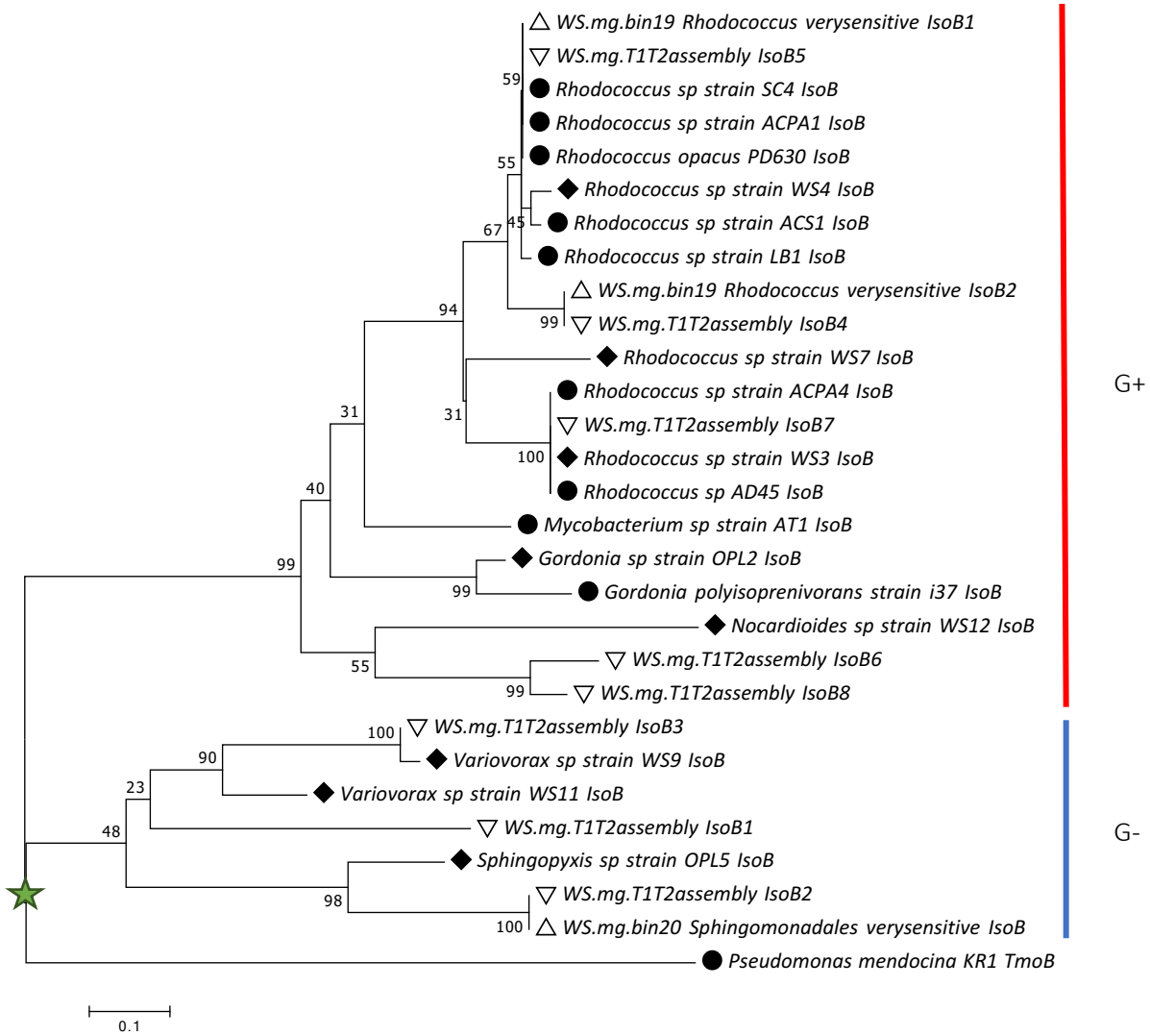

**Supplementary Figure 10** 16S rRNA gene relative abundance of genera present from oil palm soil (OPS) from native (un-fractionated) and enriched $^{13}$C-isoprene pooled heavy DNA fractions. Genera with lower abundance than 1% were established as “others”.



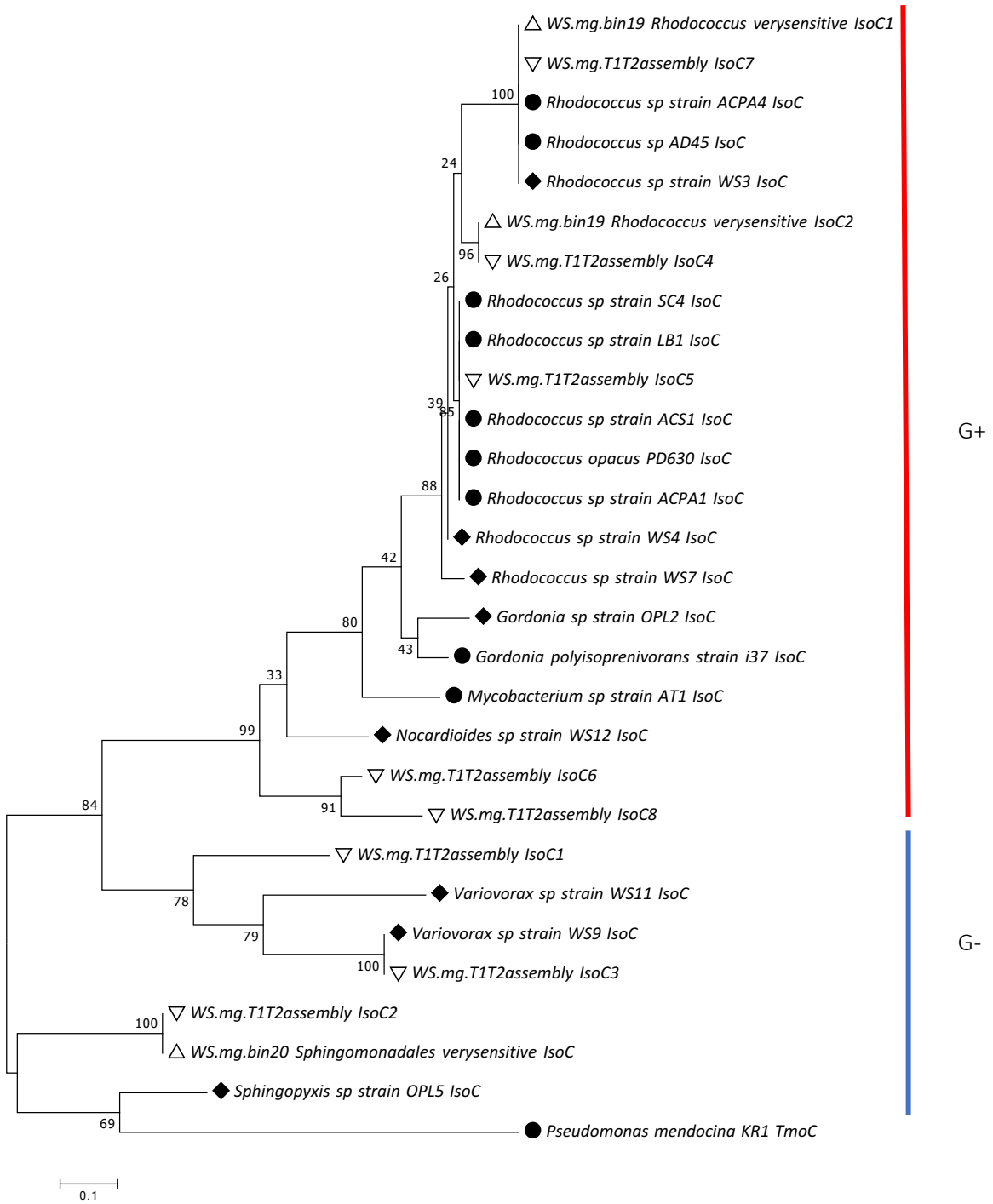

**Supplementary Figure 11** 16S rRNA gene relative abundance of genera present from oil palm leaf (OPL) washings from native (un-fractionated) and enriched $^{13}$C-isoprene pooled heavy DNA fractions. Genera with lower abundance than 1% were established as “others”.



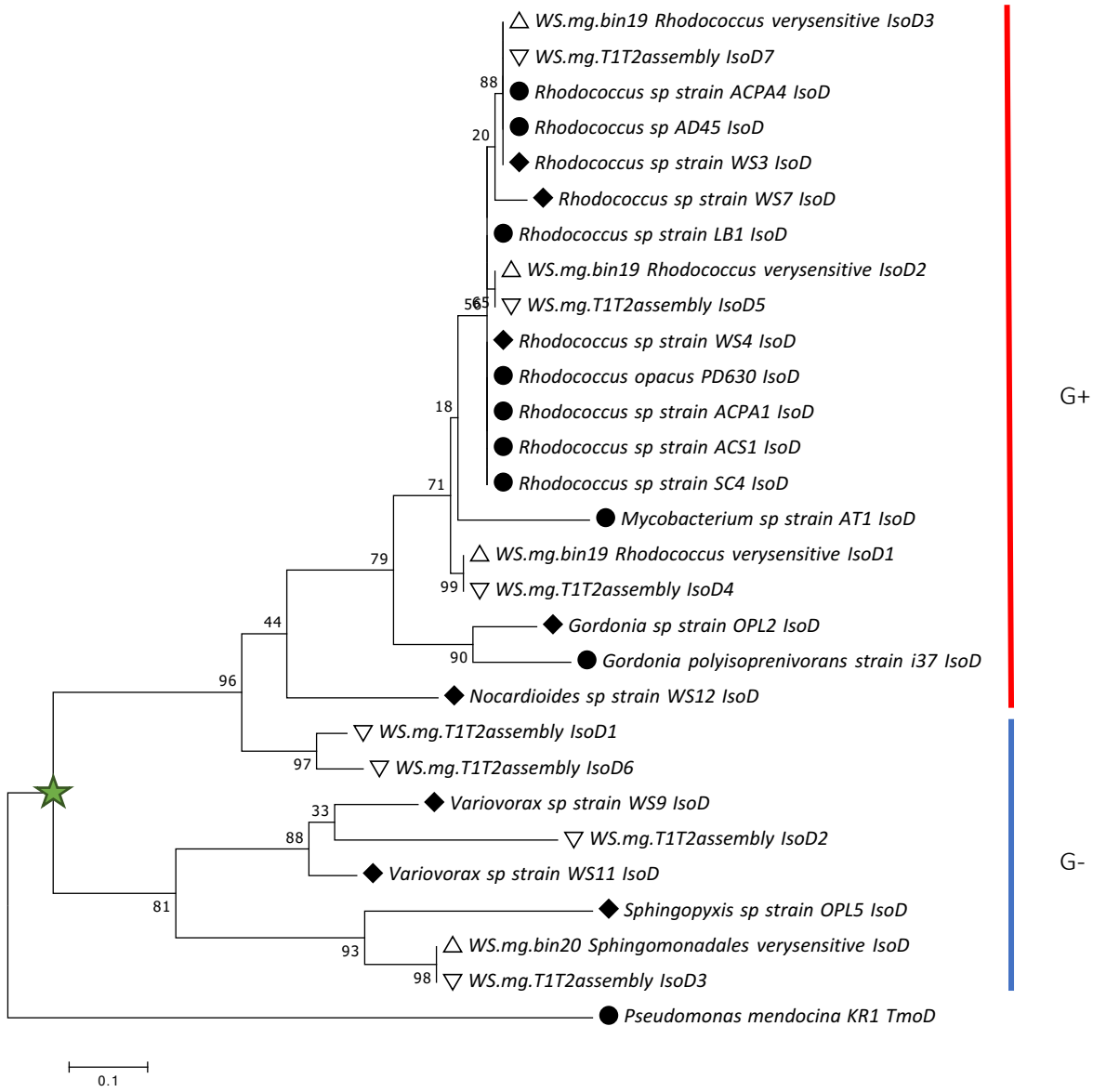

**Supplementary Figure 12** The region containing the prophage-like sequences in *Rhodococcus* sp. strain WS4.



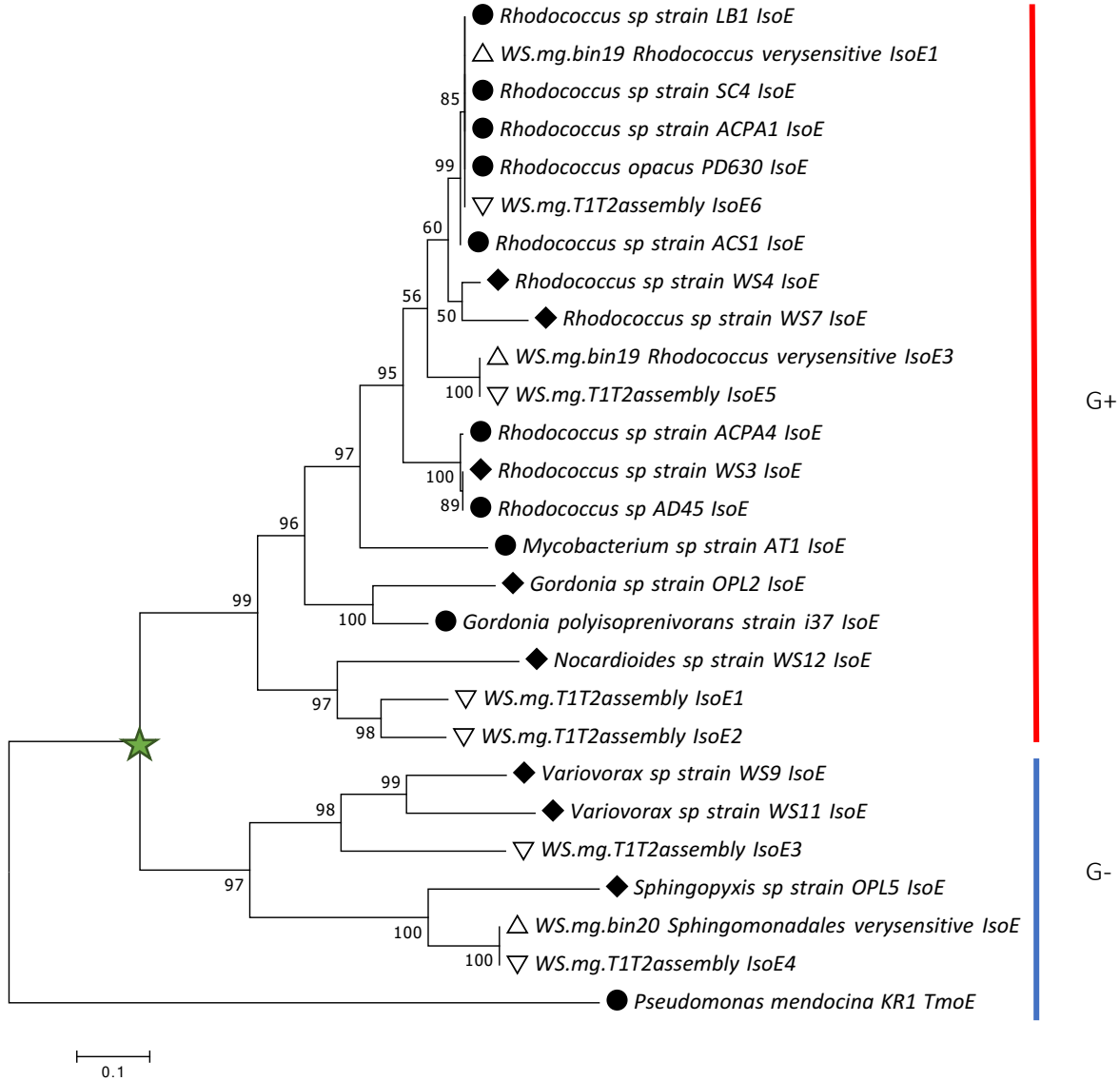

**Supplementary Figure 13** Phylogenetic analysis using IsoB sequences. Trees were constructed with maximum likelihood method and a bootstrap of 1000 replicates. The analysis was carried out with 29 different IsoB sequences. All gaps and missing data were eliminated, a total of 83 amino acid residues were included in the final dataset. Gram-positive sequences (red group), Gram-negative sequences (blue group) and the tree root (green star) that separates both groups are shown. Reference sequences (filled circle), novel isolate sequences (filled diamond), and metagenome sequences (empty triangles). The three WS.mg (willow soil metagenome) bin IsoJ sequences were obtained using the *very sensitive* settings using Metabat programme (details in **section 2.18**).



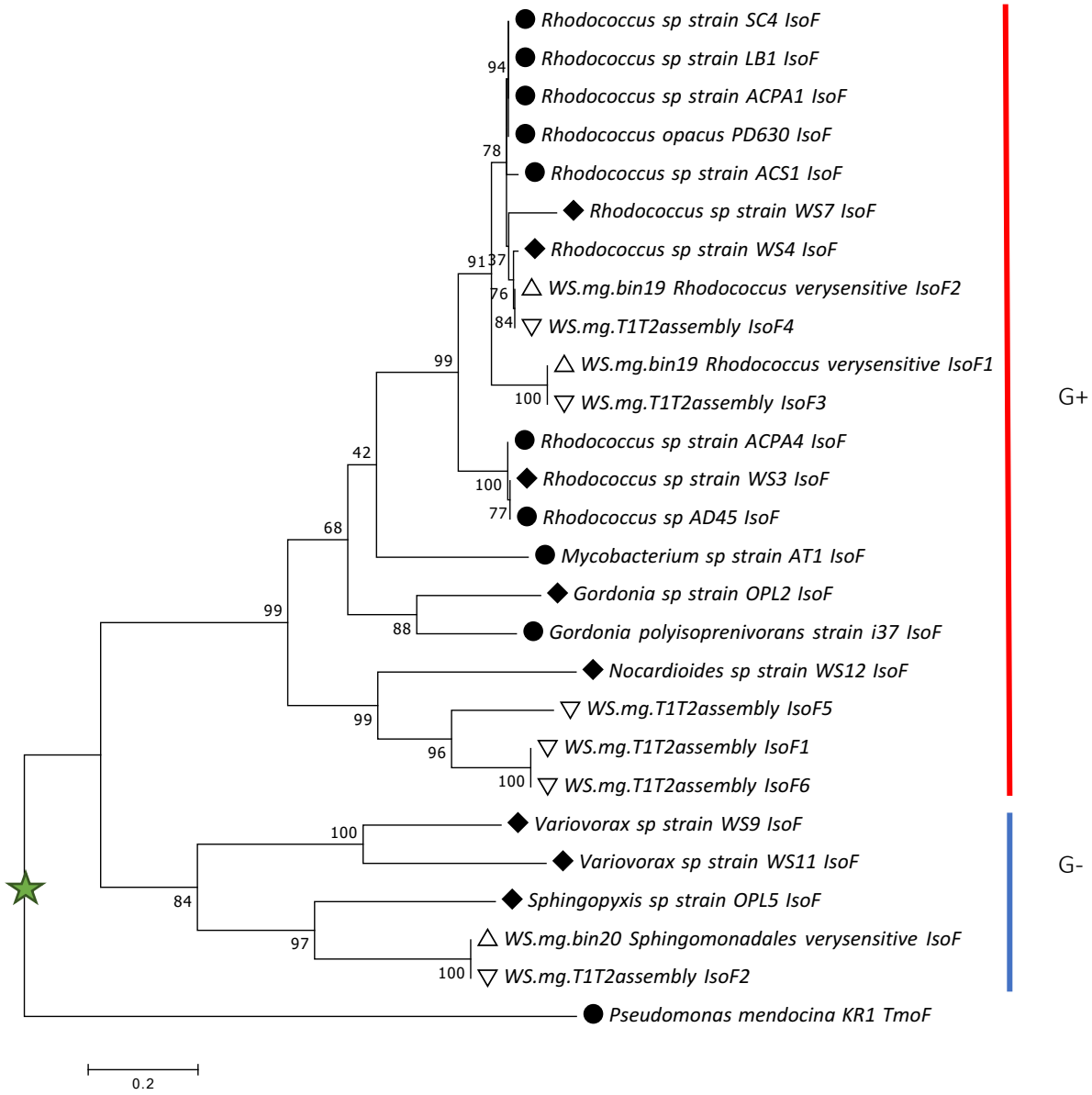

**Supplementary Figure 14** Phylogenetic analysis using IsoC sequences. Trees were constructed with maximum likelihood method and a bootstrap of 1000 replicates. The analysis was carried out with 29 different IsoC sequences. All gaps and missing data were eliminated, a total of 102 amino acid residues were included in the final dataset. Gram-positive sequences (red group) and Gram-negative sequences (blue group) and the tree root (green star) that separates both groups are shown. Reference sequences (filled circle), novel isolate sequences (filled diamond), and metagenome sequences (empty triangles). The three WS.mg (willow soil metagenome) bin IsoJ sequences were obtained using the *very sensitive* settings using Metabat programme (details in **section 2.18**).



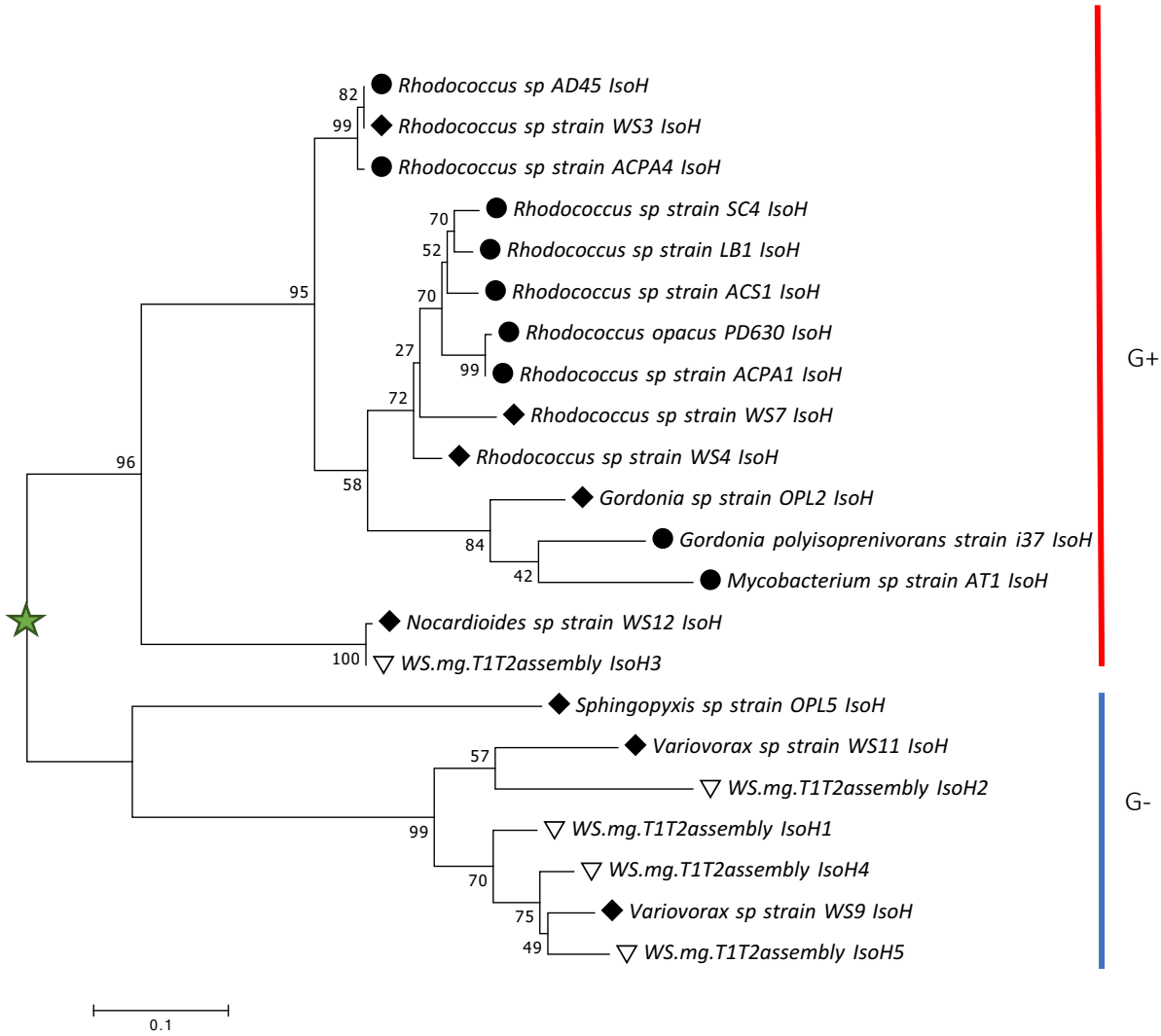

**Supplementary Figure 15** Phylogenetic analysis using IsoD sequences. Trees were constructed with maximum likelihood method and a bootstrap of 1000 replicates. The analysis was carried out with29 different IsoD sequences. All gaps and missing data were eliminated, a total of 99 amino acid residues were included in the final dataset. Gram-positive sequences (red group), Gram-negative sequences (blue group) and the tree root (green star) that separates both groups are shown. Reference sequences (filled circle), novel isolate sequences (filled diamond), and metagenome sequences (empty triangles). The four WS.mg (willow soil metagenome) bin IsoJ sequences were obtained using the *very sensitive* settings using Metabat programme (details in **section 2.18**).



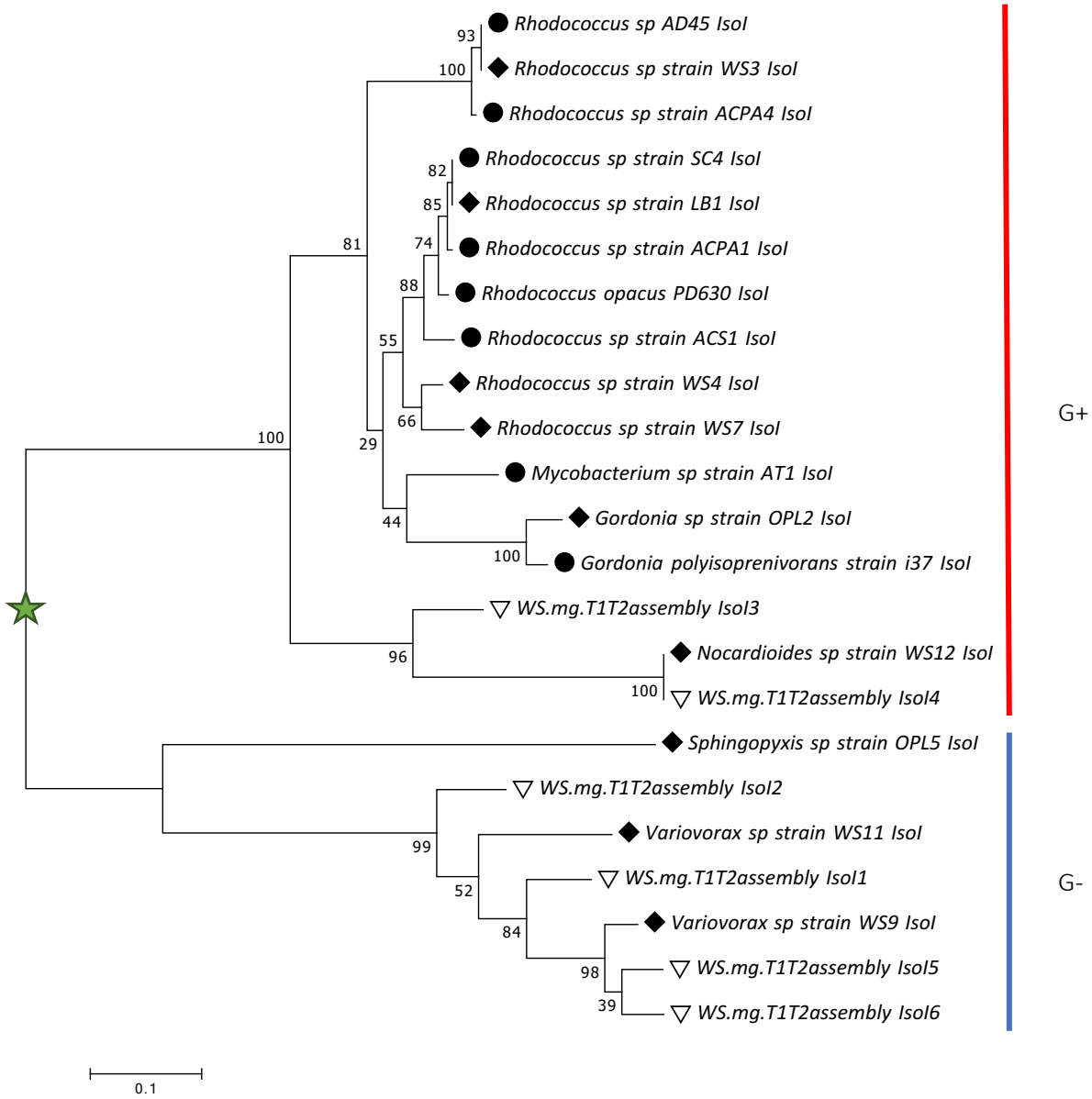

**Supplementary Figure 16** Phylogenetic analysis using IsoE sequences. Trees were constructed with maximum likelihood method and a bootstrap of 1000 replicates. The analysis was carried out with 27 different IsoE sequences. All gaps and missing data were eliminated, a total of 319 amino acid residues were included in the final dataset. Gram-positive sequences (red group), Gram-negative sequences (blue group) and the tree root (green star) that separates both groups are shown. Reference sequences (filled circle), novel isolate sequences (filled diamond), and metagenome sequences (empty triangles). The two WS.mg (willow soil metagenome) bin IsoJ sequences were obtained using the *very sensitive* settings using Metabat programme (details in **section 2.18**).



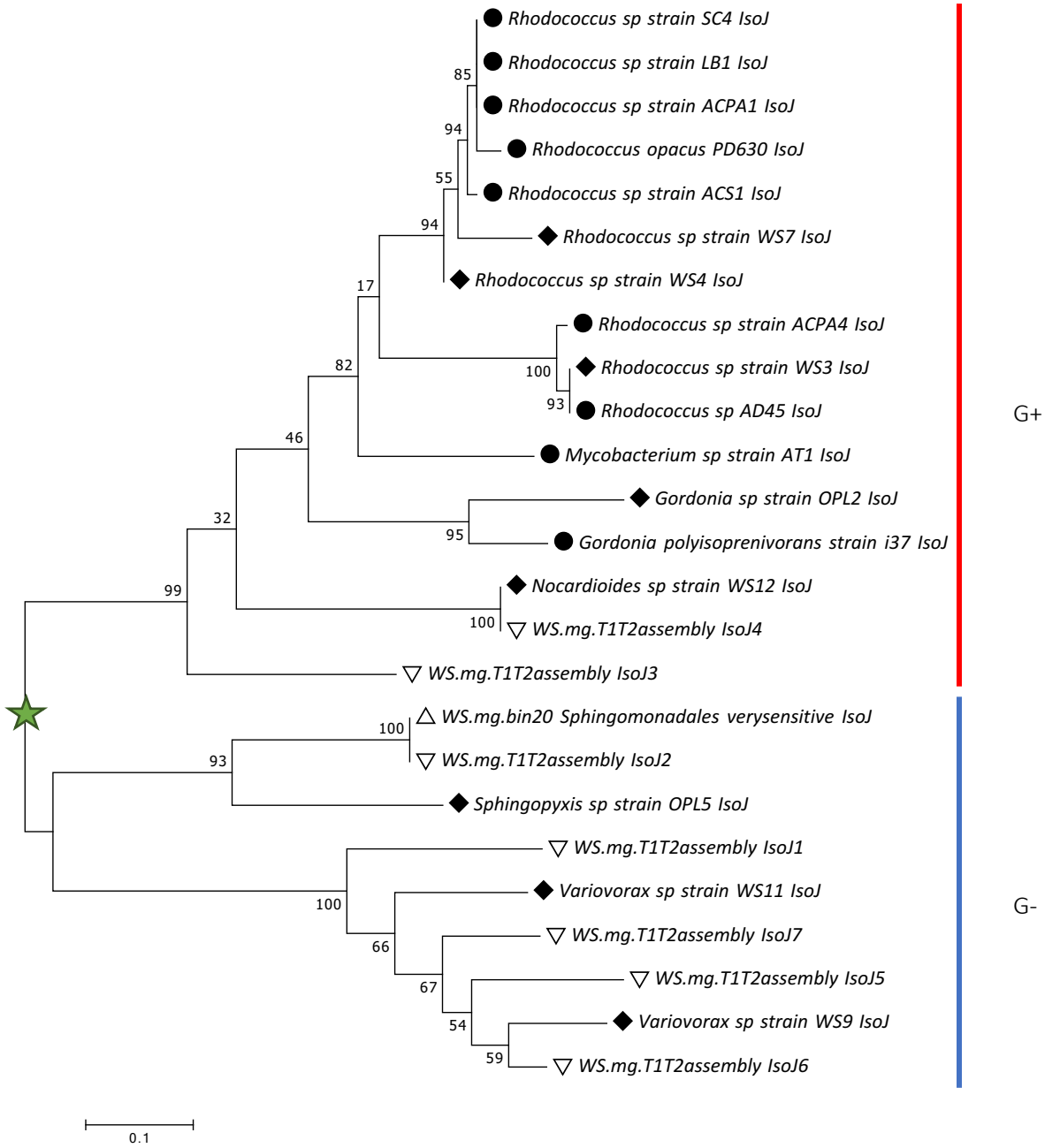

**Supplementary Figure 17** Phylogenetic analysis using IsoF sequences. Trees were constructed with maximum likelihood method and a bootstrap of 1000 replicates. The analysis was carried out with 27 different IsoF sequences. All gaps and missing data were eliminated, a total of 242 amino acid residues were included in the final dataset. Gram-positive sequences (red group), Gram-negative sequences (blue group) and the tree root (green star) that separates both groups are shown. Reference sequences (filled circle), novel isolate sequences (filled diamond), and metagenome sequences (empty triangles). The three WS.mg (willow soil metagenome) bin IsoJ sequences were obtained using the *very sensitive* settings using Metabat programme (details in **section 2.18**).




