The characterisation of root exudation and colonisation in the rhizosphere of land plants

Joshua David Roworth

Linacre College

Department of Plant Sciences

University of Oxford

Supervised by Philip Simon Poole

Thesis submitted for the degree of Doctor of Philosophy

Trinity 2017

For Gramp

Acknowledgements

To Phil for all his guidance, help and support throughout this work. Without whom none of this work would have been possible.
I also thank my second supervisor Liam Dolan for his help and support, and the rest of the Dolan lab for their support and friendship throughout my time in Oxford.

I am grateful to all members of the Poole lab for their help, advice and friendship throughout my time in Oxford. Special thanks go to Vinoy Ramachandran for all his technical help and encouragement throughout my work, and for critically reading my thesis in detail. Special thanks to Barney Geddes for all his encouragement and support. I thank Helen, Lida and Pauline, for running such a tight lab.

I am deeply obliged to the national environmental research council who generously funded this work.

Last but not least I thank my whole family for their support throughout this endeavor. To my parents and grandparents without whom I would not have attained this point in my life. Finally my fiancée, Emma Wernham, without whom none of this would have been possible.

The characterisation of root exudation and bacterial colonisation in the rhizosphere of land plants

Joshua David Roworth
Linacre College
Department of Plant Sciences
University of Oxford

Thesis submitted for the Doctor of Philosophy
Trinity 2017

Abstract

The rhizosphere of land plants is a diverse and complex environment that is critical to the life cycles of plants. Plants release photosynthetically fixed carbon into the rhizosphere that can act to solubilise nutrients, detoxify metals, and modify soil microbial communities. Within these soil microbial communities a diverse range of microbes exist, ranging from the beneficial plant growth promoting rhizobacteria to plant pathogens. These microbes must be able to effectively colonise the rhizosphere and plant rooting systems for their survival. To study root exudation and the bacterial colonisation of the roots two separate techniques were used.
To study the spatial and temporal release of exudation in *Oryza sativa*, Zea mays, *Arabidopsis thaliana,* and *Marchantia polymorpha*, luciferase based bioreporters, which respond to plant derived carbon were used. These bioreporters based in *Rhizobium leguminosarum* allowed for the production of maps showing the spatial nature of exudation from plant roots, alongside changes in exudation temporally. In this work I showed that the release of the C4-Dicarboxylic acids was conserved across all plant species tested. Moreover, the release of these exudates occurred on specific rooting structures, such as lateral roots, and not across the whole rooting system. This work has shown that *R. leguminosarum* based bioreporters can be flexibly used across a range of different plant species.
For studying the genetic basis of *Azorhizobium caulinodans* colonisation of the *O. sativa* rhizosphere, a technique known as insertion sequencing was used. This technique highlighted bacterial genes involved in rhizosphere and root colonising growth. Using these data, several interesting gene operons were identified that played a crucial role in rhizosphere growth. One example was a putative propionate catabolism operon, indicating that *A. caulinodans* is using plant-derived propionate as a carbon source for growth. Insertion sequencing was also used to probe the genetic basis of nitrogen fixation in *A. caulinodans*, here I showed that insertion sequencing was applicable to this study and could identify a number of genes involved in low nitrogen growth.

ABBREVIATIONS

ABC	ATP-binding cassette
AMF	Arbuscular mycorrhizal fungi
Amp	Ampicillin
Amp^r	Ampicillin resistant
AR	Advantageous root
ALMT	Aluminium induced malate transporter
ATP	Adenosine triphosphate
BNF	Biological nitrogen fixation
BLAST	Basic local alignment search tool
cfu	Colony-forming units
Cps	Counts per second
CR	Crown root
cv	Cultivar
dpi	Days post inoculation
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
eCR	Embryonic crown root
EDTA	Ethylenediaminetetraacetic acid
EPS	Exopolysaccharides
ES	Essential
FAO	**Food and agricultural organization of the United Nations**
FP	Fahraeus medium
GA	Growth advantageous
GC	Gas chromatography
GD	Growth defective

GI	Growth impaired
Gm	Gentamycin
Gm^r	Gentamycin resistant
HPLC	High performance liquid chromatography
Hr	Hour
INSeq	Insertion sequencing
Km	Kanamycin
Km^r	Kanamycin resistant
LB	Luria Bertani broth
LC-MS	Liquid chromatography mass spectrometry
LCO	Lipochitooligosacharides
LMWOA	Low molecular weight organic acid
MATE	Multidrug and toxic compound extrusion transporter
MCS	Multiple cloning site
MS	Mass spectrometry
N	**Nitrogen**
NAD+	Nicotinamide adenine dinucleotide (oxidised form)
NADH	Nicotinamide adenine dinucleotide (reduced form)
NADP+	Nicotinamide adenine dinucleotide phosphate (oxidised form)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
NE	Non-Essential
Nm	Neomycin
Nm^r	Neomycin resistant
OD	Optical density
ORS571	*Azorhizobium caulinodans* ORS571
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
P	**Phosphorus**

Pi	Inorganic phosphate
***Rlv*3841**	Rhizobium leguminosarum bv. viciae 3841
RNA	Ribonucleic acid
GFP	Green fluorescent protein
Spc	Spectinomycin
Spc^r	Spectinomycin resistant
SR	Seminal root
Str	Streptomycin
Str^r	Streptomycin resistant
TCA	Tricarboxylic acid
Tc	Tetracycline
Tc^r	Tetracycline resistant
TIR	Terminal inverted repeat region
Tm	Melting temperature
TY	Tryptone yeast
UMA	Universal Minimal Salts agar
UMS	Universal Minimal Salts
UN	United Nations
UV	Ultraviolet
v/v	Volume of solute/volume of solution
Wt	Wild-type
w/v	Mass of solute/volume of solution

Table of Contents

Chapter 1

Introduction

1.1 - Introduction

Since the dawn of agriculture there has been a concerted effort to improve the yields of food crops to feed the ever-burgeoning population. The world population is predicted to reach 9.7 billion by 2050 and over 11 billion by 2100 as outlined in the United Nations (UN) 2017 World Population Prospects report. Increasing population sizes put ever-increasing strain on global resources, most notably the consumption of food. Not only will food consumption increase, but also the demand for energy and space, with both of these having deleterious effects on global food supply. Furthermore, due to global warming compounds these problems are compounded, with the changing climate already affecting the growth of staple food crops (Wheeler and von Braun, 2013).

It is estimated by the Food and Agricultural Organization of the United Nations (FAO) that up to 795 million people are undernourished across the globe, as outlined in their 2015 Food Insecurity report. With developing nations contributing most to these numbers, sub-Saharan Africa is one of the hardest hit areas on the globe (Sanchez and Swaminathan, 2005; Mueller *et al.*, 2012). One of the main goals in the fight for food security is the sustainable increase in crop yields in underperforming nations. Currently, the most developed nations maintain a ceiling of high crop productivity via the mass input of nutrients/fertilisers into arable land. This input maintains the yields of rice (*Oryza sativa*), wheat (*Triticum spp.*), and maize (*Zea mays*) close to the theoretical limit (Tilman *et al.*, 2011; Mueller *et al.*, 2012). There continues to be a large disparity between yields in the developed and developing nations, largely due to fertiliser inequality. One of the main factors is the prohibitive cost of nitrogen- and phosphate-based fertilisers. In developing nations where soil quality is poor and the need for these fertilisers readily apparent, affordable access to these resources is lacking.

1.1.1 – Fertilizers and agriculture

The nitrogen (N) economy remains one of the key driving forces behind modern agriculture. According to the FAO, in 2015 the global demand for reactive nitrogen compounds (Nr) reached 113 million tons. It was the development of the Haber-Bosch process in the early 20th century, which allowed the industrial production of ammonia for use as an agricultural fertiliser to be viable. Today this process is the major source of global Nr (Galloway *et al.*, 2008). In fact, the industrial fixation of atmospheric N into Nr has reached over 120 million tons a year, which is greater than fixation by all of the natural processes combined (Rockström *et al.*, 2009). Use of Nr comes with a large cost, both economically and environmentally. As stated earlier, the economic cost involved in the production and transport of Nr makes it prohibitively expensive for poorer farmers and nations. However, it is the environmental cost which poses a long-term threat. Current rates of Nr application have caused a large perturbation in the terrestrial N cycle. The majority of Nr that is applied to agricultural land is washed out of soil into the water systems (rivers, lakes, seas, oceans). Influx of nutrients causes eutrophication of these water systems, resulting in a range of adverse environmental effects (Rockström *et al.*, 2009). Furthermore, this anthropogenic shift of the N cycle (compounded by shifts in the carbon (C) cycle) has caused an increase in atmospheric concentrations of nitrous oxide (N_2O) to levels not seen for over 650 000 years (Gruber and Galloway, 2008). In terms of climate change, the warming potential of N_2O is second only to that of methane (CH_4) (with carbon dioxide (CO_2) the highest contributor) (Ravishankara *et al.*, 2009). These effects are a result of the Nr being applied in agriculture and not associated with the direct production of Nr. What this says is that there is an excess of Nr being applied in the developed world and a deficit in the developing world. Perhaps more importantly is that whilst in the short term inorganic Nr is necessary to maintain global food supplies, a shift to more sustainable agriculture is key.

Phosphorus (P) is another key nutrient for the growth of all plants. Unlike Nr, phosphate, a so called fossil-mineral, is mined from rock (Rockström *et al.*, 2009) and is therefore a limited resource. Global demand is much lower than for N-fertilisers, with global production at approximately 20 million tons a year. Whilst Nr is involved in the eutrophication of water systems, P is by far the biggest contributor with large scale anoxic ocean events of the past linked to an increase in available P in these water systems (Handoh and Lenton, 2003). It is clear that use of both N and P-fertilizers must be reduced for a more sustainable world.

Since the middle ages crop-rotation has been used for the upkeep of soil quality, to help maintain soil structure, as well as to reduce the depletion of key nutrients. Leguminous crops, such as clover and alfalfa, have been used across the globe for this purpose. The legume family of crops form a beneficial symbiosis with soil-borne bacteria. These bacterial species, known collectively as rhizobia, establish structures on the surfaces of roots known as nodules. In return for carbon supplied by the plant, these rhizobia fix atmospheric N into NH_3, which is then supplied directly to the plant. In essence these rhizobia act as bio-fertilizers, with legumes growing without the need of exogenously supplied N (Oldroyd *et al.*, 2011; Wielbo, 2012). By using legumes as a rotational crop, the biologically available nitrogen content of the soil is replenished. These rotations act as a sustainable method of N fertilization without the expenditure of large amounts of energy and adverse environmental effects. This symbiosis is an example of one the key interactions occurring below ground in the plant rhizosphere.

1.1.2 – The rhizosphere.

For the past 100 years there has been a push to understand the above and below ground interactions between plants and their environments. These interactions range from the non-biological (e.g. nutrients, pH, metals, water, O_2 and CO_2) to the

biological (e.g. microbiome, root exudation, herbivores, parasites, and plant-plant signalling). Such interactions can be both beneficial or detrimental to plant health (Lynch, 1987; El-Shatnawi and Makhadmeh, 2001; Hirsch *et al.*, 2003; Badri and Vivanco, 2009; Dennis *et al.*, 2010).

Below-ground interactions are arguably the most important to plant health. The interface between the roots and soil is home to a myriad of interactions; perhaps most importantly it is the site of many microbial interactions with the rooting system. The soil-root interface, known as the rhizosphere was described by Lorenz Hiltner in 1904 as “the soil compartment influenced by the roots”. The emphasis of his work was on the importance of soil microorganisms and their direct effect on plant health and nutrition (Hiltner, 1904). Today the rhizosphere is a well described area which has now been established to contain three distinct zones: 1) Ectorhizosphere; the soil immediately surrounding roots; 2) Rhizoplane; the root surface; 3) Endosphere; the inner root surface, including the cortex and endodermis. The definition of the rhizosphere has evolved with our understanding to encompass the complex microbiological and biogeochemical effects that are influenced by roots (Lynch, 1987, 1990; Hirsch *et al.*, 2003; Dennis *et al.*, 2010; McNear Jr., 2013). Many of the key interactions within the rhizosphere involve soil-based microbes, including, but not limited to, mycorrhizal associations, plant growth-promoting rhizobacteria (PGPR), pathogenic interactions (e.g. fungi, bacteria), as well as nitrification and denitrification of soil. The soil microbiome and subsequently interactions therein, vary with plant species and genotype, soil type, biotic/abiotic stresses and anthropogenic effects (Berendsen *et al.*, 2012; Lakshmanan *et al.*, 2014). That is to say, there are a myriad of different factors that govern the soil microbiota. The arbuscular mycorrhizal fungi (AMF) and rhizobial-legume symbioses, are some of the most well-known and studied interactions that occur between plants and microbes. The AMF interactions are believed to be some of, if not the, earliest symbiotic relationships between plants

and microbes (Tisserant *et al.*, 2013). During symbiosis with AMF, plants supply photosynthetically fixed C to the fungi in return for biologically accessible P (Bonfante and Genre, 2008; Oldroyd, 2013). The rhizobial-legume symbiosis follows a similar pattern to that of the AMF, in so much that photosynthetically fixed C is traded for biologically fixed N in the form of NH_3 (Oldroyd *et al.*, 2011). This is an over-simplification of these two associations, but highlights two well-studied beneficial interactions that are of agricultural importance. Today, rhizosphere biology encompasses an ever-growing area of study covering symbiosis, disease, nutrient solubilisation, heavy-metal toxicity and plant-plant signalling.

The key to all of these interactions lies in communication between plants and their environment. This communication is achieved via the release of plant-derived C into the rhizosphere, a process known as root exudation (Badri and Vivanco, 2009).

1.2 – Root architecture

This thesis has a focus on the rooting systems of four different plant species: *Arabidopsis thaliana, Oryza sativa, Zea mays* and *Marchantia polymorpha*. This review acts to highlight the structural differences between the rooting systems of these plant systems.

Rooting systems are one of the key interfaces between plants and their environments. Their development was one of the key events during the colonization of land by plants, providing a mechanism for anchorage, water uptake and nutrient acquisition (Bateman *et al.*, 1998). Roots have the same basic function in all plant species, despite this, plants exhibit wide species to species variation in their root architecture (RA), as well as significant natural variation in RA within species

(Cannon, 1949; Fitter, 1987; Loudet *et al.*, 2005; Osmont *et al.,* 2007; Péret *et al.*, 2014).

The earliest rooting systems were made up of single cellular, tip-growing filamentous cells known as rhizoids (similar to root hairs) that develop on the free-living gametophytes of both non-vascular and vascular plants (Bateman *et al.*, 1998; Jones and Dolan, 2012). Figure 1.1 highlights the occurrence of rhizoids in the lineages of extant land plants. *Marchantia polymorpha* is an example of an extant liverwort whose rooting system is comprised single cellular rhizoids extending from the ventral surface of the gametophyte (thallus). The RA of *M. polymorpha* is relatively simple in comparison to the architecture seen in vascular seed plants, despite this, the role of rhizoids in nutrient and water uptake is still relatively understudied (Jones and Dolan, 2012). Figure 1.2A shows a simplified diagram of *M. polymorpha* root architecture.

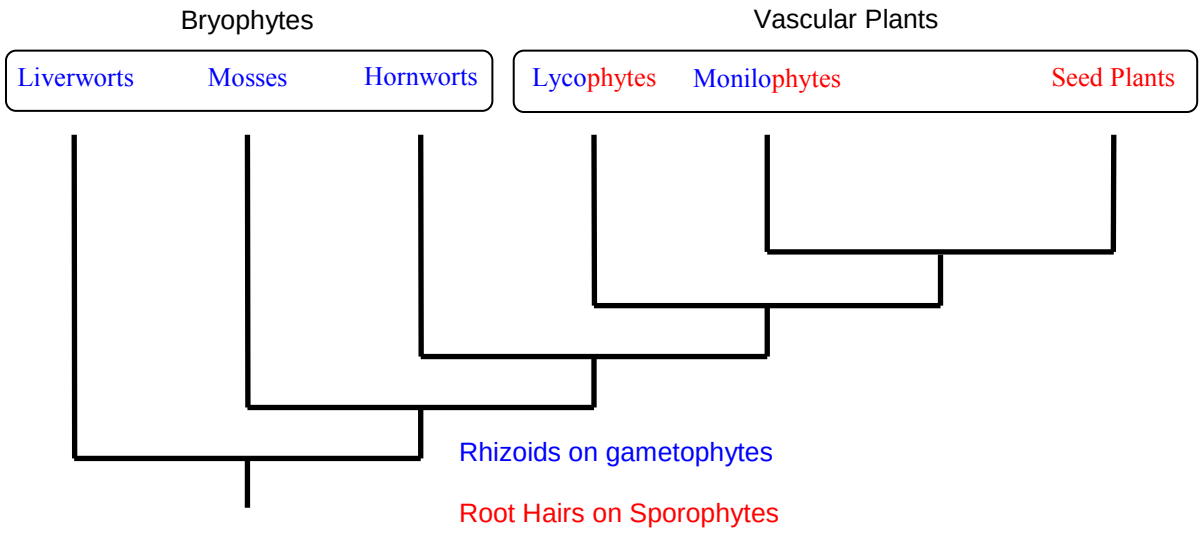


Figure 1.1 Divergence of the land plant lineages alongside the occurrence of rhizoids. Root hairs form on the roots of vascular plants, whilst rhizoids develop on the gametophytes of the bryophytes, lycophytes and monilophytes. Tree modified from (Jones and Dolan, 2012)

In contrast to the 'simple' rooting systems of non-vascular plants (such as *M. polymorpha*), vascular plants have developed complex branching root systems often referred to as 'true' roots. True roots are characterized by the formation of primary

and secondary roots, with single cellular root hairs which form and extend from the root epidermal cells (Brundrett, 2002; Osmont *et al*., 2007). Within the seed plants, two basic architectural morphologies can be seen: allorhizic and secondary homorhizic (Cannon, 1949; Osmont *et al.,* 2007). Allorhizic roots are typically exhibited by the dicots, whilst secondary homorhizic roots are typical of the monocots (Osmont *et al.,* 2007). *Arabidopsis thaliana* is an example of a dicot which forms allorhizic roots. These root systems are characterised by a long primary root (PR) which originates from the embryo, lateral roots (LR) which form from the PR and extend into the surrounding area. LRs can undergo further branching with the formation of secondary and tertiary LRs (Osmont *et al.,* 2007; Nibau *et al*, 2008). Figure 1.2B shows a simplified diagram of *A. thaliana* root architecture.

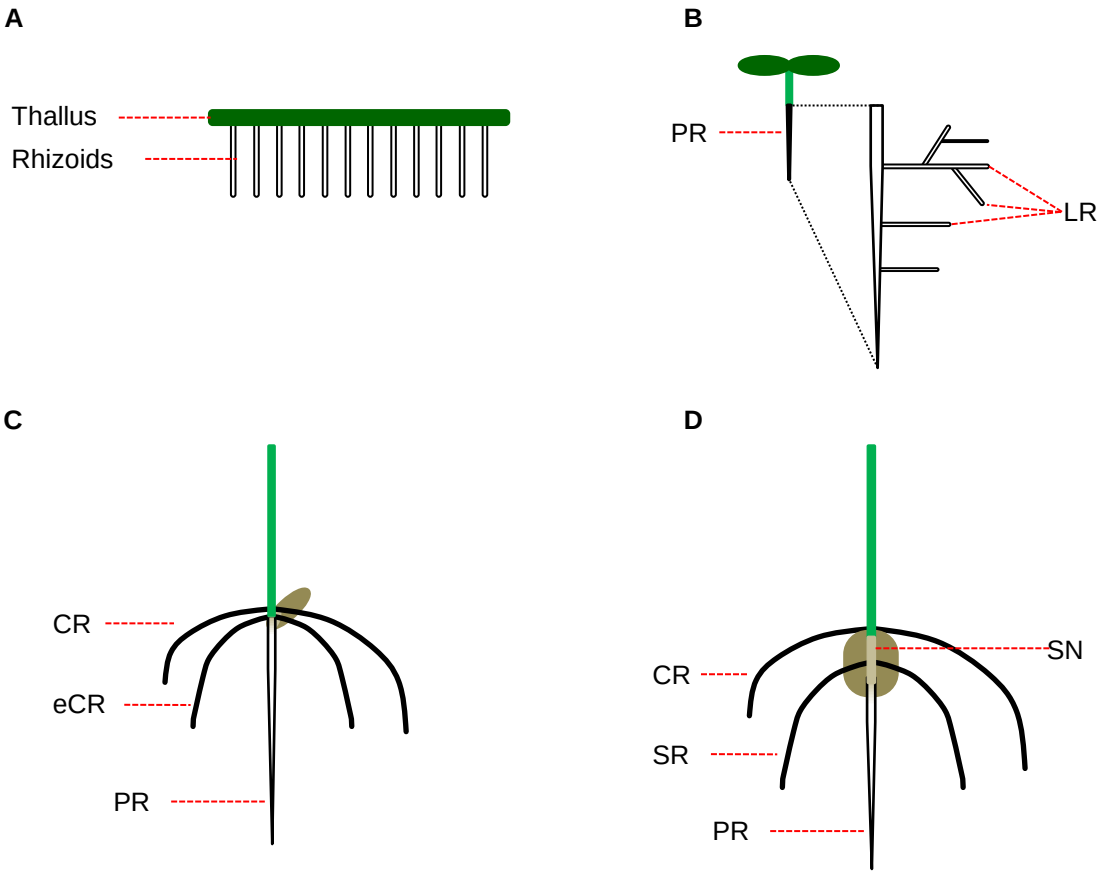


Figure 1.2 Simplified root architecture for four land plant species: *M. polymorpha*, *A. thaliana*, *O. sativa* and *Z. mays*; root hairs are not shown. A – *M. polymorpha,* root architecture is made up of tip-growing filamentous cells (rhizoids) which extend from the ventral side of the thallus. B – *A. thaliana,* produces a single PR from which LRs form, these roots can undergo secondary and tertiary branching. C – *O. sativa*, produces a single PR during germination, up to five eCRs can form post germination, post embryonic CRs form from the belowground stem. Both PRs and CRs produce lateral roots and undergo secondary to tertiary branching. D – *Z. mays*, produces a single PR during germination, SRs form on the scutellar node post-germination, post-embryonic CRs form from the belowground stem. Both PRs and CRs produce lateral roots and undergo secondary to tertiary branching. CR – crown root; eCR – embryonic crown root; LR – lateral root; PR – primary root; SR – seminal root; SN – scutellar node.

Oryza sativa and *Z. mays* exhibit secondary homorhizic roots (Osmont *et al.,* 2007; Rebouillat *et al.*, 2009). In these root systems a PR originates from the embryo, from which LRs can form (much like in allorhizic roots). Where secondary homorhizic roots differ from allorhizic is with the formation of adventitious roots (AR) which can undergo branching with formation of LRs (Brundrett, 2002; Hochholdinger *et al.*, 2004; Osmont *et al.,* 2007).

In *O. sativa* and *Z mays* ARs form on the belowground base of the stem and are called crown roots (CR) in later stages of growth ARs can also form on the aboveground stem and are named brace roots (Hochholdinger *et al.*, 2004; Osmont *et al.,* 2007; Rebouillat *et al.*, 2009). *Oryza sativa* and *Z. mays* RA differs slightly in the formation of some ARs. In *O. sativa* up to five CRs form from the embryo approximately 7 days after germination (Rebouillat *et al.*, 2009). In *Z. mays* a number of embryonic seminal roots (SR) form on the scutellar node (zone that separates the primary root and mesocotyl) approximately seven days after germination (Hochholdinger *et al.*, 2004; Paschold *et al.*, 2010). Diagrams of the rooting structures of *O. sativa* and *Z. mays* are shown in figure 1.2C and 1.2D respectively. In all root true systems the PR and LRs have a tip of dividing cells known as the apical meristem, immediately behind this is the elongation zone where newly formed cells elongate and add to overall root length.

1.3 – Root exudation and its effects

Root exudation is the mechanism by which plants communicate with and alter their soil environment. Exudates are made up of water, ions, various enzymes and mucilage, along with a diverse and vast array of carbon containing metabolites. The rate of exudation is so great that it is estimated up to 30% of all photosynthetically fixed carbon makes its way into the rhizosphere (Bais *et al.*, 2006; Kaiser *et al.*, 2015). This carbon is released in the form of low- and high-molecular weight organic compounds including sugars, organic-acids, phenolics, amino-acids, flavonoids, fatty-acids, terpenoids, alkaloids, growth regulators, enzymes and other secondary metabolites (Bertin *et al.*, 2003; Bais *et al.*, 2006; Baetz and Martinoia, 2014; Kaiser *et al.*, 2015). All of these compounds are released and cause changes in the rhizosphere, be it modifying the microbiome, modulating symbioses or chelating nutrients and metals.

1.3.1 – Exudates and signalling

Root exudates play an important role in how plants select and communicate with different soil microbial species, and is critical for the establishment of successful relationships between plants and microbial communities (Bais *et al.*, 2006; Berendsen *et al.*, 2012). The diversity of nutrients released during exudation produces a rich but also competitive environment for microbial growth. The key to successful rhizosphere colonisation by a microbe is its ability to outcompete other microbial species. This can be achieved through a variety of methods: chemotactic responses, metabolic specialisation for specific carbon pools or resistance to plant-derived defence compounds. An example is that many bacterial species inhabiting the rhizosphere produce flagella allowing free movement throughout the soil environment and to respond to chemical gradients such as those produced by exudation.

AMF are major benefactors to plant health, helping to supply water and nutrients to the plant. This symbiosis between plant and fungi is a result of complex cross-talk between the two organisms. This is mediated primarily through the use of exudates acting as signalling molecules within the rhizosphere. The first step in this communication is the release of strigolactones from roots. These diffuse into the rhizosphere and are detected by AMF, which results in an increase in hyphal branching, alongside an upturn in spore germination. Lipochitooligosacharides (LCOs) are released from AMF, these LCOs have been shown to be involved in the initiation of a variety of plant cellular processes involved in formation of mycorrhizal associations (Besserer *et al.*, 2006; Maillet *et al.*, 2011; Oldroyd, 2013). Interestingly AMF-produced LCOs show structural similarity to rhizobial Nod factors (Maillet *et al.*, 2011), which are key for legume-rhizobial symbioses. This similarity is due to both AMF-produced LCOs and Nod factors having tetrameric and pentameric backbones made up of *N*-acetylglucosmaine residues (Maillet *et al.*, 2011; Oldroyd, 2013).

The rhizobia-legume symbiosis is another key interaction initially governed by plant root exudates. Biological nitrogen fixation (BNF) accounts for up to 50% of the total N available to the biosphere (Batut *et al.*, 2004; Peoples *et al.*, 2009). Plant-produced flavonoids are key compounds involved in signalling between plants and rhizobia initiating the N-fixing symbiosis, and are divided into subgroups e.g. chalcones, flavanones, anthocyanidins and flavones (Aoki *et al*, 2000). One of the most important effects of flavonoid secretion is initiation of expression of nodulation (*nod*) genes in rhizobia (Moscatiello *et al.*, 2010). This is achieved through recognition of flavonoids by one or more bacterial NodD proteins activating gene transcription (Shaw *et al*., 2006). A study looking at Ca^{2+} induced signalling by plant released flavonoids in pea and its N-fixing symbiont *Rhizobium leguminosarum* bv. *viciae* 3841 (*Rlv*3841) showed that naringenin, luteolin and hesperetin initiated Ca^{2+} signalling whilst daidzen and genistein did not (Moscatiello *et al.*, 2010). This shows that *Rlv*3841 exhibits specialisation to flavonoids secreted by its plant host, and the plant host has specialised secretion of flavonoids in establishing a successful N-fixing symbiosis.

1.3.2 – Exudates as microbial carbon pools

Exudates play a significant role in supply of carbon to microbial species across the entire rhizosphere. These pools can be generalised *i.e.* aimed at a broad-range of microbial species; or more specialised *i.e.* aimed at specific microbial species. The pea rhizosphere has been widely studied due to the importance of symbiosis with N-fixing rhizobia. Here the effects that plant-derived carbon has on the growth and colonisation by bacteria (mainly symbiotic rhizobia) is better understood than in many other plant species. Previous studies have linked homoserine catabolism to the growth and competitiveness of *R. leguminosarum* in the rhizosphere of peas (Hynes and O'Connell, 1990). These studies have shown that the rhizobial strain present in

the nodules of pea can catabolise homoserine, while the most prevalent rhizobial strains in Faba bean and lentil nodules could not (Hynes and O'Connell, 1990). This shows that specific exudates contribute to the effectiveness one specific strain (*R. leguminosarum*) had on colonisation of these legume roots. Recent transcriptomic analysis of *Rlv*3841 with different plant species (pea, sugar beet and alfalfa) showed specific changes in gene expression depending on the plant host (Ramachandran *et al.*, 2011). The gene pRL80071 encodes a putative homoserine dehydrogenase complex that was up-regulated specifically in the pea rhizosphere suggesting that *Rlv*3841 may be able to catabolises homoserine present in the rhizosphere (Ramachandran *et al.*, 2011; Vanderlinde *et al.*, 2014). A more recent study looked at a gene cluster required for homoserine catabolism (pRL80080 – pRL80088), which is present on the symbiotic plasmid pRL1JI (Vanderlinde *et al.,* 2014). In a strain with a polar mutation in pRL80083 (encoding a dehydrogenase) there was no growth on minimal media containing homoserine as the sole carbon source, while a wild-type strain showed strong growth. Furthermore, both strains grew equally well on minimal media containing mannitol. A strain lacking pRL1JI showed no growth on homoserine, but once pRL1JI was conjugated in, it gained the ability to grow on homoserine. Further experimentation showed that strains mutated in homoserine catabolism became less competitive when compared to wild-type strains, however this mutation did not affect nodule formation or N_2-fixation. Whilst homoserine is not required for N_2-fixation, it is required for *Rlv*3841 to successfully out-compete other microorganisms in the pea rhizosphere (Vanderlinde *et al*, 2014). Catabolism of homoserine gives *R. leguminosarum* a metabolic advantage over other microorganisms in the rhizosphere, highlighting specialisation in both plant and bacterium. This also demonstrates that complex symbiotic interactions are reliant on both partners; without homoserine *R. leguminosarum* would struggle to out-compete other microbes, and without the N-fixing symbiosis with *R. leguminosarum* the plant's

growth in N-limited conditions would be hindered (Berendsen *et al.*, 2012; Turner, James, *et al.*, 2013).

Galactose is prevalent in root exudates from a variety of plants, including pea and alfalfa (Knee *et al.*, 2001). During infection of alfalfa by *Sinorhizobium meliloti*, an enzyme involved in cell wall degradation, polygalacturonase, was induced, indicating that galactose may be involved in the infection phase of the *S. meliloti*-alfalfa symbiosis (Muñoz *et al.*, 1998). Interestingly, *S. meliloti* mutants incapable of galactose metabolism were observed to have increased competitiveness for nodule occupancy (Geddes and Oresnik, 2012).

Release of exudates does not just influence single microbial species but can effect changes across the whole rhizosphere (Badri and Vivanco, 2009; Badri *et al.*, 2009; Dennis *et al.,* 2010; Turner *et al.*, 2013; Haichar *et al.*, 2014). The roles of root exudates in altering the microbial ecology of the rhizosphere have been studied in great detail. These studies have increased in scope since the development of next-generation sequencing technologies. These technologies, coupled with exudate analysis, have allowed unprecedented study of whole microbial communities within the rhizosphere, how these are affected by different plant species, as well as how these communities shift as plants mature (Badri *et al.*, 2009; Chaparro *et al.*, 2013; Turner, Ramakrishnan, *et al.*, 2013). It has been shown that there are significant differences between the ecology of bulk soil and the rhizosphere, as well as large differences between the microbiomes of differing plant species (Kuske *et al.*, 2002; Turner *et al.,* 2013). Using terminal restriction fragment length polymorphism analysis of 16S rRNA on samples taken from the soil of three separate grass species native to the arid grassland environments of Colorado, the microbial composition of the rhizosphere was studied (Kuske *et al.*, 2002). This study showed that there was a stark difference between the bacterial compositions of each rhizosphere, which

differed from bulk soil. A more recent study looked at the effects wheat, oat and pea had on the whole microbial communities (not just the bacterial communities) of soil collected from an agricultural field site in the UK (Turner *et al.*, 2013). The study showed that the rhizosphere microbiomes of each plant were different from one another, despite being grown in the same soil. Kingdom-level shifts were seen in community composition between plant species e.g. pea showed a much higher enrichment for fungal species; whilst pea and oat showed a five-fold increase in their eukaryotic communities over wheat. The authors also highlighted an oat line that was mutated in its ability to produce antifungal compounds. The microbiome of this mutant line, when compared to that of wild-type oat, showed little change in its overall fungal community, however, a large change in the non-fungal eukaryotic community was seen. This highlighted that a relatively simple change in exudate composition can have large, possibly unpredictable, effects within the rhizosphere microbiome, further solidifying the complexity of the rhizosphere and the role of root exudates.

1.3.3 – Exudates and defence

Root exudates can act to aid in plant defence against pathogens. The processes by which this defence occurs spans from the production of antimicrobial compounds, defence signalling and the recruitment of beneficial microbes that act as biocontrol agents (Mendes *et al.*, 2011; Baetz and Martinoia, 2014).

The roots of the tree-legume *Leucaena leucocephala* secrete mimosine, which is toxic to a number of microbial species (Fox and Borthakur, 2001; Ramachandran *et al.*, 2011). This compound acts as a selective pressure against microbial species that cannot degrade or otherwise detoxify it. It is interesting to note that the secretion of this compound allows *L. leucocephala* to select for specific microbial species within

its rhizosphere. *Rhizobium sp.* strain TAL1145 is a nodulating symbiont of this species. TAL1145 is capable of degrading and catabolising mimosine as a nitrogen source for growth (Fox and Borthakur, 2001; Ramachandran *et al.*, 2011). The ability of TAL1145 to degrade mimosine lies in transport via a specific ABC transporter and catabolism by an aminotransferase (MidABC & MidD respectively). This shows that plants release antimicrobial compounds into the rhizosphere to inhibit growth of unwanted microbes, whilst acting as specific compounds to aid growth of beneficial bacteria (Fox and Borthakur, 2001; Pandey and Dwivedi, 2007).

The compound canavanine is synthesized and released from the roots and seeds of leguminous crops such as alfalfa. This molecule exhibits antimicrobial properties for a range of rhizosphere-inhabiting microbes. However, certain rhizobial species have adapted to become resistant this compound (Emmert *et al.*, 1998; Cai *et al.*, 2009). There is some evidence that exporters present in rhizobia act to remove canavanine from within the cell. For example, it was shown by Ramachandran *et al.,* (2011) that during colonisation of alfalfa by *Rlv*3841 a gene encoding a canavanine exporter was up-regulated (Ramachandran *et al.*, 2011).

Root exudates can act indirectly in plant defence by recruiting bacteria that act as biocontrol agents. Selection of these bacteria is driven by a variety of organic acids present in root exudates (Weert *et al.*, 2002; Kamilova *et al.*, 2006; Liu *et al.*, 2014). An example of this is plant selection of various *Pseudomonas sp*. There are many different species and strains of *Pseudomonas* which act to defend a plant against pathogenic microbes (McSpadden-Gardener and Weller, 2001; Weert *et al.*, 2004; Bakker *et al.*, 2007). In tomato, a large proportion of the carbon secreted is in the form of organic acids. It has been shown that organic acids and certain amino acids increase motility in *Pseudomonas fluorescens* WCS365, however various sugars do not (Weert *et al.*, 2004). Further to this, colonisation and competition assays were

performed with a strain mutated in *cheA* (a gene which controls flagella-driven chemotaxis). These assays showed that *cheA* mutants had reduced colonisation ability when compared to wild-type (Weert *et al.*, 2002, 2004). This shows that a major cornerstone of competition for some bacterial species, in particular *Pseudomonas sp*., lies in their motile response to certain root exudates. A further study showed the change in root exudates in response to two microbes: WCS365 (*P. fluorescens* associated with biocontrol of pathogens) and *Fusarium oxysporum* f.sp *radices-lycopersici* (pathogenic fungus of tomato) (Weert *et al.*, 2004). This study showed two interesting results; firstly, the composition/concentration of root exudates changed when their rhizosphere was colonised by different microbial species. Secondly, when colonised by WCS365, tomato plants showed a lower incidence of disease. Upon colonisation with WCS365, the concentration of organic acids in the rhizosphere increased, in particular citric acid, while organic acid concentration drops with pathogen infection (e.g. *F. oxysporum*). With colonisation of both microbes there was an overall increase in concentration of organic acids, with the incidence of disease dropping dramatically, thus indicating that *P. fluorescens* is acting effectively to defend the plant against pathogens. More importantly perhaps, this also indicates that when both microbes colonise the rhizosphere, the plant increases secretion of organic acids in a bid to help recruit specific microorganisms that can aid in pathogen defence.

1.3.4 – Exudates and soil chemistry

Differences in a plant's soil environment lead to a marked shift in exudate profiles. As plants interact with different soil types, the quantity and composition of exudates can shift. In soils that are deficient in key nutrients and minerals, a shift in overall exudate composition can be seen (Bolan *et al.*, 1994; Hell and Stephan, 2003; Badri and Vivanco, 2009; Carvalhais *et al.*, 2011; Richardson *et al.*, 2011; Castrillo *et al.*, 2017).

These shifts differ from plant species to plant species, as well as being specific to the nutrient deficiency detected by the plant.

Phosphate is a key nutrient for plant growth; it is however often mineralised in soils making it inaccessible for plant uptake. Plants have devised a complex mechanism to increase phosphate uptake in deficient soils. These mechanisms involve altering root architecture (Péret *et al.*, 2014), forming beneficial interactions with other microbes such as mycorrhizal fungi (Bulgarelli *et al.*, 2013) and changes in exudation (Treseder, 2013; Kaiser *et al.*, 2015; Castrillo *et al.*, 2017). It has been observed that under phosphate-limiting conditions the exudation of organic acids increases, when compared to non-limited plants. In chickpea and white lupin there was a correlation between the release of certain organic acids and an increase in phosphate solubilisation and, ultimately, plant growth (Veneklaas *et al.*, 2003). This correlation between the increased presence of carboxylates and low phosphate is well known (Badri and Vivanco, 2009; Carvalhais *et al.*, 2011; Castrillo *et al.*, 2017). It is believed that these carboxylates released in the form of anions can complex with iron (Fe), calcium (Ca) and aluminium (Al) bound forms of phosphate, thus releasing it for uptake by plants (Jones, 1998). Another key in phosphate solubilisation is the acidification of the rhizosphere. It has been shown that rhizosphere pH is often lower than that of bulk soil (Hinsinger, 2001; Dakora and Phillips, 2002; Chen *et al.*, 2017). This acidification helps reduce metal-phosphate complexes and releasing phosphate from its mineralised form. The methods by which plants acidify the rhizosphere is still controversial, it is likely a combination of anions, H^+, organic acids and compounds released from soil microbes (Jones, 1998).

Soil acidification, along with release of organic acids, plays a diverse role in the acquisition of key metals present in soils (Chen *et al.*, 2017). Iron is one of the key metals for plant growth and in soil it is often complexed in the form of Fe(III)-oxides

that are inaccessible to plants. Plants grown in iron deficient soils show a marked increase in soil acidification, alongside an increase in the release of citrate (primarily) and other organic acids (Jones, 1998). As pH in the rhizosphere drops, inaccessible Fe(III)-oxides can complex readily with citrate and malate to form Fe(III)-citrate/malate which can be utilised by plants via ferric reductases which release the acid whilst reducing Fe^{3+} to Fe^{2+} for plant use (Zhang *et al.*, 1989; Jones, 1998; Badri and Vivanco, 2009). Alongside release of organic acids, certain phenolic coumarins are also exuded. Coumarins have been detected in a variety of plants, including *A. thaliana* in iron deficient alkaline soils (Fourcroy *et al.*, 2014; Clemens and Weber, 2016). Coumarins act to chelate and reduce Fe(III)-oxide complexes to Fe^{2+} for plant uptake. A separate method of iron uptake is the release of phytosiderophores of the mugineic acid family (MAs). These have been characterised in graminaceous plants including *O. sativa*, *Z. mays*, and barley (Clemens, 2001; Kobayashi and Nishizawa, 2012; Chen *et al.*, 2017). The synthesis and release of these compounds under Iron deficiency is specific to graminaceous plants and is heavily conserved. Interestingly, release of MAs into the rhizosphere follows a diurnal pattern with peaks in exudation seen in the morning (Takagi *et al.*, 1984).

A major hurdle to successful plant growth can be heavy-metal toxicity in soils. Heavy-metal toxicity is often more pronounced in soils with lower pH as metals become solubilised and accumulate in the rooting systems of plants. This can often be a “Catch-22” situation as during rhizosphere acidification (for phosphate and mineral solubilisation), metals such as zinc (Zn), Al and cadmium (Cd), become solubilised and thus toxic to plants (Ma *et al.*, 2001; Yokosho *et al.*, 2011; Chen *et al.*, 2017). Aluminium is one of the most abundant metals in the Earth’s crust, which under acidic conditions can exist as a metal ion (Al^{3+}) that can be toxic to plants. With up to 40% of arable land being classified as acidic, Al-toxicity often limits crop growth (Kochian, 1995; Ma *et al.*, 2001; Chen *et al.*, 2017). Release of low molecular weight

organic acids (LMWOAs), such as malate, citrate and oxalate, have been directly correlated to Al^{3+} exposure in *O. sativa*, wheat, *Z. mays* and *A. thaliana* (Kochian *et al.*, 2004; Ligaba *et al.*, 2006; Chen *et al.*, 2017). Across all plant species, citrate and malate make up the bulk of LMWOAs being released for Al tolerance. Wheat was discovered to release malate via Al-induced malate transporters (ALMT). These transporters have also been identified in oilseed rape (Ligaba *et al.*, 2006) and *A. thaliana* (Hoekenga *et al.*, 2006), localised at the root cap and activated almost immediately in response to the presence of Al (Sasaki *et al.*, 2004). The transfer of the *ALMT1* gene from wheat to *O. sativa* and soybean conferred an increase in Al-induced malate release and subsequent tolerance to the metal. A separate mechanism by which LMWOAs are released is via multidrug and toxic compound extrusion (MATE)-family transporters. These genes have been identified in *O. sativa*, and like the ALMT genes, are induced during Al stress (Yokosho *et al.*, 2011). In these conditions an increase in the quantity of citrate present in the rhizosphere was recorded. By mutating the gene encoding this transporter, a decrease in rhizosphere citrate concentrations together with a decrease in Al tolerance was seen (Yokosho *et al.*, 2011). Whilst these tolerance systems function independently, both can occur within the same plant species. In *A. thaliana,* both ALMT and MATE Al tolerance systems have been identified (Liu *et al.*, 2009). The *AtALMT1* gene conferred Al tolerance by malate release into the rhizosphere (Hoekenga *et al.*, 2006), while a subsequently identified Al-induced MATE transporter was involved in citrate release during Al stress. Single-knockout mutants of this gene showed a reduced tolerance to Al, while a double-knockout of the MATE transporter and *AtALMT1* showed a further reduction in tolerance to Al (Hoekenga *et al.*, 2006; Liu *et al.*, 2009). This confirmed that these two independently operating systems each confer a degree of tolerance to the toxic effects of Al.

1.3.5 – Collecting exudates

Analysis of root exudates has relied heavily on their isolation, followed by identification and quantitation of their constituent components. Numerous methods exist for the collection of root exudates, each with benefits and disadvantages (Oburger *et al.*, 2013).

The most common method of exudate collection is from hydroponically grown plants. In these experiments plants are germinated and grown in aerated hydroponic growth solutions. For exudate sampling, the plants are transferred into a sterile capture solutions e.g. water, for a period of time ranging from a few hours to days (Neumann and Römheld, 1999; Aulakh *et al.*, 2001; Badri. *et al.*, 2009; Jaitz *et al.*, 2011). This method is relatively simple and the transfer does not damage the rooting system. Hydroponic systems however are not ideal for plant growth. Due to lack of mechanical impedance (e.g. because of soil particles) there are changes in root architecture and alongside this, the nutrient availability is far removed from that of a soil or substrate system. One of the main drawbacks is that once root exudates have been collected, the plants are discarded because moving plants between collection solutions and growth mediums can alter plant growth and thus exudation (Badri *et al.*, 2009; Li *et al.*, 2009; Chaparro *et al.*, 2013). The main advantage to this method is its simplicity.

Methods combining growth in a substrate (e.g. soil, sand, vermiculite and combinations therein) with hydroponics have been explored. In these methods, plants are grown in a substrate e.g. soil, before exudates are collected. To collect the exudates, plants are removed from their substrate, rinsed in sterile water, and placed in a sterile capture solution. Once in the capture solution these methods follow the same principles as full hydroponic setups (Aulakh *et al.*, 2001; Li *et al.*, 2013). These techniques allow the plant to grow in a more natural environment, and thus

determination of the composition of a more realistic exudate profile. In contrast to full hydroponic growth, the mechanical impedance provided by growth in such a substrate allows development of a more realistic root architecture (Oburger *et al.*, 2013). However, removal of the plants from the growth substrate can lead to significant damage to fine root structures e.g. root hairs. Furthermore, the rapid change in environment from soil to a capture solution can alter both the rate and composition of exudation. Again, this method of sampling is destructive.

