Metabolic Investigation of Nitrogen Fixation in the Free-Living Plant Associative Bacteria, *Klebsiella oxytoca*

Thesis submitted by
Matthew Carey

For the degree of Doctor of Philosophy of Imperial College London
and the Diploma of Imperial College London

Supervised by
Dr Jacob Bundy
Dr Jörg Schumacher
Dr Virag Sagi-Kiss
Professor Martin Buck

Department of Metabolism, Digestion and Reproduction
Faculty of Medicine
Imperial College London

2023

Declaration of originality

I certify that this thesis, and the research to which it refers, are the products of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

The following samples or datasets were collected or analysed with collaborators:

- The Scanning electron microscopy (SEM) of extracted cell samples was carried out by Dr Ecaterina Ware (Department of Materials, Imperial College).

- LCMS targeted method development (Chapter 3) was led by Dr Virag Sagi-Kiss (Jake Bundy group, Imperial College) with each metabolite included in the methods tested by either Virag Sagi-Kiss, Yufeng Li (Jake Bundy group, Imperial College) or myself.

- Analysis of biofilm samples by TOF-SIMS was carried out by Dr Sarah Fearn (Department of Materials, Imperial College).

- The RNA Sequencing carried out in Chapter 6 was carried out by Vertis Biotechnologie AG (Freising, Germany).

Copyright statement

Abstract

Biologically available nitrogen is among the most limiting factors to crop growth. Currently this is overcome by the widespread use of inorganic fertilisers supplemented with ammonium, which comes at a great environmental cost. Biological nitrogen fixation, the conversion of atmospheric dinitrogen to ammonium by specialist microorganisms, can reduce reliance on this fertiliser and the impact on the climate. Here, the nitrogen fixing bacteria *Klebsiella oxytoca* is used as a model organism to investigate nitrogen fixation and its effect on global cell metabolism.

In the third chapter of this thesis, a series of sampling and liquid chromatography mass spectrometry methods were tested to investigate how to accurately assess the metabolic state of *K. oxytoca* under nitrogen fixing conditions. In Chapter 4 *K. oxytoca* were cultured under fixing and non-fixing conditions and Nuclear Magnetic Resonance (NMR) analysis was carried out on the supernatant. The results showed butan-2-3-diol, a metabolite known to improve draught resistance in plants, was released by cells fixing nitrogen.

In Chapter 5, analysis of *K. oxytoca* batch cultures using the methods developed in Chapter 3 revealed distinct intracellular metabolic profiles for nitrogen fixing and non-fixing cells. Recent work has shown that some species of nitrogen fixing bacteria can form biofilms and fix nitrogen in air. To the best of the authors knowledge, the work in Chapter 6 is the first reporting of *K. oxytoca* biofilms fixing nitrogen in air. Targeted LC-MS analysis found the metabolic profiles of these biofilms varied over time and between nitrogen fixing and non-fixing biofilms.

Collectively, the results presented in this thesis provide insights into the profound impact nitrogen fixation has on global cell metabolism in *K. oxytoca*. These results can be utilised to improve biotechnological solutions to the nitrogen crisis, helping to reduce the environmental impact of agricultural practices as we expand food production.

Acknowledgements

Firstly, I would like to thank my supervisors Dr Jake Bundy and Dr Jörg Schumacher for their support, guidance and teachings during my PhD study. It's been a pleasure to work with you. A very special thank you to Professor Martin Buck, who has gone over and above the call of duty to help in the completion of this thesis. Your calm guidance is truly appreciated. Thank you to my funders the BBSRC, for making this work possible.

I'm very grateful to those who were always there to be a sounding board and help out in the lab at all hours, without whom I would not have been able to complete any of this work. Namely Virag Sagi-Kiss, Chris Waite and Rob Bradley. I wish you all the best in your future scientific careers, be that in or out of academia, I will always be here to repay the (many) favours!

I would also like to say a massive thank you to all my friends and colleagues old and new that have been a source of support and knowledge, as well as making this whole experience a fun one. 'My boys' Simone Zuffa and Elena Alambriti as well as Mike Olanipekun, Duncan Roberts, Kate Gallagher and Nadeen Habboub. You are all dear friends, and I couldn't have done this without you. Chris Chapman and Connor Jones, I am so glad to have met you all those years ago in sunny Falmouth. Stefan Store and Greg Wills, who have supported me since the insectary days, thank you. Thank you also to Liz Want and Tim Ebbels, you are a credit to the college and your guidance of many aspiring scientists, including myself, is noted and appreciated. Likewise thank you to Sebastian Bosca for the endless support and guidance throughout this process. I am a different person because of you.

A special shout out to the best flat mates ever Izzy Coales and Iain Chudleigh, you have seen me at my best and worst through this process but your friendship has never wavered, I know that we will be friends for life. Also to Grace Barker, my person, I would not be here without your love and support. 'If someone asks, this is where I'll be'.

Finally, thank you to all my family, who have been a constant support throughout, despite never fully understanding what it was I've been doing. Mum, Dad and Ryan, I hope I've made you proud. You all inspire me every day to achieve more than I thought possible. And always remember 'Nothin' for nothin'.

Table of Contents

Abbreviations

Abbreviation	Meaning
AAc	Acetylacetone
Ac-CoA	acetyl-Co-enzyme A
AccQTag	6-Aminoquinoloyl-N-hydroxysuccinidimyl carbamate
ACN	Acetonitrile
ADP	Adenosine diphosphate
amtB	Ammonium Transporter
AMW	Acetonitrile-Methanol-Water
APCI	Atmospheric Pressure Chemical Ionisation
ARA	Acetylene Reduction Assay
ATP	Adenosine Triphosphate
bEBP	Bacterial Enhancer Binding Protein
BNF	Biological Nitrogen Fixation
CDP	Cytidine Diphosphate
CE	Collision Energy
CFU	Colony Forming Units
$CHCl_3$	Chloroform
CID	Collision Induced Dissociation
CO	Centrifuge Only
CO_2	Carbon Dioxide
Crc	Catabolite Repression Control
CV	Cross Validation
D_2O	Deuterium Oxide
DOPA	3,4-dihydroxyphenylalanine
EI	Electron Ionisation
EPS	Extracellular Polymeric Substances
ESI	Electrospray Ionisation
FA	Formic Acid
FAD	Flavin Adenine Dinucleotide
$FADH_2$	Flavin Adenine Dinucleotide reduced
FDR	False Discovery Rate
Fe	Iron

FID	Free Induction Decay
GABA	γ-Aminobutyric acid
GDH	Glutamate Dehydrogenase
GDP	Guanosine diphosphate
GlnA	Gene for Nitrogen Regulation
GlnB	PII-like protein
GlnK	PII protein
GlySal	Glycerol Saline
GMP	Guanosine monophosphate
GOGAT	Glutamate-oxoglutarate aminotransferase or glutamate synthase
GS	Glutamine synthetase
GSSG	Glutathione Disulphide
HAc	Acetic acid
HPLC	High Performance Liquid Chromatography
IPA	Isopropanol
IPC	Ion-Pairing Chromatography
IS2	Internal Standard 2
K_m	Michaelis-Menten constant
LB	Lysogeny Broth
LC	Liquid Chromatography
LC-MS	Liquid Chromatography Mass Spectrometry
m/z	Mass to Charge Ratio
MALDI	Matrix Assisted Laser Desorption Ionisation
MCW	Methanol-Chloroform-Water
MeOH	Methanol
Mg	Magnesium
mM	Millimolar
MoFe	Molybdenum Iron
MS	Mass Spectrometry
MSMLS	Mass Spectrometry Metabolite Library of Standards
N	Nitrogen
NADPH	Nicotinamide Adenine Dinucleotide Phosphate, reduced
NFDM	Nitrogen Free Davis & Mingioli
NH_4Cl	Ammonium Chloride

nif	Genes encoding the nitrogenase enzyme
NMR	Nuclear Magnetic Resonance
NOSEY	Nuclear Overhauser Enhancement Spectroscopy
nt	Nucleotide
Ntr	Nitrogen Regulation system
NtrB	Nitrogen Regulatory protein
NtrC	Nitrogen Regulatory Operon
OD_{600}	Optical Density
OPLS-DA	Orthogonal Projections to Latent Structures-Discriminant Analysis
OSC	Orthogonal Signal Correction
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PGPR	plant growth promoting rhizobacteria
PII(-type protein)	Homotrimetric signal transduction proteins
PLS	Partial Least Squares
Q 123	Quadrupole 123
QC	Quality Control
QqQ	Triple Quad
QTOF	Quadrupole Time-of-Flight
RP	Reversed-Phase
RT	Retention Time
S	Sulphur
SEM	Scanning Electron Microscopy
SOB	Super Optimal Broth
TBA	Tributylamine
TCA (Cycle)	tricarboxylic acid (TCA) cycle
TOF-SIMS	Time of Flight Secondary Ion Mass Spectrometry
TSP	Trimethylsilylpropanoic acid
UPLC-MS	Ultra High Performancce Liquid Chromatography Mass Spectrometry
UR	Uridylyl Removase
UT	Uridylyl Transferase
VFe	Vanadium-iron

WT	Wild-Type
XCMS-MRM	X C Mass Spectrometry - Multiple Reaction Monitoring
µg	Microgram
µL	Microliter
µM	Micromolar

List of figures

List of tables

Chapter 1 Introduction

This thesis presents a body of work inspired by the need to seek sustainable solutions to energy intensive agricultural practices that damage the environment and present societal challenges. My work has addressed how bacterial cells that can synthesise ammonia from dinitrogen gas can be characterised for their ability to carry out this process of biological nitrogen fixation. To do so I have used a range of analytical techniques to probe the molecular physiology of bacterial biological nitrogen fixation, across a range of growth regimes including in biofilms.

1.1 Agriculture and the nitrogen fertilizer problem

1.1.1 The Need for Nitrogen in Agriculture

The UN reports that by the year 2050, the human population will have exceeded 9.8 billion people [1]. Current global food shortages and poor distribution means we are likely to face a global crisis. The food and agriculture organisation predicts that food production must increase by 70% to meet growing demand [2]. The availability of biologically available, reactive nitrogen is often one of the most variable and limiting factors to crop growth; nitrogen is used to build amino acids and proteins essential for growth. Given atmospheric nitrogen is unavailable for crops to use, they require reactive nitrogen for growth (e.g. ammonium), the two main methods of providing nitrogen to crops in modern agriculture are crop rotation and using chemical fertilisers [3], alongside the farming of crops that have a symbiotic relationship with nitrogen fixing bacteria e.g., *Glycine max , the* soya bean.

Crop rotation is the practice of growing different crops overtime on the same land. Typically, this practice sees nitrogen fixing variety of plants, also known as 'green manure', grown in place of the desired crop for a given period. This practice has been shown to increase yield compared to standard monoculture, due to increases in biologically available nitrogen in the soil [4]. Additional benefits of crop rotation include soil pH regulation, weed and pest control, and disease prevention [3,4]. However, due to increased demand for higher yields along with shortened growth periods, chemical fertiliser use has become increasingly common. This practice allows for continuous monocultures while maintaining the benefits of crop rotation [5].

Currently, the Haber Bosch reaction, which sees atmospheric dinitrogen converted to ammonia, is the method employed to generate biologically available nitrogen for synthetic chemical fertilisers. The Haber Bosch process is thought to be responsible for feeding approximately 50% of the global population and is considered one of the greatest human interferences in the nitrogen cycle [6,7]. The

reaction requires a catalyst in conjunction with high pressures and temperatures, resulting in a reaction that is extremely energy intensive. It is estimated that 1.5% of global energy is dedicated to fertiliser production [8], with this expected to rise to over 2% by 2050 [9]. The Haber-Bosch process is also responsible for up to 3% of global CO_2 emissions, and up to 24% of global greenhouse emissions, which contributes to climate change [8,9].

1.1.2 The Nitrogen crisis

A changing climate and increased frequency of extreme weather events has resulted in a close examination of process that emit large volumes of greenhouse gas [10,11]. Broadly, climate change is predicted to dramatically impact the conditions of crop growth resulting in new and difficult challenges to farmers [11,12]. Were industrial farming to rely heavily on quick fixes, which reduce vulnerability to changes in climate but increase greenhouse emissions the problem would only be exacerbated [11]. Although there is historical precedent of quick fixes in farming practices having long lasting unseen consequences [12,13], it could be argued we are at an inflection point where industry players, government and the public are starting to address the issue head on [7,11,14].

For example, inorganic fertiliser use has also been linked to exacerbating global wealth inequality [15]. A lack of access to inorganic fertilisers in developing countries has caused lower crop yields and resulted in poor wealth distribution. Developed countries have abundant fertilizer, and often overuse it, while regions like sub-Saharan Africa produce only 16% of the mean baseline maize yield of that of North America for the same land size [15,16]. To put the difference into context, the US uses 13 tonnes of fertiliser per hectare per year, or 13 times the sustainable limit, while in sub-Saharan Africa only one in twenty farmers can afford to buy fertilisers [16]. However, a recent push by scientific research as well as government bodies has looked to challenge the current status quo.

Furthermore, the environment impacts of synthetic fertiliser use is not just limited to climate change. In spite of the best efforts by farmers, approximately 66% of applied nitrogen is lost to the environment, as a result of low nitrogen use efficacy by crops, resulting in run off into rivers and streams [6,7]. The downstream effects of these environmental effects are severe and include environmental degradation and coastal dead zones [17]. For example, areas which have high fertiliser use have greater instances of algal blooms in ponds, lakes and seas due to the influx of nitrogen to the aquatic environment [8,17,18]. Agal blooms cause disruption to ecosystems and widespread death of species through oxygen deprivation, greatly changing how that habitat operates and impeding it's ability to recover [17,18].

A report by Rockstrom *et al.,* (2009) set out environmental boundaries for the planet that if exceeded would cause irreversible environmental change. The concentration of atmospheric carbon dioxide is currently 10.6% over the proposed limit, in comparison to human interference with the nitrogen cycle which is 245.7% over the limit. This will be in large part due to the effect of Haber Bosch reaction on the nitrogen cycle. One study suggested cutting reactive nitrogen production by 50% globally to alleviate some of the environmental damage [8]. However, given nitrogen's key role in crop production, a reduction in emissions is unlikely unless a more sustainable but equally efficient means of reactive nitrogen production for agricultural purposes is developed.

1.1.3 Biological nitrogen fixation

One such solution may stem from biological nitrogen fixation (BNF). This process is carried out by specialist bacteria and archaea, called diazotrophs, and sees atmospheric nitrogen converted into reactive nitrogen, ammonia through the action of the enzyme nitrogenase. BNF is thought to be responsible for around 45% of reactive nitrogen produced globally, but this is usually in natural systems and not an agricultural setting [8,19]. The ammonia produced by BNF is less susceptible to leaching and volatilisation, as it is produced and used *in situ*, meaning that this process is a sustainable means of nitrogen input within agriculture. Legumes are an excellent example of BNF providing sufficient nitrogen for crop growth. This symbiotic relationship is well studied and found in commercially important crops such as *Pisum sativum* (pea), *Phaseolus vulgaris* (common beans) and *Glycine max* (soybean) [20,21]. Legumes and their nitrogen fixing symbionts are often used as 'green manure' in crop rotations due to their positive effect on the soil and sustainable levels of nitrogen they put into the soil [20].

Furthermore, research shows that terrestrial BNF has an important role in the marine nitrogen cycle, in stark contrast to the harmful effects of synthetic chemical fertilizer run off [7]. A study by Vitousek *et al.,* [22] found that land-based nitrogen fixation may effect atmospheric CO_2 sequestration by ocean biota, due to atmospheric reactive nitrogen fluxes between terrestrial and aquatic environments supporting greater carbon capture capabilities.

BNF can occur either in symbiosis, as in legumes, or free-living environments. Studies of both types of BNF have shown bacteria in symbiosis meet the crops nitrogen requirement for growth [19,23], although the evidence for free living rhizobacteria is less established than in nodules [24,25]. However, a recent review found that nitrogen assimilation by free living bacteria, while variable, may be underestimated and could exceed symbiotic nitrogen fixation, and as such be able to support crop growth on a wider

scale [25]. The genes responsible for the nitrogen fixation are highly conserved and are called the *nif* genes. The classical *nif* regulon of *Klebsiella* species consists of around 20 genes divided into 7 operons, which are responsible for their regulation, assembly, maintenance and function of the nitrogen fixing enzyme, nitrogenase [26].

Klebsiella to study biological nitrogen fixation

Klebsiella, a genus of the Enterobacteriaceae family, are rod-shaped Gram-negative bacteria. This genus is non-spore forming, non-motile bacteria and facultatively anaerobic, and are found in a wide range of environments including soil, water and in human microbiomes [27–29]. Their ubiquity and importance in these environments mean that they have been used as model organisms across a wide range of research topics including medical research and synthetic biology [30].

The genus *Klebsiella* is well characterised both in terms optimal culturing conditions and their genomes, with multiple strain entries in the gene bank [27,31–34]. This has made them ideal candidates for molecular biological approaches which have facilitated investigation into gene function and regulatory networks [26,35]. Two species of *Klebsiella* which have been well studied in relation to nitrogen fixation are *Klebsiella pneumoniae* and *Klebsiella oxytoca*, and their genetic regulation during nitrogen fixation is discussed below. Both species are plant associative nitrogen fixers and have been isolated from the roots of important non-legume agricultural crops such as wheat and rice [26,27,29].

The work in this thesis will focus on *K. oxytoca*, this species has been found in various geographies (Asia, North and South America, and Europe) and soil types and shown to naturally occur in the soil around the roots of plants and crops including rice, bananas and wheat [34,36–38]. The strain used in this work was originally from Ruth Schmitz (Christian Albrechts University, Kiel) and has been used for research by the Martin Buck laboratory prior to the research carried out in this thesis [39–41]. The last time the genome was sequenced there were no mutations although this was before the start of the work carried out in this thesis.

1.2 How does nitrogen fixation work?

1.2.1 The Nitrogenase

The nitrogenase is an oxygen sensitive, adenosine triphosphate (ATP) hydrolyzing, redox active enzyme complex. This enzyme complex is composed of two different proteins, the reductase Fe protein and the catalytic MoFe protein. The Fe protein is synthesised by four genes, *nifH, nifU*, *nifS* and *nifM*. A *nifH* homodimer, the Fe protein consists of two Mg.ATP binding sites and a [4Fe-4S] cluster (Fig. 1.1).In the building of this protein, *nifM* is a chaperone for *nifH*, while *nifU* acts as the scaffolding and transporter of the [4Fe-4S] cluster and *nifS* acts as a S atom donator [26,42].

Conversely, *nifD* and *nifK* are the primary drivers of MoFe protein synthesis [42] which is a heterotetramer composed of a P cluster [8Fe-7S] and a FeMo-co cluster [$MoFe_7S_9$-C-homocitrate] from *nifD*. To initiate the making of the P cluster, NifU delivers two [4Fe-4S] clusters to NifD and NifK after which the P cluster is formed by reactive coupling driven by *nifH*. *nifU* is also necessary for FeMo-co cluster synthesis [26]. NifU delivers NifB a [4Fe-4S] cluster, where a central carbon is added forming NifB-co. After, NifB-co is transferred to the NifEN scaffold, either directly or by NifX, the scaffold already contains a [4Fe-4s] cluster from NifU to produce the VK cluster [8Fe-9S-C]. The VK cluster then becomes the FeMo-co cluster when NifEN is provided homocitrate from NifV and Mo from NifQ, as part of a [Mo-3Fe-4S] cluster, afterwards either NifEN or NifY then transfers the FeMo-co to NifDK finally forming the MoFe protein (Fig. 1 b & c) [19,26]. A summary of the *nif* genes and their functions are given in Table 1.1.

While there is a diverse array of diazotrophs, a nitrogenase that uses the MoFe cluster is the most common, however there are two other known catalytic components. These are substitutes to the MoFe cluster and see Mo substituted for vanadium and/or iron, forming Vanadium-iron [VFe] and/or iron-iron [FeFe] nitrogenases respectively, under Mo limited conditions [43,44]. The different forms of nitrogenase vary in affinity and efficacy, with studies indicating that the Mo-nitrogenase is the most efficient out of the three. Bellenger *et al.,* [45] found that alternative forms of nitrogenase have lower R ratios (the ratio between a measurement of the rates of nitrogen fixation via acetylene reduction and rates measured by the incorporation of ^{15}N) relative to Mo-nitrogenases. The work in this thesis will focus on the diazotroph *Klebsiella oxytoca,* an obligate Mo-nitrogenase, and as such much of the discussion will focus exclusively on Mo-nitrogenases.

Table 1.1 Summary of nif genes and their function.

Gene	Identity and/or role of gene product
Regulatory	
*nif*A	σ^{54}-dependant transcription initiator of *nif* genes
*nif*L	*nif* gene negative regulator, forms inhibitory complex with *nif*A in response to oxygen and nitrogen status
Structural	
*nif*H	Forms the homodimer Fe protein, of the nitrgoenase. NifH is an obligate electron donor to the MoFe protein and is required for FeMo-co biosynthesis as well as apo-MoFe protein maturation.
*nif*D	α chain of the MoFe protein, an α2β2 tetramer. Substrate reduction takes place within the α unit
*nif*K	β chain of an α2β2 tetramer MoFe protein. P clusters are present at each αβ interfaces.
*nif*J	Pyruvate oxidoreductase as an electron donor to the Fe protein
*nif*F	Flaxodoxin, electron donor to the Fe protein
Cluster biosynthesis	
*nif*Q	Involved in FeMo-cofactor biosynthesis by incorporating Mo.
*nif*B	Involved in FeMO-co biosynthesis. Cofactor acts as a specific donor to Fe and S
*nif*N	Involved in FeMO-co biosynthesis. β chain of an α2β2 tetramer scaffolding complex
*nif*E	As above
*nif*V	Homocitrate synthase, exact role remains unknown
*nif*S	Cysteine desulphurase. Has role in FeS cluster biosynthesis and repair
*nif*Z	Involved in MoFe protein maturation
*nif*U	Involved in FeS cluster biosynthesis. My have redox capabilities involving releasing Fe for FeS biosynthesis
*nif*X	Involved in FeMo-cofactor biosynthesis. Not essential for nitrogen fixation
*nif*T	Involved in FeMo-co biosynthesis, exact role remains uknown.
Processing	
*nif*M	Accessory protein for *nif*H
*nif*Y	Chaperone of the apo-MoFe protein
Other	
*nif*W	Exact role is unknown, potentially involved in regulation of MoFe protein.

The formation, maintenance and functioning of the nitrogenase poses a huge metabolic burden on the cell. One mole of dinitrogen is converted to two moles of ammonia at the expense of 16 moles of ATP [26]. This energy allows electrons accepted from ferredoxins or flavodoxins by [4FE-4S] to be passed to the P cluster and then on to the FeMo-co where the triple bond between the two nitrogen atoms is reduced (Fig. 1 b) [19]. The catalytic turnover of the enzyme is slow compared to other enzymes; each electron transfer sees the proteins making up the enzyme complex disassociate and associate [19]. As a result, nitrogenase is very highly expressed when synthesised. It can account for up to 20% - 40% of cell protein, with NifH alone making up 10% of cell protein [26,39,46]. The metabolic cost of nitrogen fixation, in conjunction with severe oxygen sensitivity of the metal ion clusters, means that expression of the *nif* genes is tightly regulated.

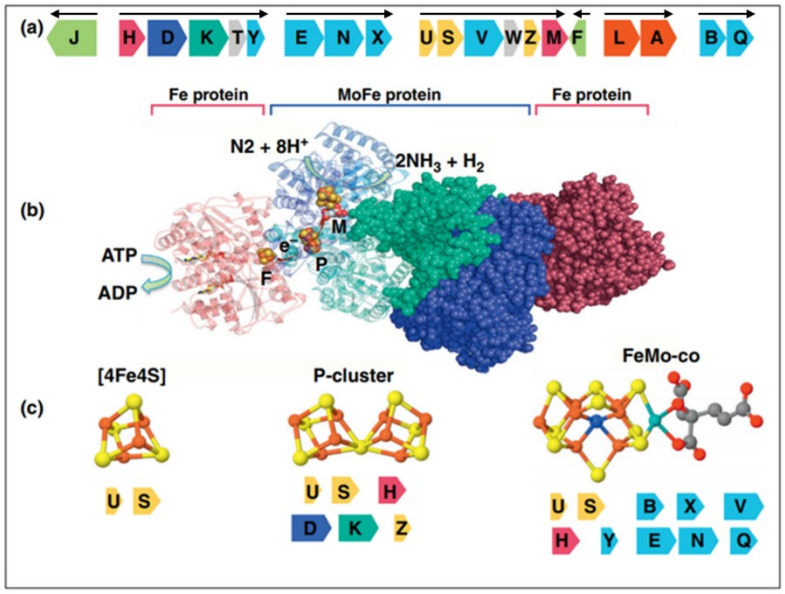


Figure 1.1 The nif regulon and nitrogenase structure. (a) The black arrows above the genes represent the seven operons of the nif regulon. (b) Crystal structure of nitrogenase with illustration of electron transfer between metal ion clusters. (c) Metal ion clusters and the genes required for synthesis. Adapted from Oldroyd and Dixon (2014) (19).

1.2.2 Regulation of the nif cluster in *K. oxytoca*

Nitrogen metabolism in bacteria is well studied and is regulated by three key enzymes; glutamine synthetase (GS), glutamate-oxoglutarate aminotransferase or glutamate synthase (GOGAT) and glutamate dehydrogenase (GDH) [47]. Additionally, there are two key metabolites that regulate these enzymes and signal nitrogen status to the cell; glutamine and α-ketoglutarate. Like many bacteria, ammonium is the preferred nitrogen source of *K. oxytoca* and GS and GDH assimilate most of the ammonium by the cell, although their capabilities are not equivalent [48].

When GS assimilates ammonium through ammonium amination to glutamine, GOGAT catalyses the formation of glutamate (Fig. 1.2) [49]. In the GOGAT reaction glutamine is used to aminate α-ketoglutarate to glutamate ready for use in downstream processes [50]. The GS-GOGAT pathway is

energy intensive and in *Escherichia coli* has been shown to use up to 15% of the cells ATP capacity [47,48]. Alternatively, ammonium can be assimilated by GDH which is much less energy intensive. Here, α-ketoglutarate is aminated to glutamate at the cost of 1 molecule of NADPH (Fig. 1.2) [26,47]. Typically during nitrogen depleted conditions, ammonium is assimilated by the GS-GOGAT pathway, as GS has a much lower K_m for ammonium compared to GDH [47]. It is this process that signals nitrogen status to the bacteria due to the changing ratio of glutamine to α-ketoglutarate under nitrogen limited conditions. As ammonium runs out, GS assimilates more ammonium resulting in increased intracellular levels of glutamine, this signals to the cell that ammonium or nitrogen is becoming limited and sets off a series of regulatory cascades known at the Ntr regulatory cascade.

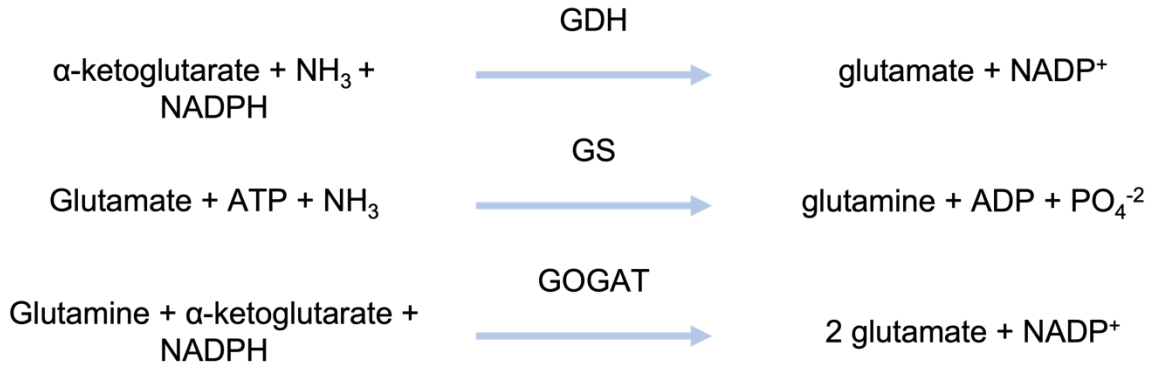


Figure 1.2 Ammonium assimilation and glutamate production in K. oxytoca.

In *K. oxytoca* when the cell becomes nitrogen depleted, the Ntr regulatory cascade is up-regulated and the following occurs. Glutamine levels are sensed by the bifunctional uridylyl transferase-uridylyl removase (UT/UR). UT/UR uridylates PII-type protein, GlnB, which alters its interaction with the target protein, NtrB [26,49]. This means that NtrB can switch from phosphatase to kinase via autophosphorylation. NtrB phosphorylates NtrC, a bacterial enhancer binding protein (bEBP) which in turn regulates *glnK* (from the *glnK-amtB* operon) as well as the rest of the *ntr* regulon and upregulation of GlnA via the NtrC activated PII promoter. The *ntr* regulon is comprised of approximately 75 genes most of which relate to nitrogen scavenging transport systems. For example, *amtB* a high affinity ammonium transporter that is regulated by GlnK [51]. PII (*glnB*) and GlnK (*glnK*) regulate nitrogen assimilation at the transcriptional level in addition to the enzymatic level [49] (Fig. 1.3 A). The *ntr* regulon also encompasses the master regulator of the *nif regulon*, and thus control of nitrogen fixation depends on *nifA* and the co-expressed *nifL*.

As discussed earlier the nitrogenase enzyme is oxygen sensitive and NifL plays a key role in reducing any potential oxygen damage of the metal ions. This is achieved by NifL forming an inhibitory complex

with NifA when the FAD cofactor senses oxygen. In the absence of oxygen, FAD is reduced to $FADH_2$ resulting in a conformational change, allowing NifA expression. NifL is then sequestered to the membrane (Fig. 1.3 b) [26].

Therefore, in low nitrogen and oxygen levels, UT/UR is responsible for uridylating GlnK which prevents it from interacting with AmtB, increasing cytoplasmic GlnK levels. GlnK then interacts with and disrupts the NifLA complex, facilitated by $FADH_2$ and low oxygen levels, resulting in an active NifA . The expression of an active form of *nifA* then drives the other 6 *nif* operons forming a functioning nitrogenase as outlined above. As and when external nitrogen levels increase, GlnK becomes de-uridylated via UT/UR and again becomes subject to membrane sequestration by AmtB. The reduced levels of cytoplasmic GlnK limits the capacity for disruption of the NifLA complex, as such reducing NifA expression and the rest of the *nif* operon as a result (Fig. 1.3) [26,47].

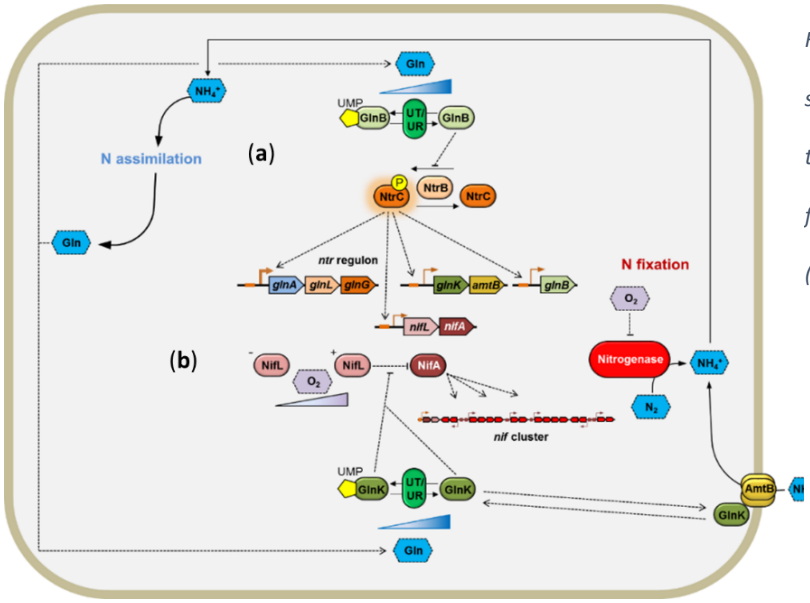


Figure 1.3 Regulation of nitrogenase synthesis. This includes pre (a) and post transcriptional (b) regulation. Adapted from Dr Chris Waite presentation (Imperial College London, unpublished).

1.3 Current studies of nitrogen fixation

There are a range of different techniques employed to study nitrogen fixation and its effect on plant growth and agricultural practices [9,30,52–54]. The work carried out, while having the same goal, also employs a variety of approaches from working with just the *nif* cluster [46], to the bacteria themselves [55] and finally the plants [53]. One interesting problem that all these studies have to overcome is how to measure the 'success' of their efforts, as there is no universal best practice of measuring nitrogen fixation.