**Supplementary Figure 18** Phylogenetic analysis using IsoH sequences. Trees were constructed with maximum likelihood method and a bootstrap of 1000 replicates. The analysis was carried out with 22 different IsoH sequences. All gaps and missing data were eliminated, a total of 214 amino acid residues were included in the final dataset. Gram-positive sequences (red group), Gram-negative sequences (blue group) and the tree root (green star) that separates both groups are shown. Reference sequences (filled circle), novel isolate sequences (filled diamond), and metagenome sequences (empty triangles).




**Supplementary Figure 19** Phylogenetic analysis using IsoI sequences. Trees were constructed with maximum likelihood method and a bootstrap of 1000 replicates. The analysis was carried out with 22 different IsoI sequences. All gaps and missing data were eliminated, a total of 214 amino acid residues were included in the final dataset. Gram-positive sequences (red group), Gram-negative sequences (blue group) and the tree root (green star) that separates both groups are shown. Reference sequences (filled circle), novel isolate sequences (filled diamond), and metagenome sequences (empty triangles).




**Supplementary Figure 20** Phylogenetic analysis using IsoJ sequences. Trees were constructed with maximum likelihood method and a bootstrap of 1000 replicates. The analysis was carried out with 25 different IsoJ sequences. All gaps and missing data were eliminated, a total of 229 amino acid residues were included in the final dataset. Gram-positive sequences (red group), Gram-negative sequences (blue group) and the tree root (green star) that separates both groups are shown. Reference sequences (filled circle), novel isolate sequences (filled diamond), and metagenome sequences (empty triangles). The WS.mg (willow soil metagenome) bin IsoJ sequence was obtained using the *very sensitive* settings using Metabat programme (details in **section 2.18**).

FEMS Microbiology Ecology, 93, 2017, fix044

doi: 10.1093/femsec/fix044
Advance Access Publication Date: 3 April 2017
Perspective

PERSPECTIVE

# Fifty important research questions in microbial ecology

Rachael E. Antwis[1,*], Sarah M. Griffiths[2], Xavier A. Harrison[3], Paz Aranega-Bou[1], Andres Arce[4], Aimee S. Bettridge[5], Francesca L. Brailsford[6], Alexandre de Menezes[1], Andrew Devaynes[7], Kristian M. Forbes[8], Ellen L. Fry[9], Ian Goodhead[1], Erin Haskell[10], Chloe Heys[11], Chloe James[1], Sarah R. Johnston[5,†], Gillian R. Lewis[7], Zenobia Lewis[11], Michael C. Macey[12], Alan McCarthy[11], James E. McDonald[13], Nasmille L. Mejia-Florez[12], David O'Brien[14], Chloé Orland[15], Marco Pautasso[16], William D. K. Reid[17], Heather A. Robinson[9], Kenneth Wilson[18] and William J. Sutherland[19]

[1]School of Environment and Life Sciences, University of Salford, The Crescent, Salford M5 4WT, UK, [2]School of Science and the Environment, Manchester Metropolitan University, Manchester, Greater Manchester M1 5GD, UK, [3]Institute of Zoology, Zoological Society of London, London, London NW1 4RY, UK, [4]Silwood Park, Faculty of Natural Sciences, Imperial College London, London, London SW7 2AZ, UK, [5]School of Biosciences, Cardiff University, Cardiff, South Glamorgan CF10 3XQ, UK, [6]School of Environment, Natural Resources and Geography, Bangor University, Bangor, Gwynedd LL57 2DG, UK, [7]Biosciences, Edge Hill University, Ormskirk, Lancashire L39 4QP, UK, [8]Department of Virology, University of Helsinki, Helsinki 00014, Finland, [9]School of Earth and Environmental Sciences, Faculty of Science and Engineering, University of Manchester, Manchester M13 9PT, UK, [10]Department of Biology, University of York, York, North Yorkshire YO10 5DD, UK, [11]Institute of Integrative Biology/School of Life Sciences, University of Liverpool, Liverpool, Merseyside L69 3BX, UK, [12]School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK, [13]School of Biological Sciences, Bangor University, Bangor, Gwynedd LL57 2DG, UK, [14]Scottish Natural Heritage, Inverness IV3 8NW, UK, [15]Department of Plant Sciences, University of Cambridge, Cambridge, Cambridgeshire CB2 1TN, UK, [16]Animal and Plant Health Unit, European Food Safety Authority, Parma 43126, Italy, [17]School of Biology, Newcastle University, Newcastle upon Tyne, Tyne and Wear NE1 7RU, UK, [18]Lancaster Environment Centre, Lancaster University, Lancaster, Lancashire LA1 4YW, UK and [19]Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge, Cambridgeshire CB2 1TN, UK

*Corresponding author: School of Environment and Life Sciences, University of Salford, Room 336, Peel Building, University of Salford, The Crescent, Salford M5 4WT, UK. Tel: +44 161 295 4641; E-mail: r.e.antwis@salford.ac.uk

**One sentence summary:** We identify research questions in the field of microbial ecology, with emerging themes that recognise vast microbial functions that could benefit humanity, and the need to integrate knowledge across organisms.

**Received:** 9 November 2016; **Accepted:** 31 March 2017

**Editor:** Marcus Horn
†Sarah R. Johnston, http://orcid.org/0000-0002-3806-3416

## ABSTRACT

Microbial ecology provides insights into the ecological and evolutionary dynamics of microbial communities underpinning every ecosystem on Earth. Microbial communities can now be investigated in unprecedented detail, although there is still a wealth of open questions to be tackled. Here we identify 50 research questions of fundamental importance to the science or application of microbial ecology, with the intention of summarising the field and bringing focus to new research avenues. Questions are categorised into seven themes: host–microbiome interactions; health and infectious diseases; human health and food security; microbial ecology in a changing world; environmental processes; functional diversity; and evolutionary processes. Many questions recognise that microbes provide an extraordinary array of functional diversity that can be harnessed to solve real-world problems. Our limited knowledge of spatial and temporal variation in microbial diversity and function is also reflected, as is the need to integrate micro- and macro-ecological concepts, and knowledge derived from studies with humans and other diverse organisms. Although not exhaustive, the questions presented are intended to stimulate discussion and provide focus for researchers, funders and policy makers, informing the future research agenda in microbial ecology.

**Keywords:** environmental processes; evolutionary processes; functional diversity; host–microbiome interactions; priority setting, research agenda

## INTRODUCTION

In recent years there has been an explosion in microbial ecological research, which is reflected in broad-scale research projects such as the Human Microbiome Project and the Earth Microbiome Project, as well as in the peer-reviewed literature (e.g. Boers, Jansen and Hays 2016). Recent rapid technological advances, including next-generation sequencing, (meta)genomics, metabolomics, (meta)transcriptomics and (meta)proteomics, have vastly increased our ability to study microbial community complexity and function (Morris *et al.* 2002; Hiraoka, Yang and Iwasaki 2016). These provide unprecedented opportunities to assess genomic potential, gene regulation, expression and function *in situ* (Schneider *et al.* 2012; Franzosa *et al.* 2015), especially when combined with detailed knowledge of natural history and environmental parameters (Peay 2014). Such techniques have been applied to a vast range of fields within the scope of 'microbial ecology' in order to better understand how microorganisms interact with and affect their environment, each other and other organisms.

With an overwhelming and ever-growing number of potential and critical research avenues in microbial ecology, it is timely to identify major questions and research priorities that would progress the field. Here we present the results of a workshop hosted by the British Ecological Society's Microbial Ecology Special Interest Group in June 2016, which used a discussion and voting-based system to identify 50 research questions of importance to the field of microbial ecology. Similar exercises identifying important research questions have been conducted in conservation (Sutherland *et al.* 2009; Dicks *et al.* 2012), pure ecology (Sutherland *et al.* 2013a), marine biodiversity (Parsons *et al.* 2014), sustainability (Dicks *et al.* 2013; Jones *et al.* 2014) and non-ecological subjects including UK poverty (Sutherland *et al.* 2013b). These papers have been widely accessed and are directly applicable to the development of policy, as highlighted by Jones *et al.* (2014).

## METHODS

### Participants

The methods used here were based broadly on those presented in Sutherland *et al.* (2011). A 1-day workshop was held by the British Ecological Society's Microbial Ecology Special Interest Group at the University of Salford (UK) in June 2016. Invitations to attend the meeting were distributed via the British Ecological Society's membership mailing list and through social media (Twitter and Facebook). In total, 34 participants from 20 institutions attended and contributed to the development of the 50 questions listed below, with the majority listed as authors on this paper.

### Questions

Prior to the workshop, attendees were asked to submit questions via an online form that they thought most closely met the following brief:

> We are aiming to identify 50 questions that, if answered, will make a considerable difference to the use of microbial ecology by practitioners and policy makers, or to the fundamentals of the field of microbial ecology. These should be questions that are unanswered, could be answered, and could be tackled by a research programme. This is expected to set the agenda for future research in the field of microbial ecology.

A total of 244 questions were submitted by attendees (see Supplementary Information), and assigned (by R.E. Antwis and S.M. Griffiths) to the following themes:

i. Host–microbiome interactions
ii. Health and infectious diseases
iii. Human health and food security
iv. Microbial ecology in a changing world
v. Environmental processes

vi. Functional diversity
vii. Evolutionary processes

An additional eighth theme named 'society and policy' was created to encompass questions that were generally applicable across the biological sciences, as well as those specific to the field of microbial ecology, which could not necessarily be addressed through laboratory based microbial ecology research, per se.

### Question selection process

Prior to the workshop, participants were asked to identify the top ~20% of questions in each theme that most closely aligned with the brief (selection of 5–11 questions from a total of 26–57 questions per theme via an online form; Supplementary Information). Participants were asked to consider all questions within a theme and to select questions based on the theme's context and the brief for the workshop. Some questions were included in more than one theme to encourage discussion and to increase the likelihood that pertinent questions remained in the selection process. Questions were then ranked according to the number of online votes they received, and this formed the material for the workshop.

Parallel sessions to discuss each theme were run at the workshop, with participants free to select which theme sessions they attended. Questions were discussed in order of lowest ranking to highest, with duplicates removed and questions reworded as necessary. For each theme, a final set of 'gold' (~15% of questions, total of 47 questions across all themes) and 'silver' questions (~10% of questions, total of 29 questions) were identified. Where necessary, a show of hands was used to ensure the democratic process was upheld.

A final plenary session was held in which all gold and silver questions were discussed. For gold questions, duplicates among categories were removed and questions reworded to reflect the discussion in the room, resulting in 43 gold questions. A similar process was then completed for silver questions, and a show of hands used to vote for seven questions that could be elevated to gold status to form the final set of 50 questions.

### Limitations

All but four participants were from British universities, although there were representatives from a range of nationalities and research areas. The manner in which this paper was developed (i.e. through a physical workshop and via the British Ecological Society) means that, without a substantial travel budget, a bias towards UK institutions was inevitable. However, many participants have worked on, or currently collaborate in, research projects on non-UK ecosystems and species, and therefore the questions proposed are drawn from considerable knowledge and experience of the field internationally. Additionally, although most individuals were from academic institutions, many individuals had previous or ongoing collaborations with industrial partners and governmental/non-governmental organisations.

## RESULTS

The following 50 questions are presented by theme, and are not ordered according to relevance or importance. Due to the nature of the process, some questions may appear similar across themes, but within the context of each theme can take on a different meaning. Some questions may relate to research areas that are already somewhat active, and these serve to highlight the importance of and encourage further work in these areas. Some of these questions apply across multiple biomes and ecosystems, and can be considered in the context of multiple host organisms and across varying temporal and spatial scales.

### Host–microbiome interactions

Host–microbiome interactions determine many host life history traits such as behaviour, reproduction, physiological processes and disease susceptibility (Archie and Theis 2011; Koch and Schmid-Hempel 2011; Willing, Russell and Finlay 2011; Daskin and Alford 2012; King *et al.* 2016). Increasingly, we are discovering that host–microbiome interactions produce complex and dynamic communities that fluctuate in compositional abundance correlated with factors as diverse as host genotype, developmental stage, diet and temporal changes, among others (e.g. Spor, Koren and Ley 2011). Even in otherwise well-studied organisms, very little is known about the consequences of microbiome variation for host processes, particularly across different spatial and temporal scales. Considerations of host microbiomes are also likely important for global issues, such as the efficacy of conservation efforts including species reintroduction programmes (reviewed in Redford *et al.* 2012; McFall-Ngai 2015). Additionally, interactions between native and non-native species are correlated with transmission of microbiota, often determined by relatedness or diet type (Ley *et al.* 2008), and the microbiome plays a key role in the control and competence of insect crop pests and vectors of disease (reviewed in Weiss and Aksoy 2011). The following questions aim to address the shortfall in our understanding of the interactions between microbiomes and their human and non-human hosts.

1. What are the primary mechanisms within a host that mediate microbe–microbe and host–microbe interactions?
2. What are the relative contributions of host-associated and environmental factors in determining host microbial community composition?
3. How do microbial communities function to affect the phenotype of the host?
4. Can compositional or evolutionary changes in microbiomes help hosts adapt to environmental change within the lifetime of the host?
5. What is the role of the microbiota in host speciation processes?
6. How can the associated microbiota be effectively included in risk assessments of invasive non-native species?
7. How does the microbiome of captive animals affect the success of reintroduction programmes?
8. How can a 'systems biology' approach improve our understanding of host–microbe interactions?

### Health and infectious diseases

The last 50 years have seen the emergence of several hypervirulent wildlife pathogens in animals (reviewed in Tompkins *et al.* 2015) and plants (Pautasso *et al.* 2015). Although the role of microorganisms as pathogens is well known, the importance of host-associated microbiomes in regulating disease susceptibility is becoming more apparent (Koch and Schmid-Hempel 2011; Daskin and Alford 2012; King *et al.* 2016). A major outstanding research goal is to understand how within-host interactions

among microbes and invading pathogens may shape patterns of infection intensity and disease progression (see also 'evolutionary processes'). Several studies have sought to determine how manipulation of host microbiomes may ameliorate the spread and impact of such diseases (e.g. reviewed in Rebollar *et al.* 2016).

While for many disease states the paradigm holds true that one microorganism causes one disease, polymicrobial infections are becoming more apparent through metagenomic and metatranscriptomic sequencing of disease-associated microbial communities (Gilbert *et al.* 2016). Consequently, the 'pathobiome' concept, where a disease state is influenced by complex interactions between commensal and pathogenic microorganisms, presents new challenges for applying Koch's postulates to diseases arising from polymicrobial interactions (Vayssier-Taussat *et al.* 2014), such as black band disease in corals (Sato *et al.* 2016) and olive knot disease (Buonaurio *et al.* 2015).

In this theme, we have identified research questions relating to the microbial ecology of infectious diseases and host health. Although much can be learnt from the comparatively high number of studies in the human and biomedical literature (e.g. using network approaches in epidemiology), the questions selected in this theme predominantly relate to non-human animals and plants, as humans are covered later ('human health and food security').

9. How can we better track the source and dispersal of particular microorganisms in real time?
10. Many microorganisms are unculturable, and many microbiome studies reveal that diseases are polymicrobial; how can we re-evaluate Koch's postulates in this context?
11. Which factors trigger 'covert' infections to become 'overt', impacting host health?
12. At the population level, how is the burden and shedding intensity of intracellular microbes affected by co-infection by extracellular parasites?
13. What is the ecological relevance of the internalisation of bacterial pathogens by protozoa in terms of their survival and spread?
14. How can network theory best be used to predict and manage infectious disease outbreaks in animals and plants?
15. Can microbiomes of wildlife (plants and animals) be used or manipulated to enhance health and/or disease resistance?

## Human health and food security

With the human population due to exceed eight billion by 2024, food security and human health are high on political and scientific agendas. The human microbiome has been the focus of intense research efforts in recent years, (e.g. Spor, Koren and Ley 2011; Walter and Ley 2011; Mueller *et al.* 2012), because gut symbionts shape the immune response (Round and Mazmanian 2009), and diversity fluctuates through chronic conditions and infectious diseases including diabetes, obesity (Ridaura *et al.* 2013; Baothman *et al.* 2016; Serino *et al.* 2016), asthma (Smits *et al.* 2016) and HIV (Lozupone *et al.* 2013). Improving our understanding of the core human microbiome and individual variation will underpin pharmomicrobiomics, enabling development of novel therapeutic treatments and, ultimately, personalised medicine (e.g. Ubeda *et al.* 2013).

Antibiotic resistance resulting from selective pressures generated by the use and misuse of antibiotics is a global threat to public health (Levy 1997; Tam *et al.* 2012). The volume of antibiotics used in agriculture now exceeds the amount used in human medicine in many countries (WHO 2011). Antibiotics are still widely used in livestock for prophylaxis and growth promotion, often at subtherapeutic concentrations, exacerbating resistance (Krishnasamy, Otte and Silbergeld 2015). The impact of the leaching of antibiotics into the natural environment and subsequent impacts on natural microbial communities remains poorly characterised (Franklin *et al.* 2016). Current practices of growing high-intensity monoculture crops have a negative impact on the microbial biodiversity of soils through a combination of tillage, subsequent erosion and chemical applications (Helgason *et al.* 1998; Jacobsen and Hjelmsø 2014; Zuber and Villamil 2016), which imposes selection pressures on pathogenic microbes, fungal symbiotic partners and plant growth promoting bacteria (Chaparro *et al.* 2012; Hartmann *et al.* 2015). Thus, there is a need to maintain and enhance microbial populations of crop ecosystems, especially in light of antibiotic resistance (Ellouze *et al.* 2014). As antibiotic resistance increases, along with our concern about potential impact on both human and animal health, there is an increasing drive to find new forms of antibiotics.

Though the remit for this section is relatively broad, the questions focus around two main areas: (i) studying the human microbiome to improve the treatment of disease, including the development of personalised medicine and novel antibiotics; and (ii) understanding how 'current' antibiotic regimes and farming practices may negatively impact the diversity of the environmental microbiome and food production capacity.

16. How can human microbiome studies improve personalised medicine?
17. What ecological principles can be applied in the search for new antibiotics and alternatives?
18. What are the main determinants of waterborne infection outbreaks, and what is the best strategy to control these in water distribution systems?
19. What are the consequences of antibiotic and pharmaceutical use in human medicine on microbial communities in freshwater and soil environments?
20. To what extent are microbial species distributions influenced by climate, and what are the consequences for food security and human health?
21. How much microbial diversity in the soil has been lost through monoculture and what is the importance of this?
22. Intensive farming may involve high levels of agrochemicals and broad-spectrum antibiotic usage: what will be the long-term effects on microbial communities?
23. How best can we harness microbial communities to enhance food production?

## Microbial ecology in a changing world

Global changes resulting from human activity impact almost every habitat on earth. It is imperative that we focus efforts on understanding the impacts of human activities such as climate change, urbanisation, agriculture and industrial processes on microbial communities, ecosystem functioning equilibrium and host health. Microbial populations have a tremendous capacity to adapt to changes in their abiotic environment, yet the functional implications of these transitions in microbial ecology are still poorly understood and characterised (Bissett *et al.* 2013), and the role of microbes in mediating the response of larger organisms to change is equally understudied. Global environmental changes (GECs) are complex and multifaceted. Human

activities such as urbanisation, land-use change and introduction of invasive species have played a role in shifting global ecosystems via desertification, climate change and habitat degradation. Although such changes have been quantified in aquatic and terrestrial habitats (e.g. Haberl *et al.* 2007; Halpern *et al.* 2008), their effects on microbial communities and impacts on ecosystem function are often hindered by a lack of characterisation of communities, or limited understanding of microbial functional traits. Shifts in basic nutrients and gases such as $CO_2$, along with temperature fluctuations and water availability, greatly influence the distribution and behaviour of species (Tylianakis *et al.* 2008). GECs can alter host fitness or ecosystem functioning (Shay *et al.* 2015; Webster *et al.* 2016) and are likely to occur in combination. While there is a great deal of research into the effects of each of these on microbial communities (e.g. Schimel *et al.* 2007; Shurin *et al.* 2012; Lloret *et al.* 2014), literature considering the effect of multiple GECs is sparser, and these have complicated and often unpredictable consequences when combined (see Boyd & Hutchins, 2012; Ryalls *et al.* 2013). In this section, we consider how human activities directly and indirectly influence the microbial world. Where applicable, these questions can be considered across multiple biomes and ecosystems, with reference to resulting trophic cascades, in addition to the impacts on multiple biogeochemical processes. We also consider how microbes can be used as a tool for mitigation or bioremediation of human-induced environmental changes, and the ways in which microbes can be included in current evaluations of global change.

24. How can we integrate microbial communities into models of global change?
25. Will ocean acidification, temperature increases and rising sea levels lead to changes in microbial diversity or function, and what will the cascading effects of this be?
26. How do human activities, such as oil and gas drilling, influence the sub-surface microbiome(s)?
27. How will increasing urbanisation affect environmental and host-associated microbial communities?
28. How resilient are different microbial functional groups to ecosystem disturbance?
29. Can we manipulate microbial succession in species-poor soils to encourage repopulation by flora and fauna?

## Environmental processes

Microbes play a fundamental role in environmental processes and ecosystem services, including nutrient cycling and organic matter decomposition (Wieder, Bonan and Allison *et al.* 2013; Creamer *et al.* 2015; Chin, McGrath and Quinn 2016), bioremediation of contaminated habitats or waste systems (Haritash and Kaushik 2009; Oller, Malato and Sanchez-Perez 2011) and influencing greenhouse gas emissions (Singh *et al.* 2010; Bragazza *et al.* 2013; Hu, Chen and He 2015). The ability to use and manipulate these processes has great potential for societal and environmental applications, particularly in extremophiles, which frequently reveal metabolic capabilities and evolutionary solutions not witnessed elsewhere in the microbial world (Coker 2016). However, it is rarely possible to directly link the presence of a specific microbial taxon to a particular ecological process. Other methodological challenges include establishing the relative importance of biotic and abiotic factors in microbial ecosystem function, and determining the appropriate spatial and temporal scale necessary to discriminate links between microbiota and their ecological functions (Bissett *et al.* 2013). Concurrently, a deeper understanding is required of human-induced impacts on the global microbiome through urbanisation, habitat degradation, climate change and the introduction of invasive species, amongst others.

30. How do we successfully establish microbial communities used in bioremediation?
31. How important is the rare microbiome in ecosystem function, and how does this change with stochastic events?
32. To what extent is microbial community diversity and function resilient to short- and long-term perturbations?
33. What is the importance of spatial and temporal variation in microbial community structure and function to key environmental processes and geochemical cycles?
34. How can we accurately measure microbial biomass in a reproducible manner?
35. Which mechanisms do extremophiles use for survival and how can they be exploited?

## Functional diversity

Ecologists are increasingly turning their attention to classifying species based on their activity (function) within an ecosystem, rather than their genotype (Crowther *et al.* 2014). This is particularly relevant for microbial ecology, in which species are hard to define, horizontal gene transfer is rife and taxonomy is often blurred. Understanding how membership within complex and dynamic microbial communities relates to the function of that community is one of the key challenges facing microbial ecology (Widder *et al.* 2016). This is true across a vast range of spatial scales, from microbial dyads to the gut of a *Drosophila* fly to ancient trees and their associated ecosystems, right through to global biogeochemical processes. There is an urgent need to understand how the genome of a microbial community (and in some cases, its host) relates to metabolic capacities. Conversely, there is also a need to understand how ecosystems depend on a particular organism or group of organisms for any given process and function. This section describes the need to move from simply describing microbial diversity to understanding what these organisms are doing, how they are doing it, and what biotic and abiotic drivers are controlling their activity. Each question may derive a suite of different answers, depending on the group of organisms, the habitat and the process.

36. What are the mechanisms driving microbial community structure and function, and are these conserved across ecosystems?
37. What is the relative importance of stochastic vs determinative processes in microbial community assembly?
38. How conserved are microbial functions across different spatial and temporal scales?
39. What is the relative importance of individual 'species' for the functioning of microbial communities?
40. How much functional redundancy is there in microbial communities, and how does functional redundancy affect measures of diversity and niche overlap?
41. How often are functional traits of microbes successfully conferred through horizontal gene transfer?
42. What methods can we use to marry microbial diversity with function; how do we link transcriptomics, proteomics and metabolomics?

43. How do we move beyond correlation to develop predictive models that advance our understanding of microbial community function and dynamics?"
44. How useful are synthetic communities for testing theories about microbial community dynamics and function?

### Evolutionary processes

The role of microorganisms in determining evolutionary outcomes of hosts is being investigated in increasing detail (McFall-Ngai *et al.* 2013). Experimental evolution studies represent a powerful means of quantifying host–microbe and microbe–microbe coevolution, and have highlighted the extraordinary capacity of microbes to act as key mediators of host fitness (e.g. King *et al.* 2016). Whilst experimental coevolution studies provide a framework for linking dyadic interactions to community-scale dynamics (Brockhurst and Koskella 2013), evolutionary principles stemming from macroecology are being applied to microbial communities of humans (Robinson, Bohannan and Young 2010). However, fundamental biological questions that are well studied in macrobiology remain controversial for microbial ecology, for example, the species concept remains a source of debate (Freudenstein *et al.* 2016). The operational taxonomic unit (OTU) has become the standard unit for identifying bacteria at the highest taxonomic resolution possible, yet it is hard to clearly define where taxonomic boundaries lie between two bacteria, and what an OTU really represents in biological terms. This is especially problematic in the context of horizontal gene transfer, which is commonly observed in bacteria and has turned our understanding of evolutionary processes upside down. This section relates to how general ecological principles influence microbial evolution and vice versa, what this means for global biodiversity, and whether evolutionary principles can be utilised for anthropogenic gain.

45. How can a bacterial 'species' be defined?
46. To what extent is faunal and floral biodiversity influenced by microbial communities?
47. To what extent do microbial communities have an equivalent to keystone 'species'?
48. Does the structure of microbial communities conform to the same ecological rules/principles as in other types of communities?
49. How do fundamental shifts in environmental conditions impact the trajectory of microbial evolution?
50. What are the relative selective forces favouring microbial genome expansion or reduction?

### Society and policy

We need to find ways to apply fundamental biological research to the benefit of society and policy. For example, collaboration with social scientists is crucial when investigating public understanding of microbial ecology, as well as using citizen science approaches to tackle microbial ecology research questions. Many questions relating to this area were discussed at the workshop, and here we present four additional questions that were developed at the meeting that relate to societal and policy-based aspects of microbial ecology.

i. How can we best address supply and demand of information about microbial ecology between researchers, clinicians, policy makers and practitioners?
ii. How can we best use social and traditional mass media for early identification of emerging threats to animal and plant health?
iii. How can we develop an open access data repository or integrate existing databases to create a centralised and standardised method for data and methods sharing in microbial ecology?
iv. How can we replace fear-based regulation with risk-based regulation, specifically with regard to the use of microbes in bioremediation and bioaugmentation?

## DISCUSSION

Here we present 50 important research questions across a number of themes relating to the field of microbial ecology. Although there are many other research issues worthy of investigation, it is intended that these questions will be used to inform and direct future research programmes and agendas, particularly in areas where microbial ecology has not previously been considered or applied. In many cases, these questions are deliberately broad to allow researchers to adapt them to their own areas of interest, for example across different systems, or to varying spatial scales. Across many questions there was strong recognition of the vast metabolic capabilities of microorganisms and microbial communities, and the need to utilise this power to improve human and animal health and wellbeing. Some themes addressed various existing mechanisms for exploiting microbial processes, namely bioremediation, soil improvement, water treatment and probiotic suppression of pathogen resistance. As these are already active areas of research, the questions posed here are structured to provide a framework by which these efforts can be directed in the future.

A predominant theme that emerged was the need to integrate knowledge between different research areas, for example, the application of information from human microbiome studies to the study of other non-model host organisms, and the potential to apply macroecological frameworks to microecological concepts. Many fundamental biological questions that are well studied in classical ecology remain controversial for microbial ecology, and the species concept (Freudenstein *et al.* 2016), taxonomy, and how the OTU should be defined for microorganisms, generated multiple questions (e.g. see 'evolutionary processes' theme). Classical community ecology concepts should not be overlooked when considering microbial dynamics (Rynkiewicz, Pedersen and Fenton 2015) and, conversely, microbial communities may prove useful models for general ecology due to their short generation times, reproducibility and ease of use in the laboratory environment (Brockhurst and Koskella 2013; Libberton, Horsburgh and Brockhurst 2015; King *et al.* 2016). There have been a number of calls for the medical profession to look to ecological and evolutionary tools when seeking to understand epidemiology (Johnson, de Roode and Fenton 2015), investigating novel antibacterial agents (Vale *et al.* 2016), and considering multihost, multiagent disease systems (Buhnerkempe *et al.* 2015).

The 'host–microbiome interactions' theme considered the need to understand factors influencing microbiome composition, which in turn have consequences for a myriad of host traits, including disease susceptibility and host evolution (Chisholm *et al.* 2006; Archie and Theis 2011; Spor, Koren and Ley 2011; Cho and Blaser 2012; Zilber-Rosenberg and Rosenberg 2008; McFall-Ngai *et al.* 2013; McFall-Ngai 2015). As this theme considered microbiota from the perspective of the host, there was some overlap with the 'health and infectious diseases' and

'evolutionary processes' themes. Probiotics were discussed as a viable and promising alternative to current strategies in a number of contexts in these themes, to improve individual health; to decrease disease susceptibility of humans and other animals; to enhance nutritional quality of food; and to mitigate the negative impacts of antibiotic use across humans, livestock, aquaculture and agriculture (Martín *et al.* 2013; Newaj-Fyzul, Al-Harbi and Austin 2014; Smith 2014; Fox 2015). Developing personalised probiotic-based therapies requires complementary diversity and functional-based studies in order to elucidate the specific roles of microbiota in health and disease, and thus how microbial communities can be manipulated.

Questions considered in both the 'functional diversity' theme and the 'environmental processes' theme raised a common need to understand changes in microbial community structure and function across spatial and temporal scales (Carmona *et al.* 2016). Establishing appropriate spatial scales for studying microbial processes is an outstanding challenge: microorganisms can orchestrate ecosystem functioning across whole biomes (Sheffer *et al.* 2015), yet fungi exhibit low mobility on tree barks (Koufopanou *et al.* 2006; Robinson, Pinharanda and Bensasson 2016), and an air void in soil can be an insuperable barrier for a bacterium. Similarly, drawing meaningful conclusions about microbial processes requires understanding of their temporal variability, for example, diurnal influences (Shurpali *et al.* 2016) or lags behind changes in ecosystem drivers (Allison and Martiny 2008).

A subject common to a number of themes was the role of individual species versus consortia in community functioning. The question of defining bacterial species is a contentious topic, and the issue remains whether some microbial taxa act as keystones in ecosystem functions. Many microbial surveys carry the implicit assumption that the most abundant taxa are also the most important, yet rare species can be hugely significant if they are highly active and/or monopolise a particular process (Lynch and Neufeld 2015). The collective metabolic capabilities of microorganisms have great potential for *in situ* applications such as bioremediation, particularly when used in multispecies consortia (Mikesková *et al.* 2012). Successful bioremediation and environmental management requires the introduction of new assemblages into an established community, or stimulation of key members of the community *in situ* (Rillig *et al.* 2006). In turn, predicting the successful establishment of deliberately introduced organisms depends on an understanding of the principles underlying microbial community formation and structure. Despite these challenges, functional diversity modelling has successfully been applied to the ecological restoration of some plant communities (Laughlin 2014). Closely linked to this is the issue of functional redundancy, and to what extent it is possible to lose species without affecting ecosystem functions. Already there is evidence that microbial communities may be less functionally redundant than macroorganism communities (Delgado-Baquerizo *et al.* 2016). This issue ties into fundamental ecological concepts, such as niche theory (Carmona *et al.* 2016); if multiple organisms are carrying out the same process, apparently interchangeably, how do they avoid competitively excluding one another? The concept of keystone species has been shown to be applicable to microbes (Neufeld *et al.* 2008; Pester *et al.* 2010; Ze *et al.* 2012; Yu *et al.* 2016), yet further work is needed to characterise the extent to which keystone functions occur in different environments and whether these can be consistently identified (Anderson 2003; Pester *et al.* 2010).

The need for open access databases and repositories, both in the context of data sharing and for methods and protocols, was reflected in the questions shortlisted for the 'society and policy' theme. Discussions included the benefits of forming collaborative and open research communities, and the need to ensure the legacy of academic research through improving regulation and policy and engagement with the public. Fear-based regulation of research, grounded in alarmist or populist campaigns, as opposed to risk-based regulation built upon evidence, was identified as a possible obstacle to progress, which could be addressed through greater interaction between microbial ecologists and the public at both governmental and grass roots levels. Large-scale assessments of ecosystem services and degradation acknowledge the paucity of data on microbial impacts, presumably because there are no convincing large-scale messages that can be derived at this stage (Norris *et al.* 2011). Microbial diversity is therefore rarely considered when estimates of biodiversity are required for policy or management decisions. That said, the increasing recognition of the fundamental impact of the microbial world on the functioning of larger-scale processes has made the deliberate manipulation of the microbial world a controversial subject, which was reflected in the number of draft questions submitted related to bioremediation and bioaugmentation (see Supplementary Information). Collaboration with social scientists was identified as crucial in gauging public understanding of microbial ecology, and citizen science approaches were considered as tools to tackle key microbial ecology research questions.

The 50 questions identified here cover a broad range of topics, but some over-arching themes recur across multiple questions, including a recognition that microbes play an important role in a variety of different processes and systems, which may be exploited to solve real-world problems. There were some similarities between the questions identified here and those identified by previous workshops of a similar nature. For example, questions relating to soil health and biodiversity (Dicks *et al.* 2013), a requirement for developing a theoretical understanding of micro- and macroecological concepts (Prosser *et al.* 2007; Sutherland *et al.* 2013a) and disease dynamics (Prosser *et al.* 2007; Sutherland *et al.* 2013a) have a degree of commonality with this list. This indicates that the ecological theory underpinning many research questions transcends scientific disciplines, and that there is still much work to be done at both theoretical and applied levels. Within these 50 questions, we have tried to provide a focus for researchers addressing scientific questions from a microbial perspective, regardless of their background. It is expected that these questions will facilitate interesting discussion and new, exciting, interdisciplinary research. The list is by no means exhaustive, and we recognise that the questions presented here are relatively community-centric, primarily due to the recent expansion in methodological approaches that have improved our understanding of microbial community diversity and function. That said, other areas of microbial ecology should not be ignored or forgotten. Given the rapidly evolving field of microbial ecology, it is expected that future workshops with a wide draw will be held to ensure that the identification of research priorities and areas of interest is a continuing process.