There are methods of *in situ* exudate collection available which combine growth of a plant in a substrate with the ability to collect exudates non-destructively, thus causing minimal interference to plant growth (Dessureault-Rompré *et al.*, 2006; Schulz and Vetterlein, 2007; Oburger *et al.*, 2013; Weidenhamer *et al.*, 2014). This allows for the temporal analysis of exudation, as well as some limited spatial data. These systems also allow for the natural growth of the rooting system, with the use of soil and vermiculite allowing nutrient availability to be matched to real-world conditions. Most importantly these are non-invasive methods, which allow for temporal analysis. An example of one of these systems is the growth of a plant in a rhizobox (also called a rhizotron) with a substrate of choice (e.g. soil or vermiculite) (Dessureault-Rompré *et al.*, 2006; Weidenhamer *et al.*, 2014). Rhizoboxes are constructed with transparent front plates that allow observation of the growing rooting system. To take an exudate sample small holes are drilled into these plates and a micro-suction device (e.g ceramic capillary) inserted. Using a syringe attached to these capillaries, exudates can be collected (Dessureault-Rompré *et al.*, 2006). The experimental setups for *in situ*-capture are complex and expensive, making it difficult to apply to large sample sizes. The sampling zone is very small, and whilst this can allow the collection of spatial data (dependent on the experimental setup), it does mean that sample size is very small and may overlook global scale changes in exudation.

1.3.6 – Detecting exudation

While the roles and importance of root exudation for plant growth have been described above, identification of specific compounds, as well as determination of the total exudate composition is key to understanding how plants are manipulating their rhizosphere. This information forms the backbone for studying the rhizosphere as without this knowledge, rhizosphere colonisation, nutrient acquisition, disease tolerance and stress responses, would be poorly understood.

It is not in the scope of this review to discuss the fine details and differences between techniques but rather to discuss their uses for root exudation detection and quantification. There are numerous methods used to detect and quantify exudation from plants. Three of the most widely used techniques to identify the exudates are High Performance Liquid Chromatography (HPLC), Liquid Chromatography Mass Spectrometry (LC-MS) and Gas Chromatography Mass Spectrometry (GC-MS) (Grayston *et al.*, 1997; Aulakh *et al.*, 2001; Badri. *et al.*, 2009; Carvalhais *et al.*, 2011; Chaparro *et al.*, 2013).

The use of chromatography coupled with mass spectrometry in particular seems to have become the standard for identification and quantification of root exudates. The use of standalone HPLC has greatly declined in favour of using tandem approaches such as GC-MS and LC-MS. Both GC-MS and LC-MS have and are used extensively to detect, identify and quantify exudates released into the rhizosphere (Jaitz *et al.*, 2011; Chaparro *et al.*, 2013). Using these two methods, a plethora of different compounds have been readily identified as present in root exudates, including: organic acids, phenolics, sugars, amino acids, fatty acids and a variety of other organic compounds (Erickso*n et al*., 2001; Dakora and Phillips, 2002; Bacilio-Jiménez *et al*., 2003; Dayakar V. Badri *etal*., 2009; Jaitz *et al.*, 2011; Chaparro *et al*., 2013). The advantage of chromatography-coupled mass spectrometry is in its ability

to identify compounds, even at low concentrations. This has allowed the creation of detailed exudate profiles in a variety of plant species, including *A. thaliana* and *O. sativa* (Aulakh *et al.*, 2001; Chaparro *et al.*, 2013). It was shown in *A. thaliana* that as plants aged, as measured at different developmental stages, there was a drop in the total exudation of sugars, conversely the total exudation of amino acids and organic acids increased (Chaparro *et al.*, 2013). A major advantage, especially for GC-MS based analysis, is that the spectra produced from any given analysis can be compared against large metabolomics databases allowing the rapid identification of known compounds and aiding in identification of unknown compounds (Kopka *et al.*, 2005).

These techniques are not devoid of downsides however. Both GC-MS and LC-MS are labour-intensive methods. In particular, sample preparation can be complex; an example for GC-MS, is the derivatization compounds used. These differ depending on what class of compounds are being identified for example, amino acids and organic acids differ from each other (Williams *et al.*, 2010). Ultimately this increases the number of samples that have to be analysed to produce a dataset encompassing the majority of exudates released. One of the major downsides to these methods is that plants are often grown hydroponically for their whole growth cycle or for the exudate collection period (Jones *et al*., 1994; Van Nieuwenhove *et al*., 2000; Chaparro *et al*., 2013). As stated earlier, there exists a multitude of different environmental factors which affect root exudation and it is well understood when analysing exudates, their composition is indicative of a plants growth environment, so it can be considered that hydroponic growth will result in a different exudate profile than when the same plant is grown in a solid substrate. This method of sample collection also hinders resolution of the temporal nature of exudation. As exudates are collected over 1-4 days (Erickson *et al.*, 2001; Jaitz *et al.*, 2011; Chaparro *et al.*, 2013) and then analysed, the day-to-day, hour-to-hour changes in exudation cannot

be quantified. Whilst standalone HPLC has been largely superseded by HPLC/LC-MS coupling, it is still used when a specific single class of compounds is investigated e.g. organic acids (Aulakh *et al.*, 2001; Gherardi and Rengel, 2004; Yuan *et al.*, 2015).

Total organic carbon analysis (TOC) is used to look at the total amount of C exuded independent of its constituent parts (Uselman *et al.*, 2000; Aulakh *et al.*, 2001; Liu *et al.*, 2005). While this method is useful for looking at global changes in exudate release into the rhizosphere as it takes into account all organic C present, it has been largely superseded by the chromatographic and mass spectrometry approaches.

1.4 – Biosensors

Bioreporters are living cells engineered to produce a measurable signal in response to defined stimuli. This allows them to be used to measure cell numbers and/or a defined environmental stimulus. The function of a bioreporter lies in the construction of a system where a marker gene(s) (e.g. *GFP* or *lux*) transcription (thus expression) is fused with a promoter of interest. Upon transcription of the promoter sequence, marker gene(s) are co-transcribed. Thus the marker gene(s) act as a reporter for the stimuli inducing expression of the promoter. This allows for the construction of two distinct types of bioreporters: 1) Constitutive, where marker genes are continually transcribed and expressed. 2) Inducible, where an exogenous signal activates the promoter and thus expression of the marker (Hansen and Sørensen, 2001; Close *et al.*, 2012). In these discussions an environmental signal is defined as a signal exogenous to the bioreporter.

1.4.1 – Biosensor overview

Some of the first bioreporters designed made use of genes encoding bioluminescence. Two of the most well-known are the *luc* genes from firefly and the *lux* genes from the bacterium *Vibrio fischeri* (Meighen, 1993; Close *et al.*, 2012). The *lux* genes (*luxCDABE*) make up an ideal reporter system as all the required substrates and enzymes needed for light production are contained within the gene cassette and the host cell (Close *et al.*, 2012). This removes the need for an exogenous application of D-luciferin, which is required for light production in eukaryotic bioluminescence systems e.g. the *luc*-system. The self-contained nature of the *lux*-system allows for its flexible use in a variety of conditions, including plants, bacteria, humans and other mammals (Close *et al.*, 2011). The *luc*-based system is considered to give a brighter signal than *lux*, making it useful in systems where there are a low number of bioreporter cells (Close *et al.*, 2012). However, due to the need to apply D-luciferin for light production, the *luc*-system has been limited to experimental systems were luciferin can be easily applied. The *lux* genes have remained the 'go-to' system for bioluminescent-based reporters in bacteria. An example of a *lux*-based system was the development of a bioreporter capable of detecting environmental naphthalene (King *et al.*, 1990). In this study the *lux* genes were inserted into the naphthalene catabolism pathway, specifically into the gene *nahG*. On exposure to naphthalene or the regulatory inducer, salicylate, the *lux* cassette was transcribed and light produced. This study showed that a light signal could be detected within 15 minutes of exposure to naphthalene or salicylate. This highlighted how biosensors utilising *lux* could be used to rapidly detect compounds in the environment. Constitutively expressed luciferase-based bioreporters can be used for monitoring bacterial growth, including their ability to colonise an environment (Meighen, 1993; Close *et al.*, 2011, 2012; Sun *et al.*, 2012). They have been used extensively in animal models to monitor progression and activity of a pathogen causing disease. An example of this was the development of a constitutive *lux*-based

bioreporter in *Yersinia pestis* (Sun *et al.*, 2012). This reporter was used to track the progression of infection through a mouse model, as well as quantification of the pathogen load.

There are a range of advantages to using bioluminescence as a reporting system. One of the main advantages is that, unlike fluorophores, luciferase-based systems do not require excitation by an exogenous light source. This simplifies the systems needed for signal detection, removing the need for an external light source such as a laser or UV light, and relevant filters. The lack of excitation makes this system less invasive than fluorescence systems, where the high energy light used for excitation can damage biological systems (Ge *et al.*, 2013). Another effect of exogenous excitation is auto-fluorescence that can occur within the microbial cell and environment, making detection of weak signals difficult; this does not occur in bioluminescent systems. Luciferase systems benefit from their rapid rate of induction by an inducer to allow real-time measurement of an environmental signal (King *et al.*, 1990). Furthermore, as bioluminescence requires several substrates and proteins to produce light, once the inducer is removed or the cell dies, the signal is lost almost immediately (Hakkila *et al.*, 2002; Close *et al.*, 2012). Despite these advantages, bioluminescence is not devoid of disadvantages. The energy requirements of bacterial luciferase are very high, reaching 20 ATP molecules per photon of light produced (Close *et al.*, 2012; Herron *et al.*, 2013). Without sufficient carbon for growth, these energy requirements can impart a significant metabolic burden upon the cell which can, and does, negatively affect growth. With firefly *luc* based system, the need to exogenously apply D-luciferin limits the use to systems where this compound can readily reach the biosensor cells. Overall luciferase allows for the development of biosensors that allow for rapid, high-resolution and non-invasive detection of environmental signals.

Fluorescence based bioreporters have also been developed. Green fluorescent protein (GFP) was first identified in 1962 (Shimomura *et al.*, 1962) and in 1994 was successfully used as a marker for gene expression in *Escherichia coli* (Chalfie *et al.*, 1994). Since the identification of GFP, a whole range of different fluorescent proteins have been identified and developed which are now used every day in laboratories as marker systems. Like the *lux* sytem, fluorescent proteins have been used as the backbone of bioreporters. An example of this was the production of a bioreporter capable of detecting genotoxic agents in environmental systems (Kostrzynska *et al.*, 2002). In this study, a transcriptional fusion was made between the *recA* promoter and GFP. As *recA* is induced by DNA damage, and genotoxic compounds damage DNA, this fusion allowed the detection of DNA damage and thus the detection of genotoxic compounds in an environmental sample. Reporters utilizing yellow fluorescent protein (YFP) have also been developed. In *Methylobacterium extorquens* this was used to detect and quantify methyl halides such as chloromethane produced by *A. thaliana* (Ul Haque *et al.*, 2013). This provides an excellent example of biosensors being used for the detection of biologically released caron, and more relevantly for this project, the release of caron from a plant. There are a range of advantages and disadvantages for fluorescence. One of the main advantages of using fluorescence is the high spatial resolution that can be achieved. This allows signals to be resolved down to single cells, depending on the imaging setup used e.g. confocal microscopy or flow cytometry. Unlike luciferase, fluorophores can persist inside cells for lengthy periods of time (dependent on fluorophore). This can help turn an otherwise undetectable signal into a detectable one. However, this stability can prove problematic; the persistence of these fluorophores means that temporal changes in compound detection are harder to see. These effects can be reduced or mitigated by using fluorophores with faster degradation rates or degradation tags (which promote protein degradation) fused to the fluorophore. Another disadvantage of fluorescent-based bioreporters is their

induction rate. As fluorophores need to be translated and folded before they can function, the time between signal induction, protein expression, and excitation/emission is quite large. This reduces their effectiveness to detect rapid changes in environmental signals. Perhaps the biggest disadvantage to fluorescence-based signal detection is the need for large and expensive equipment, making these unsuitable for use in the field or under financial constraints (Doi and Yanagawa, 1999; Wang *et al.*, 2008; Herron *et al.*, 2013).

Bioluminescence and fluorescence are both non-destructive reporters. There are other systems available, e.g. *gusA*, *lacZ*, *celB, and inaZ.* These however are very invasive techniques and require the destruction of samples, making them unsuitable for most uses in the plant rhizosphere, especially for temporal analyses (Hansen and Sørensen, 2001).

1.4.2 – Biosensors in the rhizosphere

A variety of different bioreporters have been designed to measure conditions in the soil and rhizosphere (Gage *et al.*, 2008). These have been used to measure nitrate availability (DeAngelis *et al.*, 2005), detection and spatial localisation of sucrose and tryptophan exudation (Jaeger *et al.*, 1999), alfalfa galactosides' exudation (Bringhurst *et al.*, 2001) and the detection of n-acyl homoserine lactones involved in quorum sensing (DeAngelis *et al.*, 2007). These studies made use of either GFP-based reporters or *inaZ* reporters (*inaZ* is involved in bacterial ice nucleation, causing ice nucleation at higher temperatures than normal). There has been a significant lack in the use of bioluminescent reporters to measure rhizosphere processes such as root exudation. A study was published by Darwent *et al.,* (2003) where a constitutively expressed *lux*-based bioreporter was used in *P. fluorescens.* It was previously shown that this reporter showed a reduction in luminescence corresponding to a drop in

available carbon from root exudation. This was applied to determine if there was a change in exudation corresponding to nitrogen availability and if a constitutive *lux*-based reporter could detect this (Darwent *et al.*, 2003).

Recently in our lab, a suite of *lux*-based reporters was developed (Pini *et al.*, 2017). Development of these reporters was based on a comparative transcriptomic study in *Rlv*3841 of colonisation of the rhizospheres of pea, alfalfa and sugar beet. A core set of rhizosphere-induced genes was identified, several of these genes were specific to plant species, and a number were specific to legumes and showed no expression in the sugar beet rhizosphere. Using these data, fifteen bioreporters were constructed based on genes that were up-regulated during rhizosphere growth. For example, a reporter capable of detecting C4-dicarboxylates was developed using the promoter region for *dctA*, a gene up-regulated in the transcriptomic study (Ramachandran *et al.*, 2011; Pini *et al.*, 2017). The bioreporters developed were shown to be capable of mapping root exudation, along with carbon allocation to nodules in pea and vetch (Pini *et al.*, 2017).

1.5 – Exudation conclusions

The importance of root exudation in plant health cannot be understated. As discussed above the roles that exudates play across the rhizosphere are incredibly diverse. Methods have been developed for the collection and identification of root exudates across numerous plant species, ranging from grasses to trees (Badri and Vivanco, 2009). Understanding the spatial and temporal patterns of exudation can help to elucidate its effects on the rhizosphere and how the plant governs carbon release.

The bioreporters as described by Pini *et al.* (2017) provide an exciting opportunity to investigate exudation and root colonisation in more detail. These bioreporters can be applied to other plant systems outside of the legumes and those already tested. In this project, I aimed to use these bioreporters in the spatiotemporal mapping of exudation in four different plant species with different root architectures: the monocots, *O. sativa Z. mays,* the dicot*, A. thaliana* and the non-vascular bryophyte *M. polymorpha*. Furthermore, these bioreporters can help show how a bacterial species in the rhizosphere can colonise the surface of roots and even non-host organisms.

1.6 – Colonisation of the rhizosphere and root surfaces

The colonisation of the rhizosphere by microorganisms is key for the life cycles of all the organisms regardless of whether the interactions are beneficial or detrimental, making the study of the mechanisms involved an exciting prospect. Previously I have discussed in great detail the impact of root exudation for the colonisation of the rhizosphere by microbial species and the important roles that colonisation plays for plant health. The focus now is to discuss the methods by which the genetic basis of bacterial colonisation can be elucidated.

Despite the importance of colonisation of biological systems by bacteria, the underlying mechanisms governing these processes are poorly understood. It is the complexity of this process and lack of adequate technology that make it difficult to unravel. There are a multitude of different processes that bacteria must balance in order for successful colonisation to occur and these include stress responses, nutrient acquisition, competitiveness, quorum sensing and chemotaxis. There has been a concerted effort to further understand the genes, interactions and systems used by bacteria to allow them to successfully colonise their environment. Studies

have been performed in a variety of systems with examples including the human gut, plant roots and even extreme environments such as high salinity lakes and hydrothermal vents (Rothschild and Mancinelli, 2001; Goodman *et al*., 2009; Compant *et al.,* 2010; Yuan *et al*., 2015). Determining the genetic basis of colonisation is the key to understanding bacterial adaptation, survival, and competition during colonisation. A range of genomic techniques to study the genetic basis of colonisation have been developed, ultimately these techniques will lead to the identification of novel genes involved in colonisation.

1.6.1 – Methods of gene discovery

There are a variety of methods for the discovery/identification of genes, including transcriptomic studies looking at gene expression and activity of expressed proteins. One of the principal methods is the mutation of candidate genes which can then be phenotypically screened to identify if a mutation is having a detectable effect on the bacterial lifecycle e.g. is the mutation lethal, does the mutation affect motility. One of the key techniques underpinning this is the generation and use of large mutant libraries. Using these libraries as part of a phenotypic screen allows for identification of new and possibly novel genes involved in a process without prior knowledge of their role/function.

A number of methods to generate large mutant libraries are available. One of the most common techniques is the use of transposable elements for insertion mutagenesis. These systems make use of transposition to insert a segment of DNA into the genome of a target organism (Kleckner, 1981), disrupting the DNA sequence at the point of insertion to alter the function of a gene. One of the most wide-spread systems used for the generation of insertion-based mutant pools in a Gram-negative bacterium is the Tn*5* system (De Lorenzo *et al.*, 1990), which catalyses the random

transposition of a defined transposable element into a bacterial genome. Depending on how the transposable element is marked (e.g. antibiotic resistance), these mutants can be easily selected. This ultimately allows for generation of large insertion mutant pools, easily selected based on antibiotic resistance markers. The defined transposable element allows for the identification of a mutated gene by either probing for the location of the inserted DNA sequence or by arbitrarily primed PCR (AP-PCR). This form of analysis has been used in *Rlv*3841 to identify genes involved in a variety of processes (Vanderlinde *et al.*, 2011, 2014). Vanderlinde *et al.,* (2014) used Tn*5* mutants to identify a metabolic regulator and subsequent metabolic genes involved in homoserine metabolism. Subsequent mutation of these metabolic genes showed these mutants were significantly impaired in their ability to nodulate peas and lentils. These mutant screens suffer from a major disadvantage, in that these methods are very low-throughput. They require the generation of libraries, growth in an environment of interest, identification of desired phenotypes and screening for the insert location. However, the largest disadvantage to these techniques is that a single mutant must be analysed at a time (Mazurkiewicz *et al.*, 2006). Furthermore, it is difficult to apply these techniques for *in vivo* analysis as it is nearly impossible to separate mutants in a mixed population using traditional techniques.

To combat the problems of using mutant libraries for *in*-vivo analysis, signature tagged mutagenesis (STM) was developed to combine insertion mutagenesis with *in vivo* negative selection. This method allows for the identification of genes specifically required for growth in a host environment by negative selection (Hensel *et al.*, 1995; Mazurkiewicz *et al.*, 2006). STM uses short sequences known as signature tags to mark individual mutants and allows for the generation of mutant pools containing hundreds of individually-tagged mutants (the Tn*5* system is used). By marking mutants, it is possible to grow a STM library in a condition of interest then screen for the attenuation of a mutant population. In essence, a mutant that is lost during

growth in a test condition is assumed to be necessary for bacterial growth in that condition. The major advantage of STM is that a large number of mutants can be screened at once with a single passage over the condition of interest. This technique was originally developed for the use in *Salmonella typhimurium* where it was used to successfully identify new genes involved in the virulence of the strain in mice models (Hensel *et al.*, 1995). Since its initial development there have been a variety of advancements in the use of STM. One of the advancements was the combination of STM and microarray technology. Microarray chips allow all tagged mutants from both input and output pool to be analysed at once (Groh *et al.*, 2005; Pobigaylo *et al.*, 2006).This significantly reduced the time and cost to study the composition of input and output pools.

The most recent development in STM methods is its combination with next-generation sequencing (NGS) technologies, allowing for the massive parallel sequencing of STM mutant pools for insertion-site determination (Serrania *et al.*, 2016). This study was carried out using *S. meliloti*. Their previous studies identified 5089 insertion sites using more standard STM methodology, but by combining STM with NGS, a further 4473 insertion sites were identified for a total of 9562 sites (Pobigaylo *et al.*, 2006; Serrania *et al.*, 2016). The combination of STM and NGS has thus increased the depth of analysis that can be performed.

STM has proved to be a powerful tool to identify genes involved during growth of bacteria in complex environments. Its major advantage is that large mutant libraries can be screened in pooled samples, rapidly and cheaply. Due to the nature of signature-tagging, mutants of interest can be easily isolated without the need for them to be reconstructed for further downstream analysis. The downsides inherent to any insertional approach are present in this system e.g. efficiency of the transposition system, randomness of insertions, probability to achieve 90% genome coverage and

the downstream effects of the insertion. The time in developing individually-tagged mutants is also significant. Furthermore, optimisation of the technique can be complicated; tags that give weak or cross-reacting signals need to be removed and, depending on the inoculation density, individually weak signals can be difficult to detect (Perry, 1999). STM and NGS combination has alleviated some of these problems. A recent example has shown many fitness determinants in *S. meliloti* 1021 during colonisation of alfalfa and pea (Salas *et al.*, 2017).

Using large mutant pools is not the only method of identifying genes involved in colonisation. Promoter-trapping technologies can also be used for the positive selection of genes involved in any given process. A process known as *in vivo* expression technology (IVET) based on promoter tapping techniques, was used to identify *Xanthomonas campestris* genes involved during the infection of turnip (Osbourn *et al.*, 1987; Mahan *et al.*, 1993). IVET relies on the generation of a conditionally compromised strain *i.e.* a strain lacking an essential growth factor (EGF). The second key to IVET is a plasmid-borne promoter trap; this trap contains a promoter-less copy of the EGF that is fused to a reporter gene. The bacterial genome is subsequently fragmented and cloned in front of the promoter trap. Libraries containing thousands of strains can be generated in this way. These mutants are then grown under a condition of interest and their survival is dependent on expression of the EGF. Recovered strains are (putatively) identified as containing an environmentally induced loci (EIL). Screening for the activity of the linked reporter gene eliminates constitutive promoters during growth on a general growth medium (Rediers *et al.*, 2005). The IVET strategy has been used in numerous bacterial species, growing in a range of eukaryotic hosts. In terms of soil based bacteria, IVET has been used to identify genes involved in the: colonisation of sugar beet by *P. fluorescens* SBW25 (Rainey, 1999; Jackson *et al.*, 2005), soil colonisation by *P. fluorescens* Pf0-1 (Silby and Levy, 2004), colonisation of *O. sativa* by *Pseudomonas*

stutzeri A15 (Rediers *et al.*, 2003) and the identification of genes involved in the early stages of the symbiosis between *S. meliloti* and alfalfa (Oke and Long, 1999). IVET has been applied to look at gene expression during the colonisation of the pea rhizosphere by *R. leguminosarum* bv. *viciae* A34 (Barr *et al.*, 2008). This study identified 29 loci that were induced during colonisation of the rhizosphere.

IVET has been combined with suppressor analysis (SPyVET). This method follows the initial IVET approach to identify EIL. In this study the EGF was a deletion of *dapB* making the strain auxotrophic for diaminopimelic acid (DAP). A subsequent suppressor analysis was then performed to identify both positive and negative regulators of EILs (Giddens *et al.*, 2007). Using miniTn*5* (with kanamycin as a marker) or IS-Ω-Km/hah, suppressor mutants for each IVET-fusion strain were produced. If insertions allowed the strain to grow on minimal medium deficient in DAP then one of three things have occurred: 1) the insertion disrupted a negative regulator, allowing transcription of *dapB*, 2) the outward-firing *nptII* promoter present on the insert drives expression of a positive regulator, 3) the insert occurred directly upstream of the *dapB* gene where the *nptII* drives its expression. This allowed SPyVET to identify both positive and negative regulators of EILs, as well as the wealth of information provided by the standard IVET approach.

Whist IVET based strategies can provide information about genes transcribed during colonisation, the generation of promoter traps and subsequent analysis is very time consuming, especially for methods such as SPyVET. Furthermore, these methods can only identify upregulated (inducible) genes and not genes that are down-regulated during colonisation.

1.6.2 – Next generation gene discovery

Over the past decade the rapid development and improvement in NGS (Levy and Myers, 2016) has given researchers the ability to rapidly, cheaply and accurately sequence the genomes of a huge number of different organisms across all kingdoms of life (Metzker, 2010; Van Dijk *et al.*, 2014). Moreover, these technologies have allowed development of so-called 'omics' approaches such as genomics, metagenomics and advanced transcriptomics. This has given us unprecedented abilities to probe the genetic underpinning of a wide number of processes. Due to the breadth of these technologies their role across the whole of biology will not be discussed, I will only focus on new techniques for studying of bacterial colonisation.

A technique known as insertion sequencing (INSeq) has been developed which is a rapid, high-throughput method allowing analysis of genes essential for a bacterial species during growth (Goodman *et al.*, 2009) under a variety of conditions (Barquist, Boinett and Cain, 2013; Opijnen and Camilli, 2013). The basis of INSeq and its power lies in assessing the presence of unique transposon insertions within a pool of mutants whose transposition events have saturated the given organism's genome. Mutagenesis is achieved through the use of a MmeI-adapted *mariner* transposon. The *mariner* class of transposons insert specifically into thiamine-adenine (TA) sites within a genome. This insertion preference makes it possible to determine the total number of possible insertion sites within an organism's genome and model the total number of insertions needed to saturate it. The vector used in this thesis, pSAM-RI (Fig 1.3), utilizes a himar1C9 transposase driven by an *rpoD* promoter of *R. leguminosarum* (Perry and Yost, 2014).

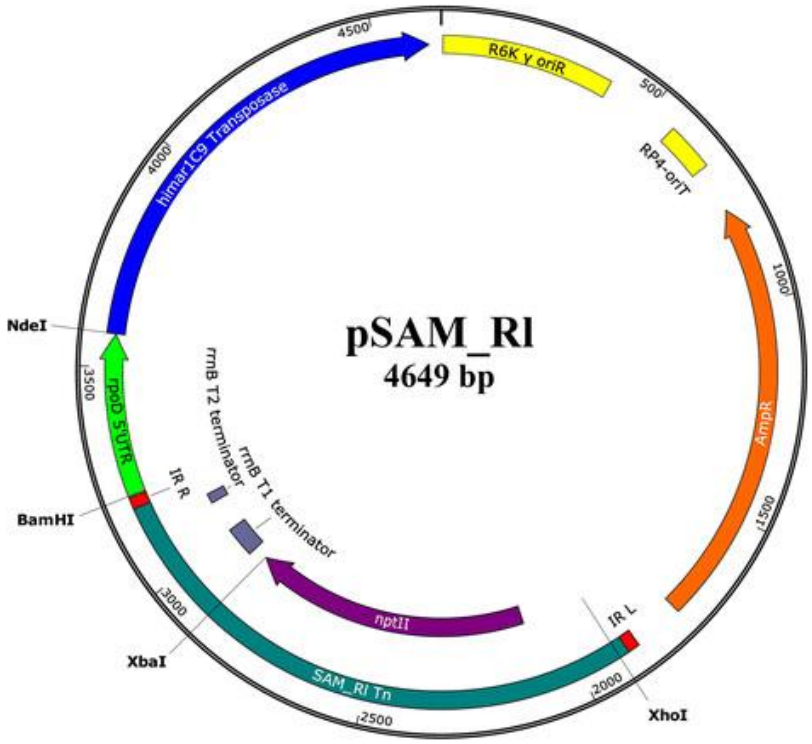


Figure 1.3 Plasmid map of pSAM_R1. Transposable element is outlined by inverse repeat regions (IR_R & IR_L). Reproduced from Perry & Yost (2014).

These classes of transposons follow a non-replicative 'cut and paste' method of transposition. A transposase recognizes terminal inverted repeat regions (TIRs) that flank and subsequently designate a transposable element. Two transposase monomers then bind to the TIRs and subsequently cleave 5' end TIRs hydrolytically; the two monomers are then brought together forming a transposase dimer known as the paired end complex (PEC). At this point the 3' ends of the TIRs are hydrolytically cleaved, freeing the transposable element. The PEC then binds to a TA site leading to insertion of the transposable element, including part of the TIRs (Munoz-Lopez and Garcia-Perez, 2010). Plasmid, pSAM-RI, is as a suicide delivery vector as it lacks the ability to replicate once transferred into a target cell. This is because pSAM_RI carries the R6K origin of replication which only allows for replication in bacterial strains carrying the *pir* gene (e.g. *E.* coli SM10ʎpir). It is this origin of replication that defines the plasmid as a suicide delivery vector, as once it is transferred into a bacterial strain that does not carry the *pir* gene it will not replicate.

This single unique transposition event is what gives INSeq the ability to indicate the 'essentiality value' for any given TA site and subsequently genes, that is to say, how essential is any given gene for growth in the condition tested. A population of rhizobia adequately saturated with these insertions is allowed to grow through 'n' number of generations under a condition of interest. The number of generations is dependent on the inoculation density and the saturation point of the experimental system. The number of insertions present in the end-population can then be analysed via NGS and using a non-parametric analysis, such as a Bayesian model or hidden Markov model (HMM) (Goodman *et al*., 2009; Perry and Yost, 2014; Perry *et al.,* 2016), the essentiality of TA sites and genes are determined. In simple terms, a gene whose TA sites are wholly under-represented can be said to be essential or defective in growth as the cells carrying that insertion delayed or failed to grow. INSeq was first used to analyse the genes underpinning colonisation of the human gut by *Bacteroides thetaiotaomicron* (Goodman *et al.*, 2009). INSeq has further been used for the study of gene-fitness during growth of *Rlv*3841 in a number of free-living conditions, including low oxygen, a variety of carbon sources and growth on complex and minimal media (Perry and Yost, 2014; Perry *et al.*, 2016; Wheatley *et al.*, 2017).

Alongside INSeq, three separate, yet very similar techniques were developed around the same time. These include transposon sequencing (Tn-seq), high-throughput insertion tracking using deep sequencing (HITS) (Gawronski *et al.*, 2009), and transposon-directed insertion site sequencing (TraDIS) (Langridge *et al.*, 2009). Out of these three techniques Tn-seq is most similar to INSeq (van Opijnen *et al.*, 2009). Tn-seq and INSeq both make use of the Himar I *mariner* transposon resulting in random insertions into 'TA' motifs, and both make use of the type IIS restriction enzyme *Mme*I restriction site to cut a defined 16 bp fragment of genomic DNA flanking the insert. These techniques differ only in the methods for sample preparation (van Opijnen and Camilli, 2013). HITS and TraDIS do not make use of

type a IIS restriction enzyme, rather they use random DNA shearing to capture the genomic DNA adjacent to the insert. This results in fragments of differing lengths that can lead to size-based biases in PCR amplification steps. INSeq, Tn-seq and HITS all make use of the Himar I *mariner* transposon, whilst TraDIS makes use of a modified Tn*5* system. Ultimately these techniques all provide a similar function in that they can assess the essentiality of individual genes for growth under given conditions.

Transcriptomics has benefited tremendously from the development of cheaper and more powerful sequencing technologies. Microarrays were, for years, the gold standard for the measurement of gene expression during cellular growth (Zhao *et al.*, 2014). They have been widely used across many different organisms, from prokaryotes to eukaryotes. Of particular interest, microarrays have been used for the transcriptional study of adaptations between rhizobia and plants (Ramachandran *et al.*, 2011). Microarrays provide a method to identify genes that are both up- and down-regulated during an organism's growth, relying on hybridisation of target sequences to DNA-probes present on an array chip. Design of these probes is based on genome sequences and known/predicted ORFs for a given organism. This limits the microarrays to measure only transcription of known sequences, and does not allow for the detection of unknown genes or regulatory elements (Zhao *et al.*, 2014). A new technique known as RNA-sequencing (RNA-Seq) has been developed (Mortazavi *et al.*, 2008). If a near-accurately sequenced genome is available then RNA-Seq provides a method to analyse the transcriptome of a target organism and identify new genes and regulatory elements. RNA-Seq derives its power from the ability of NGS to rapidly and accurately sequence millions of cDNA fragments without the need for specifically designed probes (unlike microarrays). These sequencing reads are individually mapped to a reference genome, allowing for quantification of reads for each transcribable element in a genome (Mortazavi *et al.*, 2008; Zhao *et*

al., 2014). RNA-Seq has been used to probe the transcriptome of numerous different bacterial organisms (Croucher and Thomson, 2010), including a variety of rhizobial species (López-Leal *et al.*, 2014; Pérez-Montaño *et al.*, 2016). RNA-Seq provides a powerful technique for the deep analysis of gene expression alongside identification of new or previously uncharacterised genes. Due to its advantages over microarrays it has largely replaced them for transcriptomic analyses.

1.6.3 – Colonisation conclusion

This review has highlighted methods for studying bacterial colonisation of roots and the rhizosphere. INSeq provides an opportunity for deep genetic analysis of root colonisation and has been used in this project to study colonisation of the *O. sativa* rhizosphere by the N_2-fixing diazotroph *Azorhizobium caulinodans* ORS571. INSeq has been used to study *Rlv*3841 (Perry and Yost, 2014) and could be adapted to examine *A. caulinodans* ORS571. INSeq (along with TN-seq, HITS, and TraDIS) has not been previously applied to the analysis of root colonisation by beneficial bacteria, providing an exciting opportunity to shed light on the complex genetic control surrounding bacterial colonisation of roots.

1.7 – *Azorhizobium caulinodans*

The N-fixing diazotroph *A. caulinodans* was first isolated from its leguminous host plant *Sesbania rostrata* (Dreyfus *et al.*, 1988) where it is able to form both root and shoot nodules during symbiosis (Van Nieuwenhove *et al.*, 2000). At the same time *A. caulinodans* has shown the ability to colonise the root cortex and move into the xylem of *S. rostrata*, thus colonising the whole plant (O'Callaghan *et al.*, 1997). *A. caulinodans* ORS571 was fully sequenced in 2008 identifying a 5.37 Mb circular chromosome containing 4717 putative protein coding sequence, and a symbiosis island of 87.6 Kb (Lee *et al.*, 2008; Ling *et al.*, 2016). It is capable of fixing nitrogen in

a free-living state, as well as during symbiosis, making the *in vitro* analysis of fixation possible. *A. caulinodans* belongs to the family *Xanthobacteraceae*, making it taxonomically different from other more studied rhizobial species (Lee *et al.*, 2005). With access to a complete genome together with its *in vitro* (free-living) N-fixing ability*, A. caulinodans* ORS571 is an excellent choice for studying root colonisation as well as N_2-fixation.

As well as colonising its host *S. rostrata, A. caulinodans* ORS571 is able to colonise a range of plants including *O. sativa*, wheat and *A. thaliana* (Webster *et al.*, 1997; Stone *et al.*, 2001). ORS571 can endohpytically colonise *S. rostrata* through crack-entry at lateral root junctions before moving into the xylem and leafy tissues of plants. This method of colonisation and entry is conserved in *O. sativa*, *Triticum aestivum*, and *A. thaliana*. The plant-produced flavonoid, naringenin, has been implicated in activation of *nod* genes in ORS571 during symbiosis. This flavonoid has also been shown to dramatically increase the ability for ORS571 to colonise *O. sativa*, *T. aestivum* and *Z. mays* (Webster *et al.*, 1998; Stone *et al.*, 2001; Jain and Gupta, 2003; Senthilkumar *et al.*, 2009). In *O. sativa*, ORS571 can colonise the entire plant, starting from the root and ending in leafy tissue (Chi *et al.*, 2005). As ORS571 is capable of free-living N_2-fixation it was believed that intracellular spaces in *O. sativa* and *T. aestivum*, once colonised, would provide an adequate environment for fixation. However, in a range of different studies, no N_2-fixation has been detected (Christiansen-Weniger, 1996; Van Nieuwenhove *et al.*, 2000; Stone *et al.*, 2001; Tsukada *et al.*, 2009). In *O. sativa*, auxin-induced root tumours have been shown to provide a limited environment for ORS571 growth and N_2-fixation (Christiansen-Weniger, 1996), showing that *O. sativa* could provide an environment suitable for N_2-fixation, albeit at a slow rate. A recent study showed that during colonisation of *S. rostrata* roots, *A. caulinodans* could transfer its genetic symbiosis island to other rhizobia present in soil, allowing them to nodulate *S. rostrata* (Ling *et al.*, 2016). This

element was named an integrative and conjugal element (ICE) and transfer was mediated by plant-produced naringenin. On detection of this flavonoid, the ICE was excised from the genome, circularised and transferred to other rhizobial species via horizontal gene transfer. It was shown this ICE was transferred to three *Mesorhizobium* species and a single *Sinorhizobium* species, but not to *Bradyrhizobium* or *Rhizobium* species (Ling *et al.,* 2016)

There has been a lack of whole-genome analysis in *A. caulinodan*s, with studies so far limited to a single comparative transcriptomic study used to compare the growth of ORS571 under symbiotic and N-fixing conditions using microarrays and subsequent Tn*5* mutant analysis (Tsukad*a et al*., 2009). Although this study gave insight into the genes and mechanism by which ORS571 forms a successful symbiosis, it did not examine colonisation and was focused on *A. caulinodans'* native host *S. rostrata.*

In summary, *A. caulinodans* is a versatile soil-borne bacterium, capable of forming a N_2-fixing symbiosis with *S. rostrata*, able to transfer this ability to other organisms and to internally and externally colonise an array of cereal plants. It acts as a PGPR of *O. sativa*, although the exact mechanism by which it promotes the growth is unknown. This *A. caulinodans* is an excellent candidate for further study into rhizosphere and root colonisation, and nitrogen fixation in free-living conditions.

1.8 – Conclusions and research goals

Root exudation plays numerous and critical roles in the lifecycles of plants and other soil-borne organisms (Badri and Vivanco, 2009), ranging from microbiome selection to nutrient uptake. The study, understanding and manipulation of root exudation to increase nutrient acquisition, disease resistance, beneficial symbioses and combat

heavy-metal toxicity, will be important in the increasing crop yields in the fight for global food security in coming decades. Despite this importance, there has been little development in technologies that allow for high-resolution *in vivo* mapping of exudation in a spatiotemporal manner until very recently. Pini *et al.,* (2017) demonstrated construction and use of luciferase-based bioreporters that allow *in vivo* mapping of key exudates from legume roots.

The first step in this project was to demonstrate the use of these bioreporters in non-leguminous plants. Their application in mapping exudation and colonisation of *O. sativa*, *Z. mays*, *A. thaliana* and *M. polymorpha* was investigated. Alongside this exudate mapping, these biosensors were tested against a suite of *O. sativa* mutant lines, as well as a variety of *A. thaliana* accessions. My aim in this study was to determine whether the suite of bioreporters could be used to measure exudation and, in the future, identify mutant lines or accessions which exhibit alterations in exudation compared to wild-type/control plants.

My second aim was to determine the genetic basis of *O. sativa* root colonisation and free-living N_2-fixation by the diazotrophic *A. caulinodans*. This was achieved by a genome-wide analysis of gene essentiality using INSeq. INSeq analysis was performed under a number of different growth conditions; free-living at 21% and 3% oxygen, N_2-fixing growth at 3% oxygen, rhizosphere growth and during root attachment. These conditions allowed identification of genes involved in root colonisation, root attachment, growth under low-oxygen conditions and during N_2-fixation.

The combination of these two experimental approaches will elucidate underlying roles that exudates play during bacterial colonisation of the rhizosphere. In particular,

linking metabolite exudation and gene essentiality will allow modelling of key determinants of colonisation.

Chapter 2

Materials and methods

2.1 - Antibiotics, media

2.1.1 – Antibiotics

	Stock concentrations	***E. coli***	***A. caulinodans* ORS571**	***R. leguminosarum* bv *viciae*. 3841**
Ampicillin (Amp)	100 mg mL^{-1}; water	100 µg mL^{-1}	100 µg mL^{-1}	-
Gentamycin (Gent)	50 mg mL^{-1}; HEPES	10 µg mL^{-1}	20 µg mL^{-1}	20 µg mL^{-1}
Kanamycin (Kan)	50 mg mL^{-1}; water	20 µg mL^{-1}	50 µg mL^{-1}	50 µg mL^{-1}
Neomycin (Neo)	80 mg mL^{-1}; HEPES	20 µg mL^{-1}	-	80 µg mL^{-1}
Nitrofurantoin (Nit)	5 mg mL^{-1}; DMF	-	5 µg mL^{-1}	5 µg mL^{-1}
Spectinomycin (Spec)	50 mg mL^{-1}; HEPES	50 µg mL^{-1}	-	100 µg mL^{-1}
Streptomycin (Str)	250 mg mL^{-1}; HEPES	25 µg mL^{-1}	-	500 µg mL^{-1}
Tetracycline (Tet)	10 mg mL^{-1}; water	10 µg mL^{-1}	5 µg mL^{-1}	5 µg mL^{-1}

Table 2.1 Antibiotic stock and working concentrations.

Antibiotics were used to select for appropriate bacterial species and strains. HEPES 1M buffer supplied by Thermo Scientific.

Rhizobium leguminosarum bv. *viciae* 3841 has chromosomal resistance to streptomycin. As such streptomycin was used for standard selection. *Azorhizobium caulinodans* ORS571 has chromosomal resistance to ampicillin and nitrofurantoin. Either ampicillin or nitrofurantoin was used for selection using concentrations shown in Table 2.1.

2.1.2 – Media

Tryptone yeast (TY) media was used for growth of *R. leguminosarum* and *A. caulinodans*. This consisted of: 3 g L^{-1} yeast extract, 5 g L^{-1} NaCl, 5 g L^{-1} tryptone (Beringer 1974). For solid medium, 17.5 g L^{-1} agar was added before autoclaving.

Universal minimum salts (UMS) was used for bacterial growth on minimal medium. Preparation for one litre of medium is as follows: 0.5 g L^{-1} $MgSO_4.7H_2O$, 0.2 g L^{-1} NaCl, 4.19 g L^{-1} MOPS, 0.5mM K_2HPO_4, and 1 mL of trace elements (0.375 g L^{-1} EDTA-Na_2, 016g L^{-1} $ZnSO_4.7H_2O$, 0.2g L^{-1} $NaMoO_4$, 0.25 g L^{1} H_3BO_3, 0.2g L^{-1} $MnSO_4.4H_2O$, 0.02g L^{-1} $CuSO_4.5H_2O$, 1 g L^{-1} $CoCl_2.6H_2O$). The pH was adjusted to 7 and medium then autoclaved. The medium was supplemented after autoclaving with 1 mL of calcium stock solution (75 g L^{-1} $CaCl_2.2H_2O$), 1 mL of iron stock solution 12 g L^{-1} $FeSO_4.7H_2O$ dissolved in 1M HCl), and 1 mL of vitamin stock solution (1 g L^{-1} Thiamine hydrochloride, 2 g L^{-1} D-Pantothenic acid calcium salt, 100 mg L^{-1} Biotin). For solid medium, 17.5 g L^{-1} agar was added before autoclaving, this was subsequently referred to as Universal minimum salts agar (UMA).

UMS and UMA were supplemented with different carbon sources dependent on the experimental set up. For growth of *R. leguminosarum,* media was supplemented to contain: 10 mM glucose and 10 mM NH_4. For growth of *A. caulinodans*, media was supplemented to contain: 20 mM succinate, 10 mM NH_4, and 300 µM nicotinic acid.

Luria-Bertani broth (LB) was used for the growth of *E. coli* strains. LB consists of: 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 5 g L^{-1} NaCl. For solid media, 14 g L^{-1} agar was added.

Distilled water agar (DWA) was used for plant germination. This consisted of milliQ water and 8 g L^{-1} agar.

O. sativa, Z. mays, Brachypodium distachyon, T. aestivum were grown axenically on Fahraeus medium (FP). This consists of: 0.1g L_{-1} $CaCl_2.2H_2O$, 0.12 g L^{-1} $MgSO_4.7H_2O$, 0.099 g L^{-1} KH_2PO_4, 0.149 g L^{-1} $Na_2HPO_4.12H_2O$, 0.005 g L^{-1} $FeC_6H_5O_7$. For every litre of medium, 1 mL of Gibsons trace was added (Gibsons trace: 2.86 g L^{-1} H_3BO_3, 2.03 g L^{-1} $MnSO_4.4H_2O$, 0.220 g L^{-1} $ZnSO_4.7H_2O$, 0.08 g L^{-1} $CuSO_4.5H_2O$, 0.08 g L^{-1} H_2MoO_4). The pH was adjusted to 6.6. High clarity bacteriological agar was used at a concentration of 10 g L^{-1}.

A. thaliana was germinated and grown axenically on modified Murashige and Skoog (MS). This medium consists of: 2.15 g L^{-1} MS medium basal salts, 10 g L^{-1} sucrose. The pH was adjusted to 6.6 using NaOH. For solid medium 10 g L^{-1} agar was added before autoclaving. Plants were germinated for 10 days before transfer to fresh plates as this media dries out quickly.