1.3.1 Reporting and estimating nitrogen fixation

Nitrogen fixation has proven difficult to quantify directly. Fixed nitrogen is rapidly incorporated into a range of metabolites via GDH and GS-GOGAT and since total nitrogen quantification requires cell lysis it is poorly suited to long term experiments that require cell viability over a prolonged period [56]. The most common technique to measure rates of nitrogen fixation is acetylene. The acetylene reduction assay (ARA) exploits the ability of the nitrogenase to reduce triple bonds somewhat non-discriminately, to convert acetylene to ethylene [25,56]. Both gases are measurable using gas chromatography, and if in a properly sealed container can be measured repeatedly. One drawback of this approach is the need to convert acetylene reduction rates to nitrogen reduction rates since they are not equivalent. A factor of 3 is commonly used but there is a range of conversion rates dependant on species of bacteria and type of nitrogenase among other things [45]. However, this technique still provides a good proximation of nitrogen fixation and is widely used throughout the literature to measure fixation.

1.3.2 Synthetic biology and BNF

While BNF by bacteria has received much attention, more so recently as a means to abate the impact of inorganic fertiliser, the current best example of crop growth enhancement by bacteria is via symbiosis [21,57]. However, only certain species of crop, namely legumes, have the capacity to produce root nodules and as expected of symbiosis only select bacteria can colonise a given species root nodule. While legumes are important, arguably no crop is more important on the global stage than cereals, as they provide more food than any other type of crop [21]. To illustrate this point, over 50% of global caloric intake comes directly from rice and wheat [58].

As a result of their importance there has been significant invest in improving cereal yield through BNF. One approach has been to replicate the legume symbiotic relationship in cereal crops [9]. Alternative approaches include inserting the *nif* genes directly into cereal crops, enabling them to make their own nitrogenase and fix their own nitrogen [42]. Both these approaches would see a reduction in inorganic

fertiliser use but are in the early stages of development so would greatly benefit from a greater understanding of BNF in both symbiotic systems as well as free living nitrogen fixers. Yet, exactly how free-living nitrogen fixers interact with plant roots in the rhizosphere – the area at or near the root surfaces – remains poorly studied.

Other synthetic biology groups are focussing on refactoring the *nif* cluster, an approach that aims to remove native regulation of the *nif* genes by rewriting the DNA sequence while retaining the same function to make the genes portable between different hosts [30]. The first paper to address this, from the Voigt group (38), achieved 26% of wild-type (WT) nitrogenase activity. A later paper from the same group saw this improve to 57% after they optimised the clustering using combinatorial design and high-throughput parallel cloning [46]. Wang *et al.,* [55] removed NifLA and expressed the *nif* cluster through a T7 RNA polymerase-LacI system and achieved 42% of a heterologously expressed *nif* regulon in *E. coli*. The key advantage of removing native regulation of the *nif* cluster is that a major barrier to transferring this operon to plants is removed [54]. Further, you can look at creating a biofertilizer capable of inducible nitrogen fixation. It has been suggested that the refactoring and plant-based approaches would benefit from a greater understanding of *nif* protein stoichiometry, maturity and abundance [42].

Another more 'direct' approach is to engineer bacteria that excrete ammonium into the environment. This could be done in several ways discussed in detail in [59], in short though these can be summarised by disrupting GS or GOGAT activity, forcing *nif* gene expression, or stopping *nif* switch-off mechanisms. There are several examples of such approaches in the literature, yet the arguably the most promising results come from deletion of the *nifL* gene, which inhibits *nifA* activity, in *Azotobacter vinelandii* [60]. The trade-off to the increased ammonium excretion comes in the form of a growth penalty, potentially due to the increased energetic demands of constitutive expression of high levels of nitrogenase. Although these approaches have yielded some growth benefits in cereal crops in ideal experimental conditions, including in wheat [61,62],whether these results are replicable in real world field conditions remains to be seen.

Additionally, the approaches outlined above have a narrow focus, i.e. just ammonium excretion and the direct path to maximise this. At the bacterial level this appears unviable and focussing purely on one compound will miss much of the intricacies that make the legume-rhizobia symbiosis a success, thus limiting our ability to reproduce it. In other words, the wider role of metabolism in fixation is little understood, and by viewing the *nif* genes in isolation you could be missing other key parts to bacteria plant interaction and 'communication', not least a full appreciation of the nitrogen to carbon balance

that operates for growth and replication of bacteria. A more holistic approach that considers wider metabolic changes could aid in synthetic biological approaches by highlighting potentially important, previously overlooked, metabolic processes that are involved in fixation.

1.3.3 Other biological approaches

Much attention has been invested into understanding the biological processes that regulate nitrogen fixation in diazotrophs. One of the underlying themes of research into nitrogen fixation is to 'bridge the gap' between BNF and inorganic fertilisers such that a biofertilizer could provide sufficient nitrogen to support crop growth. As such, there has been limited research into ammonium release of free living diazotrophs [25], although two promising mutants, a *nifL* knock out and a partial GS knockout, have been reported to show increased ammonia secretion relative to WT [63,64].

However, accurate ammonium quantification data is lacking with no clear standardised method employed throughout the field. In fact, previous work found that some commercial kits are unusable for laboratory research in microbiology, because of interfering effects from other nitrogen sources (Carey, MRes thesis, 2018). It is unknown what other nitrogen sources may be excreted by bacteria during fixation even in defined laboratory media, and in complex environments such as the rhizosphere, many N sources will be present. Robust analytical methods are needed to measure both organic and inorganic N compounds; mass spectrometry (MS) is one such method that can report on ammonium levels as well as other nitrogen sources in a single run. MS can also be used to track ^{15}N label, to identify sources and dynamics of nitrogen metabolism. Furthermore, given the high energy demands of fixation it follows that quality and type of carbon source are important to nitrogenase function and thus ammonia production. Indeed, early work did show that *Clostridium pasteurianum* was unable to fix nitrogen unless supplied with sufficient carbon [65]. Yet, a recent review argued that this still warrants further investigation [25].

1.3.4 Nitrogen fixing biofilms

A recent development in BNF research, which could have more immediately transferable agricultural applications is nitrogen fixing biofilms [52,66]. A biofilm is a three dimensional structure made up of an accumulation of permanently attached bacteria to an interface, enclosed by a matrix of extracellular polymeric substances (EPS), which is self-produced [67]. Some studies estimate that up to 90% of a biofilm can be made of EPS [68]. Biofilms have received considerable attention due to the health risk they pose, notably when microbial communities colonize wounds, or surgical implants. In these situations, biofilms gain a fitness advantage relative to 'lone' bacteria due to antibiotic resistance,

where the inner layers are protected by the bacteria on the outer surface of the biofilm, in addition to changes in the physiology of cells in biofilms [67,68]. Recently, there has been a shift in scientific focus, from regarding biofilms as just a problem, towards understanding this fundamentally different physiological state of bacteria and its possible beneficial applications.

Wang *et al.,* [66] was the first to show that some nitrogen fixing bacteria are able to form biofilms given the right environmental conditions. Their research found the outer layers of a biofilm form an oxygen barrier, creating the microaerobic, low oxygen, environment necessary for nitrogen fixation. They went on to show differing morphologies between bacteria on the inner and outer layers of the biofilm resulting from different environments and metabolic capabilities [66]. When coupled with previous work that illustrated biofilms facilitating nutrient retention and acquisition, via extracellular enzymes as well as communication within the biofilm community through extracellular DNA, biofilms present an exciting opportunity for biotechnical solutions to the nitrogen crisis. Although the research is in a nascent stage, there are already some examples of biotechnological approaches successfully utilizing the nitrogen fixing biofilm. Recent research used genetically engineered rice plants to produce root exudates that promoted biofilm formation and maintenance, resulting in improved BNF in rice [54]. In that study, *Gluconacetobacter diazotrophicus* was shown to form nitrogen fixing biofilms. However, *K. oxytoca* have also been shown to form biofilms in air, in currently unpublished work carried out by the Buck laboratory inspired by the findings of others, and investigations into their ability to fix nitrogen will form part of this thesis.

Something that unites all the current approaches is the need for robust whole system data on the physiological state of the bacteria in the lead up to and during fixation. There are still fundamental gaps in our understanding of free living diazotrophy adaptation, specifically relating to the impact of nitrogenase production on wider genetic regulation and subsequent metabolism within the cell. These insights would be hugely beneficial to both future biotechnical approaches, such as modifying cereal crop roots to fix nitrogen, in addition to intermediary steps such as root biofilm interactions. Combined this knowledge could help move us away from synthetic chemical fertiliser dependence towards a ‘biofertiliser’.

1.4 Considering metabolism and BNF

In addition to the specific changes that occur as diazotrophs transition to fixation, there are changes in global gene expression and metabolism that have been found at the transcript and protein level. These changes are important as they not only inform what is occurring within the cell during fixation but also how a given diazotroph interacts with its environment and subsequently the plant roots. As discussed earlier, the NIF proteins can account for to 20-40% of cell protein, which has a profound effect on global cell function [26].

Studies carried out on *Bradyrhizobium japonicum*, which forms a nitrogen fixing symbiosis with the soybean, found that its transcriptome and proteome differed between symbiotic and free-living conditions [69]. Delmotte *et al.,* [69] found that at least 43% of *B. japonicum's* genome was expressed during symbiosis, including most of the proteins involved in carbon and nitrogen metabolism showing the extent to which BNF alters global cell metabolism, and quite how much of that adaptation might be obligatory or not. A follow up study by the same group looked at the downstream effects of these transcriptional differences finding that the level of amino acids differed between nitrogen fixing and non-nitrogen fixing symbionts [70]. Specifically, levels of the glutamate, glutamine and proline, among others, were found to be in greater abundance in nodule tissues with nitrogen fixing *Bradyrhizobium diazoefficiens*. This further illustrates other pathways and compounds, not just ammonium, are also important to plant bacteria interactions.

Whether these amino acid exchanges occur with free living plant associative rhizobacteria like *K. oxytoca* remains unclear. However, there are examples in the literature of free living species also exhibiting changes in gene expression and that these are variable across the *nif* operon at any one time as well as across time [64]. In *A. vinelandii* peak gene expression was seen at different times for different genes, inline with their function relating to the nitrogenase. Firstly peak *nifA* expression was seen between 10-30 minutes after nitrogen depletion, followed by genes involved in FeMo-co biosynthesis showing maximum mRNA levels after 1 hour, finally *nifY* reached maximum expression after 2 hours [64]. These relatively rapid changes in cell function led to a staggering 211.4 µM of nitrogenase structural proteins (*nifs H, D and K*) per cell. Given the energetic demands of not only producing the nitrogenase but also the process of fixation itself, the effects this process has on the cell are wide ranging and will effect everything from carbon metabolism to available amino acids for protein synthesis.

In fact, a study by Waite *et al.,* [39] was one of the first to use a multi-omic approach to investigate the onset of nitrogen fixation in *K. oxytoca*. They estimate that the nitrogenase accounts for approximately

40% of cell protein for fixing cells in batch culture, and that fixation influences free amino acid pools within the cell. They were also amongst the first to suggest that fixed nitrogen is a requirement of further *nif* expression and a complete transition to diazotrophic growth [39]. Their work illustrates how global cell metabolism is altered beyond just ammonium production and excretion by fixation in free living diazotrophs, during diazotrophic growth. Further, transcriptomic analysis found that over 50% of the genes were differentially expressed in fixing cells compared to the non-fixing control [39]. While the *nif* genes were amongst the most upregulated, the downstream effects of all gene expression will have a large effect on global cell metabolism including interactions with the environment. Given that it is the products of metabolism that will effect cell function, and ultimately interact with the plant, there is a clear need for approaches to analyse these products to aid in our understanding of the whole system.

Much of the work discussed above relates to either root nodule-bacteria symbiosis or free living diazotrophs in batch culture, how metabolism changes during BNF in biofilms is less well studied. While the effects of general biofilm formation on metabolism have been well researched and large changes in gene expression are seen between growth phenotypes [66,68], there is yet to be a focus how fixation effects metabolism within a biofilm. Of the limited research carried out, flavonoid type has been show to impact biofilm formation in *G. diazotrophicus* [52]. Flavonoids are plant root exudates that aid in plant adaptation to new environments and have been associated with control of the plant-soil interactions [71]. That different flavonoids promote differing degrees of biofilm formation, suggests altered metabolic states during formation. Moreover, biofilm formation was found to modify the root microbiome by facilitating diazotrophic bacteria recruitment. While the mechanism of this is unknown, this shift in structure highlights the effect of metabolism on the environment [52].

There is a general lack of metabolomic profiling data for free living nitrogen fixing bacteria even under standard laboratory conditions, yet one must assume that there would be differences in the metabolic profiles of fixing and non-fixing cells. More specifically, we know of compounds that are vital for nitrogenase formation but lack knowledge of how their metabolism is effected by fixation and/or nitrogen stress. For example, sulphur containing compounds are known to be important in the formation of the nitrogenase, aiding in NifS and NifU function [26], as well as being involved in election transfer through iron-sulphur clusters [72]. Yet assays focussed on different groups of compounds have yet to be used on either symbiotic or free-living nitrogen fixers [73,74]. Nitrogen-containing compounds are often of high biological importance, and given their direct link to fixation, assays that investigate

abroad range of these compounds could add vital depth to how fixation works for free living nitrogen fixers.

1.5 Metabolomics

Metabolomic approaches, which have been employed to great success in other fields, for example medical research [75,76], have been largely underutilised in nitrogen fixation research, especially for free living diazotrophic bacteria. In general, metabolomics is an untargeted technique that quantifies a variety of metabolites – the intermediates and products of metabolism – for a given biological system e.g. bacterial cells, supernatant etc. [77]. However, there are also targeted approaches that allow you to investigate specific groups of metabolites; this approach is generally less common compared to untargeted [78]. Metabolomics can also provide insights into environmental effects on bacterial cell function, as changes in the environment, among other things, can alter metabolite composition [77,79]. This approach also has utility as a phenotyping tool [80], investigating biomarkers [81] and as a tool for discovery of key metabolites involved in important biological processes [82]. This final tool is already being utilised to understand biological processes in important plant species [82], so its application to free living plant associative bacteria is timely.

A key part of studying metabolomics is sample preparation, which is dependent on the analytical technique. When profiling intracellular metabolites, metabolic and enzymatic activity must be stopped rapidly by a process known as quenching. This process is typically achieved by quick freezing or addition of a quenching solution, for example cold methanol or acid but can vary by sample type [83]. Metabolites are extracted from the quenched samples using some combination of cold solvents, for example methanol, methanol/chloroform, or acetonitrile/water [84]. There is no exclusive quenching protocol that can stop metabolic reactions universally and approaches can be optimised to preserve specific metabolites or sample types [85–87]. The current best practice for most Gram negative bacteria is to use cold organic solvents [85], but the specific method should be optimised for each model system, and with a view on which metabolites are of most importance.

Alternatively, one can investigate the intercellular metabolites within a given system. This has the advantage of not requiring a quenching step as only metabolites excreted by the system are analysed. This approach is known as 'metabolic footprinting' and can be used as a phenotyping tool [88]. Metabolic footprinting has been used in many different fields including microbial growth medium for a variety of bacterial species. For example, it has been used to distinguish between *Burkholderia sensu lato* strains, bacteria found in the rhizosphere of cereal crops and known to be a plant growth promoting rhizobacteria (PGPR) [89]. In systems difficult to simplify such as the soil microbiome, this technique could prove useful in looking for specific signatures of soil good for crop growth.

There are two main platforms used in metabolomics, namely Nuclear Magnetic Resonance (NMR) Spectroscopy and Mass Spectrometry (MS). The approach one employs is usually a result of the sensitivity and specificity required as well as practical capabilities, such as ease of sampling or biomass. NMR profiling, is non-destructive and requires minimal sample prep, therefore, it is commonly used for metabolic footprinting as it allows for the identification and quantification of a large range of metabolites typically in the millimolar (mM) range. Conversely, MS is destructive but much more sensitive compared to NMR, able to detect down as low as the femtomolar range rather than millimolar [90]. This gives it great utility for samples of low biomass or investigation of low abundant metabolites. Both platforms are used in this thesis are discussed in detail below.

1.5.1 ^{1}H Nuclear Magnetic Resonance Spectrometry

This technique exploits a phenomenon call nuclear spin, which is determined by the protons and neutrons in the nucleus, an odd number results in isotopes with a magnetic moment, the foundation of NMR analysis. There are a range of isotopes used in NMR, however ^{1}H is the one most frequently used. ^{1}H isotopes have two spin states, 'spin-up' or 'spin down', which have the same energy levels and frequency under normal conditions. However, when a strong magnetic field is applied to a sample, precession occurs about the field, with nuclei in the spin state aligned with the field at a lower energy state, compared to those opposed to the field. The state of precession is dictated by the strength of the magnetic field, which also impacts the sensitivity of the assay; higher magnetic fields results in greater sensitivity.

Next, in a process known as resonance, a radiofrequency pulse is applied to the sample, which excites some of the low energy nuclei to the higher energy state opposed to the magnetic field. The hydrogen atoms chemical environment determines how much energy is required to get them to flip energy states.; electrons that orbit each nuclei in each sample shield the nuclei from the effects of the magnetic field. More isolated, or de-shielded, nuclei will be subjected to the magnetic field to a greater degree and thus require more energy. After this pulse nuclei will naturally return to their lower energy states, known as relaxation. The energy that is released due to the shifting state is detected and subsequently results in NMR spectra. The signal produced during relaxation is called the 'Free Induction Decay' (FID), with each nuclei resonating at its own frequency contributing to the overall signal. Overtime the signal 'decays' and this combined with the signal intensity is Fourier-transformed to produce the NMR spectra we are familiar with and is used to determine metabolite identification and abundance.

Each peak represents a hydrogen atom in a specific chemical environment, adjacent nuclei can cause peaks to split into doublets, triplets or multiplets, providing key information on chemical structure. The position on the spectra, as well as splitting information allows for identification compared to known libraries of compounds, either using commercial software or publicly available databases. The concentration of a given metabolite is calculated using the integrals of all peaks associated to a given metabolite.

1.5.2 Liquid Chromatography Mass Spectrometry

While not essential, it is common to couple MS with a technique to sperate out metabolites [90]. The work carried out in this thesis uses liquid chromatography (LC), which separates out metabolites in mixed samples so that they enter the mass spectrometer sequentially. LC takes advantage of differing chemical properties of metabolites and how they interact with two different phases. The first phase is the 'stationary phase' which is a column into which the sample is injected, along with a solvent or 'mobile phase'. When a mixed sample is injected into the column, each metabolite will travel through at different speeds based on its affinity to each phase, therefore separating them in the process. The time take for a metabolite to travel through the column is known as the retention time (RT) and contributes to metabolite identification. There are different forms of LC that can affect the RT of a metabolite, these being High Performance Liquid Chromatography (HPLC) and Ultra Performance Liquid Chromatography (UPLC). HPLC, uses higher solvent pressure to reduce the time taken separate metabolites. UPLC, the technique employed in this work, advances on HPLC by using even higher pressures in addition to smaller, more tightly packed column particles. This facilitates shortened runs with increased peak resolution for quantification, as well as lower limits of detection for low abundance metabolites.

Once metabolites have been divided, they need to be ionised and converted into gas phase ions so that they can be analysed by the mass spectrometer. This is another point in the process where a variety of options are available to choose from, for example; matrix assisted laser desorption ionisation (MALDI), electron ionisation (EI) and atmospheric pressure chemical ionisation (APCI). Here, electrospray ionisation (ESI) was employed, where the liquid sample passes through a metal capillary which is subject to a high voltage, and a potential difference is created between the capillary and the opening to the mass spectrometer [91,92]. As the liquid sample passes through the tip of the capillary it is nebulised and ionised by the strong magnetic field, aided by a sheath gas that surrounds the capillary (Fig. 1.4). The charged, nebulized particles pass through a heated capillary to the mass spectrometer, removing the droplet leaving only the charge ions. ESI has the advantage of avoiding unwanted

fragmentation compared to other methods such as EI where this can be the case [92]. The main disadvantage of ESI is that complex samples can suffer from ion suppression – a reduction in detection capacity due to impaired ionisation of metabolites of interest, typically resulting from a matrix effect [78].

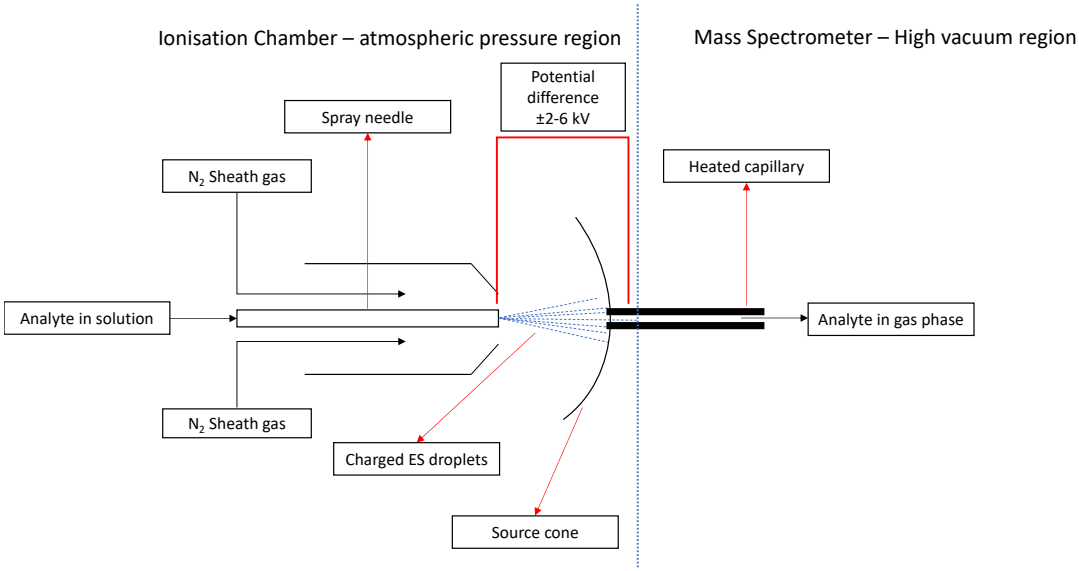


Figure 1.4 Simplified schematic representation of the ESI-ion source. Adapted from Banerjee and Mazumdar (2012) (80).

Once ionised gaseous samples enter the mass spectrometer, ions are separated based on their mass to charge ratio (m/z). Ions are detected using a mass analyser to produce a mass spectra which coupled with the retention time gives you enough information to being data evaluation and metabolite identification. The type of mass analyser that is used will depend on the intended outcome of the research, as one can carry out untargeted or targeted MS.

1.5.3 Untargeted vs Targeted MS

Untargeted MS, sometimes referred to as global profiling or simply untargeted metabolomics, is like NMR in that it indiscriminately detects measurable metabolites present in the sample, including unknowns. This is usually achieved by coupling the LC system to a quadrupole time-of-flight (QTOF) or orbitrap mass analyser. Given the complexity of most sample types, and the comprehensive nature of untargeted MS, a variety of techniques must be employed to help make sense of the data. This starts with the instrument itself with both QTOF and orbitrap mass analysers producing accurate masses that are highly resolved with additional option to collect MS2 spectra (Ion fragments already analysed by MS1 are then further fragmented and analysed by MS2) [90,91,93]. Additionally, multivariate analysis is

employed to reduce the data to smaller more tangible signals which coupled with the MS2 data from the mass analyser, are annotated through use of reference libraries or experimentation [75,94]. While advantageous to have as much information as possible about your sample this approach does come with drawbacks. Chief amongst these is the time, computational power, and analytical techniques to process these data and identify metabolites. Additionally, there is the potential of biasing your output towards highly abundant compounds.

Alternatively, one can employ targeted MS, where only known chemically characterised metabolites are measured. This is carried out on triple quad (QqQ) mass analysers, which lack the m/z range, mass accuracy and resolution of QTOF and orbitrap analysers, but are used as the gold standard to detect and quantify known metabolites in clinical settings [93]. As ionised metabolites enter the mass spectrometer the three quadrupoles of the QqQ perform different functions to facilitate detection. The first quadrupole (Q1) allows a given known precursor to pass isolating it from the others, next Q2 acts as a collision cell fragmenting the target ion in a process known as collision induced dissociation (CID). Finally, the fragmented ion particles then move through Q3, which filters for fragments of a known mass to hit the detector. Each metabolite is experimentally tested to produce a known precursor and product(s) ion, known as the transition, which coupled with the retention time provides a high degree of surety in metabolite identification. This is highly advantageous when looking for novel associations between metabolites, as analytical artifacts are not carried to downstream analysis possibly confounding results [78]. Unfortunately, all metabolites of interest must have their precursor and product ions worked out experimentally prior to starting any study which is a time-consuming process.

Whether to use an untargeted or targeted approach depends on the question you are trying to answer. It is perhaps useful to divide the possible types of questions into two categories: biomarkers and understanding. Untargeted MS is most powerful when you are looking for biomarkers, for example, looking for a specific metabolite in association with a phenotype e.g. cancer biomarkers in serum [95,96] or a metabolic signatures as biomarkers [97]. These studies are often exploratory in nature and suited to more diagnostic applications as in the case of the breast cancer metabolic signature [97]. However, as stated earlier metabolite identification in untargeted MS is difficult and time consuming. It is only through knowing what metabolites are present and how they change in relation to biological processes, and each other, that we can begin to understand systems. This is where targeted MS has great utility, answering the understanding questions. In tackling the understanding questions targeted MS also allows for practical applications of results, such as, improving synthetic biology approaches in

bacteria plant root interactions [52] or improving host responses to disease [98]. The work in this thesis aims to help better our understanding of nitrogen fixation and its effect on global metabolism in an effort to inform practical solutions to the nitrogen crisis, therefore targeted MS used throughout.

1.5.4 Metabolomic Data analysis

All metabolomic methods generate multivariate datasets, of varying degrees of complexity. The scale of this data, 1000s of datapoints for a single sample, makes interpretation difficult and the use of multivariate statistics is necessary to make sense of it [75]. These techniques allow for the identification of patterns and how these patterns relate to variables of interest. There are two types of multivariate statistics that will be used in this thesis unsupervised and supervised. Unsupervised methods are used to explain variance in the data without prior knowledge of group membership or other metadata. In contrast, supervised techniques make models that explain patterns in the data while taking data structure and metadata into account. Both are powerful approaches and vital to help understand the output of metabolomics experiments.

A benefit of employing targeted MS is that is facilitates the use of univariate statistics. This provides granular insights into differences between treatment groups that should not be overlooked. In this thesis, all of these techniques will be employed and are described in more detail in Chapter 2.

1.6 Free living plant associative diazotrophic bacteria; nitrogen fixation, metabolism, and metabolomics

Free living plant associative nitrogen fixing bacteria are an important part of the rhizosphere, having a significant impact on crop growth and yield, including for commercially important species. While it is known nitrogen fixation clearly has a profound effect on global metabolism in bacteria [26,55,64], translating our knowledge of fixation into viable solutions has proven difficult [19,30,46]. Further, to date there is only two known species of nitrogen fixing biofilms in air and how fixation effects their metabolism is understood even less [52,66]. Given that biofilms are more resistant to outside influence, nitrogen fixing biofilms could present a huge opportunity for synthetic biological intervention.

Targeted metabolomic approaches that address the 'understanding' question could help improve our knowledge of the wider changes in cell metabolism resulting from *nif* expression in batch culture and biofilm alike, complementing the existing transcriptomic and proteomic data [39]. The advantage of metabolomics is that it informs on what is actually happening in biological systems, looking at metabolites specifically and not just changes in the genes that control metabolism. This is useful because regulation of the *nif* cluster and associated genes occurs at the post-translational and post-transcriptional level [26,64]. In addition, approaches such as proteomics may tell us that 40% of cell protein is made up on NIF proteins but it fails to inform us what other changes occur in the cell [39]. Addressing what happens to metabolism at a global level will speak to some of the fundamental aspects of free living diazotrophy still poorly understood. For example, how can a bacteria invest so heavily in nitrogenase development and remain competitive in the rhizosphere under nitrogen limiting conditions. Additionally, the regulatory system governing nitrogen fixation is highly conserved and likely evolved to minimise excretion of fixed nitrogen, whilst ensuring a supply of fixed N for GS to assimilate. How does disturbing this equilibrium in favour of higher levels of excretion, better for plant growth, perturb the cell function and viability?

While there have been some applications of metabolomics to study nitrogen fixing bacteria, this has almost exclusively focussed on nitrogen fixation in symbiosis [74,99–103]. The work carried out in this thesis has already contributed to some of the first metabolomics carried out on free living nitrogen fixing bacteria [39] and to the authors best knowledge there was none available in the public domain at the onset of this work.

Chapter 2 General methods

2.1 Strains

Table 2.1 shows the *K. oxytoca* strain and mutants of this parent strain used throughout this thesis. The original WT was obtained from Ruth Schmitz (Christian Albrechts University, Kiel) and the mutants were created as in Waite *et al.,* 2021 [39], using the Lambda Red recombineering [104] method outlined below. All mutants were maintained using Kanamycin antibiotic at 15ug/ml.

Table 2.1 The strain and mutants derived from this parent strain used in this work.

Strain name	Gene knockout	Summary
Wild type (WT) M5a1	N/A - Amp^r	Has a normal growth phenotype and is able to fix nitrogen
△*nifLA*	△*nifLA* *ΔnifLA::nptII*. Amp^r Kan^r	Unable to fix nitrogen, no diazotrophic growth
△*nifH*	△*nifH* *ΔnifH::nptII*. Amp^r Kan^r	Unable to fix nitrogen, no diazotrophic growth
△*nifK*	△$nifK_{1\text{-}1203}$ *truncation (△nifK)* *nifKΔ1204-1545::nptII*. Amp^r Kan^r	Unable to fix nitrogen; C-terminal domain of NifK necessary for nitrogenase activity removed; *nif* expression otherwise unimpaired
△*amtB*	△*amtB* *ΔamtB::nptII*. Amp^r Kan^r	Able to fix nitrogen and exhibits diazotrophic growth; AmtB is necessary to regulate ammonium levels in the cell

Additionally, *E. coli* NCM 3722 was used to trial analytical techniques, see Brown and Jun [105].

2.2 Generating Mutants

All mutants were created using the Lambda Red recombineering method outline in Datsenko and Wanner [104]. In brief, specific oligonucleotides were designed to amplify a kanamycin resistance cassette (nptII) from the pGEM-T-KanFRT plasmid [106]. The oligonucleotides were designed to have 60 nucleotides homologous to the flanking regions of the gene to be knocked out in the WT strain's genome (Primers seen in Table 2.2). PCR's were carried out using 25 ng template plasmid and an annealing temperature of 55°C. Template DNA was removed by DpnI digestion and the linear PCR product extracted from the agarose gel using Qiagen gel extraction kit (Qiagen, Manchester, UK). Next, the pKD46 plasmid was transformed into the WT strain selecting for Kanamycin resistance, this plasmid expressed the Lambda Red genes required for homologous recombination and is subject to an arabinose-inducible promoter.

Overnight cultures of Red-recombinase cells were grown in super optimal broth (SOB) [0.5% (w/v) yeast extract, 2% (w/v) tryptone, 10 mM NaCl, 2.5 mM KCl, 20 mM MgSO4], and sub-cultured the following morning to 0.1 OD_{600}, at which point L-arabinose was added to a final concentration of 10 mM, cells were further grown to 0.4-0.6 OD_{600}. 1 mL cultures were washed three times in cold sterile 300 mM sucrose solution and resuspended in 20 µL 300 mM sucrose solution and thus concentrated approximately 50x. For the recombination step, approximately 1 µg of purified PCR product was added to 100 µL of electrocompetent cells and incubated on ice for 30 minutes prior to electroporation. Half of the transformed cells were recovered at 30°C for 3 hours before being plated out on Kanamycin supplemented LB plates. The remains cells were recovered overnight at room temperature before being plated in the same way. This allowed for potential failures and increased success rate.

Colonies that were found to be Kanamycin resistant were used to generate overnight cultures from which glycerol stocks were taken and genomic DNA was extracted and purified using a QIAamp DNA kit (Qiagen, Manchester UK). Diagnostic PCR was performed to screen for insertion of nptII at the correct chromosomal locus. Primers specific to the regions flanking the knockout locus were combined with those specific to the nptII cassette to achieve this (See Table 2.2 for list of primers). Finally, Sanger sequencing was carried out by Genewiz (Stortford, UK) for further confirmation. Small scale batch culture growth assays were carried out using a using a CLARIOstar plate reader (BMG LABTECH), in which a microaerobic atmosphere (0.2% O2, 1% CO2, and 98.8% N2) was maintained using an atmospheric control unit and gas cylinders. This was done to check growth rate and where applicable loss of fixation function.