## SUPPLEMENTARY DATA

Supplementary data are available at *FEMSEC* online.

## ACKNOWLEDGEMENTS

The authors thank Francis Brearley and Michael Cunliffe for their submissions to the pre-workshop questions, and Jessica

Hall, Daniel Henk and Matt Lloyd Jones for their contributions at the workshop.

## FUNDING

This work was supported by contributions from the British Ecological Society and the University of Salford towards funding the workshop. KMF was funded by the Finnish Cultural Foundation, NLMF Colciencias, MCM by Earth and Life Systems Alliance, and WJS by Arcadia. The positions and opinions presented in this article are those of the authors alone and are not intended to represent the views or scientific works of the European Food Safety Authority.

***Conflict of interest.*** None declared.

## REFERENCES

Allison S, Martiny J. Resistance, resilience, and redundancy in microbial communities. *P Natl Acad Sci USA* 2008;**105** S1:11512–9.

Anderson TH. Microbial eco-physiological indicators to assess soil quality. *Agr Ecosyst Environ* 2003;**98**:285–93.

Archie EA, Theis KR. Animal behaviour meets microbial ecology. *Anim Behav* 2011;**82**:425–36.

Baothman OA, Zamzami MA, Taher I *et al.* The role of gut microbiota in the development of obesity and diabetes. *Lipids Health Dis* 2016;**15**:108.

Bissett A, Brown MV, Siciliano SD *et al.* Microbial community responses to anthropogenically induced environmental change: towards a systems approach. *Ecol Lett* 2013;**16**:128–39.

Boers SA, Jansen R, Hays JP. Suddenly everyone is a microbiota specialist. *Clin Microbiol Infect* 2016;**22**:581–2.

Boyd PW, Hutchins DA. Understanding the responses of ocean biota to a complex matrix of cumulative anthropogenic change. *Mar Ecol Prog Ser* 2012;**470**:125–35.

Bragazza L, Parisod J, Buttler A *et al.* Biogeochemical plat-soil microbe feedback in response to climate warmings in peatlands. *Nat Clim Change* 2013;**3**:273–7.

Brockhurst MA, Koskella B. Experimental coevolution of species interactions. *Trends Ecol Evol* 2013;**28**:367–75.

Buhnerkempe MG, Roberts MG, Dobson AP *et al.* Eight challenges in modelling disease ecology in multi-host, multi-agent systems. *Epidemics* 2015;**10**:26–30.

Buonaurio R, Moretti C, Passos da Silva D *et al.* The olive knot disease as a model to study the role of interspecies bacterial communities in plant disease. *Front Plant Sci* 2015;**6**:434.

Carmona C, de Bello F, Mason N *et al.* Traits without borders: integrating functional diversity across scales. *Trends Ecol Evol* 2016;**31**:382–94.

Chin JP, McGrath JW, Quinn JP. Microbial transformations in phosphate biosynthesis and catabolism, and their importance in nutrient cycling. *Curr Opin Chem Biol* 2016;**31**:50–7.

Chisholm ST, Coaker G, Day B *et al.* Host-microbe interactions: shaping the evolution of plant immune response. *Cell* 2006;**126**:803–14.

Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. *Nat Rev Genet* 2012;**13**:260–70.

Coker JA. Extremophiles and biology: current uses and prospects. *F1000 Res* 2016;**5**: F1000 FacultyRev-396.

Creamer CA, de Menezes AB, Krull ES *et al.* Microbial community structure mediates response of soil C decomposition to litter addition and warming. *Soil Biol Biochem* 2015;**80**:175–88.

Crowther TW, Maynard D, Crowther TR *et al.* Untangling the fungal niche: the trait-based approach. *Front Microbiol* 2014;**5**:579.

Daskin JH, Alford RA. Context-dependent symbioses and their potential roles in wildlife diseases. *Proc Biol Sci* 2012;**279**:1457–65.

Delgado-Baquerizo M, Giaramida L, Reich P *et al.* Lack of functional redundancy in the relationship between microbial diversity and ecosystem functioning. *J Ecol* 2016;**104**:936–46.

Dicks LV, Abrahams A, Atkinson J *et al.* Identifying key knowledge needs for evidence-based conservation of wild insect pollinators: a collaborative cross-sectoral exercise. *Insect Conserv Diver* 2012;**6**:435–46.

Dicks LV, Bardgett RD, Bell J *et al.* What do we need to know to enhance the environmental sustainability of agriculture? A prioritisation of knowledge needs for the UK food system. *Sustainability* 2013;**5**:3095–115.

Ellouze W, Esmaeili-Taheri A, Bainard LD *et al.* Soil fungal resources in annual cropping systems and their potential for management. *BioMed Res Int* 2014:531824.

Freudenstein JV, Broe MB, Folk *et al.* Biodiversity and the species concept-Lineages are not enough. *Syst Biol* 2016;**10**.1093.

Fox JL. Agricultural probiotics enter spotlight. *Nat Biotechnol* 2015;**33**:122.

Franklin AM, Aga DS, Cytryn E *et al.* Antibiotics in agroecosystems: introduction to the special section. *J Environ Qual* 2016;**45**:377.

Franzosa E, Hsu T, Sirota-Madi A *et al.* Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat Rev Microbiol* 2015;**13**:360–72.

Gilbert JA, Quinn RA, Debelius J *et al.* Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* 2016;**535**:94–103.

Haberl H, Erb KH, Krausmann F *et al.* Quantifying and mapping the human appropriation of net primary production in earth's terrestrial ecosystems. *P Natl Acad Sci USA* 2007;**104**:12942–7.

Halpern BS, Walbridge S, Selkoe KA *et al.* A global map of human impact on marine ecosystems. *Science* 2008;**319**:948–52.

Haritash AK, Kaushik CP. Biodegradation aspects of Polycyclic Aromatic Hydrocarbons (PAHs): A review. *J Hazard Mater* 2009;**169**:1–15.

Hartmann M, Frey B, Mayer J *et al.* Distinct soil microbial diversity under long-term organic and conventional farming. *ISME J* 2015;**9**:1177–94.

Helgason T, Daniell TJ, Husband R *et al.* Ploughing up the wood-wide web? *Nature* 1998;**394**:431.

Hiraoka S, Yang CC, Iwasaki W. Metagenomics and bioinformatics in microbial ecology: current status and beyond. *Microbes Environ* 2016;**31**:204–12.

Hu HW, Chen D, He JZ. Microbial regulation of terrestrial nitrous oxide formation: understanding the biological pathways for prediction of emission rates. *FEMS Microbiol Rev* 2015;**39**:729–49.

Jacobsen CS, Hjelmsø MH. Agricultural soils, pesticides and microbial diversity. *Curr Opin Biotechnol* 2014;**27**:15–20.

Johnson PT, de Roode JC, Fenton A. Why infectious disease research needs community ecology. *Science* 2015;**349**: 1259504.

Jones AC, Mead A, Kaiser MJ *et al.* Prioritization of knowledge needs for sustainable aquaculture: a national and global perspective. *Fish Fisheries* 2014;**16**:668–83.

Koch H, Schmid-Hempel P. Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *P Natl Acad Sci USA* 2011;**108**:19288–92.

Koufopanou V, Hughes J, Bell G *et al*. The spatial scale of genetic variation in a model organism: the wild yeast Saccharomyces paradoxus. *Phil Trans B* 2006;**29**:1941–6.

King KC, Brockhurst MA, Vasieva O *et al*. Rapid evolution of microbe-mediated protection against pathogens in a worm host. *ISME J* 2016; doi: 10.1038/ismej.2015.259.

Krishnasamy V, Otte J, Silbergeld E. Antimicrobial use in Chinese swine and broiler poultry production. *Antimicrob Resist Infect Control* 2015;**4**:17.

Laughlin D. Applying trait-based models to achieve functional targets for theory-driven ecological restoration. *Ecol Lett* 2014;**17**:771–84.

Levy SB. Antibiotic resistance: an ecological imbalance. In: *Antibiotic resistance: Origins, Evolution, Selection and Spread*, vol. 207. Wiley, Chichester: Ciba Foundation Symposium, 1997, 1–17.

Ley RE, Hamady M, Lozupone C *et al*. Evolution of mammals and their gut microbes. *Science* 2008;**320**:1647–51.

Libberton B, Horsburgh MJ, Brockhurst MA. The effects of spatial structure, frequency dependence and resistance evolution on the dynamics of toxin-mediated microbial invasions. *Evol Appl* 2015;**7**:738–50.

Lloret F, Mattana S, Yuste J-C. Climate-induced die-off affects plant-soil-microbe ecological relationship and functioning. *FEMS Microbiol*. 2014;**91**:1–12.

Lozupone CA, Li M, Campbell TB *et al*. Alterations in the gut microbiota associated with HIV-1 infection. *Cell Host Microbe* 2013;**14**:329–39.

Lynch M, Neufeld J. Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* 2015;**13**:217–29.

Martín M, Miquel S, Ulmer J *et al*. Role of commensal and probiotic bacteria in human health: a focus on inflammatory bowel disease. *Microb Cell Fact* 2013;**12**:71.

McFall-Ngai M, Hadeldb MG, Boschc TCG *et al*. Animals in a bacterial world, a new imperative for the life sciences. *P Natl Acad Sci USA* 2013;**110**:3229–36.

McFall-Ngai MJ. Giving microbes their due–animal life in a microbially dominant world. *J Exp Biol* 2015;**218**:1968–73.

Mikesková H, Novotný C, Svobodová K. Interspecific interactions in mixed microbial cultures in a biodegradation perspective. *Appl Microbiol Biot* 2012;**95**:861–70.

Morris CE, Bardin M, Berge O *et al*. Microbial biodiversity: approaches to experimental design and hypothesis testing in primary scientific literature from 1975 to 1999. *Microbiol Mol Biol R* 2002;**66**:592–616.

Mueller K, Ash C, Pennisi E *et al*. The gut microbiota. *Science* 2012;**336**:1245.

Newaj-Fyzul A, Al-Harbi H, Austin B. Review: Developments in the use of probiotics for disease control in aquaculture, *Aquaculture* 2014;**431**:1–11.

Neufeld JD, Chen Y, Dumont MG *et al*. Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. *Environ Microbiol* 2008;**10**:1526–35.

Norris K, Bailey M, Baker S *et al*. Biodiversity in the context of ecosystem services. *The UK National Ecosystem Assessment Technical Report*. UK National Ecosystem Assessment, UNEP-WCMC, Cambridge, 2011.

Oller I, Malato S, Sanchez-Perez JA. Combination of advanced oxidation processes and biological treatment for wastewater decontamination- a review. *Sci Total Environ* 2011;**409**:4141–66.

Pautasso M, Petter F, Rortais A *et al*. Emerging risks to plant health: a European perspective. *CAB Reviews* 2015;**10**:21.

Parsons ECM, Favero B, Aguirre AA *et al*. Seventy-one important questions for conservation of marine biodiversity. *Conserv Biol* 2014;**28**:1206–14.

Peay K. Back to the future: natural history and the way forward in modern fungal ecology. *Fungal Ecol* 2014;**2**:4–9.

Pester M, Bittner N, Deevong P *et al*. A 'rare biosphere' microorganism contributes to sulfate reduction in a peatland. *ISME J* 2010;**4**:1591–602.

Prosser JI, Bohannan BJM, Curtis TP *et al*. The role of ecological theory in microbial ecology. *Nature* 2007;**5**:384–92.

Rebollar EA, Antwis RE, Becker MH *et al*. Using 'Omics' and integrated multi-omics approaches to guide probiotic selection to mitigate chytridiomycosis and other emerging infectious diseases. *Front Microbiol* 2016;**7**:68.

Redford KH, Segre JA, Salafsky N *et al*. Conservation and the microbiome. *Conserv Biol* 2012;**26**:195–7.

Ridaura VK, Faith JJ, Rey FE *et al*. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* 2013;**341**:1241214.

Rillig MC, Mummey DL. Mycorrhizas and soil structure. *New Phyto*. 2006;**171**:41–53.

Robinson CJ, Bohannan BJ, Young VB. From structure to function: the ecology of host-associated microbial communities. *Microbiol Mol Biol R* 2010,**74**:453–76.

Robinson HA, Pinharanda A, Bensasson DB. Summer temperature can predict the distribution of wild yeast populations. *Ecol Evol* 2016;**27**:1236–50.

Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immun* 2009;**9**:313–23.

Ryalls JMW, Riegler M, Moore BD *et al*. Effects of elevated temperature and CO2 on aboveground-belowground systems: a case study with plants, their mutualistic bacteria and root/shoot herbivores. *Front Plant Sci* 2013;**4**:445.

Rynkiewicz EC, Pedersen AB, Fenton A. An ecosystem approach to understanding and managing within-host parasite community dynamics. *Trends Parasitol* 2015;**31**:212–21.

Sato Y, Civiello M, Bell S *et al*. Intergrated approach to understanding the onset and pathogenesis of black band disease in corals. *Environ Microbiol* 2016;**18**:752–65.

Schimel J, Balser TC, Wallenstein M. Microbial stress-response physiology and its implications for ecosystem function. *Ecology* 2007;**88**:1386–94.

Schneider T, Keiblinger K, Schmid E *et al*. Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *ISME J* 2012;**6**:1749–62.

Serino M, Nicholas M, Trabelsi MS *et al*. Young microbes for adult obesity. *Pediatr Obes* 2016; 10.1111.

Shay PE, Winder RS, Trofymow JA. Nutrient-cycling microbes in coastal Douglas-fir forests: regional-scale correlation between communities, in situ climate, and other factors. *Front. Microbiol*. 2015;**6**:1097.

Sheffer E, Batterman S, Levin S *et al*. Biome-scale nitrogen fixation strategies selected by climatic constraints on nitrogen cycle. *Nat Plants* 2015;**1**:15182.

Shurin JB, Clasen JL, Greig HS *et al*. Warming shifts top-down and bottom-up control of pond food web structure and function. *Phil Trans B* 2012;**367**:3008–17.

Shurpali N, Rannik U, Jokinen S *et al*. Neglecting diurnal variations leads to uncertainties in terrestrial nitrous oxide emissions. *Sci Rep* 2016;**6**:25739.

Singh BK, Bardgett RD, Smith P *et al*. Microorganisms and climate change: terrestrial feedbacks and mitigation options. *Nat Rev Microbiol* 2010;**8**:779–90.

Smith JMA. Review of avian probiotics. *J Avian Med Surg* 2014;**28**:87–94.

Smits HH, Hiemstra PS, Prazeres da Costa C *et al.* Microbes and asthma: opportunities for intervention. *J Allergy Clin Immunol* 2016;**137**:690–7.

Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* 2011;**9**:279–90.

Sutherland WJ, Adams WM, Aronson RB *et al.* One hundred questions of importance to the conservation of global biological diversity. *Conserv Biol* 2009;**23**:557–67.

Sutherland WJ, Armstrong-Brown S, Armsworth PR *et al.* Identification of 100 fundamental ecological questions. *J Ecol* 2013a;**101**:58–67.

Sutherland WJ, Fleishman E, Mascia MB *et al.* Methods for collaboratively identifying research priorities and emerging issues in science and policy. *Methods Ecol Evol* 2011;**2**:238–47.

Sutherland WJ, Goulden C, Bell K *et al.* 100 Questions: Identifying research priorities for poverty prevention and reduction. *J Poverty Soc Justice* 2013b;**21**:189–205.

Tam CC, Rodrigues LC, Viviani L *et al.* Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut* 2012;**61**:69–77.

Tompkins DM, Carver S, Jones ME *et al.* Emerging infectious diseases of wildlife: a critical perspective. *Trends Parasitol* 2015;**31**:149–59.

Tylianakis JM, Didham RK, Bascompte J *et al.* Global change and species interactions in terrestrial ecosystems. *Ecol Lett* 2008;**11**:1351–63.

Ubeda C, Bucci V, Caballero S *et al.* Intestinal microbiota containing Barnesiella species cures vancomycin-resistant Enterococcus faecium colonization. *Infect Immun* 2013;**81**:965–73.

Vale PF, McNally L, Doeschl-Wilson A *et al.* Beyond Killing: Can we find new ways to manage infection? *Evol Med Public Health* 2016, doi: 10.1093/emph/eow012.

Vayssier-Taussat M, Albina E, Citti C *et al.* Shifting the paradigm from pathogens to pathobiome: new concepts in the light of meta-omics. *Front Cell Infect Microbiol* 2014;**4**:29.

Walter J, Ley R. The human gut microbiome: ecology and recent evolutionary changes. *Annu Rev Microbiol* 2011;**65**:411–29.

Webster NS, Negri AP, Botte ES *et al.* Host-associated coral reef microbes respond to the cumulative pressures of ocean warming and ocean acidification. *Sci Rep* 2016;**6**:19324.

Weiss B, Aksoy S. Microbiome influences on insect host vector competence. *Trends Parasitol* 2011;**27**:514–22.

WHO. *Tackling Antibiotic Resistance from a Food Safety Perspective in Europe*. Copenhagen, Denmark: WHO Regional Office for Europe, 2011. http://www.euro.who.int/en/publications/abstracts/tackling-antibiotic-resistance-from-a-food-safety-perspective-in-europe (date last accessed, 19 July 2016).

Widder S, Widder S, Allen RJ *et al.* Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J* 2016, doi: 10.1038/ismej.2016.45.

Wieder WR, Bonan GB, Allison SD. Global soil carbon projections are improved by modelling microbial processes. *Nat Clim Change* 2013;**3**:909–12.

Willing BP, Russell SL, Finlay BB. Shifting the balance: antibiotic effects on host–microbiota mutualism. *Nat Rev Microbiol* 2011;**9**:233–43.

Yu Z, Krause SMB, Beck DAC *et al.* Synthetic ecology perspective: how well does behavior of model organisms in the laboratory predict microbial activities in natural habitats? *Front Microbiol* 2016;**7**:1–7.

Ze X, Duncan SH, Louis P *et al.* *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J* 2012;**6**:1535–43.

Zilber-Rosenberg I, Rosenberg E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol Rev* 2008;**32**:723–35.

Zuber SM, Villamil MB. Meta-analysis approach to assess effect of tillage on microbial biomass and enzyme activities. *Soil Biol Biochem* 2016;**97**:176–87.

# Genetics and Ecology of Isoprene Degradation

Andrew T. Crombie, Nasmille L. Mejia-Florez, Terry J. McGenity, and J. Colin Murrell

## Contents



**Abstract**

Approximately 550 million tonnes of the monoterpene, isoprene, are emitted to the atmosphere annually, principally from terrestrial plants. In contrast to methane, which is emitted in similar quantities, little is known about the biodegradation of isoprene. However, 30 years ago, bacteria capable of living on isoprene as a sole source of carbon and energy were described, although they were not investigated in detail. Recently there has been renewed interest in the potential of bacteria living in soils, marine sediments, and on the leaves of plants to degrade isoprene. Isolates capable of isoprene metabolism use a multicomponent soluble monooxygenase, which contains a diiron center at the active site, to oxidize isoprene to the epoxide, and all isolates described to date depend on glutathione for subsequent metabolic steps. The diversity of isoprene degraders

A. T. Crombie (✉)
School of Biological Sciences, University of East Anglia, Norwich, UK
e-mail: a.crombie@uea.ac.uk

N. L. Mejia-Florez · J. C. Murrell
School of Environmental Sciences, University of East Anglia, Norwich, UK

T. J. McGenity
School of Biological Sciences, University of Essex, Colchester, UK



F. Rojo (ed.), *Aerobic Utilization of Hydrocarbons, Oils and Lipids*, Handbook of Hydrocarbon and Lipid Microbiology, https://doi.org/10.1007/978-3-319-39782-5_27-1

has been investigated in terrestrial and marine environments using DNA-stable isotope probing (DNA-SIP), together with the use of gene probes targeting the monooxygenase active-site subunit. Gaps in our knowledge and future research directions are described.

## 1 Introduction

Isoprene (2-methyl-1,3-butadiene) is one of the most highly produced biogenic volatile organic compounds (BVOCs) emitted to the atmosphere, accounting for approximately 550 Tg C $y^{-1}$, or 1/3 of total BVOCs (Guenther et al. 2006, 2012). This is similar in magnitude to the release of methane, with all other BVOCs comprising the remaining third. Isoprene is volatile, with a boiling point of 34 °C and can be considered as a trace gas; its high reactivity in the atmosphere has a major influence on Earth's climate. The effects of isoprene on climate are complex and not fully understood. It reacts with hydroxyl radicals, reducing the oxidative capacity of the atmosphere and resulting in a slower turnover of methane, a potent greenhouse gas, leading to global warming. Isoprene also reacts with oxides of nitrogen in the atmosphere, resulting in increased ozone levels, with effects on air quality and health (Sanderson et al. 2003; Pacifico et al. 2009). The oxidation products can result in secondary organic aerosols which promote increased cloud formation and global cooling (Carlton et al. 2009). Therefore under different circumstances, isoprene can act as both a global warming and a global cooling gas.

Isoprene is a key building block for isoprenoids, which consist of two or more isoprene units and are produced by all free-living organisms. This large family of molecules includes, for example, carotenoids, sterols, chlorophyll, quinones, archaeal lipids, and hopanoids. The precursor molecules for isoprenoids are dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP), which are synthesized in animals, fungi, archaea, some bacteria, and the cytosol of plants using the mevalonate (MVA) pathway or by the methylerythritol phosphate (MEP) pathway in chloroplasts and most bacteria (Rohmer 1999). The vast majority of isoprene (approximately 90%) is produced globally by terrestrial plants by the action of isoprene synthase on DMAPP in chloroplasts (Sharkey et al. 2008). Interestingly, not all plants produce isoprene, with both high and low producers, even among closely related species – for example, all American oaks emit isoprene whereas many European oaks do not (Loreto et al. 1998). High isoprene-emitting trees typically divert 2% of fixed carbon to isoprene production, and in some cases considerably more (Sharkey et al. 1996), so it is striking that the reasons for this outlay of carbon and energy are not fully understood. There is good evidence that isoprene can alleviate heat and oxidative stress (Sharkey et al. 2008; Zeinali et al. 2016) and other proposed roles for isoprene biosynthesis include plant signaling, prevention from herbivory and as a way to dissipate excess energy from photosynthesis (Magel et al. 2006; Loivamäki et al. 2008). Certain emerging crop plants, for example palm oil, are high-isoprene emitters, and there is an increasing interest in the effects on air quality of the development of isoprene-emitting agroforestry (Hewitt et al. 2009).

The remaining 10% of isoprene produced in the biosphere is attributed to bacteria, fungi, algae, and animals in both terrestrial and aquatic environments (Gelmont et al. 1981; Fares et al. 2008; Bäck et al. 2010; Exton et al. 2015). For example, *Bacillus subtilis* has been shown to produce isoprene, maybe as a consequence of stress in this bacterium (Kuzma et al. 1995). In the marine environment, macro- and microalgae are the major producers, responsible for a poorly constrained flux of 0.1–10 Tg C $y^{-1}$ (Palmer and Shaw 2005; Luo and Yu 2010; Shaw et al. 2010; Srikanta Dani et al. 2017). As seen with plants, isoprene emissions by marine microalgae increase in response to higher temperature and light intensity and so may protect these organisms during periods of stress (Exton et al. 2013). Isoprene is also produced industrially (approximately 0.8 Tg $y^{-1}$) and used primarily for polyisoprene elastomer (synthetic rubber) production (Morais et al. 2015). Since little is known about bacterial isoprene synthases, strategies have been developed to express isoprene synthase genes from plants in heterologous systems, including *Escherichia coli*, *Saccharomyces*, and *Synechocystis* (Marienhagen and Bott 2013; Lv et al. 2014).

In comparison to the global methane cycle, information regarding the biogeochemical cycling of isoprene is rather sparse and there are many unknowns, particularly estimates of both production and consumption in the biosphere and the internal recycling of isoprene in soils, on the surface of plants, and in the marine environment. Microbial production and/or consumption of isoprene have previously been reviewed by Fall and Copley (2000), Shennan (2006), and McGenity et al. (2017). The global isoprene cycle is shown in Fig. 1. Here we review the mechanisms by which microbes degrade isoprene, and their diversity in the environment.

## 2 Microbial Consumption of Isoprene

Soils have been recognized as a sink for isoprene for over 20 years. In the terrestrial environment, temperate, tropical, and boreal soils were shown to rapidly consume isoprene at concentrations of 500 ppbv in laboratory-based experiments. Field chamber experiments in temperate forests also revealed that soils could deplete isoprene added to chambers at ~400 ppbv down to below the detection limits of about 5 ppbv within 1 h (Cleveland and Yavitt 1997, 1998). Temperate agriforest and model tropical rainforest mesocosm experiments revealed rapid in situ consumption of isoprene (Pegoraro et al. 2005, 2006). More recently, continuous flow experiments with temperate forest soils revealed that these systems consume isoprene over a range of concentrations (2–200 ppbv), with substantial rates of isoprene removal even at low (20 ppbv) concentrations (Gray et al. 2015). The first demonstration of microbial consumption of isoprene in the marine environment was by Acuña Alvarez et al. (2009), who observed isoprene-degrading bacteria in estuarine, coastal, and open marine waters. They also demonstrated that isoprene produced by marine microalgae cultures could be consumed by isoprene-degrading bacteria, an important observation proving that marine microbes could benefit directly from isoprene, produced by microalgae and without using artificially high laboratory

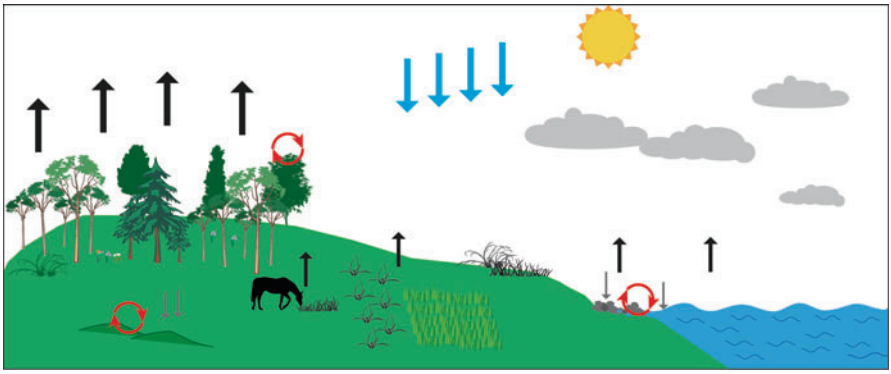


**Fig. 1** Isoprene is emitted to the atmosphere by trees, plants, algae, animals, fungi, and bacteria in the terrestrial and marine environment (black arrows), where it is rapidly photochemically oxidized (blue arrows). Isoprene may also be taken up from the atmosphere by microbes in soils and aquatic ecosystems and sediments (grey arrows). Consumption of isoprene by microbes at the point of release is shown as red circular arrows

concentrations. This study also resulted in the isolation of a number of isoprene-degrading bacteria from these marine environments (see later).

## 3 Bacterial Degradation of Isoprene

The first reports of isolation of bacteria growing on isoprene as sole carbon and energy source were by van Ginkel et al. (1987a, b), Ewers et al. (1990), and Cleveland and Yavitt (1997). Soil enrichments with isoprene under aerobic conditions yielded bacteria assigned to the actinobacterial genera *Rhodococcus, Nocardia*, and *Arthrobacter* and to the proteobacterial genus *Alcaligenes*. More recently aerobic isoprene-degrading *Pseudomonas*, *Klebsiella*, and *Alcaligenes* strains have been reported (Srivastva et al. 2015), but, as with earlier studies, these have not been characterized in any detail. Our recent work has provided collections of aerobic isoprene-degrading bacteria for further study with a number of *Rhodococcus* species being isolated from soils and the leaves of isoprene-producing trees such as Poplar and Willow (El Khawand et al. 2016; Murphy 2017). Marine sediments have also yielded a variety of Gram-positive and Gram-negative aerobic isoprene degraders (Acuña Alvarez et al. 2009; Johnston et al. 2017) (Fig. 2). To our knowledge, no anaerobic bacteria, archaea, or fungi growing on isoprene have yet been reported.

The most well-characterized isoprene-degrading bacterium described so far is *Rhodococcus* sp. AD45, isolated from freshwater sediment by Janssen and colleagues (van Hylckama Vlieg et al. 1998). In this aerobe, the initial oxidation of isoprene to 1,2-epoxyisoprene is carried out by the enzyme isoprene monooxygenase (IsoMO), a multicomponent soluble diiron monooxygenase (SDIMO) belonging to the same large family of enzymes as soluble methane monooxygenase, toluene monooxygenase, and alkene monooxygenase (Leahy et al. 2003; Holmes

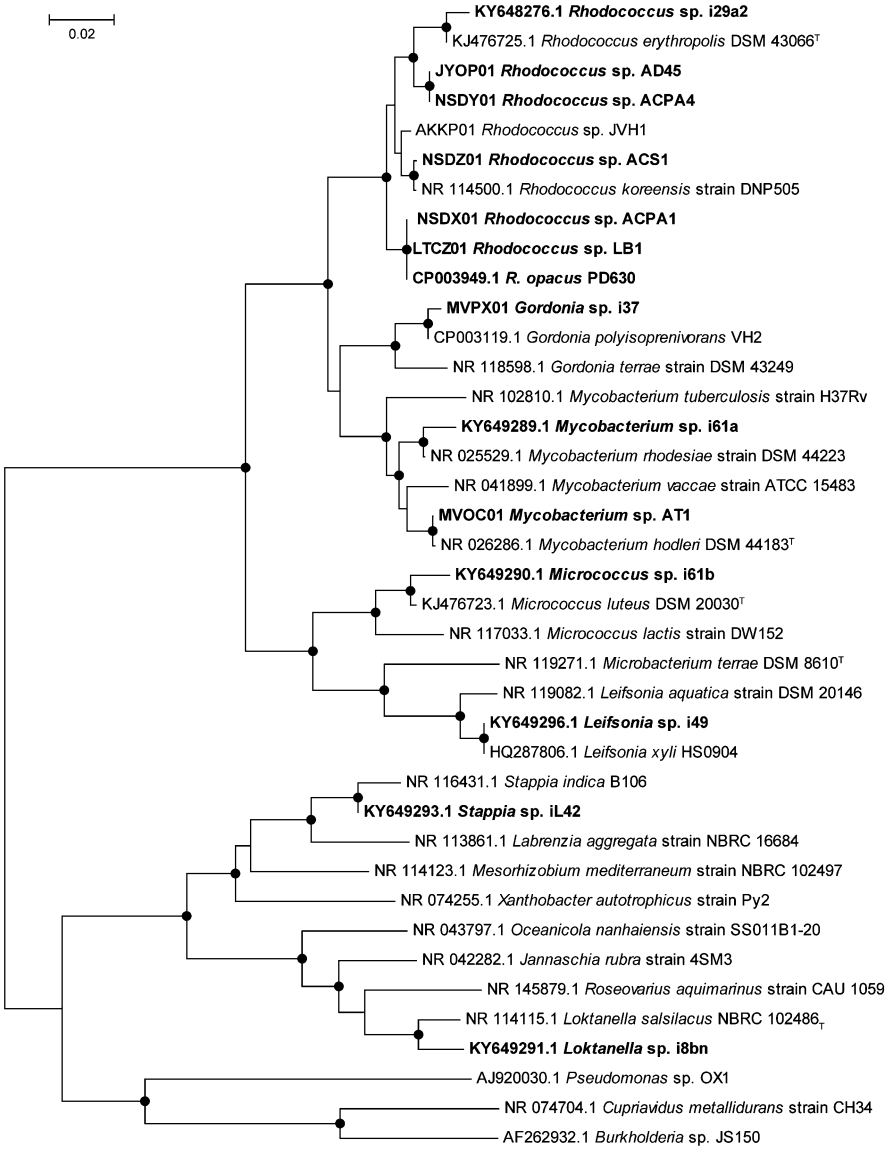



**Fig. 2** The 16S rRNA gene-based phylogeny of isoprene degraders (shown in bold) together with other representative strains. The tree was drawn using the Neighbor Joining method in Mega6 (Tamura et al. 2013) with pairwise deletion, resulting in 1594 nucleotide positions in the analysis. Bootstrap values (1000 replications) greater than 75% are shown by black circles at the nodes. The scalebar shows base substitutions per site

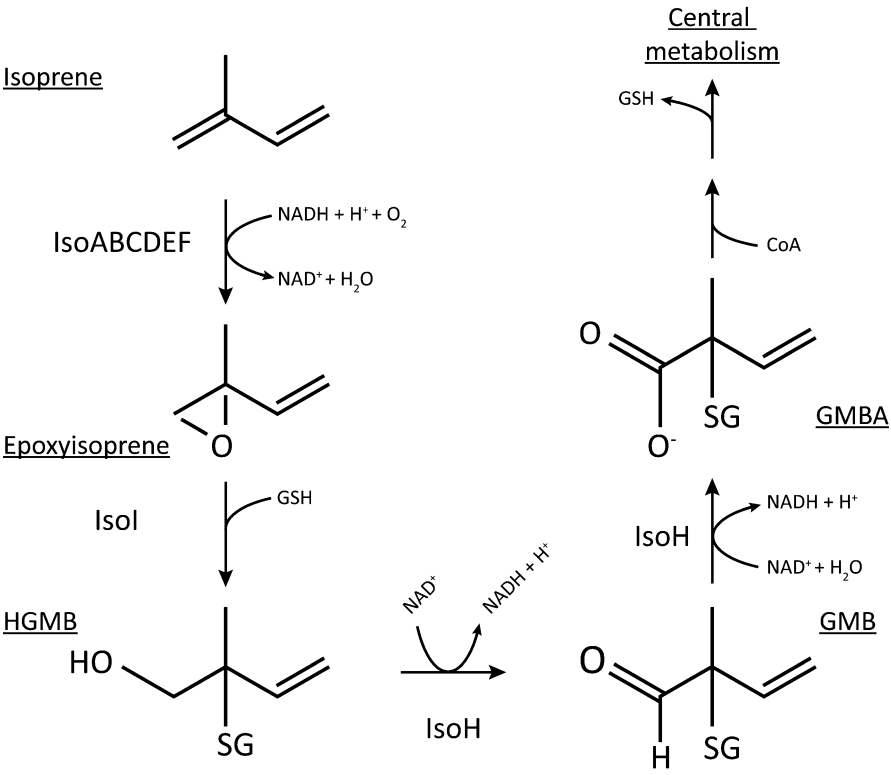



**Fig. 3** The isoprene degradation pathway. HGMB, 1-hydroxy-2-glutathionyl-2-methyl-3-butene; GMB, 2-glutathionyl-2-methyl-3-butenal; GMBA, 2-methyl-2-glutathionyl butenoic acid; GS, glutathione.

and Coleman 2008). The reactive epoxide produced in the first step of isoprene metabolism is conjugated with glutathione by a glutathione-*S*-transferase to form 1-hydroxy-2-glutathionyl-2-methyl-3-butene (HGMB). HGMB is further metabolized by a dehydrogenase to glutathionyl-2-methyl-3-butenoate (GMBA) (van Hylckama Vlieg et al. 1998, 1999). The subsequent fate of GMBA remains to be elucidated but it can be assumed that the subsequent removal of the glutathione moiety and β-oxidation of the intermediates of isoprene metabolism allows *Rhodococcus* sp. AD45 to grow on isoprene as a sole carbon and energy source (Fig. 3).

van Hycklama Vleig et al. (2000) showed that the isoprene monooxygenase is encoded by the genes *isoABCDEF*. Subsequent sequencing of the 6.8 Mbp genome of *Rhodococcus* sp. AD45 (Crombie et al. 2015) revealed that all of the genes necessary for the metabolism of isoprene are carried on a 300 kbp megaplasmid in this *Rhodococcus* strain (Fig. 4). This clustering of isoprene metabolic genes seems to be a common feature in all isoprene degraders studied to date, although in some other isoprene-degrading *Rhodococcus* strains (e.g., *Rhodococcus opacus* PD630)

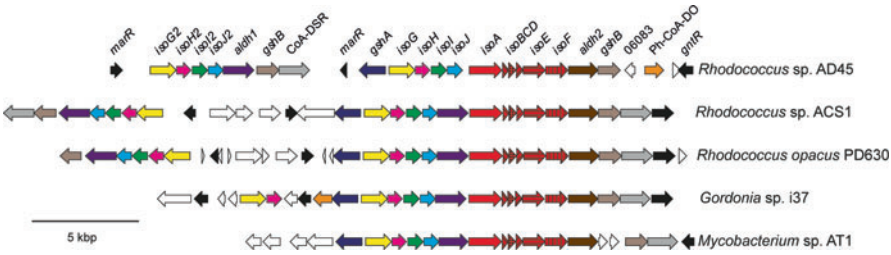



**Fig. 4** The isoprene metabolic gene cluster from representative terrestrial and marine strains. The isoprene monooxygenase genes are shown in red, and homologous genes are in corresponding colors

these isoprene-related genes appear to be chromosomally located rather than on a plasmid (Chen et al. 2014; Crombie et al. 2015). Genes *isoA*, *isoB*, and *isoE* encode the diiron ($\alpha_2\beta_2\gamma_2$) oxygenase component; *isoF* encodes a flavoprotein NADH reductase; *isoC* encodes a Rieske-type ferredoxin; and *isoD* encodes a coupling protein, which together form the multicomponent isoprene monooxygenase. The four genes *isoGHIJ*, preceding the IsoMO genes, encode a putative coenzyme A transferase, a dehydrogenase, and two glutathione transferases (van Hylckama Vlieg et al. 1998, 1999, 2000). An additional copy of this *isoGHIJ* cluster is located on the *Rhodococcus* sp. AD45 megaplasmid (Crombie et al. 2015), and these and other isoprene metabolic genes are duplicated in many isoprene degraders (Fig. 4). Upstream of *isoG*, *gshA* encodes glutamate cysteine ligase, which catalyzes the first step of glutathione biosynthesis. Two copies of *gshB*, encoding glutathione synthetase, are also present. The cluster also contains genes predicted to encode two aldehyde dehydrogenases (*aldh1* and *ald2*), both of which are present in the genomes of all other sequenced isoprene degraders. In *Rhodococcus opacus* PD630, the former of these, *aldh1*, was shown to have glyceraldehyde-3-phosphate and $NADP^+$-dependent activity (MacEachran and Sinskey 2013), possibly suggestive of a requirement for NADPH in isoprene metabolism, although, to our knowledge, alternative substrates of this enzyme were not investigated in detail. Genes encoding a putative coenzyme A disulfide reductase and three putative transcriptional regulators (*marR*1, *marR*2, and *gntR*) are also present. These *iso* genes are essential for isoprene metabolism since removal of the megaplasmid of *Rhodococcus* sp. AD45 by "curing" resulted in loss of the ability to grow on isoprene, as did deletion of *isoA* by mutagenesis (unpublished observation; Crombie et al. 2015).