Marchantia polymorpha was grown on sterile Johnsons media consisting of 6 mM KNO_3, 1 mM $MgSO_4$, 4 mM $Ca(NO_3)_2.4H_2O$, 25 µM KCl, 10 µM H_3BO_4, 1 µM $MnSO_4.4H_2O$, 1 µM $ZnSO_4.7H_2O$, 250 nM $CuSO_4.5H_2O$, 250 nM $(NH_4)6Mo_7O_{24}.4H_2O$, 25 µM $FeSO_4.7H_2O$, 25 µM Na_2EDTA, 0.6 mM $NH_4H_2PO_4^{3-}$, 0.4 mM $(NH_4)_2SO_4$, 55.5 mM inositol, 0.533 g L^{-1} (w/v) MES. The pH was adjusted to 5.8 using 1 M KOH. For wild-type plants 1.4% agar (Sigma Aldrich) was used, but for mutants 0.8% agar was used.

2.2 – Bacterial strains, plasmids, and primers

2.2.1 – Strains

Stain	***Description***	***Reference***
*Rlv*3841	*Rhizobium leguminosarum* bv. *viciae*; Str^r	Johnston and Beringer, 1975
LMB483	*Rlv*3841 [pLMB577]; Str^r Tet^r	Pini *et al.,* 2017
LMB552	*Rlv*3841 [pLMB649]; Str^r Tet^r	Pini *et al.,* 2017
LMB590	*Rlv*3841 [pLMB684]; Str^r Tet^r	Pini *et al.,* 2017
LMB592	*Rlv*3841 [pLMB686]; Str^r Tet^r	Pini *et al.,* 2017
LMB593	*Rlv*3841 [pLMB687]; Str^r Tet^r	Pini *et al.,* 2017
LMB608	*Rlv*3841 [pLMB708]; Str^r Tet^r	Pini *et al.,* 2017
LMB610	*Rlv*3841 [pLMB710]; Str^r Tet^r	Pini *et al.,* 2017
LMB612	*Rlv*3841 [pLMB712]; Str^r Tet^r	Pini *et al.,* 2017
LMB613	*Rlv*3841 [pLMB713]; Str^r Tet^r	Pini *et al.,* 2017
LMB614	*Rlv*3841 [pLMB714]; Str^r Tet^r	Pini *et al.,* 2017
LMB617	*Rlv*3841 [pLMB705]; Str^r Tet^r	Pini *et al.,* 2017
LMB619	*Rlv*3841 [pLMB717]; Str^r Tet^r	Pini *et al.,* 2017
LMB638	*Rlv*3841 [pLMB737]; Str^r Tet^r	Pini *et al.,* 2017
LMB639	*Rlv*3841 [pLMB738]; Str^r Tet^r	Pini *et al.,* 2017
LMB667	*Rlv*3841 [pLMB757]; Str^r Tet^r	Pini *et al.,* 2017

LMB672	*Rlv*3841	Pini *et al.,* 2017
OPS0100	*Rlv*3841 [pIJ11282]; Tc^r	This work
ORS571	*A. caulinodans* wild-type	Dreyfus and Dommergues, 1981
Δ*nifA*	ORS571-Δ*nifA*	Minhyung Ryu, unpublished
DH5alpha	*E. coli*; F- *deoR endA1 recA1 relA1 gyrA96 hsdR17(rkmk+) supE44 thi-1 - phoA* Δ(*lacZYA-argF*)U169 Φ80lacZΔM15 λ	Bioline
SM10*λpir*	the the leu tonA lacY supE rica::RP4-2-Tc::Mu Km λpir	Simon *et al.*, 1983
S17-1	*E. coli* conjugative donor strain; *recA* pro *hsdR* RP4-2- Tc::Mu-Km::Tn*7* integrated into the chromosome	ATCC47055

Table 2.2 All strains used in this work

R. leguminosarum strains were grown at 28 ^{o}C. *A. caulinodans* strains were grown at either 37 ^{o}C or 28 ^{o}C. *E. coli* strains were grown at either 28 ^{o}C or 37 ^{o}C. For growth in liquid medium, cultures were shaken at 200 rotations per minute (RPM).

2.2.2 – Plasmids

Plasmid	***Description***	***Reference***
pIJ11282	Constitutive lux fusion, utilizing a *nptII* promoter; Tet^r	Frederix *et al.,* 2014
pLMB649	Constitutive lux fusion, utilizing *pTac* promoter; Tet^r	Pini *et al.,* 2017
pLMB577	Phenyalanine bioreporter; Tet^r	Pini *et al.,* 2017
pLMB684	Xylose biosensor; Tet^r	Pini *et al.,* 2017
pLMB686	*Myo*-inositol biosensor; Tet^r	Pini *et al.,* 2017
pLMB687	Sucrose biosensor; Tet^r	Pini *et al.,* 2017
pLMB708	Tartrate biosensor; Tet^r	Pini *et al.,* 2017
pLMB710	Formate biosensor; Tet^r	Pini *et al.,* 2017
pLMB712	Hesperetin biosensor; Tet^r	Pini *et al.,* 2017
pLMB713	Salicylic acid biosensor; Tet^r	Pini *et al.,* 2017
pLMB714	C4-dicarboxylate biosensor; Tet^r	Pini *et al.,* 2017
pLMB705	Mannitol biosensor; Tet^r	Pini *et al.,* 2017
pLMB717	Erythritol biosensor; Tet^r	Pini *et al.,* 2017
pLMB737	Malonate biosensor; Tet^r	Pini *et al.,* 2017
pLMB738	GABA biosensor; Tet^r	Pini *et al.,* 2017
pLMB757	Fructose biosensor; Tet^r	Pini *et al.,* 2017

pSAM_RI	*Mariner* transposon vector; Amp^r Kan^r	Perry and Yost *et al.,* 2014
pRK2013	Helper plasmid; triparental conjugation; Kan^r	(Figurski and Helinski, 1979)

Table 2.3. All plasmids used in this work

2.2.3 – Primers and adapters

Name	Sequence	Description
ARB-1-ORS571-2	GGCCACGCGTCGACTAGTCANNNNNNNNNNCATCG	First Round Primer for Arbitrary Primed PCR - ORS571
ARB-2	GGCCACGCGTCGACTAGTCA	Second round primer for Arbitrary Primed PCR
oxp0340	GCTAGACTGGGCGGTTTTATG	Sense primer pSAM_Rl mapping; Ta: 67 ^{o}C
oxp0341	CTGCAGGTAGAAACGCAAAAAG	Antisense primer pSAM_Rl mapping; Ta: 67 ^{o}C
oxp0434	CGGTTCGCTTGCTGTCCATAAAACC	Ion Torrent BioSAM with 5' Biotin TEG; Ta: 68 ^{o}C
oxp0435	CTGTCCGTTCCGACTACCCTCCCGAC	Sense M12 adapter sequence; Ta – gradient from 98 – 4 ^{o}C (Goodman *et al*., 2011)
oxp0436	GTCGGGAGGGTAGTCGGAACGGACAG	Antisense M12 adapter sequence; Ta – gradient from 98 – 4 ^{o}C (Goodman *et al*., 2011)

oxp0437	AGATCGGAAGAGCGTCGTGTAGGGAA	INSeq Adapter Sense; Ta – gradient from 98 – 4 ^{o}C (Goodman *et al*., 2011)
oxp0438	TTCCCTACACGACGCTCTTCCGATCTNN	INSeq Adapter Antisense; Ta – gradient from 98 – 4 ^{o}C (Goodman *et al*., 2011)
oxp0439	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGATATAAAACCGCCCAGTCTACTCGAGGG	IT_A_FP_1 (Ion express barcode 1); Ta: 68 ^{o}C
oxp0440	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACGATATAAAACCGCCCAGTCTACTCGAGGG	IT_A_FP_2 (Ion express barcode 2); Ta: 68 ^{o}C
oxp0441	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCGATATAAAACCGCCCAGTCTACTCGAGGG	IT_A_FP_3 (Ion express barcode 3); Ta: 68 ^{o}C
oxp0442	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCGATATAAAACCGCCCAGTCTACTCGAGGG	IT_A_FP_4 (Ion express barcode 4); Ta: 68 ^{o}C
oxp0443	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGAAGGAACGATATAAAACCGCCCAGTCTACTCGAGGG	IT_A_FP_5 (Ion express barcode 5); Ta: 68 ^{o}C
oxp0444	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCAAGTTCGATATAAAACCGCCCAGTCTACTCGAGGG	IT_A_FP_6 (Ion express barcode 6); Ta: 68 ^{o}C
oxp0445	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATTCGATATAAAACCGCCCAGTCTACTCGAGGG	IT_A_FP_7 (Ion express barcode 7); Ta: 68 ^{o}C
oxp0446	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCGATAACGATATAAAACCGCCCAGTCTACTCGAGGG	IT_A_FP_8 (Ion express barcode 8); Ta: 68 ^{o}C
oxp0447	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGCGGAACGATATAAAACCGCCCAGTCTACTCGAGGG	IT_A_FP_9 (Ion express barcode 9); Ta: 68 ^{o}C
oxp0448	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACCGAACGATATAAAACCGCCCAGTCTACTCGAGGG	IT_A_FP_10 (Ion express barcode 10); Ta: 68 ^{o}C

oxp0449	CCATCTCATCCCTGCGTGTCTCCGAC TCAGTCCTCGAATCGATATAAAACCG CCCAGTCTACTCGAGGG	IT_A_FP_11 (Ion express barcode 11); Ta: 68 ^{o}C
oxp0450	CCATCTCATCCCTGCGTGTCTCCGAC TCAGTAGGTGGTTCGATATAAAACCG CCCAGTCTACTCGAGGG	IT_A_FP_12 (Ion express barcode 12); Ta: 68 ^{o}C
oxp0451	CCTCTCTATGGGCAGTCGGTGATTTC CCTACACGACGCTCTTCCGATCT	IT_trP1_FP (universal reverse primer for all Inseq forward primers); Ta: 68 ^{o}C
oxp0687	CCATCTCATCCCTGCGTGTCTCCGAC TCAGTCTAACGGACGATATAAAACCG CCCAGTCTACTCGAGGG	IT_A_FP_13 (Ion express barcode 13); Ta: 68 ^{o}C
oxp0688	CCATCTCATCCCTGCGTGTCTCCGAC TCAGTTGGAGTGTCGATATAAAACCG CCCAGTCTACTCGAGGG	IT_A_FP_14 (Ion express barcode 14); Ta: 68 ^{o}C
oxp0689	CCATCTCATCCCTGCGTGTCTCCGAC TCAGTCTAGAGGTCGATATAAAACCG CCCAGTCTACTCGAGGG	IT_A_FP_15 (Ion express barcode 15); Ta: 68 ^{o}C
oxp0690	CCATCTCATCCCTGCGTGTCTCCGAC TCAGTCTGGATGACGATATAAAACCG CCCAGTCTACTCGAGGG	IT_A_FP_16 (Ion express barcode 16); Ta: 68 ^{o}C
oxp1475	ACACGACGCTCTTCCG	Antisense primer to amplify adapter region in final INSeq PCR product; Ta: 55 ^{o}C
oxp1476	TACTCGAGGGCGCGC	Antisense primer to amplify cassette region in final INSeq PCR product; Ta: 55 ^{o}C
oxp1477	CATCTCATCCCTGCGTGTCT	Sense primer to amplify ion torrent adapter region in final INSeq PCR product; Ta: 55 ^{o}C
oxp1478	CTCTATGGGCAGTCGGTGA	Antisense primer to amplify universal adapter region in final INSeq PCR product; Ta: 55 ^{o}C

Table 2.4. All primer and adapter sequences used in this work. Ta – primer annealing temperatures.

2.3 – Molecular biology techniques

2.3.1 – DNA isolation

For all genomic DNA extractions, between 1 x 10^9 and 2 x 10^9 cells were used (unless otherwise stated). DNA samples that required purification and/or concentration, the DNA Clean and Concentrator Kit (Zymo research) was used. Genomic DNA was isolated from *R. leguminosarium* and *A. caulinodans* using the DNeasy Blood and Tissue Kit (Qiagen) following manufacture's protocol. Plasmid DNA was extracted from *E. coli* using a GeneJET Plasmid Miniprep Kit (Thermo Scientific) following manufacturer's protocol.

For DNA extraction and purification from agarose gels the GENEJet Gel Extraction kit (Thermo Scientific) was used following manufacturer's protocol.

2.3.2 - PCR

All primers for PCR reactions were designed using Geneious R8 or above. The melting temperature (T_m) was determined by Geneious software, and annealing temperatures calculated as 5 ^oC below the T_m. Synthesis of primers was performed by Eurofins MWG Operon. A Verti® Applied Biosystems thermocycler was used for all reactions.

GoTaq® Green Master Mix (Promega) was used for all mapping PCR reactions. The manufactures suggested volumes and concentrations were used, for a 25 µl reaction: 12.5 µl GoTaq® Green Master Mix, 1 µl reverse primer (final concentration 0.5 µM), 1 µl forward primer (final concentration 0.5 µM), 100 ng DNA template, nuclease-free water to reach final volume of 25 µl. For different reaction volumes the quantities of reaction components was adjusted accordingly (e.g. a 50 µl reaction used 25 µl Green Master Mix), template DNA was maintained at 100 ng and final concentrations of primers was maintained at 0.5 µM. Thermocycler conditions were set according to

manufacturer's specification: Denaturation temperature of 98 ^{o}C, annealing temperature was set according to the T_m of primers used, extension temperature of 72 ^{o}C , (60 s was allowed for each 1 kb of DNA to be amplified).

Phusion® High-fidelity polymerase (Thermo Scientific) was used for error-free amplification of DNA products. The manufacturer's suggested volumes and concentrations were used. For a 20 µL reaction: 0.2 µL Phusion DNA polymerase, 4 µL 5x Phusion GC buffer, 0.4 µL dNTPs, 50 – 100 ng template DNA, 2.5 µL reverse primer (final concentration 0.5 µM), 2.5 µL forward primer (final concentration 0.5 µM), nuclease-free water to reach final volume of 50 µL. For different reaction volumes the quantities of reaction components was adjusted accordingly (e.g. a 50 µl reaction used 1.5 µl Phusion polymerase etc.), template DNA was maintained at 100 ng and final concentrations of primers was maintained at 0.5 µM. Thermocycler conditions were set according to manufacturer's specification: Denaturation temperature of 98 ^{o}C, annealing temperature was set according to the T_m of primers used, extension temperature of 72 ^{o}C , (30 s was allowed for each 1 kb of DNA to be amplified).

2.3.3 – Arbitrary-primed PCR

Arbitrary-primed PCR was used to determine the site of *mariner* transposon insertions in the genome of *A. caulinodans*. For this, primers specific to one end of the transposon (oxp0340 & oxp0341) were designed, along with a nested arbitrary primer (ARB-1-ORS571-2) that can bind across the whole *A. caulinodans* genome at low annealing temperatures. This was achieved by adding to the 3' end of the arbitrary primer the most frequent, non-repetitive and non-palidromic, 5 bp sequence from the ORS571 genome. GoTaq® Green Master Mix (Promega) was used for all reactions. The first round of PCR utilized: 2 µL of arbitrary-primer one (final concentration of 0.5 µM), 2 µL transposon specific primer (final concentration of 0.2

μM), 25 μL master mix and mQH_2O for a final volume of 50 μL. Thermocycler conditions: 95 ^oC for 5 minutes, followed by 6 cycles of 95 ^oC for 30 seconds, 30 ^oC for 30 seconds and 72 ^oC for 90 seconds. This was followed by 30 cycles of 95 ^oC for 30 seconds, 45 ^oC for 30 seconds and 72 ^oC for 2 minutes, followed by a single step of 72 ^oC for 4 minutes. DNA was purified using the DNA Clean and Concentrator Kit (Zymo Research). The second round of PCR utilizes a primer specific for the transposon, and a nested arbitrary-primer (ARB-2) specific to the arbitrary-primer used in round 1. This reaction utilised 1/10 of the final concentration of product from round one. 0.2 μM of each primer was used, and a reaction volume of 50 μL. The Thermocycler was set for: 95 ^oC for 1 minute, followed by 30 cycles of 95 ^oC for 30 seconds, 52 ^oC for 30 seconds, and 72 ^oC for 2 minutes, followed by a single step of 72 ^oC for 4 minutes. Products were purified using gel electrophoresis and extracted using the GENEJet Gel Extraction kit (Thermo Scientific).

2.3.4 - Gel electrophoresis

Agarose gel electrophoresis was used to confirm plasmid construction (e.g. a cloned fragment inserted), as well as the separation of PCR products and restriction digests. Electrophoresis made use of 0.9% agarose gels (Sigma Aldritch) in TAE buffer (400 mM Tris acetate, 1 mM EDTA). Gels were run at 90 mV for 30 – 45 minutes. For very small fragments the agarose concentration was raised to 1.3 % and run at 85 mV for 50 minutes. DNA was stained with either ethidium bromide for a final concentration of 0.5 μg mL^{-1} or Sybr® Safe at 1:100000 (Invitrogen). Stains were added to molten agarose before casting. All gels were visualised using a GelDoc EZ System (Bio-Rad).

2.3.5 – Restriction digests and DNA ligations

All DNA (genomic, plasmid, PCR) was purified to remove any contaminants (buffers, primers, enzymes) prior to restriction digests. Restriction endonucleases and

respective buffers (NEB or Thermo Scientific) were used following manufacturer's recommended protocols.

T4 DNA Ligase and requisite buffers (Thermo Scientific) was used for all DNA ligations following the manufacturer's recommended protocol. Ligations were carried out at room temperature overnight.

2.3.6 – Conjugations

For conjugal transfer of plasmids from *E. coli* to *R. leguminosarum*, tri-parental mating was used. The helper plasmid pRK2013, maintained in *E. coli,* (maintained in DH5alpha) (Figurski and Helinski, 1979) was used in concert with the donor and recipient strains, and appropriate antibiotics were used as needed. *Rlv*3841 was grown for 3 days on a TY slope prior to conjugation. Both the *E. coli* donor and helper strains were grown for 24 hours prior to conjugation in 10 mL of LB medium. On the day of conjugation 400 µL of each *E. coli* strain was sub-cultured into 10 mL of LB and grown at 37 ^{o}C for 4 hours shaking at 100 RPM. *Rlv*3841 slopes were washed in 3 mL of sterile TY. The *E. coli* and *Rlv*3841 were pelleted at 4500 RPM (~3500 x g) for 10 minutes. Supernatants were discarded and the pellets re-suspended in fresh TY, before re-centrifugation. This was repeated once more to ensure all antibiotics were removed. 1 mL of TY was used for the final resuspensions. The recipient rhizobium (400 µL), donor *E. coli* (400 µL) and helper (200 µL) were mixed. This mixture was centrifuged at 6000 RPM (~3500 x g) for 10 minutes. The resulting pellet was re-suspended in 30 µL of TY. This was then pipetted onto a sterile nitrocellulose filter placed onto a TY plate. The plate was incubated overnight at 28 ^{o}C. Using sterile tweezers, the filter was transferred to a 15 mL Falcon tube and vortexed to re-suspend the bacterial culture. This suspension was then serially diluted and 100 µL of each dilution spread onto TY plates containing selection antibiotics. These plates were grown for 3 days at 28 ^{o}C. Control

plates of the donor *E. coli*, recipient rhizobia and helper were also spread. For *Rlv*3841 Streptomycin 500 µg mL^{-1} was used to counter-select against *E. coli*. The antibiotic resistance marker carried on the donor plasmid was used for the counter selection of *Rlv*3841 which had not received a plasmid. Positive single colonies from the dilutions were purified and the presence of the plasmid further confirmed by PCR. Positive strains were stocked at -80 ^{o}C in 22.5% glycerol.

For conjugations into *A. caulinodans,* the same protocol as above was used. Instead of placing the mixture of recipient, donor and helper onto a nitrocellulose filter, they were placed directly onto a TY plate. Nitrofurantoin 5 µg mL^{-1} was used to counter-select against *E. coli*. The antibiotic resistance marker carried on the donor plasmid was used for the counter selection of *A. caulinodans*. Positive single colonies from the dilutions were purified and the presence of the plasmid further confirmed by PCR. Positive strains were stocked at -80 ^{o}C in 22.5% glycerol.

2.4 – Plant experiments

All bacterial inocula were grown on UMA slopes with appropriate antibiotics, for 3 days prior to inoculation. These were washed 3 times before inoculation to remove any trace of antibiotics. The inocula was diluted in UMS to a final concentration of 5 x 10^{7} cells mL^{-1}, using a pipette, roots were inoculated with up to 1 mL of inoculum (depending on plant system).

2.4.1 – Growth of *Oryza sativa*

Seeds were de-husked and ~20 seeds were placed in a 50 mL Falcon tube. Seeds were sterilised in 70% ethanol for 20 – 30 seconds. The ethanol was drained and the seeds immersed in 2.6% sodium hypochlorite (bleach) solution with a drop of Triton X100 (Sigma Aldrich) and placed on an orbital rotator for 20 minutes. The

bleach/Triton solution was removed, and the seeds washed a minimum of 5 times in sterile distilled water to ensure all traces of bleach/Triton was removed.

For growth on agar, seeds were germinated on 0.8% DWA for 3 days at 28 ^{o}C with no light until roots reached a minimum length of 20 mm. FP slant plates were prepared; these are 120 x 120 mm square Petri plates with FP poured in such a way that a slope is made. Once germinated, 3 seedlings were embedded into the middle line of the slope plates so that the roots lay across the thickest part of the agar and the shoots across the thinnest. The seeds were sealed with molten agar, this stopped inoculum pooling in the wells created by seeds. Roots were then inoculated with 200 µL of 5 x 10^{7} cells mL^{-1} of bacterial inoculum. Roots were then covered with a sheet of sterile, clear cellophane and the plates sealed with Micropore tape. The lower half of the plate was wrapped in aluminium foil to cover the roots. Plates were placed upright, at a slight angle, in a growth cabinet set to 28/23 ^{o}C (day:night) with a photoperiod of 16:8 (day:night).

For growth in boiling tubes, *O. sativa* was sterilized as above. A boiling tube was half-filled with washed fine vermiculite and 30 mL of rooting solution added. Rooting solution comprised 4 mM Na_2HPO_4, 3.7 mM K_2PO_4, 1 mM $CaCl_2$, 800 µM $MgSO_4$, 100 µM KCl, 35 µM H_3BO_3, 10 µM Fe EDTA, 9 µM $MnCl_2$, 0.8 µM $ZnCl_2$, 0.5 µM $Na2MoO_4$, 0.3 µM $CuSO_4$, 10 mM KNO_3. Excess rooting solution was poured off. Tubes were sealed with a foam bung and autoclaved. A single sterile *O. sativa* seed was planted a maximum of 5 mm below the surface of the vermiculite and the tubes resealed with a foam bung. Tubes were placed in a growth cabinet set to 28/23 ^{o}C (day:night) with a photoperiod of 12:12 (day:night). After 10 days of growth, each tube was inoculated with 1 mL of 5 x 10^{7} cells mL^{-1} of bacterial inoculum.

2.4.2 – Growth of *Zea mays*

Zea mays was sterilised using the same protocol as *O. sativa*. Ten seeds were sterilised in a 50 mL Falcon tube following the protocol outlined for rice. A single *Z. mays* seedling was placed on each FP slope. A hole was cut in the middle of the top side of the plate and the shoot pulled through. Roots were then inoculated with 1 mL of 5×10^7 cells mL^{-1} of bacterial suspension. Roots were covered with a sheet of sterile, clear cellophane. Plates were sealed with Micropore tape, and the lower half of the plate wrapped in aluminium foil to cover the roots. Plates were placed upright, at a slight angle, in a growth cabinet set to a 28:23 ^{o}C (day:night) with a photoperiod of 12:12 (day:night).

2.4.3 – Growth of *Arabidopsis thaliana*

Arabidopsis thaliana seeds were sterilised using chlorine gas. A small number of seeds were placed into a 1.5 mL microfuge tubes and the lids left open. Tubes were then placed in a gas-tight box in a fume hood. 30 mL of 13% sodium hypochlorite was added to a glass beaker and placed into the box. Using a serological pipette 3 mL of 34% hydrochloric acid was added and the box quickly sealed. The addition of HCl to NaClO produces Cl_2 gas. Seeds were left to sterilise for 5 hours. Seeds were aseptically transferred to MS plates and incubated at 4 ^{o}C for 48 hours to break seed dormancy. Plates were then transferred to a tissue culture growth room set to 23 ^{o}C with a photoperiod of 16:8(day:night). During growth fresh MS plates were prepared with sterile filter paper (Whatman) covering the agar surface. After 10 days of growth the seedlings were transferred to the fresh plates so that their roots lay across the surface of the filter paper, three plants were placed onto each plate. The filter paper was then evenly inoculated with 1 mL of 5×10^7 cells mL^{-1} of bacterial inoculum. Plants were transferred back into the tissue culture growth room.

2.4.4 – Growth of *Marchantia polymorpha*

Sterile *M. polymorpha* cultures were acquired from the Dolan Lab (University of Oxford). To propagate cultures, a small section of the growing plant was cut away and placed onto Johnsons media. Plates were sealed with Micropore tape and placed in a growth cabinet 23 ^{o}C with a photoperiod of 16:8 (day:night). For long-term storage plates were stored in a growth cabinet at 16 ^{o}C with 24 hour light.

For experiments, plants were grown from gemmae (asexual vegetative propagules). Using sterile tweezers, single gemmae were removed from gemmae cups on a mature plant and placed in 25 µL water droplets on agar plates. Sixteen droplets were evenly spaced on 120 x 120 mm square Petri plates. Plates were sealed with Micropore tape and placed in a growth cabinet 23 ^{o}C with a photoperiod of 18:6 (day:night). Plants were commonly inoculated between 14 to 35 days of growth, each plant was inoculated with 200 µL of 5 x 10^{7} cell mL^{-1} of bacterial culture.

2.4.5 – Luciferase imaging

All plant-based luciferase imaging was performed using either the NightOWL or NightSHADE cooled CCD camera systems (Berthold Technologies). Plants were removed from their growth environments immediately before imaging and placed back immediately afterwards. Plates were centred in the imaging system using a custom 3D printed frame. Imaging conditions were as follows: 120 second exposure, high gain, slow readout, cosmic suppression enabled, background correction enabled. Data were analysed using the Indigo software (Berthold Technologies).

2.5 – INSeq

2.5.1 – Mutant generation

Mutant libraries of ORS571 were generated using the pSAM_RI plasmid (Fig 2.1). pSAM_RI was conjugally transferred into ORS571 as outlined above. Three biological replicates were produced in total. The conjugation drop was transferred

into 1 mL of TY using a sterile inoculation loop before selection. 450 µL of glycerol was added to the re-suspended culture, 100 µL taken for serial dilutions, and the remaining culture stored at -80 ^{o}C. Dilutions of the conjugations, alongside controls were spread onto TY Kan^{50} Nit^{5} plates in triplicate and grown at 28 ^{o}C for 48 hours. After 48 hours, the number of colonies was counted and the concentration of cells in the stocks determined. Stocks were selected using large square Petri plates (240 x 240 mm) of TY Kan^{50} Nit^{5}. A maximum of 2 x 10^{5} cells was spread onto each large plate, and multiple plates used until the whole volume of the stocks used. These were then incubated at 28 ^{o}C for 48 hours, until single colonies could be seen. The selected libraries were removed using 2 mL of UMS and gently scraping the colonies from the plates using a sterile razor blade. Each library was pooled, and the appropriate amount of glycerol added to reach a concentration of 22.5%. Serial dilutions of each library were used to determine their final concentrations. These libraries were then used for all subsequent INSeq experiments, and were labelled as Input libraries 1, 2 or 3 for sequencing.

2.5.2 – *Oryza sativa* growth with INSeq libraries

Oryza sativa plants were grown in boiling tubes as outlined in 2.4.1. Sixty plants were grown and split into three fractions of twenty plants. Each fraction was then inoculated with a separate INSeq mutant library (library 1, 2 or3) to yield 3 biological replicates (1, 2 and 3). Each individual plant was inoculated with 1 x 10^{5} cells. Bacteria were isolated 5 days post inoculation from the vermiculite and the roots. Bacterial isolation from roots was done in batches of five plants: firstly, roots were removed from the boiling tubes and vermiculite, dipped into sterile PBS (this removed excess vermiculite on the root surface) and placed into a 50 mL Falcon tube containing 25 mL of PBS. Roots were then vortexed for 1 minute before being transferred to (along with the 25 mL of PBS) a pestle and mortar and macerated. Macerated roots were filtered through two layers of sterile muslin cloth into a 50 mL

Falcon tube to remove large pieces of plant matter and vermiculite. The resulting cultures were spun at 1000 RPM (~150 x g) for 1 minute to pellet large particles of vermiculite and plant matter, the supernatant containing bacteria was transferred to a 2 L conical flask. Each 20 plant fraction was pooled into a single flask and TY medium added for a final culture volume of 500 mL. In total this yielded three 500 mL cultures labelled Root-attached 1, 2 and 3. Bacteria were isolated from the vermiculite (rhizosphere) after the roots were removed, 15 mL^{-1} of PBS was added to each boiling tube and each tube vortexed for 1 minute. The vortexed vermiculite/PBS was strained through two layers of sterile muslin cloth into 50 mL falcon tubes. The resulting cultures were spun at 1000 RPM (~150 x g) for 1 minute to pellet large particles of vermiculite and plant matter, the supernatant containing bacteria was transferred to a 2 L conical flask. Each 20 plant fraction was pooled into a single flask and TY medium added for a final culture volume of 500 mL. In total this yielded three 500 mL cultures labelled Rhizosphere 1, 2 and 3. To increase levels of bacterial DNA, all cultures were grown at 28 ^{o}C for 15 hours. In total six libraries were produced: Three from the vermiculite (Rhizosphere) 1r, 2r and 3r; and three from the roots (Root-attached) 1ra, 2ra and 3ra.

2.5.3 – Library preparation

Library perpetration followed the protocol outlined by Goodman *et al*. (2011), with a few minor modifications. During linker ligation, the LIB_AdaptT_(barcode) primers (used to barcode each sample individually) were replaced with oxp0437 and oxp0438, which were used for all samples. The primers used during the final PCR amplification were replaced with a universal primer (oxp0451) and each sample was barcoded uniquely using primers (oxp0439 to 0450).

Library preparation utilised a maximum of 2000 ng of bacterial DNA isolated from libraries grown under a condition of interest. This DNA was subjected to linear PCR

with a biotinylated primer (oxp0434), which acts to amplify the end-segment of the transposon (including the MmeI containing terminal inverted repeat regions) and adjacent genomic DNA. The biotinylation of target DNA allows these single-stranded fragments to be bound to streptavidin-coated magnetic beads. These beads then allow for concentration and purification of target DNA away from unwanted background genomic DNA. Once this initial purification was completed, single-stranded fragments underwent double-strand synthesis before being digested with the type IIS restriction enzyme MmeI. This enzyme cuts 16 bp downstream of its recognition site, resulting in a fragment of a specific length, containing the insert and 16 bp of genomic DNA. After digestion, 3'-adapter sequences (oxp0437 and oxp0438) were ligated to the fragment onto which Ion-Torrent-specific sequencing adapters were added. A sample-specific barcode (oxp0439-450 and oxp0688-690) was then added to the 5'-end of the sequence before a final 5' ion-torrent-specific adapter (oxp0451) was added. Addition of adapters and barcodes was done via a single PCR reaction, which also acts to amplify the DNA concentration. Thermocycler conditions: 94 °C for 2 min, followed by 18 cycles of 94 °C for 15 s, 60 °C for 1 min, 68 °C for 2 min and a final cycle of 68 °C for 4 min. This method resulted in 187 bp fragments that consist of the transposon, genomic DNA, adapter, and barcodes. After the final PCR step, samples were purified using the DNA Clean and Concentrator Kit (Zymo Research). The 187 bp fragment was then isolated using an E-Gel SizeSelect 2% Agarose Gel (Invitrogen) following the manufacturer's instructions. After size selection, the DNA Clean and Concentrator Kit (Zymo Research) was used to purify DNA.

DNA concentrations were checked using an Agilent 2100 Bioanalyzer System and the High Sensitivity DNA Kit (Agilent), following the manufacturer's protocol. Spectra were analysed using the 2100 Expert Software (Agilent). Samples were diluted to 75

pM. Diluted samples were combined for a final volume of 50 µL and concentration of 75 pM. Only 25 µL was used for sequencing.

2.5.4 – INSeq sequencing and analysis

Library preparation was fully automated using the Ion Chef System (Thermo Scientific) and the Ion PI Chip Kit V3 (Thermo Scientific). Reagents and sample concentrations were used according to manufacturer's instructions. Sequencing was performed using an Ion Proton System (Thermo Scientific). A run plan was created on an Ion PGM Server (Thermo Scientific) that contained information regarding the type of sequencing, number of samples, unique barcodes associated with samples and kit used.

Once sequenced all reads containing the same 5'-specific barcode were consolidated and cross-referenced to the run plan so that each set of reads was associated with a known sample. Once consolidation was completed, the Ion-Torrent software trimmed the 3'- and 5'-Ion-Torrent adapter, along with the sample specific barcode for each sample file. Data was subsequently output as FASTQ files for each library defined in the run plan, these files contain a list of 177 bp reads, made up of the transposon region, captured genomic DNA and adapters. Using the FASTQ output files as the source, reads were trimmed using 'cutadapt' to remove the 3' adapters followed by the removal of the 5' *mariner* element. Before each trimming step any previously untrimmed reads were discarded. Once trimmed, a 16-17bp read is left that is comprised solely of the captured genomic DNA. Using Bowtie these trimmed reads were aligned against the ORS571 genome. Using 'SAMtools' and 'bamToBed', the data files were converted into a '.bed' format. These files now contain TA site genomic coordinates for each insertion, using this the number of insertions mapping to each individual TA site in the genome was calculated using a custom written perl script (personal communication - Vinoy Ramachandran). A

Hidden Markov Model (HMM) analysis was used to categorise TA sites into one of four essentiality states: essential, growth defective, growth neutral and growth advantageous states, this analysis was achieved by utilising the 'tn-hmm.py' script. Genes were categorised into one of the four states listed above using the script 'process_genes.py', genes were linked to their product by using the script 'match_id.pl'. All HMM data were organised within Microsoft Excel, the raw '.wig' files (containing read counts for each TA site) were visualised within Integrative Genomic Viewer (IGV).

All INSeq analysis was performed on an in-house UNIX server. Scripts used are outlined in Supplementary 6. TRANSIT (DeJesus *et al.*, 2015) was used for resampling analysis and was performed following recommended procedures. A 5% trim was used for each gene to remove reads on the borders of each gene.

2.6 – RNA-Seq

2.6.1 – Plant, bacterial growth and RNA isolation

Free-living *A.caulinodans* ORS571 cultures were grown for 48 hrs in 50 mL of liquid TY medium at 28 ^{o}C and shaken at 200 RPM. Three separate cultures were grown to produce three biological replicates.

Oryza sativa cv. IR64 was grown in boiling tubes as outlined in 2.4.1. Sixty plants were grown and split into three fractions of twenty plants. Each fraction was inoculated with a separate *A. caulinodans* ORS571 culture (grown on TY medium), this yielded 3 biological replicates labelled 1, 2 and 3. Each individual plant was inoculated with 5 x 10^5 cells, bacteria were isolated 5 days post inoculation from the vermiculite and the roots.

Prior to bacterial isolation all work surfaces were sterilised with 70% ethanol and sprayed down with RNAaseZap™ (Invitrogen). Homemade RNA Later (700g/L^{-1} ammonium sulphate, 25 mM sodium citrate and 10 mM EDTA) was used to precipitate out proteins thus stopping RNAse activity during bacterial and RNA isolation.

Before isolation of bacteria from the vermiculite or roots, 15 mL of RNA Later was added to each boiling tube. Roots were first removed from the vermiculite and dipped into RNA Later (this removed excess vermiculite on the root surface), the roots were then placed into 50 mL Falcon tubes (5 roots per falcon tube) and vortexed for 5 minutes in 20 mL of RNA Later, the roots were removed and the resulting cultures were spun at 1000 RPM (~150 x g) for 1 minute to pellet large particles of vermiculite and plant matter. The supernatants were carefully transferred to sterile 36 mL Nalgene centrifuge tubes. Bacteria was isolated from the vermiculite by vortexing for 1 minute with 15 mL of RNA Later, the vortexed vermiculite/RNA Later was strained through two layers of sterile muslin cloth into 50 mL Falcon tubes. The resulting cultures were spun at 1000 RPM (~150 x g) for 1 minute to pellet large particles of vermiculite and plant matter. The supernatants were carefully transferred to sterile 36 mL Nalgene centrifuge tubes. All cultures were then spun-down at 10000 RPM (~12000 x g) at 4 ^{o}C for 10 minutes; the supernatant was carefully discarded so as not to disturb the pellet.

For the free-living cultures, 12 mL of culture was transferred to 36 mL Nalgene bottles and 24 mL of RNA Later added. The cultures were centrifuged at 10000 RPM (~12000 x g) at 4 ^{o}C for 10 minutes and the supernatants discarded.

All samples were treated the same in regards to RNA isolation. RNA was isolated using an RNeasy plus Mini Kit (Qiagen). Prior to isolation, 4 mL of RLT plus buffer

was transferred to a 15 mL Falcon tube and 40 µL of β-mercaptoethanol (ME) added. Pellets were resuspended in 250 µL of 10 mM Tris-cl (pH 8.0), and transferred to ribolyser tubes chilled to 4 ^{o}C containing: 0.1 mm diameter glass, 0.1 mm diameter zirconium beads and 350 µL RLT + ME. Cells were lysed in a ribolyser (FastPrep FP120, Thermo scientific) at speed 6.5 for 30 seconds then chilled in ice for 1 minute, this step was repeated for a total lysis time of 90 seconds. Tubes were incubated on ice for 3 minutes before being centrifuged at 13000 RPM (~15000 x g) at 4 ^{o}C for 3 minutes. The supernatant (homogenate) was carefully transferred to a gDNA eliminator spin column and centrifuged at 10000 RPM (~8000 x g) for 30 seconds. To the flow through 1.5 times its volume in 100% ethanol was added. Up to 700 µL of the sample (including any precipitate) was transferred to an RNease Mini spin column and centrifuged for 15 seconds at 10000 RPM (~8000 x g), the flow through was discarded, this step was repeated until the whole sample had passed though the column. The column was placed into a clean collection tube and the membrane washed with 500 µL of buffer RPE by centrifugation at 10000 RPM (~8000 x g) for 2 minutes, this step was repeated once with the centrifugation time reduced to 15 seconds for a total of two washes. The column was placed into a clean collection tube and centrifuged at 10000 RPM (8000 x g) for 1 minute to dry the membrane. To elute the RNA the column was placed in a sterile, RNase-free 1.5 mL microfuge tube, 30 µL of RNase-free water added to the column and the column centrifuged at 10000 RPM (~8000 x g) for 1 minute. Each plant fraction was pooled, in total nine libraries were produced: Rhizosphere 1, 2 and 3, Root-attached 1, 2 and 3, and Free-living 1, 2 and 3. RNA was quantified using an Experion™ Automated Electrophoresis Station (Bio-Rad) and Experion™ RNA High Sensitivity analysis kit (Bio-Rad).

2.6.2 – RNA-Sequencing library preparation

RNA samples underwent purification and conversion to cDNA as outlined below, Figure 2.1 shows an overview of the workflow. After each step RNA concentrations were checked using an Experion™ Automated Electrophoresis Station (Bio-Rad) and Experion™ RNA High Sensitivity analysis kit (Bio-Rad).

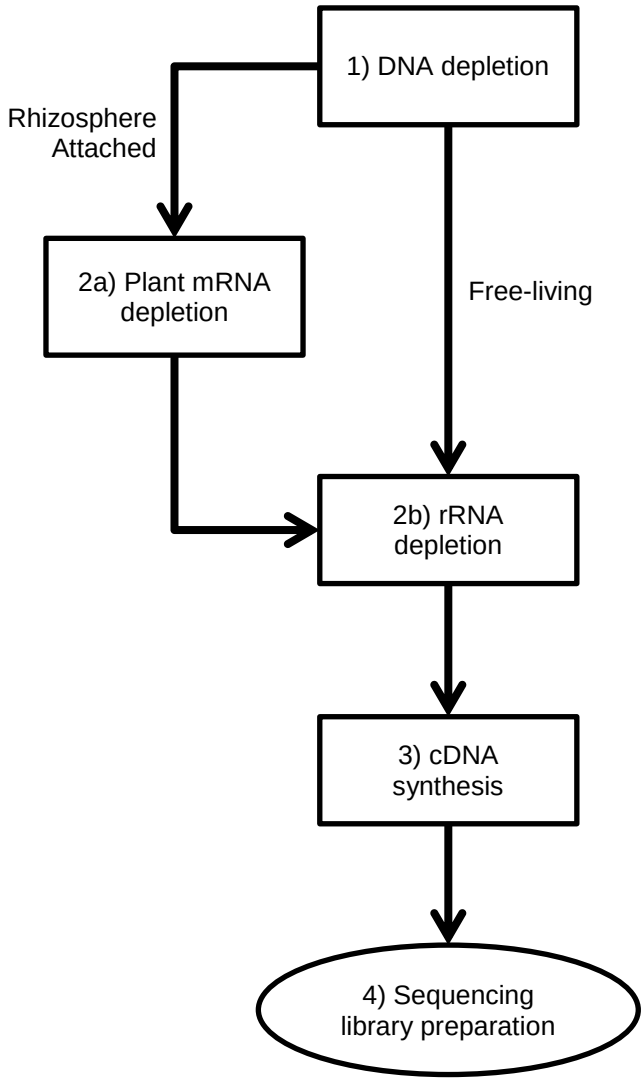


Figure 2.1 Work-flow of purification of RNA samples for RNA-Seq.

Any contaminating DNA present in the RNA samples was depleted using a TURBO DNA-*free*™ Kit (ambion®), the manufacturer's protocol and suggested volumes and concentrations were used. The three Rhizosphere samples were combined due to

low RNA yields, as were the three Root-attached samples, this yielded five samples: Free-living 1, 2 and 3, Rhizosphere, and Root-attached.

After DNA depletion, the Rhizosphere and Root-attached samples were depleted of any contaminating plant RNA. This was achieved through the use of Oligo $d(T)_{25}$ magnetic beads (NEB) that bind the poly A tail of eukaryotic mRNA. A binding buffer was prepared using RNase free water, its final composition was 20 mM Tris-HCl, 1 M LiCl and 2mM EDTA. For every 50 µg of total RNA, 100 µL of Oligo d(T) beads were used. Beads were first washed with 200 µL of binding buffer three times using a magnetic particle concentrator. An equal volume of RNA and binding buffer, a maximum of 50 µL each, was mixed and the RNA denatured at 70 ^{o}C for 5 minutes and then cooled on ice. The binding buffer was removed from the washed beads and the RNA/buffer mix added, these were then incubated at room temperature (~23 ^{o}C) on an orbital shaker at 10 RPM for 20 minutes. Beads and supernatant were separated using a magnetic particle concentrator and the supernatant transferred to RNAse-free tubes. The RNA was subsequently concentrated using a RNA Clean & Concentrator kit (Zymo Research). The manufacturer's protocol and recommended volumes were used.

All samples underwent rRNA depletion using a Ribo-Zero® rRNA Removal Kit (illumina®), the manufacturer's protocol and suggested volumes and concentrations were used. These steps yielded pure bacterial mRNA.

An Ion Total RNA-Seq Kit v2 (Thermo Scientific) and an Ion Xpress™ RNA-Seq Barcode 01-16 Kit (Thermo Scientific) was used to synthesise and barcode cDNA libraries from the purified mRNA. These kits are used in tandem and the manufactures protocol for rRNA-depleted RNA samples was followed. This yielded five barcoded samples: Free-living 1, 2 and 3, Rhizosphere, Root-attached.

Once cDNA libraries were produced the DNA concentrations were checked using an Agilent 2100 Bioanalyzer System and the High Sensitivity DNA Kit (Agilent), following the manufacturer's protocol. Spectra were analysed using the 2100 Expert Software (Agilent). Samples were then diluted to 75 pM and combined for a final volume of 50 µL and concentration of 75 pM.

2.6.3 – RNA-Seq sequencing and analysis

Library preparation was fully automated using the Ion Chef System (Thermo Scientific) and the Ion PI Chip Kit V3 (Thermo Scientific). Reagents and sample concentrations were used according to manufacturer's instructions. Sequencing was performed using an Ion Proton System (Thermo Scientific). A run plan was created on an Ion PGM Server (Thermo Scientific).

Once sequenced, all reads containing the same 5'-specific barcode were consolidated and cross-referenced to the run plan so that each set of reads was associated with a known sample. Once consolidation was completed, the Ion-Torrent software trimmed the 3'- and 5'-Ion-Torrent adapters, along with the sample specific barcode for each sample file. Data was subsequently output as FASTQ files for each library defined in the run plan. These FASTQ files contained sequenced reads up to 200 bp in length. Reads were mapped to the *A. caulinodans* ORS571 genome and the number of reads for each gene counted using EDGE-pro (Magoc *et al.*, 2013). This analysis was performed on a UNIX server.

2.7 – Miscellaneous assays

2.7.1 – Bacterial growth

A. caulinodans growth was measured under atmospheric oxygen, 3% oxygen (non N_2-fixing) and 3% oxygen (N_2-fixing). In all these assays ORS571 was used. UMS or UMA medium was supplemented with 20 mM succinate, 10 mM ammonium chloride, vitamin solution (1 g L^{-1} thiamine hydrochloride, 2 g L^{-1} D-pantothenic acid calcium salt, 100 mg L^{-1} biotin) and 300 µM nicotinic acid. For measurement of N_2-fixation, ammonium chloride was excluded. For solid medium 17.5 g L^{-1} agar was added, if N_2-fixation was being measured, agar was replaced with 6 g L^{-1} agarose. In all cases, ORS571 was grown at 37 ^{o}C.