Table 2.2 List of primers and oligonucleotides used in this work.

Plasmid	**Description**	**Reference/Source**
pGEM-T-KanFRT	pGEM-T derivative containing a nptII-FRT selection marker cassette. Ampr Kanr	Zumaquero *et al.,* 2010
pKD46	Lambda Red recombinase expression plasmid carrying nucleotides 31088-33241 of phage Lambda (gam, bet, exo) under the control of Para. Temperature sensitive replication (repA101ts). Ampr . GenBank accession: AY048746	Datsenko and Wanner, 2000
Oligonucleotide	**Sequence (binding region in upper case)**	**Target**
nifLA_mutF	ccgcggcgtccctgtcacggtgtcggacaaattgtcataactgcgacacaggagtttgcg GTGTAGGCTGGAGCTGCTTC	pGEM-T-KanFRT
nifLA_mutR	Acaaattgtcgcaattccgccgcgctggcgacaatgtcctgaatctcacataaggcttca TCCTCCTTAGTTCCTATTCCG	pGEM-T-KanFRT
nifH_mutF	cggaattgccagctggggcgccctctggtaaggttgggaagccctgcaaagtaaactgga TGGCTTTCTTGCCGCCAAGGATC	pGEM-T-KanFRT
nifH_mutR	gcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgc CCGCCTTCTATGAAAGGTTGG	pGEM-T-KanFRT
nifK_mutF	gagctgggctgcgagccaacggtgatcctgagccataacgccaacaaacgctggcaaaaa GTGTAGGCTGGAGCTGCTTC	pGEM-T-KanFRT
nifK_mutR	catactccctcctggccccgcatgacgcggggcacctgatggttaacggacgagatcgaa TCCTCCTTAGTTCCTATTCCG	pGEM-T-KanFRT
amtB_mutF	cggaattgccagctggggcgccctctggtaaggttgggaagccctgcaaagtaaactgga TGGCTTTCTTGCCGCCAAGGATC	pGEM-T-KanFRT
amtB_mutR	gcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgct CCGCCTTCTATGAAAGGTTGG	pGEM-T-KanFRT

2.3 Nitrogen Free Davis and Mingioli Medium

Nitrogen Free Davis & Mingioli (NFDM) medium is a minimal medium widely used in nitrogen fixation research [39,40,60]. The medium is used throughout this thesis and is outlined in Table 2.3. The medium has final concentrations of 69 mM K2HPO4, 25 mM KH2PO4, 0.1 mM Na2MoO4, 90 µM FeSO4, 0.8 mM MgSO4, 2% (w/v) glucose and is supplemented with NH4Cl as a nitrogen source is supplemented with varying amounts of ammonium chloride (NH_4Cl) for either low nitrogen (0.5 mM for batch culture and 1mM for biofilm growth and induces diazotrophy) or high nitrogen (10 mM for batch culture and 20 mM for biofilm growth and does not induce diazotrophy). The concentration of ammonium added for the low and high nitrogen treatments was originally taken from the literature and validated by other research carried out in the Buck laboratory prior to this research [25,39]. Each component is weighed using a microbalance and added to 1 L of demineralised water in a laboratory flask. The medium is then sterilised in the autoclave at 125°C for 20 minutes.

Table 2.3 NFDM recipe.

Component	Concentration (g/L)
$MgSO_4$	0.1
$NaMoO_4.7H_2O$	0.025
$FeSO_4.7H_2O$	0.025
K_2HPO_4	12.06
KH_2PO_4	3.4
Glucose	20

2.4 Microbial growth

Bacteria were plated on Lysogeny Broth (LB) medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) with 1.5 % w/v agar. Each component is weighed using a microbalance and added to 1 L of demineralised water in a 1L laboratory flask. The medium is then sterilised by autoclaving at 125°C for 20 minutes. Individual plates (circular 90 x 16 mm petri plates, VWR or square 120 x 120 x 17 mm, Greiner Bio-One) were then poured in a laminar flow hood and allowed to set. Plates were spread with a sterile L-shaped spreader. Plates were incubated at 30°C overnight before use. For general use, cloning and transformations *E. coli* and *K. oxytoca* strains were cultured in (LB) medium

(10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) at 37°C or 30°C respectively with constant shaking at 200 rpm overnight. Appropriate antibiotics for plasmid selection and maintenance.

2.4.1 Liquid culture experiments

Ammonium run-out was used to reliably induce diazotrophic growth and investigate the effect of nitrogen fixation on cell metabolism. For these experiments, bacteria were cultured overnight from a single colony in NFDM medium supplemented with 20 mM NH_4Cl, to ensure cellular nitrogen status. Cultures were centrifuged at 4000 x g and 4°C, after cell pellets were washed once to remove excess ammonium and resuspended in NFDM medium to 0.1 OD_{600}. NDFM medium was supplemented with either 0.5 mM (low nitrogen) or 10 mM (high nitrogen) NH_4Cl for experimental growth. Cells were grown at 25°C for diazotrophy.

Small scale growth assays were carried out using 200 µL cultures in sterile 96-well plates. OD_{600} was measured using a CLARIOstar plate reader (BMG Labtech) at 10 minute intervals. A microaerobic atmosphere was maintained at 0.2% O_2, 1% CO_2, and 98.8% N_2, using an atmospheric control unit and gas cylinders. For batch culture experiments, 30 mL cultures were sealed in 150 mL Erlenmeyer flasks using Suba-Seal septa (Sigma-Aldrich), chilled on ice to prevent growth and flushed with N_2 gas for approximately 45 minutes. After 45 minutes, cultures were anticipated to be microaerobic and were tested using O2xyDot sensors (Oxysense), that had been fixed inside flasks. Readings of under 10 are known to be microaerobic. After, cultures were gently warmed to 25°C in a water bath, before being incubated at the same temperature and shaken at 200 rpm for the duration of the experiment. All samples were taken using a gas tight syringe with an equivalent volume of oxygen free N_2 gas being injected into the headspace. After the sample was taken, OD_{600} was measured and samples were spun at 13,000 g for 5 minutes, supernatant was taken and aliquoted before freezing and cells were either stored separately or discarded depending on the experiment.

For experiments where the bacteria were cultured aerobically the process is the same as the above apart from the following differences; cultures were not flushed with nitrogen, flasks were stoppered with sponge bungs and sampling was carried out using a pipette.

2.4.2 Biofilm experiments

Biofilms were grown on NFDM medium with 1.5% w/v agarose and supplemented with either 1 mM (low nitrogen) or 20 mM (high nitrogen) NH_4Cl (CAS number: 12125-02-9 Sigma Aldrich, Gillingham UK). The medium was prepared in 50 mL square petri plates (120 x 120 x 17 mm, Greiner Bio-One) for metabolomic analysis or 60 mL sealable jars (External diameter 55 mm, neck internal diameter 30.5

mm, height 49 mm, EP Scientific) for assessment of nitrogen fixation. For both approaches, the lids were loosely placed on to allow for growth of biofilms in aerobic conditions. While trying to grow biofilms anaerobically was trialled, it was ultimately not within the scope of this thesis. The day of Acetylene Reduction Assay (ARA, outlined below), jars were screwed tightly shut to contain the acetylene.

Biofilms were generated from an overnight culture in NFDM medium supplemented with 20 mM NH_4Cl, to ensure abundant cellular nitrogen. Cultures were centrifuged at 4000 x g, with cell pellets washed once to remove excess ammonium and resuspended in NFDM medium to an OD_{600} of 0.1. At this step two different approaches were taken. To grow biofilms from a single colony (single colony biofilms), cultures were further diluted to approximately 20 cells per mL. For the 50 mL square plates, 50 µL of the washed and resuspended culture was added to the medium on the plate and evenly spread across the medium using an sterile L-shaped spreader. For the jars, 25 µL of the washed and resuspended culture was added to the surface of the medium in the jar and evenly spread using a sterile L-shaped spreader. The reduced amount reflected the smaller surface area of the medium the jars relative to the surface areas of the medium in the plates. Plates and jars were incubated at 25°C for up to 7 days.

Alternatively, to grow larger biofilms for analysis, 10 µL of the washed culture diluted to 0.1 OD_{600} was spotted onto sterilised polycarbonate disks placed onto the media. The disks have been shown to not impede growth or nutrient uptake (Buck lab, unpublished) and aids in biofilm transference. Plates and jars were incubated at 25°C for up to 7 days.

2.5 Acetylene Reduction assay

As already discussed, the ARA exploits the ability of the nitrogenase to reduce triple bonds relatively indiscriminately by converting acetylene to ethylene, allowing for an assessment of nitrogen fixation capacity at a given time. While the sampling time and number differs between experiments carried out in this thesis, the general methodology remained constant and is outlined below.

One milliliter pure, oxygen free, acetylene was injected into the headspace of the airtight culture vessels; flasks for batch culture and jars for biofilms. For the batch culture ammonium run out experiments this was timepoint zero. While some papers saturate vessels with acetylene [30,107], here the partial pressure of acetylene used was sub-saturating and approximately equivalent to the nitrogenase's K_m. After the appropriate incubation time for the experiment 500 µL of the headspace

was sampled and via gas tight syringe. The sampled headspace gas was then analysed via gas chromatography through a Hayesep N column (Agilent) at 90°C in nitrogen carrier gas. Peaks of both acetylene and ethylene were detected by flame ionisation at 300°C. The detectors output was collected using Chemstation software, where single peak areas were manually integrated. Pure standards of acetylene and ethylene were injected prior to samples to check the retention time and experiment specific negative controls were used to assess background signals.

2.6 NMR spectroscopy

2.6.1 Sample preparation

Samples supernatants were thawed at room temperature for one hour, vortexed for 10 s and centrifuged at 14,000g for 10 minutes at 4°C. These samples were then prepared to a total of volume of 650 µl using a 7:2:1 ratio of sample to phosphate buffer to IS2 (Chenomx, USA). The phosphate buffer was prepared via combination of 0.5 M NaH_2PO_4 and 0.5M Na_2HPO_4 at a ratio of 39:61, the pH was then checked to ensure the buffer was at pH 7. Each sample was centrifuged a final time for five minutes at 14000 g, after which 600 µl of the supernatant was transferred into a 5 mm NMR tube (Bruker, Rheinstetten, Germany) ready for analysis.

2.6.2 Data acquisition

^{1}H NMR spectra were acquired as described in Beckonert et. Al., [108] with some minor modifications, using a Bruker 600 MHz spectrometer (Bruker, Rheinstetten, Germany). The instrument was calibrated prior to analysis, using Bruker supplied samples, via the following steps. The NMR spectrometer was calibrated to 300K for supernatant analysis using 99.8% methanol sample, as methanol's chemical shift is temperature dependant. Next water suppression was optimised using a 2 mM sucrose (in 90% H_2O and 10% D_2O) sample. This allows for the water peak intensities to be minimised improving resolution of peaks around the water peak. Post calibration, spectra were acquired using a 1D Nuclear Overhauser enhancement spectroscopy (NOSEY) pulse sequence, using eight dummy scans and from 32-64 scans per free induction decay (FID). The modification was to the total recycle time, which was set to 5 s, d1 set to 2.3 s and the acquisition time set to 2.7 s. Spectral width was 20 ppm, receiver gain 90.5 and the transmitter frequency offset (O1) was 2822.15 Hz.

2.6.3 Data processing

Spectra were automatically phased, referenced to TSP at ^{1}H δ0.00 and baseline-corrected in Topspin software (Version 4.0.9, Bruker). Two approaches were used in tandem to facilitate accurate data

processing. Processed NMR spectra were imported into MATLAB (R2021a, MathWorks) and digitised into 32k data points with full resolution using in-house developed scripts [109]. Next, both the water peak (^{1}H δ 4.7–5.14 ppm) and TSP peak (^{1}H δ -0.02–0.02 ppm) were removed from spectra to minimise the impact of baseline disorder due to water suppression. Probabilistic quotient normalization was carried out on the resulting data sets, to account for dilution of complex biological mixtures [110]. Peaks were then integrated and resulting output used for data analysis.

To aid in this analysis and add biological context to the findings, individual metabolites were identified using Chenomx NMR suite 5.0 (Chenomx, Edmonton, Canada). This software has a reference library of standards which allows for the identification of compounds based on the TSP peak.

2.7 Data analysis

2.7.1 Data normalisation

Processed data were normalised before analysis in two main ways. Where possible, data was normalised to biomass as discussed in [39,111]. Alternatively, median fold change normalisation was carried out across all variables, as in [112,113]. Normalisation was carried out in Microsoft Excel for Mac (v 16.62).

2.7.2 General statistical analysis

Univariate analysis and presentation of processed data was carried out in Python (Python 3.8.13) or R Studio (R version 4.2.1). Parametric and non-parametric univariate analysis were used to determine changes in specific metabolites between treatment groups with a p value of <0.01 considered significant. Tests used include T-tests, Mann-Whitney U test, one-way anova, Kruskal-Wallis test. Q values corrected for false discovery rate (FDR) using the Benjamini-Hochberg procedure were also calculated [114].

2.7.3 Multivariate analysis

All multivariate analysis was carried out in Python (Python 3.8.13) or R Studio (R version 4.2.1).

2.7.3.1 Unsupervised analysis

Unsupervised multivariate analysis, which doesn't take into data structure into account, was carried out on all applicable datasets. Principal component analysis (PCA) was carried out as outlined in Wold *et al.,* [115] and first applied to metabolomics by Nicholson *et al.,* [75]. PCA reduces the dimensions of multivariate data down to a set of uncorrelated variables that capture the most variance possible,

these variables are known as principal components (PC) and the number of PC's is decided by the individual. The first PC explains the most amount of variance, with each subsequent component explaining as much of the remaining variance as possible, but always less of the total variance than the previous component. PCA results in two matrices, one a matrix of scores summarising the original variables and a second matrix of loadings which illustrates the influence of each variable. With regards to metabolomics PC's are a vector of metabolite contributions to a given sample. Typically, data is visualised from the scores plot matrix, allowing for the relationship between groups to be investigated any patterns in the data to be seen. The loadings matrix can be used to generate a loadings plot which visualises the contribution of the variables, metabolites in the case of this thesis, to the position of observations in the scores plot.

2.7.3.2 Supervised analysis

A variety of supervised techniques were used in this thesis. The first was partial least squares regression (PLS), which is used to find the relationship between two matrices; the X matrix – typically metabolite values – and the Y matrix – typically the grouping e.g. fixing or not fixing. PLS works similarly to PCA in that it fits a series of components each explaining the most covariance possible between the X and Y matrices, but each component explains less of the total covariance than the previous. PLS is well suited to analyse datasets where the X matrix contains more variables than observations. Additionally, when working with data that has a Y matrix of only two groups PLS-DA (discriminant analysis) is used in place of PLS.

Sometimes with complex samples, there is variation in the X matrix that is uncorrelated, or orthogonal, to Y. Sometimes called structured noise this was removed from the analysis by using orthogonal signal correction (OSC) in conjunction with both PLS and PLS-DA (OPLS and OPLS-DA respectively). OSC works by sorting the variation in the X matrix into two categories, variation linearly related to Y thus contributing to grouping or orthogonal to Y and therefore not contributing to grouping. The orthogonal signal is removed and the regression models are employed as described above, with enhanced model quality allowing for previously masked but important variables to be known.

2.7.3.3 Assessing multivariate model validity

All multivariate models were assessed to check their validity. This was achieved by investigating how well a given model explained the variance in the data, also known as robustness. There are two metrics of interest relating to robustness, these are the R^2 and Q^2 values. The R^2 value is the sum of the R^2X and R^2Y and indicates the total amount of variance explained by the model and will have a value between zero and one. An R^2 value closer to one would indicate more explained variance and as more

components are added, more of the variance is explained resulting in a higher R^2 value. However, this can result in over fitting – where the model generated has poor predictive performance on new unseen data. To avoid over fitting one must also consider the Q^2 value, which is a measure of the predictive ability of the model on novel unseen data and is calculated by cross validation (CV). Overfitting may have occurred when the R^2 value is in the region of 25% larger than the Q^2 value, with some arguing that the difference between the R^2 value and the Q^2 value should; not exceed 0.3.

CV works by splitting the data up into different sets, typically the training dataset is split into 5 sets at random. The model is fitted using all but one of the sets. The values for the excluded set of data are then predicted using the model and cross referenced to real values. This process is repeated so that all sets are left out and predicted one time, with a larger Q^2 indicating that your model has good predictive power. However, it is best to consider the R^2 and Q^2 value in concert, with a small difference between these values indicating a robust model that has a good chance of explaining variation in your biological system.

Permutation testing was also carried out to further test the validity of the model. This technique removes any true relationship between the X and Y matrices by mixing them randomly. For each permutation a new model is calculated with R^2 and Q^2 values, which will assess the likelihood of the model arising by chance. For the original model to be considered useful, its Q^2 value should be distinct and higher from those generated by permutation analysis. Typically a *p* value is generated which indicates the probability of the actual Q^2 being calculated by chance.

Chapter 3 Method development for accurate and reproducible analysis of the nitrogen fixing metabolome

As discussed previously in chapter 1.5.3, there are merits to both targeted and untargeted metabolomic approaches and sampling plays an important role in relation to both approaches. In this chapter I explore both sampling approaches and the development of targeted LCMS assays for cells experiencing nitrogen limited growth and diazotrophy.

The work carried out in this chapter was not intended to be an exhaustive set of sampling and targeted LCMS assays. Rather, it was meant to be building on previous research and applying it to my model diazotrophic bacterial system for which there is little to no research for either on these techniques.

All the projects in this chapter were performed in a collaborative effort as stated where a range of culturing conditions and genetic backgrounds were being used for physiological studies of 53iazotrophy.

The Scanning electron microscopy (SEM) was carried out by Ecaterina Ware (Department of Materials, Imperial College). The LCMS targeted method development was led by Virag Sagi-Kiss with each metabolite included in the methods tested by either Virag Sagi-Kiss, Yufeng Li (Imperial College) or myself. Our combined efforts were used to develop the targeted methods.

3.1 Introduction

Metabolomics has been employed to study metabolism in a wide array of biological systems, answering fundamental biological and biochemical questions alike [116]. The turnover of metabolism can be extremely rapid, in the order of seconds or faster, therefore the snapshots provided by metabolomics provide unique insights in metabolic pathways [117,118]. Given the fast turnover times, the sampling process needs to also be rapid, with metabolic reactions being stopped in the shortest time possible, particularly in microbial cultures. Therefore, the quenching of a sample and the subsequent extraction are highly important parts of metabolomic analysis [87,119]. These approaches can bias the prevalence of metabolite classes and therefore can have huge impact on the data generated. Biofluids such as blood or urine typically have simple sample preparation steps prior to analysis, such as dilution or protein removal. In contrast, metabolite extraction of microbial intracellular metabolites is more complex, and must consider the supernatant as well as cell disruption to access the metabolites [86,118,120].

Sampling of bacteria is well studied yet, the method best suited to accessing the metabolites contained by the cell membrane remains a subject of much debate [79,118,120–123]. Generally, metabolites are extracted using hot or cold solvents (methanol, ethanol, acetonitrile, water, acids, bases), by making the cell membrane permeable enough for the quenching solution to enter the barrier, stop metabolism and recover the metabolites [122]. Some methods include the addition of a bead beating step, which has been shown to increase extraction repeatability [124]. However, there is limited research into the best quenching and extraction methods for nitrogen fixing bacteria such as *K. oxytoca*. Park *et al.,* evaluated extraction methods for *K. oxytoca* finding that three – cold methanol, methanol/chloroform and hot water yielded the most peaks [125]. These cells were not under nitrogen fixing conditions however and the quenching and extraction steps were carried out separately. Given the sensitivity of the nitrogenase and the metabolic reactions associated with it a method that simultaneously quenched and extracted samples would be more optimal for obtained the best snapshot of metabolism.

Post quenching and extraction, the samples are ready to be analysed. As previously stated, the extraction method can bias metabolite composition, so it is important to consider these steps of the experimental process in concert. Further, when trying to improve the understanding of a biological system, as in this thesis, coupling the best sampling method with a targeted metabolomic approach should yield the best insights. As such, alongside evaluation of sampling methods a range of targeted LCMS methods were also developed so that the two steps could properly inform the other.

Reversed-phase (RP) chromatography, is considered the standard LC separation technique as it produces reliable and repeatable results and is compatible with aqueous sample types. The main drawback to this approach is it's poor retention of highly polar metabolites, which are often highly biologically important [90]. To overcome this, several approaches have been developed that build on the foundations of RPLC but either modify the mobile phase or the metabolite in such a way to increase their retention time (RT).

Ion-pairing chromatography (IPC) modifies the mobile phase of RPLC by addition of an amphiphilic ion pair reagent for example, alkylsulfate or alkylsulfonate. The addition of an ion pair reagent alters the RT of ionic metabolites, by forming ion-pairs with metabolites. For example, an anion would have its RT increased by addition of positively charged surfactant molecule and visa-versa. One must be mindful that the ion pair reagent chosen is sufficiently volatile and compatible with the mass spectrometer [126].

Rather than changing the mobile phases one can modify the metabolites themselves through derivatization. There are a number of different derivatization methods available which target a variety of different classes of metabolites [127,127–129]. While many different metabolite classes are of interest in relation to fixation, 6-Aminoquinoloyl-N-hydroxysuccinidimyl carbamate (AccQ-Tag) was chosen as it derivatizes the amino acids along with other amino group containing compounds. The compounds are of course highly relevant to nitrogen fixation but also amino containing compounds are found across a wide range of metabolic pathways, in many different organisms meaning this approach has high applicability to other areas of research also. In fact, a paper by Gray *et al.,* using this method detected and quantified 25 metabolites in human plasma [127].

3.1.1 Aims and objectives:

The aim of this chapter was to test a variety of different sampling methods, while simultaneously developing targeted LCMS assays in order to detect as many biologically relevant metabolites as possible. This was done to develop the best collective targeted metabolomic approach for nitrogen fixing *K. oxytoca*.

Objectives:

- Assess a variety of known quenching and extraction techniques that have been shown to be effective for bacteria species but not *K. oxytoca* under nitrogen fixing conditions.
- Develop and evaluate targeted LCMS methods to be used in combination with each other, namely IPC and AccQ-Tag derivatization, to cover as much of the metabolome as possible.
- Use information gained from each technique to inform the best collective approach for this specific model system.

3.1.2 Design of Extraction Experiment

When sampling for metabolomic assays one can extract only, quench and extract separately, or quench and extract simultaneously. For the purposes of this chapter these three approaches will be grouped and be called sampling methods/approaches/techniques. The success of these sampling approaches is generally assessed by three criteria.

Metabolite similarity – Which metabolites are unique and those that are common between extraction methods.

Reproducibility – A comparison of within-replicate variability.

Yield – Both in terms of the number of metabolites detected for a given sampling method in addition to the amount of any given metabolite.

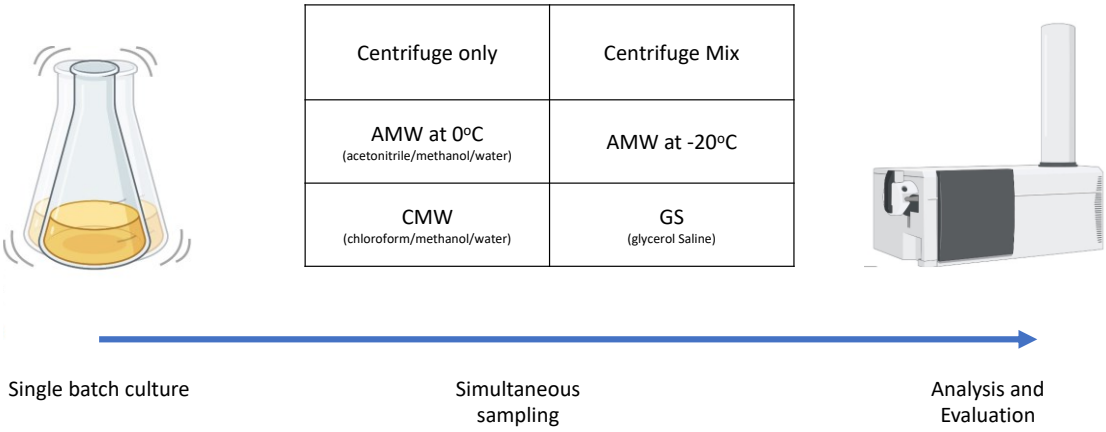


Figure 3.1 500 mL culture grown either aerobically or anaerobically, very initial experiments were carried out aerobically for simplicity. All samples were taken during late exponential growth. Cultures were sampled 'simultaneously' with total sampling time taking less than 5 minutes for all treatments. Full details are outlined below.

Sampling techniques were a mix of extraction, quenching and extraction and simultaneous quenching and extraction. The review of each technique was on going throughout the process and some techniques were dropped earlier than others.

3.2 Methods

3.2.1 Chemicals and Reagents

The basis of these methods was developed using a large library of standards, the mass spectrometry metabolite library of standards (MSMLS) was obtained from IROA Technologies (NJ, USA). Other chemical standards not in the MSMLS library, formic acid (FA), chloroform ($CHCl_3$), acetonitrile (WHich), deuterium oxide (D_2O), tributylamine (TBA), acetylacetone (AAc), acetic acid (HAc), sodium phosphate monobasic and dibasic, and isotopically labelled internal standard, l-phenyl-d_5-alanine, were obtained from Sigma-Aldrich (Gillingham, U.K.). AccQ-Tag Ultra reagent was obtained from Waters UK (Wilsmlow, UK). LC-MS grade water, water with 0.1% FA (v/v), awhichACN with 0.1% FA (v/v) were purchased from Fisher Scientific (Leicester, U.K.). Methanol (MeOH) and isopropanol (IPA, LC-MS grade) were obtained from Honeywell (Charlotte, NC, U.S.A.).

3.2.2 *Quenching*

3.2.2.1 Microbial growth conditions

K. oxytoca were grown in batch culture as outlined in Chapter 2.4.1 for aerobic and anaerobic cultures in NFMD media. The initial testing of sampling methods was carried out using aerobic cultures for ease and to speed up initial tests, setting up anaerobic cultures can take up to 75 minutes longer including sparging time and bringing cultures back to 25°C gradually. Once initial experiments had been carried out and results analysed, subsequent experiments were carried out on anaerobic cultures. Total culture volume was 500 mL and sampling was carried out during late exponential growth. OD_{600} was measured in the lead up to and the time of sampling.

3.2.2.2 Quenching solutions

Three different quenching/extraction solutions were used in this chapter. Acetonitrile-methanol-water (AMW) to a final ratio of 40:40:20 (v/v), methanol-chloroform-water (MCW) to a final ratio of 2:2:2. Finally glycerol-saline (GlySal) made up of pure glycerol and NaCl solution (13.5 g/L) to a 3:2 ratio, a GlySal washing solution was also made up similarly to the GlySal quench solution but to a final ratio of 1:1 (v:v).

3.2.2.3 Sampling techniques

All sampling methods were carried out on a single bacterial culture with all methods and reps (n=3) taking no longer than 3 minutes total time. Methods that were tested are outlined in brief below in Table 3.1.

Table 3.1 An overview of the sampling techniques used in this chapter, to be used a reference. Full methodology for each method is outlined below.

Name	Description
Centrifuge only	5 mL of culture was centrifuged at 13,000g for 5 minutes at ambient temperature. Supernatant removed and stored for later analysis. Pellets extracted using either AMW at -20°C or CMW at ambient temperature. Supernatant and pellet analysed separately.
Centrifuge mix	As above but the extracted pellet is mixed back in with supernatant.
AMW at 0°C	AM + 0.1M Formic acid added to culture for a final ratio of 40:40:20 AMW. A simultaneous quench and extraction.
AMW at -20°C	As above apart from the solvent was at -20°C
MCW	CMW quenching solution at -20°C added to the sample for a final ratio of 2:2:2 CMW.
GlySal at -20°C	Glycerol saline (GlySal) solution made up of pure glycerol and NaCl solution (13.5 g/L) to a 3:2 ratio. To quench 4 parts GlySal solution to 1 part culture. Pellets were stored for later extraction.

Centrifuge only

The pellets generated by the CO method were extracted in two ways outlined below.

For pellet extraction using AMW: 500 uL of AMW solution and approximately 0.2 mL 0.1 mm zirconium beads were added to pellets and mixed thoroughly. The mixture was transferred to a bead beating tube, and the tube containing the pellet was rinsed with a further 0.5 mL of extraction solution. Samples were processed in a Precellys Evolution bead beater (0.1 mm zirconium beads (CK01), Stretton Scientific, Stretton UK) at 6000 rpm for 3 x 20s. Processed samples were transferred to a fresh 2.0 mL centrifuge tube and the bead beating tube was rinsed several times for a final volume of 2 mL in the centrifuge tube. Samples were then centrifuged (16,162 g, 10 minutes at ambient temperature). The supernatant was removed, split into equal 400 µL aliquots and dried overnight at 30 °C using a vacuum concentrator in aqueous mode at 250 g (Rotor: F-45-48-11, Eppendorf, Stevenage UK).

For pellet extraction using MCW the method was adapted from *Belle et al.,* [130] but in short: 0.6 mL of prechilled methanol chloroform solution (-20°C, 2:1 v:v) was added to pellets along with 0.2 mL 0.1 mm zirconium beads (Stretton Scientific, Stretton UK) and mixed thoroughly. Samples were then transferred to a bead beating tube and processed in a Precellys Evolution bead beater (Stretton Scientific, Stretton UK) at 6000 rpm for 3 x 20s. An additional 0.2 mL of water and chloroform were added to samples to separate the phases. The samples were mixed thoroughly for 1 minute and centrifuged (14,800 g for 10 minutes). 300 µL of the upper aqueous layer was removed, split into equal 100 µL aliquots, and dried overnight at 30 °C and dried overnight at 30 °C using a vacuum concentrator in aqueous mode at 250 g (Rotor: F-45-48-11, Eppendorf, Stevenage UK).

Acetonitrile methanol water

Five mL of culture was added to 20 mL of quenching solution for a final ratio of 40:40:20 AMW. Quenching solution was at either 0°C or -20°C. Mixture was vortexed thoroughly and frozen on dry ice. Samples were sonicated while being thawed and centrifuged (14,800 g for 10 minutes). Supernatant was then removed, aliquoted equally and dried overnight at 30 °C using a vacuum concentrator in aqueous mode at 250 g (Rotor: F-45-48-11, Eppendorf, Stevenage UK). Pellets were desiccated and stored at 4°C for scanning electron microscopy (SEM) analysis, to investigate damage to the cell membrane.

Methanol chloroform water

A 2:1 MeOH/$CHCl_3$ solution was prepared, and 2 mL of culture was added to 6 mL of MC solution, at -20°C and mixed thoroughly. Samples were snap frozen and sonicated while thawed. An additional 2 mL of water and chloroform were added, and the sample was mixed again. Samples were centrifuged (13,000 g, 10 minutes) and the upper aqueous layer was removed, split into equal aliquots and dried overnight. Again, Pellets were desiccated and stored at 4°C for SEM analysis.

Glycerol saline

Sampling using this method was carried out as in Villa-Boas *et al.,* (2007) [87]. In brief, 5 mL of culture was added to 20 mL of GlySal solution at -20°C and mixed thoroughly. Samples were acclimatised for approximately 5 minutes before being centrifuged (18,000g, 20 minutes, at -20°C). Supernatants were removed and stored and pellets were washed in the GlySal washing solution, and again centrifuged 18,000g, 20 minutes, at -20°C). Supernatants were removed and pellets were extracted using MC as above.

3.2.2.4 Scanning Electron Microscopy (SEM)

Pellets post extraction were prepared for SEM by resuspension in water before air drying on imaging platform. Images were taken with a FEIHelios DualBeam NanoLab 600 system at 5 kV by Ecaterina Ware Department of Materials, Imperial College London. Images were taken at X 10,000 to X 100,000 magnification. Sampling methods selected for SEM analysis were CO MCW pellets, AMW and MCW.

3.2.2.5 Amino acid spike in experiment

100 uL of an amino acid standard (Sigma Aldrich, Gillingham UK) was spiked into 900 µL of water or NFDM media. The NFDM medium was then either stored for analysis, dried down and reconstituted to the same volume, or dried down and concentrated 10 X. From here samples were derivatised using AccQ-Tag as outlined above.