Actinobacteria such as *Rhodococcus* are metabolically versatile and it appears that isoprene metabolism in *Rhodococcus* sp. AD45 and other isoprene degraders tested to date (*Gordonia* and *Mycobacterium*) is an inducible trait (Crombie et al. 2015; Johnston et al. 2017). Carbon sources such as glucose and succinate repress isoprene metabolism but isoprene monooxygenase polypeptides are readily observed in cell extracts of isoprene-grown *Rhodococcus* sp. AD45. The regulation of expression of isoprene metabolic genes in this strain has been studied in some detail. A replicated time course experiment was conducted in which succinate-grown

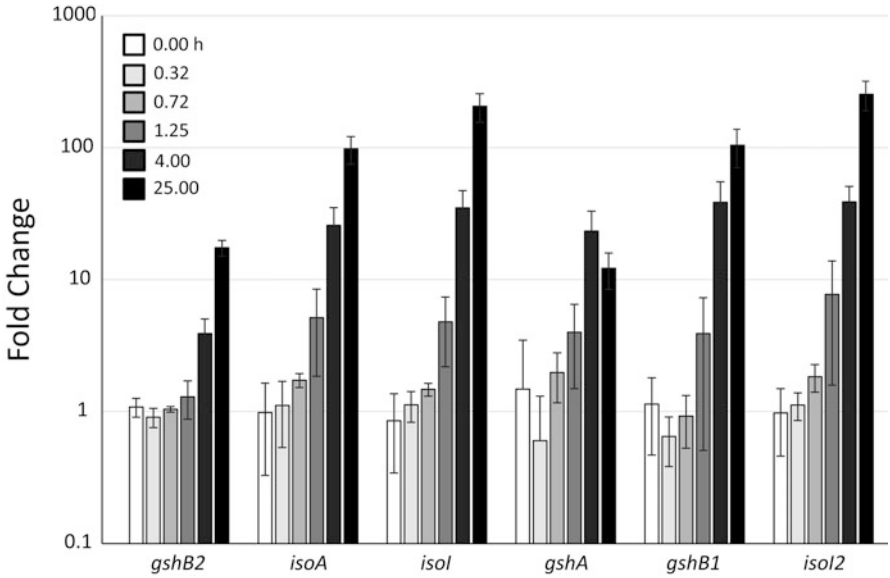



**Fig. 5** Induction of representative isoprene-degradation genes in *Rhodococcus* sp. AD45 after incubation with isoprene, at six timepoints from zero to 25 h, quantified by RNAseq. The bar chart shows the fold change at each time point relative to controls incubated without carbon substrate (Data from Crombie et al. 2015)

cells were subcultured into medium containing isoprene, epoxyisoprene, glucose, succinate, or no substrate (Crombie et al. 2015). Isoprene- and epoxyisoprene-induced expression of *iso* genes were examined by sequencing the transcriptome and comparison with cells grown on glucose and succinate (at 0, 19, 43, 75, 240 min and 25 h). Both isoprene and epoxyisoprene induced high levels of expression of *isoABCDEF, isoGHJI*, and all other putative isoprene metabolism genes present in the 22-gene cluster on the megaplasmid of *Rhodococcus* sp. AD45 (Fig. 5). Under isoprene-induced conditions, these genes represented over 25% of all transcripts observed. These results were subsequently confirmed by targeted RT-PCR. There was no significant over-expression of genes on the chromosome in response to isoprene or epoxyisoprene, suggesting that all isoprene metabolic genes reside on this plasmid (Crombie et al. 2015). The dynamics of gene transcription in response to isoprene or epoxyisoprene indicated that the inducer was not isoprene itself, but rather epoxyisoprene or a subsequent metabolic product of isoprene oxidation, subsequently confirmed by transcriptional analysis of a mutant with an inactivated IsoMO in which isoprene (which could not be metabolized) did not induce *iso* gene transcription. These data provide further targets for mutagenesis and expression studies in order to elucidate the full pathway of isoprene metabolism in *Rhodococcus* sp. AD45. The use of reporter strains to analyze transcriptional regulation and analysis of putative regulators encoded by *marR* and *gntR* may also be a fruitful approach.

## 4 Ecology of Isoprene Degraders

Analysis of the genomes of a number of other isoprene-degrading *Rhodococcus, Gordonia, Mycobacterium*, and *Variovorax* strains has revealed that the clustering of *isoABCDEF* along with *isoGHIJ* and genes involved in glutathione biosynthesis seems to be a common feature in isoprene degradation (Fig. 4) (Crombie et al. 2015; El Khawand et al. 2016; Johnston et al. 2017). The genes encoding isoprene monooxygenase are often misannotated as toluene or alkene monooxygenases, other members of the SDIMO family, but the close proximity of *isoGHIJ* to *isoABCDEF* is a good indication of isoprene metabolism in newly isolated strains. The availability of the *isoABCDEF* sequence has also provided tools for cultivation-independent studies aimed at assessing the distribution, diversity, and activity of isoprene degraders in the environment. A fruitful approach in molecular ecology studies of methane-oxidizing bacteria has been to use methane monooxygenase marker genes ("functional genes") encoding key components of methane monooxygenases (*pmoA* or *mmoX*) to examine different environments (Dumont and Murrell 2005b; McDonald et al. 2008). We have used a similar approach with isoprene monooxygenase. The homologue of *mmoX* (which encodes the large subunit of the oxygenase component of the soluble methane monooxygenase) in IsoMO is *isoA*, which encodes the putative active site component. This polypeptide appears to be highly conserved in all isoprene degraders studied to date. Phylogenetic analysis of IsoA of isoMO from known isoprene degraders and comparison with the corresponding components of SDIMO enzymes such as toluene monooxygenase and alkene monooxygenases (Fig. 6) has confirmed that *isoA* is a suitable marker gene for cultivation-independent studies and that derived IsoA homologs can be readily distinguished from those of non-isoprene-degraders containing SDIMOs other than IsoMO. Alignment of IsoA from bona fide isoprene degraders has allowed the design of *isoA* PCR primer sets targeting these bacteria. These primers did not amplify SDIMO genes from non-isoprene degraders but gave *isoA* gene products with DNA from a range of isoprene-degrading isolates and enrichment cultures originating from various soils, sediments, and leaf samples (El Khawand et al. 2016). Alignments of the IsoA sequences retrieved from environmental samples, with those of characterized isoprene degraders, showed that the IsoA sequences were relatively highly conserved (>86% identity) and could be broadly separated into two groups, those from marine environments and those from terrestrial environments, predominantly actinobacterial isoprene degraders (El Khawand et al. 2016). However, the subsequent isolation and analysis of more isoprene degraders suggests that there is variation in their *isoA* sequences which will, as has been the case with methane monooxygenase functional gene PCR primers, necessitate redesign of *isoA* PCR primer sets. This functional gene probing approach has extended knowledge of the diversity of isoprene-degraders in both terrestrial and marine environments (El Khawand et al. 2016; Johnston et al. 2017).

In order to determine which isoprene degraders are active in the environment, other cultivation-independent techniques such as DNA Stable Isotope Probing (DNA-SIP) (Dumont and Murrell 2005a) need to be used. DNA-SIP has been used to identify active isoprene degraders in both the terrestrial and marine

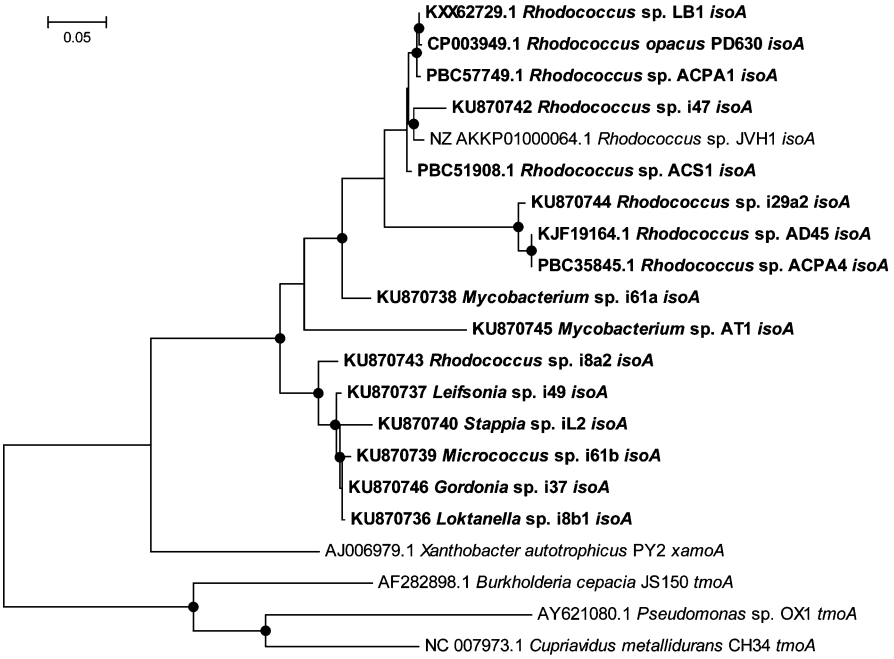



**Fig. 6** The relationship between the *isoA* gene sequences of isoprene degraders (shown in bold) and other homologous sequences. The tree was drawn in Mega6 (Tamura et al. 2013) using the Maximum Likelihood method. Sites with less than 95% coverage were deleted, resulting in 1005 nucleotide positions in the analysis. Bootstrap values (1000 replications) greater than 75% are shown by black circles at the nodes. The scalebar shows base substitutions per site. Based on the presence of adjacent *isoGHIJ* genes, we can predict that *Rhodococcus* sp. JVH1 can grow on isoprene, although this has not, to our knowledge, been tested

environment. For example, surface soil samples from around Willow trees (*Salix fragilis*) were supplied with $^{13}$C–labelled isoprene in microcosms and isoprene uptake was monitored by gas chromatography. After sufficient "heavy" isoprene was incorporated into biomass, $^{13}$C–labelled DNA was isolated by buoyant density gradient centrifugation and used as template in PCR reactions with primers targeting 16S rRNA genes. The analysis revealed a considerable enrichment of several species of *Rhodococcus* in these microcosms, indicating that under the incubation conditions used, *Rhodococcus* represented the majority of active isoprene degraders in these soils (El Khawand et al. 2016). Interestingly, 16S rRNA gene sequences from the Betaproteobacteria *Comamonas* and *Variovorax* were also enriched in $^{13}$C–labelled DNA. In addition to soils, DNA-SIP has also been used to identify active isoprene degraders on leaves of White Poplar (*Populus alba*). Microbes washed from the leaves of this tree were incubated in microcosms with $^{13}$C–labelled isoprene. After sufficient $^{13}$C–labelling, "heavy" DNA was isolated and used in shotgun metagenomics experiments. This enabled the reconstruction of a considerable proportion of the draft genome of an isoprene-degrading *Variovorax* species, including the *iso*

metabolic gene clusters *isoABCDEF* and *isoGHIJ* (Crombie et al., manuscript in preparation). This information has subsequently been used in the targeted isolation of isoprene-degrading *Variovorax* species (Mejia-Florez et al., unpublished).

The functionality of these *Variovorax iso* genes retrieved directly from the environment using DNA-SIP was confirmed by expression studies. The putative IsoMO genes *isoABCDE* from the reconstructed *Variovorax* genome were expressed in a heterologous expression system: a non-isoprene-degrading variant of *Rhodococcus* sp. AD45 without the 300 kbp megaplasmid carrying the *iso* genes (Crombie et al., manuscript in preparation). When expressed, the *Variovorax isoABCDEF* genes conferred the ability of the *Rhodococcus* strain to oxidize isoprene, thus proving that it is a bona fide isoprene monooxygenase. Metatranscriptome data obtained from the same isoprene incubation experiments used for DNA-SIP also confirmed that these *Variovorax* genes were expressed under the enrichment conditions. Targeted isolations have now yielded isoprene-degrading *Variovorax* strains from leaves, thus providing further *isoA* genes to refine PCR primer sets and a new model Gram-negative isoprene degrader to complement the Gram-positive strains available (Crombie et al. manuscript in preparation).

The diversity of isoprene degraders in the marine environment has also been investigated using cultivation-independent methods. Surface estuarine sediments from the Colne Estuary (UK), incubated with $^{13}$C–labelled isoprene, yielded $^{13}$C–DNA which when analyzed revealed the development of isoprene-degrading communities dominated by Actinobacteria including *Gordonia, Mycobacterium, Microbacterium*, and *Rhodococcus* (Johnston et al. 2017). Enrichments of similar environmental samples from the Colne Estuary yielded isolates *Gordonia* sp. i37 and *Mycobacterium* sp. AT1, which grew on isoprene as a sole source of carbon and energy. Analysis of their genomes revealed the same gene arrangements of *isoABCDEF* and *isoGHIJ* seen in *Rhodococcus* sp. AD45 (Fig. 4). As with *R.* sp. AD45, isoprene oxidation was inducible in the presence of isoprene (Johnston et al. 2017). A second SDIMO was identified in the genomes of both of these isoprene-degraders which had significant sequence identity to the propane monooxygenase from *Gordonia* TY-5 (Kotani et al. 2003). This second SDIMO system enabled these bacteria to grow on propane, which, interestingly, appears to be a common feature of many isoprene-degrading bacteria (although not *Rhodococcus* sp. AD45) (Acuña Alvarez et al. 2009; Johnston et al. 2017). These marine isolates, together with other isoprene degraders isolated and characterized from the Colne Estuary and other marine environments (Acuña Alvarez et al. 2009; Johnston et al. 2017), are yielding valuable genome sequence information to refine *isoA* PCR primers and confirming the presence of both *isoABCDEF* and *isoGHIJ* gene clusters in bona fide isoprene degraders.

## 5 Conclusions and Research Needs

All isoprene-degrading bacteria studied so far possess an isoprene monooxygenase of the SDIMO family, which is required for the initial oxidation of isoprene. This enzyme is induced by isoprene or a further oxidation product of isoprene metabolism. It will be

interesting to explore the structure and substrate range of this new class of enzymes in relation to others in the SDIMO family; for example, soluble methane monooxygenase found in some methanotrophs and alkene monooxygenases of propene degraders. These enzymes can also (co-)oxidize isoprene but the bacteria lack the additional metabolic machinery to allow growth on isoprene (Johnston et al. 2017). A second unifying feature of extant isoprene-degraders is the use of glutathione to detoxify epoxyisoprene, the first oxidation product of isoprene. This is in contrast to other alkene degraders, which often use coenzyme M as cofactor (Krishnakumar et al. 2008), and glutathione biosynthesis genes have so far always been found in close proximity to *iso* genes in isoprene degraders. The subsequent steps in isoprene metabolism require further study but the identification of *iso* genes in *Rhodococcus* sp. AD45 and the availability of a mutagenesis and expression system in this "workhorse" organism will now allow us to characterize the mechanisms by which bacteria regulate isoprene metabolism and subsequently incorporate carbon from isoprene into biomass. In some cases (*Rhodococcus* sp. AD45), all necessary *iso* genes reside on a megaplasmid, hinting at the possibility that they are transferred between bacteria by horizontal gene transfer. This notion is supported by the lack of congruence between the phylogeny derived from *isoA* and 16S rRNA gene sequences (Figs. 1 and 6).

The studies summarized here clearly indicate that isoprene-degrading bacteria are widespread in the environment and that soils possess the capability to deplete isoprene at environmental concentrations. Cultivation-independent techniques, such as DNA-, RNA- or protein-SIP, or single cell technologies such as Raman microspectroscopy, will help reveal new isoprene degraders (Murrell and Whiteley 2011; Wang et al. 2016). Challenges for the future include conducting these sequence-independent experiments at conditions that mimic those in the environment, which will identify isoprene-degrading microbes which may exploit specific micro niches. Given that intercellular isoprene concentrations inside leaves may be up to three orders of magnitude higher than atmospheric (Fini et al. 2017), reaching the low ppmv range, the possibility of isoprene-degrading endophytes should be explored. It will also be necessary to increase the diversity of isoprene-degrading strains in cultivation. Characterization of isolates may reveal other pathways for metabolism of isoprene by bacteria and indeed may identify other isoprene-degrading microbes such as Archaea and fungi. To our knowledge there is so far no evidence for anaerobic degradation of isoprene but this possibility should also be borne in mind. To determine the impact that biological isoprene uptake has on global fluxes it will be necessary to quantify microbial activity in the environment. Approaches being pursued in our laboratory range from construction of biosensor strains to express reporter genes under conditions where isoprene-related genes are expressed, through to purification and characterization of isoprene metabolic enzymes, including IsoMO, to determine the affinity and kinetics of isoprene degradation in different isolates.

**Acknowledgments** The authors acknowledge funding from the Earth and Life Systems Alliance (ELSA) at the Norwich Research Park, a Natural Environment Research Council (NERC) grant to JCM (NE/J009725/1) and TJM (NE/J009555/1), an ERC Advanced Grant to JCM (694578 – IsoMet), and Colciencias (Government of Colombia) and Newton Fund support for a studentship to NLMF.

## References

Acuña Alvarez L, Exton DA, Timmis KN, Suggett DJ, McGenity TJ (2009) Characterization of marine isoprene-degrading communities. Environ Microbiol 11:3280–3291

Bäck J, Aaltonen H, Hellén H, Kajos MK, Patokoski J, Taipale R et al (2010) Variable emissions of microbial volatile organic compounds (MVOCs) from root-associated fungi isolated from Scots pine. Atmos Environ 44:3651–3659

Carlton AG, Wiedinmyer C, Kroll JH (2009) A review of Secondary Organic Aerosol (SOA) formation from isoprene. Atmos Chem Phys 9:4987–5005

Chen Y, Ding Y, Yang L, Yu J, Liu G, Wang X et al (2014) Integrated omics study delineates the dynamics of lipid droplets in *Rhodococcus opacus* PD630. Nucleic Acids Res 42:1052–1064

Cleveland CC, Yavitt JB (1997) Consumption of atmospheric isoprene in soil. Geophys Res Lett 24:2379–2382

Cleveland CC, Yavitt JB (1998) Microbial consumption of atmospheric isoprene in a temperate forest soil. Appl Environ Microbiol 64:172–177

Crombie AT, Khawand ME, Rhodius VA, Fengler KA, Miller MC, Whited GM et al (2015) Regulation of plasmid-encoded isoprene metabolism in *Rhodococcus*, a representative of an important link in the global isoprene cycle. Environ Microbiol 17:3314–3329

Dumont MG, Murrell JC (2005a) Stable isotope probing – linking microbial identity to function. Nat Rev Microbiol 3:499–504

Dumont MG, Murrell JC (2005b) Community-level analysis: key genes of aerobic methane oxidation. Methods Enzymol 397:413–427

El Khawand M, Crombie AT, Johnston A, Vavlline DV, McAuliffe JC, Latone JA 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. Environ Microbiol 18:2743–2753

Ewers J, Freier-Schroder D, Knackmuss HJ (1990) Selection of trichloroethene (TCE) degrading bacteria that resist inactivation by TCE. Arch Microbiol 154:410–413

Exton DA, Suggett DJ, McGenity TJ, Steinke M (2013) Chlorophyll-normalized isoprene production in laboratory cultures of marine microalgae and implications for global models. Limnol Oceanogr 58:1301–1311

Exton DA, McGenity TJ, Steinke M, Smith DJ, Suggett DJ (2015) Uncovering the volatile nature of tropical coastal marine ecosystems in a changing world. Glob Chang Biol 21:1383–1394

Fall R, Copley SD (2000) Bacterial sources and sinks of isoprene, a reactive atmospheric hydrocarbon. Environ Microbiol 2:123–130

Fares S, Brilli F, Noguès I, Velikova V, Tsonev T, Dagli S, Loreto F (2008) Isoprene emission and primary metabolism in *Phragmites australis* grown under different phosphorus levels. Plant Biol 10:38–43

Fini A, Brunetti C, Loreto F, Centritto M, Ferrini F, Tattini M (2017) Isoprene responses and functions in plants challenged by environmental pressures associated to climate change. Front Plant Sci 8:1281

Gelmont D, Stein RA, Mead JF (1981) Isoprene – the main hydrocarbon in human breath. Biochem Biophys Res Commun 99:1456–1460

Gray CM, Helmig D, Fierer N (2015) Bacteria and fungi associated with isoprene consumption in soil. Elem Sci Anth 3:000053

Guenther A, Karl T, Harley P, Wiedinmyer C, Palmer PI, Geron C (2006) Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). Atmos Chem Phys 6:3181–3210

Guenther AB, Jiang X, Heald CL, Sakulyanontvittaya T, Duhl T, Emmons LK, Wang X (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. Geosci Model Dev 5:1471–1492

Hewitt CN, MacKenzie AR, Di Carlo P, Di Marco CF, Dorsey JR, Evans M et al (2009) Nitrogen management is essential to prevent tropical oil palm plantations from causing ground-level ozone pollution. Proc Natl Acad Sci 106:18447–18451

Holmes AJ, Coleman NV (2008) Evolutionary ecology and multidisciplinary approaches to prospecting for monooxygenases as biocatalysts. Antonie Van Leeuwenhoek 94:75–84

Johnston A, Crombie AT, El Khawand M, Sims L, Whited G, McGenity TJ, Murrell JC (2017) Identification and characterisation of isoprene-degrading bacteria in an estuarine environment. Environ Microbiol 19:3526–3537

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

Krishnakumar AM, Sliwa D, Endrizzi JA, Boyd ES, Ensign SA, Peters JW (2008) Getting a handle on the role of coenzyme M in alkene metabolism. Microbiol Mol Biol Rev 72:445–456

Kuzma J, Nemecek-Marshall M, Pollock W, Fall R (1995) Bacteria produce the volatile hydrocarbon isoprene. Curr Microbiol 30:97–103

Leahy JG, Batchelor PJ, Morcomb SM (2003) Evolution of the soluble diiron monooxygenases. FEMS Microbiol Rev 27:449–479

Loivamäki M, Mumm R, Dicke M, Schnitzler J-P (2008) Isoprene interferes with the attraction of bodyguards by herbaceous plants. Proc Natl Acad Sci 105:17430–17435

Loreto F, Ciccioli P, Brancaleoni E, Valentini R, De Lillis M, Csiky O, Seufert G (1998) A hypothesis on the evolution of isoprenoid emission by oaks based on the correlation between emission type and *Quercus* taxonomy. Oecologia 115:302–305

Luo G, Yu F (2010) A numerical evaluation of global oceanic emissions of α-pinene and isoprene. Atmos Chem Phys 10:2007–2015

Lv X, Xie W, Lu W, Guo F, Gu J, Yu H, Ye L (2014) Enhanced isoprene biosynthesis in *Saccharomyces cerevisiae* by engineering of the native acetyl-CoA and mevalonic acid pathways with a push-pull-restrain strategy. J Biotechnol 186:128–136

MacEachran DP, Sinskey AJ (2013) The *Rhodococcus opacus* TadD protein mediates triacylglycerol metabolism by regulating intracellular NAD(P)H pools. Microb Cell Factories 12:104

Magel E, Mayrhofer S, Müller A, Zimmer I, Hampp R, Schnitzler JP (2006) Photosynthesis and substrate supply for isoprene biosynthesis in poplar leaves. Atmos Environ 40:138–151

Marienhagen J, Bott M (2013) Metabolic engineering of microorganisms for the synthesis of plant natural products. J Biotechnol 163:166–178

McDonald IR, Bodrossy L, Chen Y, Murrell JC (2008) Molecular ecology techniques for the study of aerobic methanotrophs. Appl Environ Microbiol 74:1305–1315

McGenity TJ, Crombie AT, Murrell JC (2017) Microbial cycling of isoprene, the most abundantly produced biological volatile organic compound on earth. ISME J (in press)

Morais ARC, Dworakowska S, Reis A, Gouveia L, Matos CT, Bogdał D, Bogel-Łukasik R (2015) Chemical and biological-based isoprene production: green metrics. Catal Today 239:38–43

Murphy G (2017) Isoprene degradation in the terrestrial environment. PhD thesis, Department of Biological Sciences, University of Essex, Colchester

Murrell JC, Whiteley AS (2011) Stable isotope probing and related technologies. American Society of Microbiology, Washington, DC

Pacifico F, Harrison SP, Jones CD, Sitch S (2009) Isoprene emissions and climate. Atmos Environ 43:6121–6135

Palmer PI, Shaw SL (2005) Quantifying global marine isoprene fluxes using MODIS chlorophyll observations. Geophys Res Lett 32:L09805

Pegoraro E, Abrell L, Van Haren J, Barron-Gafford G, Grieve KA, Malhi Y et al (2005) The effect of elevated atmospheric $CO_2$ and drought on sources and sinks of isoprene in a temperate and tropical rainforest mesocosm. Glob Chang Biol 11:1234–1246

Pegoraro E, Rey ANA, Abrell L, Van Haren J, Lin G (2006) Drought effect on isoprene production and consumption in Biosphere 2 tropical rainforest. Glob Chang Biol 12:456–469

Rohmer M. 1999. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. Nat Prod Rep 16:565–74

Sanderson MG, Jones CD, Collins WJ, Johnson CE, Derwent RG (2003) Effect of climate change on isoprene emissions and surface ozone levels. Geophys Res Lett 30:1936. https://doi.org/10.1029/2003GL017642

Sharkey TD, Singsaas EL, Vanderveer PJ, Geron C (1996) Field measurements of isoprene emission from trees in response to temperature and light. Tree Physiol 16:649–654

Sharkey TD, Wiberley AE, Donohue AR (2008) Isoprene emission from plants: why and how. Ann Bot 101:5–18

Shaw SL, Gantt B, Meskhidze N (2010) Production and emissions of marine isoprene and monoterpenes: a review. Adv Meteorol 2010:art. ID 408696

Shennan JL (2006) Utilisation of C2–C4 gaseous hydrocarbons and isoprene by microorganisms. J Chem Technol Biotechnol 81:237–256

Srikanta Dani KG, Silva Benavides AM, Michelozzi M, Peluso G, Torzillo G, Loreto F (2017) Relationship between isoprene emission and photosynthesis in diatoms, and its implications for global marine isoprene estimates. Mar Chem 189:17–24

Srivastva N, Shukla AK, Singh RS, Upadhyay SN, Dubey SK (2015) Characterization of bacterial isolates from rubber dump site and their use in biodegradation of isoprene in batch and continuous bioreactors. Bioresour Technol 188:84–91

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725–2729

van Ginkel CG, Welten HGJ, de Bont JAM (1987a) Oxidation of gaseous and volatile hydrocarbons by selected alkene-utilizing bacteria. Appl Environ Microbiol 53:2903–2907

van Ginkel CG, de Jong E, Tilanus JWR, de Bont JAM (1987b) Microbial oxidation of isoprene, a biogenic foliage volatile and of 1,3-butadiene, an anthropogenic gas. FEMS Microbiol Lett 45:275–279

van Hylckama Vlieg JET, Kingma J, van den Wijngaard AJ, Janssen DB (1998) A glutathione *S*-transferase with activity towards cis-1,2-dichloroepoxyethane is involved in isoprene utilization by *Rhodococcus* sp. strain AD45. Appl Environ Microbiol 64:2800–2805

van Hylckama Vlieg JET, Kingma J, Kruizinga W, Janssen DB (1999) Purification of a glutathione *S*-transferase and a glutathione conjugate-specific dehydrogenase involved in isoprene metabolism in *Rhodococcus* sp. strain AD45. J Bacteriol 181:2094–2101

van Hylckama Vlieg JET, Leemhuis H, Spelberg JHL, Janssen DB (2000) Characterization of the gene cluster involved in isoprene metabolism in *Rhodococcus* sp. strain AD45. J Bacteriol 182:1956–1963

Wang Y, Huang WE, Cui L, Wagner M (2016) Single cell stable isotope probing in microbiology using Raman microspectroscopy. Curr Opin Biotechnol 41:34–42

Zeinali N, Altarawneh M, Li D, Al-Nu'airat J, Dlugogorski BZ (2016) New mechanistic insights: why do plants produce isoprene? ACS Omega 1:220–225

# The poplar phyllosphere harbours disparate isoprene-degrading bacteria

Andrew T. Crombie[1], Nasmille L. Larke-Mejia[2], Helen Emery[2], Robin Dawson[2], Jennifer Pratscher[3], Gordon P. Murphy[4], Terry J. McGenity[4], J. Colin Murrell[2]

[1]School of Biological Sciences, University of East Anglia, Norwich, United Kingdom NR4 7TJ

[2]School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom NR4 7TJ

[3]Lyell Centre, Heriot-Watt University, Edinburgh, United Kingdom EH14 4AS

[4] School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK

Author for correspondence: Andrew Crombie, School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK. Telephone +44 1603 592239. Email a.crombie@uea.ac.uk

## Abstract

The climate-active gas isoprene (2-methyl-1,3-butadiene) is released to the atmosphere in huge quantities, almost equalling that of methane, yet we know little about the biological cycling of isoprene in the environment. Although bacteria capable of growth on isoprene as sole source of carbon and energy have previously been isolated from soils and sediments, no microbiological studies have targeted the major source of isoprene and examined the phyllosphere of isoprene-emitting trees for the presence of degraders of this abundant carbon source. Here, we identified isoprene-degrading bacteria in poplar tree derived microcosms by DNA stable isotope probing. The genomes of novel isoprene-degrading taxa were reconstructed, putative isoprene metabolic genes were identified and isoprene-related gene transcription was analysed by shotgun metagenomics and metatranscriptomics. Gram-positive bacteria of the genus *Rhodococcus* proved to be the dominant isoprene degraders, as previously found in soil. However, a novel diversity of isoprene utilizers was also revealed, notably *Variovorax*, a genus not previously associated with this trait. This finding was confirmed by expression of the isoprene monooxygenase from *Variovorax* in a heterologous host. A novel *Variovorax* strain that could grow on isoprene as sole carbon and energy source was isolated. Analysis of its genome confirmed that it contained isoprene metabolic genes with an identical layout and high similarity to those identified by DNA-SIP and metagenomics. This study provides evidence of a wide diversity of isoprene-degrading bacteria in the isoprene-emitting tree phyllosphere and greatly enhances our understanding of the biodegradation of this important metabolite and climate-active gas.