For growth in liquid culture, ORS571 was grown on UMA slopes for 2 days at 37 ^{o}C, and washed with 3 mL of UMS. Bacterial suspensions were spun-down at 4500 RPM (~3500 x g)and the pellet re-suspended in 1 mL UMS. This washing step was repeated twice. Liquid cultures were inoculated to give a final OD^{600} of 0.01. These cultures were grown at 37 ^{o}C shaking at 200 RPM. Optical density was recorded over a 48 hour period.

For N_2-fixing growth, ORS571 was spread onto UMA (0.6% agarose) and grown in a controlled environment cabinet (Belle) for 48 hours at 37^{o}C, 3% oxygen, 200 RPM. N_2-fixation was measured using an acetylene reduction assay. The controlled environment cabinet was flushed with N to reach the desired oxygen concentration, and the temperature maintained with a heat bar.

R. leguminosarum was grown routinely on TY medium at 27 ^{o}C for 3 days. Bacteria containing *lux* fusions were grown on UMA supplemented with 20 mM succinate or pyruvate, 10 mM ammonium chloride and vitamins (Pini *et al*., 2017). Luminosity was checked using the NightOWL imaging system (Berthold Technologies). Imaging

conditions were as follows: 20 second exposure, high gain, slow readout, cosmic suppression enabled, background correction enabled. The data were analysed using Indigo software (Berthold Technologies).

2.7.2 – Acetylene reduction

An acetylene reduction assay was used to determine the activity of nitrogenase (thus N_2-fixation). This assay measures reduction of acetylene (C_2H_2) into ethylene (C_2H_4). ORS571 was grown on UMA slopes (20 mM succinate, vitamins solution and 300 µM nicotinic acid) solidified with 0.6% agarose. Slopes were grown under 3% oxygen at 37^{o}C for 2 days, sealed with a foam bung. After 2 days the foam bung was replaced with an airtight lid containing a resealing rubber gasket. Acetylene was introduced into each universal bottle for a final concentration of 2 %. Each bottle was incubated for 24 hours at 37 ^{o}C. After one hour and after 24 hours, 1 mL of gas was removed from each bottle using a syringe and needle. These 1 mL samples were passed through a Clarus® 480 gas chromatograph (Perkin-Elmer) in order to determine the acetylene and ethylene levels.

2.7.3 – Biosensor sensitivity assays

The sensitivity of bioreporters was determined by measuring the minimum concentration of an inducer molecule which yielded a detectable response. *Rhizobium leguminosarum* bioreporters were grown for 72 hours on UMA slopes (20 mM succinate or 20mM glucose and vitamins solution) with appropriate antibiotics before being washed with 3 mL of UMS. Bacterial suspensions were spun-down at 4500 RPM (~3500 x g) and the pellet re-suspended in 1 mL UMS. This washing step was repeated twice. Cultures were suspended in 50 mL of molten UMA (vitamins solution) at 42 ^{o}C for a final concentration of 1 x 10^{8} cells mL^{-1}. The medium was poured into a 120 x 120 mm square Petir plate and allowed to solidify. 25 µl droplets of inducing molecules ranging from 10 mm to 1 µM were spotted onto the agar

surface, five replicates for each concentration and four different concentrations per plate was used alongside five distilled water controls. Plates were incubated at 28 ^{o}C and imaged at two and four hours. The erythritol biosensor was tested differently and grown in UMS liquid culture at 28^{o}C for two days, the culture was subsequently transferred to five universal bottles for a total volume in each bottle of 10 mL with a final concentration of 1 x 10^{8} cfu/mL^{-1}. Induction was tested at final concentrations of 0.01 mM, 0.1 mM, 1 mM, 10 mM of erythritol. These bottles were imaged using the NightOWL camera after 4 hours.

2.8 – Computational Work

2.8.1 – Statistics and data management

The Microsoft Office suite was used for all data input. Microsoft Excel was used to perform statistical analysis. Graphs were produced using a either Microsoft Excel or GraphPad Prism 7.

Restriction digests, cloning, primer binding and PCR amplification were tested, prior to experiments, *in silico* using Geneious R8.

A. caulinodans ORS571 genome annotations and protein table are from the NCBI database.

Chapter 3

Developing luciferase-based root exudate mapping in land plants

3.1 – Introduction

Up to 21% of photosynthetically fixed carbon (Kaiser *et al.*, 2015) is exuded from plant roots into the rhizosphere. This carbon plays a vast and diverse role, as a source of nutrients for soil microbes, altering stress tolerance and pathogen defence in plants, to the recruitment of growth promoting bacteria (Bais *et al.*, 2006; Martínez *et al.*, 2010). Whilst the importance of root exudation is well understood, the spatiotemporal nature of exudation is relatively unknown. The identification of spatial and temporal release patterns for key exudates is important in understanding how these exudates are reaching and interacting with the rhizosphere. Ultimately this will help better our understanding of plant-microbe interactions and how this influences plant health, alongside clarifying how the chemical composition of soils is altered.

A method for the spatial and temporal mapping of root exudation using bacterial luciferase (*lux*) driven by a series of rhizosphere induced promoters has been developed. These bioreporters make use of a plasmid containing a promoter fusion to the lux operon (*luxCDABE*) (Pini *et al.*, 2017). The plasmids are environmentally stable (due to the presence of the *par* locus) with the *lux* operon providing all the components and substrates for light production without the addition of an exogenous substrates such as luciferin. Sixteen bioreporters have been made available, these reporters are based in *R. leguminosarum* bv. *viciae* 3841; these encompass fourteen specific metabolites, and two constitutive reporters in which the *lux* genes are expressed at different levels (Table 34.1). Strain numbers are shown in Materials and Methods 2.2.1.

These bioreporters have previously only been used in mapping exudation in pea and vetch (Pini *et al.*, 2017). The aim of this chapter was to explore whether these biosensors can also be used for the spatiotemporal mapping of exudation in non-

leguminous plants, specifically *O. sativa*, *Z*. mays, *A. thaliana*, and *M. polymorpha*. This will provide a platform for future studies to use plant genetics to identify genes associated with exudation. Ultimately this will give a wider picture of the effects exudation has on the life cycle of plants and the elucidation of its genetic basis.

3.2 – Results

3.2.1 – Exudate mapping

Before any mapping began, the sensitivity of each bioreporter was determined by measuring the minimum concentration of an inducer molecule which yielded a detectable response as outlined in Materials and Methods 2.7.3. The results are shown in Table 3.1 and were subsequently published as part of Pini *et al.,* (2017), images are available in Supplementary 1.

Table 3.1 Table of available lux fusions for metabolite mapping and their detectable thresholds on minimal medium supplemented with inducer molecules.

Inducer	Sensitivity / mM
Sucrose/Raffinose	0.1
Xylose	1
Fructose	0.01
Inositol	0.1
Mannitol	0.001
Erythritol	1
Malate/Succinate/Aspartate (C4-Dicarboxylates)	0.01/0.1/10 respectively
Malonate	10
Tartrate	0.1
Formate	10
Salicylic Acid	1
GABA	0.5
Phenylalanine	0.01
Hesperetin	0.001
Constitutive (nptII)	N/A
Constitutive (tac)	N/A

3.2.2 – Optimisation of mapping in *Oryza sativa*

Oryza sativa is the staple food source for around 50% of the world's current population, making it one of the most important agricultural crops (USDA, 2012) and an important target for study. One aim of this work was to develop and optimise a method that would allow for the spatiotemporal mapping of exudation in *O. sativa* using luciferase-based bioreporters.

Seeds were sterilised and germinated as outlined in Materials and Methods 2.4.1 before being inoculated with the constitutive pNeo fusion. Plants were grown at 26 ^{o}C with a photoperiod of 16:8 (Day:Night) and imaging was performed as outlined in Materials and Methods 2.4.5 at 1 & 2 days post inoculation (dpi). These experiments highlighted the importance of root spatial arrangement when studying exudation. Bacterial growth/metabolism is required for the production of light using luciferase (ATP-dependent process), if bacterial growth/metabolism is hindered or stopped then the amount of light produced will decrease or cease completely. In these experiments luminosity was lost when roots grew away from the agar surface and began to dry, indicating that *Rlv*3841 needs a 'moist' environment for growth/metabolism. At 1 dpi the primary roots lay flat against the agar surface with exudation being detected, due to the observation of luminescence, on root surfaces (Fig.3.1A). At 2 dpi the growth of lateral roots pushed the primary root away from the agar surface, resulting in an observed reduction of luminescence on the root surfaces (Fig.3.1B).

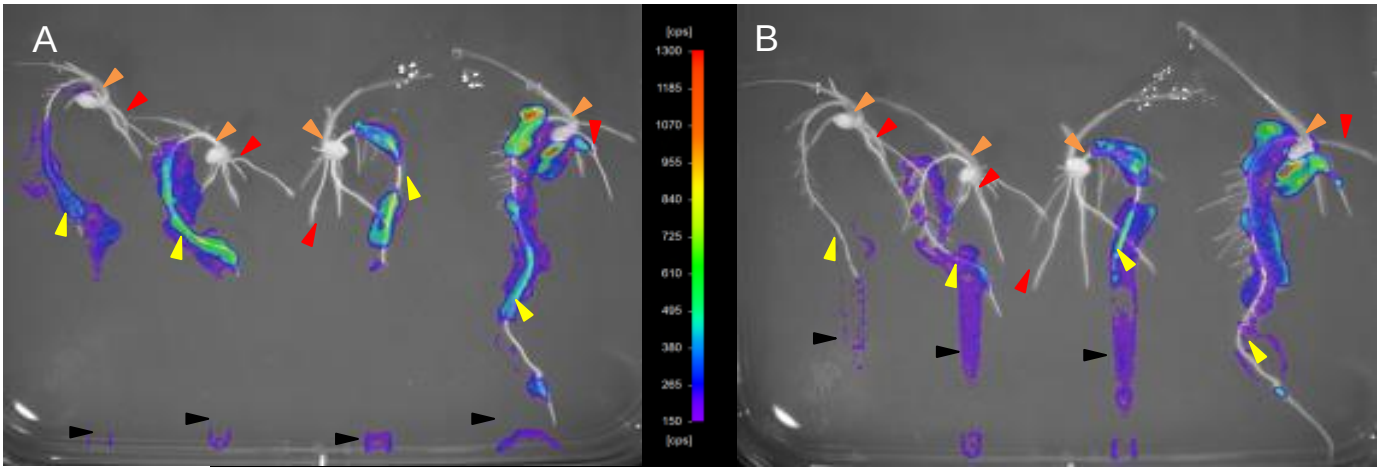


Figure 3.1. *O. sativa* cv. Dongjin inoculated with a constitutive pNeo lux bioreporter (OPS0100). A - 1 day post inoculation; B – 2 day post inoculation. Yellow arrows indicate primary roots; Red arrows indicate crown roots; Orange arrows indicate mesoctoyl; Black arrows indicate dried bacterial inoculum. CPS – counts per seconds, luminescence scale is equal for each image. Imaging and analysis was performed using the Nightowl system and IndiGO software.

These experiments highlighted the need for a new method of growth that would allow the rooting system to grow flat along the agar surface. Furthermore, spots of bacterial inoculum can be seen (with very low signal) dried to the agar surface, this limits the ability of bacteria to colonise the root surface. Four different methods of plant and bacterial growth were trialled to see if the problems seen above could be readily resolved. Three of these, which did not work, are outlined below. Primarily the pNeo fusion was used, but the final setup was tested with both the pNeo and pTac constitutive fusions, plants were imaged at 4 dpi.

1) A sterile segment of filter paper was placed over the bottom half of a slant plate, to allow the roots to grow over the surface and the roots were covered in sterile cellophane (Figure 3.2A). The aim was to determine if the filter paper could act as an anchor point for the roots so they would ‘stick’ to the papers surface, while maintaining a wet environment for bacterial growth.

2) Roots were sandwiched in-between two sheets of filter paper. This followed the same logic as point 1, however an extra layer of paper was added over the roots to ensure they stayed flat. The upper layer was removed before imaging (Figure 3.2B).

3) A single layer of sterile filter paper was placed on top of the roots. This aimed to keep the roots flat whilst allowing the bacteria to grow and colonise the roots. The paper was removed before imaging (Figure 3.2C).

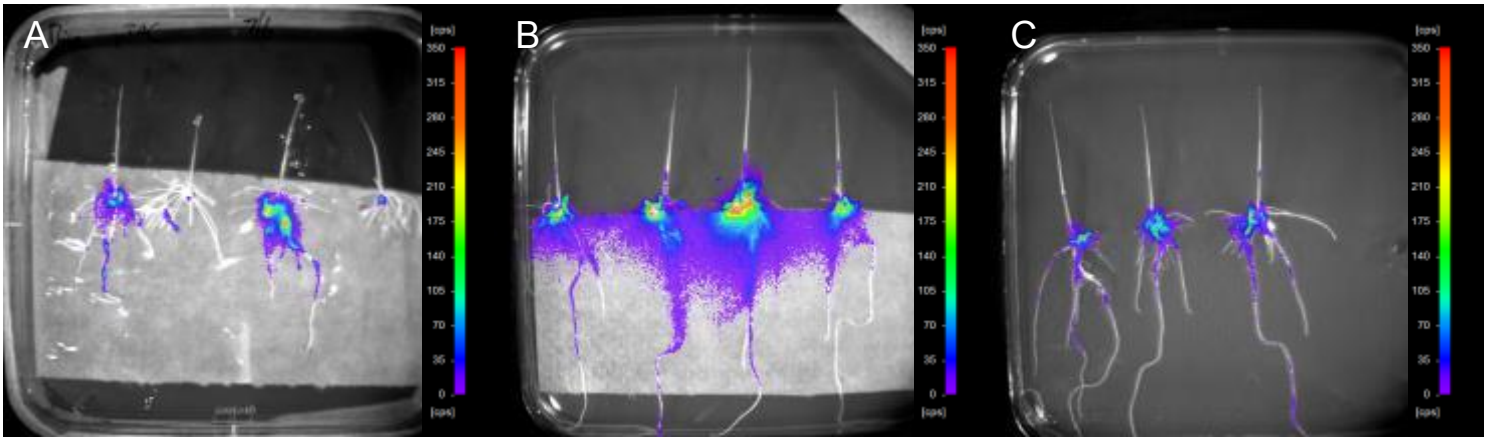


Figure 3.2 - Growth of *Oryza sativa* with constitutive lux fusion (pTac) and Imaged 4 dpi using the NightOWL system. Image analysis utilised the IndiGO software. CPS – counts per second. A – Grown on filter paper, roots covered with cellophane; B – Roots sandwiched between two sheets of filter paper (top sheet removed for imaging); C – Roots grown with paper covering (paper removed for imaging);

None of these methods proved suitable for mapping. Using paper covering the agar and cellophane as an overlay material did not keep the roots flat against the agar surface (Fig 3.2A). Sandwiching roots between two sheets of filter paper (Fig 3.2B) held the roots flat and allowed the bacteria to grow, however in removing the upper layer of filter paper the lateral roots and root hairs were damaged allowing for imaging on a single time point only. Covering the roots with a single sheet of filter paper (Fig 3.2C) caused the roots to grow into the agar, an anaerobic environment, which hindered bacterial growth (*R. leguminosarum* is aerobic), and roots that stayed above the surface were damaged on removal of the filter paper. The fourth method tested proved most successful. This involved covering the roots with a sterile sheet of clear cellophane; this method stopped roots growing away from the agar surface

whilst providing a suitable environment for bacteria to grow and colonise the root surfaces (Fig. 3.3). Furthermore, by using clear cellophane to cover the roots, the plants could be imaged over a successive period of days without disturbing the plant. As such, covering the roots with cellophane was taken forward for further optimisation using *O. sativa* cv. IR64. This cultivar was chosen as future work focused on screening a library of *O. sativa* cv. IR64 EMS mutants.

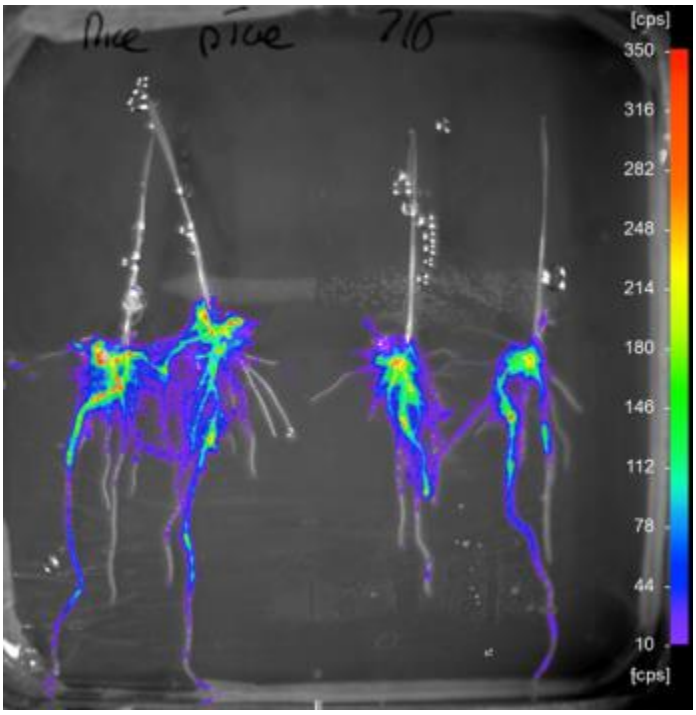


Figure 3.3. *O. sativa* cv. Dongjin inoculated with constitutive pTac lux fusion (LMB552), roots covered with cellophane and imaged at 4 days post inoculation. Imaged using NightOWL camera, analysed with IndiGO software. CPS – counts per seconds.

Once the growth system had been finalised, the pTac constitutive fusion was chosen as the standard for total exudate mapping due to its increased promoter strength over pNeo. This increased strength puts a higher metabolic burden on the bacterial cell, meaning cells require a higher amount of ATP to generate light and maintain growth. This metabolic burden proved advantageous as it led to the death of bacterial cells that were not gaining access to plant-derived carbon with the remaining cells being dependent on the carbon exudation of the plant; this allowed

for more precise maps with lower levels of background luminescence (*i.e.* luminesce caused by slow growing cells away from root surfaces).

To summarise, using cellophane for the root overlay created a high-resolution system for temporally and spatially mapping exudation without interfering with plant growth. Further, this method of growth allowed for the imaging of plants (thus exudation) over successive days, giving the opportunity to probe the temporal nature of exudation alongside its spatial nature. This system was used to map root exudation in *O. sativa* with the fourteen specific biosensors.

3.2.3 – Exudation mapping in *O. sativa*

The aim of these experiments was to utilise the system optimised in 3.2.2 to produce spatial exudation maps in *O. sativa* with the specific bioreporters that were available. Here I aimed to identify which out of the compounds were present in *O. sativa* exudates, if so whereabouts from the roots they were released and if different exudates followed the same or differing spatial patterns of release. The experimental setup was as described in Materials and Methods 2.4.1. Roots were inoculated with a biosensor and imaged over a 12-day period, images showing highest peak luminescence were used for analysis.

With the bioreporters used, six showed strong signals (Fig. 3.4B, C, H, J, K, N) indicating the presence of xylose, fructose, C4-dicarboxylates (malate, succinate and aspartate), tartrate, formate and phenylalanine. A further six showed weak signals (Fig. 3.4A, E, F, G, L, O) for sucrose/raffinose, inositol, sorbitol, erythritol, salicylic acid and hesperetin. Finally, two bioreporters showed no signal at all (Fig. 3.4 I & M) suggesting that no malonate and γ-aminobutyic acid were present on the rice roots (or at least below the threshold levels – see Table 3.1). For all of the twelve detected

metabolites the seeds of each plant gave a strong signal in comparison to the rest of the rooting system. This is likely due to the high metabolic activity of the growing embryo and the large carbon store present within seeds, both supporting bacterial metabolism and providing a large concentration of any inducing molecule. The lack of detectable luciferase activity on roots does not rule out the absence of metabolites, rather it may be that the concentrations are below the detectable threshold of the reporters (see Table 3.1). The constitutive pTac fusion (Fig 3.4P) acted as a positive control, showing that there is enough carbon in the general root exudate to support bacterial growth as shown by the presence of luminescence on all root surfaces.

A spatial pattern was seen for fructose, C4-dicarboxylates and phenylalanine exudation on the *O. sativa* roots (Fig 3.4 C, H, N). The release of these three exudates appears to be centered on the seed and crown roots, with only low levels of exudation present down the length of the primary roots, specifically on the older root surface (closest to the seed). Sugars and carboxylic acids appear to make up the bulk of the detectable metabolites being released into the rhizosphere.

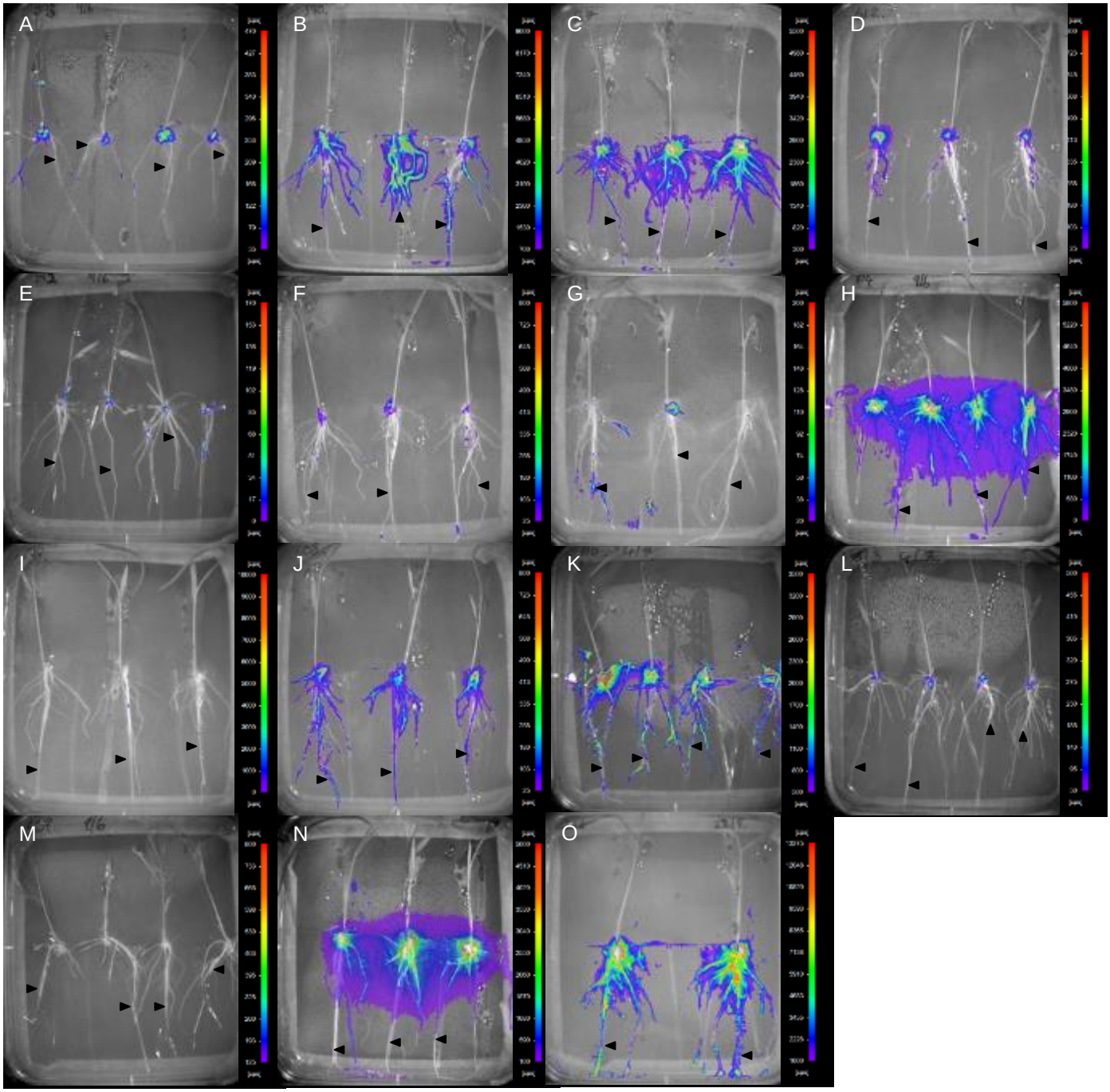


Figure 3.4 *O. sativa* cv. IR64 inoculated with specific lux fusions. Imaged using NightOWL camera, analysed with IndiGO software. CPS – counts per second. Images are representative, the day which showed peak luminescence. Black arrows indicate primary roots. A – Sucrose/Raffinose, 11 dpi. B – Xylose, 8 dpi. C – Fructose, 8 dpi. D – Hesperetin, 11 dpi. E – Inositol, 11 dpi. F – Sorbitol, 11 dpi. G – Erythritol, 12 dpi. H – Aspartate/Malate/Succinate, 11 dpi. I – Malonate, 12 dpi. J – Tartrate, 11 dpi. K – Formate, 13 dpi. L – Salicylic acid, 13 dpi. M – γ-aminobutyric acid, 11 dpi. N – Phenylalanine, 11 dpi. O – Constitutive (*pTac*), 13 dpi.

A library of 2000 *O. sativa* cv. IR64 mutant lines (produced by EMS mutation) was made available to me by the International Rice Research Institute. The aim of this work was to determine whether *lux* mapping could be used as a screening method to identify mutants altered in exudation. The constitutive pTac fusion was chosen for mapping due to its low background signal. Seeds were sterilized, germinated and grown as outlined in Materials and Methods 2.4.1. Plates were imaged after 4 dpi, with the aim to compare luminescence. A total of 600 lines were screened. Unfortunately, observed differences in in exudation were mainly attributed to variations in plant growth *i.e.* a larger root had larger exudation compared to a smaller root. This variation was expected to some degree, however the levels observed proved to be much higher than expected. Thus, a small change in growth rate shifted the levels of exudation by a larger degree (based on reporter signal strength), making any form of analysis unreliable. This method of screening has not been ruled out, rather plant growth and imaging needs to be further refined and optimised before a reliable screen can be set up. This refinement was not possible within the time scope of this project so was not pursued. All images are included in Supplementary 2.

The results in this chapter have highlighted that these biosensors are capable of spatially mapping exudation across a growing root system *in vivo*. Furthermore, I have shown that exudation of certain metabolites is not limited to the root elongation and meristematic zones and that exudation is occurring across the whole rooting system. Indicating that plants are not indiscriminately releasing exudates into the soil, rather that there is control over the localisation of exudate release.

3.2.4 – Exudate mapping in *Zea mays*

The aim of this work was to determine if *Z. mays* produces the same exudation patterns that were seen in *O. sativa*. This work focused primarily on the release of C4-dicarboxylates from the roots. C4-dicarboxyaltes, in particular malate, have been found in the exudates of many different plant species, including *Z. mays*, and have been implicated in playing important roles in nutrient acquisition, metal detoxification and microbiome manipulation (Jones, 1998). This makes the C4-dicarboxylates an interesting target for study. In this work the overarching question to be answered was: Is the exudation of C4-dicarboxylates conserved across cereals and does the spatial pattern of exudation match. If C4-dicarboxyalte exudation and any spatial patterns of release are similar across plant species (in this case cereals) with similar root architecture, then it could be argued that the mechanisms by which the exudation of these compounds is occurring is conserved.

Zea mays (L.) *ssp.* Mas cv. B73 (USDA) was used for this work. Plants were grown in conditions as set outline in Materials and Methods 2.4.2. The protocol for mapping *O. sativa* was modified to take into account the size of a growing *Z. mays* seedling. For this work maize was inoculated with either the C4-dicarboxylate reporter or the pTac constitutive reporter. Plates were imaged up to 8 dpi. Both reporters were successful in mapping exudation both spatially and temporally in this system. Figure 3.5 shows representative images for 4 & 7 dpi.

The colonisation and exudation from the primary, crown and seminal roots is shown at 4 dpi (Fig 3.5A). There was a noticeable increase in luminosity where lateral roots are beginning to form on the primary roots. At 7 dpi (Fig 3.5B) the luminescence on the primary roots dropped, with a substantial increase in luminescence on the lateral roots. The mature regions of the lateral roots *i.e.* regions closest to the parent root; showed the highest luminescence, with a significant drop in signal across the length

of the primary root. Luminescence values were taken from the lateral roots which emerged from the primary roots, three separate areas were chosen: the root tip, elongation zone, and base (~2mm from the primary root), these zones are circled in figure 3.5B. Ten lateral roots were chosen at random and this was repeated for three replicate plants, a single representative replicate is shown in figure 3.5E. Using a 2-tailed t.test assuming unequal variances, the values were compared. A significant difference between the root tip and the base of the lateral roots was seen (p-values <0.0005), as was a significant difference between the elongation zone and base (p-values <0.005). No significant difference was seen between the root tip and elongation zone. This shows that exudation is occurring across mature root surfaces not limited to the growing meristem and elongation zones. These results are mirrored in the *O. sativa* data, though in that case exudation was observed across the whole rooting system and not limited to lateral roots. The exudation of C4 dicarboxylates follows a similar pattern to that seen with the constitutive reporter. At 4 dpi (Fig 3.5C) luminescence on the primary, seminal and crown roots can be seen, with peaks in luminescence where lateral roots are forming, as well as along the elongation zone of the primary and crown roots. At 7 dpi (Fig 3.5D) the exudation has mainly shifted to the lateral roots, mirroring what was seen with the constitutive fusion. The key difference is that C4-dicarboxylates appear to be released in greater quantities from the apical meristem and elongation zone, with the levels of luminesce dropping towards the parent root. Again, luminescence values were taken from lateral roots that emerged from the primary roots. Three separate areas were chosen: the root tip, elongation zone, and base (~2mm from the primary root), these zones are circled in Figure 3.5D. Ten lateral roots were chosen at random and this was repeated for three replicate plants, a single representative replicate is shown in Figure 3.5F. Using a 2-tailed t-test assuming unequal variances, the values for each zone were compared. A significant difference between the root tip and elongation zone was seen for two of the three replicates (p-values <0.05).

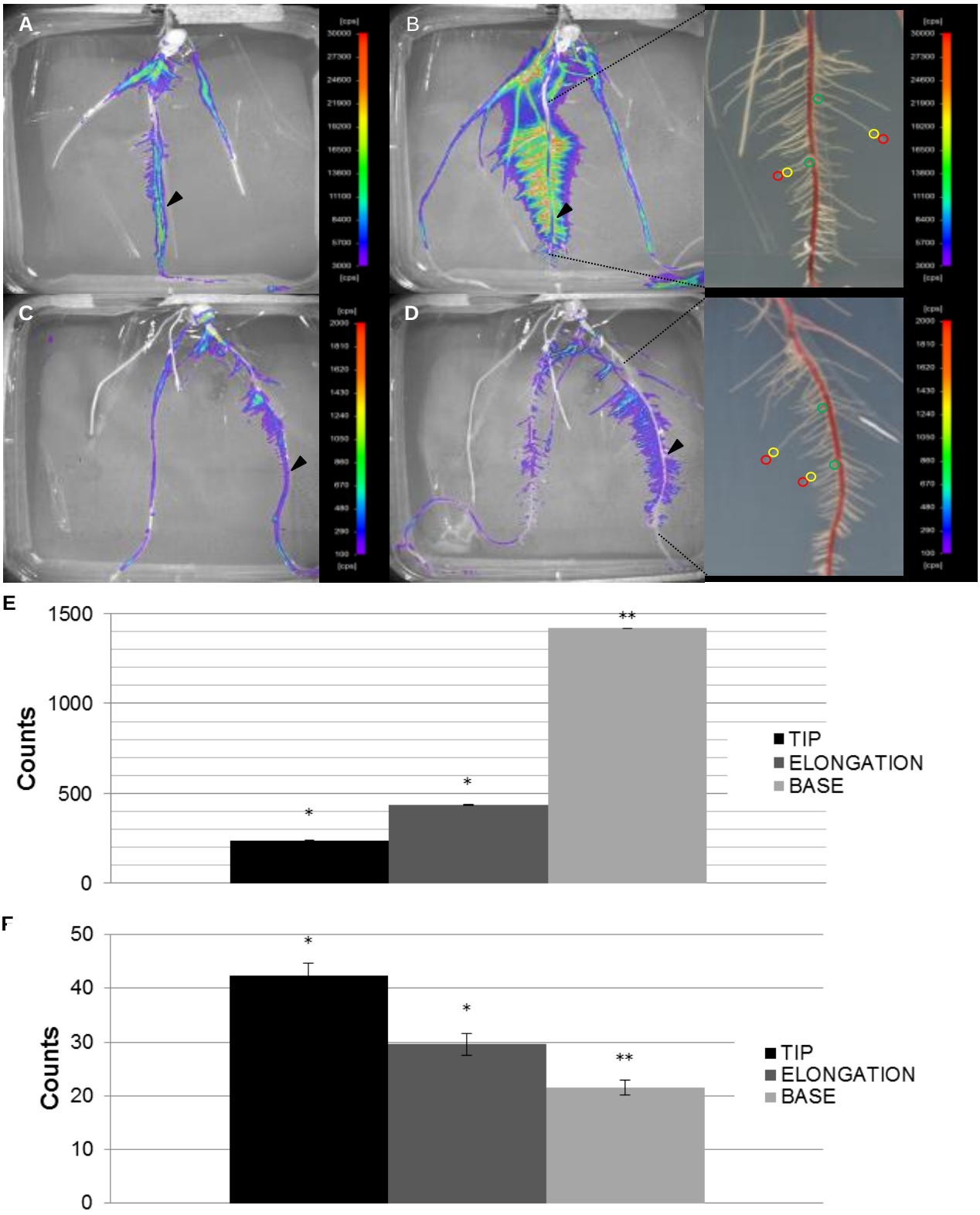


Figure 3.5 Exudate mapping in *Zea mays*. Imaged using NightOWL camera, analysed with IndiGO software. CPS – counts per second. Black arrows indicate primary roots. A&B - Constitutive fusion 4 & 7 dpi respectively. C&D – C4-dicarboxylates, 4 & 7 dpi respectively. E- Average lux counts on the tip, elongation and basal zones for ten lateral roots for the constitutive fusion; Error bars show standard error; Statistical analysis using two-tailed t-test assuming unequal variance, asterisks represent significant differences (p-value <0.05. F- Average lux counts on the tip, elongation and basal zones of ten lateral roots for the C4-dicarboxylate fusion. Error bars show standard error; Statistical analysis using two-tailed t-test assuming unequal variance, asterisks represent significant differences (p-value <0.05).

Here I have shown that in *Z. mays* exudates are detected in their greatest quantities being released from the lateral roots rather than the primary or crown roots. Using the constitutive reporter, I have shown that the more mature lateral root surfaces are releasing exudates in great quantities. In contrast, the C4-dicarboxylates are being released primarily from the elongation zone with a drop in exudation across the mature lateral surfaces. These results show general exudation occurring across the length of the whole rooting system in particular the lateral and crown roots, with C4-dicarboxylates released from specific points in the growing root system.

3.2.5 – Exudate mapping in *Marchantia polymorpha*

The bryophytes are widely accepted to be the earliest diverging lineages of extant land plants, having colonised the land around 450 Ma ago. Bryophytes are made up of three groups of non-vascular land plants: liverworts, hornworts, and mosses. The liverworts are the earliest diverging group within the bryophytes, of which *M. polymorpha* is a member. For the study of the early evolution of non-vascular land plants *M. polymorpha* is often used as a model. Its rooting system comprises of a number of tip-growing filamentous cells called rhizoids which have a similar structure and function to root hairs that are present in more complex rooting systems. Whilst root exudation plays an important and complex role in the rhizosphere of vascular plants, it is relatively understudied in non-vascular plants. *Marchantia polymorpha* gives a unique opportunity to investigate exudation within the earliest rooting structures of land plants. The aims of these experiments were to use the bioreporters to determine if exudation occurs in *M. polymorpha* and if so are these exudates released from the thallus of the plant or the rhizoids.

M. polmorpha was grown from gemmae on modified Johnson's media for 35 days before being inoculated with biosensors as outlined in Materials and Methods 2.4.4. Care was taken so that the inoculum did not spread to neighboring plants. Imaging was performed at 4 and 6 dpi. Johnson's medium contains sucrose and inositol, ruling out the use of the corresponding fusions for exudate mapping. As rhizoids emerge from the ventral side of the thallus, all imaging was performed through the bottom of the growth plate to observe this region.

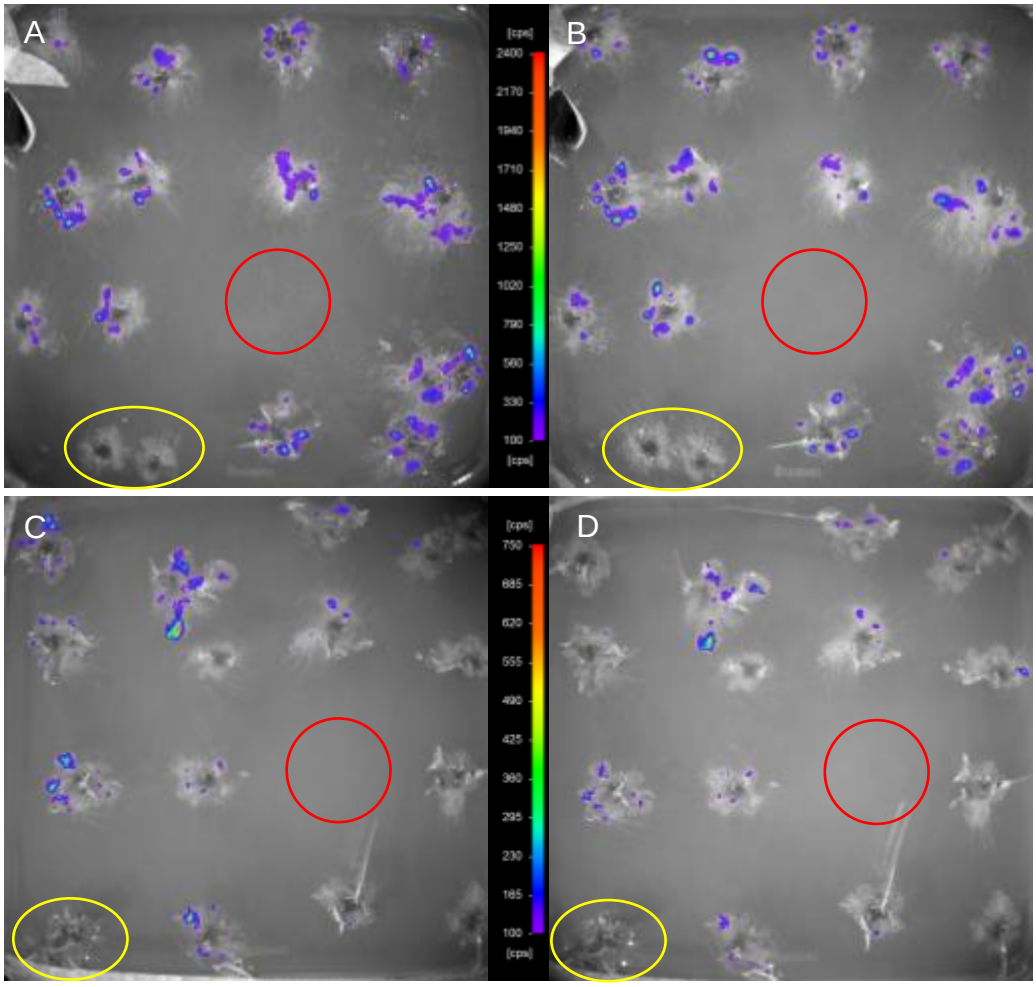


Figure 3.6 Exudate mapping in *Marchantia polymorpha*. Plants were grown for 35 days before inoculation. Imaged using NightOWL camera; analysed with IndiGO software. CPS – counts per second. A – C4-dicarboxylates at 4 dpi; B – C4-dicarboxylates at 6 dpi; C – Formate at 4 dpi; D – Formate at 6 dpi. Red circles indicate control droplets of inoculum with no plant; yellow circles indicate uninoculated control plants.

Out of twelve reporters, only those detecting C4-dicarboxylates and formate produced detectable signals (Fig 3.6). C4-dicarboxylates (Fig 3.6A and B) had a high

peak luminescence across both time points. Formate showed a weak signal at 4 and 6 dpi (Fig 3.6C and D), indicating that level of exudation was close to the detectable threshold for this sensor (10 mM). In both cases exudation appeared to be centered on the midrib and meristematic regions of the thallus. There was no luminescence present on the rhizoids themselves, indicating that the bulk of exudation was occurring from the thallus itself. However, rhizoids pull water and nutrients to the thallus by surface tension. This effect was seen during inoculation, with a portion of the inoculum being pulled towards the thallus, this would act to pull the bacteria up to the thallus giving rise to the signal seen along the midrib. To further explore this, the mutant line Mp*CSLD*1-1 (Honkanen *et al.*, 2016) which produces very short rhizoids, was used to determine if the signals seen were a result of exudation from the rhizoids or the thallus. If no change in signal was seen it could be determined that the bulk of exudation was occurring from the thallus rather than the rhizoids. If the signal decreased, then the bulk of exudation would likely be occurring on the rhizoids. In this experimental setup, no rhizoids could be seen on any of the mutant plants. Each plate contained three wild-type and twelve mutant plants. Two of the mutants were left un-inoculated, and a droplet of bacteria with no plant was used as a baseline for luminescence. Both the ventral and dorsal sides of the plants were imaged. The wild-type and mutant plants both produced detectable quantities of C4-dicarboxylates, with the bulk of exudation being detected on the ventral side of the thallus (Fig 3.7.B), with only a small signal was seen on the dorsal side of the plants (Fig 3.7A); this signal was much lower and covered a much smaller area than what was seen on the ventral side. This confirms that exudation is occurring preferentially from the ventral side of the plant. Both lines show highest peak levels of luminescence centered on the midrib and meristematic regions. The *Mp*CSLD1-1 mutant showed significantly (p-value <0.05) lower signals than the wild-type plants (Fig 3.8). These data further suggest that the bulk of exudation is occurring primarily from rhizoids

rather than directly from the thallus, with only small levels occurring at the meristematic regions of the plant.

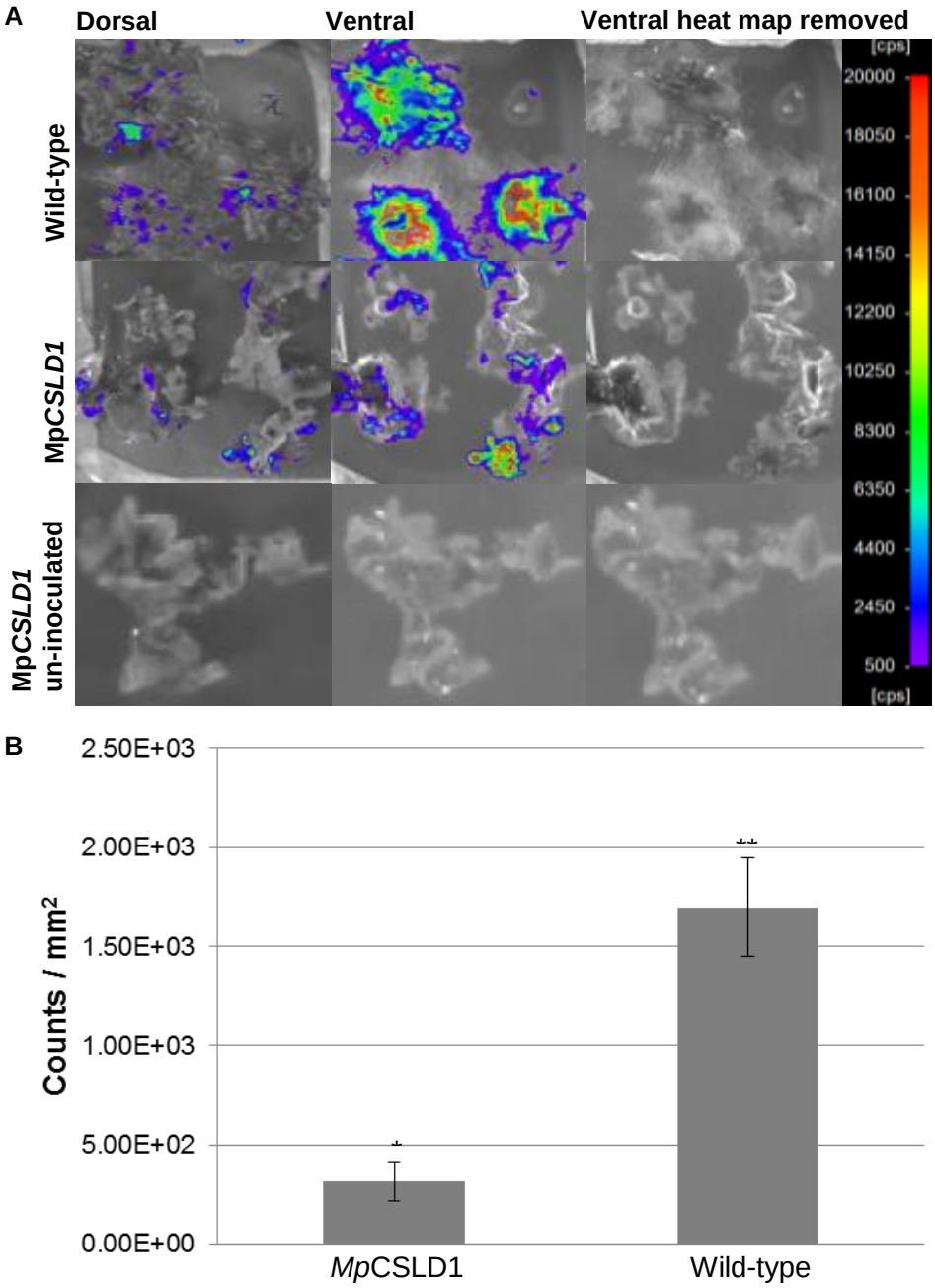


Figure 3.7. *Marchantia polymorpha* short rhizoid mutants screened for exudation of C4-dicarboxylates. Imaged using NightOWL camera, analysed using IndiGO software. CPS – counts per second. A – *M. polymorpha* wild-type and *Mp*CSLD1 inoculated with C4-dicarboxylate fusion, ventral and dorsal surfaces imaged at 4 dpi. All plants were screened on the same plate. B -Total counts per mm^2 for the C4-dicarboxylate fusion on the ventral surfaces of *Mp*CSLD1-1 and wild-type. Error bars show standard error, asterisks represent statistical significance. A two-tailed t-test assuming unequal variance was used for statistical analysis.