3.2.3 *UHPLC-MS Method Development*

3.2.3.1 Microbial cultures

E. coli were prepared and cultured as in chapter 2.4. In brief, five 30 mL experimental cultures were established from 5 separate single colonies and grown overnight in LB. The five separate overnight cultures were treated individually and washed twice in LB and initiated at 0.1 OD_{600} at 37°C with constant shaking at 200 rpm. The five washed cultures were used to establish five samples. Sampling was carried out every hour for 8 hours, with 1 mL of culture taken at each timepoint, OD_{600} measured and centrifuged (13,000g, 5 minutes). Supernatant was split into two aliquots of 400 µL and stored for further analysis and the pellet was discarded. Quality control (QC) samples were generated by pooling equal volume of all samples.

3.2.3.2 Developing the metabolite library

The process for all metabolites in the MSMLS library was similar for all the LC methods. First, the library was reviewed and metabolites that were mislabelled and duplicated were removed. Metabolite standards were made in up in water or a mix of water-methanol to a final concentration of 10 ug/mL and stored at -80°C.

Prior information on a given metabolite determined the next steps. When little information was known, metabolite standards were diluted to 1 mg/mL using water, and direct infusion was carried out, explained in more detail below. Alternatively, where possible the parent and fragment ions, and

CE, were taken from the XCMS-MRM database. This facilitated mixtures of 12 compounds (which had different masses at unit resolution) to be pooled and run to verify the RT, parent ion and fragment ion (or RT/Q1/Q3), data which is used for compound identification. The collision energy (CE) values taken from the database were converted to an approximate value relevant to the XEVO-TQS. This conversion was calculated using experimental CE values for standards compared to database values. Optimal CE was determined by ramping around the predicted value in increments of 2-5 V until the best was found. Where relevant prior experimental data were considered. For example, positive molecular ions in the XCMS-MRM database were only tested with positive mode RPLC and for the amine group containing compounds the AccQ-Tag derivatization method. An example workflow of a single metabolite can be seen in Figure 3.2

In instances where the above experimental steps failed to collect RT/Q1/Q3 data, standards were directly infused into the MS and the vendor built in optimisation process was used to determine the best values for the CE/Q1/Q3. Where the software produced data for these compounds the RT was determined in a follow up LC run.

Standards that were applicable for AccQ-Tag derivatisation underwent the same process as above. The observed parent ion was $[M+171]^+$ for monoamines and CE was optimised for the highest abundance fragment ion (171.05). Compounds with two or more amine groups behaved differently. As such, the maximum number of AccQ-Tag additions and charges was used to define the parent ion but other derivatization products were also recorded and the relevant RT/Q1/Q3 values added to the database to assist annotation in real samples by helping to identify potential interferences [131].

3.2.3.3 Derivatisation using AccQ-Tag Ultra Kit

Standards and samples were derivatised as per instructions of the AccQ-Tag Ultra Kit (Waters UK Ltd. Wilmslow, UK). In the brief, 10 µL of sample/standard were mixed with 70 µL borate buffer and 20 µL AccQ-Tag reagent and incubated at room temperature for one minute. Samples/standards were heated to 55 °C for 10 minutes to degrade the excess AccQ-Tag reagent. Dilutions were carried out as appropriate and this was typically between 10 and 100 times. Dilution was carried out if the chromatographic peaks were observed to saturate the detector.

3.2.3.4 UHPLC-MS settings

Analysis was carried out via 5 µL injection of sample with UHPLC separations performed on 2.1. mm i.d. columns of either 100 (RPLC, IPC) or 150 mm (AccQ-Tag RPLC) in length packed with the 1.8 µm

C18 bonded HSS T3 stationary phase (Waters). Mobile phases were method dependant, (A) 0.1% FA in H_2O (AccQ-Tag, RPLC) or TBA:HAc:AAc in H_2O (IPC) and (B) 0.1% WHichin ACN (AccQ-Tag, RP) or MeOH:IPA (IPC), using flow rates of 0.4 (RPLC, IPC) or 0.6 mL/min (AccQ-Tag RPLC). The gradient conditions were optimised for each method and can be seen in Sagi-Kiss et.al. [129]. Each metabolite had a 30s window either side of the measured RT, for a total window of 60s. All methods used a set-up of a binary solvent manager, sample manager, and column manager (Waters, Milford, MA, U.S.A.) interfaced to a Xevo TQ-S tandem quadrupole mass spectrometer (Waters. Corp). A dedicated instrument was used for IPC, as otherwise the ion pairing molecules cause contamination in other samples. See Appendix 1 for other instrument conditions including; ionization conditions and gas flows. The analysis times for the RPLC, IPC, and AccQ-Tag RPLC methods were 14.5, 21.0, and 13.0 min, respectively.

3.2.3.5 UHPLC-MS Data Processing

Collected data was processed via manual integration using Skyline, a freeware package. RT, ion ratios, peak shape comparison between samples and authentic standards plus signal in the blanks were all used to assign metabolites. QC samples with standard spike in were also used in contentious cases to aid in peak annotation. Relative standard deviations (RSDs) were calculated using the QC samples. Data analysis and visualisation was carried out as described in chapter 2.7.

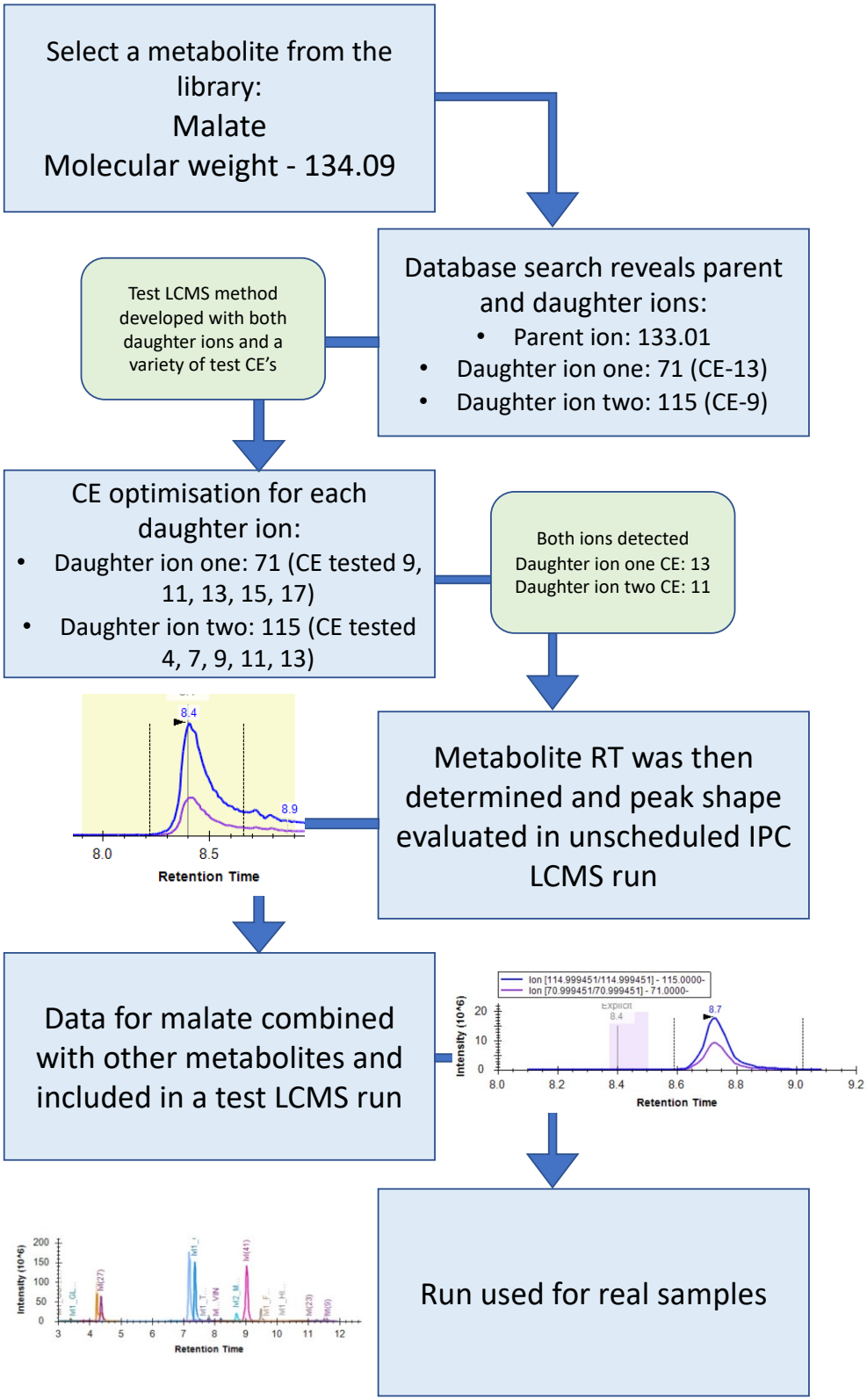


Figure 3.2 Workflow of a single metabolite during library development, steps outlined in brief.

3.3 Results

3.3.1 3.3.1 UHPLC-MS Results

3.3.1.1 Summary of separation methods

Of an initial metabolite library of 350 compounds, 111 amine metabolites were tested by AccQ-Tag derivatisation (commercial kit with 6-aminoquinoloyl-N-hydroxysuccinidimyl carbamate derivatisation). These 111 metabolites were found to spread well across the whole of the chromatogram and have generally good Gaussian peak shape. This resulted in biologically important pairs of metabolites such as leucine and isoleucine being separated and well resolved. RPLC was used as the 'control' chromatography method and 102 metabolites were detected by both the AccQ-Tag RPLC and RPLC. Figure 3.3 A compares the RT of metabolites between these methods and shows that for all metabolites RT was improved by AccQ-Tag derivatisation. Further, 74 metabolites which had a RT of less than one minute in the RPLC method, unacceptable for use in a targeted method, had their RT improved by AccQ-Tag RPLC such that they could be included in the method.

However, not all metabolites had their RT improved by AccQ-Tag RPLC sufficiently to be included in the final method. These compounds were the sugar amines and related compounds. These metabolites were not only poorly retained but also produced multiple peaks, which made downstream analysis steps difficult. As such, they were excluded from the AccQ-Tag RPLC method.

The IPC method also resulted in general improvement in RT compared to RPLC (Fig. 3.3 B), with RT data collected for 283 compounds from the initial 350 metabolite library. In contrast to the AccQ-Tag RPLC method, peak shape was found to be more problematic as some peaks were less Gaussian with broad or asymmetrical profiles. Additionally, there were several metabolites where RT data was unable to be collected, due to low sensitivity or especially poor chromatographic performance. RT data for all metabolites can be found in (116).

Figure 3.3 highlights that large number of metabolites that were poorly (RT > 1 min) or borderline poorly (RT > 1.5 min) retained by RPLC compared to the other methods. Only three metabolites, had better RT using RLPC compared to IPC, however, two of the three had their RT improved by AccQ-Tag RPLC. Of the 283 metabolites for which RT data was collected for IPC, 60 were poorly retained and 14 were borderline poorly retained. Some of the metabolites that fall into these categories had improved RT using the AccQ-Tag RPLC and as such were assigned to that method. In total, there were 57 metabolites for which RT data was collected for both IPC and AccQ-Tag RPLC. Overall, a weak

association was seen between the methods (r^2=0.12), which made metabolite assignment easier and is beneficial when combining methods.

Overall, the combined chromatography methods of IPC and AccQ-Tag RPLC can detect and determine 287 metabolites. This represents a substantial increase from 40% coverage using RPLC alone to 90% for when using the methods in combination.

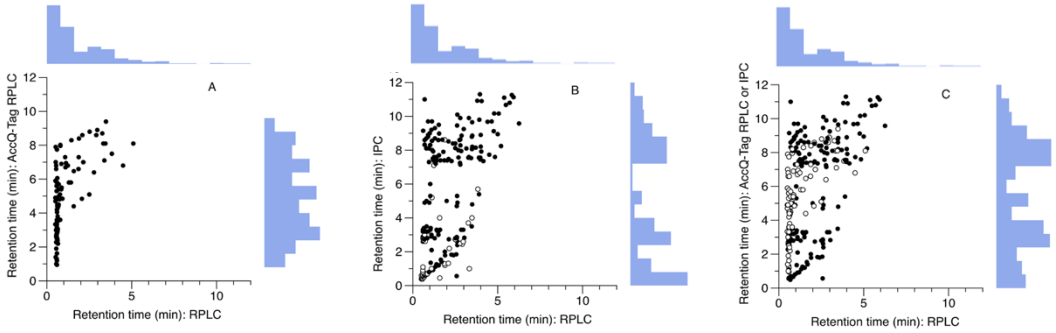


Figure 3.3 Comparison of RTs for each chromatographic method. RT is improved by AccQ-Tag and IPC with the combination of both methods providing a good coverage of metabolites with suitable RT's to be included in final method. (A) RPLC compared to AccQ-Tag RPLC. (B) RPLC compared to IPC. Compounds which were also detected by AccQ-Tag RPLC are shown by open symbols. (C) RP compared to IPC (filled symbols) and AccQ-Tag (open symbols) as a combined strategy; where a metabolite can be analysed by either technique, the AccQ-Tag data are shown. Histograms show distribution of retention time data [129].

3.3.2 3.3.2 Quenching and Sampling results

3.3.2.1 Evaluation of AccQ-Tag RPLC using biological samples

The AccQ-Tag RPLC method was tested using a set of real biological samples using the common bacterial model organism, *E. coli*. Figure 3.4 shows that *E. coli* growth on LB was normal and comparable to previous studies. Supernatant samples were taken each hour for eight hours and analysed using AccQ-Tag RPLC-MS, serving as a simple test run for the biological system of interest. This approach allowed for monitoring of metabolite levels across time, which as discussed earlier is also known a metabolic footprinting approach. Of the potential 111 metabolites, 59 were detected and suitable to be used for analysis, including all 20 of the standard proteinogenic amino acids, in addition to ammonium.

Monitoring metabolite levels overtime means that metabolites can be grouped based on how they change. In this work, 15 metabolites had no trend across time, 5 had a linear change and 2 had a

complex pattern. Cysteine was roughly biphasic starting with a linear decrease followed by a plateau, whereas 2-aminoisobutyrate increased to a maximum value before decreasing as cells began to use this metabolite. The remain metabolites – 37 – displayed changes that could be modelled by sigmoid curves as outlined previously [132]. These curves were either positive (in red – excreted during growth) or negative (in bue -depleted during growth) (Fig. 3.4). It is outside the scope of this thesis to fully investigate these changes, rather this analysis was carried out to investigate data quality and check that it could be done in principal.

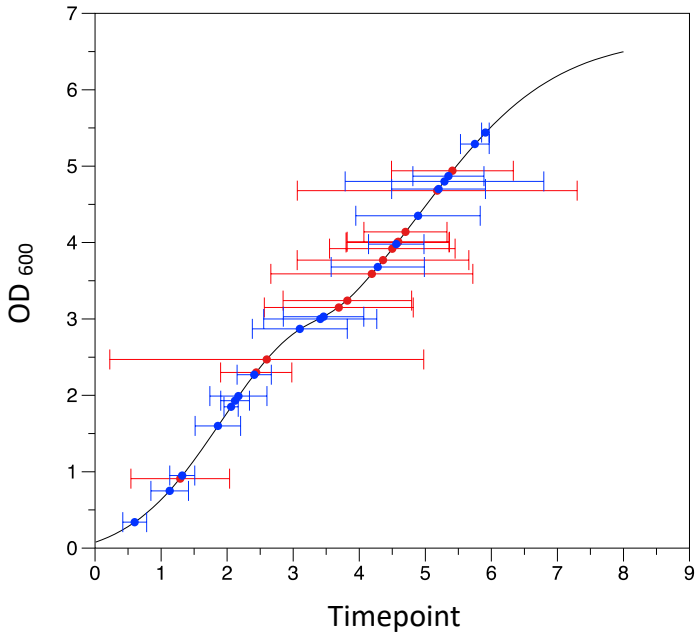


Figure 3.4 Metabolite utilisation, of E. coli *grown on LB. Plot is OD against timepoint in black, overlayed with metabolites and when they have been utilised either in red – excreted during growth or in bue -depleted during growth, n = 5.*

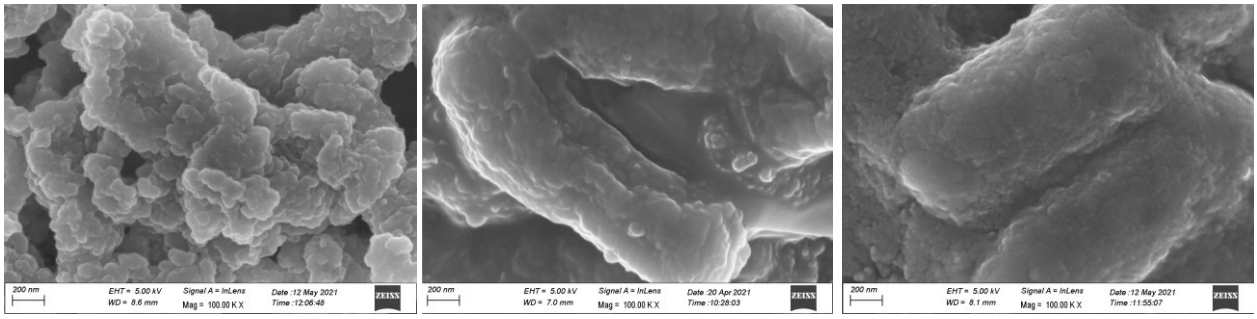
3.3.2.2 Qualitative evaluation of methods

All methods were carried out at the same time under as similar conditions as possible, with initial tests carried out using *K. oxytoca* grown in NFDM aerobically. This was done to both become familiar with the methods and as setting up aerobic cultures is less time intensive. All methods took under 3 minutes for all three replicates to be completed, excluding centrifugation time, for a total time of 9 minutes 37 seconds for all samples. The centrifuge only approach had minimal hands-on time approximately 10s before centrifugation. However reactions were likely still taking place until extraction solution was added after 5 minutes of centrifugation, meaning metabolites could be changing during this time. Similarly, the AMW and MCW approaches had very quick hands-on time, which is essentially how long it takes to mix the sample and the quenching/extraction solution, approximately 10s. The GlySal method, was similar to the AMW and MCW methods in that it had a

very quick hands-on time but this approach only quenches the samples in this first step. While the mixing of the sample and the quenching solution was achievable, the viscosity of the samples meant that downstream extraction and analysis steps were difficult. In fact, even after extraction and several washes the resulting samples were still deemed to viscous to carry forward for LCMS analysis. Similar observations have been made in previous studies [133].

3.3.2.3 Scanning Electron Microscopy (SEM) to assess cell integrity

To assess how much cells had been 'damaged' allowing for intracellular metabolite recovery SEM was carried out. These samples were collected from anaerobic cultures, after then initial assessment using aerobic cultures had excluded the GlySal approach. Figure 3.5 compares the cell structure after the extraction by either centrifuge only, AMW at -20, or MCW, with panel A showing the most cell damage likely as a result of the bead beating step. While the cell integrity appears to be relatively high after simultaneous quenching and extraction using AMW and MCW, the cells only need be made porous enough for the intracellular metabolites to be extracted. Therefore, these results need to be considered in concert with other metrics.



CO AMW MCW

Figure 3.5 SEM images of single K. oxytoca cells post extraction for (A) CO plus MCW, (B) AMW and (C) MCW. Images are a representative of all cells in the sample and are shown at a magnification of X 100,000. The CO method (A), which included bead beating, has the greatest cell wall disruption, whereas the AMW (B) and MCW (C) methods which did not include bead beating have less visible cell wall disruption.

3.3.2.4 Recovery of intracellular metabolites

After the initial testing of sampling methods, samples from anaerobic cultures were collected and analysed using the same AccQ-Tag RPLC-MS method as the LB supernatants, with 49 metabolites being detected. Prior to normalisation data were inspected to check quality and ensure there had been no issues with data collection. During this process a potential issue with the data was noticed. While within treatment variation was low and there were differences between treatments the differences

were extreme and not in line with previous data [119,129]. The centrifuge only treatments had a far greater peak integral compared to any other treatment an example of which can be seen for methionine in Figure 3.6. Of the 49 metabolites, 34 were significantly higher in the CO treatment compared to AMW and 26 metabolites between CO and MCW (Wilcox-sign ranked test, p <=0.005). Importantly, many of the proteinogenic amino acids were found to be significantly more abundant in the CO treatment. While this could be explained by the greater cell damage seen in the SEM images, that the peak area was decreased in the centrifuge only mix treatment (where extracted pellets are mixed back into the supernatant before being dried) indicated potential ion suppression.

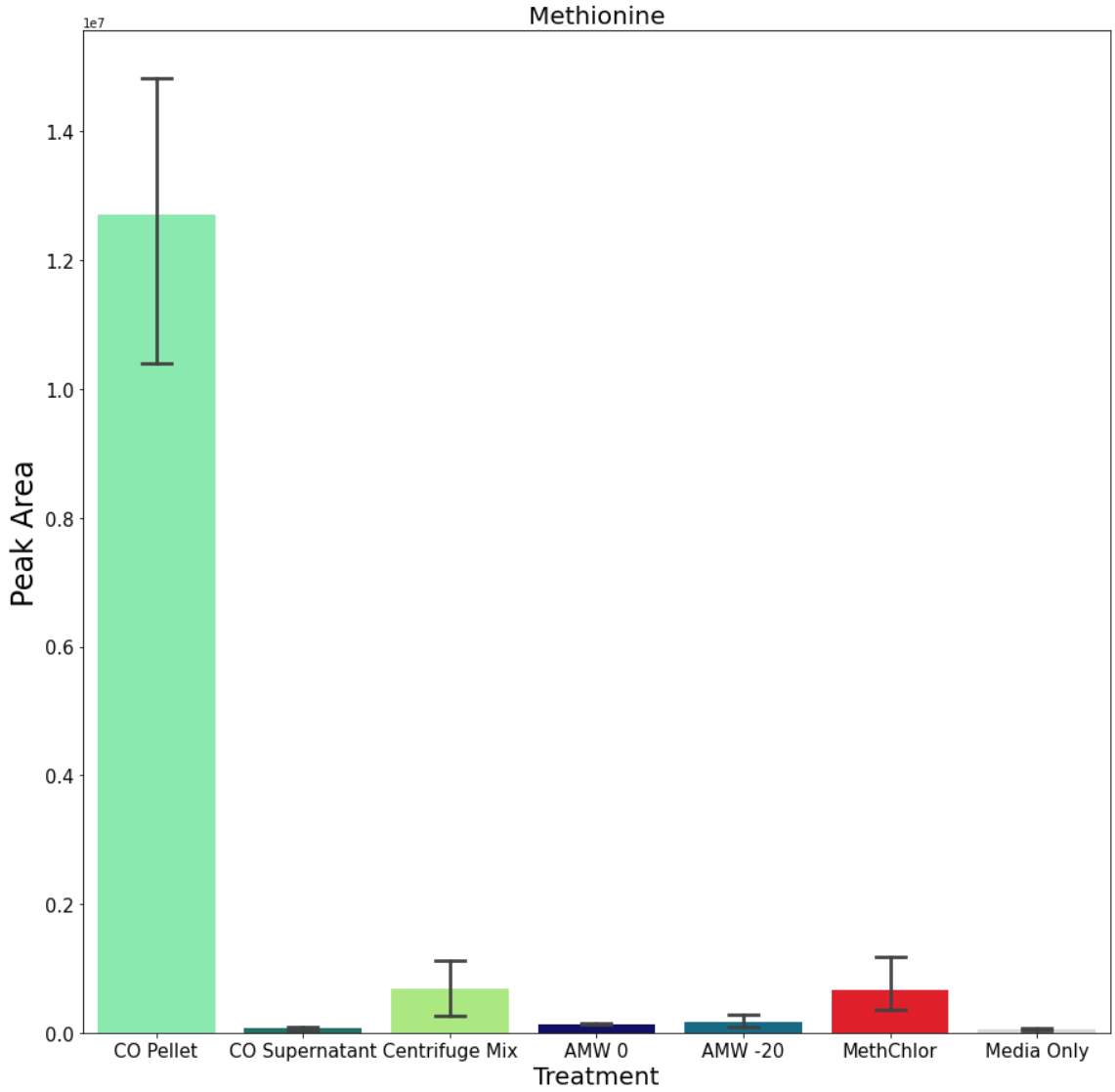


Figure 3.6 A comparison of methionine abundance from a single anaerobic culture of K. oxytoca using sampling methods outlined above. This metabolite is representative of a trend seen for nearly all metabolites detected using the AccQ-Tag RPLC-MS method, n = 3.

3.3.2.5 Amino acid abundance affected by media

To investigate possible medium induced ion-suppression a known amino acid standard was spiked into water and NFDM medium and analysed using AccQ-Tag RPLC-MS. The NFDM amino acid standard mixes were either analysed without any drying down, dried down and reconstituted to the same volume, or dried down and concentrated X 10. Results show that 14 out of the 23 metabolites tested were significantly different between the AA water and AA medium treatments (Fig. 3.7). Metabolites that are important to the nitrogen fixation process, including glutamine and glutamate (Fig. 3.7) were most abundant in the water only treatment and thus subject to ion suppression.

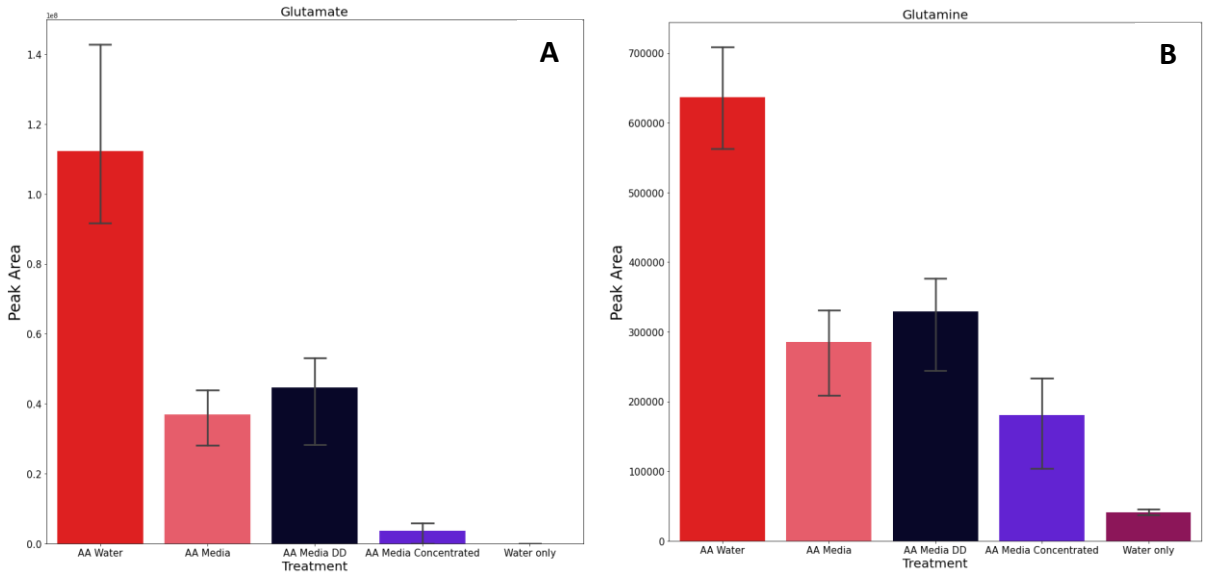


Figure 3.7 Comparison of metabolite abundance of (A) glutamate and (B) glutamine from amino acid standard spiked into water and NFDM media. NFDM was analysed as normal, or was dried down and reconstituted to initial volume, or dried down and concentrated ten times. Water blank included for comparison, n =3.

3.3.2.6 Methanol Chloroform water (MCW) extraction metabolite recovery

Due to the ion suppression resulting from the media, it was decided that the cells would have to be separated from the medium before analysis. Therefore, the CO approach was selected as the most suitable sampling method to be used for work in this thesis. Results comparing metabolite abundance of samples from anaerobic cultures indicated that metabolite recovery was better when using MCW simultaneous quenching and extraction compared to AMW (Fig. 3.6). Therefore, a trial of CO paired with MCW extraction was carried out to allow comparison to CO paired with AMW extraction. Pellets extracted using MCW performed better than the simultaneous quenching and extraction using MCW on key metabolites such as glutamate (Fig. 3.8). Further, the MCW pellet extraction appeared to be qualitatively better compared to AMW pellet extraction. Despite pellets extracted using AMW having a greater abundance than either simultaneous quenching and extraction approach, pellets extracted

using MCW had even greater metabolite abundance; as seen for glutamate in Figure 3.8. This suggests that MCW is the most efficient extraction method for *K. oxytoca* cell pellets. Additionally, the MCW approach had been used in conjunction with both AccQ-Tag RPLC and IPC by others in the context of developing the targeted LC-MS methods [129].

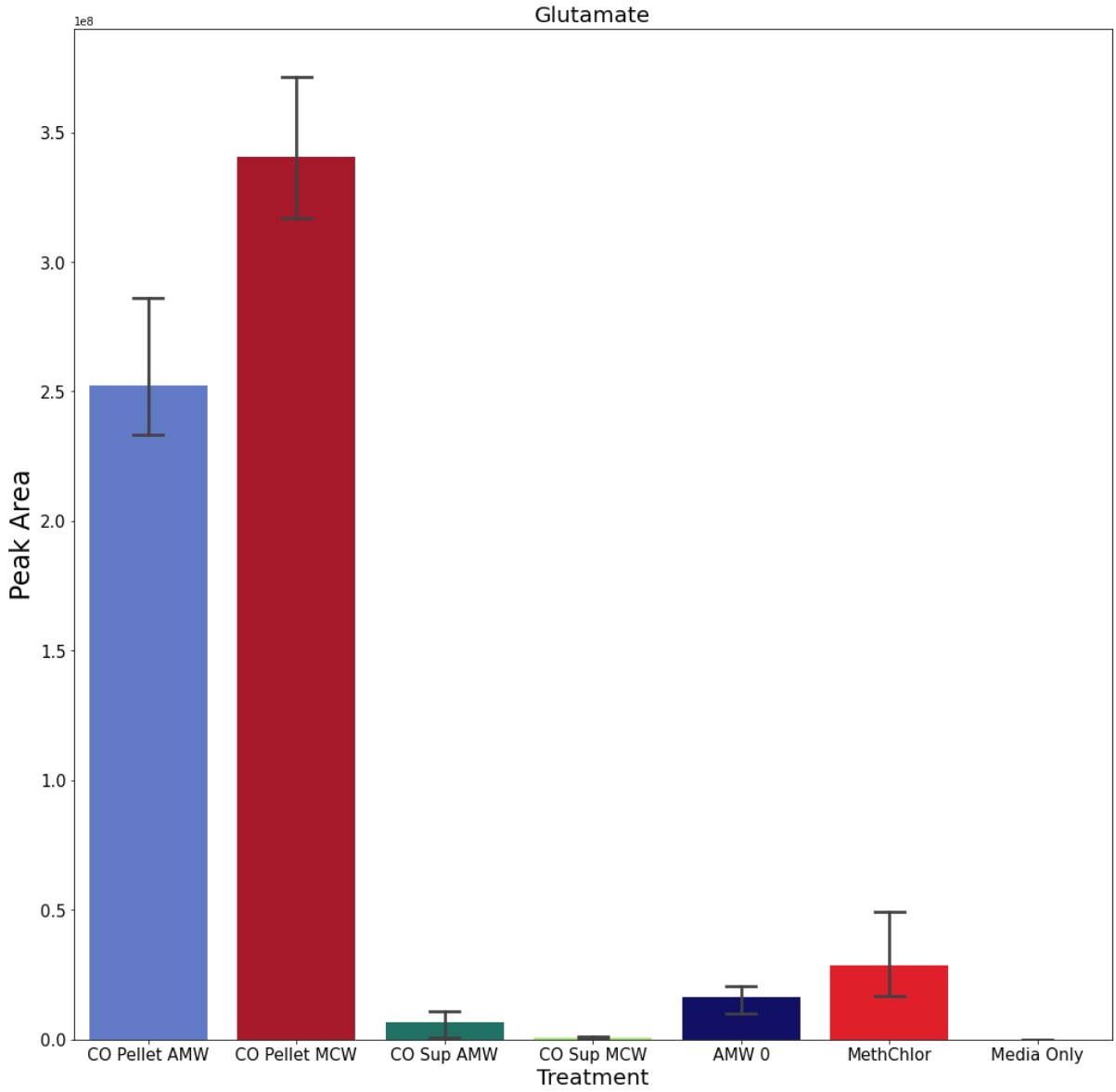


Figure 3.8 Comparison of glutamate abundance from a single culture of K. oxytoca. Sampled using CO with MCW extraction, CO with AMW extraction and simultaneous quenching and extraction using AWM, and MethChlor. A medium control was also included, n = 3.