\body

## Introduction

Isoprene (2-methyl-1,3-butadiene) is emitted to the atmosphere at a rate of approximately 500 Tg $y^{-1}$, on a par with methane and approximately one third of total volatile organic compound (VOC) emissions (1, 2). The vast majority originates from terrestrial plants (400 – 600 Tg $y^{-1}$), with a small but uncertain flux from marine algae (0.1 – 12 Tg $y^{-1}$) (2-4) and from bacteria, fungi and animals (5, 6). A reactive diene, isoprene is rapidly photochemically oxidised (1), and has a significant and complex effect on global climate (7). Hydroxyl (OH) and nitrate ($NO_3$) radicals and ozone ($O_3$) in the atmosphere react with isoprene depending on prevailing conditions (1). In pristine environments, isoprene reacts directly with ozone and hydroxyl radicals, resulting in ozone depletion. However, the high $NO_x$ levels typical of urban environments result in formation of tropospheric ozone, with important negative effects on human health and on yields of ozone-sensitive crops (8). Globally, these reactions result in a net radiative forcing of 0.09 W $m^{-2}$, with an additional indirect effect since depletion of OH radicals increases the atmospheric lifetime of methane (9). The isoprene oxidation products form secondary organic aerosols (SOA) and cloud condensation nuclei, with implications for planetary albedo, air quality and climate (10, 11).

Plants produce isoprene in the chloroplast and release it to the atmosphere from the abaxial surface of leaves via the stomata (12). Although isoprene production is not a universal trait among plants (13), it protects against heat and oxidative stress (14, 15) and has roles in plant/insect signalling and plant energy dynamics (5, 16). High isoprene-emitting trees worldwide include oil palm (*Elaeis guineensis*), eucalyptus (*Eucalyptus* spp.) and, in temperate regions, English oak (*Quercus robur*), poplar (*Populus* spp.) and willow (*Salix* spp.), with reported emissions of 77 (English oak) and 175 (oil palm) μg $g^{-1}$ (dry leaves) $h^{-1}$ (17). Due to its short atmospheric lifetime, isoprene concentrations vary from a few ppbv (18), to tens of ppbv in high-isoprene-emitting forests (19). Soils acted as a biological sink, both in closed-chambers (20, 21) and in continuous systems using isoprene concentrations of 2 – 200 ppbv (22), suggesting that soil microbes act as a significant sink for isoprene, consuming up to 4% of global emissions.

Several isoprene-consuming bacterial strains have been isolated from soils and marine sediments (6, 23-26) and recently several strains have been characterised

biochemically and genetically (27-31). Prior to this work, all strains characterised by sequence data were members of the Actinobacteria and all terrestrial examples were rhodococci. They all contain six genes (*isoABCDEF*) encoding a multi-component isoprene monooxygenase (IsoMO) (Fig. 1) with four additional genes, *isoGHIJ*, just upstream of the monooxygenase genes. van Hylckama Vlieg et al. (27, 28) showed that *isoI* encodes a glutathione-*S*-transferase that conjugates glutathione with the epoxide product of IsoMO, and *isoH* encodes a dehydrogenase that acts on the product of IsoI. IsoG and IsoJ are involved in the subsequent metabolic pathway but have yet to be characterised. In *Rhodococcus* AD45, this entire cluster *isoGHIJABCDEF* is co-transcribed as an operon (29). Significantly, all previously characterised isoprene gene clusters include glutathione biosynthesis genes *gshAB*, suggesting a specific role for glutathione in isoprene metabolism. This small thiol is not usually found in Gram-positive bacteria (32), although a role in styrene metabolism has also been suggested recently (33).

The plant microbiome plays an essential role in plant health and development (34) and the phyllosphere, although an unstable environment, constitutes an extensive habitat for microorganisms, mainly bacteria, whose diverse community typically comprises $10^6$ – $10^7$ cells $cm^{-2}$ (35). In the intercellular spaces of leaves, near the point of emission from stomata, isoprene concentrations reach 30 ppmv, approximately three orders of magnitude higher than atmospheric concentrations (12, 36-38). Apart from isoprene, plants produce a range of other VOCs. Although leaf-dwelling methylotrophs can reduce plant methanol emissions to the atmosphere (39), the extent to which the plant microbiome moderates release of other VOCs is not known (40, 41). Human intervention is likely to alter global emissions of isoprene, partly because emissions increase with temperature but are inversely related to atmospheric carbon dioxide concentrations, but also due to the development of high isoprene-emitting agroforestry (42, 43). Therefore, our aim was to better understand the diversity of microbial taxa involved in isoprene consumption and their mechanisms of action. Specifically, this study addressed the following questions: does the microbiome of a common isoprene-emitting tree harbour isoprene degraders able to take advantage of this abundant carbon source, and are novel genes expressed in response to isoprene?

## Results and Discussion

**Isoprene degraders isolated from leaves and soil**

Leaf washings from high isoprene-emitting tree species (oak, poplar, willow) (17) consumed isoprene, supplied as sole source of carbon and energy, in microcosms. After 15 days, 65% of the microcosms showed >30% depletion of added isoprene (SI Appendix, Table S1), demonstrating the presence of isoprene-degrading microbes. We obtained three *Rhodococcus* isolates which grew on isoprene as sole source of carbon and energy, two (strains ACPA1 and ACPA4) from poplar leaves and one (strain ACS1) from soil beneath trees. All strains contained an IsoMO and related isoprene metabolic genes similar to those identified previously (29, 44).

**Stable isotope probing identifies active degraders**

To extend the revealed diversity beyond those amenable to cultivation, we used DNA-stable isotope probing (DNA-SIP) in two independent experiments. Since sun-exposed leaves may produce more isoprene than shaded leaves (45), leaves of white poplar (*Populus alba*) taken from locations on the tree exposed to full sun were used. Initially, microbial cells were dislodged from the leaves and incubated in minimal medium with $^{13}$C-labelled or unlabelled isoprene (headspace concentration approximately 500 ppmv). Added isoprene was consumed rapidly and cells were harvested after 7 or 8 days (SI Appendix, Fig. S1). Extraction, fractionation (based on incorporation of $^{13}$C label), and quantification of DNA showed that labelled DNA from the $^{13}$C-isoprene incubations migrated to “heavy” and “light” regions of the ultracentrifuge tubes depending on incorporation of label, whereas DNA from $^{12}$C-isoprene incubations was restricted (>99%) to light fractions (SI Appendix, Fig. S2a). Analysis of 16S rRNA gene amplicons showed that the timepoint zero un-enriched bacterial community, highly consistent across six replicates, comprised mainly Proteobacteria (*Sphingomonas* (36% mean relative abundance (RA)), *Pseudomonas* (17%), members of the Oxalobacteraceae (8%), *Methylobacterium* (2%) and members of the Comamonadaceae (2%)), Bacteroidetes (*Hymenobacter* (17%) and Sphingobacteriaceae (1%)) and *Actinobacteria* (4%), (Fig. 2, SI Appendix, Fig. S3), in agreement with previous studies of plant-associated bacteria (46, 47). Following enrichment, the active isoprene-degrading community was dominated by *Rhodococcus* (78% ± 8% mean RA ± s.d.), although several taxa from the Proteobacteria were also labelled, principally members of the Xanthomonadaceae (8% ± 7%) and Comamonadaceae (3% ± 1.6%) (Fig. 2, SI Appendix, Fig. S3). Although, so far, all well-described terrestrial isoprene degraders in cultivation are

members of the genus *Rhodococcus*, we have previously noted the presence of sequences related to Proteobacteria in DNA-SIP incubations with labelled isoprene (30), despite the fact that no proteobacterial genomes available in the public databases contain recognizable isoprene-degradation-related gene sequences. We investigated these non-actinobacterial isoprene degraders in more detail, again using DNA-SIP.

The second DNA-SIP experiment, with isoprene at 150 ppmv, included cells incubated without isoprene (no substrate) as an additional control. RNA was also extracted from the incubations, although not subjected to ultracentrifugation or separation into labelled and unlabelled fractions. Analysis showed that, as before, labelled and unlabelled DNA were efficiently separated (SI Appendix, Fig. S2b). In contrast to the first experiment, nucleic acids were sequenced by shotgun meta-omics. Community composition was assessed by clade-specific marker genes using MetaPhlAn2 (48) (Fig. 2, SI Appendix, Fig. S3). Analysis of the same DNA, from the heavy fraction of the $^{13}$C-isoprene incubations of the first experiment, by both 16S rRNA amplicon and metagenomic sequencing, enabled direct comparison of the microbial community revealed by these two methods, both of which confirmed the dominance of *Rhodococcus* sequences (78% and 94% RA by amplicon and shotgun sequencing respectively), together with a contribution from Comamonadaceae (3% and 1.5% respectively). However, the Xanthomonadaceae, identified as 8% among 16S rRNA amplicons, were not identified in significant numbers by the shotgun approach (< 0.1% RA), possibly indicative of bias in the PCR-based method.

In the second SIP experiment the un-enriched timepoint zero community was dominated by the Bacteroidetes *Hymenobacter* and *Pedobacter* (50% RA). *Sphingomonas*, the most abundant genus of the first experiment, was a minor component (3%) in the second. Following isoprene enrichment, the labelled community was again dominated by *Actinobacteria* (mainly *Rhodococcus*, average 74% RA) together with, as before, Proteobacteria, notably Comamonadaceae, principally *Variovorax*, which averaged 16% of the labelled community, albeit with considerable inter-sample variability (SI Appendix, Fig. S3). In contrast, these genera comprised a small fraction (0.5 – 1%) of the DNA from the pooled light fractions. The community profile of control incubations without added substrate was similar to that of the light fractions, further confirming that the heavy fractions contained the labelled DNA from isoprene-consumers (Fig. 2).

The transcriptionally active taxa were characterised by profiling the mRNA transcriptome reads using Metaphlan2 (SI Appendix, Fig. S4). Interestingly, this approach, albeit based on an extremely restricted subset of taxonomically-informative marker genes, identified the Bacteroidetes genus *Pedobacter* and the Gammaproteobacterium *Pseudomonas* as transcriptionally active, not only in isoprene enrichments, but also in un-enriched timepoint-zero samples and in incubations without added substrate, suggesting that these taxa were able to scavenge nutrients from endogenous organic matter and/or dead microbial cells. Transcripts of *Rhodococcus* and *Variovorax*, as well as the Bacteroidetes *Chryseobacterium* and *Riemerella* were specifically enriched in isoprene-incubations, indicating that these taxa were directly or indirectly stimulated by isoprene (SI Appendix, Fig. S4).

**Assembly identifies novel sequences**

To identify novel isoprene-degrading sequences, reads from pooled un-enriched timepoint zero samples, from each $^{13}$C-isoprene-enriched heavy fraction, from pooled light-fraction DNA and from no-substrate incubations were co-assembled, resulting in a total of 1.84 Gbp of sequence (SI Appendix, Table S3). Using, as query, the amino acid (AA) sequence of *isoA* (encoding the alpha subunit of IsoMO), we identified 11 sequences with 40 – 100% inferred AA identity to IsoA from *Rhodococcus* AD45 (Fig. 3). These scaffolds were examined in detail and five contained sufficient sequence data to also identify one or more of the genes, specific for isoprene metabolism in *Rhodococcus* AD45 (29), immediately flanking the monooxygenase (Fig. 4), (36 – 100% AA identity with the *Rhodococcus* AD45 gene products), indicating a likely role in isoprene degradation. Similarly, reads originating from rRNA-depleted RNA were assembled resulting in 220,637 – 454,247 transcripts containing 83.0 – 223.3 Mbp of sequence for each sample (SI Appendix, Table S4). Eighteen transcripts containing distinct *isoA*-related sequences were identified (Fig. 3), eight of which contained other isoprene-related genes in addition to the IsoMO genes *isoABCDEF* (SI Appendix, Table S5). Many of these genes and transcripts were closely related to those of the *Rhodococcus* isolates obtained in this study, as well as to other strains identified previously (29, 30) (Fig. 3, Fig. 4). Within-sample quantification of global transcript abundance showed that these *Rhodococcus*-like isoprene gene transcripts were among the most highly expressed, many among the top 0.2% or 1% of community-wide transcripts (Fig. 3 and SI Appendix, Table S5). In

addition to sequences grouping with those of characterised isolates, three highly expressed transcripts, centred around metagenome scaffold MG_3829, formed a distinct cluster (97% identity to *isoA* from *R. opacus* PD630), suggesting that there is yet more diversity to discover among these isoprene-degrading Actinobacteria. Apart from these *Rhodococcus*-like *isoA* genes, more divergent sequences were identified, as described below. Significantly, low-level transcription of *isoA* was also observed at timepoint-zero (un-enriched) and in incubations without isoprene (no substrate), suggesting that there is a degree of constitutive expression of these isoprene-degradation genes.

**Genome reconstruction through binning**

To reconstruct individual genomes from the metagenome, scaffolds were assigned to 266 bins based on abundance and nucleotide composition. Genome bin quality was assessed and refined, and taxonomy was assigned, resulting in 27 genome bins with predicted completeness > 90% and contamination < 5%. From the 266 bins, 18 were identified as of interest based on the presence of an *isoA* sequence or predicted RA of > 2% in the labelled community, and these were examined in greater detail (SI Appendix, Table S6, Table S7). In total, nine *isoA*-containing scaffolds were assigned to genome bins, allowing taxonomy to be inferred for most of these sequences (Fig. 3). Several sequences, similar to those of characterised isolates, were assigned to actinobacterial bins (Fig. 3). The most abundant (average 31% RA across three $^{13}$C-isoprene enrichments), predicted 99% complete and with 1.1% contamination (SI Appendix, Table S6), had 98.7% average nucleotide identity (ANI) to isolate *Rhodococcus* sp. ACPA4. Several other (less complete) *Rhodococcus* bins were also identified. However, since DNA-SIP showed a considerable diversity of labelled organisms, we also looked at more dissimilar sequences. Genome bin 197, assigned to *Pseudonocardia,* contained two *isoA* homologues. The first of these, on scaffold MG_3829, contained a typical isoprene metabolic gene cluster with 75% – 93% AA identity to six isoprene metabolic gene products, including IsoA, from *Rhodococcus* AD45 (Fig. 4). The same genome also contained scaffold MG_720, with a less similar IsoA homologue (48% AA identity with IsoA from *Rhodococcus* AD45). This region of scaffold MG_720 contained genes with high sequence identity and an identical layout to *Rhodococcus jostii* DSM44719 and *Pseudonocardia dioxanivorans* CB1190 (49) annotated as toluene-4-monooxygenase (SI Appendix, Fig. S5). These monooxygenase genes are of low similarity and arranged in a

different order to *isoABCDEF* from isoprene degraders and are not flanked by any other recognizable isoprene-related genes, suggesting that this region was not responsible for isoprene metabolism by known or predicted pathways. Interestingly, although they could not be identified as directly involved in isoprene degradation, these divergent genes on scaffold MG_720 of genome bin 197 were transcribed at moderately high levels (Fig. 3 and SI Appendix, Table S5).

Genome bin 095 contained a sequence with homology to a putative toluene monooxygenase from *Myxococcus* bacteria sp.68-20 (OJY25058.1), with a similar gene layout. Again, although this genome contained no identifiable isoprene-related genes apart from the monooxygenase, the *isoA* homologue was represented in the transcriptome (Fig. 3 and SI Appendix, Table S5).

These two examples suggest either that isoprene is capable of inducing genes not central to its metabolism, that these monooxygenase genes were constitutively expressed at considerable levels, or that they form part of a novel and so-far undescribed isoprene metabolic pathway, topics of on-going investigation.

Genome bin 232, assigned to the betaproteobacterial genus *Variovorax*, contained an *isoA* homologue (scaffold MG_478), which was represented by a moderately expressed transcript and aligned most closely with *xamoA* of the propylene-degrader *Xanthobacter autotrophicus* Py2 (Fig. 3 and SI Appendix, Table S5). In contrast to characterized isoprene-degraders, however, *X. autotrophicus* Py2 does not use glutathione in its alkene metabolic pathway (50), does not contain any homologues to the isoprene metabolic genes surrounding *isoA-F* at the *xamoA* locus (51), and does not grow on isoprene (31). Scaffold MG_478 comprised 76,719 bp of contiguous DNA, allowing examination of the genomic context of *isoA* (Fig. 4). In addition to the six genes (*isoABCDEF*) encoding IsoMO, *isoGHIJ* and *aldh1* were present in an identical layout to those of many isoprene-degrading isolates (Fig. 4), although sequence identity of the gene products with those of *Rhodococcus* AD45 ranged from 42 – 71%, much lower than those of our characterised isolates. Glutathione biosynthesis genes *gshA* and *gshB* were not present in this isoprene cluster but were found on scaffold MG_3916 of this genome bin (85% and 76% AA identity to ACS17089.1 and ACS17095.1 from *Variovorax paradoxus* S110 respectively (52)), perhaps reflecting the general use of glutathione in Gram-negative bacteria, as opposed to its isoprene-specific use in Gram-positive strains.

The co-occurrence of both *isoABCDF* and *isoGHIJ* has previously been successful in identifying *bone fide* isoprene degraders. However, in order to verify that this sequence contained the genetic potential for isoprene oxidation, we cloned the putative IsoMO genes *isoABCDEF* into a plasmid vector and induced expression in *Rhodococcus* AD45-ID, a strain of *R.* AD45 lacking the megaplasmid containing the *iso* genes, which is incapable of isoprene oxidation. When expressed, the IsoMO from bin 232 indeed oxidised isoprene, in contrast to controls (SI Appendix, Fig. S6).

**Targeted isolation of a *Variovorax* strain**

By screening the 16S rRNA gene sequences of numerous isolates from isoprene-enrichments, we obtained *Variovorax* sp. WS11, isolated from a soil enrichment, which grew on isoprene as sole source of carbon and energy with a growth rate of 0.052 ± 0.004 $h^{-1}$ (SI Appendix, Fig. S7). The *Variovorax* sp. WS11 genome contains an isoprene gene cluster with identical layout to the metagenome-derived sequence of bin 232 (Fig. 4). Overall average AA identity (53) was compared between these genomes and 52 *Variovorax* genome assemblies available in GenBank. The metagenome-derived genome was most similar to *Variovorax* sp. CF079 (accession number GCA_900101545.1) (82.3%) whereas strain WS11 was most similar to *Variovorax* sp. B2 (accession number GCA_002891695.1) (88%) and similarity between the two strains was 79.5%, suggesting that both of these strains are novel species. All *Variovorax* genome assemblies available in GenBank were searched for isoprene-related gene sequences like those present in the *Variovorax* strains described here, but with negative results, suggesting that much diversity still exists outside the reference sequences.

**Conclusions**

Although isoprene degraders are present in all isoprene-exposed environments tested (6), no studies have identified isoprene degraders residing on the leaves of isoprene-emitting trees, at the source of emission. Here, the isoprene degraders retrieved from the poplar phyllosphere were dominated by *Rhodococcus*, but included other Actinobacteria (*Pseudonocardia*) and Proteobacteria (*Variovorax*). Taxonomy alone is insufficient to identify isoprene-degrading bacteria, with extremely closely related strains differing in terms of isoprene-degrading ability, and phylogeny based on 16S rRNA genes is not congruent with *isoA*-based analyses (31). These data suggest lateral transfer of the isoprene metabolic genes and imply

that surveys that rely on 16S rRNA gene analysis are not able to identify isoprene degraders.

We used metagenomics to reconstruct the genome of a *Variovorax* strain, and conventional methods to obtain related isolate *Variovorax* WS11. Both genomes contained the entire isoprene metabolic gene cluster. Despite being more similar in sequence to alkene monooxygenase from *Xanthobacter autotrophicus* Py2 than IsoMO from known isoprene degraders (Fig. 3), the presence of other genes unique to isoprene metabolism and the expression of the monooxygenase in a heterologous host proved that these are genuine isoprene metabolic gene clusters. *Variovorax* are metabolically-versatile bacteria capable of degradation of natural products and xenobiotics, frequently plant-associated and with plant-growth promoting effects (52, 54, 55), and have been identified as part of the core bacterial microbiome of both *Arabidopsis* and poplar (56, 57).

Interestingly, we also detected a significant level of monooxygenase transcripts, similar or identical to $^{13}$C-labelled DNA scaffolds, not so far implicated in isoprene degradation, possibly indicative of novel isoprene metabolic pathways, and this is the topic of continuing investigation. This study, the first to show that the leaves of an isoprene-emitting tree provide a habitat for taxonomically disparate isoprene degraders, is key to continued development of molecular tools to detect isoprene degraders. This is a prerequisite for quantification of isoprene-related genes and transcripts and comparison of the activity of microbes associated with isoprene-emitting and non-emitting environments, (including comparisons between high and low isoprene-emitting tree species), and hence to establish the extent to which the tree microbiome is able to take advantage, and mitigate release, of this abundant carbon source and climate-active gas.

## Materials and Methods

### Enrichment, isolation and stable isotope probing

Isoprene-degraders were enriched from soil or from cells dislodged from leaves by ultrasound and purified by standard methods. For SIP enrichments, cells were washed from 5 g leaves, re-suspended in minimal medium and supplied with $^{13}$C-labelled or unlabelled isoprene to a headspace concentration of 500 ppmv (first experiment) or 150 ppmv (second experiment). Isoprene consumption was followed by gas chromatography (GC) and cells were harvested when they had consumed

approximately 11 µmol or 6 µmol (isoprene C) $ml^{-1}$ respectively. Each treatment was carried out in triplicate. Details are given in SI Appendix.

**Nucleic acid extraction and purification**

DNA and RNA were extracted using standard methods, (SI Appendix). Total RNA was depleted of rRNA using Ribo-Zero (Illumina, San Diego, CA, USA). Labelled and unlabelled DNA was separated by density gradient ultracentrifugation and fractionation as described previously (30) and fraction density was quantified by refractometry. Fractions containing labelled ("heavy") and unlabelled ("light") DNA were identified based on the data presented in SI Appendix, Fig. S2. See SI Appendix for details.

**Sequencing of nucleic acids**

Communities were profiled using 16S rRNA gene amplicons (first DNA-SIP experiment) generated using primers 0341F/0785R (58) and sequenced using Illumina MiSeq. Amplicons were analysed using Qiime (59). Metagenomic and metatranscriptomic libraries were sequenced using Illumina HiSeq (SI Appendix, Table S3, Table S4). Quality-filtered metagenomics reads were taxonomically profiled using Metaphlan2 v2.5.0 (48). Filtered reads were co-assembled using IDBA-UD v1.1.1 (60) (SI Appendix, Table S3) and binned using MaxBin v2.2 (61). Bins were quality checked and refined using CheckM v1.0.5 (62) and RefineM v0.0.23 (63). Filtered transcript reads were *de novo* assembled using Trinity v2.3.2 (64) (SI Appendix, Table S4). Local Blast databases were constructed and searched using tblastn v2.2.28 (65). Reads were mapped to assembled transcripts and quantified using kallisto v0.43.1 (66). Normalized expression levels (TPM) of each transcript were ranked for each sample and converted to percentile. Full details are contained in SI Appendix.

**Expression of IsoMO**

The putative IsoMO genes *isoABCDEF* of metagenome bin 232 were PCR-amplified from the pooled heavy fractions of DNA-SIP enrichments, cloned into expression vector pTipQC1 (67) and expressed in a strain of *Rhodococcus* (*R.* AD45-ID) cured of the megaplasmid which contains the isoprene metabolic genes, see SI Appendix for details. Isoprene uptake of IsoMO-expressing cell suspensions was quantified by GC.

**Accession numbers**

The genome sequences of *Variovorax* sp. WS11 and megaplasmid-cured *Rhodococcus* AD45-ID have been deposited at DBJ/ENA/GenBank under accessions PXZZ00000000 and PYHL00000000. Versions described here are PXZZ01000000 and PYHL0100000 respectively. Sequence reads have been deposited at GenBank SRA (accession number SRP101805), see SI Appendix, Tables S2 – S4.


## Acknowledgements

Plasmid pTipQC1 was a kind gift from Tomohiro Tamura. This work was funded by the Earth and Life Systems Alliance (ELSA) at the University of East Anglia, Natural Environment Research Council (NERC) grants to J.C.M. (NE/J009725/1) and T.J.M. (NE/J009555/1), an ERC advanced grant to J.C.M. (694578—IsoMet) and an NERC Fellowship to J.P. (NE/L010771/1). We acknowledge receipt of a Colombian Government Scholarship to N.L.L-M.


## References


1. Atkinson R & Arey J (2003) Gas-phase tropospheric chemistry of biogenic volatile organic compounds: a review. *Atmos Environ* 37, s2(0):197-219.
2. Guenther AB*, et al.* (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. *Geosci. Model Dev.* 5(6):1471-1492.
3. Shaw SL, Gantt B, & Meskhidze N (2010) Production and emissions of marine isoprene and monoterpenes: a review. *Adv Meteorol* 2010.
4. Dani SKG*, et al.* (2017) Relationship between isoprene emission and photosynthesis in diatoms, and its implications for global marine isoprene estimates. *Mar Chem* 189:17-24.
5. Sanadze GA (2017) Biogenic isoprene emission as expression of dissipativity, a fundamental cell property. *Russian Journal of Plant Physiology* 64(2):133-140.
6. McGenity TJ, Crombie AT, & Murrell JC (2018) Microbial cycling of isoprene, the most abundantly produced biological volatile organic compound on Earth. *The ISME Journal* 12:931–941.
7. Pacifico F, Harrison SP, Jones CD, & Sitch S (2009) Isoprene emissions and climate. *Atmos Environ* 43(39):6121-6135.
8. Ashworth K, Wild O, & Hewitt CN (2013) Impacts of biofuel cultivation on mortality and crop yields. *Nature Clim. Change* 3:492–496.
9. Folberth GA, Hauglustaine DA, Lathière J, & Brocheton F (2006) Interactive chemistry in the Laboratoire de Météorologie Dynamique general circulation model: model description and impact analysis of biogenic hydrocarbons on tropospheric chemistry. *Atmos. Chem. Phys.* 6(8):2273-2319.
10. Fiore AM*, et al.* (2012) Global air quality and climate. *Chem Soc Rev* 41(19):6663-6683.
11. Carlton AG, Wiedinmyer C, & Kroll JH (2009) A review of Secondary Organic Aerosol (SOA) formation from isoprene. *Atmos. Chem. Phys.* 9(14):4987-5005.
12. Fall R & Monson RK (1992) Isoprene emission rate and intercellular isoprene concentration as influenced by stomatal distribution and conductance. *Plant Physiol* 100(2):987-992.

13. Monson RK, Jones RT, Rosenstiel TN, & Schnitzler J-P (2012) Why only some plants emit isoprene. *Plant Cell Environ* 36(3):503-516.
14. Sharkey TD, Wiberley AE, & Donohue AR (2008) Isoprene emission from plants: why and how. *Ann Bot-London* 101(1):5-18.
15. Zeinali N, Altarawneh M, Li D, Al-Nu'airat J, & Dlugogorski BZ (2016) New mechanistic insights: why do plants produce isoprene? *ACS Omega* 1(2):220-225.
16. Loivamäki M, Mumm R, Dicke M, & Schnitzler J-P (2008) Isoprene interferes with the attraction of bodyguards by herbaceous plants. *Proc Natl Acad Sci* 105(45):17430-17435.
17. Kesselmeier J & Staudt M (1999) Biogenic volatile organic compounds (VOC): an overview on emission, physiology and ecology. *Journal of Atmospheric Chemistry* 33(1):23-88.
18. Greenberg JP*, et al.* (1999) Tethered balloon measurements of biogenic VOCs in the atmospheric boundary layer. *Atmos Environ* 33(6):855-867.
19. Wiedinmyer C*, et al.* (2005) Ozarks Isoprene Experiment (OZIE): Measurements and modeling of the "isoprene volcano". *J Geophys Res-Atmos* 110(D18):D18307.
20. Cleveland CC & Yavitt JB (1997) Consumption of atmospheric isoprene in soil. *Geophys Res Lett* 24(19):2379-2382.
21. Pegoraro E*, et al.* (2005) The effect of elevated atmospheric $CO_2$ and drought on sources and sinks of isoprene in a temperate and tropical rainforest mesocosm. *Glob Change Biol* 11(8):1234-1246.
22. Gray CM, Helmig D, & Fierer N (2015) Bacteria and fungi associated with isoprene consumption in soil. *Elem. Sci. Anth.* 3:000053.
23. Srivastva N, Singh A, Bhardwaj Y, & Dubey SK (2017) Biotechnological potential for degradation of isoprene: a review. *Critical Reviews in Biotechnology*:1-13.
24. Cleveland CC & Yavitt JB (1998) Microbial consumption of atmospheric isoprene in a temperate forest soil. *Appl Environ Microbiol* 64(1):172-177.
25. van Ginkel CG, de Jong E, Tilanus JWR, & de Bont JAM (1987) Microbial oxidation of isoprene, a biogenic foliage volatile and of 1,3-butadiene, an anthropogenic gas. *FEMS Microbiol Lett* 45(5):275-279.
26. Ewers J, Freier-Schroder D, & Knackmuss HJ (1990) Selection of trichloroethene (TCE) degrading bacteria that resist inactivation by TCE. *Arch Microbiol* 154(4):410-413.
27. van Hylckama Vlieg JET, Kingma J, Kruizinga W, & Janssen DB (1999) Purification of a glutathione *S*-transferase and a glutathione conjugate-specific dehydrogenase involved in isoprene metabolism in *Rhodococcus* sp. strain AD45. *J Bacteriol* 181(7):2094-2101.
28. van Hylckama Vlieg JET, Kingma J, van den Wijngaard AJ, & Janssen DB (1998) A glutathione *S*-transferase with activity towards cis-1,2-dichloroepoxyethane is involved in isoprene utilization by *Rhodococcus* sp. strain AD45. *Appl Environ Microbiol* 64(8):2800-2805.
29. Crombie AT*, et al.* (2015) Regulation of plasmid-encoded isoprene metabolism in *Rhodococcus*, a representative of an important link in the global isoprene cycle. *Environ Microbiol* 17(10):3314-3329.
30. El Khawand M*, 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. *Environ Microbiol* 18(8):2743-2753.
31. Johnston A*, et al.* (2017) Identification and characterisation of isoprene-degrading bacteria in an estuarine environment. *Environ Microbiol* 19(9):3526-3537.
32. Johnson T, Newton G, Fahey R, & Rawat M (2009) Unusual production of glutathione in *Actinobacteria*. *Arch Microbiol* 191(1):89-93.
33. Heine T*, et al.* (2018) On the enigma of glutathione dependent styrene degradation in *Gordonia rubripertincta* CWB2. *Appl Environ Microbiol*.
34. Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, & Schulze-Lefert P (2013) Structure and functions of the bacterial microbiota of plants. *Annu Rev Plant Biol* 64:807-838.

35. Vorholt JA (2012) Microbial life in the phyllosphere. *Nat Rev Micro* 10(12):828-840.
36. Brüggemann N & Schnitzler JP (2002) 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):456-463.
37. Sun Z, Hüve K, Vislap V, & Niinemets Ü (2013) Elevated [$CO_2$] magnifies isoprene emissions under heat and improves thermal resistance in hybrid aspen. *J Exp Bot* 64(18):5509-5523.
38. Singsaas EL, Lerdau M, Winter K, & Sharkey TD (1997) Isoprene increases thermotolerance of isoprene-emitting species. *Plant Physiology* 115(4):1413-1420.
39. Abanda-Nkpwatt D, Musch M, Tschiersch J, Boettner M, & Schwab W (2006) Molecular interaction between *Methylobacterium extorquens* and seedlings: growth promotion, methanol consumption, and localization of the methanol emission site. *J Exp Bot* 57(15):4025-4032.
40. Junker RR & Tholl D (2013) Volatile organic compound mediated interactions at the plant-microbe interface. *Journal of Chemical Ecology* 39(7):810-825.
41. Bringel F & Couée I (2015) Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics. *Front Microbiol* 6.
42. Rosenstiel TN, Potosnak MJ, Griffin KL, Fall R, & Monson RK (2003) Increased $CO_2$ uncouples growth from isoprene emission in an agriforest ecosystem. *Nature* 421(6920):256-259.
43. Rosenkranz M, Pugh TAM, Schnitzler J-P, & Arneth A (2015) Effect of land-use change and management on biogenic volatile organic compound emissions – selecting climate-smart cultivars. *Plant Cell Environ* 38(9):1896-1912.
44. Crombie AT, Emery H, McGenity TJ, & Murrell JC (2017) Draft genome sequences of three terrestrial isoprene-degrading *Rhodococcus* strains. *Genome Announc* 5(45):e01256-01217.
45. Sharkey TD, Loreto F, & Delwiche CF (1991) High carbon dioxide and sun/shade effects on isoprene emission from oak and aspen tree leaves. *Plant Cell Environ* 14(3):333-338.
46. Redford AJ, Bowers RM, Knight R, Linhart Y, & Fierer N (2010) The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environ Microbiol* 12(11):2885-2893.
47. Laforest-Lapointe I, Messier C, & Kembel SW (2016) Tree phyllosphere bacterial communities: exploring the magnitude of intra- and inter-individual variation among host species. *PeerJ* 4:e2367.
48. Truong DT*, et al.* (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Meth* 12(10):902-903.
49. Sales CM*, et al.* (2011) Genome sequence of the 1,4-dioxane-degrading *Pseudonocardia dioxanivorans* strain CB1190. *J Bacteriol* 193(17):4549-4550.
50. Allen JR, Clark DD, Krum JG, & Ensign SA (1999) A role for coenzyme M (2-mercaptoethanesulfonic acid) in a bacterial pathway of aliphatic epoxide carboxylation. *Proc Natl Acad Sci* 96(15):8432-8437.
51. Krum JG & Ensign SA (2001) Evidence that a linear megaplasmid encodes enzymes of aliphatic alkene and epoxide metabolism and coenzyme M (2-mercaptoethanesulfonate) biosynthesis in *Xanthobacter* strain Py2. *J Bacteriol* 183(7):2172-2177.
52. Han JI*, et al.* (2011) Complete genome sequence of the metabolically versatile plant growth-promoting endophyte *Variovorax paradoxus* S110. *J Bacteriol* 193(5):1183-1190.
53. Konstantinidis KT & Tiedje JM (2005) Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* 187(18):6258-6264.
54. Belimov AA*, et al.* (2009) Rhizosphere bacteria containing 1-aminocyclopropane-1-carboxylate deaminase increase yield of plants grown in drying soil via both local and systemic hormone signalling. *New Phytol* 181(2):413-423.
55. Satola B, Wübbeler JH, & Steinbüchel A (2013) Metabolic characteristics of the species *Variovorax paradoxus*. *Appl Microbiol Biotechnol* 97(2):541-560.