Here I have shown that the lux bioreporters are capable of mapping exudation in *M. polymorpha*. Out of the reporter fusions tested only those detecting formate and C4-dicarboxylates showed a signal, indicating their presence as exudates. Further, using a short-rhizoid mutant it has been possible to identify that the exudation of C4-dicarboxylates is occurring primarily from the rhizoids rather than the thallus.

3.2.6 – Exudate mapping in *Arabidopsis thaliana*

Arabidopsis thaliana Columbia (Col-0) is the most widely used model plant system. Its small size and speed of growth make it an ideal target for exudate mapping. Furthermore, plants can be grown for a long period of time on Petri plates before space becomes an issue, allowing imaging over long time periods. The aims of this work were to spatially and temporally map exudation in *A. thaliana* and to determine if this mapping technique could be used as the basis for a genome wide association study (GWAS).

All *A. thaliana* seeds were sterilised, grown and inoculated with bioreporters as outlined in Materials and Methods 2.4.3. Images were taken at 24 hr intervals post inoculation as outlined in materials and methods 2.4.5. Due to the inclusion of sucrose in the growth medium, constitutive fusions could not be used to screen total exudation as sucrose would support bacterial metabolism independent of any exudates, leading to a large 'background' signal. Screening in *A. thaliana* was thus optimised with the C4-dicarboxylate fusion, the presence of malate, succinate and aspartate in root exudates has been previously documented (Badri *et al.*, 2009; Chaparro *et al.*, 2013). Plants were grown for 10 days before inoculation to allow the plant to reach a size that supported a large bacterial community and to fix an adequate amount carbon that would be detectable as exudates.

Using the C4-dicarboxylates reporter, root exudation was mapped in wild-type Col-0 plants (Fig 3.9).

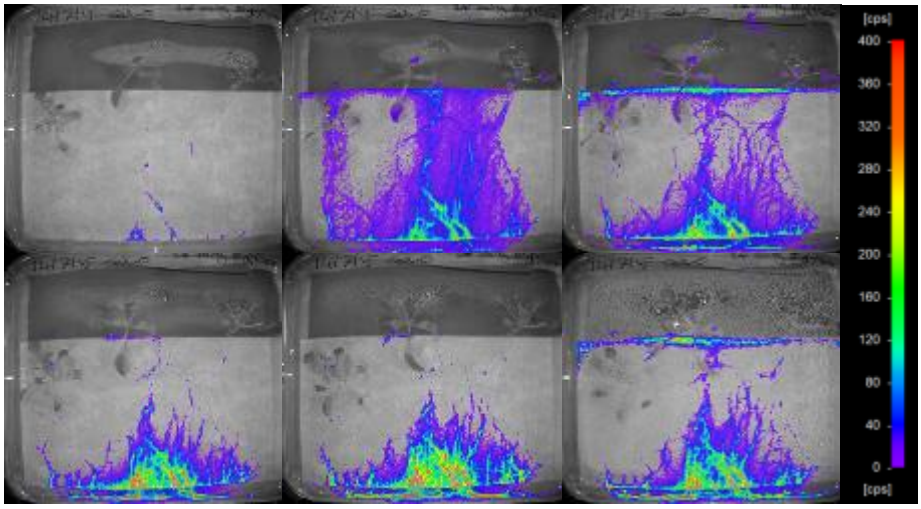


Figure 3.9 *Arabidopsis thaliana* Col-0 inoculated with C4-Dicarboxylate bioreporter. Imaged using the NightOWl camera, analysed using the IndiGo software. Imaged at 6, 7, 8, 9, 10, and 13 dpi from top left to bottom right. Images representative of two replicates.

Some spatial arrangement surrounding the release of these exudates can been seen, where exudation appears to begin around the younger areas of the rooting system before increasing on the more mature root surfaces. Inoculation bias was discounted as each piece of filter paper was evenly inoculated with 1 mL of bacteria. Due to the method of imaging and the size of the roots it was not possible to determine the exact spatial location of exudation e.g. lateral roots. The peak intensity of the signals was far less than those seen in larger vascular plants; this was expected due to the size difference between the different rooting systems. Presence of inositol (polyol), hesperetin (flavonoid), fructose (sugar) and salicylic acid (organic acid) were also screened for using the respective bioreporters (Fig 3.10). From these, only fructose yielded a detectable signal (Fig 3.10C). Unlike in the C4-dicarboxylate results, there was no increased signal from the younger rooting structures.

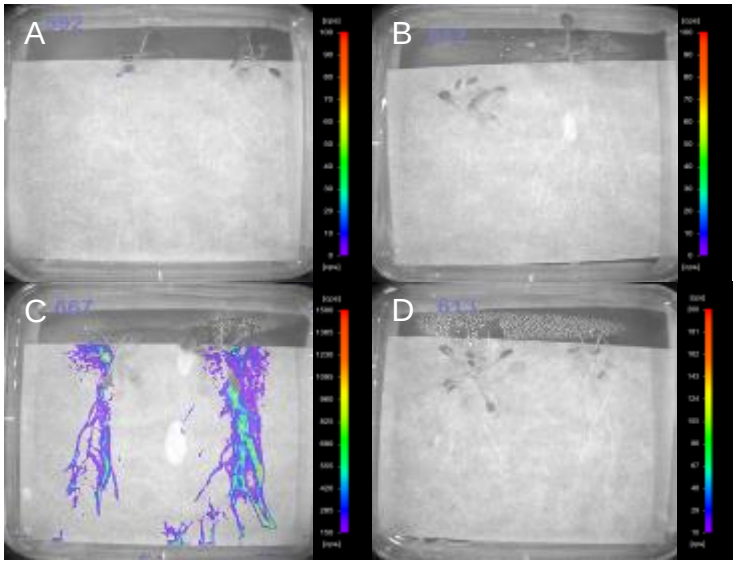


Figure 3.10 *Arabidopsis thaliana* Col-0 inoculated with a range of biosensors, imaged at 7 dpi. Imaged using NightOWL camera, analysed using IndiGO software. CPS – counts per second. A – Inositol; B – Hesperetin; C – Fructose; D – Salicylic Acid. Images representative of two replicates.

Once mapping was optimised I tested if the screening method developed above could be used to form the basis of a genome wide association study (GWAS). GWAS allow for the identification of genes and genomic regions that are linked to a phenotype of interest through the screening of hundreds of genotyped *A. thaliana* lines for a specific phenotype (e.g. variation in root exudation). Lines that show a positive phenotype are genotypically compared, if a specific locus remains the same or has very little variance across these lines but differs in the non-positive lines then it can be said to be linked to the phenotype of interest. This technique relies heavily on the ability to robustly quantify a phenotype of interest.

Six randomly selected *A*. *thaliana* lines (provided by Javier Augusti) were screened for altered C4-dicarboxylate exudation. Two plants per line were placed on each plate with two replicate plates prepared for each, bringing the total number of plates to 12, six representative plates are shown in figure 3.11. A randomly selected line

(line no. 7337) was chosen to act as an internal control for each plate. This line was placed at the side or centre of the Petri plates to check for edge effects that could cause variations in exudation. The inoculum for all plants was prepared from a single slope culture, all conditions for growth are as outlined in Materials and Methods 2.4.3. Figure 3.11 shows a single representative replicate for each line. The internal control shows a different signal across each plate. The only visible difference between each plate was the shape and small variations in the size of each control root. It is likely therefore that these changes in exudate patterns are a result of variations in plant growth. Studies such as GWAS rely on robust phenotypic analysis.

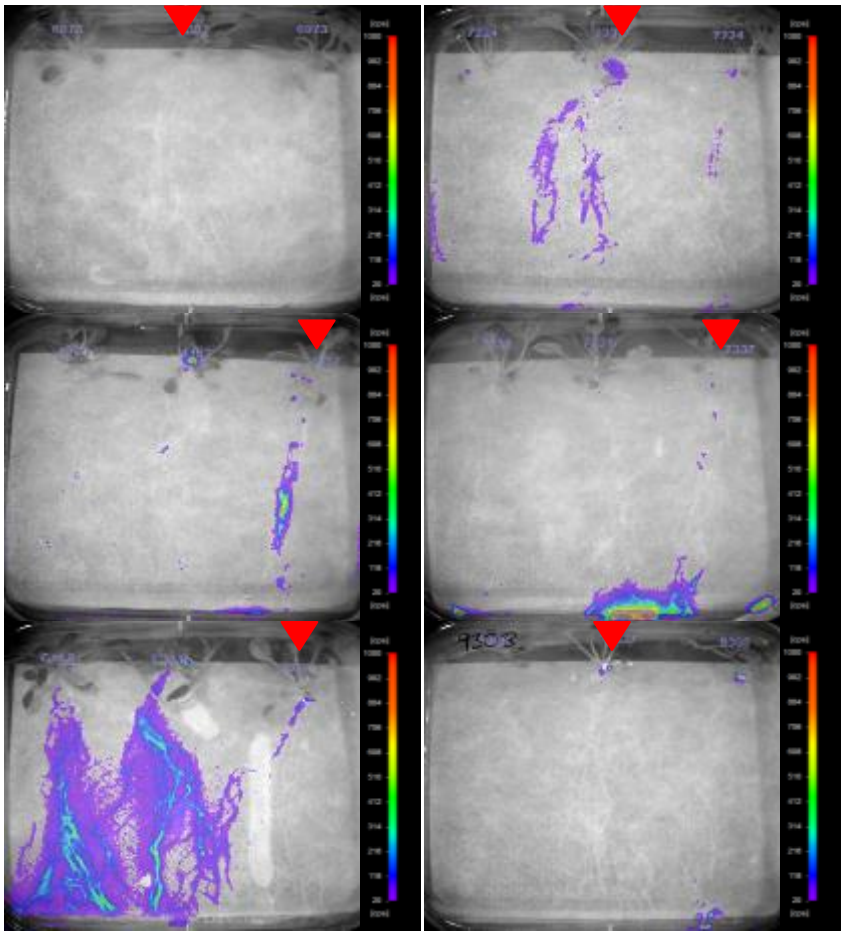


Figure 3.11 *A. thaliana* GWAS accessions inoculated with C4-Dicarboxylate fusion imaged at 8 dpi. Red arrow indicates control line: 7337. Images are representative of two replicate plates for each accession.

In this experiment the cause of the differences in exudation could not be differentiated between a genotypic effect or a growth effect (*i.e.* difference in root size and shape).

In summary these results show that these bioreporters can be used to map exudation in *A. thaliana.* The exudation of C4-dicarboxylates appears to be occurring from the younger rooting surfaces; however localisation to specific rooting structures and zones was not possible due to limitations in image resolution. Further, in its current state the use of this system is not for suitable for a GWAS of root exudation due to variations in plant growth.

.

3.3 – Discussion

Root exudation has been mapped across a variety of different land-plant species, from the some of the earliest diverging linages to higher vascular plants. Across tested plant species one common factor is exudation of C4-dicarboxylates. The reporter fusion used to map these dicarboxylates is driven by the *Rlv*3841 *dctA* gene promoter region. It has been shown that this bioreporter specifically responds to malate, succinate, and aspartate (Pini *et al.,* 2017) at concentrations of 0.01 mM, 0.1 mM and 10 mM respectively. Previous studies have shown that these three metabolites are present in the exudates of a variety of different plant species and, as such, in this study the response seen is likely a result of all three metabolites being detected. That being said, it is likely that malate and succinate are responsible for the bulk of the detected signal as the concentration thresholds for induction are 1000- and 100-fold lower, respectively, than that of aspartate (Table 3.1). These data would indicate that C4-dicarboxylate exudation occurred during the earliest phases of plant evolution and divergence, as shown by exudation in *M. polymorpha*. This exudation has then been conserved through the evolution of the non-vascular and vascular

plants, all the way to modern plant species. Future work focusing on the genetic basis of exudation of C4-dicarboxylates in different species and the evolution of these genes would be key to determining if this arose once in evolution and was conserved or if it has evolved independently multiple times. Despite this uncertainty, it is interesting that these compounds are released by such a basal plant species.

The revelation that metabolites such as C4-dicarboxylates are released from the crown, seminal and lateral roots preferentially over the primary roots in cereals is a novel and exciting result. The constitutive fusion gives some evidence that exudation occurs across the whole rooting system, along with showing the total colonisation of the roots. Current literature focuses on the release of exudates via passive or active processes from the root cap border cells (RCBs), apical root meristems and the elongation zones; there has been little evidence to show deeper spatial organization of exudation outside of these zones from more mature root surfaces (Dennis *et al.*, 2010). Here I have shown that exudation is likely occurring across the entirety of the root surface from the most mature regions to the growing meristem. More importantly the release of exudates from different root zones lends credence to the theory that plants can manipulate exudation across the whole rooting system rather than a more passive release from just the growing tips and elongation zones. Furthermore, ABC-transporters have been shown to regulate the release of carbon into the soil (Badri *et al.*, 2009), this form of transport is an active process and requires the use of energy in the form of ATP. These active processes are tightly regulated as any organism must control energy expenditure to survive. Coupling these spatial data with evidence that plants are using ABC-transporters for exudation indicates the global and local control of exudation into the rhizosphere. This control gives a plant a much finer control over its manipulation of the soil environment.

Plant produced low molecular weight compounds play diverse roles in the rhizosphere e.g. organic acid solubilisation of phosphate, manipulation of the microbiome and metal detoxification (Bolan *et al.*, 1994; Bertin *et al.*, 2003; Bais *et al.*, 2006; Badri and Vivanco, 2009). However, upon release exudates do not diffuse any great distance from the roots before being depleted by rhizosphere processes. Some estimates state that from 2 to 47 mm distant from roots the exudate concentrations drops to around 1% of that which was originally released (Dennis *et al.*, 2010). To overcome this one of two things could occur: firstly, there could be an increase in the concentration of exudates being released; and secondly, exudates could be released over a greater area. Plants already invest up to 21 % of their photosynthetically fixed carbon into exudation. Therefore it would be more advantageous for plants to manipulate how and where they release exudates to interact with as great an area as possible rather than sink more carbon into the rhizosphere. Here I have shown that the highest concentrations of C4-dicarboxylates are released from crown and lateral roots in the *O. sativa* and maize rooting systems. These rooting structures extend into a greater volume of soil than a single primary root, and by focusing exudation to these zones exudates can affect a much greater area of the growth medium. It is unknown if similar patterns of exudation occur in plants with different root architectures e.g. those with allorhizic root systems.

Release of C4-dicarboxylates from *A. thaliana* roots begins on younger root surfaces before transitioning to more mature surfaces. During this process there does not appear to be specific spatial localization on the rooting system, rather the whole system appears to be responsible with a small preference for the younger root surfaces. As *A. thaliana* grew in this imaging system its roots spread over a large area rapidly. This ultimately lead to the primary and lateral roots overlapping multiple times making it impossible to isolate any single root structure where exudation was occurring. Despite this, a strong temporal affect was seen with an apparent increase

in signal over time. This temporal effect was also seen across all plant species tested with C4-dicarboxylates, and is likely due to the increasing size of the plant allowing for the release of an increased concentration of metabolites as well as supporting a larger bacterial population.

I have also shown the release of phenylalanine is focused around the crown roots and the seed of a growing *O. sativa* plant. This localisation can be assumed to follow a similar logic as that discussed for C4-dicarboxylates, in that localisation of exudation to crown roots is to reach a larger soil volume. Phenylalanine exudation is a curious phenomenon as plants actively use nitrogen to biosynthesise phenylalanine; and that it is widely used in protein synthesis and production of a variety of secondary metabolites such as flavonoids (Tzin and Galili, 2010). So it is interesting that phenylalanine, whose biosynthesis has taken a share of available nitrogen, is released back into the rhizosphere. Bacteria and other microbes can catabolise phenylalanine as an energy source (Teufel *et al.*, 2010) and it has been seen that it is released readily into the rhizosphere along with other aromatic amino acids (Badri *et al.*, 2009; Chaparro *et al.*, 2013; Moe, 2013). As such, it is possible that bacteria present in the rhizosphere are utilising phenylalanine as a carbon source and plants are releasing it to stimulate and/or control microbial growth.

In this thesis the use of exudate mapping to screen for altered exudation across plant libraries was tested in both *A. thaliana* and *O. sativa* (GWAS and a mutant screen respectively). The aims of these screens were to act as methods to identify genes (and operons) involved in control of root exudation. In the case of *A. thaliana*, small differences in plant growth resulted in altered exudate profiles, and a similar result was seen when screening *O. sativa* mutant lines. For both GWAS and mutant screens the ability to robustly measure a phenotype of interest is critical. In this work the variations in plant growth shifted the exudate profiles to a degree where

comparative analysis between genotypes or mutants was increasingly difficult. To overcome this the growth of plants must be highly regulated in such a way that the roots and shoots of each line grow reproducibly; signals could then be normalised to a control line. This optimisation of growth was not within the time or scope of this thesis.

Overall, I have shown the ability to map the exudation of key metabolites across a range of different plant species. I focused on organic acids whose roles are far reaching in the rhizosphere and beyond. Novel spatial organisation of exudation has been identified along with confirmation that as a plant grows there is an increase in the release of exudates. A key result is showing that colonisation and exudation is occurring across the whole system (constitutive reporter) with specific spatial localisation of C4-dicarboxylates. This is novel and important result as it shows that plants spatially control exudation and it is not a random process. Further, the exudation of 14 metabolites has been mapped, however, these metabolites constitute only a fraction of the different compounds released during exudation. To develop new biosensors genomic techniques such as INSeq and RNA-seq can be used to identify genes involved in the transport and metabolism of carbon compounds in the rhizosphere which could then be used to produce new bioreporters. The application of INSeq to study rhizosphere growth is shown in chapter 4 of this thesis.

Chapter 4

Characterisation of root colonisation: A bacterial perspective

4.1 – Introduction

Bacterial colonisation of roots plays a critical role in the life cycle of land plants. These associations can be beneficial or deleterious to plant health. Despite the end result of these interactions, bacteria have had to develop a variety of strategies to survive within the rhizosphere and during colonisation of a host plant. The elucidation of key genes involved during colonisation is key to understanding how bacteria successfully colonise. Colonisation of plants by plant growth-promoting rhizobacteria (PGPR) is a key area of research. The nitrogen-fixing rhizobial species are some of the most well-known PGPRs; these species form symbiotic relationships with legumes and ultimately supply ammonia to the host legume. The importance of these species in agriculture makes them a strong focus for research and provides a model for the development of N_2-fixation in non-legume species such as cereals. Determining the genetic basis underpinning how rhizobia colonise cereal roots is one of the first steps in manipulating root colonisation to our advantage.

A technique known as insertion sequencing (INSeq), provides a rapid, high-throughput method to study the essentiality of genes and genomic regions within bacteria grown under a variety of conditions (Barquist *et al.,* 2013; Opijnen and Camilli, 2013). The basis of INSeq and its power lies in assessing the presence of unique transposon insertions within a pool of mutants whose transposition events have saturated the given organisms' genome. Mutagenesis is achieved through the use of the MmeI-adapted *mariner* based transposon pSAM-Rl. By complementing this technique with transcriptomic analyses, such as RNA-Seq, novel genes, their essentiality, and transcriptional state during colonisation can be determined.

Here I aimed to use INSeq as a method to identify new genes in *A. caulinodans* ORS571 that are involved in the colonisation of the *O. sativa* rhizosphere and roots. As with Lux analysis that directly reports on solute secretion, INSeq can also reveal

information about the chemical signals in the *O. sativa* root environment *i.e.* If genes involved in the transport and/or the metabolism of root exudates are shown to be essential for rhizosphere growth and/or successful root colonisation, then they can be tentatively described as responding to chemical signals influencing bacterial growth. By using INSeq to study which genes are essential for growth in the rhizosphere and the specific genes involved in the metabolism and transport of root derived carbon, it will be possible to produce new bioreporters for the spatiotemporal mapping of root exudation (as outlined in Chapter 3 of this thesis). To complement INSeq, a technique known as RNA-Seq was used to look at the bacterial transcriptome during rhizosphere and root colonisation, which allows the determination of a genes transcriptional state during growth under a condition of interest. By performing RNA-Seq on *A. caulinodans* grown under the same conditions as used for INSeq, then a gene's essentiality and transcriptional state can be determined and a more robust understanding of a gene's role during rhizosphere and root colonisation developed.

4.2 – Results

4.2.1 – Mutant generation in *Azorhizobium caulinodans* ORS571

The aim of this work was to determine if the MmeI-adapted *mariner* transposon, pSAM_RI (Perry and Yost, 2014), designed for use in *R. leguminosarum* is suitable for mutant generation in *A. caulinodans* ORS571. If suitable, ORS571 mutant pools would be generated and taken forward for use in an INSeq study.

Mutant pools in strain ORS571 were generated using pSAM_RI (Perry and Yost, 2014) (plasmid map shown in Figure 1.3). This plasmid was maintained in *E.coli* SM10*λpir.* pSAM_RI was introduced into ORS571 by conjugative transfer, as outlined in Materials and Methods 2.3.5. Upon transfer into ORS571, a 1345bp region of the plasmid containing the *nptII* gene, transposon borne Rho-independent

terminator and flanked by MmeI-adapted inverse repeats (IR R & IR L), undergoes transposition into a thiamine-adenine (TA) site within the host genome. The high frequency of conjugation and subsequent transposition allows the generation of millions of unique mutants via a relatively simple conjugative transfer. Perry and Yost (2014) showed a transposition frequency of 2.01 x 10^{-4} transposon mutants per recipient cell for *Rlv*3841, alongside this they showed frequencies for *S. meliloti* and *Agrobacterium tumefaciens* were 8.04 x 10^{-5} and 2.54 x 10^{-5} respectively, thus demonstrating the frequency of the transposition across multiple species. Three independent mutant libraries were generated using the pSAM_RI as outlined in Materials and Methods 2.5.1. These libraries consisted of upwards of 10^{9} cells and were subsequently used for INSeq analysis, labelled as 'Input' pools 1, 2 and 3.

To determine if transposition was successful, the presence of the mariner cassette was determined via PCR amplification. Four independent colonies were selected and PCR using insert specific primers (oxp0340 & oxp0341) generated a 1 kb fragment. A positive control (pSAM-Rl plasmid) and negative control (wild-type genomic DNA) were included. All four samples plus the positive control yielded a 1 kb fragment, with no amplified DNA present in the negative control (Supplementary 3). Once the presence of the insert was confirmed, arbitrary primed PCR (AP-PCR) was used to identify the locations of the inserts for all four samples (as outlined in Materials and Methods 2.3.3). Only three of the samples were successful in producing amplified DNA (Supplementary 3). This DNA was extracted, sequenced and used to map the location of the insert within the ORS571 genome. Insertion locations were as follows: AZC_RS19465 (Oxidoreductase), AZC_RS04985 (porin family protein) and AZC_RS05105 (3-oxoadipate CoA-transferase). These data confirmed the presence of different insertions within the ORS571 genome, indicating successful mutant generation suitable for INSeq.

4.2.2 – Determining the root carrying capacity for ORS571

To generate a reliable map of gene essentiality using the INSeq method, insertion libraries must be grown for a suitable number of generations under the condition of interest so that differences in insertion density can be easily distinguished. This work determined the number of generations ORS571 could grow through before saturating the growth system. This information was then used to determine inoculation densities and the growth period between inoculation and mutant isolation.

In this system a minimum of eight generations was chosen as a starting point as this number ensures that adequate amplification of the library is achieved, allowing for the separation of non-essential (NE), essential (ES), growth defective (GD) and growth advantageous (GA) genes. The minimum number of cells a system can be inoculated with is defined as the total number of TA sites present in the bacterial genome. ORS571 contains 58,163 unique sites and as such is the minimum number of cells required for inoculation. This inoculum is limited by the carrying capacity of the test system i.e. the number of given cells a bacterial population can reach before growth plateaus. To determine the carrying capacity of *O. sativa* cv. IR64 roots, plants were grown in fine vermiculite within large boiling tubes (outlined in Materials and Methods 2.4.1), for each time point, three replicate plants were used. This growth system allowed isolation of four fractions of bacteria:

1. Rhizosphere – vortexed vermiculite.
2. Loosely root-attached – vortexed roots.
3. Tightly root-attached – sonicated roots.
4. Endosphere – macerated roots.

For this study fractions 2,3 and 4 (loosely root-attached, tightly root-attached and endosphere respectively) were combined to provide the overall number of root-associated bacteria. Figure 4.1 shows the amplification of ORS571 during the colonisation of the *O. sativa* rhizosphere and roots from an initial inoculum of $1x10^4$ cells. These data show that the bacterial population reaches ~$5x10^8$ and ~$1x10^8$ cfu in the root-environment and root-associated fractions respectively. Thus, the bacteria in the rhizosphere, and associated with roots, passed through approximately 16 and 14 generations, respectively.

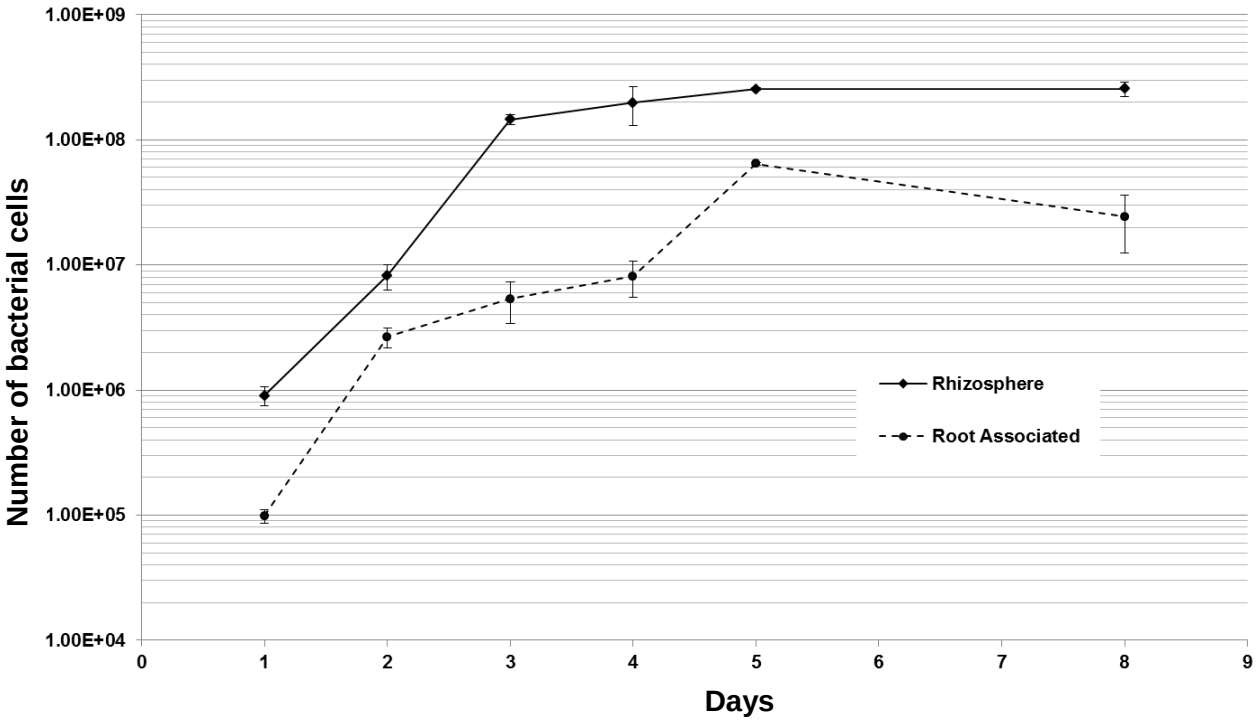


Figure 4.1. Expansion of *A. caulinodans* ORS571 population during the colonisation of the rice rhizosphere and roots; Rhizosphere represents bacteria isolated from the vermiculite plants were grown; Root-associated represents bacteria isolated from vortexed, sonicated and macerated roots. Each time point represents the average of three replicate plants. Error bars represent standard error of the mean.

4.2.3 – Library preparation

Using data from the carrying capacity experiments and the total number of TA sites in ORS571 (58,163), an inoculum density of 1 x 10^5 cells was chosen. This density allowed the mutants to grow through a minimum of 10 generations before reaching

saturation, giving a robust separation between mutants, this subsequently gives robust essentiality states for TA sites and genes.

Overall, the experimental design yielded a total of four conditions for sequencing:

1. Input – DNA isolated from mutant ORS571 library stocks.
2. Free-living – DNA isolated after mutant libraries were grown in minimal media at 28 ^{o}C.
3. Rhizosphere – DNA isolated from bacteria collected from the rhizosphere (*i.e*. fraction 1 above).
4. Root-attached –DNA isolated from bacteria collected from the root (*i.e*. fractions 2, 3 and 4 above).

As stated earlier, three independent libraries were produced (designated 1, 2 and 3) and used to obtain three biological replicates for each condition as outlined in Materials and Methods 2.5.2. In total, 12 libraries were produced: Input (Input 1-3), Free-living (Free-living 1-3), Rhizosphere (Rhizosphere 1-3) and Root-attached (Root-attached 1-3). Input libraries (1-3) act as the housekeeping/control condition e.g. a gene that is GD in the input and all other conditions can be said to be GD regardless of growth environment.

Library preparation, sequencing and analysis followed the method as outlined in Materials and Methods 2.5.4. Three of the libraries (Rhizosphere 2 and Root-attached 1 & 2) did not yield enough DNA for quantification (quantification was repeated three times for these samples) and were discarded before sequencing. The remaining 10 samples were taken forward for sequencing. On completion, a sequencing report was generated (Supplementary 4) outlining the quality of chip loading and sequencing. In total $1.27x10^{8}$ reads were achieved, quality analysis determined that 92% of these reads were usable, yielding a final total of $1.16x10^{8}$ usable reads. In total, nine FASTQ files were created; 3 x Input 1, 2 and 3, 3x Free-

living 1, 2 and 3, Rhizosphere 1 and 3 and Root-attached 3. Further quality control led to Rhizosphere library 1 being discarded due to low insertion density and read counts compared to the other libraries. The generation of sequencing libraries was repeated leading to the creation of Rhizosphere 1b & 2b and Root-attached 1b & 2b. After sequencing of these new samples, Rhizosphere 1b and Root-attached 1b were discarded due to low read values. The ten samples were taken forward for analysis were: Input 1, 2 & 3, Free-living 1, 2 & 3, Rhizosphere 2b & 3 and Root-attached 2b & 3.

4.2.4 INSeq results

The aim of this work was to use INSeq to identify genes required for growth of *A.* caulinodans in the *O. sativa* rhizosphere and during root colonisation. This information was then used to identify metabolic pathways important for bacterial growth in the rhizosphere and on *O. sativa* roots.

Analysis of sequencing data follows the analysis outlined in Materials and Methods 2.5.4. To determine the essentiality state for TA sites, a 4-state hidden Markov model (HMM) analysis as outlined by Dejesus and Loerger (2013) was performed. This categorised each TA site in the genome into one of four states: essential (ES), growth-defective (GD), neutral (NE) or growth-advantaged (GA). In simple terms, the state of any given TA site is ‘hidden’ but the state dependent output is ‘visible’. It is therefore possible to infer the state value by observing the output. In this case, the state of a site is one of ES, GD, NE or GA, whilst the output is the number of reads given for a TA site. Sites were designated as NE if the number of reads was equal or close to the mean number of reads for all non-empty sites within the genome. A site was designated as GD if it contained 100-times fewer reads than that defined by NE. A site was designated as GA if it contained at least 5-times the number of reads than NE. Finally, a site was defined as ES if the number of reads it contained was very

close to zero (1 to 2 reads). Sites were then grouped into genes and non-coding regions and the genes were then classified into one of the four states (ES, GD, NE or GA). This was achieved in two ways: Firstly, a state was assigned based on the most frequent site state that was found within the gene boundaries. Secondly, a gene was labelled as ES if the minimum run of sites labelled as ES was at least 3-standard deviations above the maximum predicted run of unoccupied sites in the gene, based on the Extreme Value Distribution. State-calls for TA sites present in non-coding regions were discarded. Data were now outputted as a text file containing a list of genes and their essentiality scores. Finally, each gene was matched to its gene product e.g. AZC_RS00010 = phosphoenolpyruvate synthase regulatory protein. This analysis was repeated after pooling (concatenating) the FASTQ files for each replicate together. This provided a final pooled data for analysis. Genes, products and essentiality for each condition is provided in Supplementary 5. All analysis of gene states calls was performed on both the un-pooled and pooled data sets, this allowed for an easier and more refined analysis, unless otherwise stated analysis is given for the pooled data sets. The accuracy of the HMM analyses state calls were confirmed by looking at the state calls given for known key ES genes. If these genes were not described as ES then the data could be described as unreliable. The states of genes encoding for ATP-synthase, ribosomal subunits, and DNA polymerases, were chosen as indicators and all were given an ES state.

The percentage of genes of ORS571 for each HMM state and for each condition is shown in Table 4.1. Analysis identified that in this experiment, the ORS571 core essential genome (overlap of these four different conditions) is made up of 574 growth-impaired (GI) genes (genes that are classified as ES or GD) (Fig. 4.2). By representing the data within a Venn diagram, it is possible to see the number of GI genes unique to each condition and multiple conditions. The number of GI genes unique to Root-attached conditions is 13 and only three genes are uniquely growth-

impaired in Rhizosphere. There are forty-six GI genes common between these two conditions which are listed in Table 4.2. These forty-six genes make up the colonisation core genome, that is to say, genes which are required for successful ORS571 colonisation of the *O. sativa* rhizosphere and roots. Within these core growth-impaired genes there are distinct sets of genes that appear to be involved in a variety of processes, including: oxalate metabolism/detoxification and propionate metabolism. These gene sets can be classified into distinct operons.

Table 4.1 Percentage of genes in ORS571 for each given HMM state. ES – growth essential; GD – growth-defective; NE – growth neutral; GA – growth advantageous.

State	Input	Free-living	Rhizosphere	Root-attached
ES	11%	13%	13%	13%
GD	1%	1%	2%	2%
NE	85%	81%	75%	76%
GA	3%	4%	10%	9%

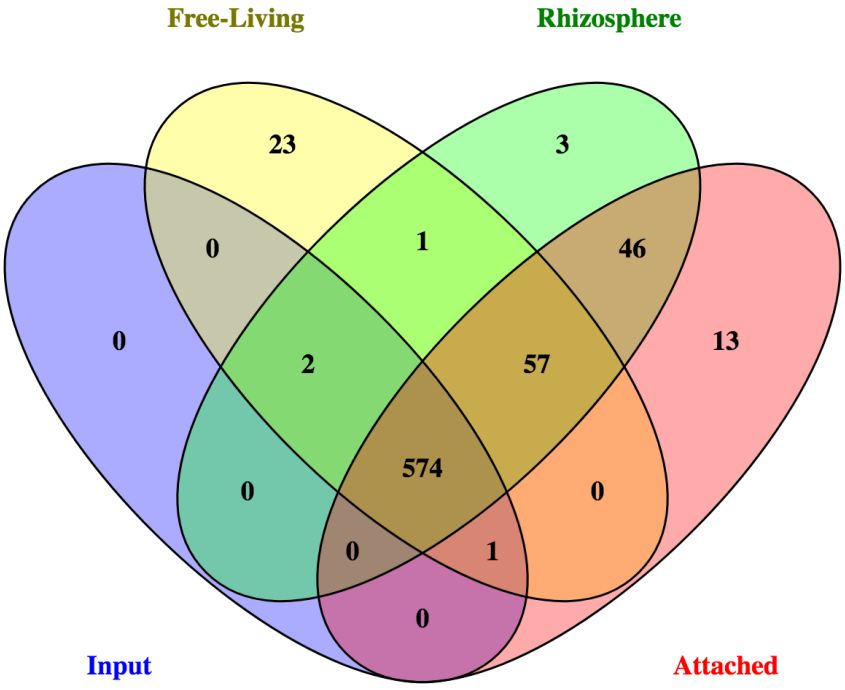


Figure 4.2 Venn diagram showing the combined number of growth-impaired (GI) genes (defined as essential (ES) and growth-defective (GD) genes) observed for *A. caulinodans* ORS571 across all four conditions; Free-living, Rhizosphere, Root-attached and Input libraries.

Table 4.2 Forty-six genes, common between Rhizosphere and Root-attached, which show growth impairment (GI) (i.e. genes shown to be essential (ES) or growth-defective (GD)), as identified by HMM analysis of INSeq data. * indicates genes and products discussed further in text.

Locus	Product
AZC_RS00225	hypothetical protein
AZC_RS00255	hypothetical protein
AZC_RS00260	methyltransferase
AZC_RS02325	hypothetical protein
AZC_RS02370	GntR family transcriptional regulator*
AZC_RS02380	Formyl-CoA decarbocylase*
AZC_RS02670	tol-pal system protein YbgF
AZC_RS05845	bifunctional glutamine-synthetase adenylyltransferase/deadenyltransferase
AZC_RS06250	Porin
AZC_RS06445	NYN domain-containing protein
AZC_RS09005	twin arginine-targeting protein translocase TatB
AZC_RS09045	hypothetical protein
AZC_RS09050	dihydrolipoyl dehydrogenase
AZC_RS09195*	methylmalonyl-CoA mutase*
AZC_RS09200*	methylmalonyl-CoA mutase*
AZC_RS10365	GTP pyrophosphokinase
AZC_RS10860	cytochrome ubiquinol oxidase subunit II
AZC_RS10865	cytochrome ubiquinol oxidase subunit I
AZC_RS11870	integration host factor subunit alpha
AZC_RS12235	hypothetical protein
AZC_RS12455	short-chain dehydrogenase
AZC_RS15130	thiol:disulfide oxidoreductase
AZC_RS15285	protoheme IX farnesyltransferase
AZC_RS16425*	endoribonuclease L-PSP*
AZC_RS16430*	4-hydroxybenzoyl-CoA thioesterase*
AZC_RS16435*	acyl-CoA dehydrogenase*
AZC_RS16440*	enoyl-CoA hydratase*
AZC_RS16445*	3-hydroxyacyl-CoA dehydrogenase*
AZC_RS18030	coenzyme A pyrophosphatase
AZC_RS18040	ATPase AAA
AZC_RS18045	hypothetical protein
AZC_RS18050	LytTR family transcriptional regulator
AZC_RS18610	SUF system Fe-S cluster assembly protein
AZC_RS18885	polyphosphate kinase
AZC_RS19165	Zn-dependent protease
AZC_RS20105*	2-polyprenylphenol 6-hydroxylase*
AZC_RS20395	two-component sensor histidine kinase
AZC_RS20590*	succinate--CoA ligase subunit alpha*
AZC_RS20595*	succinate--CoA ligase subunit beta*
AZC_RS20600	malate dehydrogenase
AZC_RS22835	geranylgeranyl pyrophosphate synthase
AZC_RS22850	thiol:disulfide interchange protein
AZC_RS22855	transcriptional regulator
AZC_RS22860	cytochrome c-type biogenesis protein
AZC_RS22875	heme ABC exporter, ATP-binding protein CcmA
AZC_RS24135	16S rRNA methyltransferase

The GI genes AZC_RS09195 and AZC_RS09200 (encoding methylmalonyl-CoA mutase subunits) form an operon with AZC_RS09205 and AZC_RS09210 (encoding two subunits of acetyl/propionyl-CoA carboxylase. Whilst the latter two genes are not classed as GI within the pooled data (and are therefore not listed in table 4.2), they are classed as GI in the Rhizosphere 3 and Root-attached 3 data. Using IGV to view the pooled reads, the number of reads appears lower in the Rhizosphere and Root-attached datasets when compared to data from the Free-living samples. To confirm this apparent difference a resampling analysis was performed using the TRANSIT software package (DeJesus *et al.,* 2015). Resampling looks to see if there is a statistically significant difference in reads for genes between two data sets and gives a statistical score based on permutation tests. It is critical to note that this analysis looks at the total number of TA sites and total number of reads per gene; it is not looking at reads for each individual TA site within a gene. Take, for example, a gene containing 10 TA sites, if the first TA site in this gene was shown to contain 500 insertions and the remaining 9 sites showed no insertions then this gene would be scored as ES using HMM analysis; in a separate condition each of the 10 TA sites contained 50 insertions (for a total of 500 insertions) the gene would be scored as NE under HMM analysis. If these two conditions were compared using resampling, no difference between the two genes would be seen as they both contain a total of 500 insertions. The Rhizosphere and Free-living datasets were compared, as were the Root-attached and Free-living datasets. In both comparisons, all four genes (AZC_RS09195, AZC_RS09200, AZC_ RS09205 and AZC_RS09210) showed a statistically significant difference (p-value <0.05) when compared to the Free-living data. In all but one case, the p-values were <0.01 (Rhizosphere vs Free-living, AZC_RS09205, p-value = 0.0156). This type of analysis only shows if there is a difference in reads between two conditions and does not give a growth-state. The combination of the pooled and un-pooled HMM data with a resampling analysis indicates this operon is required for robust bacterial growth in both the rhizosphere

and during root attachment. The predicted pathway based on genomic annotations and the KEGG database is shown in Figure 4.3.

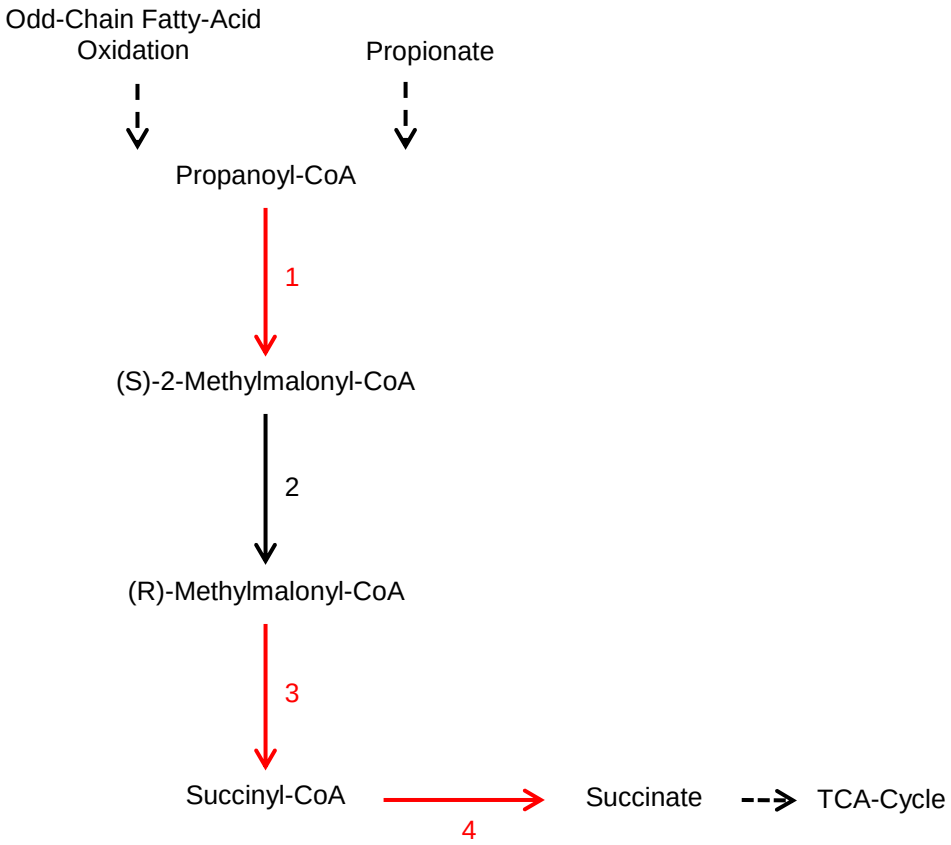


Figure 4.3. The pathway for propionyl-CoA metabolism in *A. caulinodans* ORS571. Red arrows indicate reactions carried out by genes whose growth is impaired during colonisation; Black arrows indicate growth-neutral genes. 1 = Propionyl-CoA carboxylase; 2 = Methylmalonyl-CoA epimerase; 3 = Methylmalonyl-CoA mutase; 4 = Succinyl-CoA synthetase.