3.4 Discussion

3.4.1 Increased metabolite coverage compared to RPLC

While previous studies have implemented targeted methods covering a wide range of metabolites [126,127,129], the work in this chapter expands on these studies by combining AccQ-Tag RP-LC, IPC and RPLC methods to increase the overall metabolite library coverage. Together, the chromatography methods of IPC and AccQ-Tag RPLC can detect and determine 287 metabolites. Alone, RPLC captured 40% of the metabolites in our library, compared to the combined method approach which detects 90%. The remaining 10% are metabolites that have historically always been difficult to retain using LC approaches, namely the sugars and polyols.

That there was only a small amount of metabolite overlap, 57 metabolites, and weak association between retention characteristics of these metabolites means that IPC and AccQ-Tag RPLC are well suited to be used together.

3.4.2 AccQ-Tag method performs well for *E. coli* supernatant

The key aim of this work was to check that the AccQ-Tag RPLC-MS method worked well for bacterial samples and was comparable to other studies. Hence, *E. coli* was chosen as the model organism as it has been widely used in many metabolomic studies, including by this research group [124,132,134,135]. The data collected in this experiment successfully allowed us to follow metabolite changes through time in the supernatant and was of high enough quality to allow for non-linear modelling using a TREF (Time-resolved exometabolomic footprinting) approach. The results are broadly similar to those from previous studies which found some metabolites are taken up by cells whereas others are excreted into the medium. For example, glutamate, arginine and alanine were all used up from the medium by the cells in this work and in previous work carried out by this laboratory [132].

While further investigation of the *E. coli* supernatant data could be carried out, it was considered outside the scope of this thesis and thus will not be taken further here. Moreover, a thorough investigation of all three LC-MS methods was carried out including a variety of different biological samples and validations separate to the work of this thesis, but for which this work contributed to [129].

3.4.3 Considering targeted LC-MS methods

More broadly, targeted methods are useful for investigating what is occurring a biological system at a given time. This experiment illustrates that addressing these questions is possible using the AccQ-Tag RPLC-MS method and the work by Sagi-Kiss *et al.,* [129] illustrate these methods applicability to other sample types and model systems. The ability to answer these types of questions is key when optimising biological systems and therefore having two robust methods is a strong asset. Furthermore, that these methods cover a great number of biological important metabolites means that fast moving metabolic changes within a system can be investigated in ways that would have previously proven difficult.

Of course, these methods could be further improved, with the addition of more metabolites being a clear place to start. Alternatively, adding an additional complementary method such as RPLC which would cover metabolites that lack an amine group or ionise poorly in negative mode or with IPC. Similarly, one could add a targeted complementary analysis for a currently poorly addressed sub-group of polar metabolites. One could suggest a system specific group to focus on e.g. carboxylic acids or thiols. The Thiol group would be particularly interesting for nitrogen fixation given that this group of metabolite has been show to play a role in the formation of the nitrogenase [26,136,137].

Additionally, while these methods meet the Metabolomics Standards Initiative (MSI) level 1 identification criteria [138], there is still a risk of misassigning compounds when analysing complex biological samples. For example, structural isomers which are resolved using standards may behave differently in real samples where only one might be present or the RT has shifted due to matrix effects. This can result in metabolite misassignment but is overcome by reanalysing the sample with an authentic standard spiked in.

3.4.4 Quenching of *K. oxytoca* proves to be non-trivial

Samples were generated using the six different sampling methods outlined above and were from the same *K. oxytoca* batch culture at the same time point. This included simultaneous quench and extraction methods as well as more traditional cell separation and extractions methods [87,121,128]. The hands-on time for most of the sampling methods was very quick, which is advantageous when trying to capture an accurate representation of the cells metabolic state. The obvious exception to this is the centrifuge only treatment in which cells are still metabolically active for the 5 minutes they are being spun down. There are many potential external confounding factors during centrifugation, such as temperature change and delayed extraction. However, this approach does benefit from a reduced risk

of the supernatant masking low abundant metabolites and the supernatant not having to be accounted for later in the analysis [122].

The GlySal approach first outlined by Villas-Boas was the only method not taken forward for analysis [87]. This was for two main reasons, the first is that with the wash steps this method took far longer than the other methods including the centrifuge only method. A review of quenching methods reported they found this method took five times longer than the other quenching methods trialled, which is similar to the results from the work in this chapter [133]. The increase in sampling time could be accepted should the downstream results be of sufficiently higher quality, compared to other sampling methods. However, the final sample from the extracted pellet, after many washes, still showed glycerol residue at levels which were thought to pose a risk to the LC-MS machine. One trial injection was attempted but it was clear that the sample was too viscous and therefore the possibility of many samples being analysed in a high throughput manner was negligible. As such, the decision was made to not take the GlySal quenching method forward for analysis.

3.4.5 Medium composition interferes with LC-MS analysis

The remaining sampling techniques were tested on anaerobic cultures and analysed using the AccQ-TAG RPLC-MS method which had been used successfully on the *E. coli* supernatant samples. However an initial comparison of results indicated that the sampling methods had not performed as expected. There was a stark reduction in metabolite abundance in samples that included the supernatant compared to the centrifuge only sampling method, that included the cell pellet only (Fig. 3.6). The most obvious cause of this issue was ion suppression due to the supernatant. While the possibility of ion-suppression caused by the salt content of the medium had been considered, the degree of ion-suppression was far greater than expected. Indeed, AccQ-Tag RPLC-MS was selected as the first analytical method to analyse the samples not only because it had been trialled on the *E. coli* supernatant but also because it dilutes the samples. The derivatisation process includes a 1:10 dilution, and samples were further diluted (a minimum of 1 in 10) for a final dilution of 1:100. This dilution was expected to limit the ion suppression caused by the salts, should any been seen in the results. A spike in experiment using an amino acid standard confirmed that ion suppression was the cause of the reduced signal. In the 10 X concentrated sample, the abundance of glutamate was significantly lower than not only the amino acids in water but also in the unconcentrated media, a trend seen in almost all the other metabolites tested. Given that the samples collected were concentrated 30 X the ion suppression would only be compounded, thus explaining the much lower abundances seen in the sampling methods including media.

Similarly, the high glucose content of the media, required due energy demands of fixation, posed more of a problem than expected. The sampling process along with the derivatisation steps discussed above were expected to dilute the glucose such that it would not impact the analysis. Especially considering that results discussed later in Chapter 4 indicate that up to 50% of the glucose in the bacteria is used by fixing cells. However, after the AccQ-Tag RPLC-MS analysis had finished there was a brown residue on the cone which was believed to the glucose residue due to its resemblance to a medium only dried down sample. While not tested it is possible that alternative carbon sources such as sucrose could have alleviated this issue. Alternatively, one could attempt to selectively deplete the supernatant of glucose using a filtration method.

3.4.6 Overcoming medium interference in LC-MS analysis

Two directions were considered at this junction. First was to the change the medium composition to make it more compatible with MS analysis, either through adaptation of the NFDM or trialling an altogether new media. However, neither of these options were thought viable, as this medium has been used in other studies [139–141] as well as previous research carried out in this laboratory [39,41], and for this thesis. Additionally, other minimal medium recipes which have been used for nitrogen fixation are similar too, or based on NFDM, meaning they also have high salt and carbon content [30,64]. As such, the decision was made to explore the second direction, separating the cells from media.

While the CO treatment was initially included, as a commonly used standard to compare the other methods to, the overall ratios of metabolites were similar to the other methods suggesting that the metabolic snapshot was being captured. The other separation technique commonly used is fast filtration. A single, not reported, test run was carried using the filtration equipment available in the lab, and the time taken to separate the cells from the medium exceeded that of centrifugation for the same volume of media. Also, removing the cells from the filter for extraction was not instantaneous, increasing the total time cell metabolism could be changing. Therefore, the CO approach was elected to be taken forward.

The CO awhichproach, is still wiwhichely used in many metabolomic studies despite the previously discussed limitations [39,79,120,142]. In this work, centrifugation was coupled with AMW extraction resulting in the highest signal for all metabolites in the first round of testing. However, it could be that AMW was not the best extraction method for this model system, and other commonly used extraction methods such as MCW could more effective [130]. The AMW sampling method was chosen because it

has been shown to be superior to MCW when extracting *E. coli* cells by Rabinowitz and Kimball [119]. Given the similarity between *E. coli* and *K. oxytoca* it was thought that AMW could produce superior extraction compared to MCW for this model system as well. Yet, when we compare how the AMW and MCW approaches performed for simultaneous quenching and extraction, metabolite abundance was higher in the MCW method for most metabolites, despite the sample volume being lower. This could be due to MCW being a biphasic extraction method, whereas AMW is monophasic [119]. The polar metabolites of interest are in the aqueous upper layer of the MCW extraction method which could be removing a further source of ion-suppression from the other phases. In fact, in the other experiments associated with this method development carried out by Sagi-Kiss *et al.,* [129] the MCW extraction method was used exclusively and it is widely used in other studies as well [39,79,86].

To investigate whether the MCW extraction would perform better than AMW extraction for the cell pellets generated from centrifugation a trial was carried out. In this trial metabolite abundance for the extracted cell pellets was higher than in the simultaneous quench extraction using MCW. Additionally, the metabolite abundance was higher and peak shape more Gaussian (more symmetrical and lacking shoulders or tails) for many metabolites in the MCW extracted cell pellets compared to AMW extracted pellets. Most importantly, the amino acids were among the metabolites seen to have greater abundance with MCW extraction, which given their importance in nitrogen metabolism is an important factor to consider. As such, the MCW extraction method was shown to be a better at extracting key metabolites using AccQ-Tag RPLC-MS.

The centrifuge only MCW extraction method, also benefited from better cell disruption compared to the other sampling methods. SEM images show that cell pellets extracted using a bead beating step were visibly more disrupted compared to cells using a solvent only extraction (Fig. 3.5). While cells do not have to be completely disrupted for intracellular metabolites to be detected, methods that include bead beating generally have better repeatability and yield of metabolites [130,143–146]. A direct comparison of cell pellet extraction with and without bead beating was not carried out due to the consensus in the literature of its benefits. However, it is possible that a portion of the increased metabolite abundance in the centrifuge only treatment is due to the addition of the bead beating step.

In summary, the sampling method that was selected was the centrifuge only approach outlined above using a MCW extraction of the cell pellet. This was due to the experimental limitations imposed by the resources available, namely the reagents, sampling equipment and medium composition. The supernatant and medium were not analysed due to the ion suppression and glucose issues described

above. However, had other resources been available, such as improved fast filtration or a column-based approach to reduce the salt content of the medium, than a different method may have been developed that was superior to the one chosen.

With that said, were I to have focussed exclusively on sample quenching and extraction, the optimal process would have likely been a fast filtration method. This would need to achieve near instantaneous separation of the cells from the supernatant, and would need to work with aerobic and anaerobic growth conditions. This method would also need to have utilised inert filter paper, from which the cells could be extracted directly without contamination. Thorfinnsdottir *et al.,* recently outlined the parameters one would need to test and optimise throughout the entire sampling process to accurately investigate the central carbon metabolome for *E. coli* including; filter material and pore size, sample biomass, whether to include a rinsing step and if so the temperature and composition of the solution, quenching and extraction solvents, and other purification steps prior to analysis [147]. Were I to restart the part of the thesis again, I would at a minimum have tried to address the issue of high salt content in the medium. By either changing the medium composition or removing the salts prior to analysis, I could have analysed the supernatant using LC-MS and quantified the excretion of ammonium and other plant growth promoting metabolites.

3.5 Conclusion

The method development carried out in this chapter is complementary in the sense that both sections work towards answering 'what is happening' questions as accurately as possible. The chosen sampling method of centrifugation followed by CMW extraction will preserve the cells metabolic state at a given moment in time and the complimentary UHPLC-MS methods of AccQ-Tag RPLC and IPC cover many polar metabolites found in the metabolome. Crucially, key metabolites relating to nitrogen fixation, such as glutamine and glutamate (as well as the amino acids more broadly), are quenched and extracted well, in addition to be being detected and well characterised by UHPLC-MS.

The work in this chapter was not intended to be an exhaustive list of all possible sampling and analytical approaches, rather an investigation of a curated list of methods from the literature. To that end, the work in this chapter has been successful in selecting the best approaches for sampling and in determining which metabolites to use in the UHPLC-MS methods. When trying to address problems such as improving soil nitrogen fixation, having the highest quality data relating to the 'what is happening' questions is of paramount importance. This work lays the foundation from which further experimental data can be collected to address these questions, while having confidence in both the data collected and the conclusions drawn from that data.

Chapter 4 Metabolic footprinting of supernatant samples under nitrogen fixing conditions

4.1 Introduction

Bacterial metabolism plays a vital role in how species interact with the soil environment, from other bacteria to plant roots. Therefore, understanding how nitrogen fixing bacteria, such as *K. oxytoca,* modulate their local environment through secretion of metabolites is important [148]. This information allows insights not only into cell gene expression but also metabolites that are exchanged between bacteria and between the bacteria and plant roots [102,149]. While ammonium and phosphorus secretion have received much attention [21,26,149,150], it is important to not overlook the other metabolites that bacteria supply to their local environment and plant roots.

As in symbiotic relationships between crops and soil microorganisms, free living plant associated soil microbes can, and often do, provide compounds which the recipient is unable to produce themselves [21,151]. Nitrogen fixing bacteria are obviously important to the rhizosphere and the soil microbiome at large for their ability to produce ammonium, which only a small number of bacteria and archaea can provide. However, given the plant associative nature of *K. oxytoca* it is likely that other compounds are exchanged between the bacteria and the plant roots, not just ammonium. This has been shown to be the case for well-studied symbiotic species for which there are strong parallels [21,57,74,152].

To date, much research has focussed on the ammonium output of *K. oxytoca* and similar species as the key metric of interest both *in-situ* and for engineered species [30,60,63,153,154]. Focussing solely on ammonium excretion has clear merits, particularly its direct approach to the problem of nitrogen limitation in soils, but perhaps misses important nuances relating to the exchange of other metabolites that could improve either ammonium excretion, plant/bacteria health, or increase crop yields further. For example, a recent paper investigating plant root exudates found a specialised metabolic network regulated the metabolite composition of the soil close to *Arabidopsis* roots, altering the local bacterial composition [155]. This led to a favourable rhizosphere composition around the plant roots which improved growth and plant health [155]. As such, one could expect plant associative bacteria to also engage in similar practices of modulating their metabolism to get the most from the plant and the root exudates they produce. To be clear, it is unlikely, impossible perhaps, that one bacterial species would dominate the rhizosphere or be able to provide all the metabolites that plants can't synthesise themselves. However, it is also unlikely that a bacteria species will produce a single metabolite that the plant needs. Understanding what else plant associative species like *K. oxytoca* provide the plants will inform optimisation of those systems.

As discussed in Chapter 1, synthetic biological approaches tend to focus on the nitrogen fixation aspect of diazotrophic bacteria. The NIF cluster from *K. oxytoca* has been refactored [30] and is a candidate for insertion into the plants themselves [156]. This is an important first step, but optimisation of refactored bacteria and other future steps should consider all possible metabolite exchanges. As we move towards an understanding of the complete system including mutualisms and chemotaxis, synthetic biological approaches can use this information to improve plant growth through genetic engineering of bacteria.

Metabolic footprinting techniques allow for metabolic signatures of given cell states to be uncovered. Using these approaches to investigate how the supernatant composition changes due to nitrogen fixation will help uncover other metabolites that are produced during fixation, that could also be beneficial to plants or bacteria. Using an NMR based approach will allow for broad changes in metabolism and for the detection of metabolites that accumulate in the supernatant which could be used by plants, or indeed other microorganisms in the biome of the soil.

4.1.1 Aims and objectives:

The aim of this chapter is to investigate the nitrogen fixation induced metabolic changes in the supernatant of *K. oxytoca* and to look for other metabolites resulting from fixation that could be promote plant growth and health, and give insights into the broader physiology of diazotrophs.

Objectives:

- To measure the global profiles of *K. oxytoca* supernatant before, transitioning to and during fixation using ^{1}H NMR spectroscopy.
- To analyse the spectral data of the supernatant using multivariate and univariate statistical methods.

4.2 Methods

4.2.1 Bacterial strain, mutants, and culturing

The parent strain and mutants of this strain used in the experiments carried out in this chapter can be seen in Table 4.1.

Table 4.1 Bacterial strain and mutants used in this Chapter.

Strain name	Gene knockout	Summary
Wild type (WT) M5a1	None	Has a normal growth phenotype and is able to fix nitrogen
△*nifLA*	△*nifLA*	Unable to fix nitrogen, no diazotrophic growth
△*nifK*	△$nifK_{1\text{-}1203}$ *truncation (△nifK)*	Unable to fix nitrogen; C-terminal domain of NifK necessary for nitrogenase activity removed; *nif* expression otherwise unimpaired
△*amtB*	△*amtB*	Able to fix nitrogen and exhibits diazotrophic growth; AmtB is necessary to regulate ammonium levels in the cell

Bacteria were grown as outlined in Chapter 2.4.1 'Liquid culture experiments'. In brief, the WT and three mutants were grown for 24 hours in minimal NDFM medium supplemented with either 0.5 mM (low nitrogen) or 10 mM (high nitrogen) ammonium chloride. Ammonium run out, when 0.5 mM supplied ammonia is exhausted from the growth media, was used to induce diazotrophic growth. The use of 0.5 mM ammonium chloride for the low nitrogen treatment and it reliably inducing diazotrophic growth had previously been tested and validated by earlier research carried out in the laboratory [39]. The same was also true for the supplementation of 10 mM ammonium chloride for the high nitrogen treatment and the absence of diazotrophic growth and nitrogen fixation [39]. Sampling was carried out at regular intervals, every hour for the first 4 hours and every 30 minutes between 4 – 10 hours; the time of diazotrophic growth. A final sample was collected at 24 hours for a total of 17 timepoints. During sampling, OD_{600} was measured, and cells were spun at 13,000 g for five minutes after which the supernatant was split into equal 450 μL aliquots and stored at -80°C.

In summary, the WT and three mutants were subject to 4 treatment groups, as they were grown on either low or high nitrogen under aerobic and anaerobic conditions (Fig. 4.1.). Three replicates were carried out per treatment.

Acetylene reduction was also carried out, as outlined in Chapter 2.5 after 10.5 hours to confirm that nitrogen fixation was occurring and that cells had entered diazotrophy. To allow comparison with result in Chapter 6, results are presented in percentage of fixed ethylene. This was calculated by integrating both the acetylene and ethylene peaks, summing these, and then calculating the percentage contribution of the ethylene peak.

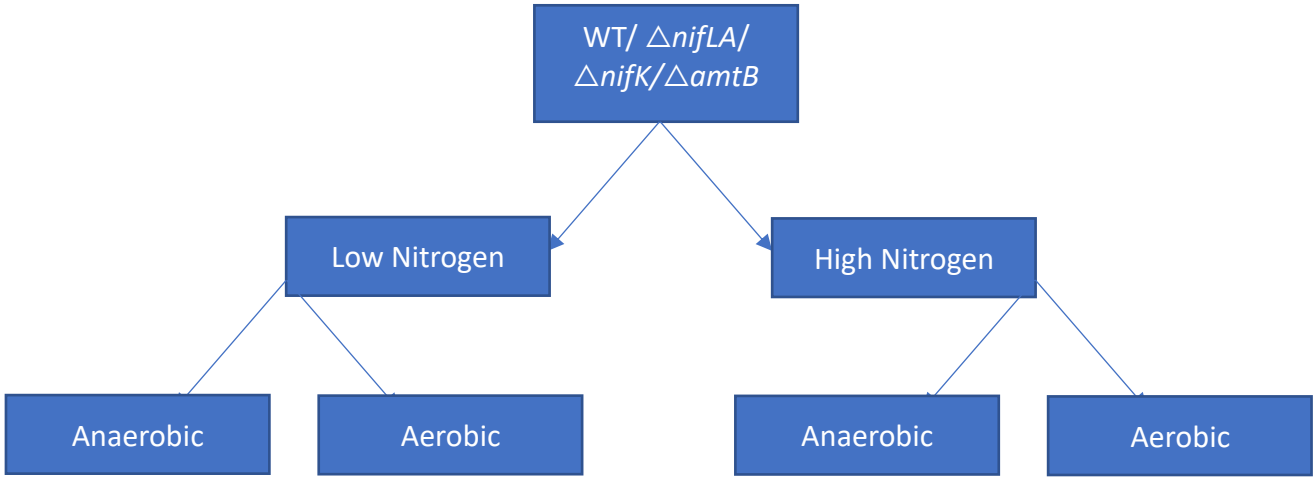


Figure 4.1 Summary of experimental treatments. Each strain is cultured on low nitrogen (0.5 mM ammonium chloride) and high nitrogen (10 mM ammonium chloride) NFDM under aerobic or anaerobic conditions.

4.2.2 ^{1}H NMR spectroscopy-based metabolic profiling

Supernatant samples were analysed using ^{1}H NMR spectroscopy as outlined in Chapter 2.6.2. Not all samples were analysed. No samples were analysed from any strain or mutant grown on high nitrogen under aerobic conditions, due to external constraints. It was considered that this was the best group to drop as the growth conditions mean that no fixation had occurred and it is the least relevant to conditions related to nitrogen fixation. Further, the growth curves for the WT and all mutants grown on high nitrogen were similar between aerobic and anaerobic conditions.

Then 16 of the 17 timepoints were analysed from all remaining treatments, the 1 hour timepoint was excluded, allowing for both time series and treatment group effects to be investigated.

4.3 Results

4.3.1 Diazotrophic growth clearly displayed by WT and △*amtB* mutant

The growth profiles of the WT and all mutants derived from the WT were similar when grown under unlimited nitrogen, aerobic conditions (Fig. 4.2 A). There are also no clear differences between the growth profiles of the WT and mutants when cultured in nitrogen rich, anaerobic conditions (Fig. 4.2 B). Further, growth profiles between the aerobic and anaerobic conditions are similar for growth on nitrogen rich medium (Fig. A&B). However, growth under high nitrogen, aerobic conditions achieved a greater OD compared to anaerobic conditions for the WT and all mutants compared to high nitrogen anaerobic growth. Overall, these profiles suggest that there is not a growth penalty resulting from gene deletion for the mutants studied, when grown under 'ideal', carbon and nitrogen replete, conditions.

Similarly, to high nitrogen conditions, no clear growth differences were observed under nitrogen limited, aerobic conditions between the WT strain and the mutants of this strain (Fig. 4.2 C). Although the △*amtB* and △*nifLA* mutants appear to have achieved a lower OD at 10 hours and exhibited a larger decline in OD at the 24 timepoint compared to WT and △*nifK*.

In contrast, growth differences that reflected capacities for nitrogen fixation and therefore diazotrophic growth were observed between the WT and mutants when grown under nitrogen limiting anaerobic conditions (Fig. 4.2. D). Both the WT and △*amtB* have functioning *nif* genes and can fix nitrogen resulting in diazotrophic growth, whereas the △*nifLA* and △*nifK* do not have functioning *nif* genes and cannot fix nitrogen meaning a second phase of growth did not occur. This means that while the WT and mutants exhibit similar growth profiles up until nitrogen run out, approximately 200 minutes, the WT and △*amtB* mutant exhibit a secondary phase of growth, which the △*nifLA* and △*nifK* mutants do not. This secondary phase starts after approximately 400 minutes for △*atmB* and 600 minutes for WT. Additionally, the △*amtB* mutant appears to have a slightly faster growth rate and higher final OD during this secondary growth phase, compared to WT. Separate unpublished studies carried out in this laboratory by Chris Waite (Imperial College London) suggest this is attributed to an earlier onset of low nitrogen status that triggers earlier *nif* gene expression.

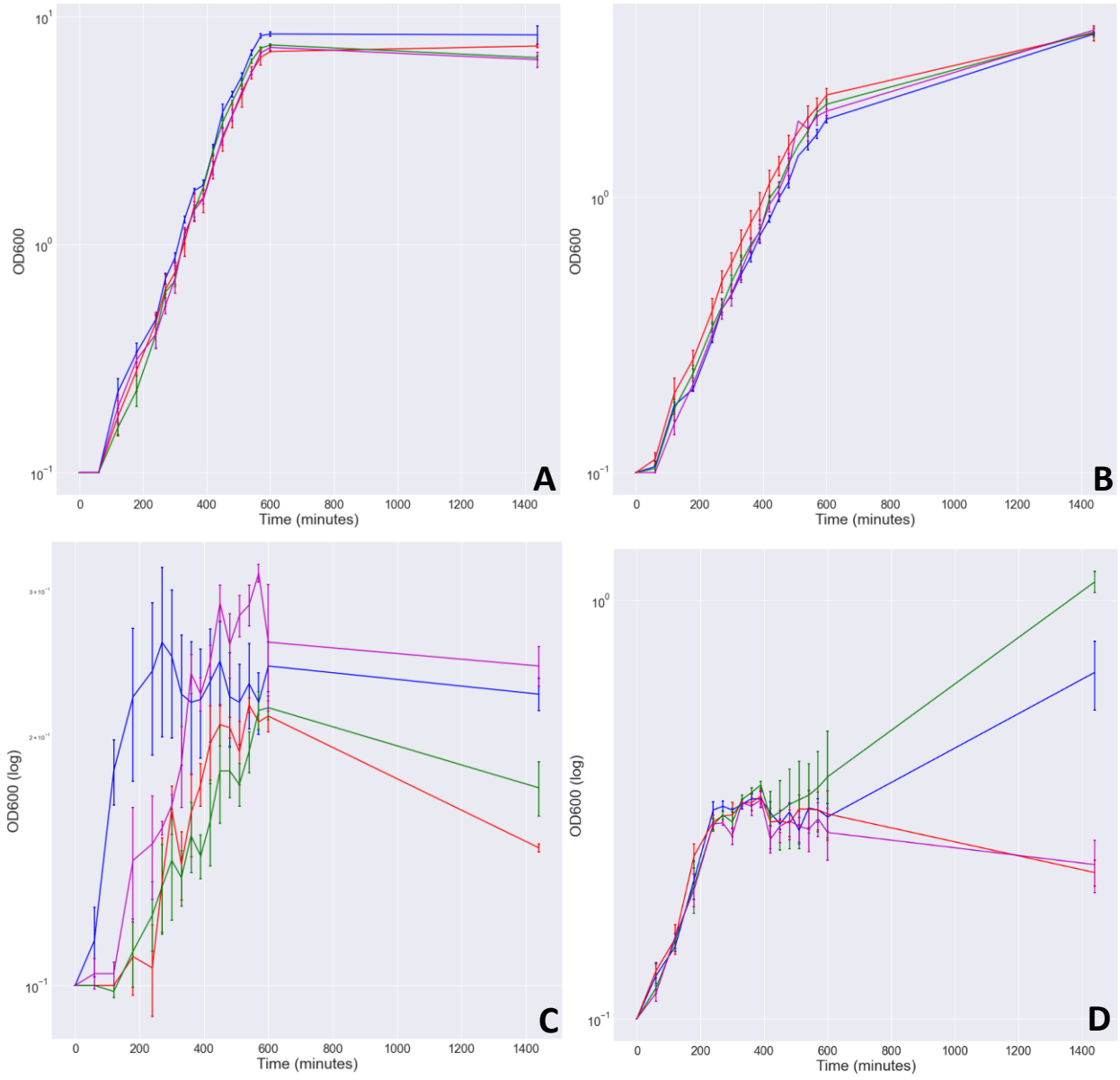


Figure 4.2 Growth curves of K. oxytoca strains WT (blue), ⊿nifLA (red), ⊿ amtB (green) and ⊿nifK (purple) for (A) nitrogen rich aerobic growth, (B) nitrogen rich anaerobic growth, (C) nitrogen poor aerobic growth, (D) nitrogen poor anaerobic growth. Treatment D is the is the only condition which should have facilitated diazotrophic growth. Log OD_{600} is plotted against time (minutes) for each treatment.

4.3.2 Acetylene reduction shows △*amtB* fixes more nitrogen compared to WT

The ARA results show that more acetylene was 'fixed' at both time points by △*amtB* compared to WT (Fig 4.3). This aligns with the greater OD_{600} observed at both timepoints and with more cells being able to fix nitrogen. However, differences persist after data is normalised to OD_{600}. There was no significant difference between △*amtB* and WT at 8.5 hours (p value = 0.39), or at 24 hours (p value = 0.066). There was a significant difference in nitrogen fixed between timepoints for the WT and △*amtB* respectively though, WT 8.5 v WT 24 p = 0.0409, and △*amtB* 8.5 v △*amtB* 24 p = 0.0221. The data trends towards △*amtB* fixing more nitrogen than WT and led me to conclude that △*amtB* transitions to diazotrophy at an earlier timepoint compared to WT.

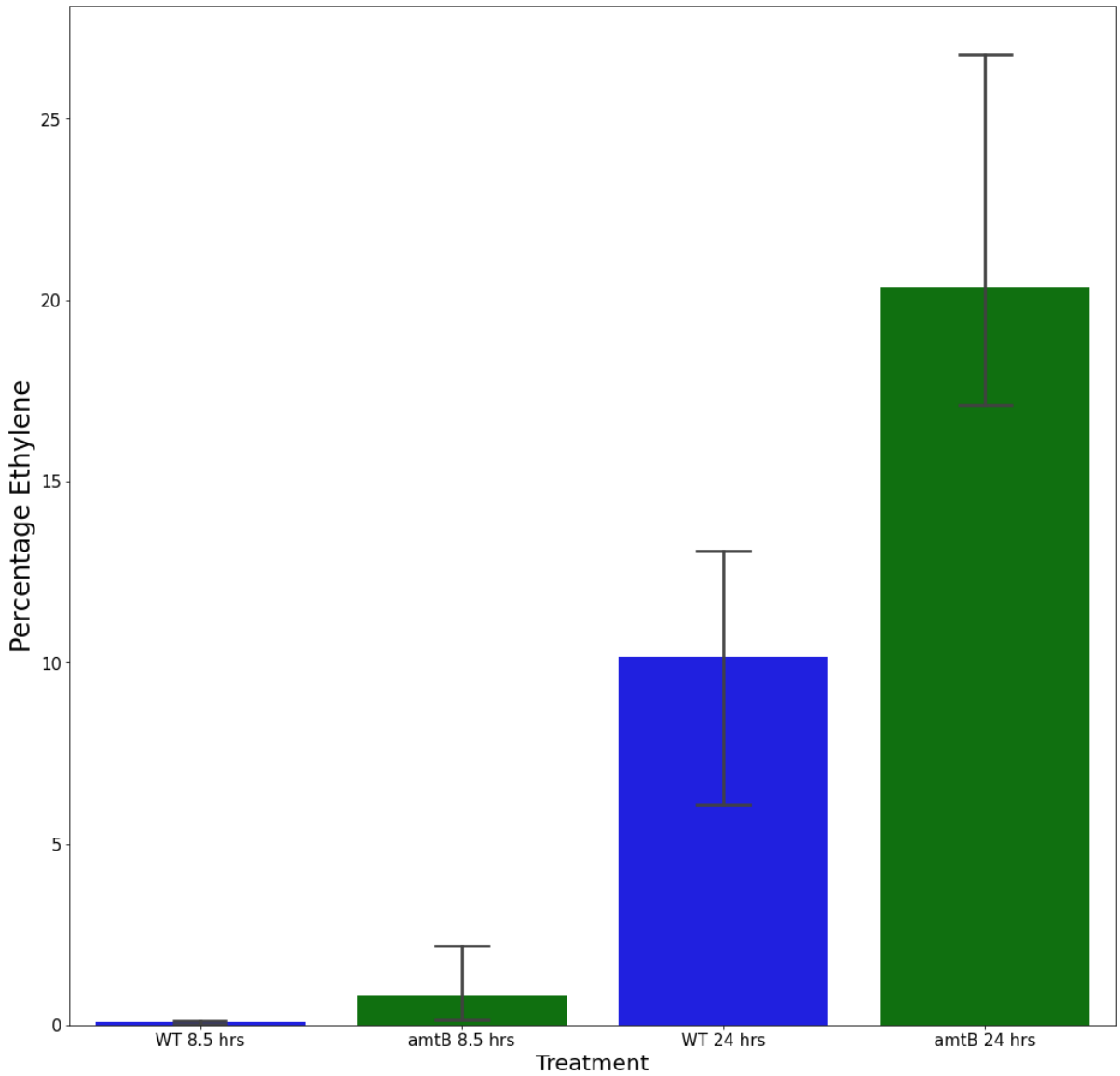


Figure 4.3 Comparison of percentage ethylene fixed by the WT and △amtB mutant during early fixation (8.5 hrs) and late fixation (24 hrs), n = 3.