56. Bodenhausen N, Horton MW, & Bergelson J (2013) Bacterial Communities Associated with the Leaves and the Roots of Arabidopsis thaliana. *PLOS ONE* 8(2):e56329.
57. Beckers B, Op De Beeck M, Weyens N, Boerjan W, & Vangronsveld J (2017) Structural variability and niche differentiation in the rhizosphere and endosphere bacterial microbiome of field-grown poplar trees. *Microbiome* 5(1):25.
58. Klindworth A*, et al.* (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41(1):e1.
59. Kuczynski J*, et al.* (2005) Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Curr Protoc Microbiol*, (John Wiley & Sons, Inc.).
60. Peng Y, Leung HC, Yiu SM, & Chin FY (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420-1428.
61. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, & Singer SW (2014) MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2(1):1-18.
62. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, & Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* 25(7):1043-1055.
63. Parks DH*, et al.* (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology* 2(11):1533-1542.
64. Haas BJ*, et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494.
65. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. *J. Mol. Biol.* 215(3):403 - 410.
66. Bray NL, Pimentel H, Melsted P, & Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34:525.
67. Nakashima N & Tamura T (2004) Isolation and characterization of a rolling-circle-type plasmid from *Rhodococcus erythropolis* and application of the plasmid to multiple-recombinant-protein expression. *Appl Environ Microbiol* 70(9):5557-5568.

## Figure Legends

**Figure 1.** The isoprene gene cluster in *Rhodococcus* sp. AD45. Genes present in all previously characterised isoprene degraders are shown in solid colours. In many isolates, *aldh1* is located between *isoJ* and *isoA*, as shown (Crombie et al., 2017). The horizontal line shows the extent of co-transcribed genes in *R.* sp. AD45 (Crombie et al., 2015).

**Figure 2.** Community profile of the un-enriched (timepoint zero) and unlabelled (light) and labelled (heavy) fractions of $^{13}$C-isoprene incubations from two DNA-SIP experiments. For the first experiment, the labelled and unlabelled bacterial communities were characterised by amplicon sequencing of the 16S rRNA gene (16S), using DNA extracted from cells at the start of the experiment (un-enriched timepoint zero) and DNA from light and heavy fractions of incubations with $^{13}$C-labelled and unlabelled $^{12}$C-isoprene (SI Appendix, Table S2). The microbial communities of the second experiment were analysed by shotgun metagenomic sequencing (MG) of the timepoint zero DNA, the DNA from the heavy fractions of each of triplicate $^{13}$C-isoprene incubations, pooled DNA from the light fractions of $^{13}$C-isoprene

563 incubations and pooled DNA from samples incubated without substrate (no-subs), as well as
564 pooled DNA from the heavy fractions of $^{13}$C-isoprene incubations from the first DNA-SIP
565 experiment. Taxa present at > 10% are shown in bold. 16S and enriched samples show the
566 mean of triplicates. For complete data, including controls, see SI Appendix, Fig. S3.
567
568 **Figure 3.** The relationship between the *isoA* genes of known isoprene degraders (in bold),
569 metagenome scaffold sequences (prefixed MG) and metatranscriptome sequences, together
570 with other representative sequences from the databases. Transcripts are prefixed MT
571 followed by sample identification (Un-enriched, timepoint zero; No-subs, incubations without
572 isoprene; $^{12}$C-1 – $^{12}$C-3, incubations with unlabelled isoprene; $^{13}$C-1 – $^{13}$C-3, incubations with
573 labelled isoprene). Scaffolds or transcripts containing isoprene-related genes in addition to
574 *isoABCDEF* are indicated with **. For each sample, transcripts were ranked by normalised
575 transcript abundance (TPM) and highly expressed transcripts are marked with four, three,
576 two, or one red circles, indicating that the *isoA*-containing transcript was among the most
577 abundant 0.2%, 1%, 10%, or 50% respectively of all transcripts from that sample (SI
578 Appendix,Table S4). Where identical *isoA* sequences were present on different transcripts
579 from the same sample, only the most highly expressed is shown. The taxonomy of genome
580 bins is shown after the scaffold identification. NA, not assigned; NB, not binned. Bootstrap
581 values over 50% (1000 replications) are shown as solid circles at the nodes. The scale bar
582 indicates nucleotide substitutions per site.
583
584 **Figure 4.** The isoprene metabolic gene clusters from known isoprene-degrading isolates (in
585 bold) together with representative sequences from the assembled metagenome (prefix MG_,
586 with predicted taxonomy) and metatranscriptomes (prefix MT_).
587
588
589

# Functional gene probing reveals the widespread distribution, diversity and abundance of isoprene-degrading bacteria in the environment.

Ornella Carrión[1], Nasmille L. Larke-Mejía[1], Lisa Gibson[1], Muhammad Farhan Ul Haque[1], Javier Ramiro-García[2], Terry J. McGenity[3], J. Colin Murrell[1*]

[1]School of Environmental Sciences, University of East Anglia, Norwich, UK

[2]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[3]School of Biological Sciences, University of Essex, Colchester, UK

[*]Corresponding author:

J Colin Murrell, [1]School of Environmental Sciences, University of East Anglia

Norwich Research Park, NR4 7TJ, UK

E-mail: j.c.murrell@uea.ac.uk

Tel: (+44) 01603 592959

Fax: (+44) 01603 591327

## Abstract

**Background:** Approximately 500 Tg of isoprene are emitted to the atmosphere annually, an amount similar to that of methane, and despite its significant effects on the climate, very little is known about the biological degradation of isoprene in the environment. Isolation and characterisation of isoprene degraders at the molecular level has allowed the development of functional probes targeting *isoA* encoding the α-subunit of the isoprene monooxygenase. This enzyme belongs to the soluble diiron centre monooxygenase family and catalyses the first step in the isoprene degradation pathway. The use of probes targeting functional genes is a successful approach in molecular ecology to study specific groups of bacteria in complex environments. Here, we developed and tested a novel *isoA* PCR primer set to study the distribution, abundance and diversity of isoprene degraders in a wide range of environments.

**Results:** The new *isoA* probes specifically amplified *isoA* genes from taxonomically diverse isoprene-degrading bacteria including members of the genera *Rhodococcus, Variovorax* and *Sphingopyxis*. There was no cross-reactivity with genes encoding related oxygenases from non-isoprene degraders. Sequencing of *isoA* amplicons from DNA extracted from environmental samples enriched with isoprene revealed that most environments tested harboured a considerable variety of *isoA* sequences, with poplar leaf enrichments containing more phylogenetically diverse *isoA* genes. Quantification by qPCR using these *isoA* probes revealed that isoprene degraders are widespread in the phyllosphere, terrestrial, freshwater and marine environments. Specifically, soils in the vicinity of high isoprene-emitting trees contained the highest number of isoprene-degrading bacteria.

**Conclusion:** This study provides the molecular ecology tools to broaden our knowledge of the distribution, abundance and diversity of isoprene degraders in the environment, which is a fundamental step necessary to assess the impact that microbes have in mitigating the effects of this important climate-active gas.

**Keywords:** Isoprene, climate, isoprene monooxygenase, *isoA*, functional gene probes.

## Background

Isoprene (2-methyl-1,3-butadiene) comprises approximately one third of the total volatile organic compounds (VOC) emitted to the atmosphere, an amount that is approximately equal to emissions of methane [1, 2]. Although isoprene has a short lifetime in the atmosphere (in the order of hours) due to rapid photochemical oxidation [1], it has a significant impact on atmospheric chemistry and hence climate [3]. In unpolluted environments with low levels of nitrogen oxides, isoprene reacts with hydroxyl radicals, thus reducing the oxidizing capacity of the atmosphere [1]. This, in turn, prolongs the lifetime of greenhouse gases such as methane and enhances global warming [4, 5]. In polluted environments, nitrogen oxides are typically present at high concentrations and react with isoprene, leading to the formation of tropospheric ozone [1], which is a greenhouse gas with important negative effects on plant and animal health [6]. Conversely, atmospheric oxidation of isoprene results in the formation of secondary organic aerosols and cloud condensation nuclei, which in turn promotes global cooling [7, 8]. The vast majority of isoprene emitted to the atmosphere is produced by terrestrial plants (~500 Tg $y^{-1}$) [2, 9], with small contributions from marine algae (0.1-12 Tg $y^{-1}$) and minor contributions from bacteria, fungi and animals [10-17].

Isoprene is also produced industrially (~0.8 Tg $y^{-1}$), where it is used primarily to synthesize polyisoprene rubber [18]. In plants, isoprene is synthesized in the chloroplast from dimethylallyl diphosphate (DMAPP), an intermediate of isoprenoid biosynthesis, in a reaction mediated by isoprene synthase [19]. It has been shown that isoprene protects plants against heat and oxidative stress [9, 20] and it has also been suggested that it might have a role in plant-insect interactions [21] and plant energy dynamics [22]. However, not all plants produce isoprene, with both high and low emitters being observed even among closely related species [23-25].

Although atmospheric levels of isoprene are low (1-4 ppb) [26]; due to its high reactivity, concentrations are significantly higher (up to 36 ppb) at ground level in high isoprene-emitting forests [27]. Closed chambers and continuous-flow experiments have shown that soils can act as a biological sink for isoprene at environmentally relevant concentrations [28-30]. These studies confirmed the potential for soil microbes to consume isoprene released locally in soils as well as from the atmosphere. In fact, bacteria that grow on isoprene as sole carbon and energy source have been isolated from soils, leaves and coastal/marine environments [31-37]. These isolates are mainly Actinobacteria, although recently more Alpha- and Betaproteobacteria strains such as *Sphingopyxis* sp. OPL5 and *Variovorax* sp. WS11 have also been isolated [38]. All known isoprene degraders contain six genes (*isoABCDEF*) encoding the isoprene monooxygenase (IsoMO) that catalyses the first step of the isoprene degradation pathway. Four additional genes, *isoGHIJ*, are located immediately upstream (5') of the IsoMO structural genes and encode enzymes involved in the subsequent steps in isoprene catabolism [32, 33, 39, 40]. The IsoMO is a four-component soluble diiron monooxygenase (SDIMO) composed of a dimeric hydroxylase, a NAD(P)H oxidoreductase, a coupling protein and a Rieske-type

ferredoxin. Other members of the SDIMO family include the soluble methane monooxygenase (sMMO), alkene monooxygenases, phenol hydroxylases and aromatic monooxygenases, which are key enzymes in the bacterial oxidation of hydrocarbons and have biotechnological applications [41, 42]. In addition, structural and genetic analyses have revealed that the hydroxylase α-subunit of SDIMOs contains a carboxylated-bridge diiron centre in a distinctive 4-helix bundle structure at the active site (reviewed in [42]).

The IsoMO catalyses the initial oxidation of isoprene to 1,2-epoxy-isoprene. The epoxide is converted to 1-hydroxyl-2-glutathionyl-2-methyl-3-butene (HGMB) by a glutathione *S*-transferase (IsoI) and then by a dehydrogenase (IsoH) to 2-glutathionyl-2-methyl-3-butenoate (GMBA) [33]. The fate of GMBA is uncertain. It is assumed that subsequent removal of glutathione and β-oxidation of these intermediates enable isoprene degraders to grow on isoprene as a carbon source but the final steps in the catabolism of isoprene remain to be elucidated.

The identity and availability of genes essential for isoprene metabolism provide tools for cultivation-independent studies to assess the distribution, diversity and activity of isoprene degraders in the environment. Functional gene probes are important molecular ecology tools to study functional groups of interest in complex environments. For example, primers targeting the α-subunit of sMMO have been used to extend our knowledge of the diversity and abundance of methane-oxidising bacteria in many environments [43-46]. The *isoA* gene, encoding the α-subunit harbouring the active site of IsoMO, is highly conserved in all isoprene degraders studied and is a suitable marker gene for isoprene degradation [36]. Primers targeting *isoA*, tested negative with genes encoding the corresponding active site of SDIMOs from non-isoprene degraders, but amplified *isoA* from extant isoprene-degrading bacteria and

from a range of environmental samples enriched with isoprene, generating sequences which were >86% identical to those of bona-fide isoprene degraders [36]. However, the increasing number and variety of isoprene degraders has revealed a higher diversity of *isoA* sequences, which emphasizes the need to refine *isoA* primers to cover all the isoprene-degrading strains characterised to date. Here, we designed new *isoA* probes that amplified *isoA* genes from all extant isoprene degraders, but did not amplify *isoA* homologues of related enzymes from non-isoprene degraders. These new *isoA* probes were then used to investigate the distribution, diversity and abundance of isoprene degraders in phyllosphere, soils, freshwater and marine environments to better understand their role in the isoprene biogeochemical cycle.

## Results and Discussion

### Design and validation of *isoA* primers

In order to cover the diversity of all well-characterised isoprene degraders, we designed new functional probes targeting *isoA*, as this gene encodes the α-subunit containing the active site of the IsoMO and is an excellent marker gene for isoprene degradation [36]. We aligned the *isoA* genes from 38 bona-fide isoprene degraders available from Genbank and strains recently isolated by our group (Additional File 1; Table S1). Microorganisms included in the analysis belonged to the classes Actinobacteria (e.g. *Gordonia, Mycobacterium, Rhodococcus*), Alphaproteobacteria (e.g. S*phingopyxis*) and Betaproteobacteria (e.g. *Variovorax*). Eighteen *isoA* sequences detected in the metagenomes from isoprene enrichments of willow soil, willow leaves and poplar leaves (unpublished data) were added to the *isoA* database to design the new probes. These metagenome-derived sequences had a minimum

query coverage of 98% and an identity of ≥ 85% at the derived amino acid level to ratified IsoA sequences. Finally, genes encoding the α-subunit of other SDIMOs such as sMMO, alkene monooxygenase or toluene monooxygenase from non-isoprene-degrading microorganisms (Additional File 1; Table S1) were also included in the alignment to guide the specific amplification of *isoA* genes by the new probes. Conserved positions within the *isoA* gene were identified and various sets of primers spanning different regions were manually designed (Additional File 1; Table S2; Figure S1). Eleven different combinations of primers were initially tested, including isoA14F and isoAR3, which have been previously investigated [38]. In this preliminary validation of the *isoA* probes, *Rhodococcus* sp. AD45 and *Variovorax* sp. WS9 were selected as representative Gram-positive and Gram-negative isoprene-degrading bacteria. *Xanthobacter autotrophicus* Py2 was chosen as a negative control since the α-subunit of its alkene monooxygenase is closely related to IsoA (70% amino acid identity to IsoA from *Rhodococcus* sp. AD45).

Five out of the eleven combinations of primers tested yielded a PCR product of the expected size from the positive control strains, but there was no amplification from *Xanthobacter autotrophicus* Py2 (Additional File 1; Table S3). Combinations isoA136F+isoA1019R, isoA300F+isoA1019R, isoA379F+isoA862R and isoA379F+isoA1019R also generated non-specific amplification products (data not shown). However, the combination of primers isoA14F and isoA511R, which spans the first iron centre of the IsoMO α-subunit (Additional File 1; Figure S1), yielded a specific PCR product of 497 bp. Therefore, we selected primers isoA14F and isoA511R for further validation on genomic DNA from additional isoprene-degrading and non-degrading isolates (Additional File 1; Table S4). A specific amplification product of the expected size was obtained for all 30 positive control strains used in

this study. To check the cross-reactivity of the primers isoA14F and isoA511R, we used as negative controls 12 non-isoprene-utilising strains with related oxygenases to IsoMO that grow on alkanes, alkenes or aromatic compounds. Examples include bacteria containing sMMO (*Methylococcus capsulatus* Bath), toluene monooxygenase (*Pseudomonas mendocina* KR1) or alkene monooxygenase (*Rhodococcus rhodochrous* B276). We also studied as negative controls two strains that belong to the same genera as the bona-fide isoprene-degraders *Rhodococcus* (*Rhodococcus opacus* DSM 1069) and *Variovorax* (*Variovorax paradoxus* EPS), but do not oxidise isoprene (Additional File 1; Table S4). No PCR products were obtained with the primers isoA14F and isoA511R with template DNA from any of the negative control strains. To check that the lack of amplification was not due to the quality of the DNA, a 16S rRNA gene PCR was performed and strong amplification products were obtained with DNA from all 14 negative control strains (data not shown). Therefore, we conclude that the primers isoA14F and isoA511R are specific for *isoA* encoding the α-subunit of the IsoMO and ratify that *isoA* is an excellent marker gene to determine if isoprene-degrading bacteria contain IsoMO.

Diversity of *isoA* genes in environmental samples

To test the specificity of the new *isoA* primer set and to investigate the diversity of *isoA* genes, and thus isoprene degraders in various environments, we enriched 11 samples from phyllosphere, soils, freshwater and marine environments with isoprene (Additional File 1; Table S5). DNA extracted from these enrichments was subjected to PCR amplification with the *isoA* primer set isoA14F and isoA511R. A single PCR product of the correct size (497 bp) was obtained with all the enrichments. To confirm

191 that these PCR products contained only *isoA* genes, before committing to high-
192 throughput sequencing of *isoA* amplicons, 9 out of the 11 enriched environmental
193 samples were selected to construct *isoA* libraries from purified PCR amplicons.
194 Seventy-one clones from these *isoA* libraries were sequenced (Additional File 1: Table
195 S5; Figure 1). Bioinformatic analysis using BLASTx [47] confirmed all sequences as
196 *isoA* genes. No hits to proteins distinct from IsoA were detected in any enrichment.

197 The diversity of *isoA* sequences in DNA extracted from enriched samples from leaves
198 varied according to the type of tree sampled. For example, *isoA* genes from ash leaf
199 enrichments were similar to *isoA* from *Rhodococcus*, whereas *isoA* genes from oil
200 palm leaf samples were phylogenetically close to *isoA* from *Gordonia* (Figure 1). It was
201 not surprising to find *Rhodococcus isoA* homologues in leaf enrichments, as several
202 isoprene-degrading bacteria from this genus have been obtained from poplar, willow,
203 oil palm and horse chestnut leaves [35, 36, 38]. In addition, *Gordonia* strains able to
204 grow on isoprene have been previously isolated from an estuarine environment and
205 oil palm leaves [31, 38].

206 Most of the *isoA* sequences retrieved from soils were phylogenetically close to *isoA*
207 from *Variovorax*, although *Rhodococcus isoA* homologues were also found in DNA
208 from tyre dump soil enrichments (i.e. tyre dump soil clone 14; Figure 1). *Variovorax*
209 strains are common inhabitants of soil and water [48-50], and are frequently
210 associated with the phyllosphere [51, 52] and rhizosphere [53, 54] of plants. Indeed,
211 some species of this genus, such as *Variovorax paradoxus*, are considered plant
212 growth-promoting rhizobacteria that exert beneficial effects on plant growth [54].
213 *Variovorax* species are metabolically versatile bacteria that can degrade a wide range
214 of natural and xenobiotic compounds such as alkanesulfonates, polychlorinated
215 biphenyls or trichloroethylene [55]. Larke-Mejía [38] has shown that two novel

*Variovorax* strains, isolated from soil surrounding a willow tree can grow on isoprene. The presence of *Rhodococcus isoA* homologues in soil collected from a tyre dump site is consistent with the recent isolation of isoprene-degrading *Rhodococcus* strains from these samples [38].

In freshwater sediment enrichments, the predominant *isoA* sequences were phylogenetically closer to *isoA* from *Rhodococcus*, though sequences similar to *isoA* from *Sphingopyxis* were also present (i.e. freshwater sediment clone 5; Figure 1). In fact, *Rhodococcus* sp. AD45, the most well-characterised isoprene degrader, was isolated from freshwater sediment [32]. *Sphingopyxis* is a genus commonly associated with the phyllosphere [51, 52], although *Sphingopyxis* species have also been isolated from soils, freshwater and seawater samples [56-59]. It was not until the isolation of *Sphingopyxis* sp. OPL5 [38], however, that members of this genus were shown to metabolise isoprene.

Finally, in salt marsh sediment enrichments, only *isoA* sequences from *Rhodococcus* were found, whereas enrichments with coastal sediment yielded more diverse *isoA* sequences, with both *Rhodococcus* and *Variovorax isoA*-like genes being obtained from the clone libraries (i.e. coastal sediment clones 4 and 5; Figure 1). Indeed, the presence of *Rhodococcus isoA* homologues in marine environments has been confirmed with the recent isolation of two isoprene-degrading *Rhodococcus* strains from salt marsh and coastal sediments (unpublished data).

Since the sequences obtained from the clone libraries were specific to *isoA* and showed variability within and across the different ecosystems studied, we explored in more detail the diversity of *isoA* genes in the environment using high-throughput sequencing. Purified *isoA* PCR products from DNA isolated from 11 enriched

environmental samples were sequenced using Illumina MiSeq technology (Additional File 1; Table S5). *isoA* amplicon sequencing yielded a total of 136,986 quality-filtered sequences with an average of 12,453 reads per sample. These *isoA* sequences, when analysed by the DADA2 pipeline [60], grouped into 136 unique operational taxonomic units (OTUs) that were manually checked by BLASTx (see Methods). Two OTUs, representing 0.3% and 2.5% of the sequences from DNA extracted from enriched poplar leaves and freshwater sediment, respectively, were discarded for the analysis as they had hits to proteins not related to IsoA or other SDIMO α-subunits. Therefore, a final set of 134 OTUs was used for downstream analysis (Figure 1; Additional File 2; Table S6).

The *isoA* amplicon sequencing analysis revealed that the phyllosphere yielded most variability between samples when compared to other environments (Figure 2). For example, 99.8% of the *isoA* sequences from ash leaf enrichments were similar to *isoA* from *Rhodococcus*. *isoA* genes similar to those of *Gordonia* and *Mycobacterium* were also found, although at low relative abundance (<1%). Conversely, DNA from oil palm leaf enrichments yielded predominantly *Gordonia* and *Sphingopyxis isoA* homologues, even though *isoA* genes from *Rhodococcus* and *Mycobacterium* were also present (<1%). Indeed, we have recently isolated isoprene-degrading *Gordonia* and *Sphingopyxis* strains from oil palm leaves [38]. The predominant sequences in willow leaf enrichments were phylogenetically close to *isoA* from *Mycobacterium* (50.7%) and *Rhodoccocus* (49.3%). Finally, poplar leaf enrichments yielded more phylogenetically diverse *isoA* genes, with homologues to *isoA* from *Gordonia* (2.3%)*, Mycobacterium* (40%)*, Rhodococcus* (50%)*, Sphingopyxis* (1.9%) and *Variovorax* (5.8%, Figure 2). Although these genera are common inhabitants of the phyllosphere and soils [50, 51,

54, 61], no *Variovorax* or *Mycobacterium* strains from leaves able to metabolise isoprene have been reported so far.

The most abundant *isoA* sequences identified in willow, oil palm and tyre dump soil enrichments were similar to *Rhodococcus* and Variovorax *isoA* genes (Figure 2). *isoA* sequences from *Mycobacterium, Nocardioides* and *Sphingopyxis* were also detected in willow soil (<1%). Sequences similar to *isoA* from *Gordonia, Mycobacterium* and *Sphingopyxis* were present at low relative abundance (<1%) in DNA extracted from tyre dump soil. These results are consistent with the isolation of *Nocardioides, Rhodococcus* and *Variovorax* isoprene-degrading strains from willow and tyre dump soils [35, 38].

Finally, *isoA* homologues from *Sphingopyxis* (46.8%) and *Rhodococcus* (53.2%) dominated DNA from freshwater sediment enrichments, whereas *isoA* sequences retrieved from coastal and salt marsh enriched samples had highest identity to *isoA* from *Rhodococcus* (Figure 2).

In addition, several sequences were distinct from those of *isoA* of bona-fide isoprene degraders, indicating that there is likely novel diversity of isoprene-utilising bacteria yet to be discovered. For example, most of the *isoA* clones and OTUs originating from enriched freshwater sediment samples (i.e. freshwater sediment clones 5 and 8; OTUs 34 and 38) occupied a distinct position in the IsoA phylogenetic tree as shown in Figure 1. BLASTx analysis of these sequences revealed that they had 83-91% identity at the derived amino acid level to IsoA from extant isoprene-degraders of the genera *Rhodococcus* and *Sphingopyxis*, suggesting that this environment harbours novel isoprene-utilising strains. The sequencing information obtained using the new *isoA* probes isoA14F and isoA511R, can now be used to design targeted enrichment and

isolation strategies to isolate novel species of isoprene degraders from various environments and expand the diversity of existent isoprene-degrading bacteria.

Distribution and abundance of isoprene degraders in the environment

*isoA* primer set isoA14F and isoA511R was used to study the distribution and abundance of *isoA*-containing bacteria using qPCR. *isoA* primers were first optimised and validated on DNA extracted from environmental samples enriched with isoprene (Additional File 1: Figure S3). Subsequently, clone libraries from qPCR products obtained from both enriched and natural (non-enriched) environmental samples were constructed to ensure that *isoA* qPCR products were absolutely specific for *isoA*. Fifty-two clones from these *isoA* libraries were sequenced. All sequences had 89-100% identity at the derived amino acid level to IsoA from bona-fide isoprene degraders and no hits to proteins distinct from IsoA were detected in any sample (data not shown).

After qPCR assay validation, the abundance of isoprene degraders across a wide range of natural samples, including leaves, soils, freshwater and marine sediments was studied (Figure 3). Interestingly, leaves from high isoprene-emitting trees such as willow, poplar and oil palm [62, 63] contained similar *isoA* numbers as those from an ash tree, a low isoprene emitter (between 11.6 ± 3.6 and 23.7 ± 6.1 *isoA* genes per million copies of 16S rRNA gene in DNA extracted from high isoprene-emitting trees, versus 22.4 ± 5.6 *isoA* genes per million copies of 16S rRNA gene in DNA from ash leaves; Figure 3). However, a larger number of environmental samples will be required to confirm statistically that there are no significant differences in the abundance of isoprene degraders between high and low-emitting trees. Moreover, soils sampled in the vicinity of high isoprene-emitting trees yielded the highest numbers of isoprene

312 degraders (with values ranging from 122.2 ± 5.0 to 303.3 ± 60.3 *isoA* genes per million
313 copies of 16S rRNA gene; Figure 3). The fact that there were 10-fold more isoprene
314 degraders in soils than in the leaves of the same trees, could be explained by the flux
315 of bacteria from the phyllosphere to the pedosphere during rainfall, which has been
316 estimated to be up to 1.5 x $10^{16}$ cells $ha^{-1}$ $y^{-1}$ in a subtropical oak-cedar forest [64]. It
317 was not surprising to see relatively high numbers of *isoA* genes (67.7 ± 14.4 *isoA*
318 genes per million copies of 16S rRNA gene) in soil sampled from a tyre dump, since
319 tyres mainly consist of polyisoprene rubber [65, 66]. Although several bacteria and
320 fungi can degrade polyisoprene rubber [67, 68], some studies have suggested that
321 actinomycetes are the key microorganisms carrying out this process [69-72]. The
322 cleavage of polyisoprene rubber by Lcp and RoxA oxygenases results in the
323 production of oligoisoprene molecules [73] that could potentially be used by the
324 isoprene degraders for growth. Indeed, rubber-contaminated soils have yielded
325 isoprene-utilising bacteria from different genera, including *Rhodococcus* [38, 74],
326 although these strains were not characterised in detail. Grassland and landfill soils
327 also contained isoprene degraders, although at lower relative abundance (from 12.4
328 ± 1.3 to 25.4 ± 2.8 *isoA* genes per million copies of 16S rRNA gene) than soils
329 surrounded by high isoprene-emitting trees, as anticipated (Figure 3). The presence
330 of isoprene-degrading microorganisms was also studied in freshwater and marine
331 natural samples. Surprisingly, these samples yielded similar copy numbers of *isoA*
332 genes (ranging from 12.0 ± 1.9 to 25.1 ± 2.8 *isoA* copies per million copies of 16S
333 rRNA gene) as leaves (**Figure 3**). This was unexpected since isoprene emissions from
334 marine environments (0.1-12 Tg $y^{-1}$) [2, 11, 16] are very much lower than from
335 terrestrial plants (500 Tg $y^{-1}$) [2, 9]. In marine environments, isoprene is synthesised
336 by phytoplankton, heterotrophic bacteria and seaweeds [16]. However, only a few

studies have directly measured the isoprene concentration in the euphotic zones of the oceans and the exact mechanism behind marine isoprene production is still unknown [75]. Another poorly understood aspect of the marine isoprene cycle is the role of microbial degradation. Ocean depth profiles of isoprene concentrations have suggested that isoprene is biologically consumed [75-78]. Indeed, Acuña Alvarez *et al*. [31] and Johnston *et al.* [37] found isoprene-degrading bacteria in estuarine and marine water samples, most of which were Actinobacteria. More importantly, Acuña Alvarez et al. [31] also showed that isoprene-utilising bacteria degraded isoprene from the headspace of microalgae cultures at environmentally relevant concentrations. A recent study by Steinke et al. [17] has reported that freshwater lakes also emit isoprene to the atmosphere. Therefore, additional laboratory experiments and field studies as well as more accurate models are required to better understand the isoprene cycle in marine and freshwater environments.

## Conclusions

Although we cannot exclude the possibility that *isoA* genes from novel uncultivated isoprene-degrading bacteria have been missed, or that other pathways of isoprene metabolism exist, new probes targeting the *isoA* gene encoding the active site of the IsoMO, have proven to be a successful tool to study the diversity, distribution and abundance of isoprene degraders in a wide range of environments. This now facilitates the development of targeted strategies to isolate novel genera of isoprene degraders, monitor them under natural conditions and to determine how isoprene degradation is regulated in the environment. This study provides molecular probes to investigate the

significance of the isoprene biological sink and how bacteria might mitigate the effect on the atmosphere of this abundant climate-active gas.

## Methods

Enrichments of environmental samples with isoprene

To study the distribution, abundance and diversity of isoprene degraders in the environment, a wide range of terrestrial, phyllosphere, freshwater and marine samples were collected (Additional File 1: Table S5). Terrestrial samples comprised soils from the vicinity of high isoprene-emitting trees (oil palm and willow) and soils without nearby high isoprene-emitting vegetation (grassland and landfill soils). Soil from a tyre dump was also studied since several isoprene-degrading strains have been isolated from rubber-contaminated soils [38, 74]. Samples from the phyllosphere included leaves from high isoprene-producing trees as for example, willow, poplar and oil palm, and low emitters such as ash trees. Freshwater samples were collected from a local lake. Samples from marine environments consisted of coastal and salt marsh sediments (Additional File 1: Table S5).

In the case of soils, freshwater and marine sediments, only the top 3 cm were collected after removing vegetation or macroscopic algae. 1 g of material was used to set up microcosms enrichments in 120 ml sealed vials containing 10 ml of water for soils, Ewers medium [79] for freshwater sediments or marine basal medium (MBM) [80]; for marine samples. Salinity of MBM was adjusted to 20 practical salinity units (PSU) in coastal sediment samples and to 35 PSU in salt marsh sediments to mimic natural conditions.

The surface of 5-10 leaves from each tree (depending on leaf size) was sampled by washing leaves with Ewers medium or using cotton swabs. For leaf washings, leaves were immersed in 50 ml of Ewers medium in 250 ml conical flasks and submerged in a water bath for sonication (5 min, 50 kHz; Mettler ME2), followed by shaking on an orbital shaker (1 h, 150 rpm) to detach microbial cells from the surface of the leaves. Leaves were then removed from the flasks and medium was centrifuged (5,000 x *g*, 20 min) to separate cells and particulate material from the supernatant. Supernatant was filtered through CellTrap CT402LL001N00 filters (Mem-Teq). Pellets of cells and particulate material were combined with the filtrated supernatant, resuspended in 10 ml of fresh minimal medium and transferred to 120 ml sealed vials. When leaf surfaces were sampled with cotton swabs, the swabs were placed in 250 ml conical flasks containing 50 ml of Ewers medium and sonicated in a water bath as above. After shaking in an orbital shaker for 1 h at 150 rpm, cotton swabs were removed and medium was aliquoted into 120 ml sealed vials.

All enrichments were set up in triplicate and incubated at 25 ˚C with 25 ppm isoprene, except for oil palm leaf samples from Malaysia, which were incubated at 30 ˚C with 50 ppm of isoprene. Consumption of isoprene was monitored daily by gas-chromatography as described in [34]. When isoprene was depleted in the headspace, samples were spiked again with 25 ppm of isoprene or 50 ppm in the case of oil palm leaf samples from Malaysia. Enrichments were subcultured at two-week intervals three times by making 1/10 dilutions of the samples in fresh medium.

## DNA extraction from bacterial strains and environmental samples

Genomic DNA from positive and negative control strains was extracted from cultures grown in rich media using the Wizard Genomic DNA Purification Kit (Promega), according to the manufacturer's instructions.

To extract DNA from environmental samples, the FastDNA™ SPIN kit for Soil (MP Biomedicals) was used following the protocol described by the manufacturer.

## Amplification of *isoA* genes

50 ng of genomic DNA or 20 ng to 1 µg of environmental DNA were used as a template in a 50 µl PCR reaction containing 4 µM of isoA14F and isoA511R primers. The PCR program consisted of an initial step of 94 °C for 2 min, followed by 31 cycles of 95 °C for 15 s, 54 °C for 30 s, 72 °C for 1 min and a final extension step of 72 °C for 7 min. In the case of freshwater and marine DNA samples, 40 cycles were carried out to obtain an amplicon visible on an agarose gel stained with ethidium bromide.