Genes encoding for succinyl-CoA synthetase were also determined to be GI in Rhizosphere and Root-attachment, and are neutral during free-living growth. When the original libraries were grown in their free-living state 10 mM succinate was used as the carbon source. Under such conditions mutation of succinyl-CoA synthetase would not detrimentally affect the TCA cycle as succinate is still directly available to the cell. However, in the Rhizosphere samples where the concentrations of succinate are likely to be low, it becomes key that succinyl-CoA is converted to succinate to maintain the TCA cycle and bacterial growth. Tying this back into propionate degradation, I have shown that the key genes involved in the conversion of

propionate to succinyl-CoA are GI for rhizosphere growth and root attachment. This indicates that under these growth conditions ORS571 is metabolising propionyl-CoA for the production of succinyl-CoA for use within the TCA cycle (Figure 4.3).

The GI genes AZC_RS02370 and AZC_RS02380 (GntR family transcriptional regulator and formyl-CoA transferase, respectively) form an operon with two other genes, AZC_RS02375 and AZC_RS02385 (oxalyl-CoA decarboxylase and formly-CoA transferase, respectively). The latter two genes whilst not classed as GI in the pooled data (and are therefore not listed in Table 4.2), they are classed as GD in Rhizosphere 3 and Root-attached 3. Furthermore, when looking at the number of reads in IGV the number of reads for these genes drops, indicating that they are 'borderline GI'. It is likely that this operon is putatively involved in oxalate metabolism/detoxification. A resampling analysis does not show a statistically significant difference between AZC_RS02375 and AZC_RS02380 when comparing the Rhizosphere or Root-attached to the Free-living data. In fact, when comparing Root-attached to Free-living, a statistical difference is not seen, even though the HMM analysis gave a GD phenotype. This highlights how HMM and resampling can give two conflicting scores, with neither being incorrect. Despite AZC_RS02375 and AZC_RS02385 being labelled as NE, when taking into account the number of reads and comparing with the data from Free-living samples using IGV, these two genes appear to be 'borderline GI'. These data suggest that the metabolism/degradation of oxalate is important during the colonisation of the *O.* sativa rhizosphere and roots.

A third set of genes putatively forming part of an operon is also seen as GI in the rhizosphere and during root attachment. These genes have not been attributed to any specific pathway in ORS571, and their annotations are based on structural and sequence homology. The genes AZC_RS16425, RS16430, RS16435, RS16440 and RS16445 encode: *ridA* family protein (endoribonuclease L-PSP), 4-hydroxybenzoyl-

CoA thioesterase (4HBT), acyl-CoA dehydrogenase, enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase, respectively (Table 4.2). A sixth gene, AZC_RS16450 encoding a salicylyl-CoA 5-hydroxylase (S5H), is located at the start of the operon, but was not classed as GI. When using IGV to examine AZC_RS16450, there does appear to be a reduction in the number of reads towards the C-terminus when comparing the Rhizosphere and Root-attached with the Free-living and Input data. This difference is not significant with the resampling analysis method. Interestingly AZC_RS16425 is translationally-coupled with the C-terminus of AZC_RS16430. The function of this operon is hard to predict as these gene products have not been experimentally confirmed, and no obvious function for the operon is apparent. Three of the genes (AZC_RS16435, AZC_RS16440 and AZC_RS16445) appear to be involved in the β-oxidation of fatty acids. Salicylyl-CoA 5-hydroxyalse (AZC_RS16450) is involved in the conversion of salicylyl-CoA into gentisyl-CoA in the salicylate degradation pathway present in *Streptomyces sp.* strain WA46 (Ishiyama *et al.*, 2004). 4-HBT converts 4-hydrobenzoyl-CoA into 4-hydroxybenzoate (4-HB) and CoA. 4-HB is used in a variety of pathways; most notably it is a key compound for the biosynthesis of ubiquinone. Separate from ubiquinone biosynthesis, during the degradation of aromatic compounds, 4-HB is degraded via the protocatechuate branch of the β-ketoadipate pathway. None of the genes involved in this pathway, that can be identified in the ORS571 genome, show any growth impairment and were classified as NE. Importantly 4-hydroxybenzoate hydroxylase (4-hydroxybenzoate 3-monooxygenease) is NE under all conditions; this enzyme converts 4-HB into protocatechuate, before entering the β-ketoadipate pathway for production of TCA cycle-intermediates. Furthermore, a homologue of this hydroxylase was shown to be involved in 4-HB transport into *R. leguminosarum* (Wong *et al.,* 1994, Ramachandran *et al.,* 2011) and it is unlikely that the operon seen is involved in protocatechuate degradation.

The gene AZC_RS20105, from sequence and structural homology with known proteins, is predicted to encode for a 2-polyprenylphenol 6-hydroxylase. This enzyme is key in the aerobic conversion of 2-polyprenylphenol to 2-polyprenyl-6-hydroxyphenol during ubiquinone biosynthesis. This was classified as GI in the Rhizosphere and Root-attached samples, whilst being NE in Free-living and Input samples. It should be noted that there is a clear reduction in the number of reads across all TA sites within this gene compared to adjacent genes for both Input and Free-living conditions, which would indicate the gene is becoming GD and thus is truly ES in the rhizosphere and during root attachment. Interestingly all other genes involved in ubiquinone biosynthesis show a ES phenotype across all tested conditions. This result suggests that the aerobic biosynthesis of 2-polyprenyl-6-hydroxyphenol is critical for the survival of ORS571 in the rhizosphere.

4.2.5 – RNA-Seq

RNA-Seq identifies the transcriptional state of genes during growth of bacteria under conditions of interest which can be complementary to results produced by an INSeq study. This work focussed on determining if RNA-Seq was suitable for the study of rhizosphere and root colonisation, and if so, combining the results of RNA-Seq and INSeq to further probe the roles of genes during colonisation.

For this study, *O. sativa* was grown under the same conditions as those used for INSeq (Materials and Methods 2.4.1) and inoculated with wild-type *A. caulinodans* ORS571. In total 60 plants were grown and inoculated, these were split into three fractions to produce biological replicates. Bacteria were isolated and RNA extracted as outlined in Materials and Methods 2.6.2. This yielded six RNA samples: Rhizosphere 1, 2 and 3, and Root-Attached 1, 2 and 3. RNA was also extracted from ORS571 grown under free-living conditions in TY medium (Materials and Methods 2.4.1), this acted as a control for both gene expression and that the library

preparation was successful. Each sample underwent bacterial mRNA purification and cDNA synthesis as outlined in Materials and Methods 2.6.2. Due to low RNA yields, the rhizosphere samples were combined into one sample prior to purification, as were the Root-attached samples. In total, five RNA samples were produced and labelled as Free-living 1, 2 and 3, Rhizosphere and Root-attached. RNA concentrations were checked multiple times during purification and deemed high enough to continue, the final cDNA libraries showed similar concentrations for all samples (Supplementary 7). Once the cDNA libraries were produced they were each diluted to 75 pM, all five libraries were then combined in equal volumes for a final volume of 50 µL which was used for sequencing as outlined in Materials and Methods 2.6.3. Once sequencing was completed, a sequencing report was generated (Supplementary 8). In total, $7.69x10^7$ reads was achieved, however, only 55% of these were usable giving a final read count of $4.23x10^8$ reads. Reads were grouped into samples based on their unique barcodes, for each sample the total number of reads, usable reads and mean read length is shown in Table 4.3. Around 7% of the total number of reads did not have barcodes so were discarded before analysis. It should be noted that the Rhizosphere and Root-attached samples showed low read numbers, 2.8- and 2.5-times fewer reads, respectively, than Free-living 2 (Free-living sample with lowest number of reads). All data was output as FASTQ files.

Table 4.3 Number of reads and mean read length generated from Ion-torrent sequencing of RNA-Seq libraries. Libraries were generated from *A. caulinodans* ORS571 grown under free-living conditions and during colonisation of the *O. sativa* rhizosphere and roots.

Sample	Total Reads	Useable Reads
No Barcode	5.01E+06	2.76E+06
Rhizosphere	5.93E+06	3.26E+06
Root-attached	6.47E+06	3.56E+06
Free-Living 1	1.85E+07	1.02E+07
Free-Living 2	1.68E+07	9.23E+06
Free-Living 3	2.42E+07	1.33E+07

EDGE-pro (Magoc *et al*., 2013) was used to map reads to the ORS571 genome and for counting reads, this data is shown in Supplementary 9. After mapping and counting, Free-living 1, 2 and 3 were concatenated to form a single Free-living data set. For both the Rhizosphere and Root-attached the average number of counts across all genes was very low at 6 and 9 reads respectively, whilst the average number of reads for the concatenated Free-living was 914 reads (non-concatenated Free-living 1, 2 and 3 was 100, 343, 470 respectively). The extremely low number of reads for the Rhizosphere and Root-attached data made meaningful analysis almost impossible and only genes that showed a large increase in reads, a minimum of15-fold, were deemed as 'up-regulated' and analysed. These genes are highlighted in Table 4.4.

Table 4.4 Genes of *A. caulinodans* that up-regulated in a) the rhizosphere and b) during root colonisation/attachment of *O. sativa*, identified by RNA-Seq. Essentiality identified using INSeq is shown for each gene.

Rhizosphere	**Gene Product**	**Essentiality**
AZC_RS05040*	arginine decarboxylase	NE
AZC_RS05520*	HlyD family secretion protein	NE
AZC_RS05980	Transglutaminase	NE
AZC_RS09425	circularly permuted type 2 ATP-grasp protein	NE
AZC_RS12165*	nitrate ABC transporter substrate-binding protein	NE
AZC_RS14270*	MFS transporter	NE
AZC_RS19180*	GGDEF domain-containing protein	GA
AZC_RS19625*	chitooligosaccharide synthase NodC	NE
AZC_RS20970*	NAD(P)H dehydrogenase (quinone)	GA
AZC_RS23110	NAD-glutamate dehydrogenase	NE
AZC_RS23350	UDP-N-acetylmuramate--L-alanine ligase	ES

Root-attached	**Gene Product**	
AZC_RS05025*	PRC-barrel domain containing protein	NE
AZC_RS05965	formate dehydrogenase-N subunit alpha	NE
AZC_RS14270*	MFS transporter	NE
AZC_RS21125*	methylmalonate-semialdehyde dehydrogenase (CoA acylating)	NE
AZC_RS22780*	FAD-binding oxidoreductase	NE
AZC_RS23350	UDP-N-acetylmuramate--L-alanine ligase	ES

This cut-off was determined so that read value for each ‘up-regulated’ gene was more than three standard deviations greater than the mean value for the data set. Thus, genes in Table 4.4 can be considered up-regulated, although the very low read numbers meant that this analysis is fraught with difficulty.

In Table 4.4, genes which were ‘down-regulated’ in the Free-living data are marked with asterisks (these genes showed a minimum of two-fold decrease in reads compared to the average for Free-living data).Although the high number of reads acquired in the Free-living data sets highlight that adequate RNA can be isolated from ORS571 for RNA-Seq, this data was not analysed in detail (Supplementary 9). Due to the low number of reads, annotation quality and time-constraints, a meaningful in-depth analysis of the up-regulated genes identified in the Rhizosphere and Root-attached samples is not possible without further experimental repeats and better quality data. For these reasons this data was given only a cursory analysis. The majority of the genes shown in Table 4 were down-regulated during free-living growth, which might lead to the theory that they may be expressed specifically for rhizosphere and root-attached growth, but this would require further data before any conclusions could be drawn.

A methylmalonate-semialdehyde dehydrogenase (AZC_RS21125) was identified in the Root-attached data as being upregulated. This dehydrogenase is involved in the conversion of methylmalonate-semialdehyde (produced as an intermediate in valine metabolism) to propionyl-CoA. This links back to the INSeq study which showed the conversion of propionyl-CoA to succinate as being essential for growth.

In both the Rhizosphere and Root-attached data a gene encoding for a UDP-N-acetylmuramate--L-alanine ligase (AZC_RS23350) was up-regulated, but not during free-living growth. Based on annotations this ligase forms part of the peptidoglycan

biosynthesis pathway for cell wall synthesis/strengthening. Further, it is categorised as ES by INSeq, under all conditions tested (Input, Free-living, Rhizosphere and Root-attached) indicative of its importance for bacterial growth in general.

4.3 – Discussion

While INSeq has already been shown to be a powerful tool to determine which genes are essential during bacterial growth in a diverse range of environments, here I have shown the application of this technique to study the colonisation of the roots of *O. sativa* using *A. caulinodans* ORS571. I have demonstrated that the use of the modified pSAM-Rl vector (Perry and Yost, 2014) designed for use in *R. leguminosarum* is compatible for use in *A. caulinodans,* without the need for modifications to the plasmid. In my experiments, the high rate of transposition frequency obtained and the mutant pools constructed were sufficiently large for successful INSeq analysis in an experimental system examining root colonisation.

This work showed that ORS571 genes involved in bacterial colonisation and *O. sativa* root attachment can be identified. In these experiments the number of genes specific for growth in any one condition was relatively low. For free-living, rhizosphere and root- attached growth, only 23, 3 and 13 genes, respectively, resulted in impaired bacterial growth (Fig. 4.1). This is interesting as a higher proportion of genes were expected to be important for growth under these conditions. Regardless distinct operons were isolated which appear to be key in colonisation of the *O. sativa* rhizosphere by ORS571. The loss of library 1 for the rhizosphere and attached fractions highlights one of the drawbacks to the INSeq method, in that the method relies on isolating an adequate concentration of target DNA for sequencing. Whilst this is straightforward for ‘free-living’ cultures of bacteria, it can become more difficult if the complexity of a system is increased *i.e.* adding other organisms to the

experimental condition. In the case of this work, isolation and amplification of bacterial DNA was hindered by the presence of large concentrations of plant DNA, which necessitated the growth of mutant libraries isolated from the rhizosphere and roots for 15 hours (roughly 3 generations) in complex media to increase cell numbers and thus DNA concentration.

Most notably genes involved in propionate-CoA metabolism and the subsequent conversion of succinyl-CoA to succinate indicate that the metabolism of propionate-CoA to provide succinate for the TCA cycle is mandatory for ORS571 growth within the rhizosphere and on root surfaces. This indicates that the levels of succinate are insufficient to maintain bacterial growth on their own. I have shown that *O. sativa* releases the C4-dicarboxylates: malate, succinate and aspartate into the rhizosphere indicating they are available for bacterial growth, though the concentrations of succinate released is unknown in our system. It has however been shown in 24 day old Alfalfa that the concentrations of succinate in root exudates did not exceed 2.84 µmol/g of dry root (Lipton *et al.,* 1987), whilst from *O. sativa* cv. IR64 seedlings the rate of succinate release was around 1.6 µmol plant^{-1} d^{-1} (Aulakh *et al.,* 2001). If similar concentrations are being released in the experimental set up I used, then over the 15 days of *O. sativa* growth I would expect no more than 24 µM of succinate to be present in the rhizosphere. This concentration is below what is required to support the growth of ORS571, indicating that succinate is being supplied from another pathway, in this case possibly propionate-CoA metabolism. Further, the presence of propionate in *O. sativa* exudates has been shown experimentally in the literature (Breidenbach *et al.*, 2016). Furthermore, ORS571 has been shown to grow on propionate as a sole carbon source (Dreyfus *et al.*, 1988; Kitts *et al.*, 1989) though this growth was poorer than was seen on malate or succinate. This would lead to the assumption that ORS571 is capable of metabolising propionate to ultimately produce succinate for use in the TCA-cycle. The gene *prpE* encodes for propionyl-CoA

synthetase an enzyme that catalyses the conversion of propionate to propionyl-CoA across many different species which is a key step for the metabolism of propionate. This gene in ORS571 shows no negative phenotype, and based on total read numbers provides an advantage to growth when mutated. This would suggest that the propionyl-CoA used for succinate production is not derived from propionate itself; rather it is being supplied as a product of other metabolic reactions within the cell. There are however, two acyl-CoA synthetases present within the ORS571 genome. Acyl-CoA synthetases can use a variety of short carbon chains, and it is possible that they are acting as a redundancy when *prpE* is non-functional. One other possible source of propionyl-CoA is via the oxidation of odd-chain fatty acids, which leads to the production of propionyl-CoA that can be used to produce succinate following the pathway outlined in figure 4.4. (Halarnkar and Blomquist, 1989; Dunn *et al.*, 2004). With *prpE* being scored as NE there is some credence to this theory that fatty acids are supplied by the plant, but there is no direct evidence for this in *O. sativa*.

Oxalate is produced in one form or another as a metabolic end product within plant cells. It is produced in soluble forms (oxalate) as well as insoluble forms (calcium oxalate), and its presence has been detected in *O. sativa* (Libert and Franceschi, 1987; Sahin, 2003). A number of oxalate metabolising bacteria have been isolated and characterised, these so called oxalotrophs can utilise oxalate as a sole-carbon source for growth. A number of rhizobia including the genera: *Mesorhizobia*, *Bradyrhizobiaceae* and *Xanthobacter* have been identified to grow on oxalate (Koch *et al.*, 2014). In particular *Azorhizobium oxalatiphilum* was recently identified (Sahin, 2003, 2005; Lang *et al.*, 2013), but interestingly ORS571 was shown to be unable to grow on oxalate as a sole carbon source (Dreyfus *et al.*, 1988). This is interesting as genes putatively involved in the conversion of oxalate to formate were identified from this INSeq study. These genes encoded for: formyl-CoA transferases, oxalyl-CoA decarboxylase, and an associated transcriptional regulator. If this operon has been

correctly identified it would be involved in the conversion of oxalate to formate following the pathway shown in figure 4.5.

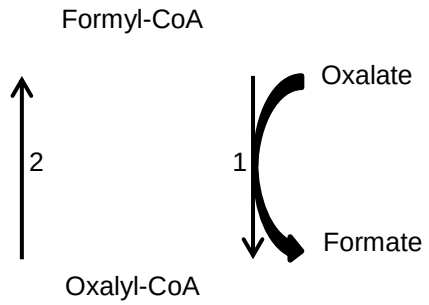


Figure 4.5. Proposed pathway for oxalate metabolism in *A. caulinodans* ORS571. 1 = Formyl-CoA transferase; 2 = Oxalyl-CoA decarboxylase

During the conversion of oxalyl-CoA to formate, a small percentage of the produced oxalyl-CoA can be directed to the production of glyoxylate via an oxalyl-CoA reductase. This glyoxylate can then be converted to glycerate and finally glycerate-3-phosphate, for use in glycolysis. The conversion of glyoxylate to glycerate can proceed by one of two intermediary pathways: One, the conversion to tartronic semialdehyde; Two, the conversion to hydroxypyruvate (Sahin, 2003). These pathways are outlined in a review by Sahin, (2003). None of the genes involved in these pathways could be positively identified in ORS571 and any genes tenuously linked by weak homology were categorised as neutral. Oxalate/formate MFS antiporters are present in ORS571 (AZC_RS18720, RS18725, and RS19270) suggesting that transport of oxalate can occur. AZC_RS18720 and RS18725 were scored neutral across all conditions, whilst AZC_RS19260 was scored as GA. These results indicate that plant-derived oxalate is not being transported into the cell during rhizosphere growth and root attachment. This makes it unlikely that oxalate metabolism is progressing in the same manner as is seen during the growth of other oxalatrophs. In *Bradyrhizobium japonicum* the catabolism of L-arabinose proceeded to formate with oxalate acting as an intermediate (Koch *et al.*, 2014). There is no

evidence that this pathway occurs in ORS571. Only one gene from this pathway was identified based on annotations and homology and was ranked as NE across all conditions. This gene encodes for a glycolate oxidase which catalyses the oxidation of glycolate to glyoxylate, making it a key step in the pathway proposed for *B. japonicum*. This suggests that oxalate is not being produced via L-arabinose metabolism in ORS571.

RNA-Seq was used to examine gene expression during ORS571 colonisation of the rhizosphere and roots of *O. sativa*. Growth of ORS571 under free-living conditions was intended to be used as a baseline for gene expression and as a control to show if the methods of RNA extraction, purification, library preparation and sequencing were successful. This showed that the protocols for RNA extraction to sequencing, as well as subsequent mapping and counting, were successful. However, due to the extremely low reads in the rhizosphere and root-attached samples, which arose from difficulties in extracting RNA from Rhizosphere/Root attached samples an in-depth analysis was impossible and no clear conclusions could be drawn from these initial experiments. As the average number of reads was so low, no down-regulated genes could be identified, and only a handful of up-regulated genes were identified. Furthermore, genes identified by INSeq as essential for rhizosphere and root-attached growth were cross-referenced with this data and none of were up-regulated.

Out of the putatively up-regulated genes identified by RNA-Seq, two were of interest. The first of these encoded a methylmalonate-semialdehyde dehydrogenase, this enzyme is involved in the conversion of methylmalonate-semialdehyde (MMS) to propionyl-CoA. MMS is produced as an intermediate during valine metabolism and in *P. aeruginosa* used as a source of propionyl-CoA (Bannerjee *et al.*, 1970; Steele *et al.*, 1992). INSeq highlighted that the conversion of propionyl-CoA to succinyl-CoA was critical for growth of ORS571 in the rhizosphere and during root-attachment. It is

possible that MMS dehydrogenase is a source of propionyl-CoA for metabolism to succinyl-CoA, however INSeq deemed MMS dehydrogenase as NE during rhizosphere and root-attached growth. This would indicate that it is not the sole source of propionyl-CoA for metabolism to succinyl-CoA. The second gene of interest encoded for a UDP-N-acetylmuramate-L-alanine ligase (MurB) which is involved in the production of peptidoglycan for cell wall biosynthesis/strengthening. This was up-regulated in both the rhizosphere and root-attached data. Interestingly *murB* appeared to be down-regulated in free-living growth conditions even though it was deemed essential for general bacterial growth under all conditions tested with INSeq. Indicating that *murB* is an essential part of bacterial growth, though it does not need to be expressed at high levels continually. This would suggest that ORS571 was growing rapidly during rhizosphere and root colonisation and/or an increase in peptidoglycan was needed to strengthen the cell walls in response to environmental stress, both of which are expected when any bacteria is growing and colonising plant systems.

In conclusion I have shown the suitability of the INSeq approach to identify novel genes and putative pathways which are involved in the colonisation of the plant rhizosphere and roots. The next step in this study would be to experimentally confirm the roles of these genes during colonisation. The first step would be to produce mutants of these genes and perform growth studies both as free-living organisms and during growth with *O. sativa*; this would confirm that the conclusions about fitness drawn from the INSeq are correct. Once they have been confirmed, then more in depth metabolic studies could be performed to elucidate their exact role and the metabolic pathways involved. Application of RNA-Seq to study gene expression in ORS571 has also been tested. This work has shown that the methods to isolate, extract and purify RNA from *A. caulinodans* are viable. However, to acquire adequate RNA concentrations for the analysis of rhizosphere and root colonisation, the number

of bacteria isolated need to be increased substantially; this should easily be achieved by increasing the number of plants grown and inoculated. Once adequate RNA has been isolated and sequenced, a differential sequencing analysis such as DESeq2 (Love *et al.*, 2014) could be used to robustly look at the fold-change in gene expression between a control condition (free-living growth) and test conditions (rhizosphere and root-attached growth).

Chapter 5

Characterisation of nitrogen-fixing growth

5.1 - Introduction

Nitrogen is one of the most important elements for the growth of any biological organism. It is present in the majority of all molecules required for biological growth, including: nucleic acids, amino-acids, and a variety of secondary products/metabolites. Whilst dinitrogen (N_2) makes up 78% of the earth's atmosphere, the strength of the triple bond between the two nitrogen atoms makes N_2 incredibly stable and thus inaccessible for the majority of biological organisms due to the energy required to break this bond. Some bacteria, termed diazotrophs, are capable of fixing atmospheric dinitrogen into biologically accessible ammonia. Much of the biologically accessible nitrogen in soil is produced from this fixation. To overcome the effects of nitrogen limitation, plants of the Fabaceae family (legumes) have evolved a symbiosis with a group of diaztotrophic bacteria known as rhizobia. Some of these species, such as *A. caulinodans*, are capable of fixing nitrogen during symbiosis as well as during free-living growth (Dreyfus, Garcia and Gillis, 1988; Peoples et al., 2009). The study of biological nitrogen fixation (BNF) is of great interest in terms of global food security, and a major goal is the transfer of BNF to non-leguminous crops such as the cereals (Geddes et al., 2015). Dissecting the molecular and genetic basis of BNF is key in improving the process and the possible transfer of symbioses to non-leguminous crops. Physiological and biochemical approaches have identified a range of key processes involved in BNF. Over the past decade the combination of genomic and transcriptomic analyses has helped further elucidate a range of genes and loci involved in the transport and metabolism of rhizobia during symbioses and fixation. These studies have often been focused heavily on fixation during symbiosis. To understand the basal genetics behind fixation separated from symbiosis, free-living diazotrophs must be used, *A. caulinodans* provides the opportunity to study fixation in free-living culture. When this species is grown microaerobically on minimal medium lacking a nitrogen source, BNF can occur (Dreyfus, Garcia and Gillis, 1988).

In this study, INSeq was used to study N_2-fixation in *A. caulinodans* ORS571. This same approach was also used in Chapter 4. This allowed for the identification of new genes involved in BNF. Importantly, this work highlighted that a comparative genome-wide technique can be used to identify new genes involved in a complex and agriculturally important process.

5.2 – Results

5.2.1 – Libraries and growth conditions

Before INSeq could be performed the conditions under which the mutant libraries were grown had to be shown to allow free-living N_2-fixation. To confirm this, *A. caulinodans* ORS571 wild-type and a Δ*nifA* incapable of N_2-fixation were grown under the test conditions, and N_2-fixation assessed by acetylene reduction (Materials and Methods 2.7.2). Both strains were grown minimal medium (Materials and Methods 2.7.1) at 3% oxygen. Growth was assayed visually after 48 hours. Acetylene reductions showed that N_2-fixation was occurring (Supplementary 10) in wild-type ORS571 but not in ORS571-Δ*nifA*, establishing that these conditions were suitable for N_2-fixation. Growth curves were produced for ORS571 and ORS571-Δ*nifA* grown under N_2-fixing conditions in UMS (Materials and Methods 2.7.1). ORS571-Δ*nifA* showed substantially lower growth than wild-type ORS571, indicative that ORS571-Δ*nifA* cannot synthesise NH_3^+ for growth via N_2-fixation. The small amount of growth seen for ORS571-Δ*nifA* under these conditions is likely due to stored cellular nitrogen.

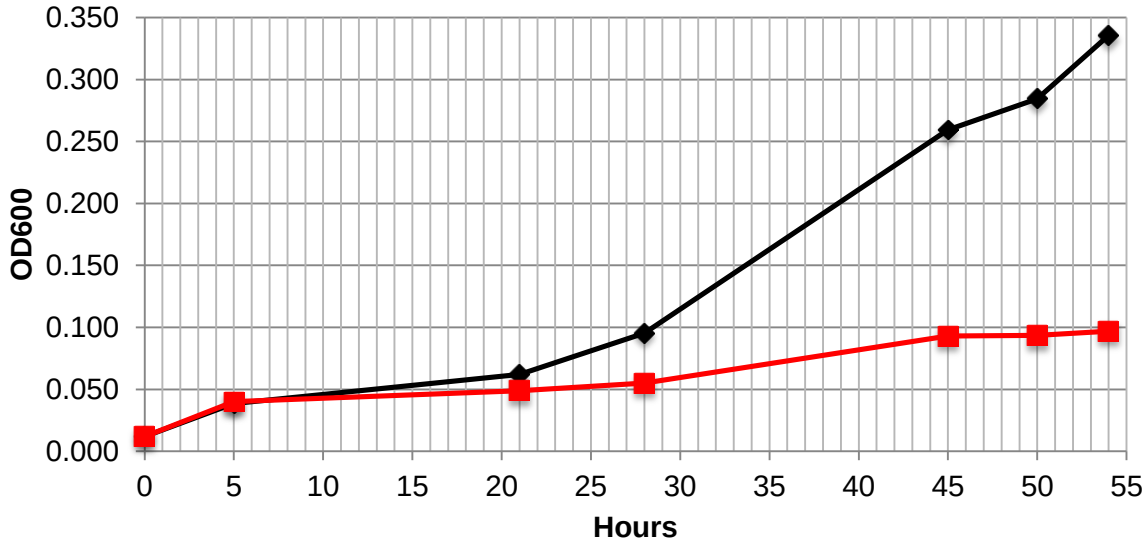


Figure 5.2 Growth of *Azorhizobium caulinodans* ORS571 and ORS571-Δ*nifA* under nitrogen fixing conditions. Measured by the increase in absorbance of 600 nm light, values are the average of two replicates.

5.2.2 – Insertion Sequencing

All growth was performed on solid medium using 120 mm x 120 mm square Petri plates. A maximum of 5 x 10^4 separated colonies could be grown on these Petri plates, using this, and the number of TA sites in ORS571 (58,163), the inoculation for each plate, and thus the overall experiment was determined. A total of 2.5 x 10^4 cells were spread onto a single plate (using half of the maximum number of cells ensured individual colonies did not overlap), five replicate plates were produced for each library for a total inoculation of 1.25 x 10^5 cells. The three independent mutant libraries produced for use in chapter 4 were reused here to produce three biological replicates for each growth condition.

Overall, the experimental design yielded a total of three conditions for sequencing:

1. Ambient – cells grown at 21% oxygen (supplemented with NH_4^+)
2. Microaerobic – cells grown at 3% oxygen (supplemented with NH_4^+)
3. Fixing – cells grown at 21% oxygen (not supplemented with NH_4^+)

As stated earlier, three independent libraries were produced and used to produce three biological replicates for each condition. The input data set produced as outlined in chapter 4 was reused here. After the libraries were isolated from the plates, replicates were pooled to yield nine total libraries (three replicates for each condition). Library preparation followed the protocol outlined Materials and Methods 2.5.3. Before analysis, replicates for each condition were concatenated together to yield a total of three FASTQ files covering ambient, microaerobic and N_2-fixing conditions.

5.2.3 – Sequencing Analysis

This work aimed to use INSeq to identify genes that are essential for *A. caulinodans* growth under low nitrogen conditions and during N_2-fixation. This was achieved by utilizing a HMM analysis to determine ES, GD, NE and GA states for all data sets and a resampling analysis to look at differences in reads between data sets. All data is outlined in Supplementary 11.

HMM analysis outlined a core functional genome of 641 genes which were classified as GI (genes classified as either ES or GD) across all conditions. Thirteen genes were identified as GI during N_2-fixing growth, and four genes labelled as GI under 3% O_2 non-N_2-fixing growth. Sixteen genes were identified as GI for generalised microaerobic growth. These data are represented as a Venn diagram (Fig 5.1), with a list of identified genes and their products shown in Table 5.1 for the non-fixing, fixing and microaerobic growth conditions.

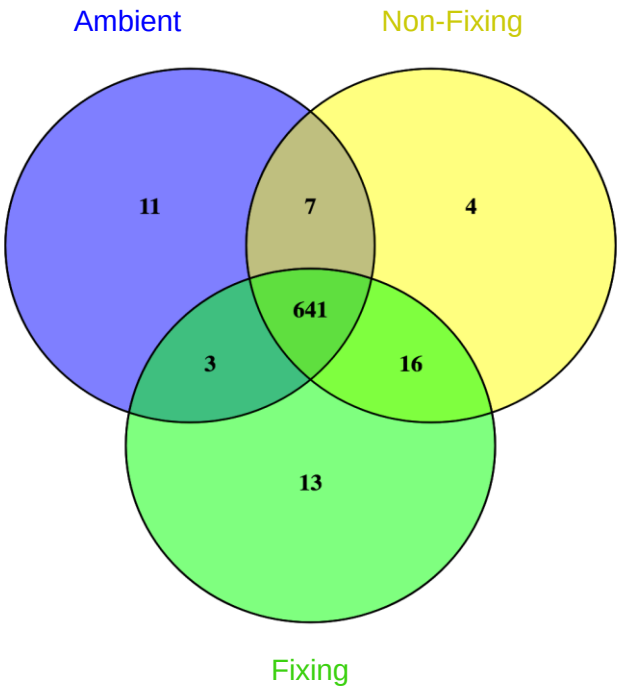


Figure 5.1 Venn diagram showing the combined number of growth impaired genes observed for *A. caulinodans* across all three conditions. Ambient – Growth at 21% O_2; Non-fixing – Growth at 3% O_2 with supplemented nitrogen; Fixing – Growth at 3% O_2 with no supplemented nitrogen.

Table 5.1 Growth impaired genes for *A. caulinodans* ORS571 under fixing, non-fixing, and microaerobic conditions, as identified by a Hidden Markov Model analysis. Fixing conditions represent growth under 3% O_2 with no nitrogen source; Non-fixing conditions represent growth under 3% O_2 with supplemented nitrogen; Microaerobic conditions represent growth under 3% O_2 independent of nitrogen availability.

Fixing conditions	
AZC_RS00470	aldehyde-activating protein
AZC_RS04515	elongation factor Tu
AZC_RS04600	hypothetical protein
AZC_RS05845	bifunctional glutamine-synthetase
AZC_RS12315	3'(2'),5'-bisphosphate nucleotidase CysQ
AZC_RS14570	hypothetical protein
AZC_RS15920	nitrogen regulation protein NR(I)
AZC_RS15925	PAS domain-containing sensor histidine kinase
AZC_RS15930	tRNA dihydrouridine synthase DusB
AZC_RS18380	hypothetical protein
AZC_RS20500	efflux RND transporter periplasmic adaptor
AZC_RS21230	hypothetical protein
AZC_RS24670	hypothetical protein

Non-Fixing conditions	
AZC_RS01695	type III PLP-dependent enzyme
AZC_RS05295	DUF1153 domain-containing protein
AZC_RS13310	tRNA (adenosine(37)-N6)-dimethylallyltransferase MiaA
AZC_RS14530	MBOAT family protein

Microaerobic conditions	
AZC_RS01685	translational GTPase TypA
AZC_RS03600	hypothetical protein
AZC_RS05205	complex I NDUFA9 subunit family protein
AZC_RS06175	molecular chaperone DnaJ
AZC_RS06445	NYN domain-containing protein
AZC_RS08810	phosphatidate cytidylyltransferase
AZC_RS09060	ubiquinone-binding protein
AZC_RS10870	cytochrome o ubiquinol oxidase subunit III
AZC_RS11645	lipopolysaccharide heptosyltransferase II
AZC_RS12455	NAD(P)-dependent oxidoreductase
AZC_RS16060	glycosyltransferase family 2 protein
AZC_RS18705	6-phosphogluconolactonase
AZC_RS21880	ATP F0F1 synthase subunit B
AZC_RS23300	ribosomal RNA small subunit methyltransferase H
AZC_RS23305	transcriptional regulator MraZ
AZC_RS24110	GNAT family N-acetyltransferase

In order to assess if the growth conditions were suitable to separate between fixing and non-fixing growth , I analysed the *nif* and *fix* genes as these are critical for N_2-fixation. These genes should have been GD or ES under N_2-fixing conditions and NE under all other conditions tested. All *nif* and fix genes were labelled as NE under all growth conditions (this will be elaborated upon in the discussion). To determine if there was a significant drop in the total number of reads for these genes under fixing conditions compared to non-fixing conditions (microaerobic and ambient), a resampling approach was used. When comparing the N_2-fixing data set against 3%-non-fixing and 21%-non-fixing, all genes that showed significantly ($p<0.05$) fewer reads under N_2-fixation are outlined in table 5.2. These data showed the reads for a number of *nif* genes were significantly lower in the N_2-fixing dataset compared to the non-fixing dataset; when comparing fixing growth versus ambient growth the number of *nif* genes showing a significant drop in reads increased. Importantly *nifA*, the regulator of N_2-fixation, showed a significant drop in reads during fixation when compared to both non-fixing and ambient datasets. The *suf* genes also showed a significant reduction in the number of reads in the N_2-fixation dataset, these genes are involved in the construction of Fe-S complexes that are critical to the function of nitrogenase. The *fix* genes were scored NE using the HMM analysis, and only *fixL* showed a significant reduction in reads using resampling. However, using IGV to visualise the read counts for each 'TA' site, the number of reads for each site in the *fixABCX* operon could be seen to be lower during N_2-fixation when compared comparison to non-fixing and ambient conditions.

Table 5.2 Resampling analyses of INSeq data, showing genes with significantly fewer reads ($p<0.05$) in the N_2-fixation dataset compared to datasets from i) non-fixing and ii) ambient conditions

Fixing vs Non-Fixing	
AZC_RS01315	Hypothetical
AZC_RS04970	Hypothetical
AZC_RS05320	Nitrogenase iron-molybdenum cofactor biosynthesis protein - NifE
AZC_RS05325	Nitrogenase molybdenum-iron protein subunit beta - NifK
AZC_RS05350	Fe-S cluster assembly protein - SufB
AZC_RS05355	ABC transporter ATP-binding protein - SufC
AZC_RS05360	Fe-S cluster assembly protein - SufD
AZC_RS05365	Cysteine desulfurase - SufS
AZC_RS05375	ATPase AAA - NifA
AZC_RS05850	superoxide dismutase
AZC_RS06925	Transcriptional regulator araC
AZC_RS07410	Branched chain amino acid ABC transporter permease
AZC_RS08140	NAD-dependent dehydratase
AZC_RS08240	pantoate-beta-alanine ligase
AZC_RS08585	mltA-interacting protein - MipA
AZC_RS10855	MFS transporter
AZC_RS11275	Hypothetical protein
AZC_RS11305	mureine hydrolysis effector protein
AZC_RS13890	dihydrodipicolinate synthetase
AZC_RS14860	3-hydroxyacyl-CoA dehydrogenase
AZC_RS15920	Nitrogen regulation protein NR(I) - NtrC
AZC_RS15925	PAS domain containing sensor histidine kinase - NtrB
AZC_RS15930	tRNA dihydrouridine synthase - DusB
AZC_RS16890	Galactarate dehydratase
AZC_RS17420	VWA domain containing
AZC_RS17550	Cysteine desulfurase - NifS
AZC_RS23990	putative glyoxalase protein

Fixing vs Ambient	
AZC_RS00525	S-(hydroxymethyl)glutathione dehydrogenase
AZC_RS01405	YggS family pyridoxal phosphate enzyme
AZC_RS01685	GTP-binding protein TypA
AZC_RS01720	[protein-PII] uridylyltransferase
AZC_RS02085	hypothetical protein
AZC_RS02200	NAD synthetase
AZC_RS03515	2-methylisocitrate lyase
AZC_RS03600	hypothetical protein
AZC_RS04600	hypothetical protein
AZC_RS04760	adenylyl-sulfate kinase
AZC_RS05315	Nitrogenase molybdenum-cofactor biosynthesis protein - NifN
AZC_RS05320	Nitrogenase iron-molybdenum cofactor biosynthesis protein - NifE
AZC_RS05325	Nitrogenase molybdenum-iron protein subunit beta - NifK
AZC_RS05330	Nitrogenase molybdenum-iron protein alpha chain - NifD
AZC_RS05355	ABC transporter ATP-binding protein - SufC

AZC_RS05365	Cysteine desulfurase - SufS
AZC_RS05375	ATPase AAA - NifA
AZC_RS05955	formate dehydrogenase subunit gamma
AZC_RS06000	formate dehydrogenase subunit beta
AZC_RS06925	AraC family transcriptional regulator
AZC_RS08240	pantoate--beta-alanine ligase
AZC_RS08280	2-keto-4-methylthiobutyrate aminotransferase
AZC_RS09635	ribonucleas R
AZC_RS10870	cytochrome o ubiquinol oxidase subunit III
AZC_RS11405	molybdopterin molybdenumtransferase MoeA
AZC_RS12405	hypothetical protein
AZC_RS13890	dihydrodipicolinate synthase family protein
AZC_RS15165	VWA domain-containing protein
AZC_RS15395	hypothetical protein
AZC_RS15920	Nitrogen regulation protein NR(I) - NtrC
AZC_RS15925	PAS domain containing sensor histidine kinase - NtrB
AZC_RS16060	hypothetical protein
AZC_RS16655	pilus biosynthesis protein
AZC_RS16890	galactarate dehydratase
AZC_RS16910	amino acid ABC transporter permease
AZC_RS17565	Nitrogenase cofactor biosynthesis protein - NifB
AZC_RS17610	aminotransferase DegT
AZC_RS19405	glycosyl transferase family 1
AZC_RS20500	hypothetical protein
AZC_RS20540	acyltransferase
AZC_RS21695	ammonia channel protein (AmtB)
AZC_RS21835	pseudouridine synthase
AZC_RS22025	hypothetical protein
AZC_RS22570	ABC transporter permease
AZC_RS23250	acyl-CoA synthetase
AZC_RS23750	glycosyl transferase
AZC_RS23840	PAS domain containing sensor histidine kinase - FixL
AZC_RS23990	putative glyoxalase protein
AZC_RS24565	hypothetical protein

The genes *ntrB* (AZC_RS15925) and *ntrC* (AZC_RS15920) were classified by the HMM as ES only during N_2-fixation. These genes form a signal transduction pathway involved in the expression of *nifA* during periods of low nitrogen availability (Pawlowski et al., 1991; Martinez-Argudo et al., 2005). In ORS571 these genes form part of an operon with a third gene, *dusB* (AZC_RS15930), also classified as ES during N_2-fixing growth.

A gene encoding for an ammonium transporter, *amtB* (AZC_RS21695), showed a significant drop in the number of reads in the N_2-fixing dataset. *amtB* forms an operon with a P-II homologue gene called *glnK* which is involved in the regulation of *ntrB* and thus expression of *nifA*. Resampling showed no significant difference in reads between *glnK* (AZC_RS21700) when comparing the N_2-fixing against non-fixing (plus ammonium at 3% O_2) and ambient (plus ammonium at 3% O_2); using IGV the number of reads can be seen to drop in the fixing dataset indicating *that glnK* is involved in N_2-fixation.

Genes involved in flagellar biosynthesis and motility showed an increase in essentiality under non-fixing and ambient growth conditions, but not during N_2-fixing conditions (Table 5.3). In fact, there is a marked increase in the number of reads for growth under N_2-fixing conditions, indicating that motility is dispensable during fixation. For example, in the case of AZC_RS03335 there was a 21-fold increase in the number of reads during fixation when compared to non-fixing growth. This is indicative of genes becoming GA.

Table 5.3 Flagellar biosynthesis and motility genes, which showed significantly higher reads ($p<0.05$) in the N_2-fixing dataset when compared to non-fixing and ambient data. Using a resampling analysis.

Fixing vs non-fixing		Fold Change
AZC_RS03205	flip (flagellar biosynthetic protein)	10
AZC_RS03235	FlgG (basal body rod protein)	5
AZC_RS03250	flgB	7
AZC_RS03260	FlhB (flagellar biosynthesis protein)	6
AZC_RS03280	FliM (Flagellar motor switch)	8
AZC_RS03285	MotA (chemotaxis protein)	6
AZC_RS03300	FlgE (flagellar hook protein)	7
AZC_RS03305	FlgK (flagellar hook protein)	5
AZC_RS03310	FlgL (flagellar hook associated protein)	5
AZC_RS03335	FlhH (flagellar biosynthesis protein)	21

Fixing vs ambient		Fold Change
AZC_RS03190	flagellar motor protein	5
AZC_RS03200	FliF (m-ring protein)	5
AZC_RS03205	flip (flagellar biosynthetic protein)	3
AZC_RS03235	FlgG (basal body rod protein)	5
AZC_RS03250	flgB	8
AZC_RS03260	FlhB (flagellar biosynthesis protein)	5
AZC_RS03280	FliM (Flagellar motor switch)	9
AZC_RS03285	MotA (chemotaxis protein)	4
AZC_RS03300	FlgE (flagellar hook protein)	6
AZC_RS03305	FlgK (flagellar hook protein)	3

This analysis has highlighted that INSeq can be used to study *A. caulinodans* during N_2-fixation and low nitrogen growth. This was achieved by combining a HMM analysis with resampling analyses to identify genes that showed increases in essentiality when grown under the conditions tested.

5.3 – Discussion

In a previous chapter I have shown the use of insertion sequencing in determining gene essentiality during the colonisation of the *O. sativa* rhizosphere and roots by A. caulinodans ORS571. Here I have highlighted that this technique is also capable of

being used to study free-living N_2-fixation by *A. caulinodans*. This work has also shown that INSeq can be used on cells grown on solid media, something that has not been shown before. As all previous studies have made use of liquid media to grow cells.

The first step of this study was to determine if the growth conditions were suitable for the INSeq analysis of N_2-fixation. To confirm this, the *nif* and *fix* genes were used as control genes. That is to say, if the *nif* and/or *fix* genes were not identified as GI or having significantly fewer reads during N_2-fixing growth compared to non-fixing microaerobic and non-fixing ambient growth then the experiment could be considered a failure. Whilst the HMM analysis did not score these genes as GI, resampling analysis confirmed that the *nif* and *suf* genes had significantly fewer reads under N_2-fixing conditions when compared against non-fixing and ambient growth. Other than *fixL*, the *fixABCX* operon was not identified by either HMM or resampling. Using IGV however, a drop in the number of reads under N_2-fixation is clear, indicative that the fix genes are important under these experimental conditions.