4.3.3 Multivariate statistical analysis showed no strain, mutant or fixing specific metabolic profiles in supernatant samples

Of the four treatment groups, only three were analysed using NMR spectroscopy these being the aerobic and anaerobic low nitrogen treatments and the anaerobic high nitrogen treatment. All spectra obtained were included in the final analysis. The PCA scores plots (Fig. 4.4), show that no changes in the supernatant resulted in separation due to strain or mutants derived from this parent strain. Figure 4.4 (A) shows that all samples were largely grouped together. There is some separation along PC1 seemingly caused by early timepoints of the NifK deletion mutant, WT and NifLA deletion mutant. When looking for a specific metabolic profile of nitrogen fixation it is important to consider what happens in low nitrogen conditions. In Figure 4.4 (B), we can see that separation between the fixing metabolic footprints of WT and △*amtB* grown in nitrogen poor anaerobic conditions from the other treatment groups and the △*nifLA* and △*nifK* mutants. However, Figure 4.4 (C) shows that most of the separation from Figure 4.4 (B) can be explained by oxygen status of the treatment group, anaerobic (AN) or aerobic (AE) for cells grown on low nitrogen.

When looking at the loadings of the PCA analysis carried out on all samples (Fig. 4.4 A), separation is mainly driven by oxoglutarate and citrate in the aerobic condition and lactate, succinate and formate in the anaerobic condition. Additionally, levels of butan-2,3-diol were seen in comparatively high abundances in 24 hour samples for nitrogen fixing WT and △*amtB* samples, grown under anaerobic nitrogen limited conditions, with the highest abundances seen in the 24 hr △*amtB* supernatants. Butan-2-3-diol is produced as part of butanoate metabolism from pyruvate. Pyruvate is involved in many different pathways including glycolysis and the TCA cycle, so could be increased in cells using lots of carbon for growth, resulting increased butan-2-3-diol production [157]. Up to 50% (55 mM) of the available glucose is being used during nitrogen fixation, due to the energetic demands of nitrogenase synthesis. The distribution of samples on the scores plots of panels B and C of Figure 4.4 , appeared to be in part due to timepoint or growth. As such, PC1 was plotted against OD_{600} (Fig 4.4 D), which shows that some of PC1 can be explained by growth as well as the other metabolic changes resulting from oxygen status.

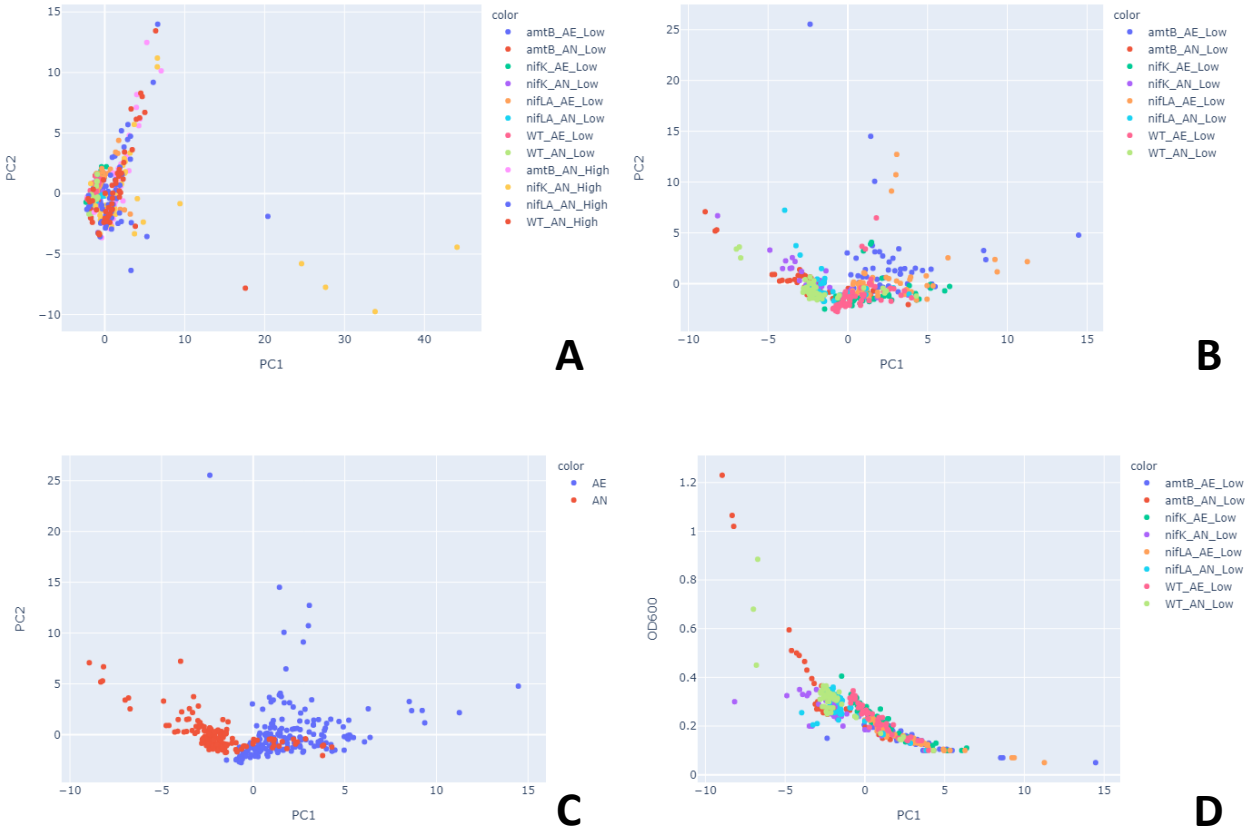


Figure 4.4 (A-C) PCA scores plot showing the supernatant metabolic profiles of WT, ⊿amtB, ⊿nifLA and ⊿nifK, grown under various conditions. A) Including all data coloured by treatment, namely anaerobic (AN) high and low nitrogen and aerobic (AE) low nitrogen B) Low nitrogen conditions only coloured by treatment and C) Low nitrogen conditions coloured by oxygen status. D) PC1 of low nitrogen PCA against sample OD_{600}.

4.3.4 Supernatant alpha-Ketoglutarate abundance shows inverse trend between high and low nitrogen anaerobic conditions

Of the 22 metabolite peaks integrated, most exhibited the same trend, either increase or decrease, in the supernatant independent of medium nitrogen status when grown in anaerobic conditions. However, alpha-Ketoglutarate levels in the medium remained low/unchanged in the high nitrogen condition but increased in the supernatant of the WT and all mutants grown on low nitrogen. The effect is most profound for the △*nifLA* mutant which had no fixation capacity at all compared to the WT and △*amtB* mutant in the low nitrogen anaerobic treatment. This reflects the increased nitrogen stress that the △*nifLA* mutant is under. Interestingly, the increase in peak intensity appears to be less

than in the △*nifK* mutant compared to △*nifLA* mutant despite that the fact it is also unable to fix nitrogen. The change in alpha-ketoglutarate abundance for the △*amtB* mutant was most similar to change seen in the high nitrogen anaerobic conditions, potentially due to its increased fixation capacity compared to the WT. A Kruskal-Wallis test found no WT or mutant effect on alpha-Ketoglutarate peak intensity at timepoint 17 ($p = 0.104$).

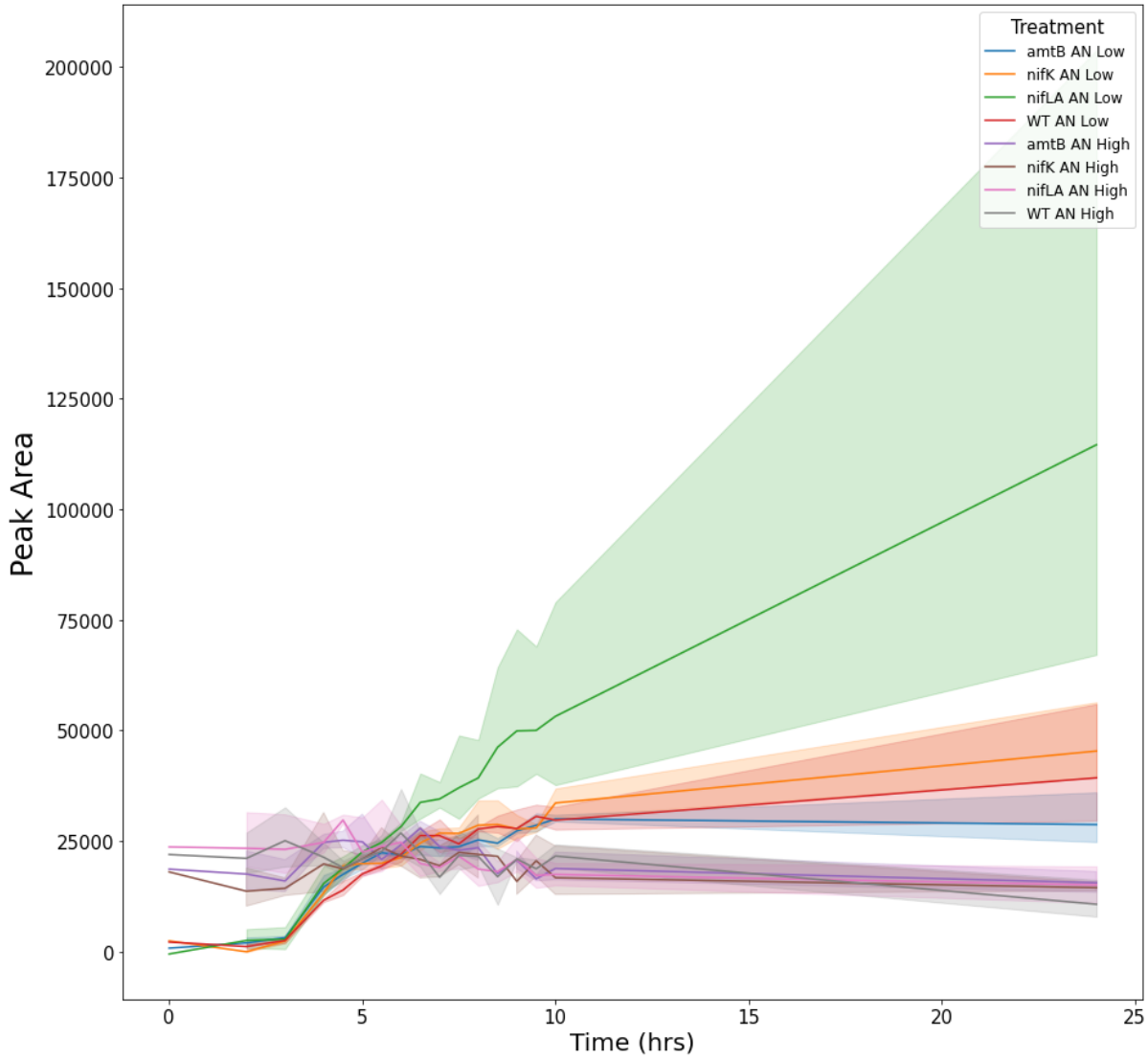


Figure 4.5 Time series plot alpha-ketoglutarate oxoglutarate. Plot shows changes in peak integral over time. Lines represent group mean curves and shaded areas denote 95% confidence bands, n=3.

4.3.5 Univariate analysis showed a significant strain and mutant effect on key metabolite abundance at the 24-hour time point for the WT and mutants grown under fixing conditions

Of the 22 peaks integrated, 11 were found to have a WT and mutant effect at the 24-hour time point. The most significant WT and mutant effect was seen in acetoin ($p => 0.001$), other metabolites of interest that were found to differ significantly between the WT and mutants were butan-2-3-diol ($p = 0.007$), formate ($p => 0.0001$) and lactate ($p => 0.0001$). Figure 4.6 shows that for all of these metabolites nitrogen the fixing phenotype, △amtB and WT, had the greatest peak integral in that order. The non-fixing mutants, △*nifLA* and △*nifK*, had lower abundance of the metabolites at the 24-hour timepoint.

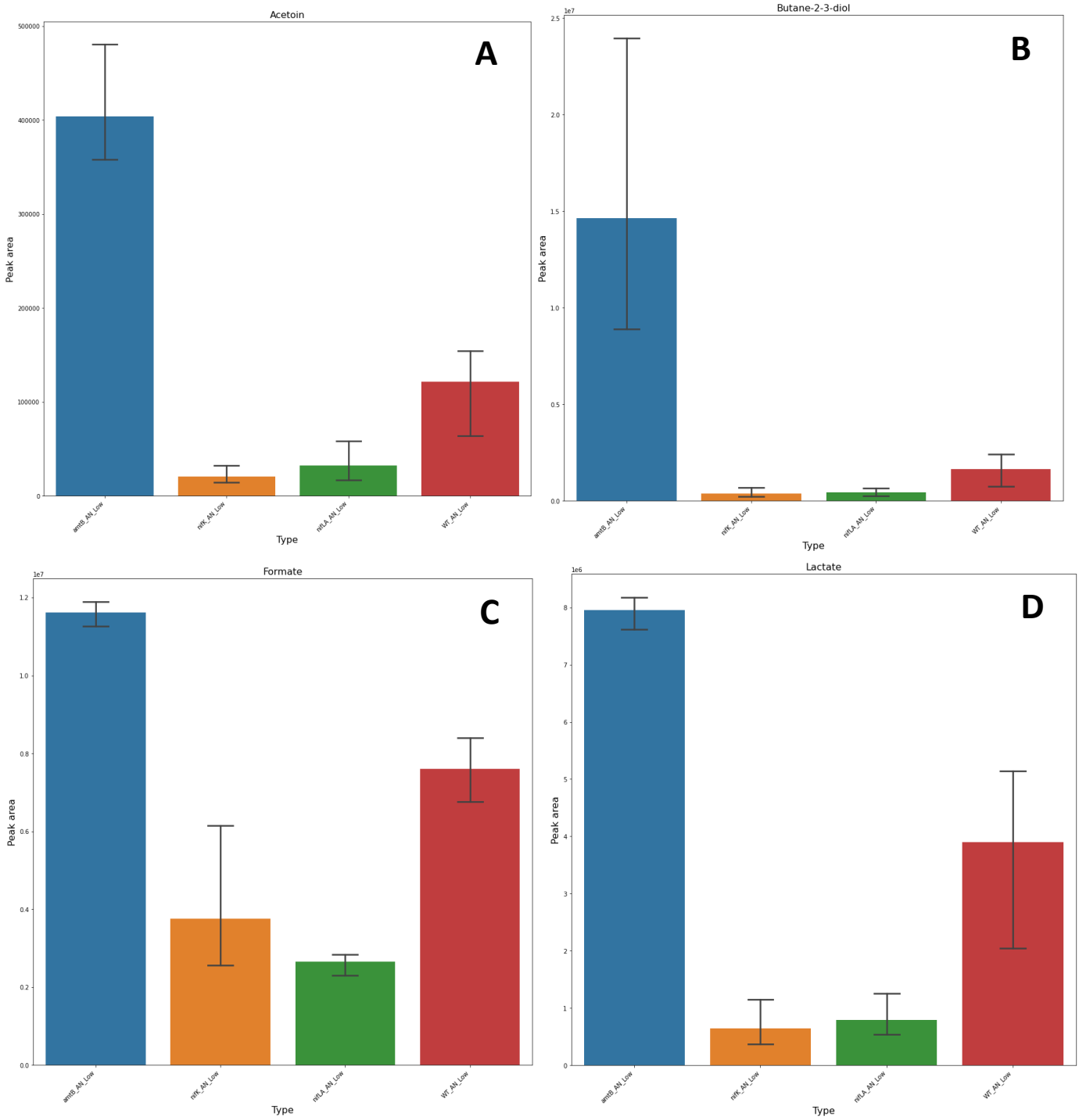


Figure 4.6 Plot showing peak integral at the 24-hour timepoint for the low nitrogen anaerobic condition. Plot shows ⊿amtB in blue, ⊿nifK in orange, ⊿nifLA in green and WT in red, and the metabolites show are (A) Acetoin, (B) Butan-2-3-diol, (C) Formate and (D) Lactate.

4.4 Discussion

The results from this work were broadly in line with expectations and with similar experiments from the literature [30,60,63,153]. The WT strain and mutants derived from this parent strain all grew consistent with expectations and the positive △*amtB* control and negative △*nifLA* and △*nifK* controls worked well. The work in this chapter is among the first to carry out metabolic footprinting on *K. oxytoca* under nitrogen fixing conditions and to characterise a △*amtB* mutant in this species. This work has also shown that *K. oxytoca* can produce Butan-2-3-diol a plant growth promoting metabolite under nitrogen fixing conditions [158].

4.4.1 Growth profiles vary depending on environment

Under nitrogen rich conditions the WT and mutants grew similarly in an exponential manner and no growth penalty associated with deletion mutations for △*amtB*, △*nifLA* and △*nifK* was observed (Fig. 4.2 A & B). This was true for both aerobic and anaerobic high nitrogen conditions, although the WT and mutants grew to a lower OD_{600} in the anaerobic condition. In contrast, the WT and mutants grew as expected, but not in the same manner when grown in nitrogen depleted conditions (Fig. 4.2 C & D). When grown under nitrogen limited aerobic conditions, the WT and mutants reached a similarly low $OD_{600,}$ approximately 0.3, and no nitrogen fixation occurred. This is because the nitrogenase is extremely oxygen sensitive [26], and needs microaerobic or anaerobic conditions to not degrade [26,159]. However, when grown under nitrogen poor anaerobic conditions the WT strain and △*amtB* mutant displayed diazotrophic growth, displaying a secondary phase of growth on fixed atmospheric nitrogen as expected. The △*nifLA* and △*nifK* mutants did not display the secondary phase of diazotrophic growth, presumably due to lack of nitrogenase activity.

The *nifLA* mutant is unable to fix nitrogen because the deletion removes the master regulator of the Nif operon [26,63,153]. This means that NifA is unable to initiate and drive the remaining six operons of the Nif cluster and as such no nitrogenase formation should occur at all. While the △*nifK* mutant is also unable to fix nitrogen it is believed that an incomplete assembly might form that is catalytically inactive for reducing nitrogen to ammonia. This is because *nifA* will drive the other six operons and the *nifH* homodimer should be able to form, whereas the homodimer made from *nifD* and *nifK* could be unable to form or if it does form it will not be able to function properly. While this is in part conjecture, the lack of a secondary growth phase and absence of fixation seen in the ARA goes a long way to supporting this.

The WT strain and △*amtB* mutant were found to be able to fix nitrogen (Fig. 4.2 D and Fig. 4.3), with the △*amtB* mutant fixing more nitrogen at both timepoints measured as well as starting the secondary growth phase earlier compared to the WT. This is likely because of the defective ammonium assimilation caused by the △*amtB* deletion placing the cell under increased nitrogen stress compared to WT [26]. Under nitrogen stress the intracellular levels of glutamine are increased, signalling to the cell that ammonium is limited and initiating the Ntr regulatory cascade. Typically, under nitrogen stress GlnK is uridylated meaning it no longer interacts with AmtB and can disrupt the NifLA complex and start NifA expression [26,59]. As a result of the deletion AmtB, does not interact with GlnK at all and the levels of cytoplasmic GlnK are likely further increased resulting in higher levels of NifLA disruption and NifA expression. The increased levels of *nifA* have been shown to cause increased nitrogenase production and higher levels of nitrogen fixation in other nitrogen fixing species such as *Azotobacter vinelandii* [60,153].

4.4.2 Metabolic changes resulting from fixation

Differences in metabolic profiles between the WT strain, the mutants derived from this parent strain, and growth conditions were not as pronounced as might have been expected. This led me to conclude that a homeostasis occurs during diazotrophy to limit effects of extreme nitrogen starvation on cell function. Alternatively, subtle changes escape detection using the analytical methods employed. The only clear separation between metabolic profiles was between the anaerobic and aerobic growth conditions grown on low nitrogen, this is due to the difference in metabolism required for growth in each respective oxygen condition. Previous studies have found similar differences in metabolic profile resulting from oxygen status in other species of bacteria [160]. With alpha-Ketoglutarate, citrate and isobutyrate accumulating in the aerobic supernatant compared to lactate, ethanol, succinate and formate accumulating in the anaerobic supernatant in this study and one carried out by Sun *et al.,* [160].

Other than the metabolic differences caused by oxygen status, no clear separation was seen based on mutation or fixation status. The lack of separation is seen when comparing all metabolic profiling data, i.e. anaerobic growth conditions with high and low nitrogen to aerobic growth conditions with low nitrogen (Fig. 4.4 A), and when comparing the low nitrogen treatments only (Fig. 4.4 B). While one might expect nitrogen fixation to result in an altered metabolic footprint, the low sensitivity of NMR compared to mass spectrometry means that only broad changes in the supernatant were likely to be detected. Given the low biomass of these cultures, the starting OD_{600} was 0.1, it is possible that the subtle changes in supernatant composition were missed.

While is true that both nitrogen starvation and fixation are thought to have a large impact on global cell metabolism, most of the metabolites that were detected relate to general growth and thus one could expect them be detected in similar levels between strains as they grow. Indeed, work that looked at metabolic footprinting in *Bacillus licheniformis* found that the key factor in differentiating samples was growth [161]. As such, OD_{600} was plotted against the first four components of the PCA analysis and was found to correlate well with PC1 (Fig. 4.4 D). This means that most of the variation explained by PC1 is as a result of growth rather than strain or nitrogen fixation and goes some way to explaining the lack of grouping.

4.4.3 Key metabolites show treatment and WT/ mutant specific differences

While limited changes were seen through multivariate analysis, a focused look using univariate analysis showed differences at the 24-hour timepoint for the anaerobic low nitrogen treatment between the WT and mutants derived from this strain. This is the treatment in which the WT strain and △*amtB* mutant fix nitrogen and is the only one that had a significant differences in metabolite abundance between the WT and mutants. One of the strongest effects on metabolite abundance was for lactate, which as stated earlier is known to accumulate in the supernatant of anaerobically grown bacteria [160,161]. The accumulation of lactate is in line with the respective OD_{600} of the sample, i.e. △*amtB* has the greatest abundance and OD_{600} and △*nifK/*△*nifLA* has the lowest abundance and OD_{600}. While unlikely to be the case here, future studies of longer duration should take into consideration that lactate may inhibit growth if at sufficiently high concentrations [162].

Butan-2-3-diol and its precursor acetoin were also found to have a significant difference in abundance at the 24-hour timepoint between the WT and mutants when grown in anaerobic low nitrogen conditions. The △*amtB* mutant and WT strain had accumulated higher levels of both metabolites in the supernatant, compared to the non-fixing mutants. It was not previously clear if *K. oxytoca* produced butan-2-3-diol under nitrogen fixing conditions [157,163]. △*amtB* had much higher abundance of butan-2-3-diol compared to WT a difference that persists even when accounting for the differences in OD_{600}, the differences between the fixing and non-fixing mutants also persist when accounting for OD_{600} suggesting a possible fixation effect. However, the other analysed treatments also saw both metabolites accumulate in the supernatant, meaning that the accumulation is not solely due to nitrogen fixation but also a result growth of *K. oxytoca* in general [157]. Indeed, butan-2-3-diol production in *K. oxytoca* has been shown previously, however with a focus on commercial production of the metabolite, rather than in relation to nitrogen fixation [125,163]. The WT strain and mutants showed similar amounts of both metabolites accumulate in the supernatant, when grown under unlimited

nitrogen anaerobic conditions across all timepoints; there was greater abundance at the 24-hour timepoint compared to the low nitrogen anaerobic condition. In the low nitrogen aerobic condition, the △*amtB* mutant had again accumulated the most of both metabolites but this was at much lower levels compared to both the high and low nitrogen anaerobic conditions.

These metabolites are of interest in relation to nitrogen fixation because acetoin and butan-2-3-diol have been shown to elicit induced systemic biotic resistance in plants [158,164]. For example, these metabolites have been shown to cause stomatal closure in two plant species (*Arabidopsis thaliana* and *Nicotiana benthamiana)* and this was most efficient through root absorption [164]. Furthermore, a separate study found that plant associative bacteria was best maintained in the plant root rhizosphere when it was a butan-2-3-diol super producer compared to WT and non-producing mutant [165]. Therefore, the possible increase in acetoin and butan-2-3-diol from fixation could mean these metabolites are also shared during free-living plant associative bacteria and plant root interactions. That *K. oxytoca* is able to fix nitrogen and produce these additional plant growth promoting metabolites warrants further investigation into possible chemical exchange between the bacteria and the plant root.

In general, most metabolites changed in similar ways across time between treatments for the WT strain and the mutants derived from the WT strain. However, there was one notable exception to this, Alpha-ketoglutarate which is an indicator of nitrogen stress [47]. The abundance remained constant in the supernatant over time for the high nitrogen anaerobic condition and increased over time for the low nitrogen conditions. For the aerobic condition the increase was greatest in the △*amtB* mutant, whereas in the anaerobic condition the increase was greatest in the non-fixing mutants, particularly △*nifLA.* Conversely, △*amtB* sees the abundance of alpha-ketoglutarate remain relatively constant after an initial increase once nitrogen begins to be fixed. While the WT strain can also fix nitrogen there is still an increase over time, with a jump between the 10 hr and 24-hour timepoint not seen for the △*amtB* mutant. This would suggest that the △*amtB* mutant is able to fix nitrogen at a sufficiently high rate to reduce nitrogen stress on the cell whereas the WT strain is less able to do so. Whether this translates into greater ammonium excretion from the cell is not known from this data but has been shown in other 'super fixers' [55,60,63,107].

4.4.4 Future work: Sensitivity and investigating specific metabolites

The benefit of using an untargeted metabolic profiling approach here was that bias in the types of metabolites detected was minimised. However, the relatively low sensitivity of NMR compared to mass

spectrometry-based techniques means that metabolites in low abundance were likely missed. Additionally, metabolites such as ammonium are not able to be detected by NMR and therefore cannot be included in analysis. As discussed in the Chapter 3, the original intention had been to analyse the supernatant using the LC-MS techniques developed as part of that chapter in conjunction with NMR. This would have given the analysis greater sensitivity and possibly elucidated other changes in the supernatant due to fixation. Therefore future work investigating changes in the supernatant should include either de-salting of the supernatant [166,167] or revision of the medium composition to still allow fixation but with a lower salt concentration. Both approaches would allow for the more sensitive mass spec methods to be employed resulting in more in-depth analysis of supernatant metabolite composition and how *K. oxytoca* interacts with its environment. Alternative medium compositions were investigated when the high salt interreference was discovered and indeed other studies use different recipes, such as B medium [168]. The salt content in B medium was also high, meaning there was no guarantee that it would be any more suited to LC-MS analysis. Perhaps more importantly, the wider research carried out on nitrogen fixation on *K. oxytoca* in the laboratory had been carried out using NFDM such that it's use in this thesis allowed for more direct comparisons between studies [39,41].

It is also worth noting that ammonium production was nowhich measured inwhichs work, which would have been an interesting addition to the results. The LCMS methods developed as part of this thesis in combination with other does include ammonium and had also been shown to differentiate between labelled and unlabelled ammonium [129]. Thus, ammonium would have been measured using this method and been a means of potential further distinction between the △*amtB* mutant and WT strain. Previous work carried out found that many commercially available kits have low specificity for ammonium and other amino acids, such as serine, will also be measured as ammonium (M Carey Masters Thesis). The kits trialled as part of this work were also found to have this interference and therefore were unsuited for use. Instead, ARA results will have to serve as a proxy for ammonium excretion here, but future work should include a measure of ammonium excretion by the WT strain and mutants derived from the WT strain.

More broadly, understanding all the metabolites that could be being exchanged between free living nitrogen fixing bacteria and their environment remains important. In symbiotic systems, the nitrogen fixing *Bradyrhizobium diazoefficiens* was found to have host specific metabolic profiles for specific nodules [70]. Additionally, other research has found that amino acids and osmoprotectants (e.g. trehalose) were also significantly increased in root nodules containing nitrogen fixing symbionts compared to the roots and negative controls [70,74,169,170]. Future efforts should look for other plant

growth promoting metabolites, such as butan-2-3-diol, that free living nitrogen fixers produce either in general growth or under fixing conditions.

One way to better understand the metabolites bacteria might excrete is by studying the cells themselves. This would alleviate some of the methodological issues surrounding the high salt content of nitrogen fixing medium and LCMS analysis, and also provide insights into the cell state and their metabolic phenotype [81,101,154]. Targeted LCMS analysis is arguably the best approach for metabolomic analysis of bacterial cells, allowing pathway analysis to be carried out without having to ID metabolites as when using an untargeted approach [171]. Given that the other amino acids have been implicated with an symbiotic system [70], targeted methods that include them, such as those discussed earlier, would be of high utility in looking for nitrogen fixing phenotypes.

4.5 Conclusion

The results here show that the △*amtB* mutant is a suitable positive control that does not appear to incur a growth penalty when grown under a variety of treatments, including non-limiting nitrogen and carbon conditions and nitrogen fixing conditions. Even though the analyses carried out in this chapter did not explicitly find a nitrogen fixing metabolic footprint, there were some interesting mutant specific results for a select number of metabolites such as butan-2-3-diol and alpha-ketoglutarate. This led me to conclude that the changes fixing bacteria can affect on their environment are likely subtle and at low concentration. The use of both more sensitive and targeted methods could allow for these changes to be detected. Targeted analysis of the cell pellet will be useful to investigate possible similarities and difference between free living and symbiotic nitrogen fixers and how these can be utilised both *in situ* and in synthetic biological approaches.

Chapter 5 Investigation of metabolic changes in *K. oxytoca* cells in batch culture

5.1 Introduction

The metabolic changes that occur within a cell resulting from altered environmental conditions or growth states can have an impact on cell growth and the products of metabolism [47]. This will occur through the complex feed forward and feedback systems operating to control gene expression and enzyme activities, with many of the key effector molecules being metabolites. Some such metabolites have global significance for cell biology, as occurs with carbon and nitrogen metabolism, whereas others are far more limited in their scope of action, for example end point products in some amino acid synthetic pathways. Therefore, it is important to not only consider metabolic footprinting but also what metabolic changes occur within the cells themselves. Metabolic profiling of bacteria provides insights into metabolism that can be used to alter biosynthetic pathways and change outcomes to beneficial ones. For example, studies which investigated the biosynthesis of butan-2-3-diol in *K. oxytoca,* for the potential commercial value, used insights gained from cell carbon metabolism to alter biosynthetic pathways and increase production [125,163]. This shows the importance of metabolic profiling of cells and how the information can form the foundations of biosynthetic efforts. Moreover, that synthetic biological approaches have already been applied to both *K. oxytoca* and other nitrogen fixing bacteria with some success [19,30,46,172], means that information gained will be of high value for future efforts.

As explained in more detail in Chapter 1.4 bacterial metabolism has been shown to vary greatly depending on cell state, e.g. carbon deplete, in batch culture growth versus chemostatic steady states [163,173]. Within the context of nitrogen fixation this has been primarily focussed on either the symbiotic bacteria in batch culture [23] or on root nodules containing bacteria [21,70,169]. This research has revealed an accumulation in a variety of compounds in the root nodules and cells including amino acids and certain sugars [70]. Indeed, work carried out by Waite *et al.,* [39] profiled the amino acids in *K. oxytoca* using a adapted version of the LC-MS method developed in Chapter 3. This work found that histidine accumulated in WT cells and proline accumulated in non-fixing controls. However, the conditions required for symbiotic nitrogen fixation are thought to be stricter compared to free living nitrogen fixers [25]. For example, the carbon concentration needs to be especially high and diversity of carbon sources limited for symbiotic fixers compared to free living ones [21,25]. Therefore, the metabolic changes in the free-living system are likely different to compensate for this greater adaptability and thus metabolic profiling that includes a range of metabolite classes is needed.

It is also important to note that when free living nitrogen bacteria transition to diazotrophic growth and fix nitrogen it is in part triggered and maintained by nitrogen stress [26,39]. Therefore, while the two metabolic states occur simultaneously, nitrogen limited but not fixing, and fixing and nitrogen limited, they could each have their own effect, which needs to be considered in potential biotechnological applications [19]. Targeted analysis that looks at metabolites across a broad range of metabolic pathways could allow for distinctions to be made and allow for useful insights. Additionally, given the important role that the amino acids [70] have in nitrogen fixing root nodules analytical methods with a particular focus on these and other amine group containing metabolites would be of high value and potentially insightful. The AccQ-Tag method discussed in Chapter 3, includes the amino acids in addition to other amine compounds and the IPC method contains a broad range of metabolite classes and combined these methods should prove useful when investigating nitrogen fixing cells in batch culture.

5.1.1 Aims and objectives:

Here the aim is to look at the metabolomic profile of *K. oxytoca* under fixing and non-fixing conditions, and explore how these profiles change across time, using targeted LCMS analysis.