## Clone libraries

*isoA* PCR products from environmental samples were purified using the NucleoSpin gel and PCR Clean-up kit (Macherey-Nagel) and cloned into the pGEM®-T easy vector system (Promega) following the manufacturer's instructions prior to transformation into *Escherichia coli* TOP10 cells. Positive clones were screened by PCR using the M13F and M13R primers. Clones yielding a PCR product were sent for sequencing using M13 primers.

430 *isoA* amplicon sequencing

431 Duplicate PCR products from each environmental sample were pooled before DNA
432 purification using the NucleoSpin gel and PCR Clean-up kit (Macherey-Nagel). The
433 quality of DNA was assessed by gel electrophoresis and the Qubit dsDNA High
434 Sensitivity Assay Kit (ThermoFisher) according to the manufacturer's instructions.
435 Purified *isoA* amplicons from enriched environmental samples were subjected to
436 Illumina Mi-Seq sequencing by MrDNA (Shallowater, TX, USA) using an Illumina
437 MiSeq platform.

438 *isoA* amplicon sequencing data were analysed using DADA2 pipeline [60] with default
439 filtering parameters. Reads were truncated at 275 nucleotides and quality-filtered if
440 their expected error was higher than two. After denoising the sequences using the
441 estimated error rates, forward and reverse reads were merged. Resultant sequences
442 were screened for chimeras and then manually checked by BLASTx [47]. Those OTUs
443 with a top hit distinct from a ratified *isoA* sequence were discarded, obtaining a final
444 set of 134 unique OTUs for downstream analysis.

445

446 Quantitative real-time PCR

447 Quantification of isoprene degraders in environmental samples was estimated by
448 qPCR targeting the *isoA* gene using primers isoA14F and isoA511R (Additional File
449 1; Table S2). qPCR assays were carried out using a StepOne Plus real-time PCR
450 instrument (Applied Biosystems). qPCR reactions (20 µl) contained 2-20 ng of DNA,
451 400 nM of each primer and 10 µl of SensiFast SYBR Hi-ROX kit (Bioline). The qPCR
452 reaction consisted of an initial denaturation step at 95 ºC for 3 min, followed by 40
453 cycles of 95 ºC for 20 s, 60 ºC for 20 s and 72 ºC for 30 s. Data were acquired at 88

ºC for 15 s to avoid quantification of primer dimers. Specificity of qPCR reactions was determined from melting curves obtained by increasing the temperature in 0.3 ºC increments from 60 ºC to 95 ºC, followed by gel electrophoresis and clone library construction from several qPCR products.

The copy number of *isoA* genes was determined from qPCR of ten-fold dilution series ($10^0$-$10^8$ copies per µl) of DNA standards. Standards were prepared by cloning the *isoA* gene of *Rhodococcus* sp. AD45 into the pGEM®T Easy vector (Promega) and using this as template DNA. The detection limit of the *isoA* qPCR assay was $10^2$ copies per 20 µl reaction.

Finally, *isoA* copies were normalised to 16S rRNA gene copy number in order to estimate the abundance of isoprene degraders in different environmental samples.

Number of copies of 16S rRNA genes was determined by qPCR using 519F and 907R primers [81]. Reactions (20 µl) contained 10-70 pg DNA, 400 nM of each primer and 10 µl of SensiFast SYBR Hi-ROX kit. The qPCR reaction consisted of an initial denaturation step at 95 ºC for 3 min, followed by 40 cycles of 95 ºC for 20 s, 55 ºC for 20 s and 72 ºC for 30 s. Data collection was performed at 72 ºC for 15 s. Specificity of the qPCR reaction and quantification of 16S rRNA gene copy number were determined as above.

## Additional files

**Additional file 1: Table S1.** Sequences of hydroxylase α-subunits of soluble diiron monooxygenases used in the design of *isoA* probes. **Table S2.** Primers used in this study targeting the *isoA* gene. **Table S3.** Combinations of *isoA* primers tested in this study. **Table S4.** Control strains used in this study to validate the *isoA* gene probes isoA14F and isoA511R. **Table S5.** Environmental samples used in this study. **Table S7.** Alpha diversity of enriched environmental samples subjected to *isoA* amplicon sequencing. **Figure S1.** Alignment of IsoA sequences from representative isoprene-degrading bacteria and position of the new *isoA* probes. **Figure S2.** Diversity and abundance of Operational Taxonomic Units obtained by amplicon sequencing from enriched environmental samples. **Figure S3.** Relative abundance of isoprene degraders in enriched environmental samples estimated by qPCR. (docx 517 kb).

**Additional file 2: Table S6.** Operational taxonomic units (OTUs) retrieved from enriched environmental samples targeting *isoA* (xlsx 23 kb).

## Acknowledgements

We thank Elizabeth Wellington and Ian Lidbury from the University of Warwick for importing soils from Malaysia to their laboratory and extracting DNA. We also thank Mohd Shahrul Mohd Nadzir from the National University of Malaysia for providing oil palm leaves and soil samples. We thank Nick Coleman, Scott Ensign and Greg Whited for providing bacterial strains to validate the *isoA* probes. We thank Paul Wilmes for adavise on *isoA* amplicon sequencing analysis. We thank Andrew Crombie for insightful comments on the manuscript and Robin Dawson, Colin Lockwood, Simone Payne and Leanne Sims for useful scientific discussions.

## Funding

The work on this project was funded through an ERC advanced grant to J.C.M. (694578—IsoMet), Natural Environment Research Council (NERC) grants to J.C.M. (NE/J009725/1) and T.J.M. (NE/J009555/1), a Colombian Government Scholarship (No. 646, Colciencias/Newton Fund (2014)) to NLLM and the Earth and Life Systems Alliance (ELSA) at the University of East Anglia.

## Availability of data and materials

*isoA* amplicon sequencing data generated in this study were deposited to the sequence read archives (SRA) under project number (submitted). Clone library sequence data were deposited to NCBI Genbank under accession numbers (submitted) to (submitted).

## Authors' contributions

OCF, TJM, NLLM and JCM planned the experiments. OCF, LG and NLLM carried out the experimental work and analysed results. MFUH planned and analysed qPCR experiments. JRG conducted *isoA* amplicon sequencing analysis. OCF and JCM wrote the manuscript with contributions from all authors. All authors read and approved the manuscript before submission.

## Ethics approval and consent to participate

Not applicable

## Consent for publication

Not applicable

## Competing interests

The authors declare that they have no competing interests.

## References


1. Atkinson R, Arey J. Gas-phase tropospheric chemistry of biogenic volatile organic compounds: A review. Atmos Environ. 2003;37,s2:197-219.
2. Guenther AB, Jiang X, Heald CL, Sakulyanontvittaya T, Duhl T, Emmons LK, Wang X. The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions. Geosci Model Dev. 2012;5:1471–92.
3. Pacifico F, Harrison SP, Jones CD, Sitch S. Isoprene emissions and climate. Atmos Environ. 2009;43:6121–35.
4. Collins WJ, Derwent RG, Johnson CE, Stevenson DS. The Oxidation of Organic Compounds in the Troposphere and their Global Warming Potentials. Clim Change. 2002;52:453–79.
5. Folberth GA, Hauglustaine DA, Lathière J, Brocheton, F. Interactive chemistry in the Laboratoire de Météorologie Dynamique general circulation model : model description and impact analysis of biogenic hydrocarbons on tropospheric chemistry. Atmos Chem Phys. 2006;6:2273–319.
6. Ashworth K, Wild O, Hewitt CN. Impacts of biofuel cultivation on mortality and crop yields. Nature Clim Change. 2013;3:492-6.
7. Carlton AG, Wiedinmyer C, Kroll JH. A review of Secondary organic aerosol (SOA) formation from isoprene. Atmos Chem Phys. 2009;9:4987–5005.
8. Fiore AM, Naik V, Spracklen DV, Steiner A, Unger N, Prather M. *et al.* Global air quality and climate. Chem Soc Rev. 2012:41;6663–83.
9. Sharkey TD, Wiberley AE, Donohue AR. Isoprene emission from plants: Why and how. Ann Bot-London. 2008;101:5–18.
10. Bäck J, Aaltonen H, Hellén H, Kajos MK, Patokoski J, Taipale R *et al.* Variable emissions of microbial volatile organic compounds (MVOCs) from root-associated fungi isolated from Scots pine. Atmos Environ 2010;44:3651–59.
11. Dani SKG, Benavides AMS, Michelozzi M, Peluso G, Torzillo G, Loreto F. Relationship between isoprene emission and photosynthesis in diatoms, and its implications for global marine isoprene estimates. Mar Chem. 2017;189:17–24.

12. Exton DA, Suggett DJ, McGenity TJ, Steinke, M. Chlorophyll-normalized isoprene production in laboratory cultures of marine microalgae and implications for global models. Limnol Oceanogr. 2013;58:1301–11.
13. Fall R, Copley SD. Bacterial sources and sinks of isoprene, a reactive atmospheric hydrocarbon. *Environ Microbiol*. 2000;2:123–30.
14. Kuzma J, Nemecek-Marshall M, Pollock WH, Fall R. Bacteria produce the volatile hydrocarbon isoprene. Curr Microbiol. 1995;30:97–103.
15. McGenity TJ, Crombie AT, Murrell JC. Microbial cycling of isoprene, the most abundantly produced biological volatile organic compound on Earth. ISME J. 2018;12:931-41.
16. Shaw SL, Gantt B, Meskhidze N. Production and emissions of marine isoprene and monoterpenes: a review. Adv Meterol. 2010:408696.
17. Steinke M, Hodapp B, Subhan R, Bell TG, Martin-Creuzburg D. Flux of the biogenic volatiles isoprene and dimethyl sulphide from an oligotrophic lake. Sci Rep. 2018;8:630.
18. Morais ARC, Dworakowska S, Reis A, Gouveia L, Matos CT, Bogdal D *et al*. Chemical and biological-based isoprene production: Green metrics. Catal Today. 2015;239:38-43.
19. Logan BA, Monson RK, Potosnak MJ. Biochemistry and physiology of foliar isoprene production. Trends Plant Sci. 2000;5:477–81.
20. Zeinali N, Altarawneh M, Li D, Al-Nu'Airat J, Dlugogorski BZ. New mechanistic insights: why do plants produce isoprene? ACS Omega. 2016;1:220–25.
21. Loivamäki M, Mumm R, Dicke M, Schnitzler JP. Isoprene interferes with the attraction of bodyguards by herbaceous plants. Proc Natl Acad Sci. 2005; 105:17430–35.
22. Sanadze GA. Biogenic isoprene emission as expression of dissipativity, a fundamental cell property. Russian Journal of Plant Physiology. 2017;64:133–40.
23. Loreto F, Ciccioli P, Brancaleoni E, Valentini R, De Lillis M, Csiky O *et al*. A hypothesis on the evolution of isoprenoid emission by oaks based on the correlation between emission type and *Quercus* taxonomoy. Oeacologia. 1998;115:17430-35.
24. Monson RK, Jones RT, Rosenstiel TN, Schnitzler JP. Why only some plants emit isoprene. Plant Cell Environ. 2013;36:503-16.
25. Sharkey TD. Is it useful to ask why plants emit isoprene? Plant Cell Environ. 2013;36:517-20.

26. Greenberg JP, Guenther A, Zimmerman P, Baugh W, Geron C, Davis K *et al*. Tethered balloon measurements of biogenic VOCs in the atmospheric boundary layer. Atmos Environ. 1999;33:855–67.
27. Wiedinmyer C, Greenberg J, Guenther A, Hopkins B, Baker K, Geron C *et al.* Ozarks Isoprene Experiment (OZIE): Measurements and modeling of the "isoprene volcano." J Geophys Res-Atmos. 2005;110:1–17.
28. Cleveland CC, Yavitt JB. Microbial consumption of atmospheric isoprene in a temperate forest soil. Appl Env Microbiol. 1998;64:172–77.
29. Gray CM, Helmig D, Fierer N. Bacteria and fungi associated with isoprene consumption in soil. Elem Sci Anth. 2015;3:000053.
30. Pegoraro E, Abrell L, Van Haren J, Barron-Gafford G, Grieve KA, Malhi Y *et al*. The effect of elevated atmospheric $CO_2$ and drought on sources and sinks of isoprene in a temperate and tropical rainforest mesocosm. Glob Change Biol. 2005;11:1234–46.
31. Acuña Alvarez LA, Exton DA, Timmis KN, Suggett DJ, McGenity TJ. Characterization of marine isoprene-degrading communities. Environ Microbiol. 2009;11:3280–91.
32. van Hylckama Vlieg JET, Kingma J, van den Wijngaard AJ, Janssen DB. A glutathione *S*-transferase with activity towards cis-1, 2-dichloroepoxyethane is involved in isoprene utilization by *Rhodococcus* sp. strain AD45. Appl Environ Microbiol. 1998;64:2800–05.
33. van Hylckama Vlieg JET, Kingma J, Kruizinga W, Janssen DB. Purification of a glutathione *S*-transferase and a conjugate-specific dehydrogenase involved in isoprene metabolism in *Rhodococcus* sp. strain AD45. J Bacteriol. 1999;181:2094-101.
34. Crombie AT, Khawand ME, Rhodius VA, Fengler KA, Miller MC, Whited GM *et al.* Regulation of plasmid-encoded isoprene metabolism in *Rhodococcus*, a representative of an important link in the global isoprene cycle. Environ Microbiol. 2015;17:3314-29.
35. Crombie AT, Emery H, McGenity TJ, Murrell JC. Draft genome sequences of three terrestrial isoprene-degrading *Rhodococcus* strains. Genome Announc. 2017;5:e01256-17.
36. El Khawand M, Crombie AT, Johnston A, Vavlline DV, McAuliffe JC, Latone JA *et al.* Isolation of isoprene degrading bacteria from soils, development of *iso*A gene probes and identification of the active isoprene-degrading soil community using DNA-stable isotope probing. Environ Microbiol; 2016;18:2743–53.

630 37. Johnston A, Crombie AT, Khawand ME, Sims L, Whited GM, McGenity TJ *et al.* Identification
631 and characterisation of isoprene-degrading bacteria in an estuarine environment. Environ
632 Microbiol. 2017;19:3526–37.
633 38. Larke-Mejía NL. Microbial ecology of isoprene degraders in the terrestrial environment. PhD
634 thesis. School of Environmental Sciences, University of East Anglia, Norwich, UK. 2018.
635 39. van Hylckama Vlieg JET, Leemhuis H, Spelberg JHL, Janssen DB. Characterisation of the
636 gene cluster involved in isoprene metabolism in *Rhodococcus* sp. strain AD45. J Bacteriol.
637 2000;182:1956-63.
638 40. Crombie AT, Mejía-Florez NL, McGenity TJ, Murrell JC. Genetics and ecology of isoprene
639 degradation. In: Rojo F, editor. Aerobic utilization of hydrocarbons, oils and lipids. Springer
640 International Publishing; 2018. p. 1-15.
641 41. Coleman NV, Bui NB, Holmes AJ. Soluble di-iron monooxygenase gene diversity in soils,
642 sediments and ethene enrichments. Environ Microbiol. 2006; 8:1228–39.
643 42. Leahy JG, Batchelor PJ, Morcomb SM. Evolution of the soluble diiron monooxygenases. FEMS
644 Microbiol Rev. 2003;27:449-479.
645 43. Dumont MG, Murrell JC. Community-level analysis: key genes of aerobic methane oxidation.
646 Methods Enzymol. 2005;397:413-427.
647 44. McDonald IR, Kenna EM, Murrell JC. Detection of methanotrophic bacteria in environmental
648 samples with the PCR. Appl Environ Microbiol. 1995;61:116-21.
649 45. McDonald IR, Bodrossy L, Chen Y, Murrell JC. Molecular ecology techniques for the study of
650 aerobic methanotrophs. Appl Environ Microbiol. 2008;74:1305–15.
651 46. Farhan Ul Haque M, Crombie AT, Ensminger SA, Baciu C, Murrell JC. Facultative
652 methanotrophs are abundant at terrestrial natural gas seeps. Microbiome. 2018;6:118.
653 47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol
654 Biol. 1990;215:403-10.
655 48. Lee J, Lee CS, Hugunin KM, Maute CJ, Dysko RC. Bacteria from drinking water supply and
656 their fate in gastrointestinal tracts of germ-free mice: a phylogenetic comparison study. Water
657 Res. 2010;44:5050-8.
658 49. Willems A, Mergaert J, Swings J. Genus X. *Variovorax*. In: Brenner DJ, Krieg NR, Staley JT,
659 Garrity GM, editors. Bergey's manual of systematic bacteriology. 2nd edition, volume 2: the

660 Proteobacteria. Part C: the Alpha-, Beta-, Delta and Epsilonproteobacteria. New York: Springer;
661 2005. p. 732-5.
662 50. Woo HL, Hazen TC, Simmons BA, DeAngelis KM. Enzyme activities of aerobic lignocellulolytic
663 bacteria isolated from wet tropical forest soils. Syst Appl Microbiol. 2014;37:60-7.
664 51. Lambais MR, Crowley DE, Cury JC, Büll RC, Rodrigues RR. Bacterial diversity in tree canopies
665 of the Atlantic forest. Science. 2006;312:1917.
666 52. Vorholt JA. Microbial life in the phyllosphere. Nat Rev Microbiol. 2012;10:828-40.
667 53. Bodenhausen N, Horton MW, Bergelson J. Bacterial communities associated with the leaves
668 and roots of *Arabidopsis thaliana*. PLos One. 2013;8:e56329.
669 54. Han JI, Choi HK, Lee SW, Orwin PM, Kim J, Laroe SL *et al.* Complete genome sequence of
670 the metabolically versatile plant growth-promoting endophyte *Variovorax paradoxus* S110. J
671 Bacteriol. 2011;193:1183-90.
672 55. Satola B, Wübbeler JH, Steinbüchel A. Metabolic characteristics of the species *Variovorax*
673 *paradoxus*. Appl Microbiol Biotechnol. 2013;97:541-60.
674 56. Baik KS, Choe HN, Park SC, Hwang YM, Kim EM, Park C *et al. Sphingopyxis rigui* sp. nov. and
675 *Sphingopyxis wooponensis* sp. nov., isolated from wetland freshwater, and emended
676 description of the genus *Sphingopyxis*. Int J Syst Evol Microbiol. 2013;63:1297-1303.
677 57. Choi JH, Kim MS, Jung MJ, Rosh SW, Shin KD, Bae JW. *Sphingopyxis soli* sp. nov., isolated
678 from landfill soil. Int J Syst Evol Microbiol. 2010;60:1682-6.
679 58. Kim BS, Lim YW, Chun J. *Sphingopyxis marina* sp. nov. and *Sphingopyxis litoralis* sp. nov.,
680 isolated from seawater. Int J Syst Evol Microbiol. 2008;58:2415-9.
681 59. Srinivasan S, Kim MK, Sathiyaraj G, Veena V, Mahalakshmi M, Kalaiselvi S *et al*. *Sphingopyxis*
682 *panaciterrulae* sp. nov., isolated from a ginseng field. Int J Syst Evol Microbiol. 2010;60:2358-
683 63.
684 60. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: high-
685 resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581-3.
686 61. Beckers B, Op De Beeck M, Weyens N, Boerjan W, Vangronsveld J. Structural variability and
687 niche differentiation in the rhizosphere and endosphere bacterial microbiome of field-grown
688 poplar trees. Microbiome. 2017;5:25.

689 62. Guenther A, Zimmerman P, Wildermuth M. Natural volatile organic compound emission rates
690 estimates for US woodland landscapes. Atmos Environ. 1994;28-1197-1210.
691 63. Kesselmeier J, Staudt M. Biogenic volatile organic compounds (VOC): An overview on
692 emission, physiology and ecology. J Atmos Chem. 1999;33:23-88.
693 64. Bittar TB, Pound P, Whitetree A, Moore LD, Van Stan JT. Estimation of throughfall and stemflow
694 bacterial flux in a subtropical oak-cedar forest. Geophys Res Lett. 2018;45:1410–18.
695 65. Ciullo PA, Hewitt N. The rubber formulary. Elsevier Science; 1999.
696 66. Shuttleworth MJ, Watson AA. Synthetic polyisoprene rubbers. In: Whelan A, Lee KS, editors.
697 Developments in rubber technology-2: synthetic rubbers. Dordrecht: Springer Netherlands;
698 1981. p. 233–67.
699 67. Linos A, Steinbüchel A. Biodegradation of natural and synthetic rubbers. In: Koyama T,
700 Steinbüchel A, editors. Biopolymers – Biology, Chemistry, Biotechnology, Applications, vol. 2
701 (Polyisoprenoids), 1st ed. Weinhem: Wiley-VCH; 2001. p. 321-59.
702 68. Shah AA, Hasan F, Shah N, Kanwal N, Zeb S. Biodegradation of natural and synthetic rubbers:
703 a review. Int Biodeter Biodegr. 2013;83:145-157.
704 69. Andler R, Hiessl S, Yücel O, Tesch M. Steinbüchel A. Cleavage of poly(*cis*-1,4-isoprene) rubber
705 as solid substrate by culture of *Gordonia polyisoprenivorans*. N Biotechnol. 2018;44:6-12.
706 70. Imai S, Ichikawa K, Muramatsu Y, Kasai D, Masai E, Fukuda M. Isolation and characterisation
707 of *Streptomyces, Actinoplanes* and *Methylibium* strains that are involved in degradation of
708 natural rubber and synthetic poly(*cis*-1,4-isoprene). Enzyme Microb Technol. 2011;49:526-31.
709 71. Jendrossek D, Tomasi G, Kroppenstedt RM. Bacterial degradation of natural rubber: a privilege
710 of actinomycetes? FEMS Microbiol Lett. 1997;150:179-88.
711 72. Linos A, Berekaa MM, Reichelt R, Keller U, Schmitt J, Flemming HS *et al*. Biodegradation of
712 *cis*-1,4-polyisoprene rubbers by distinct actinomycetes: microbial strategies and detailed
713 surface analysis. Appl Environ Microbiol. 2000;66:1639-45.
714 73. Bode HB, Zeeck A, Pluckhahn K, Jendrossek D. Physiological and chemical investigations into
715 microbial degradation of synthetic poly(*cis*-1,4-isoprene). Appl Environ Microbiol.
716 2000;66:3680-5.

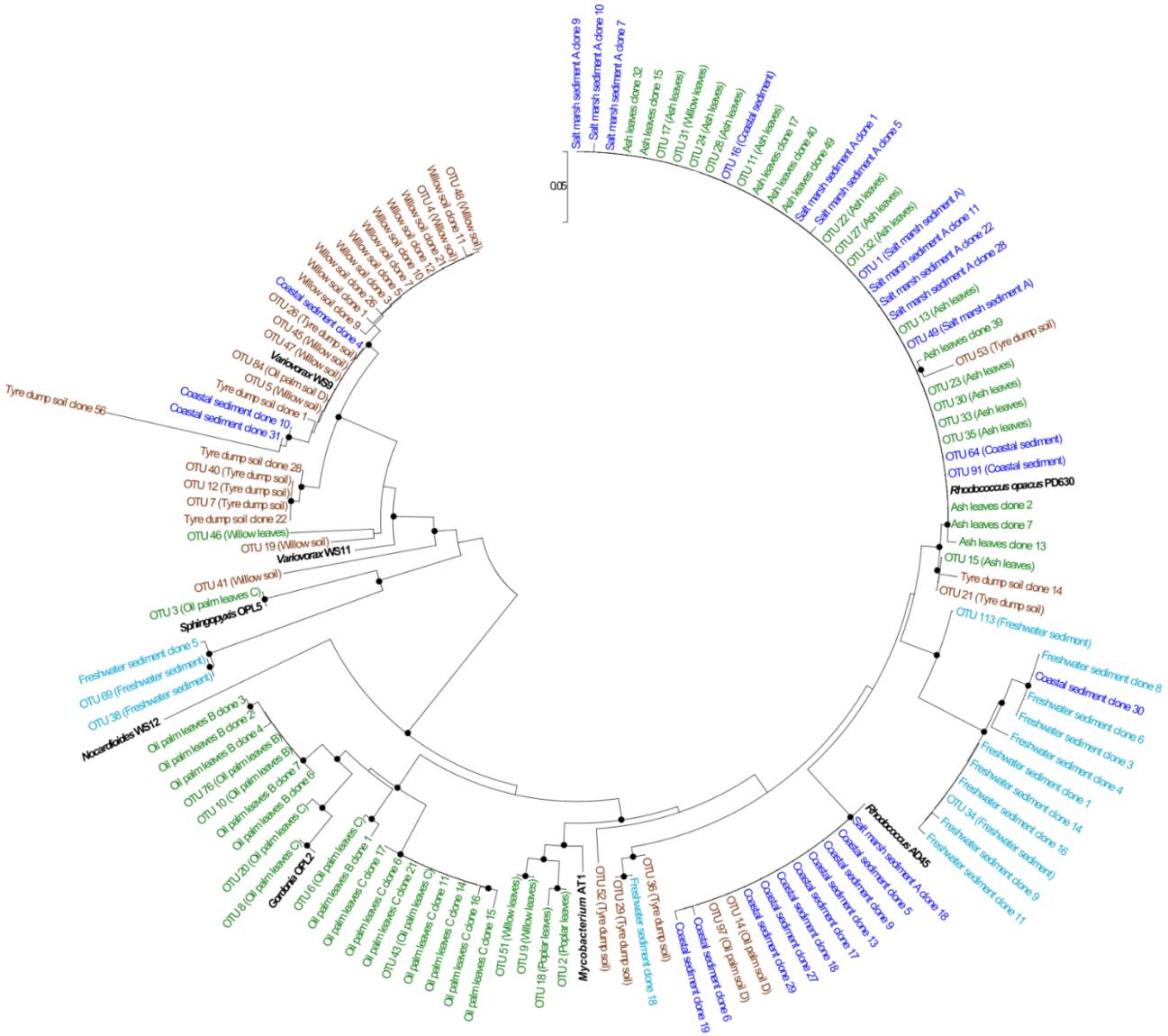
74. Srivastva N, Shukla AK, Singh RS, Upadhyay SN, Dubey SK. Characterisation of bacterial isolates from rubber dump site and their use in biodegradation of isoprene in batch and continuous bioreactors. Bioresour Technol. 2015;188:84-91.
75. Booge D, Schlundt C, Bracher A, Endres S, Zäncker B, Marandino CA. Marine isoprene production and consumption in the mixed layer of the surface ocean – a field study over two oceanic regions. Biogeosciences. 2018;15:649-667.
76. Booge D, Marandino C, Schlundt C, Palmer PI, Saltzman ES, Wallace DW. Can simple models predict large scale surface ocean isoprene concentrations? Atmos Chem Phys. 2016;16:11807-21.
77. Moore RM, Wang L. The influence of iron fertilisation on the fluxes of methyl halides and isoprene from ocean to atmosphere in the SERIES experiment. Deep-Sea Res Pt II. 2006;53:2938-409.
78. Palmer PI, Shaw SL. Quantifying global marine isoprene fluxes using MODIS chlorophyll observations. Geophys Res Lett. 2005;32:L09805.
79. Dorn E, Hellwig M, Reineke W, Knackmuss HJ. Isolation and characterization of a 3-chlorobenzoate degrading pseudomonad. Arch Microbiol. 1974;99:61–70.
80. Baumann P, Baumann L. The marine Gram-negative Eubacteria: Genera *Photobacterium, Beneckea, Alteromonas, Pseudomonas and Alcaligenes*. In: Starr MP, Stolp H, Trüper HG, Balows A, Schlegel HG, editors. The Prokaryotes: A Handbook on Habitats, Isolation and Identification of Bacteria. New York: Springer-Verlag; 1981. p. 1302–31.
81. Lane DJ. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, editors. Nucleic acid techniques in bacterial systematics. Chichester: John Wiley & Sons; 1991. p. 115-75.
82. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33:1870–4.

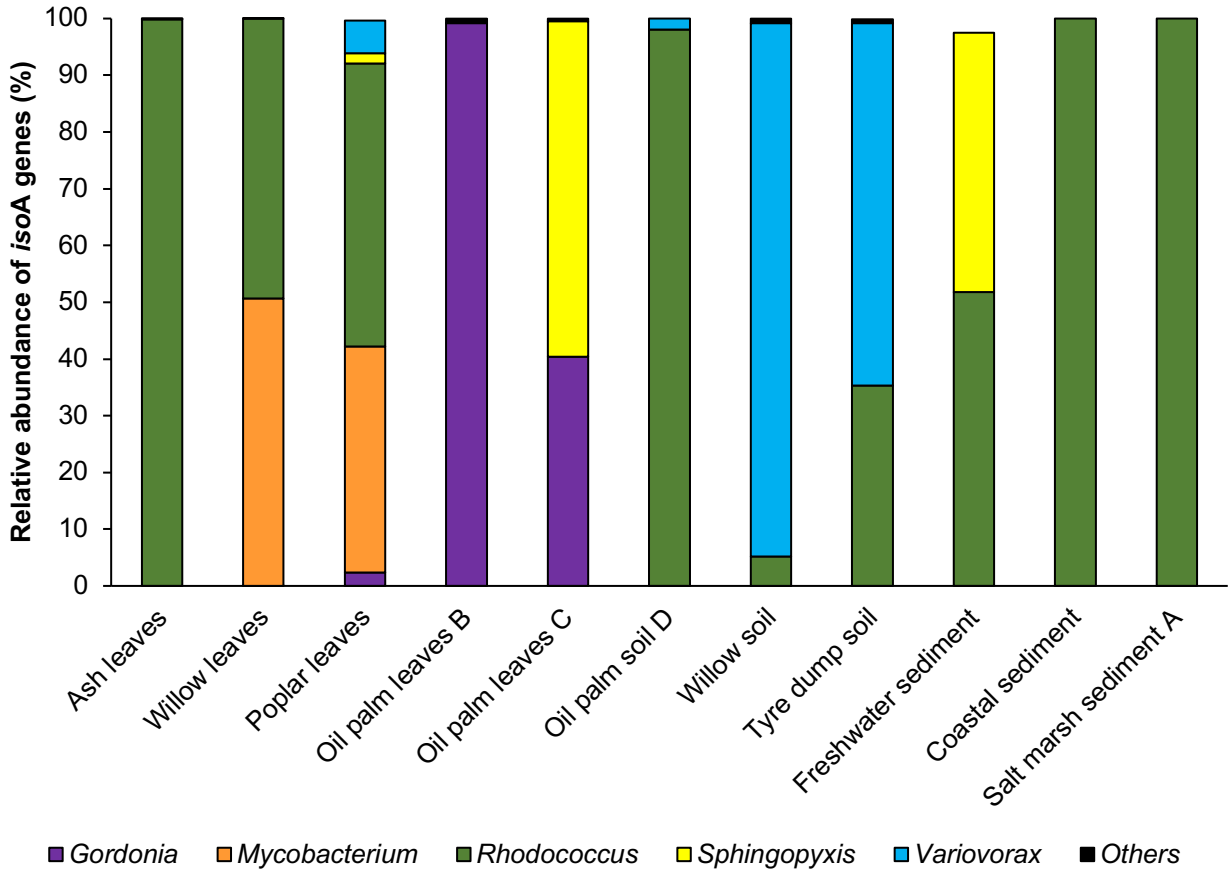
## Figure legends

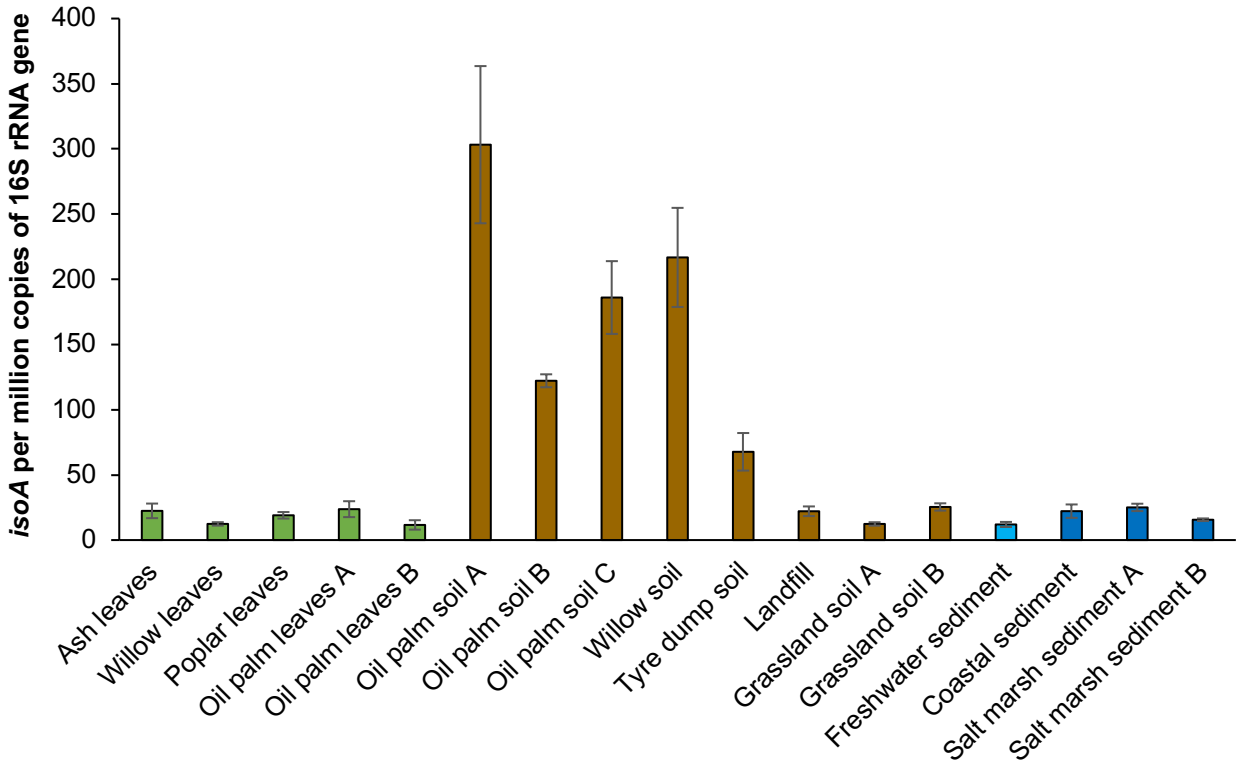
**Figure 1. Phylogenetic tree of IsoA sequences retrieved from enriched environmental samples.** Sequences in black represent IsoA sequences from bona-fide isoprene degraders. Sequences obtained from clone libraries and amplicon sequencing of leaf samples are represented in green, soils in brown, freshwater environments in light blue and marine sediments in dark blue. Environments where a particular OTU is abundant are shown in brackets. The tree was drawn in Mega7 [82] using the neighbour-joining method and the Jones-Taylor-Thornton model. Scale bar indicates 0.05 substitutions per site. Bootstrap values ≥50% (based on 1,000 replicates) are represented with dots at branch points.