The reason many of the known genes involved in fixation were not clearly identified as causing growth impairment is due to how ORS571 grows under N_2-fixing conditions. The ability of ORS571 to fix nitrogen under these experimental conditions was tested using acetylene reductions (Supplementary 10) and comparing wild-type ORS571 against ORS571-Δ*nifA*. Whilst the ORS571-Δ*nifA* showed no acetylene reduction, some limited growth was seen over the first day of bacterial growth. This mirrored what was seen in growth curves utilising wildtype and ORS571-Δ*nifA* grown under N_2-fixing conditions, where a small amount of growth was seen for ORS571-ΔnifA before it plateaued (Fig 5.2). This growth was likely a result of stored nitrogen sources present in the bacterial cells alongside trace ammonium in the growth medium. Trace ammonium in the growth medium will be present due to the

breakdown of amino acids, nucleic acids, and as a small amount of contamination in the agarose, this will contribute to growth alongside any cellular nitrogen stores. These sources of nitrogen have allowed the bacterial cells to grow through a number of generations before N_2-fixation becomes critical for growth. Though the growth resulting from trace nitrogen was small (approximately 3 generations) it will have allowed for mutants in fixation related genes, such as the *nif* and *fix* genes, to grow. During INSeq this has led to increased numbers of reads in these genes, effectively shifting genes from ES/GD to NE when a HMM is used. Looking at the total number of reads for a gene and comparing across conditions using resampling has helped highlight differences in insertions between conditions to be detected. However, the *fixABCX* genes all showed a NE score and had no statistical difference in the number of reads during fixation compared to the non-fixing conditions. This is surprising as the *fix* genes are required for N_2-fixation across all rhizobial species. By using IGV to look at and compare reads across each condition, a drop in the number of reads for these genes can be seen. This indicates that *fixABCX* are necessary for N_2-fixation and it is growth due to stored nitrogen that pushed the read numbers over the thresholds for HMM and resampling analysis. Despite these setbacks, these data have confirmed that the experimental setup has allowed for genes involved in N_2-fixation to be identified. It is important to note that although at the time point the cells were harvested for study, N_2-fixation was starting to support growth of ORS571, the bulk of bacterial growth was likely to be the result of other trace nitrogen sources. Therefore the nitrogen fixing growth condition should in fact be labelled as low-nitrogen growth.

Interestingly, *ntrB* and *ntrC* gave a clear ES phenotype during N_2-fixation. In previous studies, *ntrC* has been shown to regulate *nifA* during times of low intracellular nitrogen in *A. caulinodans* and other rhizobial species (Ratet et al., 1989; Dixon and Kahn, 2004). NtrC has been shown to play an important role in the regulation of N_2-

fixation in *A. caulinodans*, with mutants showing an 85-90% reduction in N_2-fixation in free-living culture when compared to wildtype (Pawlowski et al., 1991). During symbiotic growth with *S. rostrate* ORS571 *ntrC* mutants have shown a small reduction in N_2-fixation in nodules, and during free-living growth the mutants ability to utilise L-arginine and L-histidine as a primary source was impaired. (Pawlowski et al., 1987). The phosphorylated form of NtrC is its active form; this phosphorylation is controlled by NtrB. The phosphorylation activity of NtrB is controlled by the uridylylated form of the P-II protein (GlnB or GlnK). GlnB has been shown to be involved in the activation of glutamine synthase (GlnA) (Michel-Reydellet et al., 1997) and is in fact essential across all conditions tested, including during rhizosphere colonisation and root attachment. This is unsurprising if GlnB is involved in the activation of GlnA as mutating either would prevent synthesis of glutamine and starve the cell. The second P-II protein gene (*glnK*) has been identified in ORS571 as AZC_RS21700 and forms an operon with the ammonium transporter gene *amtB*. Both showed a reduced number of reads only during fixation (*glnK* is observable using IGV, whilst *amtB* was identified during resampling). It is likely that during nitrogen limitation GlnK is uridylylated thus activating NtrB, NtrC, and ultimately increases the expression of *nifA*. This pathway for *nifA* regulation is similar to that seen in *K. pneumoniae* (Dixon and Kahn, 2004). However, this does not fully explain why *ntrB* and *ntrC* showed such a strong ES phenotype, when mutations in genes such as *nifA* did not. It is likely that *ntrBC* plays a much larger role in the overall nitrogen status of the cell during nitrogen limitation, and by mutating these genes the cell is crippled in its ability to access stored nitrogen and trace environmental nitrogen. In essence, NtrBC are the key controllers for overall nitrogen regulation during low nitrogen growth. This is further confirmed by *amtB* showing a significant drop in reads during low nitrogen growth. AmtB is key in the transport of environmental ammonium into the cells and its transcription is activated by NtrC. Therefore under N_2-fixing conditions, it can be posited that low levels of

environmental ammonium (present in the agarose, or from the breakdown of nucleic and amino acids) were supporting the growth of ORS571 and that NtrBC are key in the regulation of transporters and nitrogen use within the cells. This highlighted that the rate of N_2-fixation in these conditions was not great enough to support the growth of the cells on its own, and at the time point taken for this analysis the cells maintained growth from intracellular and environmental nitrogen sources.

Forming an operon with *ntrBC* is a third gene called *dusB* (tRNA-dihydrouridine synthase) that is translationally linked to *ntrB*. This gene was identified as GI only during N_2-fixing growth (Table 5.1), and showed significantly fewer reads when compared to non-fixing growth (Table 5.2). The role DusB plays is unknown, the protein itself reduces C5-C6 double bonds of uridines on tRNA, with preference to the D-strand (Dalluge et al., 1997). The role dihydrouridines play in tRNA is unknown, but they are possibly involved in increasing conformational flexibility (Dalluge et al., 1996). Regardless of the role of dihydrouridine, it appears likely that DusB could be involved in the posttranscriptional modification of *ntrB* and/or *ntrC* mRNA during nitrogen limitation and possibly during N_2-fixation. This gene has not been identified in the past, nor has it been insinuated to play a role during fixation, making it an interesting target for further study.

Here I have shown the suitability of INSeq to identify novel genes involved during low-nitrogen growth rather than my original aim of during N_2-fixation, of *A. caulinodans* ORS571. The next step in this study would be to experimentally confirm the identified genes e.g. *dusB* for their role during low nitrogen growth. The first step would be to produce mutants of these genes and perform growth studies under fixing and non-fixing conditions; this would confirm that the conclusions about fitness drawn from INSeq are correct. Once they have been confirmed, more in-depth metabolic studies could be performed to further elucidate their roles and metabolic pathways.

N_2-fixing growth can be optimised by growing cultures for a longer period of time under fixing conditions, thus increasing the number of bacterial generations dependent on fixing nitrogen. This would allow for a clearer identification of essential and growth defective genes during nitrogen fixing growth.

Chapter 6

Discussion and future perspectives

6.1 - Introduction

The release of photosynthetically fixed carbon from plant roots is a crucial and far-reaching process. These exudates play critical roles in numerous interactions, not least, the colonisation of roots by microorganisms. The determination of the spatial and temporal release of key exudates is a key step in understanding how plants are able to shape the rhizosphere through root exudation. This colonisation of roots by bacterial species is critical in the life cycles of plants and the mechanisms by which these bacteria colonise roots is key to understanding species-specific interactions e.g. *A. caulinodans* and *O. sativa*; and more globally conserved mechanisms e.g. rhizobia and legumes. In this project, I have aimed to highlight new techniques for the spatiotemporal mapping of root exudation and the genome-wide study of colonisation.

6.2 – Exudate mapping

The ability to spatially map the release of root exudates has been limited by both the lack of specific knowledge and technology. Large-scale spatial information has been gleaned by the use of rhizotrons and micro-suction techniques to collect exudates. The resolution of these methods is limited to the macro scale of exudation e.g. mature root surface and the growing surfaces. The other major limitation is that the micro-suction devices cannot be moved around the system allowing only a single snapshot of a growing tip as the root passes the collection point. This can limit any temporal analysis to the whole root system rather than allowing for spatial and temporal mapping of a specific root surface. Methods such as HPLC and GCMS on hydroponically collected exudates have allowed for the large-scale temporal mapping of exudation along with the composition of the exudate profile, but lack the ability of spatial mapping. Here I have highlighted the use of a luciferase-based system for the

spatiotemporal mapping of exudation in a variety of land plant species. Luciferase based bioreporters developed in *R. leguminosarum* (Ramachandran *et al.*, 2011; Pini *et al.*, 2017) have allowed for the mapping of fourteen key metabolites from the roots of *O. sativa*, *Z. mays*, *A. thaliana* and *M. polymorpha*, showing that *R. leguminosarum*-based reporters are applicable to the study of exudation outside of leguminous species.

6.2.1- Spatiotemporal nature of Exudation

In *Z. mays* the use of a constitutive reporter showed exudation occurring across the whole rooting system. The bulk of this exudation/luminescence was detected on the lateral root surfaces with increasing luminescence moving from the lateral root tip to the junction with the primary root. The signals from the primary root surfaces were lower than those seen on the lateral root surfaces indicating that there was more exudation occurring from the lateral root surfaces compared to the primary roots, and/or there was preferential colonisation occurring on the lateral roots leading to a higher bacterial load and thus signal. The C4-dicarboxylate fusion revealed a similar pattern of exudation (more signal from lateral surfaces than primary). The major difference seen was that the majority of luminescence (thus exudation) was seen towards the lateral root tips, with the signal dropping towards the junctions with primary root. In the past it has been assumed that the growing root tips, and elongation zones were responsible for the bulk of exudates released into the rhizosphere (Dennis *et al.*, 2010). Here I have shown experimentally that exudation is occurring across the whole rooting system, with specific localisation for some key exudates. In *O. sativa* the exudation of phenylalanine and C4-dicarboxylates was centred along the crown roots and seeds, with low exudation occurring down the length of the primary roots. The constitutive bioreporter highlighted again that

exudation occurred across the whole rooting system, and that there is spatial localisation of the release of key specific metabolites. In contrast to these stark spatial effects, in *A. thaliana* the exudation of C4-dicarboxylates and fructose was seen across the whole rooting system and not localised to any specific areas. A temporal affect was also seen with C4-dicarboxylate exudation, which started at low levels and increased over time.

6.2.2- C4-dicarboxylate exudation is conserved across plant species

An interesting facet of the exudate mapping applied in this thesis is the conservation of C4-dicarboxylate exudation across plants. The bryophytes form the most basal diverging land plant lineage, their rooting system is made up of single cellular rhizoids that are somewhat analogous to root hairs. However, there is little information surrounding the release of photosynthetically carbon from bryophytes such as *M. polymorpha*, yet their importance in studying land plant evolution is well known. Here I have shown the release of C4-dicarboxylates from the thallus and rhizoids of *M. polymorpha*, indicating that the mechanism for the exudation of these compounds occurred early in the colonisation of land. It would be naïve to assume that during the early Devonian 450 Ma ago when the colonisation of land by plants began (Algeo *et al.*, 1998), that the earths biosphere was perfectly suited for their growth. As plants spread across this landscape it seems likely that the mechanisms underpinning the scavenging of nutrients, detoxification of metals, and the formation of primitive symbioses with microorganisms, would have developed. This work has highlighted that the earliest diverging land plant lineage (bryophytes) released C4-dicarboxylates into their environments. These LMWOAs are associated with nutrient solubilisation, and metal chelation (Jones *et al.*, 2009) in the vascular plants, and as such it is likely that these compounds are playing a similar role in *M. polymorpha*.

6.2.3 - Exudates conclusion

Here I have shown bioreporters expressed in *R. leguminosarum* map exudation across a range of non-leguminous plants including the two most important cereal crops (*O. sativa* and *Z. mays*). This work suggests that exudation occurs across the entire rooting system and is not confined to specific regions. Furthermore, it is likely that the localisation of specific exudates to lateral and crown roots enables the plant to increase the surface area of the chemical soil-root interface. While the use of the current set of bioreporters in *R. leguminosarum* has been very successful there is a need to extend the range of compounds sensed. Furthermore, it would be useful to extend the species of bacteria used, particularly to include species that are excellent at colonising cereals and can exist as endophytes. Using high throughput genomic techniques such as INSeq it is possible to identify new transport and metabolic systems in bacteria that can be engineered to produce an ever-increasing number of bioreporters for exudate mapping and the *in-vivo* mapping of root colonisation.

6.3 - Insertion Sequencing and RNA-Seq

With the advent of next-generation sequencing technologies a range of new opportunities have arrived to understand the genetics underpinning biological processes. In this project insertion sequencing (INSeq) has been applied to study gene essential for the colonisation of the rhizosphere and roots of *O. sativa* by nitrogen-fixing *Azorhizobium caulinodans*. Further to the study of colonisation, I have also shown the applicability of INSeq for the study of free-living N_2-fixation. The mechanisms by which bacteria colonise an environment are incredibly complex. Identifying genes that play critical roles during colonisation is the first step in identifying the signalling and metabolic pathways underpinning it. Using INSeq I have identified a number of gene candidates in *A. caulinodans* that are essential for

bacterial growth in rhizosphere and during the colonisation of root surfaces. Furthermore, the application of RNA-Seq for the study of colonisation has been probed.

6.3.1 - Propionate metabolism is required for colonisation

The conversion of propionyl-CoA to succinate was identified as an essential pathway during *A. caulinodans* colonisation of the *O. sativa* rhizosphere. I showed that genes involved in the conversion of propionyl-CoA to succinate (via methylmalonyl-CoA as an intermediate) are essential for rhizosphere and root attached growth. This would indicate that *A. caulinodans* is utilizing propionate as carbon source. It has been shown that *A. caulinodans* is capable of growing on propionate as a sole carbon source, albeit poorly (Dreyfus *et al.*, 1988), and there is some evidence that propionate is present in *O. sativa* root exudates (Breidenbach *et al.*, 2016). Whilst the *prpE* gene encoding for propionate-CoA synthetase was labelled as growth neutral, there are two separate acyl-CoA synthetases present in *A. caulinodans.* As acyl-CoA synthetases can interact with a variety of short-chain lengths they may be acting as a redundant system when *prpE* is mutated. The activation of these genes upon the detection and metabolism of propionate could be used to form the basis of a new bioreporter based in *A. caulinodans*. By creating a promoter fusion with the *lux* genes, a luciferase-based reporter could be developed. Further confirmation of these genes role in propionate metabolism would be their mutation and subsequent growth of mutant strains on propionate.

6.3.2 - Oxalate metabolism/degradation is required for rhizosphere colonisation

The conversion of oxalate to formate was also identified as a key process during rhizosphere colonisation and root attachment. Oxalate has been identified as one of the key organic acids making up *O. sativa* root exudates (Hoffland *et al.*, 2006). It has been shown that *A. caulinodans* ORS571 cannot use oxalate as a sole carbon source (Dreyfus *et al.*, 1988) and any toxicity it poses to *A. caulinodans* is unknown. Regardless of the exact role oxalate is playing in bacterial metabolism, it is nonetheless important for rhizosphere growth. To confirm the role of the genes identified during colonisation, mutants should be created and grown in the *O. sativa* rhizosphere. Competition studies between mutant and wild-type strains would further confirm that oxalate is required for successful colonisation. Furthermore, these genes could be used to engineer an oxalate bioreporter, which could be used to probe oxalate exudation into the rhizosphere.

6.3.3 - INSeq can be used to study nitrogen fixation

INSeq has been successfully used to study gene essentiality in bacteria across a range of growth conditions and species. All of these studies have made use of liquid cultures, or the isolation of bacteria from an aqueous environment. In this thesis the study of N_2-fixation was performed on a solid medium and appears to be the first insertion sequencing study (including Tn-seq, HITS, and TraDIS) to do so. This is important as it broadens the range of conditions/environments that these techniques could be used to study. Whilst this study highlighted new genes that are involved in N_2-fixation, it has not been without its setbacks. Most notably the small amount of growth resulting from stored nitrogen within the cells resulted in the *nif* and *fix* genes being scored as neutral for growth during fixation. This was mitigated by using resampling and careful analysis of each gene using integrative genome viewer.

Though these results were not perfect, they proved that INSeq could be used for the study of N_2-fixation. The method of growth could be further optimised to reduce or remove stored nitrogen by sub culturing; this would however be most effective in liquid cultures. One method to achieve this would be the growth of cultures on solid medium (as outlined in Chapter 5) to initiate N_2-fixation, these cultures could then be resuspended and grown in liquid minimal medium under 3% oxygen for a minimum of ten generations. This method would allow for a larger separation between low nitrogen growth and nitrogen fixing growth, ultimately giving a much more robust method to study N_2-fixation

6.3.4 - *ntrBC* and *dusB* are essential for low nitrogen growth

The *ntrBC* /*dusB* operon was highlighted to be essential for growth under limiting nitrogen conditions by HMM analysis, something that was not seen for the *nif* or *fix* genes. As stated in Chapter 5, it has been determined that the growth condition for this experimental setup was primarily low-nitrogen growth with only a small influence from N_2-fixation. This is likely due to the fact that ORS571 was capable of growing for a small number of generations on cellular nitrogen stores and trace nitrogen in the growth medium. Once this nitrogen was abolished, only a small number of generations were grown under true N_2-fixing conditions, giving the results seen in this study. This was confirmed by examining *ntrBC* and *amtB* that are tied heavily with low nitrogen growth. AmtB transports NH_4^+ into the cell during times of low cellular nitrogen and is activated by *ntrC*. By mutating *ntrC, ntrB*, or *amtB* the transport of ammonium into the cell is abolished, and any growth from trace ammonium in the medium is stopped. Furthermore *ntrC* mutants in ORS571 have been shown to be reduced in their ability to utilise histidine and arginine as nitrogen sources (Pawlowski *et al.*, 1987). This severely limits the ability for the cells to utilise nitrogen stored in

these amino-acids and ultimately would halt cellular growth. Future directions would be to use INSeq to study these genes under 21% oxygen low nitrogen growth, as well as under more robust N_2-fixing conditions. The role of *dusB* is unknown but based on its annotation it is likely involved in the post-transcriptional regulation of *ntrBC*. To confirm this, mutants of *dusB* must be constructed and their effect on the expression, translation and stability of *ntrBC* studied.

6.3.5 – RNA-Seq

The use of transcriptomics to probe gene expression has increased rapidly over the past decade due to the advancement and falling prices of sequencing technologies. In this project the use of RNA-Seq to study *A. caulinodans* gene expression during the colonisation of the rhizosphere and roots of *O. sativa* was probed. RNA-Seq produces data which is complimentary to that produced my INSeq, and one of the aims of this project was to use these two datasets in concert to understand the genetics behind rhizosphere and root colonisation in greater detail. Whilst this work showed that RNA-Seq could be performed on bacteria isolated from the rhizosphere and roots of plants, the extremely low read numbers made analysis of this data almost impossible. As such only a few genes were identified as upregulated during colonisation and not much could be guessed about their function. To overcome this problem, the number of plant replicates grown must be increased by two or three times. This would allow for the isolation of more bacteria and thus RNA, in turn this would hopefully give a more robust and usable RNA-Seq dataset.

6.3.6 - Conclusions

This project has shown the successful application of INSeq to identify the fitness determinants of *Azorhizobium caulinodans* during root colonisation and low nitrogen growth. The pSAM_Rl plasmid designed for use in *R. leguminosarum* (Perry and Yost, 2014) was successfully used for production of insertion mutant libraries in *A. caulinodans*. These libraries were successfully used to study the aforementioned conditions highlighting hitherto unidentified metabolic and regulatory networks required for growth. The next step will be to produce mutants of identified genes to confirm their phenotypic effects on rhizosphere growth and N_2-fixation. Furthermore genes such as those involved in propionate and oxalate metabolism/degradation can be used to produce new biosensors capable of detecting these compounds on plant roots *in-vivo*.

Bibliography

Algeo, T. J., Scheckler, S. E. and Scott, A. C. (1998) 'Terrestrial-Marine Teleconnections in the Devonian: Links between the Evolution of Land Plants, Weathering Processes, and Marine Anoxic Events [and Discussion]', *Philosophical Transactions: Biological Sciences*, pp. 113–130.

Aoki, T., Akashi, T. and Ayabe, S. (2000) 'Flavonoids of Leguminous Plants: Structure, Biological Activity, and Biosynthesis', *Journal of Plant Research*, 113(4), pp. 475–488.

Aulakh, M. S., Wassmann, R., Bueno, C., Kreuzwieser, J. and Rennenberg, H. (2001) 'Characterization of root exudates at different growth stages of ten rice (Oryza sativa L.) cultivars', *Plant Biology*, 3(2), pp. 139–148.

Bacilio-Jiménez, M., Aguilar-Flores, S., Ventura-Zapata, E., Pérez-Campos, E., Bouquelet, S. and Zenteno, E. (2003) 'Chemical characterization of root exudates from rice (Oryza sativa) and their effects on the chemotactic response of endophytic bacteria', *Plant and Soil*, 249(2), pp. 271–277.

Badri., Quintana, N., El Kassis, E. G., Kim, H. K., Choi, Y. H., Sugiyama, A., Verpoorte, R., Martinoia, E., Manter, D. K. and Vivanco, J. M. (2009) 'An ABC Transporter Mutation Alters Root Exudation of Phytochemicals That Provoke an Overhaul of Natural Soil Microbiota', *Plant Physiology*, 151(4), pp. 2006–2017.

Badri, D. V. and Vivanco, J. M. (2009) 'Regulation and function of root exudates', *Plant, Cell and Environment*, 32(6), pp. 666–681.

Badri, D. V., Weir, T. L., van der Lelie, D. and Vivanco, J. M. (2009) 'Rhizosphere chemical dialogues: plant-microbe interactions', *Current Opinion in Biotechnology*, 20(6), pp. 642–650.

Baetz, U. and Martinoia, E. (2014) 'Root exudates: The hidden part of plant defense', *Trends in Plant Science*, 19(2), pp. 90–98.

Bais, H. P., Weir, T. L., Perry, L. G., Gilroy, S. and Vivanco, J. M. (2006) 'The Role of Root Exudates in Rhizosphere Interactions With Plants and Other Organisms', *Annual Review of Plant Biology*, 57(1), pp. 233–266.

Bakker, P. A. H. M., Pieterse, C. M. J. and Loon, L. C. van (2007) 'Induced Systemic Resistance by Fluorescent Pseudomonas spp.', *The American Phytopathological Society*, 97(2), pp. 239–243.

Bannerjee, D., Sanders, L. E., John, R. and Chem, J. B. (1970) 'Properties of Purified Methylmalonate Semialdehyde Dehydrogenase of Pseudomonas aeruginosa', *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, 245(7), pp. 1828–1835. Available at: http://www.jbc.org/content/245/7/1828.full.pdf (Accessed: 20 April 2018).

Barquist, L., Boinett, C. J. and Cain, A. K. (2013) 'Approaches to querying bacterial genomes with transposon-insertion sequencing', *RNA Biology*, 10(7), pp. 1161–1169.

Barr, M., East, A. K., Leonard, M., Mauchline, T. H. and Poole, P. S. (2008) 'In vivo expression technology (IVET) selection of genes of Rhizobium leguminosarum biovar viciae A34 expressed in the rhizosphere', *FEMS Microbiology Letters*, 282(2), pp. 219–227.

Bateman, R. M., Crane, P. R., DiMichele, W. A., Kenrick, P. R., Rowe, N. P., Speck, T. and Stein, W. E. (1998) 'EARLY EVOLUTION OF LAND PLANTS: Phylogeny, Physiology, and Ecology of the Primary Terrestrial Radiation', *Annual Review of Ecology and Systematics*, 29(1), pp. 263–292.

Batut, J., Andersson, S. G. E. and O'Callaghan, D. (2004) 'The evolution of chronic infection strategies in the alpha-proteobacteria', *Nature Reviews Microbiology*, 2(12), pp. 933–945.

Berendsen, R. L., Pieterse, C. M. J. and Bakker, P. A. H. M. (2012) 'The rhizosphere

microbiome and plant health', *Trends in Plant Science*, 17(8), pp. 478–486.

Bertin, C., Yang, X. and Weston, L. A. (2003) 'The role of root exudates and allelochemicals in the rhizosphere', *Plant and Soil*, 256(1), pp. 67–83.

Besserer, A., Puech-Pagès, V., Kiefer, P., Gomez-Roldan, V., Jauneau, A., Roy, S., Portais, J. C., Roux, C., Bécard, G. and Séjalon-Delmas, N. (2006) 'Strigolactones stimulate arbuscular mycorrhizal fungi by activating mitochondria', *PLoS Biology*, 4(7), pp. 1239–1247.

Bolan, N. S., Naidu, R., Mahimairaja, S. and Baskaran, S. (1994) 'Influence of low-molecular-weight organic acids on the solubilization of phosphates', *Biology and Fertility of Soils*, 18(4), pp. 311–319.

Bonfante, P. and Genre, A. (2008) 'Plants and arbuscular mycorrhizal fungi: an evolutionary-developmental perspective', *Trends in Plant Science*, 13(9), pp. 492–498.

Breidenbach, B., Pump, J. and Dumont, M. G. (2016) 'Microbial community structure in the rhizosphere of rice plants', *Frontiers in Microbiology*, 6(JAN), p. 1537.

Bringhurst, R. M., Cardon, Z. G. and Gage, D. J. (2001) 'Galactosides in the rhizosphere: Utilization by Sinorhizobium meliloti and development of a biosensor', *Proc Natl Acad Sci U S A*, 98(8), pp. 4540–4545.

Brundrett, M. C. (2002) 'Coevolution of roots and mycorrhizas of land plants', *New Phytologist*, 154(2), pp. 275–304.

Bulgarelli, D., Schlaeppi, K., Spaepen, S., van Themaat, E. V. L. and Schulze-Lefert, P. (2013) 'Structure and Functions of the Bacterial Microbiota of Plants', *Annual Review of Plant Biology*, 64(1), pp. 807–838.

Cai, T., Cai, W., Zhang, J., Zheng, H., Tsou, A. M., Xiao, L., Zhong, Z. and Zhu, J. (2009) 'Host legume-exuded antimetabolites optimize the symbiotic rhizosphere', *Molecular Microbiology*, 73(3), pp. 507–517.

Cannon, W. A. (1949) ‘A tentative classification of root systems.’, *Ecology*, 30(4), pp. 542–548.

Carvalhais, L. C., Dennis, P. G., Fedoseyenko, D., Hajirezaei, M. R., Borriss, R. and Von Wir??n, N. (2011) ‘Root exudation of sugars, amino acids, and organic acids by maize as affected by nitrogen, phosphorus, potassium, and iron deficiency’, *Journal of Plant Nutrition and Soil Science*, 174(1), pp. 3–11.

Castrillo, G., Teixeira, P. J. P. L., Paredes, S. H., Law, T. F., de Lorenzo, L., Feltcher, M. E., Finkel, O. M., Breakfield, N. W., Mieczkowski, P., Jones, C. D., Paz-Ares, J. and Dangl, J. L. (2017) ‘Root microbiota drive direct integration of phosphate stress and immunity’, *Nature*, 543(7646), pp. 513–518.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. and Prasher, D. (1994) ‘Green fluorescent protein as a marker for gene expression’, *Science*, 263(5148), pp. 802–805.

Chaparro, J. M., Badri, D. V., Bakker, M. G., Sugiyama, A., Manter, D. K. and Vivanco, J. M. (2013) ‘Root Exudation of Phytochemicals in Arabidopsis Follows Specific Patterns That Are Developmentally Programmed and Correlate with Soil Microbial Functions’, *PLoS ONE*. Edited by K. Wu, 8(2), p. e55731.

Chen, Y. T., Wang, Y. and Yeh, K. C. (2017) ‘Role of root exudates in metal acquisition and tolerance’, *Current Opinion in Plant Biology*, 39, pp. 66–72.

Chi, F., Shen, S. H., Cheng, H. P., Jing, Y. X., Yanni, Y. G. and Dazzo, F. B. (2005) ‘Ascending migration of endophytic rhizobia, from roots to leaves, inside rice plants and assessment of benefits to rice growth physiology’, *Applied and Environmental Microbiology*, 71(11), pp. 7271–7278.

Christiansen-Weniger, C. (1996) ‘Endophytic establishment of Azorhizobium caulinodans through auxin- induced root tumors of rice (Oryza sativa L.)’, *Biology and Fertility of Soils*, 21(4), pp. 293–302.

Clemens, S. (2001) ‘Molecular mechanisms of plant metal tolerance and homeostasis’, *Planta*, 212(4), pp. 475–486.

Clemens, S. and Weber, M. (2016) ‘The essential role of coumarin secretion for Fe acquisition from alkaline soil’, *Plant Signaling & Behavior*, 11(2), p. e1114197.

Close, D. M., Hahn, R. E., Patterson, S. S., Baek, S. J., Ripp, S. A. and Sayler, G. S. (2011) ‘Comparison of human optimized bacterial luciferase, firefly luciferase, and green fluorescent protein for continuous imaging of cell culture and animal models.’, *Journal of biomedical optics*, 16(4), p. 47003.

Close, D., Xu, T., Smartt, A., Rogers, A., Crossley, R., Price, S., Ripp, S. and Sayler, G. (2012) ‘The evolution of the bacterial luciferase gene cassette (lux) as a real-time bioreporter’, *Sensors*, 12(1), pp. 732–752.

Compant, S., Clement, C. and Sessitsch, A. (2010) ‘Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization’, *Soil Biology and Biochemistry*, 42(5), pp. 669–678.

Croucher, N. J. and Thomson, N. R. (2010) ‘Studying bacterial transcriptomes using RNA-seq’, *Current Opinion in Microbiology*, 13(5), pp. 619–624.

Dakora, F. D. and Phillips, D. A. (2002) ‘Root exudates as mediators of mineral acquisition in low-nutrient environments’, *Plant and Soil*, 245(1), pp. 35–47.

Darwent, M. J., Paterson, E., McDonald, A. J. S. and Tomos, A. D. (2003) ‘Biosensor reporting of root exudation from Hordeum vulgare in relation to shoot nitrate concentration’, *Journal of Experimental Botany*, 54(381), pp. 325–334.

DeAngelis, K. M., Firestone, M. K. and Lindow, S. E. (2007) ‘Sensitive whole-cell biosensor suitable for detecting a variety of N-acyl homoserine lactones in intact rhizosphere microbial communities’, *Applied and Environmental Microbiology*, 73(11), pp. 3724–3727.

DeAngelis, K. M., Ji, P., Firestone, M. K. and Lindow, S. E. (2005) ‘Two novel

bacterial biosensors for detection of nitrate availability in the rhizosphere', *Applied and Environmental Microbiology*, 71(12), pp. 8537–8547.

DeJesus, M. A., Ambadipudi, C., Baker, R., Sassetti, C. and Ioerger, T. R. (2015) 'TRANSIT - A Software Tool for Himar1 TnSeq Analysis', *PLoS Computational Biology*, 11(10).

Dennis, P. G., Miller, A. J. and Hirsch, P. R. (2010) 'Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities?', *FEMS Microbiology Ecology*, 72(3), pp. 313–327.

Dessureault-Rompré, J., Nowack, B., Schulin, R. and Luster, J. (2006) 'Modified micro suction cup/rhizobox approach for the in-situ detection of organic acids in rhizosphere soil solution', *Plant and Soil*, 286(1–2), pp. 99–107.

Van Dijk, E. L., Auger, H., Jaszczyszyn, Y. and Thermes, C. (2014) 'Ten years of next-generation sequencing technology', *Trends in Genetics*, 30(9), pp. 418–426.

Doi, N. and Yanagawa, H. (1999) 'Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution', *FEBS Letters*, 453(3), pp. 305–307.

Dreyfus, B., Garcia, J. L. and Gillis, M. (1988) 'Characterization of Azorhizobium caulinodans gen. nov., sp. nov., a Stem-Nodulating Nitrogen-Fixing Bacterium Isolated from Sesbania rostrata', *International Journal of Systematic Bacteriology*, 38(1), pp. 89–98.

Dunn, M. F., Araíza, G. and Mora, J. (2004) 'Biochemical characterization of a Rhizobium etli monovalent cation-stimulated acyl-coenzyme A carboxylase with a high substrate specificity constant for propionyl-coenzyme A', *Microbiology*, 150(2), pp. 399–406.

El-Shatnawi, M. K. J. and Makhadmeh, I. M. (2001) 'Ecophysiology of the plant-rhizosphere system', *Journal of Agronomy and Crop Science*, 187(1), pp. 1–9.

Emmert, E. A. B., Milner, J. L., Lee, J. C., Pulvermacher, K. L., Olivares, H. A., Clardy, J. and Handelsman, J. (1998) 'Effect of canavanine from alfalfa seeds on the population biology of Bacillus cereus', *Applied and Environmental Microbiology*, 64(12), pp. 4683–4688. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9835549 (Accessed: 4 August 2017).

Erickson, J., Schott, D., Reverri, T., Muhsin, W. and Ruttledge, T. (2001) 'GC-MS analysis of hydrophobic root exudates of sorghum and implications on the parasitic plant Striga asiatica', *Journal of Agricultural and Food Chemistry*, 49(11), pp. 5537–5542.

Figurski, D. H. and Helinski, D. R. (1979) 'Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans.', *Proceedings of the National Academy of Sciences of the United States of America*, 76(4), pp. 1648–52.

Fitter, A. H. (1987) 'An Architectural Approach To the Comparative Ecology of Plant Root Systems', *New Phytologist*, 106(s1), pp. 61–77.

Fourcroy, P., Sisó-Terraza, P., Sudre, D., Savirón, M., Reyt, G., Gaymard, F., Abadía, A., Abadia, J., Álvarez-Fernández, A. and Briat, J. F. (2014) 'Involvement of the ABCG37 transporter in secretion of scopoletin and derivatives by Arabidopsis roots in response to iron deficiency', *New Phytologist*, 201(1), pp. 155–167.

Fox, P. M. and Borthakur, D. (2001) 'Selection of several classes of mimosine-degradation-defective Tn*3*Ho*gus*-insertion mutants of *Rhizobium* sp. strain TAL1145 on the basis of mimosine-inducible GUS activity', *Canadian Journal of Microbiology*, 47(6), pp. 488–494.

Gage, D. J., Herron, P. M., Arango Pinedo, C. and Cardon, Z. G. (2008) 'Live reports from the soil grain - The promise and challenge of microbiosensors', *Functional Ecology*, 22(6), pp. 983–989.

Galloway, J. N., Townsend, A. R., Erisman, J. W., Bekunda, M., Cai, Z., Freney, J. R., Martinelli, L. a, Seitzinger, S. P. and Sutton, M. a (2008) 'Transformation of the Nitrogen Cycle: Recent Trends, Questions, and Potential Solutions', *Science*, 320(May), pp. 889–892.

Gawronski, J. D., Wong, S. M. S., Giannoukos, G., Ward, D. V. and Akerley, B. J. (2009) 'Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung', *Proceedings of the National Academy of Sciences*, 106(38), pp. 16422–16427.

Ge, J., Wood, D. K., Weingeist, D. M., Prasongtanakij, S., Navasumrit, P., Ruchirawat, M. and Engelward, B. P. (2013) 'Standard fluorescent imaging of live cells is highly genotoxic', *Cytometry Part A*, 83 A(6), pp. 552–560.

Geddes, B. A. and Oresnik, I. J. (2012) 'Inability to catabolize galactose leads to increased ability to compete for nodule occupancy in Sinorhizobium meliloti', *Journal of Bacteriology*, 194(18), pp. 5044–5053.

Gherardi, M. J. and Rengel, Z. (2004) 'The effect of manganese supply on exudation of carboxylates by roots of lucerne (Medicago sativa)', *Plant and Soil*, 260(1–2), pp. 271–282.

Giddens, S. R., Giddens, S. R., Jackson, R. W., Jackson, R. W., Moon, C. D., Moon, C. D., Jacobs, M. a, Jacobs, M. a, Zhang, X.-X., Zhang, X.-X., Gehrig, S. M., Gehrig, S. M., Rainey, P. B. and Rainey, P. B. (2007) 'Mutational activation of niche-specific genes provides insight into regulatory networks and bacterial function in a complex environment.', *Proceedings of the National Academy of Sciences of the United States of America*, 104(46), pp. 18247–52.

Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., Knight, R. and Gordon, J. I. (2009) 'Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat', *Cell Host and Microbe*, 6(3), pp.

279–289.

Goodman, A. L., Wu, M. and Gordon, J. I. (2011) ‘Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries’, *Nature Protocols*, 6(12), pp. 1969–1980.

Grayston, S. J., Vaughan, D. and Jones, D. (1997) ‘Rhizosphere carbon flow in trees, in comparison with annual plants: the importance of root exudation and its impact on microbial activity and nutrient availability’, *Applied Soil Ecology*, 5(1), pp. 29–56.

Groh, J. L., Luo, Q., Ballard, J. D. and Krumholz, L. R. (2005) ‘A method adapting microarray technology for signature-tagged mutagenesis of Desulfovibrio desulfuricans G20 and Shewanella oneidensis MR-1 in anaerobic sediment survival experiments’, *Applied and Environmental Microbiology*, 71(11), pp. 7064–7074.

Gruber, N. and Galloway, J. (2008) ‘An Earth-system perspective of the global nitrogen cycle’, *Nature*, 451(7176), pp. 293–296.

Haichar, F. el Z., Santaella, C., Heulin, T. and Achouak, W. (2014) ‘Root exudates mediated interactions belowground’, *Soil Biology and Biochemistry*, 77, pp. 69–80.

Hakkila, K., Maksimow, M., Karp, M. and Virta, M. (2002) ‘Reporter genes lucFF, luxCDABE, gfp, and dsred have different characteristics in whole-cell bacterial sensors.’, *Analytical biochemistry*, 301(2), pp. 235–242.

Halarnkar, P. P. and Blomquist, G. J. (1989) ‘Comparative aspects of propionate metabolism’, *Comparative Biochemistry and Physiology -- Part B: Biochemistry and*, 92(2), pp. 227–231.

Handoh, I. C. and Lenton, T. M. (2003) ‘Periodic mid-Cretaceous oceanic anoxic events linked by oscillations of the phosphorus and oxygen biogeochemical cycles’, *Global Biogeochemical Cycles*, 17(4), pp. 1–11.

Hansen, L. H. and Sørensen, S. J. (2001) ‘The use of whole-cell biosensors to detect and quantify compounds or conditions affecting biological systems’, *Microbial*

Ecology, 42(4), pp. 483–494.

Hell, R. and Stephan, U. W. (2003) 'Iron uptake, trafficking and homeostasis in plants.', *Planta*, 216(4), pp. 541–551.

Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E. and Holden, D. W. (1995) 'Simultaneous identification of bacterial virulence genes by negative selection.', *Science (New York, N.Y.)*, 269(5222), pp. 400–3.

Herron, P. M., Gage, D. J., Arango Pinedo, C., Haider, Z. K. and Cardon, Z. G. (2013) 'Better to light a candle than curse the darkness: illuminating spatial localization and temporal dynamics of rapid microbial growth in the rhizosphere', *Frontiers in Plant Science*, 4(323), pp. 1–14.

Hinsinger, P. (2001) 'Bioavailability of soil inorganic P in the rhizosphere as affected by root-induced chemical changes: A review', *Plant and Soil*, 237(2), pp. 173–195.

Hirsch, A. M., Bauer, W. D., Bird, D. M., Cullimore, J., Tyler, B., Yoder, J. I., Dietz Bauer, W. and Yoder6, J. I. (2003) 'Molecular Signals and Receptors: Controlling Rhizosphere Interactions Between Plants and Other Organisms', *Source: Ecology Ecology*, 84(844), pp. 858–868. Available at: http://www.jstor.org/stable/3108029%0Ahttp://www.jstor.org/stable/3108029?seq=1&cid=pdf-reference#references_tab_contents%0Ahttp://about.jstor.org/terms (Accessed: 30 July 2017).

Hochholdinger, F., Park, W. J., Sauer, M. and Woll, K. (2004) 'From weeds to crops: Genetic analysis of root development in cereals', *Trends in Plant Science*, 9(1), pp. 42–48.

Hoekenga, O. A., Maron, L. G., Piñeros, M. A., Cançado, G. M. A., Shaff, J., Kobayashi, Y., Ryan, P. R., Dong, B., Delhaize, E., Sasaki, T., Matsumoto, H., Yamamoto, Y., Koyama, H. and Kochian, L. V (2006) 'AtALMT1, which encodes a malate transporter, is identified as one of several genes critical for aluminum

tolerance in Arabidopsis', *Proceedings of the National Academy of Sciences of the United States of America*, 103(25), pp. 9738–43.

Hoffland, E., Wei, C. and Wissuwa, M. (2006) 'Organic anion exudation by lowland rice (Oryza sativa L.) at zinc and phosphorus deficiency', *Plant and Soil*, 283(1–2), pp. 155–162.

Honkanen, S., Jones, V. A. S., Morieri, G., Champion, C., Hetherington, A. J., Kelly, S., Proust, H., Saint-Marcoux, D., Prescott, H. and Dolan, L. (2016) 'The Mechanism Forming the Cell Surface of Tip-Growing Rooting Cells Is Conserved among Land Plants', *Current Biology*, 26(23), pp. 3238–3244.

Hynes, M. F. and O'Connell, M. P. (1990) 'Host plant effect on competition among strains of *Rhizobium leguminosarum*', *Canadian Journal of Microbiology*, 36(12), pp. 864–869.

Ishiyama, D., Vujaklija, D. and Davies, J. (2004) 'Novel Pathway of Salicylate Degradation by Novel Pathway of Salicylate Degradation by Streptomyces sp . Strain WA46', *Applied and Environmental Microbiology*, 70(3), pp. 1297–1306.

Jackson, R. W., Preston, G. M. and Rainey, P. B. (2005) 'Genetic characterization of Pseudomonas fluorescens SBW25 rsp gene expression in the phytosphere and in vitro', *Journal of Bacteriology*, 187(24), pp. 8477–8488.

Jaeger, C. H., Lindow, S. E., Miller, W., Clark, E. and Firestone, M. K. (1999) 'Mapping of sugar and amino acid availability in soil around roots with bacterial sensors of sucrose and tryptophan', *Applied and Environmental Microbiology*, 65(6), pp. 2685–2690. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10347061 (Accessed: 13 August 2017).

Jain, V. and Gupta, K. (2003) 'The flavonoid naringenin enhances intercellular colonization of rice roots by Azorhizobium caulinodans', *Biology and Fertility of Soils*, 38(2), pp. 119–123.

Jaitz, L., Mueller, B., Koellensperger, G., Huber, D., Oburger, E., Puschenreiter, M. and Hann, S. (2011) 'LC-MS analysis of low molecular weight organic acids derived from root exudation', *Analytical and Bioanalytical Chemistry*, 400(8), pp. 2587–2596.

Jones, D. L. (1998) 'Organic acids in the rhizosphere - A critical review', *Plant and Soil*, 205(1), pp. 25–44.

Jones, D. L., Edwards, A. C., Donachie, K. and Darrah, P. R. (1994) 'Role of proteinaceous amino acids released in root exudates in nutrient acquisition from the rhizosphere', *Plant and Soil*, 158(2), pp. 183–192.

Jones, D. L., Nguyen, C. and Finlay, R. D. (2009) 'Carbon flow in the rhizosphere: Carbon trading at the soil-root interface', *Plant and Soil*, 321(1–2), pp. 5–33.

Jones, V. A. S. and Dolan, L. (2012) 'The evolution of root hairs and rhizoids.', *Annals of botany*, 110(2), pp. 205–212.

Kaiser, C., Kilburn, M. R., Clode, P. L., Fuchslueger, L., Koranda, M., Cliff, J. B., Solaiman, Z. M. and Murphy, D. V. (2015) 'Exploring the transfer of recent plant photosynthates to soil microbes: Mycorrhizal pathway vs direct root exudation', *New Phytologist*, 205(4), pp. 1537–1551.

Kamilova, F., Kravchenko, L. V., Shaposhnikov, A. I., Makarova, N. and Lugtenberg, B. (2006) 'Effects of the Tomato Pathogen Fusarium oxysporum f. sp. radicis-lycopersici and of the Biocontrol Bacterium Pseudomonas fluorescens WCS365 on the Composition of Organic Acids and Sugars in Tomato Root Exudate', *Molecular Plant-Microbe Interactions*, 19(10), pp. 1121–1126.

King, J. M., Digrazia, P. M., Applegate, B., Burlage, R., Sanseverino, J., Dunbar, P., Larimer, F. and Sayler, G. S. (1990) 'Rapid, sensitive bioluminescent reporter technology for naphthalene exposure and biodegradation.', *Science (New York, N.Y.)*, 249(4970), pp. 778–781.

Kitts, C. L., Schaechter, L. E., Rabin, R. S. and Ludwig, R. A. (1989) 'Identification of

Cyclic Intermediates in Azorhizobium caulinodans Nicotinate Catabolism', *JOURNAL OF BACTERIOLOGY*, 171(6), pp. 3406–3411. Available at: http://jb.asm.org/content/171/6/3406.full.pdf (Accessed: 21 July 2017).

Kleckner, N. (1981) 'Transposable Elements in Prokaryotes', *Annual Review of Genetics*, 15(1), pp. 341–404.

Knee, E. M., Gong, F. C., Gao, M., Teplitski, M., Jones, a R., Foxworthy, a, Mort, a J. and Bauer, W. D. (2001) 'Root mucilage from pea and its utilization by rhizosphere bacteria as a sole carbon source.', *Molecular plant-microbe interactions : MPMI*, 14(6), pp. 775–84.

Kobayashi, T. and Nishizawa, N. K. (2012) 'Iron uptake, translocation, and regulation in higher plants.', *Annual review of plant biology*, 63(1), pp. 131–152.

Koch, M., Delmotte, N., Ahrens, C. H., Omasits, U., Schneider, K., Danza, F., Padhi, B., Murset, V., Braissant, O., Vorholt, J. A., Hennecke, H. and Pessi, G. (2014) 'A link between arabinose utilization and oxalotrophy in bradyrhizobium japonicum', *Applied and Environmental Microbiology*, 80(7), pp. 2094–2101.