Objectives:

- To investigate the metabolic profiles of *K. oxytoca* cells before, transitioning to and during fixation using targeted LCMS analysis developed in Chapter 3.
- To analyse the integrated data of the cells using multivariate and univariate statistical methods.

5.2 Methods

5.2.1 Bacteria strain, mutant and culturing

The parent strain and mutant of this strain in the table below were used in the experiments carried out in this chapter.

Table 5.1 List of strain and mutant used.

Strain name	Gene knockout	Summary
Wild type (WT) M5a1	None	Has a normal growth phenotype and is able to fix nitrogen
△*amtB*	△*amtB*	Able to fix nitrogen and exhibits diazotrophic growth; AmtB is necessary to regulate ammonium levels in the cell through action of the AmtB ammonia transporter.

Bacterial strains were grown as outlined in Chapter 2.4.1 'Liquid culture experiments'.

In brief, the WT strain was anaerobically grown for up to 24 hours in minimal NDFM medium supplemented with either 0.5 mM (low nitrogen) or 10 mM (high nitrogen) ammonium chloride. Ammonium run out was used to induce diazotrophic growth, and for the low nitrogen WT condition sampling of bacterial cells was carried out at 3.5-hour (non-fixing), 12.5-hours (early fixation) and 24-hours (fixing). For the high nitrogen WT condition cells were samples at the 3.5-hour and 12.5-hour timepoints only, as work carried out in Chapter 4 suggests cells may not be in ammonium replete conditions after 24-hours; they are no longer growing exponentially but still carbon replete. The △*amtB* mutant was grown anaerobically on low nitrogen (0.5 mM) and sampled at the 12.5-hours as a positive control and reference for comparison to the three low nitrogen timepoints collected. Each condition was carried out with 6 replicates and sacrificial sampling. This meant that 6 overnight cultures were used and that each culture was used to set up a replicate for all timepoints of the high and low nitrogen treatments. Sacrificial sampling also had the advantage of limiting oxygen exposure and minimising possible changes in cell metabolism due to this.

5.2.2 Timepoint selection

Earlier detailed growth curves carried out in Chapter 4 show that nitrogen run out on low nitrogen occurs between 3-4 hours therefore the 3.5-hour timepoint will show changes occurring during run out in the low nitrogen treatment but still have replete conditions in the high nitrogen condition. The 12.5-hour timepoint was selected because supernatant growth curves show that cells are typically in an early fixation phase where they are fixing nitrogen but may not yet growing therefore, they appear to be distinct from the late fixation phenotype. An ARA was also carried out at at 12.5- and 24-hour timepoints for all treatments to confirm whether nitrogen fixation was occurring.

Finally, the 24-hour timepoint was selected because previous data shows that *K. oxyotca* is still fixing nitrogen after 24-hours with previous unpublished work in the Buck laboratory indicating that cells can grow beyond 24-hours. Also, the NMR supernatant analysis from Chapter 4 shows that approximately half of the initial glucose is still available in the medium at the 24-hour timepoint meaning carbon is not limiting.

5.2.3 Sampling

The sampling method used in this chapter is the one outlined in Chapter 3. In brief, this involved centrifugation of the batch cultures followed by removal of the supernatant, quenching and extraction of cells using methanol, chloroform and water. The extracted samples were then aliquoted, dried down and stored at -80^{o}C for analysis.

5.2.4 LCMS analysis

The LC-MS carried out in this chapter implements the AccQ-Tag RPLC and ion pairing chromatography (IPC) methods outlined in Chapter 3. These methods cover a range of metabolites and were selected as the optimum pairing for a targeted assessment of the metabolome [129]. In total 121 metabolites were detected with 119 of these deemed to be of sufficient quality for analysis.

5.2.5 Data Analysis

Peaks were integrated in Skyline (MacCoss Lab software, US) and normalised to biomass using OD_{600}. Data were analysed using in house Python (Version 3.8.13) or R Studio (R version 4.2.1, Boston, USA) scripts. R Studio packages primarily included ‘mixOmics’, ‘pls’, ‘factoextra’, ‘tidyverse’, ‘viridis’ and ‘ggplot2’. Example Python and R Studio scripts can be seen in Appendix 2. Where applicable, univariate

statistics, e.g. T-Test/Mann-Whitney U test were corrected for FDR using the Benjamini-Hochberg method. Pathway analysis was carried out on metabolites with significantly different fold-changes between treatments using MetaboAnalystR Package [174,175].

5.3 Results

5.3.1 Multivariate analysis showed clear differences in the intracellular metabolite profiles between all treatment groups – timepoint, nitrogen status and strain

The PCA scores plot below (Fig. 5.1 A) shows that the metabolic profiles of the high nitrogen, non-fixing, treatments are distinct from both the low nitrogen treatments and each other. The high nitrogen 3.5-hour time point appears the most distinct grouping possibly due to the exponential growth of the cells at that time. Additionally, there is grouping of treatments within the low nitrogen condition, between those fixing nitrogen, WT *K. oxytoca* at 24-hours and △*amtB* at 12.5-hours, and those not yet fixing nitrogen or transitioning to fixation, WT at 3.5- and 12.5-hours. The WT low nitrogen 12.5-hour samples appear to show a spread, with some are closer to the fixing samples, and others are closer to the non-fixing 3.5-hour low nitrogen WT samples.

The grouping of the WT low nitrogen 24-hour samples and the low nitrogen 12.5-hour △*amtB* samples is closest. This is true of both within treatment and between treatment grouping for these two treatment groups. The OPLS-DA scores plot (Fig. 5.1 B), shows separation between the WT 24-hour △*amtB* 12.5-hour treatments. Although a relatively low amount of variation is explained, this model has a high predictive power and significant variation between these treatments was found. Furthermore, these fixing treatments fall between the nitrogen stressed, low nitrogen early timepoints and the high nitrogen 12.5-hour timepoint along PC2 on the PCA scores plot (Fig. 5.1 A). In contrast PC1 appears to separate the WT high nitrogen 3.5-hour treatment from all other treatments. As such, these results indicate the existence of marked differences between fixing and non-fixing metabolic profiles.

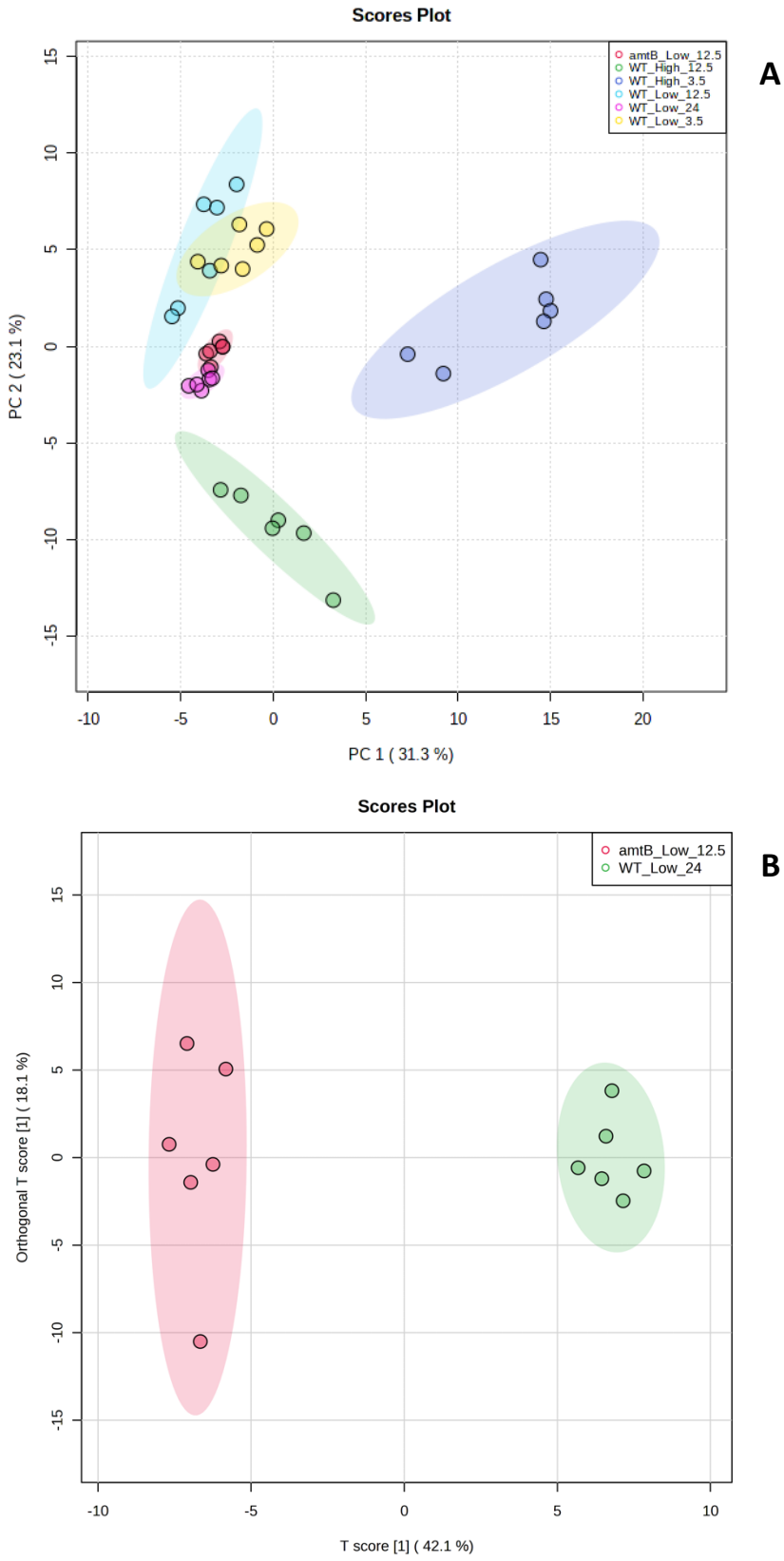


Figure 5.1 (A) PCA scores plot showing the metabolic profiles of WT and ⊿amtB mutant grown under a range of conditions. Ellipses represent 95% confidence interval. (B) Cross validated OPLS0DA scores plot comparing the metabolic profiles of WT at 24-hours and ⊿amtB at 12.5-hours grown under anaerobic nitrogen limited conditions, where fixation is occurring. R2X = 0.421, R2Y = 0.956, Q2 = 0.928, p <= 0.01, n = 6. Ellipses represent 95% confidence interval.

5.3.2 Key nitrogen status metabolites show significant differences due to time, nitrogen status and strain

Of the 119 metabolites analysed 86 were shown to have a time effect on metabolite abundance for the WT low nitrogen when using one-way Kruskal-Wallis test. Two key nitrogen status metabolites were included in this list, as time was found to have a significant effect on the abundance of glutamine and glutamate. Additionally, glutamate and glutamine had an inverse relationship across time. Glutamate increased non-significantly between the 3.5-to-12.5-hour timepoint before a large decrease to its lowest abundance at the 24-hour timepoint (Fig. 5.2 A). In comparison, glutamine levels decreased between 3.5-to-12.5-hour timepoint before increasing to the greatest abundance at the 24-hour timepoint (Fig. 5.2 B).

Additionally, there were significant differences between intracellular glutamine and glutamate levels when comparing between timepoints for the low and high nitrogen conditions. At the 3.5-hour timepoint, the abundance of glutamate and glutamine was significantly different between WT *K. oxytoca* grown on high and low nitrogen (adjusted p = 0.00379 and 0.0016 respectively, Mann-Whitney U test with FDR correction). Figure 5.2 C & D, shows that that *K. oxytoca* grown on high nitrogen at 3.5-hours had a greater intracellular abundance of glutamate and glutamine compared to not only the low nitrogen treatment at 3.5-hours, but all treatments. A similar occurrence is seen when comparing WT cells grown under high and low nitrogen conditions after 12.5-hours, with glutamate and glutamine abundance differing significantly (adjusted p = 0.0294 and 0.0411 respectively, Mann-Whitney U test with FDR correction). However, the difference in abundance of glutamate and glutamine between WT cells grown under high and low nitrogen conditions is smaller at the 12.5-hour timepoint compared to the 3.5-hour timepoint (Fig. 5.2 C&D).

When comparing the WT low and the △*amtB* low treatments at 12.5 hours, glutamate was not found to differ significantly. However, glutamine abundance was significantly different between the two treatments, with △*amtB* having the greater abundance (adjusted p = 0.379 and 0.0453 respectively, Mann-Whitney U test with FDR correction, Fig. 5.2 C & D).

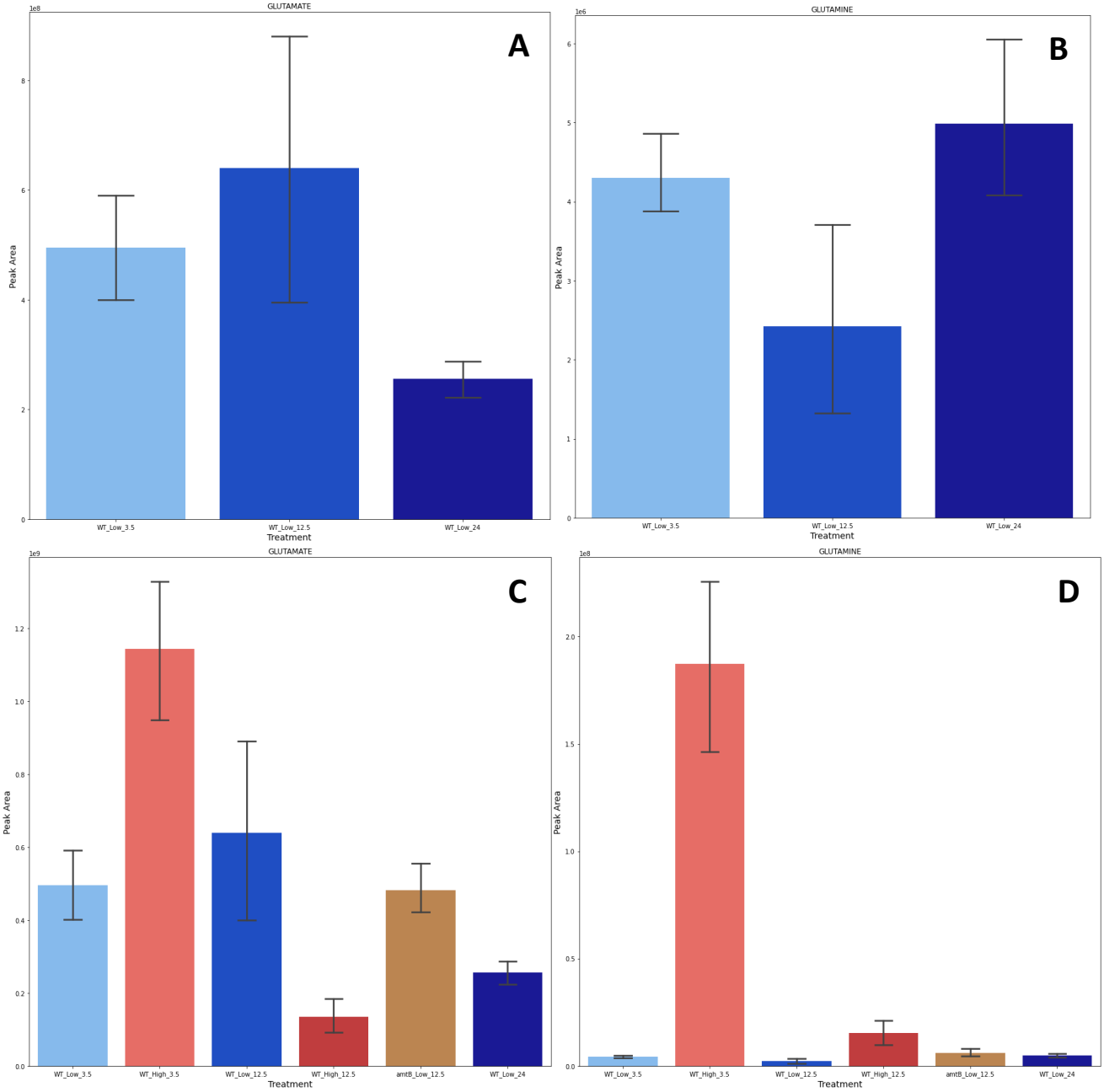


Figure 5.2 Comparison of metabolite abundance; A-B for all timepoints of the low nitrogen condition for (A) glutamate and (B) glutamine. C-D for all treatments and timepoints for (C) glutamate and (D) glutamine, n=6.

5.3.3 Intracellular amino acid abundance varies across time in WT cells grown under fixing conditions

Like glutamine and glutamate discussed above the rest of the amino acids were found to vary significantly overtime via Kruskal-Wallis test (FDR corrected p = <0.05). The three most abundant amino acids; alanine, proline and glutamate follow a similar trend across time, with abundance peaking at 12.5-hours compared (Fig. 5.3 A). Valine which is the amino acid with the next highest abundance, peaks at 3.5-hours and decreases across the 12.5- and 24-hour timepoint. Comparing the other amino acids without the most abundant four in Figure 5.3 B, we see the threonine and arginine are also found in the greatest abundance at the 12.5-hour timepoint. The remaining amino acids decrease over time such as glycine, asparagine and tryptophan, although this effect plateaus for histidine between 12.5- to the 24-hour timepoint. That the amino acids behave differently and approximately conform to 'groups', shows the metabolic changes that the WT *K. oxytoca* undergo during transition from nitrogen depletion to fixation. Although not an amino acid, ammonium is also included in Figure 5.3 A & B because its ammonium assimilation. It's intracellular abundance will strongly impact the abundance of glutamine and glutamate, due to its role in the GS-GOGAT and GDH nitrogen assimilation pathways discussed in Chapter 1. The intracellular abundance of ammonium peaks 12.5-hours, following the same trend as glutamate abundance.

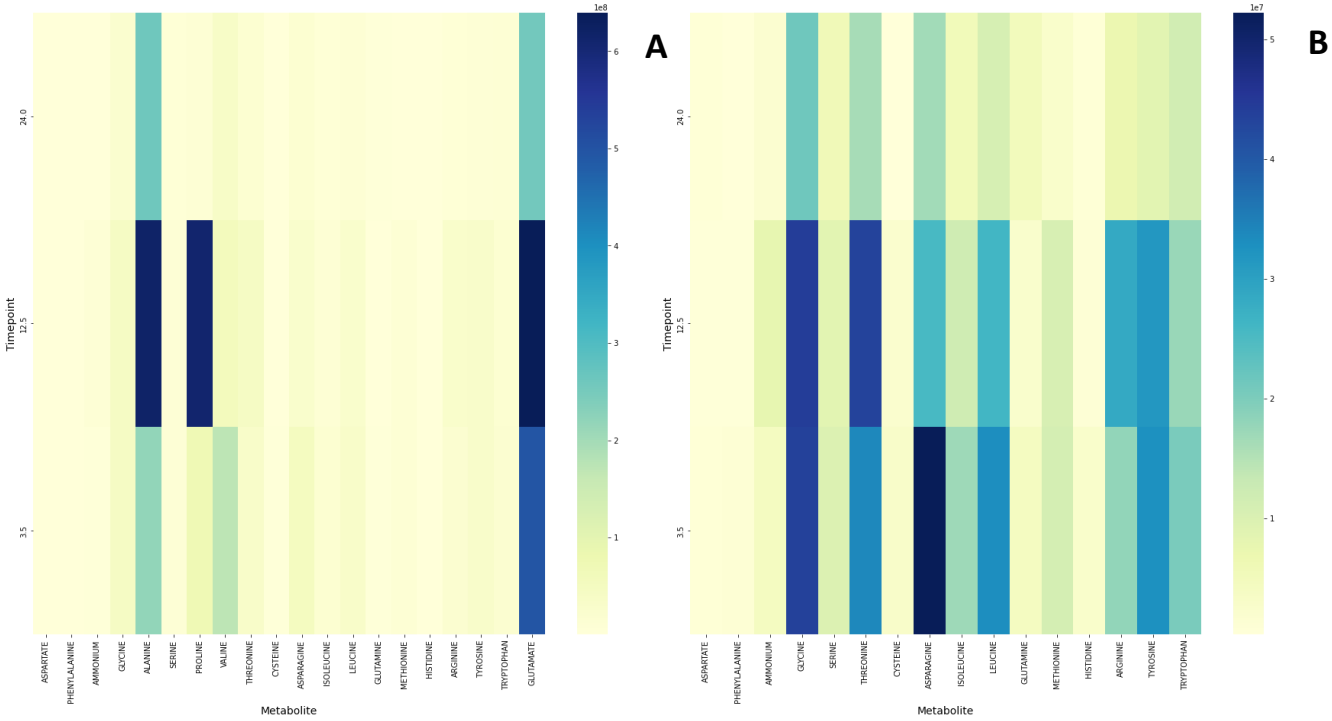


Figure 5.3 Comparison of intracellular amino acid abundance across time of wild-type K. oxytoca. (A) All detected amino acids (B) With four most abundant amino acids excluded. Heat map density representative of average abundance at a given timepoint. All amino acids were found to undergo significant changes in abundance across time (one sided Kruskal-Wallis test, n = 6, p < 0.05).

5.3.4 Metabolic profiles show clear differences in univariate analysis

Given the grouping from PCA analysis it is unsurprising that univariate analysis shows a trend of large differences in metabolite composition between treatments both across time and between either nitrogen status or strain. Here I will discuss differences between specific treatments and include pathway analysis that shows the increased and decreased metabolite pathways. Given the small sample size of each treatment group, a trade-off is necessary to include all groups, the pathway analysis should be considered tentative and descriptive in nature.

Comparison of WT low nitrogen 3.5-hours to WT low nitrogen 12.5-hours

34 metabolites were significantly differentially abundant between the WT low nitrogen 3.5- and 12.5-hour timepoints; this number was reduced from 45 pre-FDR correction. The four most significant metabolites were NADP (p =0.000160), NAD (p =0.000193), Uridine (p =0.00257) and Malate (p =0.00262), which were all lower at the 3.5-hour timepoint compared to the 12.5-hour timepoint (Fig. 5.4). While none were found to be significantly different, results of the pathway analysis led me to conclude that arginine and proline, and nicotinate and nicotinamide metabolism are reduced in the WTL 3.5 hr cells. In comparison, the purine metabolism, arginine biosynthesis and pyrimidine metabolism pathways are increased in the WTL 3.5-hour cells compared WTL 12.5. To be significantly increased or decreased a metabolite pathway would have to have a *p* value of <=0.05 post FDR correction.

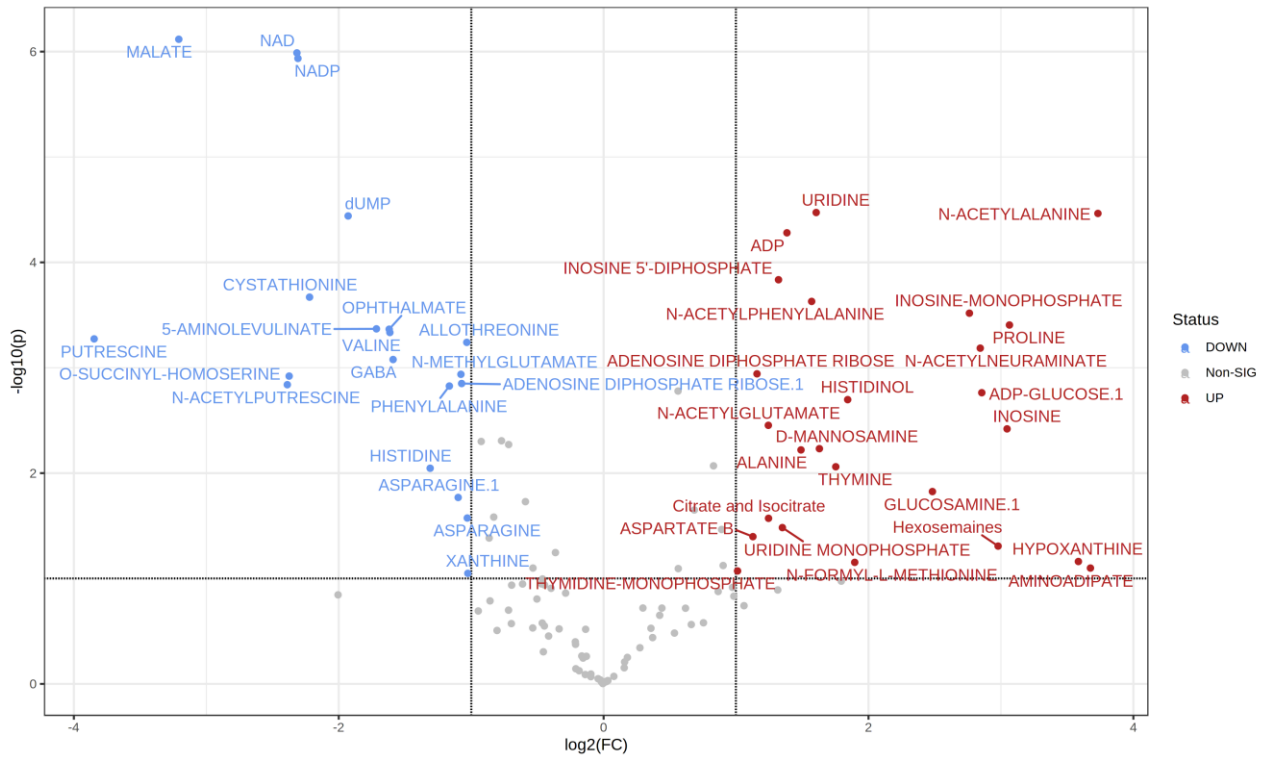


Figure 5.4 Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 3.5-hour samples compared to WTL 12.5-hours. Metabolites significantly less abundant at the 3.5-hour timepoint in blue and those significantly more abundant in the 3.5-hour timepoint compared to 12.5-hour timepoint in red. Metabolites that were not different between treatments in grey.

Comparison of WT low nitrogen 3.5-hours to WT low nitrogen-24 hours

As was suggested in the multivariate analysis, a greater number of metabolites were significantly different between the 3.5- and 24-hour timepoints than between the 3.5- and 12.5-hour timepoints (71 compared to 34). The most significantly different metabolites are reduced glutathione (2.28E-06), N-acetylglutamate (0.005), 5-Aminolevulinate (0.005) and ATP (0.005) (Fig. 5.5). The amino acids threonine, homoserine, leucine, methionine, proline, valine, cysteine and ammonium are significantly more abundant at the 3.5-hour timepoint compared the 24-hour timepoint and have a fold change of >=1. This means that they decrease in intracellular abundance across time.

The pathways that were close to being significantly decreased at the 3.5-hour timepoint compared to the 24-hour timepoint are purine metabolism, pyrimidine metabolism and arginine biosynthesis. In addition to being non-significant and had few contributing metabolites to each pathway. However, the pantothenate and CoA biosynthesis pathway was found to be significantly increased in cells at the 3.5-hour timepoint compared to the 24-hour timepoint ($p = 0.032$, 5/24 metabolite hits, Fig. 5.6). The metabolites contributing to this pathway included the amino acids valine and cysteine as well as CoA itself. Additionally, the amino acid related pathways of Valine, leucine and isoleucine biosynthesis (4/22 metabolites) and cysteine and methionine metabolism (5/40 metabolites) were both significant pre-FDR correction but not post-FDR correction. This is likely due to the small sample size and low number of metabolites detected from these pathways.

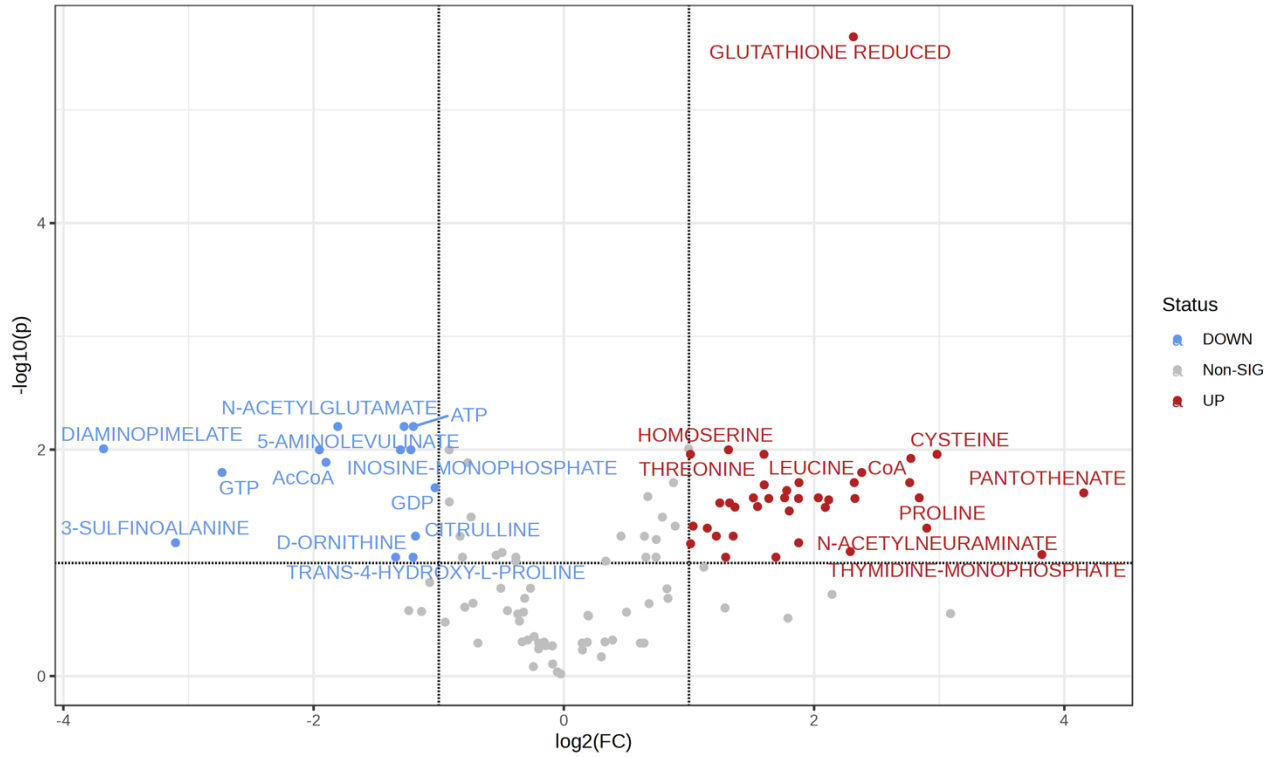


Figure 5.5 Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 3.5-hour samples compared to WTL 24-hours. Metabolites significantly less abundant at the 3.5-hour timepoint in blue and those significantly more abundant in the 3.5-hour timepoint compared to 24-hour timepoint in red. Metabolites that were not different between treatments in grey.

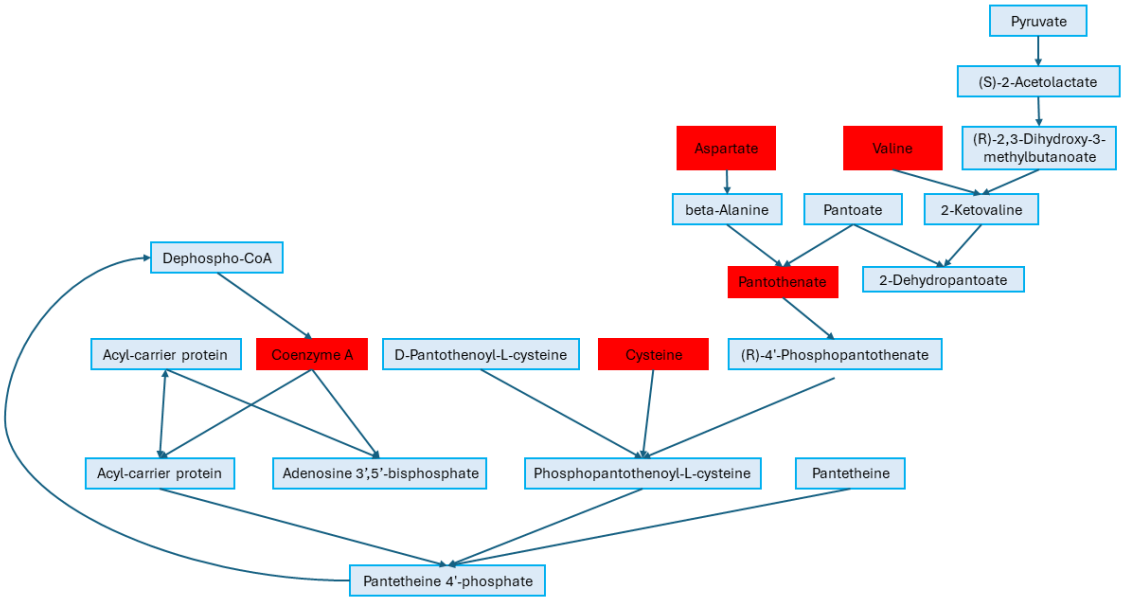


Figure 5.6 Pantothenate and CoA biosynthesis pathway for K. Oxytoca. Metabolites in red are those that were significantly more abundant in WTL 3.5-hour treatment compared to WTL 24-hour treatment and had a foldchange of =>1, those in blue had either no change or were not analysed.