**Figure 2. Relative abundance and diversity of *isoA* genes in enriched environmental samples revealed by amplicon sequencing.** *isoA* amplicon yielded an average of 12,453 quality-filtered reads per sample. After analysis with the DADA2 pipeline [60] a final set of 134 unique OTUs was obtained (Additional File 2: Table S6). Only OTUs with ≥1% abundance in at least one of the samples are represented. For relative abundance of individual OTUs in each enriched environmental sample, see Additional File 1: Figure S2. For α-diversity of enriched environmental samples estimated using Shannon index, see Additional File 1: Table S7.

**Figure 3. Relative abundance of isoprene degraders in natural (non-enriched) environmental samples estimated by qPCR.** *isoA* copies are normalised to the 16S rRNA gene copy number in each sample. Results shown are the average of triplicate samples. Errors bars represent standard deviations. Leaf samples are represented in green, soils in brown, freshwater environments in light blue and marine sediments in dark blue.


0.05
Salt marsh sediment A clone 9
Salt marsh sediment A clone 10
Salt marsh sediment A clone 7
Ash leaves clone 32
Ash leaves clone 15
OTU 17 (Ash leaves)
OTU 31 (Willow leaves)
OTU 24 (Ash leaves)
OTU 28 (Ash leaves)
OTU 16 (Coastal sediment)
OTU 11 (Ash leaves)
Ash leaves clone 17
Ash leaves clone 40
Ash leaves clone 49
Salt marsh sediment A clone 1
Salt marsh sediment A clone 5
OTU 22 (Ash leaves)
OTU 27 (Ash leaves)
OTU 32 (Ash leaves)
OTU 1 (Salt marsh sediment A)
Salt marsh sediment A clone 11
Salt marsh sediment A clone 22
Salt marsh sediment A clone 28
OTU 13 (Ash leaves)
OTU 49 (Salt marsh sediment A)
Ash leaves clone 39
OTU 53 (Tyre dump soil)
OTU 23 (Ash leaves)
OTU 30 (Ash leaves)
OTU 33 (Ash leaves)
OTU 35 (Ash leaves)
OTU 64 (Coastal sediment)
OTU 91 (Coastal sediment)
Rhodococcus opacus PD630
Ash leaves clone 2
Ash leaves clone 7
Ash leaves clone 13
OTU 15 (Ash leaves)
Tyre dump soil clone 14
OTU 21 (Tyre dump soil)
OTU 113 (Freshwater sediment)
Freshwater sediment clone 8
Coastal sediment clone 30
Freshwater sediment clone 6
Freshwater sediment clone 3
Freshwater sediment clone 4
Freshwater sediment clone 1
Freshwater sediment clone 14
Freshwater sediment clone 16
OTU 34 (Freshwater sediment)
Freshwater sediment clone 9
Freshwater sediment clone 11
Rhodococcus AD45
Salt marsh sediment A clone 18
Coastal sediment clone 5
Coastal sediment clone 9
Coastal sediment clone 13
Coastal sediment clone 17
Coastal sediment clone 18
Coastal sediment clone 27
Coastal sediment clone 29
OTU 14 (Oil palm soil D)
OTU 97 (Oil palm soil D)
Coastal sediment clone 6
Coastal sediment clone 19
OTU 36 (Tyre dump soil)
Freshwater sediment clone 18
OTU 29 (Tyre dump soil)
OTU 52 (Tyre dump soil)
Mycobacterium AT1
OTU 2 (Poplar leaves)
OTU 18 (Poplar leaves)
OTU 9 (Willow leaves)
OTU 51 (Willow leaves)
Oil palm leaves C clone 15
Oil palm leaves C clone 16
Oil palm leaves C clone 14
Oil palm leaves C clone 11
Oil palm leaves C clone 21
OTU 43 (Oil palm leaves C)
Oil palm leaves C clone 6
Oil palm leaves C clone 17
Oil palm leaves B clone 1
OTU 6 (Oil palm leaves C)
Gordonia OPL2
OTU 8 (Oil palm leaves C)
OTU 20 (Oil palm leaves C)
Oil palm leaves B clone 6
Oil palm leaves B clone 7
OTU 10 (Oil palm leaves B)
OTU 76 (Oil palm leaves B)
Oil palm leaves B clone 4
Oil palm leaves B clone 2
Oil palm leaves B clone 3
Nocardioides WS12
OTU 38 (Freshwater sediment)
OTU 69 (Freshwater sediment)
Freshwater sediment clone 5
Sphingopyxis OPL5
OTU 3 (Oil palm leaves C)
OTU 41 (Willow soil)
Variovorax WS11
OTU 19 (Willow soil)
OTU 46 (Willow leaves)
Tyre dump soil clone 22
OTU 7 (Tyre dump soil)
OTU 12 (Tyre dump soil)
OTU 40 (Tyre dump soil)
Tyre dump soil clone 28
Coastal sediment clone 31
Coastal sediment clone 10
Tyre dump soil clone 1
OTU 5 (Willow soil)
OTU 84 (Oil palm soil D)
Variovorax WS9
OTU 47 (Willow soil)
OTU 45 (Willow soil)
OTU 28 (Tyre dump soil)
Coastal sediment clone 4
Willow soil clone 9
Willow soil clone 1
Willow soil clone 26
Willow soil clone 3
Willow soil clone 7
Willow soil clone 10
Willow soil clone 12
Willow soil clone 21
OTU 4 (Willow soil)
Willow soil clone 11
OTU 48 (Willow soil)
Tyre dump soil clone 56


Relative abundance of isoA genes (%)
100
90
80
70
60
50
40
30
20
10
0
Ash leaves
Willow leaves
Poplar leaves
Oil palm leaves B
Oil palm leaves C
Oil palm soil D
Willow soil
Tyre dump soil
Freshwater sediment
Coastal sediment
Salt marsh sediment A
Gordonia
Mycobacterium
Rhodococcus
Sphingopyxis
Variovorax
Others


isoA per million copies of 16S rRNA gene
400
350
300
250
200
150
100
50
0
Ash leaves
Willow leaves
Poplar leaves
Oil palm leaves A
Oil palm leaves B
Oil palm soil A
Oil palm soil B
Oil palm soil C
Willow soil
Tyre dump soil
Landfill
Grassland soil A
Grassland soil B
Freshwater sediment
Coastal sediment
Salt marsh sediment A
Salt marsh sediment B

# Additional File 1

## Functional gene probing reveals the widespread distribution, diversity and abundance of isoprene-degrading bacteria in the environment.

Ornella Carrión[1], Nasmille L. Larke-Mejía[1], Lisa Gibson[1], Muhammad Farhan Ul Haque[1], Javier Ramiro-García[2], Terry J. McGenity[3], J. Colin Murrell[1*]

[1]School of Environmental Sciences, University of East Anglia, Norwich, UK

[2]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[3]School of Biological Sciences, University of Essex, Colchester, UK

[*]Corresponding author:

J Colin Murrell, [1]School of Environmental Sciences, University of East Anglia

Norwich Research Park, NR4 7TJ, UK

E-mail: j.c.murrell@uea.ac.uk

Tel: (+44) 01603 592959

Fax: (+44) 01603 591327

**Table S1. Sequences of hydroxylase α-subunits of soluble diiron monooxygenases used in the design of *isoA* probes.**

| Microorganism | Enzyme | Genbank Accession Number |
|---|---|---|
| *Gordonia* sp. i37 | IsoMO | KU870746.1 |
| *Gordonia* sp. OPL2 | IsoMO | Submitted |
| *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 | Submitted |
| *Rhodococcus* sp. ACPA1 | IsoMO | NSDX01000002.1 |
| *Rhodococcus* sp. ACPA4 | IsoMO | NZ_NSDY01000003.1 |
| *Rhodococcus* sp. ACS1 | IsoMO | NZ_NSDZ01000001.1 |
| *Rhodococcus* sp. ACS2 | IsoMO | Submitted |
| *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 | Submitted |
| *Rhodococcus* sp. TD2 | IsoMO | Submitted |
| *Rhodococcus* sp. TD3 | IsoMO | Submitted |
| *Rhodococcus* sp. WL1 | IsoMO | Submitted |
| *Rhodococcus* sp. WS1 | IsoMO | Submitted |
| *Rhodococcus* sp. WS2 | IsoMO | Submitted |
| *Rhodococcus* sp. WS3 | IsoMO | Submitted |
| *Rhodococcus* sp. WS4 | IsoMO | Submitted |

| | | |
|---|---|---|
| *Rhodococcus* sp. WS5 | IsoMO | Submitted |
| *Rhodococcus* sp. WS6 | IsoMO | Submitted |
| *Rhodococcus* sp. WS7 | IsoMO | Submitted |
| *Rhodococcus* sp. WS10 | IsoMO | Submitted |
| *Rhodococcus* sp. SK2ab | IsoMO | Submitted |
| *Rhodococcus* sp. SK5 | IsoMO | Submitted |
| *Rhodococcus erythropolis* i47 | IsoMO | KU870742.1 |
| *Rhodococcus opacus* PD630 | IsoMO | NZ_JH377098.1 |
| *Shinella* sp. i39 | IsoMO | KU870741.1 |
| *Sphingopyxis* sp. OPL5 | IsoMO | Submitted |
| *Stappia* sp. iL42 | IsoMO | KU870740.1 |
| *Variovorax* sp. WS9 | IsoMO | Submitted |
| *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 |

MO: monooxygenase. SDIMO: soluble diiron monooxygenase.

**Table S2. Primers used in this study targeting the *isoA* gene.**

| Primer | Sequence (5'-3')* | Nucleotide position respect to *isoA* from *Rhodococcus* sp. AD45 |
|---|---|---|
| isoA14F | GVGACGAYTGGTAYGACA | 14 |
| isoA136F | TGGGABGAACCBTTCCGSGT | 136 |
| isoA300F | CATGGTCGARCABATGGC | 300 |
| isoA379F | GTBTTCGGVATGCTCGACGA | 379 |
| isoA511F | GTVAAGAAYTTCTTYGACGA | 511 |
| isoA511R | TCGTCRAAGAARTTCTTBAC | 511 |
| isoA862R | TCSAKCATGAAYTCCTTGAA | 862 |
| isoA1019R | GCRTTBGGBTTCCAGAACA | 1019 |

*Equimolar mixtures at degenerate positions: B (C, G, T); K (G, T); R (A, G); S (C, G); V (G,A,C); Y (C, T).

**Table S3. Combinations of *isoA* primers tested in this study.**

| Combination | Amplicon size (bp) | Amplification from *Rhodococcus* sp. AD45 DNA | Amplification from *Variovorax* sp. WS9 DNA | Amplification from *Xanthobacter autotrophicus* PY2 DNA |
|---|---|---|---|---|
| isoA14F+isoA511R | 497 | + | + | - |
| isoA14F+isoA862R | 848 | + | + | + |
| isoA14F+isoA1019R | 1005 | + | + | + |
| isoA136F+isoA511R | 375 | + | - | - |
| isoA136F+isoA862R | 726 | + | - | + |
| isoA136F+isoA1019R | 883 | + | + | - |
| isoA300F+ isoA862R | 562 | - | + | - |
| isoA300F+isoA1019R | 719 | + | + | - |
| isoA379F+ isoA862R | 483 | + | + | - |
| isoA379F+isoA1019R | 640 | + | + | - |
| isoA511F+isoA1019R | 485 | - | - | + |

**Table S4. Control strains used in this study to validate the *isoA* gene probes isoA14F and isoA511R**.

| Strain | Control | Enzyme | Reference | Amplification with *iso*A primers |
|---|---|---|---|---|
| *Rhodococcus* sp. AD45 | Positive | Isoprene MO | [32] | + |
| *Rhodococcus* sp. WS1 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. WS2 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. WS3 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. WS4 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. WS5 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. WS6 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. WS7 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. WS8 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. WS10 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. TD1 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. TD2 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. TD3 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. WL1 | Positive | Isoprene MO | [38] | + |
| *Rhodococcus* sp. i47 | Positive | Isoprene MO | [83] | + |
| *Rhodococcus* sp. LB1 | Positive | Isoprene MO, Propane MO | [36] | + |
| *Rhodococcus* sp. SC4 | Positive | Isoprene MO, Propane MO | [36] | + |
| *Rhodococcus opacus* PD630 | Positive | Isoprene MO, Propane MO | [34] | + |
| *Rhodococcus* sp. ACPA1 | Positive | Isoprene MO | [35] | + |
| *Rhodococcus* sp. ACPA4 | Positive | Isoprene MO | [35] | + |
| *Rhodococcus* sp. ACS1 | Positive | Isoprene MO | [35] | + |
| *Rhodococcus* sp. ACS2 | Positive | Isoprene MO | Unpublished | + |
| *Rhodococcus* sp. SK2ab | Positive | Isoprene MO | Unpublished | + |
| *Rhodococcus* sp. SK5 | Positive | Isoprene MO | Unpublished | + |
| *Gordonia* sp. i37 | Positive | Isoprene MO, Propane MO | [31, 37] | + |
| *Gordonia* sp. OPL2 | Positive | Isoprene MO | [38] | + |
| *Nocardioides* sp. WS12 | Positive | Isoprene MO | [38] | + |
| *Variovorax* sp. WS9 | Positive | Isoprene MO | [38] | + |
| *Variovorax* sp. WS11 | Positive | Isoprene MO | [38] | + |
| *Sphingopyxis* sp. OPL5 | Positive | Isoprene MO | [38] | + |
| *Xanthobacter autotrophicus* PY2 | Negative | Alkene MO | [84] | - |
| *Methylococcus capsulatus* Bath | Negative | Methane MO | [85] | - |
| *Methylocella silvestris* BL2 | Negative | Methane MO, Propane MO | [86, 87] | - |
| *Mycobacterium* sp. NBB4 | Negative | Ethene MO, Propene MO, Group 3 SDIMO, Group 6 SDIMO. | [88, 89] | - |
| *Pseudomonas mendocina* KR1 | Negative | Toluene MO | [90] | - |
| *Rhodococcus jostii* RHA1 | Negative | Propane MO | [91] | - |
| *Rhodococcus opacus* DSM 1069 | Negative | Unknown | [92] | - |
| *Rhodococcus rhodochrous* B276 | Negative | Alkene MO | [93] | - |
| *Rhodococcus rhodochrous* PNKb1 | Negative | Alkene MO | [94] | - |
| *Rhodococcus erythropolis* JCM 3201 | Negative | Alkane MO | [95] | - |
| *Pseudomonas putida* ML2 | Negative | Benzene dioxygenase | [96] | - |
| *Rhodococcus aetherivorans* I24 | Negative | Toluene dioxygenase | [97] | - |
| *Rhodococcus rhodochrous* DSM 43241 | Negative | Alkane MO | [98] | - |
| *Variovorax paradoxus* EPS | Negative | Alkanesulfonate MO | [55]; unpublished data | - |

MO: monooxygenase; SDIMO: soluble diiron monooxygenase.

**Table S5. Environmental samples used in this study.**

| Material | Type of sample | Sampling site | Location | Analysis |
|---|---|---|---|---|
| Ash leaves | Natural and enriched | University of East Anglia | Norwich, UK | Clone library<br>*iso*A amplicon sequencing<br>qPCR |
| Poplar leaves | Natural and enriched | University of East Anglia | Norwich, UK | *iso*A amplicon sequencing<br>qPCR |
| Willow leaves | Natural and enriched | University of East Anglia | Norwich, UK | Clone library<br>*iso*A amplicon sequencing<br>qPCR |
| Oil palm leaves A | Natural | Tawau | Sabah, Malaysia | qPCR |
| Oil palm leaves B | Natural and enriched | Sepang | Selangor, Malaysia | Clone library<br>*isoA* amplicon sequencing<br>*qPCR* |
| Oil palm leaves C | Enriched | Kew Gardens | London, UK | Clone library<br>*iso*A amplicon sequencing |
| Oil palm soil A | Natural | Sepang | Selangor, Malaysia | qPCR |
| Oil palm soil B | Natural | Sepang | Selangor, Malaysia | qPCR |
| Oil palm soil C | Natural | Sepang | Selangor, Malaysia | qPCR |
| Oil palm soil D | Enriched | Kew Gardens | London, UK | *iso*A amplicon sequencing |
| Willow soil | Natural and enriched | University of East Anglia | Norwich, UK | Clone library<br>*isoA* amplicon sequencing<br>qPCR |
| Tyre dump soil | Natural and enriched | Industrial park | Fakenham, UK | Clone library<br>*iso*A amplicon sequencing<br>qPCR |
| Landfill soil | Natural | Landfill | Strumpshaw, UK | qPCR |
| Grassland soil A | Natural | Bowthorpe | Norwich, UK | qPCR |
| Grassland soil B | Natural | University of East Anglia | Norwich, UK | qPCR |
| Coastal sediment | Natural and enriched | Penarth beach | Penarth, UK | Clone library<br>*isoA* amplicon sequencing<br>qPCR |
| Salt marsh sediment A | Natural and enriched | Stiffkey salt marsh | Stiffkey, UK | Clone library<br>*isoA* amplicon sequencing<br>qPCR |
| Salt marsh sediment B | Natural | Warham salt marsh | Warham, UK | qPCR |
| Freshwater sediment | Natural and enriched | University of East Anglia lake | Norwich, UK | Clone library<br>*isoA* amplicon sequencing<br>qPCR |

**Table S7. Alpha diversity of enriched environmental samples subjected to *isoA* amplicon sequencing.** Shannon index for each enrichment was calculated using the packages phyloseq [99], ggplot2 [100] and tidyverse [101] included in R 3.4.4 [102].

| Enriched environmental sample | Shannon index |
|---|---|
| Ash leaves | 1.65 |
| Willow leaves | 1.05 |
| Poplar leaves | 1.49 |
| Oil palm leaves B | 0.25 |
| Oil palm leaves C | 1.19 |
| Oil palm soil D | 0.63 |
| Willow soil | 1.83 |
| Tyre dump soil | 2.11 |
| Freshwater sediment | 1.14 |
| Coastal sediment | 0.43 |
| Salt marsh sediment A | 0.18 |

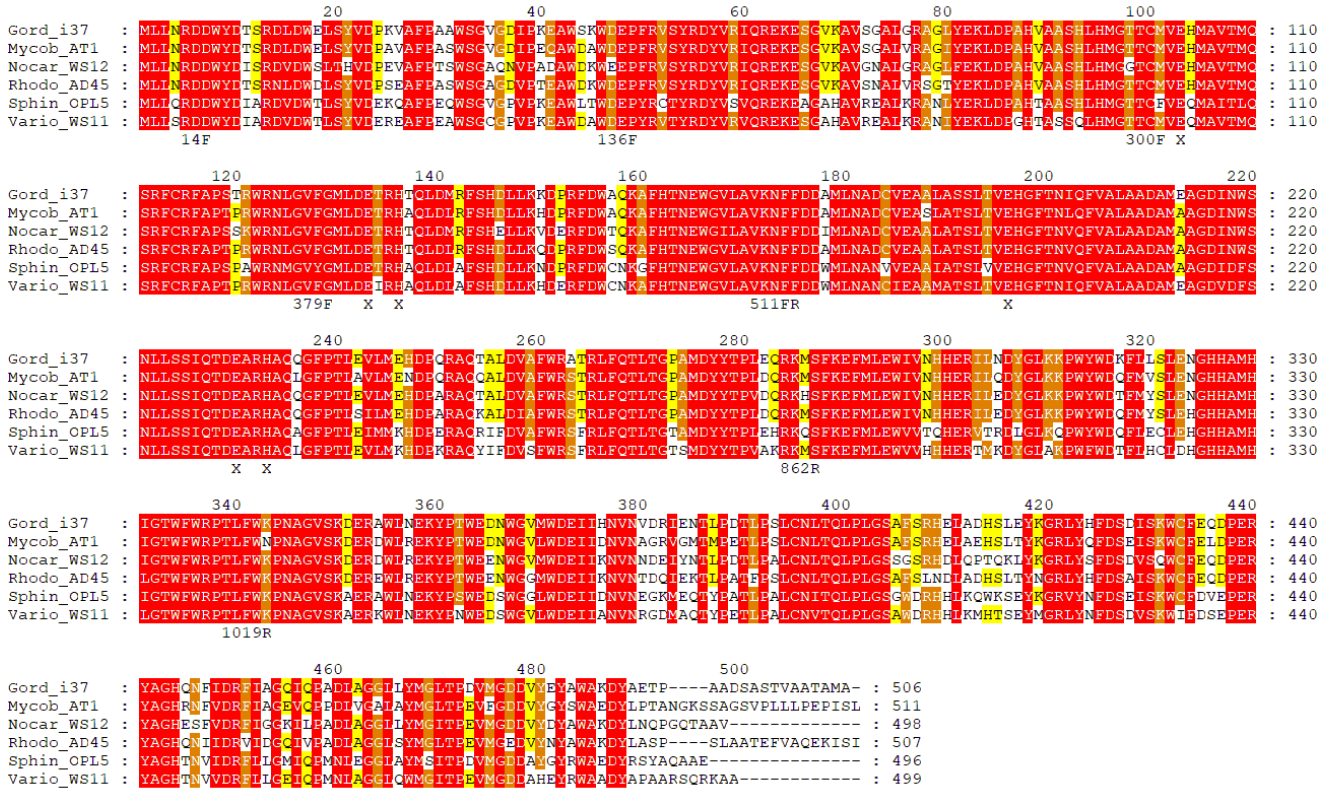


**Figure S1. Alignment of IsoA sequences from representative isoprene-degrading bacteria and position of the new *isoA* probes.** Alignment of IsoA sequences was done using the ClustalW package included in BioEdit Sequence Alignment Editor v7.2.6 [103]. Conserved domains were analysed using GeneDoc v2.5.010 [104]. 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 Additional File 1; Table S1.

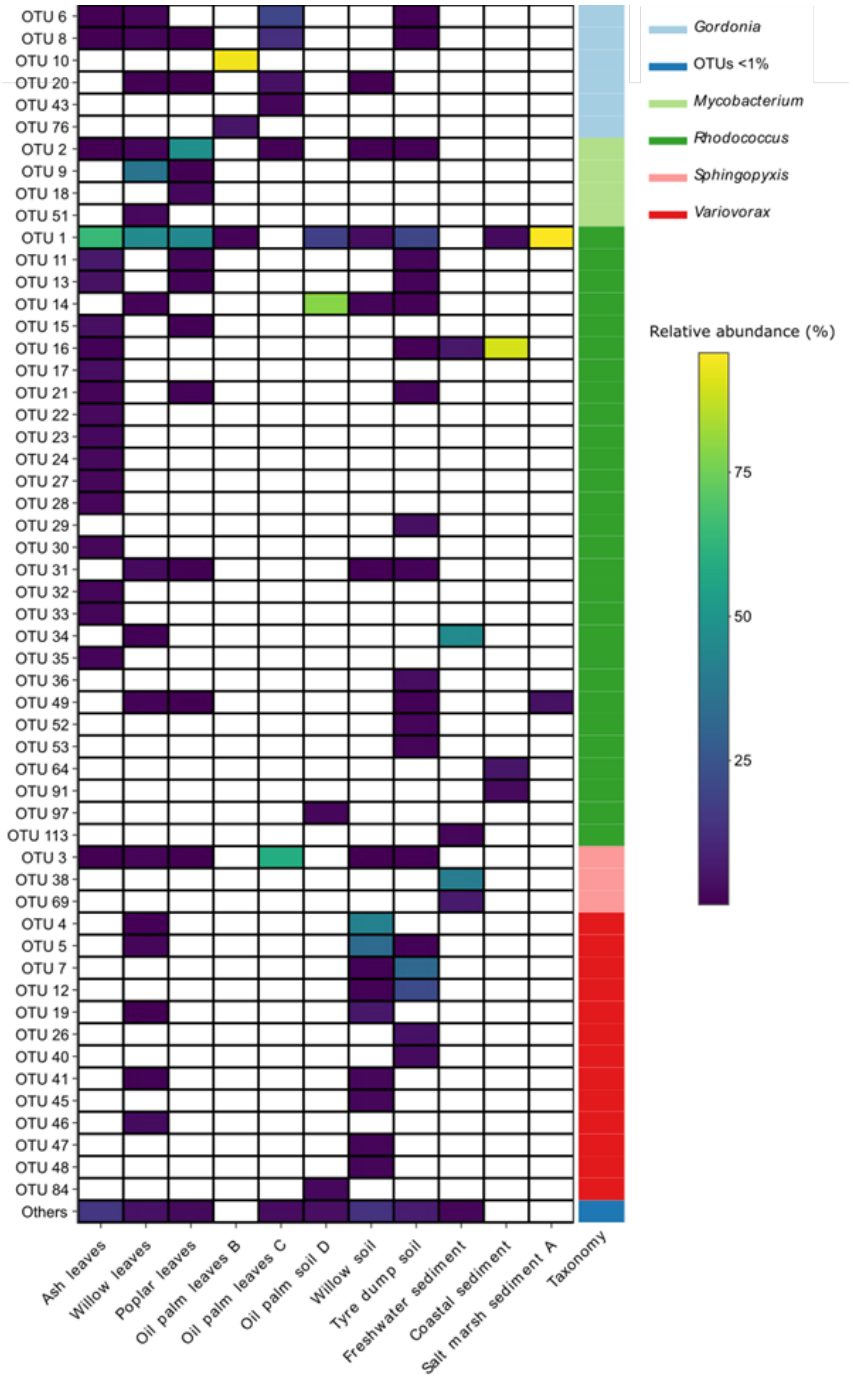



**Figure S2. Diversity and abundance of Operational Taxonomic Units obtained by amplicon sequencing from enriched environmental samples.** Analysis of *isoA* sequences from enriched environmental samples with the DADA2 pipeline [60] yielded a final set of 134 Operational Taxonomic Units (OTUs). Only OTUs with ≥1% relative abundance in at least one sample are represented. Taxonomy column indicates the genus of bona-fide isoprene degraders phylogenetically closer to a particular OTU. The heatmap was constructed using the packages plotly [105], heatmaply [106], ggplot2 [100] and tidyverse [101] included in R 3.4.4 [102].

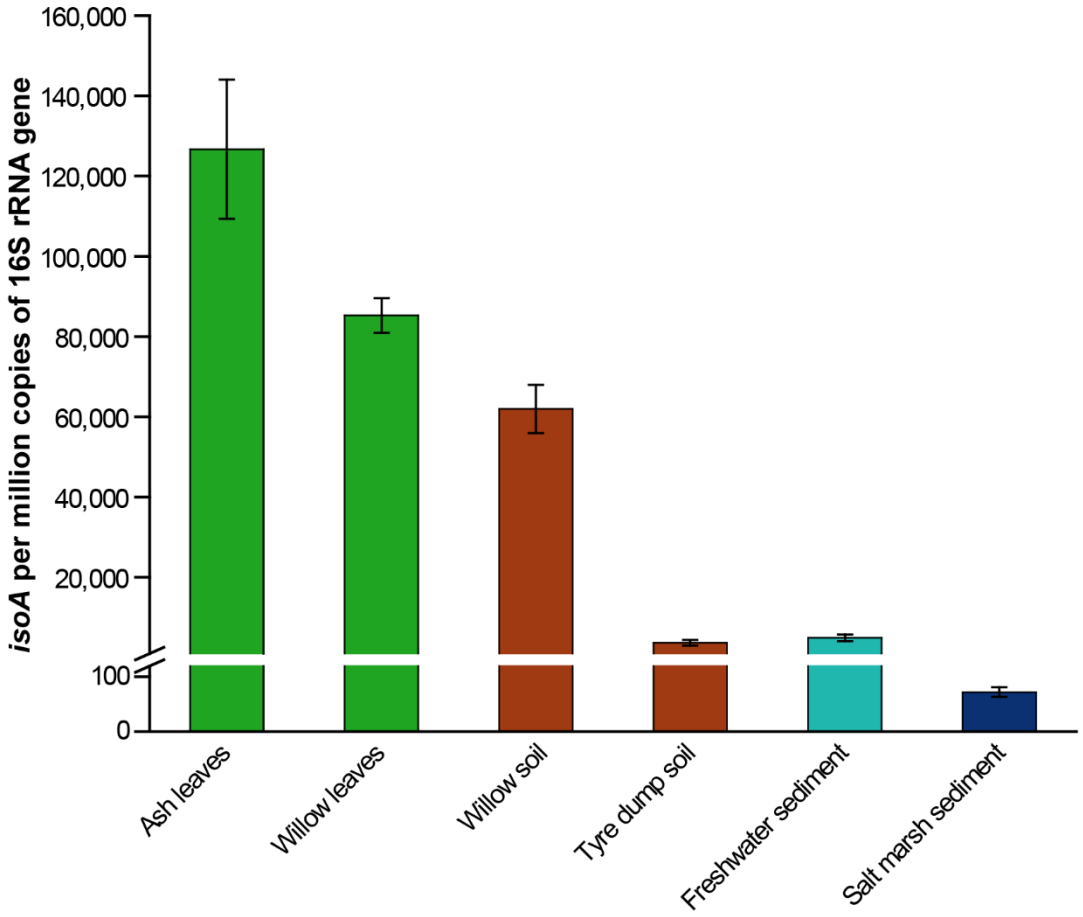



**Figure S3. Relative abundance of isoprene degraders in enriched environmental samples estimated by qPCR.** Number of *isoA* genes are normalised to copies of 16S rRNA gene in each sample. Results shown are the average of triplicate samples. Errors bars represent standard deviations. Leaf samples are represented in green, soils in brown, freshwater environments in light blue and marine sediments in dark blue.

## References


83. Johnston A. Molecular ecology of marine isoprene degradation. PhD thesis. School of Environmental Sciences, University of East Anglia, Norwich, UK. 2014.

84. Small FJ, Ensign SA. Alkene monooxygenase from *Xanthobacter* strain Py2. J Biol Chem. 1997;272:24913–20.

85. Stainthorpe AC, Lees V, Salmond, GPC, Dalton H, Murrell JC. The methane monooxygenase gene cluster of *Methylococcus capsulatus* (Bath). Gene. 1990;91:27–34.

86. Chen Y, Scanlan J, Song L, Crombie AT, Rahman MT, Schäfer H *et al*. γ-glutamylmethylamide is an essential intermediate in the metabolism of methylamine by *Methylocella silvestris*. Appl Environ Microbiol. 2010;76:4530-7.

87. Crombie AT, Murrell JC. Trace-gas metabolic versatility of the facultative methanotroph *Methylocella silvestris*. Nature. 2014;510:148–51.

88. Coleman NV, Yau S, Wilson NL, Nolan LM, Migocki MD, Ly MA *et al.* Untangling the multiple monooxygenases of *Mycobacterium chubuense* strain NBB4, a versatile hydrocarbon degrader. Environ Microbiol Rep. 2011;3:297–307.

89. Martin KE, Ozsvar J, Coleman NV. SmoXYB1C1Z of *Mycobacterium* sp. strain NBB4: A soluble methane monooxygenase (sMMO)-like enzyme, active on $C_2$ to $C_4$ alkanes and alkenes. Appl Environ Microbiol. 2014;80:5801-6.

90. Whited GM, Gibson DT. Toluene-4-monooxygenase, a three-component enzyme system that catalyzes the oxidation of toluene to p-cresol in *Pseudomonas mendocina* KR1. J Bacteriol. 1991;173:3010–16.

91. Sharp JO, Sales CM, LeBlanc JC, Liu J, Wood TK, Eltis LD *et al.* An inducible propane monooxygenase is responsible for N-nitrosodimethylamine degradation by *Rhodococcus* sp. strain RHA1. Appl Environ Microbiol. 2007;73:6930–8.

92. Eggeling L, Sahm H. Degradation of coniferyl alcohol and other lignin-related aromatic compounds by *Nocardia* sp. DSM 1069. Arch Microbiol. 1980;126:141-8.

93. Furuhashi K, Taoka A, Uchida S, Karube I, Suzuki S. Production of 1,2-epoxyalkanes from 1-alkenes by *Nocardia corallina* B-276. European J Appl Microbiol Biotechnol. 1981;12:39–45.

94. Woods NR, Murrell JC. The metabolism of propane in *Rhodococcus rhodochrous* PNKb1. J Gen Microbiol. 1989;135:2335-44.
95. Táncsics A, Benedek T, Szoboszlay S, Veres PG, Farkas M, Máthé I *et al.* The detection and phylogenetic analysis of the alkane 1-monooxygenase gene of members of the genus *Rhodococcus*. Syst Appl Microbiol. 2015;38:1–7.
96. Tan HM, Mason JR. Cloning and expression of the plasmid-encoded benzene dioxygenase genes from *Pseudomonas putida* ML2. FEMS Microbiol Lett. 1990;72: 259–64.
97. Chartrain M, Jackey B, Taylor C, Sandford V, Gbewonyo K, Lister L *et al.* Bioconversion of indene to *cis*-(1S,2R)-indandiol and *trans*-(1R,2R)-indandiol by *Rhodococcus* species. J Ferment Bioeng. 1998;86:550–8.
98. Vomberg A, Klinner U. Distribution of *alk*B genes within n-alkane-degrading bacteria. J Appl Microbiol. 2000;89:339–48.
99. McMurdie PJ, Holmes S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE. 2013;8:e61217.
100. Wickham H. ggplot2: Elegant graphics for data analysis. New York: Springer-Verlag; 2016.
101. Wickham H. tidyverse: Easily installed and load the “tidyverse”. 2017. https://CRAN.R-project.org/package=tidyverse. Accessed 11 Sept 2018.
102. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. 2018. https://www.R-project.org. Accessed 11 Sept 2018.
103. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 1999;41:95-8.
104. Nicholas KB, Nicholas HBJ. Genedoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author. 1997.
105. Sievert C. plotly for R. 2018. https://plotly-book.cpsievert.me. Accessed 11 Sept 2018.
106. Galili T, O’Callaghan A, Sidi J, Sievert C. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. Bioinformatics. 2018;34:1600-2.