Kochian, L. V (1995) 'Cellular Mechanisms of Aluminum Toxicity and Resistance in Plants', *Annual Review of Plant Physiology and Plant Molecular Biology*, 46(1), pp. 237–260.

Kochian, L. V., Hoekenga, O. A. and Piñeros, M. A. (2004) 'How Do Crop Plants Tolerate Acid Soils? Mechanisms of Aluminum Tolerance and Phosphorous Efficiency', *Annual Review of Plant Biology*, 55(1), pp. 459–493.

Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., Dörmann, P., Weckwerth, W., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A. R. and Steinhauser, D. (2005) 'GMD@CSB.DB: The Golm metabolome database', *Bioinformatics*, 21(8), pp. 1635–1638.

Kostrzynska, M., Leung, K. T., Lee, H. and Trevors, J. T. (2002) 'Green fluorescent

protein-based biosensor for detecting SOS-inducing activity of genotoxic compounds', *Journal of Microbiological Methods*, 48(1), pp. 43–51.

Kuske, C. R., Ticknor, L. O., Miller, M. E., Dunbar, M., Davis, J. a, Barns, S. M., Dunbar, J. M. and Belnap, J. (2002) 'Comparison of Soil Bacterial Communities in Rhizospheres of Three Plant Species and the Interspaces in an Arid Grassland Comparison of Soil Bacterial Communities in Rhizospheres of Three Plant Species and the Interspaces in an Arid Grassland', *Applied and Environmental Microbiology*, 68(4), pp. 1854–1863.

Lakshmanan, V., Selvaraj, G. and Bais, H. P. (2014) 'Functional soil microbiome: belowground solutions to an aboveground problem.', *Plant physiology*, 166(2), pp. 689–700.

Lang, E., Schumann, P., Adler, S., Sproer, C. and Sahin, N. (2013) 'Azorhizobium oxalatiphilum sp. nov., and emended description of the genus Azorhizobium', *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, 63(Pt 4), pp. 1505–1511.

Langridge, G. C., Phan, M. D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E., Dougan, G., Wain, J., Parkhill, J. and Turner, A. K. (2009) 'Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants', *Genome Research*, 19(12), pp. 2308–2316.

Lee, K.-B., De Backer, P., Aono, T., Liu, C.-T., Suzuki, S., Suzuki, T., Kaneko, T., Yamada, M., Tabata, S., Kupfer, D. M., Najar, F. Z., Wiley, G. B., Roe, B., Binnewies, T. T., Ussery, D. W., D'Haeze, W., Herder, J. Den, Gevers, D., Vereecke, D., Holsters, M. and Oyaizu, H. (2008) 'The genome of the versatile nitrogen fixer Azorhizobium caulinodans ORS571.', *BMC genomics*, 9(1), p. 271.

Lee, K. B., Liu, C. Te, Anzai, Y., Kim, H., Aono, T. and Oyaizu, H. (2005) 'The hierarchical system of the "Alphaproteobacteria": Description of Hyphomonadaceae

fam. nov., Xanthobacteraceae fam. nov. and Erythrobacteraceae fam. nov', *International Journal of Systematic and Evolutionary Microbiology*, 55(5), pp. 1907–1919.

Levy, S. E. and Myers, R. M. (2016) 'Advancements in Next-Generation Sequencing', *Annual Review of Genomics and Human Genetics*, 17(1), pp. 95–115.

Li, X. gang, Liu, B., Heia, S., Liu, D. dou, Han, Z. min, Zhou, K. xin, Cui, J. jie, Luo, J. yu and Zheng, Y. ping (2009) 'The effect of root exudates from two transgenic insect-resistant cotton lines on the growth of Fusarium oxysporum', *Transgenic Research*, 18(5), pp. 757–767.

Li, X. gang, Zhang, T. L., Wang, X. X., Hua, K., Zhao, L. and Han, Z. M. (2013) 'The composition of root exudates from two different resistant peanut cultivars and their effects on the growth of soil-borne pathogen', *International Journal of Biological Sciences*, 9(2), pp. 164–173.

Libert, B. and Franceschi, V. R. (1987) 'Oxalate in Crop Plants', *J. Agric. Food Chem*, 35, pp. 926–938. Available at: http://pubs.acs.org/doi/pdf/10.1021/jf00078a019 (Accessed: 22 July 2017).

Ligaba, A., Katsuhara, M., Ryan, P. R., Shibasaka, M. and Matsumoto, H. (2006) 'The BnALMT1 and BnALMT2 genes from rape encode aluminum-activated malate transporters that enhance the aluminum resistance of plant cells.', *Plant physiology*, 142(3), pp. 1294–303.

Ling, J., Wang, H., Wu, P., Li, T., Tang, Y., Naseer, N., Zheng, H., Masson-Boivin, C., Zhong, Z. and Zhu, J. (2016) 'Plant nodulation inducers enhance horizontal gene transfer of Azorhizobium caulinodans symbiosis island.', *Proceedings of the National Academy of Sciences of the United States of America*, 113(48), pp. 13875–13880.

Liu, A., Ma, B. L. and Bomke, a. a. (2005) 'Effects of Cover Crops on Soil Aggregate Stability, Total Organic Carbon, and Polysaccharides', *Soil Science Society of*

America Journal, 69(6), pp. 2041–2048.

Liu, J., Magalhaes, J. V., Shaff, J. and Kochian, L. V. (2009) 'Aluminum-activated citrate and malate transporters from the MATE and ALMT families function independently to confer Arabidopsis aluminum tolerance', *Plant Journal*, 57(3), pp. 389–399.

Liu, Y., Zhang, N., Qiu, M., Feng, H., Vivanco, J. M., Shen, Q. and Zhang, R. (2014) 'Enhanced rhizosphere colonization of beneficial Bacillus amyloliquefaciens SQR9 by pathogen infection', *FEMS Microbiology Letters*, 353(1), pp. 49–56.

López-Leal, G., Tabche, M. L., Castillo-Ramírez, S., Mendoza-Vargas, A., Ramírez-Romero, M. a and Dávila, G. (2014) 'RNA-Seq analysis of the multipartite genome of Rhizobium etli CE3 shows different replicon contributions under heat and saline shock.', *BMC genomics*, 15, p. 770.

De Lorenzo, V., Herrero, M., Jakubzik, U. and Timmis, K. N. (1990) 'Mini-Tn5 transposoon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria', *Journal of Bacteriology*, 172(11), pp. 6568–6572.

Loudet, O., Gaudon, V., Trubuil, A. and Daniel-Vedele, F. (2005) 'Quantitative trait loci controlling root growth and architecture in Arabidopsis thaliana confirmed by heterogeneous inbred family', *Theoretical and Applied Genetics*, 110(4), pp. 742–753.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550.

Lynch, J. M. (1987) 'Soil Biology: Accomplishments and Potential', *Soil Science Society of America Journal*, 51(6), p. 1409.

Lynch, J. M. (1990) 'Beneficial interactions between micro-organisms and roots', *Biotechnology Advances*, 8(2), pp. 335–346.

Ma, J. F., Ryan, P. R. and Delhaize, E. (2001) ‘Aluminium tolerance in plants and the complexing role of organic acids’, *Trends in Plant Science*, 6(6), pp. 273–278.

Magoc, Magoc, Derrick Wood and Salzberg (2013) ‘EDGE-pro: Estimated Degree of Gene Expression in Prokaryotic Genomes’, *Evolutionary Bioinformatics*, 9, p. 127.

Mahan, M. J., Slauch, J. M. and Mekalanos, J. J. (1993) ‘Selection of bacterial virulence genes that are specifically induced in host tissues’, *Science*, 259(5095), pp. 686–688.

Maillet, F., Poinsot, V., André, O., Puech-Pagès, V., Haouy, A., Gueunier, M., Cromer, L., Giraudet, D., Formey, D., Niebel, A., Martinez, E. A., Driguez, H., Bécard, G. and Dénarié, J. (2011) ‘Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza.’, *Nature*, 469(7328), pp. 58–63.

Martínez, D., Garza, E., Betancourt, R., Mónico, T., Antúnez, C., Román, A., Flores, E., Illades, B. and Fernández, G. (2010) ‘Helicobacter pylori vacA genotypes with gastric ulcer and chronic gastritis.’, *BioMed Central Gastroenterology*, 10(1), pp. 126–130.

Mazurkiewicz, P., Tang, C. M., Boone, C. and Holden, D. W. (2006) ‘Signature-tagged mutagenesis: barcoding mutants for genome-wide screens’, *Nature Reviews Genetics*, 7(12), pp. 929–939.

McNear Jr., D. H. (2013) ‘The Rhizosphere - Roots, Soil and Everything In Between’, *Nature Education Knowledge*, 4(3), p. 1. Available at: https://www.nature.com/scitable/knowledge/library/the-rhizosphere-roots-soiland-67500617 (Accessed: 30 July 2017).

McSpadden-Gardener, B. B. and Weller, D. M. (2001) ‘Changes in populations of rhizosphere bacteria associated with take-all disease of wheat changes in populations of rhizosphere bacteria associated with take-all disease of wheat’, *Applied and Environmental Microbiology*, 67(10), pp. 4414–4425.

Meighen, E. A. (1993) ‘Bacterial bioluminescence: organization, regulation, and application of the lux genes.’, *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 7(11), pp. 1016–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8370470 (Accessed: 11 August 2017).

Mendes, R., Kruijt, M., de Bruijn, I., Dekkers, E., van der Voort, M., Schneider, J. H. M., Piceno, Y. M., DeSantis, T. Z., Andersen, G. L., Bakker, P. A. H. M. and Raaijmakers, J. M. (2011) ‘Deciphering the rhizosphere microbiome for disease-suppressive bacteria’, *Science*, 332(6033), pp. 1097–1100.

Metzker, M. L. (2010) ‘Sequencing technologies - the next generation.’, *Nature reviews. Genetics*, 11(1), pp. 31–46.

Moe, L. A. (2013) ‘Amino acids in the rhizosphere: From plants to microbes’, *American Journal of Botany*, 100(9), pp. 1692–1705.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. and Wold, B. (2008) ‘Mapping and quantifying mammalian transcriptomes by RNA-Seq’, *Nature Methods*, 5(7), pp. 621–628.

Moscatiello, R., Squartini, A., Mariani, P. and Navazio, L. (2010) ‘Flavonoid-induced calcium signalling in Rhizobium leguminosarum bv. viciae’, *New Phytologist*, 188(3), pp. 814–823.

Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N. and Foley, J. A. (2012) ‘Closing yield gaps through nutrient and water management’.

Munoz-Lopez, M. and Garcia-Perez, J. (2010) ‘DNA Transposons: Nature and Applications in Genomics’, *Current Genomics*, 11(2), pp. 115–128.

Muñoz, J. a, Coronado, C., Pérez-Hormaeche, J., Kondorosi, a, Ratet, P. and Palomares, a J. (1998) ‘MsPG3, a Medicago sativa polygalacturonase gene expressed during the alfalfa-Rhizobium meliloti interaction.’, *Proceedings of the National Academy of Sciences of the United States of America*, 95(August), pp.

9687–9692.

Neumann, G. and Römheld, V. (1999) ‘Root excretion of carboxylic acids and protons in phosphorus-deficient plants’, *Plant and Soil*, 211(1), pp. 121–130.

Nibau, C., Gibbs, D. J. and Coates, J. C. (2008) ‘Branching out in new directions: the control of root architecture by lateral root formation’, *New Phytologist*, 179(3), pp. 595–614.

Van Nieuwenhove, C., Van Holm, L., Kulasooriya, S. A. and Vlassak, K. (2000) ‘Establishment of Azorhizobium caulinodans in the rhizosphere of wetland rice (Oryza sativa L.)’, *Biology and Fertility of Soils*, 31(2), pp. 143–149.

O’Callaghan, K. J., Davey, M. R. and Cocking, E. C. (1997) ‘Xylem colonization of the legume Sesbania rostrata by Azorhizobium caulinodans’, *Proceedings of the Royal Society B: Biological Sciences*, 264(1389), pp. 1821–1826.

Oburger, E., Dell’mour, M., Hann, S., Wieshammer, G., Puschenreiter, M. and Wenzel, W. W. (2013) ‘Evaluation of a novel tool for sampling root exudates from soil-grown plants compared to conventional techniques’, *Environmental and Experimental Botany*, 87, pp. 235–247.

Oke, V. and Long, S. R. (1999) ‘Bacterial genes induced within the nodule during the Rhizobium-legume symbiosis’, *Molecular Microbiology*, 32(4), pp. 837–849.

Oldroyd, G. E. D. (2013) ‘Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants.’, *Nature reviews. Microbiology*, 11(4), pp. 252–63.

Oldroyd, G. E. D., Murray, J. D., Poole, P. S. and Downie, J. A. (2011) ‘The Rules of Engagement in the Legume-Rhizobial Symbiosis’, *Annual Review of Genetics*, 45(1), pp. 119–144.

van Opijnen, T., Bodi, K. L. and Camilli, A. (2009) ‘Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms’, *Nature*

Methods, 6(10), pp. 767–772.

van Opijnen, T. and Camilli, A. (2013) 'Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms', *Nature Reviews Microbiology*, 11(7), pp. 435–442.

Osbourn, a E., Barber, C. E. and Daniels, M. J. (1987) 'Identification of plant-induced genes of the bacterial pathogen Xanthomonas campestris pathovar campestris using a promoter-probe plasmid.', *The EMBO journal*, 6(1), pp. 23–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=553351&tool=pmcentrez&rendertype=abstract (Accessed: 15 August 2017).

Osmont, K. S., Sibout, R. and Hardtke, C. S. (2007) 'Hidden Branches: Developments in Root System Architecture'.

Pandey, A. K. and Dwivedi, U. N. (2007) 'Induction, isolation and purification of mimosine degradation enzyme from newly isolated Pseudomonas putida STM 905', *Enzyme and Microbial Technology*, 40(5), pp. 1059–1066.

Paschold, A., Marcon, C., Hoecker, N. and Hochholdinger, F. (2010) 'Molecular dissection of heterosis manifestation during early maize root development', *Theoretical and Applied Genetics*, 120(2), pp. 383–388.

Pawlowski, K., Ratet, P., Schell, J. and de Bruijn, F. J. (1987) 'Cloning and characterization of nifA and ntrC genes of the stem nodulating bacterium ORS571, the nitrogen fixing symbiont of Sesbania rostrata: Regulation of nitrogen fixation (nif) genes in the free living versus symbiotic state', *MGG Molecular & General Genetics*, 206(2), pp. 207–219.

Peoples, M. B., Brockwell, J., Herridge, D. F., Rochester, I. J., Alves, B. J. R., Urquiaga, S., Boddey, R. M., Dakora, F. D., Bhattarai, S., Maskey, S. L., Sampet, C., Rerkasem, B., Khan, D. F., Hauggaard-Nielsen, H. and Jensen, E. S. (2009) 'The

contributions of nitrogen-fixing crop legumes to the productivity of agricultural systems', *Symbiosis*, 48(1–3), pp. 1–17.

Péret, B., Desnos, T., Jost, R., Kanno, S., Berkowitz, O. and Nussaume, L. (2014) 'Root architecture responses: in search of phosphate.', *Plant physiology*, 166(4), pp. 1713–23.

Pérez-Montaño, F., del Cerro, P., Jiménez-Guerrero, I., López-Baena, F. J., Cubo, M. T., Hungria, M., Megías, M. and Ollero, F. J. (2016) 'RNA-seq analysis of the Rhizobium tropici CIAT 899 transcriptome shows similarities in the activation patterns of symbiotic genes in the presence of apigenin and salt', *BMC Genomics*, 17(1), p. 198.

Perry, B. J., Akter, M. S. and Yost, C. K. (2016) 'The use of transposon insertion sequencing to interrogate the core functional genome of the legume symbiont rhizobium leguminosarum', *Frontiers in Microbiology*, 7(NOV), pp. 3389–1873.

Perry, B. J. and Yost, C. K. (2014) 'Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the Rhizobiaceae', *BMC Microbiology*, 14(1), p. 298.

Perry, R. D. (1999) 'Signature-tagged mutagenesis and the hunt for virulence factors', *Trends in Microbiology*, 7(10), pp. 385–388.

Pini, F., East, A. K., Appia-Ayme, C., Tomek, J., Karunakaran, R., Mendoza-Suarez, M., Edwards, A., Terpolilli, J. J., Roworth, J., Downie, J. A. and Poole, P. S. (2017) 'Lux bacterial biosensors for *in vivo* spatiotemporal mapping of root secretion', *In Review*.

Pini, F., East, A. K., Appia-Ayme, C., Tomek, J., Karunakaran, R., Mendoza-Suárez, M., Edwards, A., Terpolilli, J. J., Roworth, J., Downie, J. A. and Poole, P. S. (2017) 'Bacterial biosensors for in vivo spatiotemporal mapping of root secretion', *Plant Physiology*, 174(3).

Pobigaylo, N., Wetter, D., Szymczak, S., Schiller, U., Kurtz, S., Meyer, F., Nattkemper, T. W. and Becker, A. (2006) 'Construction of a large signature-tagged mini-Tn5 transposon library and its application to mutagenesis of Sinorhizobium meliloti', *Applied and Environmental Microbiology*, 72(6), pp. 4329–4337.

Rainey, P. B. (1999) 'Adaptation of Pseudomonas fluorescens to the plant rhizosphere.', *Environmental microbiology*, 1(3), pp. 243–257.

Ramachandran, V. K., East, A. K., Karunakaran, R., Downie, J. A. and Poole, P. S. (2011) 'Adaptation of Rhizobium leguminosarum to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics', *Genome Biology*, 12(10), p. R106.

Ravishankara, A. R., Daniel, J. S. and Portmann, R. W. (2009) 'Nitrous oxide (N2O): the dominant ozone-depleting substance emitted in the 21st century.', *Science*, 326(October), pp. 123–125.

Rebouillat, J., Dievart, A., Verdeil, J. L., Escoute, J., Giese, G., Breitler, J. C., Gantet, P., Espeout, S., Guiderdoni, E. and Périn, C. (2009) 'Molecular genetics of rice root development', *Rice*, 2(1), pp. 15–34.

Rediers, H., Bonnecarrère, V., Rainey, P. B., Hamonts, K., Vanderleyden, J. and De Mot, R. (2003) 'Development and Application of a dapB-Based in Vivo Expression Technology System to Study Colonization of Rice by the Endophytic Nitrogen-Fixing Bacterium Pseudomonas stutzeri A15', *Applied and Environmental Microbiology*, 69(11), pp. 6864–6874.

Rediers, H., Rainey, P. B., Vanderleyden, J. and De Mot, R. (2005) 'Unraveling the secret lives of bacteria: use of in vivo expression technology and differential fluorescence induction promoter traps as tools for exploring niche-specific gene expression', *Microbiol Mol Biol Rev*, 69(2), pp. 217–261.

Richardson, A. E., Lynch, J. P., Ryan, P. R., Delhaize, E., Smith, F. A., Smith, S. E.,

Harvey, P. R., Ryan, M. H., Veneklaas, E. J., Lambers, H., Oberson, A., Culvenor, R. A. and Simpson, R. J. (2011) 'Plant and microbial strategies to improve the phosphorus efficiency of agriculture', *Plant and Soil*, 349(1–2), pp. 121–156.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R. W., Fabry, V. J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P. and Foley, J. A. (2009) 'A safe operating space for humanity', *Nature*, 461(7263), pp. 472–475.

Rothschild, L. J. and Mancinelli, R. L. (2001) 'Life in Extreme Environments', *Nature*, 409(6823), pp. 1092–1101.

Sahin, N. (2003) 'Oxalotrophic bacteria', *Research in Microbiology*, 154(6), pp. 399–407.

Sahin, N. (2005) 'Isolation and characterization of a diazotrophic, oxalate-oxidizing bacterium from sour grass (Oxalis pes-caprae L.)', *Research in Microbiology*, 156, pp. 452–456.

Salas, M. E., Lozano, M. J., López, J. L., Draghi, W. O., Serrania, J., Torres Tejerizo, G. A., Albicoro, F. J., Nilsson, J. F., Pistorio, M., Del Papa, M. F., Parisi, G., Becker, A. and Lagares, A. (2017) 'Specificity traits consistent with legume-rhizobia coevolution displayed by *Ensifer meliloti* rhizosphere colonization', *Environmental Microbiology*.

Sanchez, P. A. and Swaminathan, M. S. (2005) 'Hunger in Africa: The link between unhealthy people and unhealthy soils', *Lancet*, 365(9457), pp. 442–444.

Sasaki, T., Yamamoto, Y., Ezaki, B., Katsuhara, M., Ahn, S. J., Ryan, P. R., Delhaize, E. and Matsumoto, H. (2004) 'A wheat gene encoding an aluminum-activated malate transporter', *Plant Journal*, 37(5), pp. 645–653.

Schulz, H. and Vetterlein, D. (2007) ‘Analysis of organic acid concentration with time in small soil-solution samples from the rhizosphere of maize (Zea mays L.)’, *Journal of Plant Nutrition and Soil Science*, 170(5), pp. 640–644.

Senthilkumar, M., Madhaiyan, M., Sundaram, S. and Kannaiyan, S. (2009) ‘Intercellular colonization and growth promoting effects of Methylobacterium sp. with plant-growth regulators on rice (Oryza sativa L. Cv CO-43)’, *Microbiological Research*, 164(1), pp. 92–104.

Serrania, J., Johner, T., Rupp, O., Goesmann, A. and Becker, A. (2016) ‘Massive parallel insertion site sequencing of an arrayed Sinorhizobium meliloti signature-tagged mini-Tn 5 transposon mutant library’, *Journal of Biotechnology*.

Shaw, L. J., Morris, P. and Hooker, J. E. (2006) ‘Perception and modification of plant flavonoid signals by rhizosphere microorganisms’, *Environmental Microbiology*, 8(11), pp. 1867–1880.

Shimomura, O., Johnson, F. H. and Saiga, Y. (1962) ‘Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea’, *Journal of Cellular and Comparative Physiology*, 59(3), pp. 223–239.

Silby, M. W. and Levy, S. B. (2004) ‘Use of in vivo expression technology to identify genes important in growth and survival of Pseudomonas fluorescens Pf0-1 in soil: Discovery of expressed sequences with novel genetic organization’, *Journal of Bacteriology*, 186(21), pp. 7411–7419.

Simon, R., Priefer, U. and Pühler, A. (1983) ‘A broad host range mobilization system for in vivo genetic engineering: Transposon mutagenesis in gram negative bacteria’, *Bio/Technology*, 1(9), pp. 784–791.

Steele, M. I., Lorenz, D., Hatter, K., Park, A. and Sokatch, J. R. (1992) ‘Characterization of the mmsAB operon of Pseudomonas aeruginosa PAO encoding methylmalonate-semialdehyde dehydrogenase and 3-hydroxyisobutyrate

dehydrogenase', *Journal of Biological Chemistry*, 267(19), pp. 13585–13592. Available at: http://www.jbc.org/content/267/19/13585.full.pdf (Accessed: 20 April 2018).

Stone, P. J., O'Callaghan, K. J., Davey, M. R. and Cocking, E. C. (2001) '*Azorhizobium caulinodans* ORS571 Colonizes the Xylem of *Arabidopsis thaliana*', *Molecular Plant-Microbe Interactions*, 14(1), pp. 93–97.

Sun, Y., Connor, M. G., Pennington, J. M. and Lawrenz, M. B. (2012) 'Development of Bioluminescent Bioreporters for In Vitro and In Vivo Tracking of Yersinia pestis', *PLoS ONE*. Edited by Y.-F. Chang, 7(10), p. e47123.

Takagi, S., Nomoto, K. and Takemoto, T. (1984) 'Physiological aspect of mugineic acid, a possible phytosiderophore of graminaceous plants.', *Journal of Plant Nutrition*, 7(1–5), pp. 469–477.

Teufel, R., Mascaraque, V., Ismail, W., Voss, M., Perera, J., Eisenreich, W., Haehnel, W. and Fuchs, G. (2010) 'Bacterial phenylalanine and phenylacetate catabolic pathway revealed', *Proceedings of the National Academy of Sciences*, 107(32), pp. 14390–14395.

Tilman, D., Balzer, C., Hill, J. and Befort, B. L. (2011) 'Global food demand and the sustainable intensification of agriculture', *Proceedings of the National Academy of Sciences of the United States of Americaof the United States of America*, 108(50), pp. 20260–20264.

Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., Charron, P., Duensing, N., Frei dit Frey, N., Gianinazzi-Pearson, V., Gilbert, L. B., Handa, Y., Herr, J. R., Hijri, M., Koul, R., Kawaguchi, M., Krajinski, F., Lammers, P. J., Masclaux, F. G., Murat, C., Morin, E., Ndikumana, S., Pagni, M., Petitpierre, D., Requena, N., Rosikiewicz, P., Riley, R., Saito, K., San Clemente, H., Shapiro, H., van Tuinen, D., Becard, G., Bonfante, P., Paszkowski, U., Shachar-Hill, Y. Y.,

Tuskan, G. A., Young, J. P. W., Sanders, I. R., Henrissat, B., Rensing, S. A., Grigoriev, I. V., Corradi, N., Roux, C. and Martin, F. (2013) 'Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis', *Proceedings of the National Academy of Sciences*, 110(50), pp. 20117–20122.

Treseder, K. K. (2013) 'The extent of mycorrhizal colonization of roots and its influence on plant growth and phosphorus content', *Plant and Soil*, 371(1–2), pp. 1–13.

Tsukada, S., Aono, T., Akiba, N., Lee, K. B., Liu, C. Te, Toyazaki, H. and Oyaizu, H. (2009) 'Comparative genome-wide transcriptional profiling of Azorhizobium caulinodans ORS571 grown under free-living and symbiotic conditions', *Applied and Environmental Microbiology*, 75(15), pp. 5037–5046.

Turner, T. R., James, E. K. and Poole, P. S. (2013) 'The plant microbiome', *Genome Biology*, 14(6), p. 209.

Turner, T. R., Ramakrishnan, K., Walshaw, J., Heavens, D., Alston, M., Swarbreck, D., Osbourn, A., Grant, A. and Poole, P. S. (2013) 'Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants', *ISME Journal*, 7(12), pp. 2248–2258.

Tzin, V. and Galili, G. (2010) 'The Biosynthetic Pathways for Shikimate and Aromatic Amino Acids in *Arabidopsis thaliana*', *The Arabidopsis Book*, 8(i), p. e0132.

Ul Haque, M. F., Nadalig, T., Bringel, F., Schaller, H. and Vuilleumier, S. (2013) 'Fluorescence-based bacterial bioreporter for specific detection of methyl halide emissions in the environment', *Applied and Environmental Microbiology*, 79(21), pp. 6561–6567.

USDA (2012) *A Global Staple*. United States Department of Agriculture. Available at: http://www.ers.usda.gov/topics/crops/rice/background.aspx.

Uselman, S. M., Qualls, R. G. and Thomas, R. B. (2000) 'Effects of increased

atmospheric CO2, temperature, and soil N availability on root exudation of dissolved organic carbon by a N-fixing tree (Robinia pseudoacacia L.)', *Plant and Soil*, 222(1–2), pp. 191–202.

Vanderlinde, E. M., Hynes, M. F. and Yost, C. K. (2014) 'Homoserine catabolism by Rhizobium leguminosarum bv. viciae 3841 requires a plasmid-borne gene cluster that also affects competitiveness for nodulation', *Environmental Microbiology*, 16(1), pp. 205–217.

Vanderlinde, E. M., Magnus, S. A., Tambalo, D. D., Koval, S. F. and Yost, C. K. (2011) 'Mutation of a broadly conserved operon (RL3499-RL3502) from Rhizobium leguminosarum biovar viciae causes defects in cell morphology and envelope integrity', *Journal of Bacteriology*, 193(11), pp. 2684–2694.

Veneklaas, E. J., Stevens, J., Cawthray, G. R., Turner, S., Grigg, A. M. and Lambers, H. (2003) 'Chickpea and white lupin rhizosphere carboxylates vary with soil properties and enhance phosphorus uptake', *Plant and Soil*, 248(1–2), pp. 187–197.

Wang, Y., Shyy, J. Y.-J. and Chien, S. (2008) 'Fluorescence proteins, live-cell imaging, and mechanobiology: seeing is believing.', *Annual review of biomedical engineering*, 10(1), pp. 1–38.

Webster, G., Gough, C., Vasse, J., Batchelor, C. a, O'Callaghan, K. J., Kothari, S. L., Davey, M. R., Dénarié, J. and Cocking, E. C. (1997) 'Interactions of rhizobia with rice and wheat', *Plant and Soil*, 194(1–2), pp. 115–122.

Webster, G., Jain, V., Davey, M. R., Gough, C., Vasse, J., Dénarié, J. and Cocking, E. C. (1998) 'The flavonoid naringenin stimulates the intercellular colonization of wheat roots by Azorhizobium caulinodans', *Plant, Cell & Environment*, 21(4), pp. 373–383.

Weert, S. De, Vermeiren, H., Mulders, I. H. M., Kuiper, I., Hendrickx, N., Guido, V., Vanderleyden, J., Mot, R. De, Lugtenberg, B. J. J., de Weert, S., Vermeiren, H.,

Mulders, I. H. M., Kuiper, I., Hendrickx, N., Bloemberg, G. V, Vanderleyden, J., De Mot, R. and Lugtenberg, B. J. J. (2002) 'Flagella-Driven Chemotaxis Towards Exudate Components Is an Important Trait for Tomato Root Colonization by Pseudomonas fluorescens', *Molecular plant-microbe interactions : MPMI*, 15(11), pp. 1173–1180.

Weert, S., Kuiper, I., Lagendijk, E. L., Lamers, G. E. M. and Lugtenberg, B. J. J. (2004) 'Role of chemotaxis toward fusaric acid in colonization of hyphae of Fusarium oxysporum f. sp. radicis-lycopersici by Pseudomonas fluorescens WCS365.', *Molecular plant-microbe interactions : MPMI*, 17(11), pp. 1185–1191.

Weidenhamer, J. D., Mohney, B. K., Shihada, N. and Rupasinghe, M. (2014) 'Spatial and Temporal Dynamics of Root Exudation: How Important is Heterogeneity in Allelopathic Interactions?', *Journal of Chemical Ecology*, 40(8), pp. 940–952.

Wheatley, R. M., Ramachandran, V. K., Geddes, B. A., Perry, B. J., Yost, C. K. and Poole, P. S. (2017) 'Role of O2 in the growth of Rhizobium leguminosarum bv. viciae 3841 on glucose and succinate', *Journal of Bacteriology*, 199(1), pp. e00572-16.

Wheeler, T. and von Braun, J. (2013) 'Climate Change Impacts on Global Food Security', *Science*, 341(6145). Available at: http://science.sciencemag.org/content/341/6145/508.full (Accessed: 29 July 2017).

Wielbo, J. (2012) 'Rhizobial communities in symbiosis with legumes: genetic diversity, competition and interactions with host plants', *Open Life Sciences*, 7(3), pp. 363–372.

Williams, T. C. R., Poolman, M. G., Howden, A. J. M., Schwarzlander, M., Fell, D. A., Ratcliffe, R. G. and Sweetlove, L. J. (2010) 'A Genome-Scale Metabolic Model Accurately Predicts Fluxes in Central Carbon Metabolism under Stress Conditions', *Plant Physiology*, 154(1), pp. 311–323.

Yokosho, K., Yamaji, N. and Ma, J. F. (2011) 'An Al-inducible MATE gene is involved

in external detoxification of Al in rice', *Plant Journal*, 68(6), pp. 1061–1069.

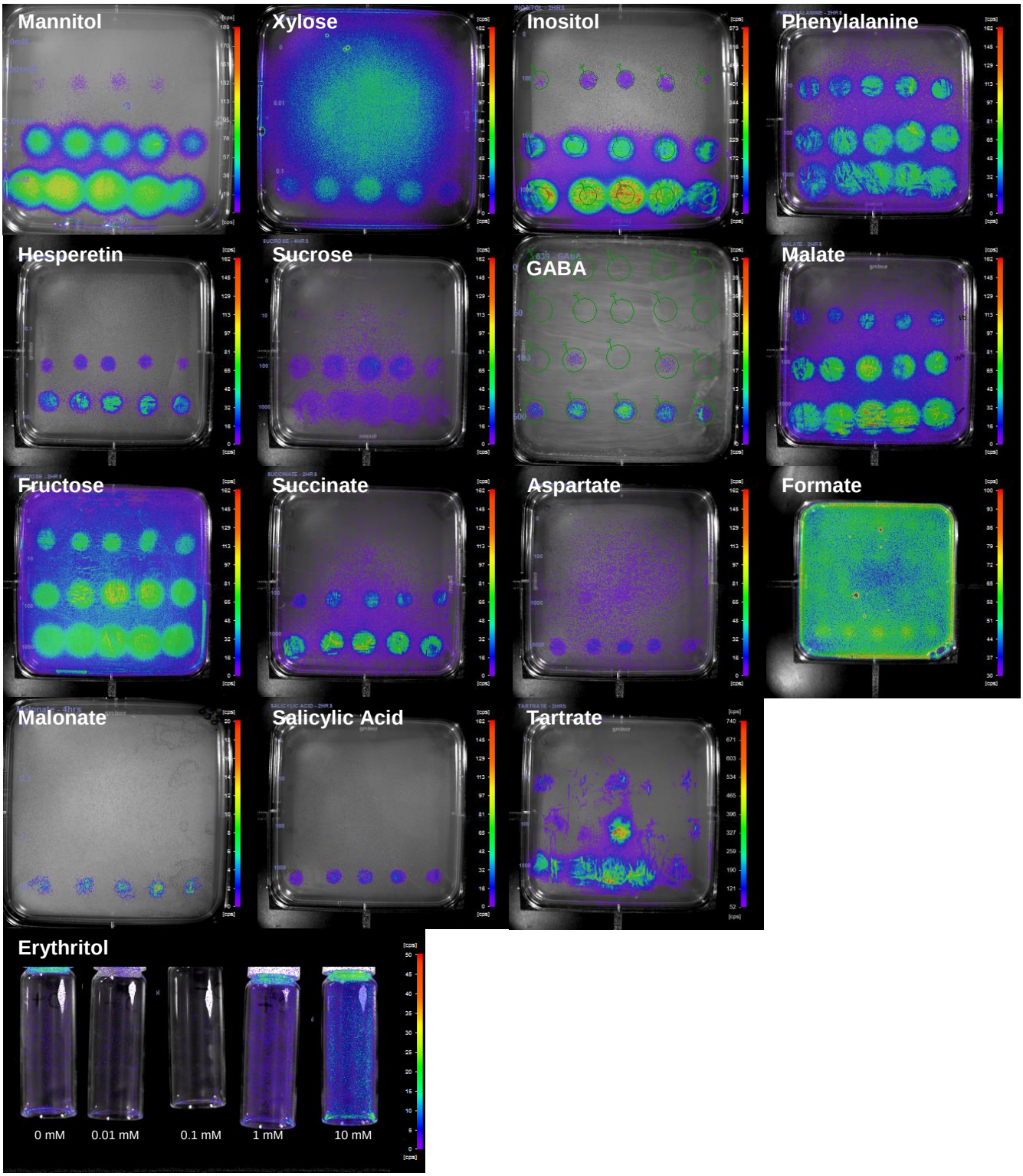
Yuan, J., Zhang, N., Huang, Q., Raza, W., Li, R., Vivanco, J. M. and Shen, Q. (2015) 'Organic acids from root exudates of banana help root colonization of PGPR strain Bacillus amyloliquefaciens NJN-6.', *Scientific reports*, 5(August), pp. 1–8.

Zhang, F., Römheld, V. and Marschner, H. (1989) 'Effect of zinc deficiency in wheat on the release of zinc and iron mobilizing root exudates', *Zeitschrift für Pflanzenernährung und Bodenkunde*, 152(2), pp. 205–210.

Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K. and Liu, X. (2014) 'Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells', *PLoS ONE*. Edited by S.-D. Zhang, 9(1), p. e78644.

Appendix

Supplementary 1

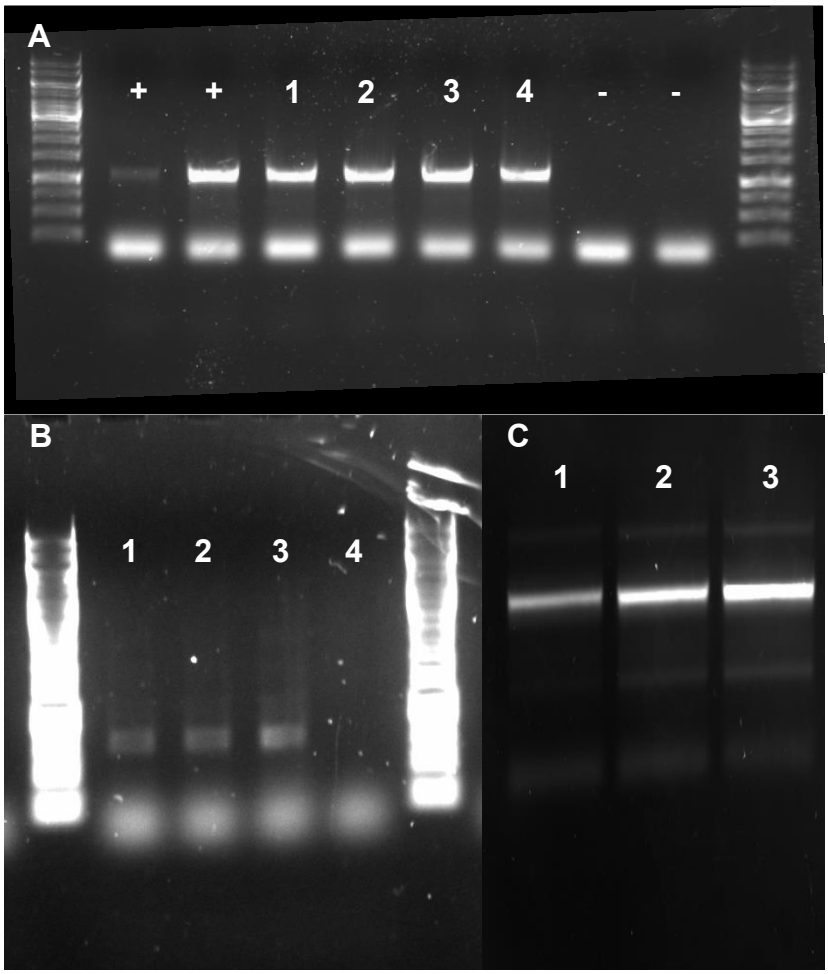


Supplementary 1. Images show inducible bioreporters inoculated with varying concentrations of their inducing molecules and the minimum concentration required to produce a detectable light signal. Biosensors were suspended in UMA and were inoculated with droplets of the inducing molecules and a water control, each row of droplets represent 5 replicates of a single concentration. The Erythritol biosensor was left suspended in UMS and spiked with different concentrations of erythritol. Concentrations in mM from top rows to bottom rows: Hesperetin – 0, 0.001, 0.01, 0.1; Xylose - 0, 0.001, 0.01, 0.1; Inositol - 0, 0.001, 0.01, 0.1; Phenylalanine - 0, 0.01, 0.1, 1; Hesperetin – 0, 0.1, 1, 10; Sucrose – 0, 0.01, 0.1, 1; GABA – 0, 0.05, 0.1, 0.5; Malate - 0, 0.01, 0.1, 1; Fructose - 0, 0.01, 0.1, 1; Succinate - 0, 0.01, 0.1, 1; Aspartate – 0, 0.1, 1, 10; Formate – 0, 0.1, 1, 10; Malonate – 0, 0.1, 1, 10; Salicylic acid - 0, 0.01, 0.1, 1, Tartrate - 0, 0.01, 0.1, 1; Erythritol – 0, 0.01, 0.1, 1, 10.

Supplementary 2: *Oryza sativa* exudate mutant screen

Oryza sativa IR64 mutant lines inoculated with constitutive pTac fusion. Imaged at 6 dpi. Each sheet contains plates grown, inoculated and images on the same days. Images contained on attached CD.

Supplementary 3



Supplementary 3. Gel electrophoresis images used to confirm successful insertional mutagenesis of *A. caulinodans* ORS571 using the plasmid pSAM_Rl. A – PCR reactions used to confirm the successful insertion of the pSAM-Rl transposon into the ORS571 genome: 1, 2, 3 & 4 show four individual ORS571-pSAM mutants; pSAM-Rl plasmid DNA used as positive controls (+); ORS571 wild-type DNA used as negative controls (-); 1 kb DNA ladders. B – AP-PCR 1st round of PCR for four ORS571-pSAM mutants; 1 kb DNA ladder. C – 2nd Round AP-PCR of the three samples seen in B.

Supplementary 4: Sequencing reports

Sequencing reports for Ion-Torrent sequencing of INSeq libraries. Images contained on attached CD.

Supplementary 5: Colonisation INSeq data

Tables containing HMM analysis results and resampling results for INSeq studies of ORS571 free-living growth, *O. sativa* rhizosphere growth, and *O. sativa* root attached growth. Contained on attached CD.

Supplementary 6: INSeq analysis scripts

All scripts used for analysis of INSeq sequencing data. Contained on attached CD.

Supplementary 7: RNA cDNA libraries

Tables containing DNA quantification results for cDNA libraries produced from *Azorhizobium caulinodans* ORS571 RNA. Contained on attached CD.

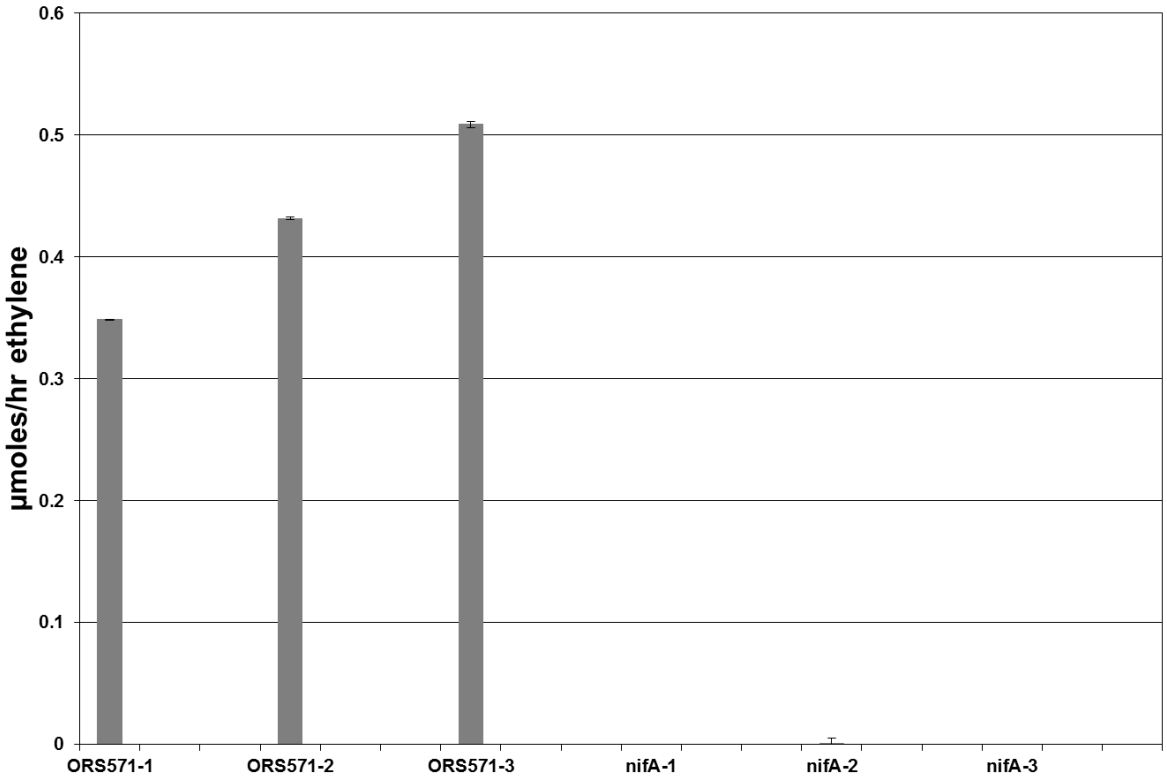
Supplementary 8: RNA-Seq sequencing report

Sequencing reports for Ion-Torrent sequencing of RNA-Seq libraries. Contained on attached CD.

Supplementary 9: Colonisation RNA-Seq data

Table containing RNA-Seq results (number of reads for each gene) for ORS571 during free-living growth, *O. sativa* rhizosphere growth, and *O. sativa* root attached growth. Contained on attached CD.

Supplementary 10



Supplementary 10. Graph showing the nitrogen fixing ability of ORS571 grown on nitrogen free UMA medium. Fixation is measured by the conversion of acetylene to ethylene by the nitrogenase enzyme, as measured by the production of ethylene in µmoles per hour. Three replicates of ORS571 wild-type are shown (ORS571-1, 2 and 3), and three replicates of ORS571-Δ*nifA* incapable of fixation (*nifA*-1, 2 and 3). Error bars show standard error of the mean.

Supplementary 11: Nitrogen-fixing INSeq data

Tables containing HMM analysis results and resampling results for INSeq studies of ORS571 growth at ambient oxygen, 3% oxygen (with supplemented nitrogen) and 3% nitrogen (no-supplemented nitrogen). Contained on attached CD.