Comparison of WT low nitrogen 12.5-hours to WT low nitrogen 24-hours

As with the 3.5-hour timepoint, there are many differences between metabolic profiles of the 12.5- and 24-hour timepoints. 46 metabolites were significantly differentially abundant between these timepoints, including several amino acid and TCA cycle related metabolites (Fig. 5.7). The metabolites most significantly different are all more abundant in the 24-hour timepoint and include NAD (6.39E-06), Malate (6.39E-06), AcCoA (9.08E-05) and 5-Aminolevulinate (0.000106). Additionally, glutamine was both significantly different but also has a fold change greater than one, being more abundant at the 24-hour timepoint. Whereas, other amino acids including, proline alanine and leucine were significantly more abundant at the 3.5-hour timepoint.

With regards to pathway analysis between these two treatments, only one was found to be significant post FDR correction; this was the aminoacyl-tRNA biosynthesis, which was decreased in *K. oxytoca* after 24-hours (p = 3.01E-08, 13/45 metabolites). However, this pathway is low impact, and the 13 metabolites detected in this pathway are all at the start of the metabolic pathway. Biologically important metabolites specific to this pathway including L-Seryl-tRNA, involved in cell wall synthesis, that were not measured. Therefore, it is difficult to say whether this is 'real' and will not be discussed

further. Two pathways were significantly decreased at 24-hours compared to 12.5-hours pre-FDR correction. These were glutathione metabolism (4/22 metabolites) and pantothenate and CoA biosynthesis (4/24 metabolites). That the pantothenate and CoA biosynthesis was close to being significantly downregulated at 24-hours compared to 12.5-hours and that this pathway was significantly downregulated at 24-hours compared to the 3.5-hour timepoint indicates a trend in the abundance of these contributing metabolites over time. No pathways were significantly upregulated at the 24-hour timepoint compared to the 12.5-hour timepoint.

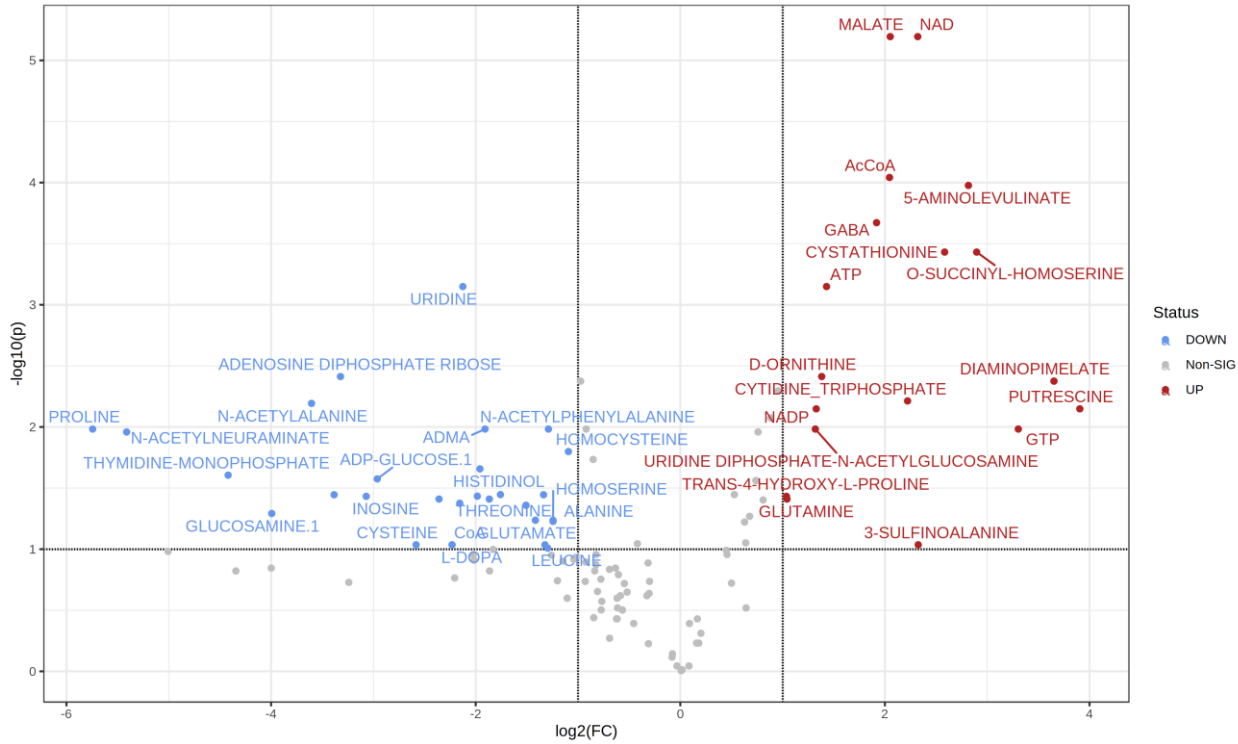


Figure 5.7 Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 24-hour samples compared to WTL 12.5-hours. Metabolites significantly less abundant at the 24-hour timepoint in blue and those significantly more abundant at the 24-hour timepoint compared to 12.5-hour timepoint in red. Metabolites that were not different between treatments in grey.

Comparison of WT low nitrogen 3.5-hours to WT high nitrogen 3.5-hours

There were large differences in the metabolic profiles between the high and low nitrogen treatments at 3.5-hours, with 81 metabolites found to be significantly different between these treatments (Fig. 5.8). In general, most metabolites were at lower intracellular abundance in the low nitrogen treatment compared to the high nitrogen treatment. Notable exceptions to this trend include GABA, malate and phenylalanine which were more abundant in the low nitrogen treatment. The fold changes between these treatments are the biggest seen in this study. The most significantly different metabolites are CDP ($p = 0.000250$), ADP-Glucose ($p = 0.000642$), D-ornithine ($p = 0.000642$) and ADMA ($p = 0.000764$).

Several pathways were significantly (adjusted $p <= 0.05$) downregulated in *K. oxytoca* in the low nitrogen treatment. These being the pathways for alanine, aspartate and glutamate metabolism ($p = 0.0085$, 6/22 metabolites, impact = 0.657), arginine biosynthesis ($p = 0.0102$, 5/16 metabolites, impact = 0.372) and purine metabolism ($p = 0.023$, 8/73 metabolites, impact = 0.253). Due to their central role in cell function these pathways also have a big impact score (out of a maximum of 1). No pathways were significantly upregulated in the low nitrogen treatment at the 3.5-hour timepoint.

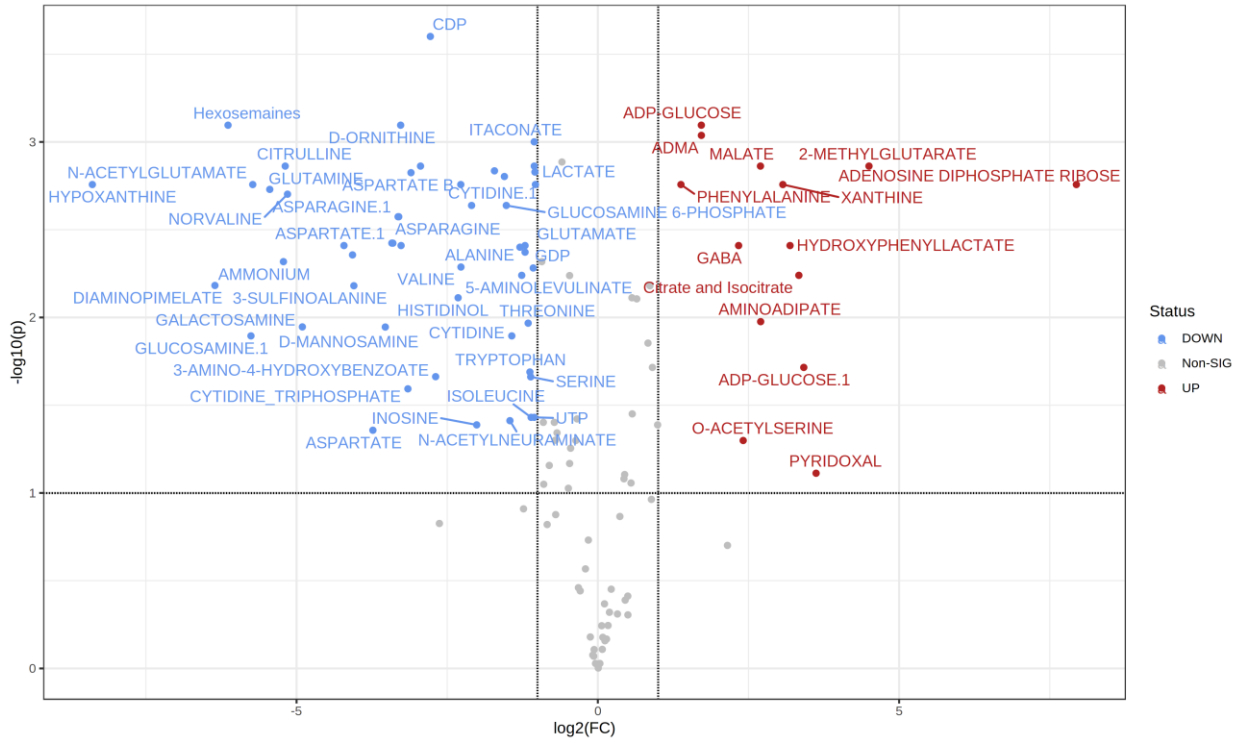


Figure 5.8 Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 3.5-hour samples compared to WTH 3.5-hours. Metabolites significantly less abundant at the low 3.5-hour timepoint in blue and those significantly more abundant at the low nitrogen 3.5-hour timepoint compared to high nitrogen 3.5-hour timepoint in red. Metabolites that were not different between treatments in grey.

Comparison of WT low nitrogen 12.5-hours to WT high nitrogen 12.5-hours

Many differences were also found between the WT low nitrogen and WT high nitrogen condition at the 12.5-hour timepoint, at 62 metabolites, this is lower than the number of differences at 3.5 hours (Fig. 5.9). The most significant differences were metabolites not found between other treatments GMP ($p = 0.00157$), methionine Sulfoximine ($p = 0.00216$), O-succinyl-homoserine ($p = 0.00216$) and xanthurenate ($p = 0.00235$). However, unlike the comparison at the 3.5-hour timepoint there is a more even spread of metabolites across the volcano plot (Fig. 5.9).

Purine metabolism was significantly upregulated in the high nitrogen treatment compared to the low nitrogen treatment ($p = 0.0170$, 9/73 metabolites, impact = 0.14). Several pathways relating to amino acid metabolism were also upregulated in the high nitrogen treatment pre-FDR correction, including those related to alanine, glutamate, arginine, proline and methionine, which could suggest a trend of amino acid utilisation. No pathways were significantly downregulated in the high nitrogen treatment compared to the low nitrogen treatments.

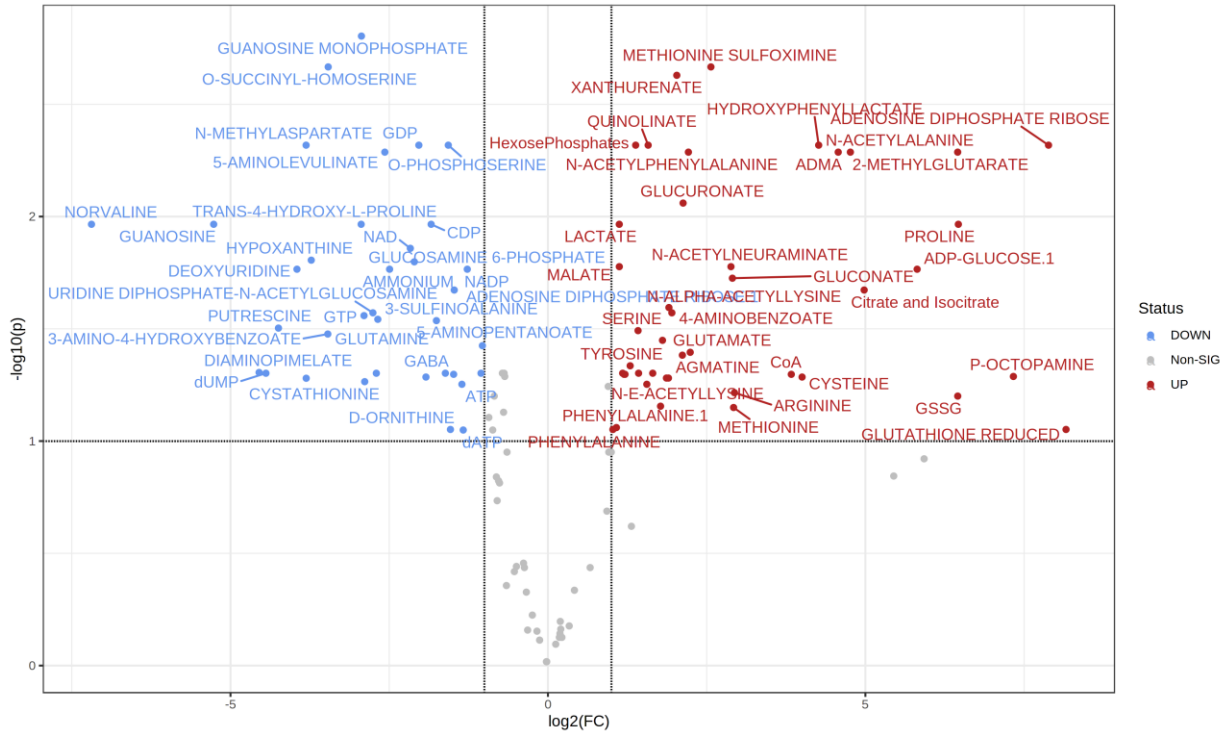


Figure 5.9 Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 12.5-hour samples compared to WTH 12.5-hours. Metabolites significantly less abundant at the low 12.5-hour timepoint in blue and those significantly more abundant at the low nitrogen 12.5-hour timepoint compared to high nitrogen 12.5-hour timepoint in red. Metabolites that were not different between treatments in grey.

Comparison of WT low nitrogen 12.5- and 24-hour timepoints to △amtB at 12.5-hours

The univariate comparison between the two 'fixing' WT timepoints, of 12.5- and 24-hours, and the △*amtB* 12.5-hour timepoint support the results from the univariate analysis. There are more significant differences between the 12.5-hour WT and △*amtB* treatments than between the 24-hour WT treatment and △*amtB* treatments, 47 and 23 respectively (Fig 5.10 A & B). The four most significantly different metabolites between the 12.5-hour WT and △*amtB* treatments are malate (p = 2.74E-06), AcCoA (p = 0.000190), UDP-acetyleglucosamine (p = 0.00377) and NAD (p =0.000385), which were all more abundant in the △*amtB* treatment. Three of these metabolites are the same as the most significantly different between the low nitrogen WT 12.5- and 24-hour treatments. The other metabolite most different between the WT treatments, 5-aminolevulinate, is amongst the most different between WT and △*amtB* at 12.5-hours, further highlighting the similarity between the △*amtB* mutant at 12.5-hours and the WT at 24-hours under low nitrogen conditions. The metabolites with the most significant differences between △*amtB* at 12.5-hours and WT at 24-hours were cyclic GMP (p = 0.00206), N-acetyl alanine (p = 0.00325), citrulline (p = 0.00325) and aminoadipate (p = 0.00325).

No pathways were significantly different between the WT and △*amtB* mutant post-FDR correction when comparing the 12.5-hour timepoints. However, pre-FDR correction the pathways for both arginine biosynthesis and metabolism were increased in △*amtB* cells, as well as pathways relating to methane, proline, lysine and butanoate metabolism. No pathways were significantly different pre- or post-FDR corrections between △*amtB* and the WT at 24-hours.

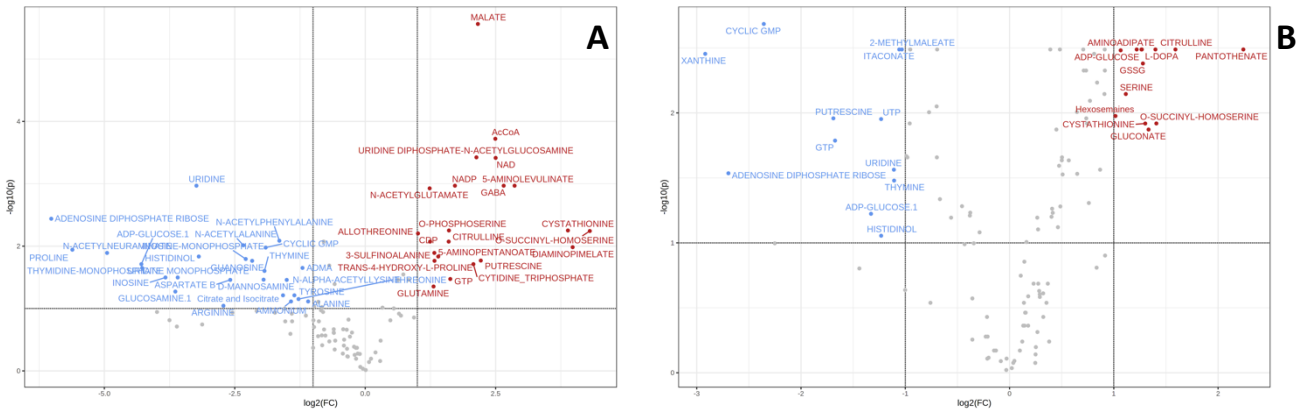


Figure 5.10 A) Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 12.5-hour samples compared to WTH 12.5-hours. Metabolites significantly less abundant at the low 12.5-hour timepoint in blue and those significantly more abundant at the low nitrogen 12.5-hour timepoint compared to high nitrogen 12.5-hour timepoint in red. Metabolites that were not different between treatments in grey. B) Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 12.5-hour samples compared to WTH 12.5-hours. Metabolites significantly less abundant at the low 12.5-hour timepoint in blue and those significantly more abundant at the low nitrogen 12.5-hour timepoint compared to high nitrogen 12.5-hour timepoint in red. Metabolites that were not different between treatments in grey.

5.3.5 Some metabolites appear to be key differentiators between fixing and non-fixing phenotypes

While there are many metabolites that differentiate the metabolic profiles of the three low nitrogen timepoints, some appear to contribute to the nitrogen fixation phenotype more than others. For example, acetyl-CoA (Ac-CoA) and 5-aminolevulinate decrease slightly between 3.5- to 12.5-hours and then accumulate to their highest abundance between the 12.5- to 24-hour timepoint (Fig. 5.11 A&B). Further, the intracellular abundance at 24-hours in the WT is similar to the abundance at 12.5-hours in △*amtB*. These metabolites were found to be amongst the most differentiating between all three timepoints being amongst the most significant differences between the non-fixing 3.5-hour timepoint and fixing 24-hour timepoint.

Previous work has shown that both malate and N-acetyl-glutamate have some role in fixation and symbiosis for rhizobacteria [176,177]. The metabolites were found to differ significantly across time in the low nitrogen treatments ($p = 3.78E-07$ and $p = 2.88E-05$ respectively). The abundance of malate followed a similar pattern to that of Ac-CoA and 5-aminolevulinate apart from the greatest abundance was at the 3.5-hour timepoint, rather than at 24-hours (Fig. 5.11 C). However, the abundance of n-acetyl-glutamate increases in a near linear fashion across time (Fig 5.11 D). This is one of the only metabolites to display this trend for the low nitrogen treatments. While one might think that n-acetyl-glutamate may accumulate in the cells with growth, there is a large decrease in intracellular abundance between the 3.5- and 12.5- hour timepoints in the high nitrogen treatments. This suggests that this metabolite does not just accumulate in the growing cells and that the accumulation in the low nitrogen condition could be due to fixation occurring in the cells.

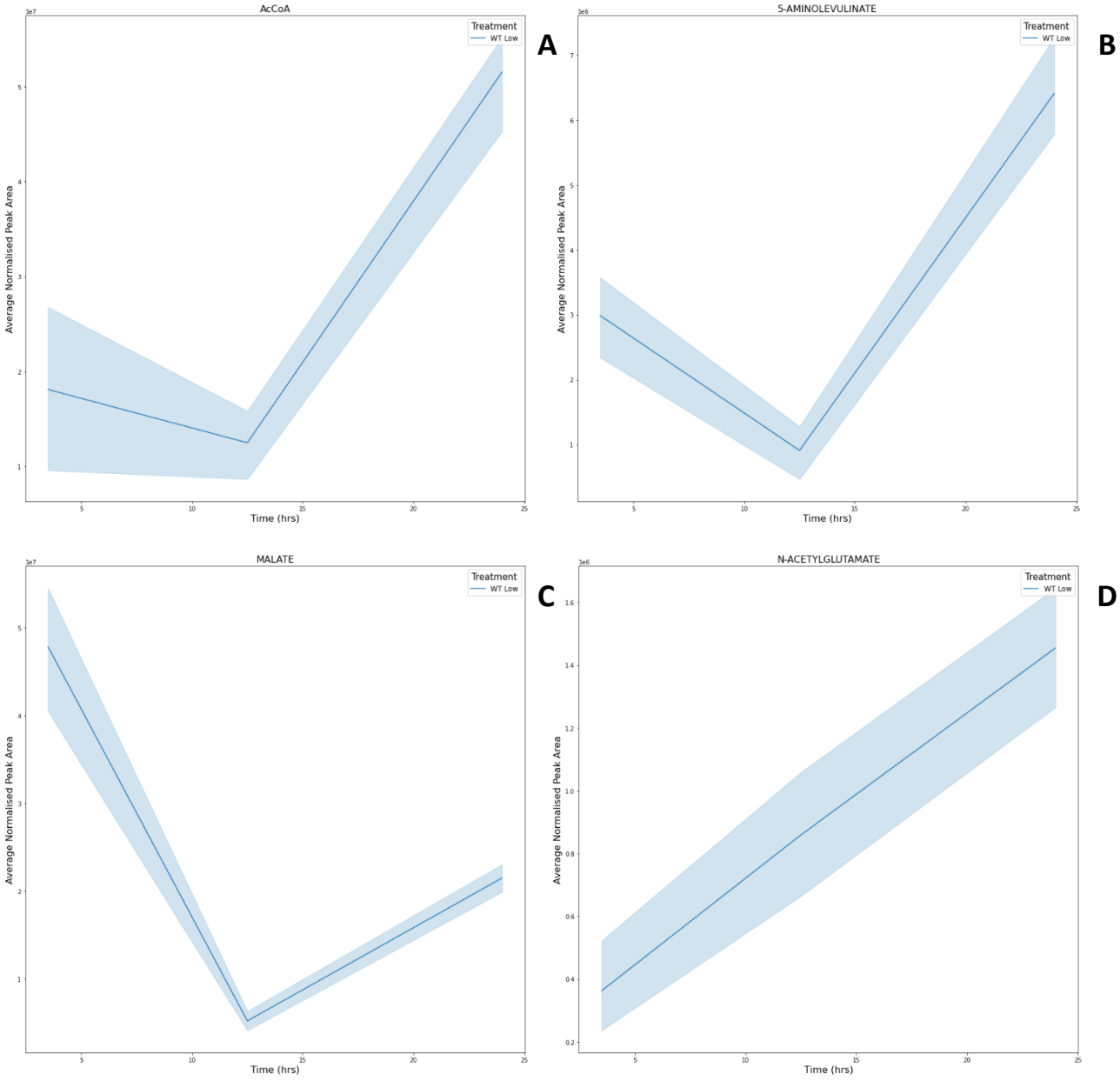


Figure 5.11 Changes in intracellular metabolite abundance over time for K. oxytoca grown under low nitrogen conditions, for (A) Ac-CoA, (B) 5-aminolevulinate, (C) malate and (D) N-acetylglutamate. Lines represent group mean and shaded areas denote 95% confidence bands, n = 6.

5.4 Discussion

To the best of my knowledge, the work carried out in this chapter is a considerable addition to the only published data investigating intracellular metabolite changes resulting from nitrogen fixation in *K. oxytoca*, in a targeted manner [39]. This study utilised the well tested method of nitrogen run out to induce nitrogen fixation in a reliable, non-intrusive manner and induce a metabolic response in the cells [49,178,179]. This approach has the advantage of allowing comparison between nitrogen statuses as well as the consumption based metabolic changes to be investigated, which medium replacement and continuous culture do not allow [64,180]. *K. oxytoca* cells were sampled 3 times over a 24-hour time period at times known to align with ammonium run out, transition to diazotrophy and established nitrogen fixation. This allowed for distinct metabolic phenotypes to be uncovered between fixing and non-fixing cells. Furthermore, the exploratory pathway analysis carried out highlights a trend of nitrogen scavenging, for example purine metabolism, for cells grown in nitrogen deplete conditions. These results are consistent with early research looking at the changes in gene expression *E. coli* undergo under nitrogen limited conditions [181].

5.4.1 Distinct metabolic phenotypes seen between fixing and non-fixing *K. oxytoca*

There is extensive evidence in the literature showing the genetic changes caused by transitioning to diazotrophy including differences in the transcriptome and proteome [39,41,180,182]. It was believed that these changes would translate into differences at the metabolite level as these are the end points of metabolism and represent the downstream consequences of the genetic changes. The results in this chapter confirm that there are profound differences in metabolite composition as *K. oxytoca* cultures transition from ammonium run out to established nitrogen fixation. What's more, cultures fixing nitrogen do not return to a metabolic state comparable to nitrogen replete conditions, as their metabolite profiles are distinct from cells growing in nitrogen replete conditions. Therefore, these results support earlier findings that fixing cells remain in a state of nitrogen stress (needed for continued expression of GS and for activation of *nif* gene expression) despite fixing sufficient nitrogen to support growth [39]. That the intracellular abundance of ammonium at the 24-hour timepoint is comparable to that of the 3.5-hour timepoint further supports this notion.

The metabolite profiles of the WT strain at 24-hours and the △*amtB* mutant at 12.5-hours were remarkably similar. Providing further proof of △*amtB* knockouts being both suitable positive controls but also a possible candidate for biotechnological solutions [60,107,153]. Indeed, one study found greater accumulation of ammonium in the supernatant for an △*amtB* knockout compared to WT in *Azotobacter vinelandii* [153]. Given its reduced ability to actively transport ammonium back into the cell,

the △*amtB* mutant used in this work will be under greater nitrogen stress compared to the WT, hence it's earlier transition to diazotrophy and apparent greater capacity for nitrogen fixation. However, that the metabolite profiles of the 24-hour WT and 12.5-hour △*amtB* treatments are so similar suggests that given sufficient time the WT strain could reach the same fixing capacity as the △*amtB* mutant. Or put another way, the △*amtB* mutant reaches a state of higher nitrogen stress faster meaning that it simply generates the nitrogenase enzyme earlier compared to the WT. It may not mean that △*amtB* deficient cells have a greater abundance of the nitrogenase or can fix any more nitrogen than the WT for the same level of nitrogen stress, rather they enter diazotrophy earlier on in time..

5.4.2 Intracellular amino acid and ammonium abundance plays a key role in differentiating metabolite profiles for *K. oxytoca* grown in fixing conditions

The changes in amino acid abundance for cells grown under low nitrogen conditions as seen in this chapter are broadly similar to those already published in the literature for free living and symbiotic species [39,180,183,184], with abundance vary significantly across the three timepoints. The decrease in intracellular glutamine abundance between the 3.5- and 12.5-hour timepoint is similar to data from *E. coli* and *K. oxytoca* [39,178,183]. Given that glutamine is used up in the GS-GOGAT pathway as an amino group donor, and thus necessary for assimilation of nitrogen, it is logical that this would decrease between these timepoints as nitrogen becomes depleted [47]. These results also show that intracellular glutamine increases once cells have been fixing nitrogen for a prolonged period, a result supported by the △*amtB* positive control. Perhaps these results suggest a small reduction in nitrogen stress as cells switch to growth on fixed atmospheric nitrogen. Although to be clear, the abundance glutamine is still orders of magnitude lower than for cells in nitrogen replete conditions and further indicates ongoing nitrogen stress is required to maintain nitrogen fixation.

Conversely, glutamate levels remained relatively similar between the 3.5- to 12.5-hour timepoints, which agrees with previously published data [178,183,185] but contrasts earlier research from this laboratory [39]. Although given that changes in glutamate abundance were not measured between 0-3.5 hours here and that the decrease in glutamate levels were seen primarily between 0-2.5-hours in that research, it is possible that the decrease was missed. Further, a decrease in intracellular glutamate abundance was observed between the 12.5- and 24-hour timepoints which does contradict other research that has looked at differences between these timepoints [183]. Combined this meant the ratio of glutamate to glutamine was smallest after 24-hours and peaked at 12.5-hours.

The changes in abundance of the other amino acids between 3.5- to 12.5-hours in this study are also similar to the work previously published by this laboratory [39]. One notable exception is histidine which accumulates in cells in the earlier study [39] but decreases in this work (Fig. 5.3). One possible explanation is that cells were in a slightly different physiological states when sampling occurred as the 12.5-hour timepoint is as cells transition to diazotrophy. This is a dynamic process and while we have done our best to standardise this in our laboratory, as evidenced by the similarity in changes between other amino acids, it could be histidine was metabolised as part of the TCA cycle [186] or had undergone deamination in these samples and not in the early work [187,188].

Given the importance of amino acids as the building blocks for growth, repair and cell wall synthesis it is not surprising that such profound changes overtime were observed across the low nitrogen treatments. While mostly non-significant, the pathways that were most often found to be altered across time were those related to amino acid metabolism and biosynthesis. Additionally, purine metabolism was found to be non-significantly upregulated at 24-hours compared to 3.5-hours. This is in agreement with exploratory transcriptome analysis carried out by this laboratory which found purine and nitrate/nitrite metabolism was upregulated during diazotrophic growth [39]. The metabolism of these nitrogen contain metabolites could be indicative of nitrogen scavenging during starvation providing another possible nitrogen source alongside fixed atmospheric nitrogen.

Purine metabolism was also upregulated in the high nitrogen treatments when compared to their low nitrogen counterparts. Although the pathways for alanine, aspartate and glutamate metabolism and arginine biosynthesis were amongst the most significantly upregulated in cells on high nitrogen at 3.5-hours compared to those on low nitrogen at 3.5-hours. It is unsurprising that glutamate metabolism is upregulated in nitrogen replete conditions as ammonium is being assimilated via the low energy GDH pathway producing glutamate via amination of α-ketoglutarate [47]. Given the clear importance of amino acid metabolism in facilitating nitrogen fixation and understanding its differentiation from non-fixing physiology future biotechnological approaches could consider these pathways as an avenue for experimentation to explore potentials for engineering an improved diazotrophy.

5.4.3 Elevated levels of TCA cycle metabolites in nitrogen fixing *K. oxytoca*

The intracellular abundance of malate and acetyl-coenzyme A (Ac-CoA) followed an interesting pattern over time. The abundance of these metabolites is at the lowest at the 12.5-hour timepoint, decreasing between 3.5- to 12.5-hours before increasing in abundance again at 24-hours. Given their role in the TCA-cycle the decrease in intracellular abundance could indicate altered carbon metabolism because

of the transition to diazotrophy and formation of the nitrogenase enzyme. Indeed, carbon metabolism has been shown to impact nitrogen fixation, including the levels of nitrogenase in symbiotic nitrogen fixers [189] and in free living nitrogen fixers at the transcriptional level [190]. Lin *et al.,* also show that in *Pseudomonas stutzeri* the *Crc* gene, responsible for carbon catabolite repression, is in fact a global regulator of nitrogen and carbon metabolism, coupling them together. This means that Crc facilitates nitrogen fixation in the root associated *P. stutzeri* [190]. In this work, the increase in these metabolites at 24-hours could indicate some recovery of normal cell metabolism, which is also evidenced by the secondary growth phase.

Additionally, both malate and Ac-CoA are involved in metabolic pathways responsible for the production of butan-2-3-diol, the plant growth promoting metabolite that was found to accumulate in the medium in Chapter 4. Malate is involved via the butanoate pathway [163,191] and Ac-CoA through the production of acetate which is converted to acetoin the precursor to butan-2-3-diol [163,192,193]. The intracellular increase is more profound in Ac-CoA compared to malate and this could be due to the central role Ac-CoA plays in many different metabolic pathways.

5.4.4 Increased intracellular levels of known plant associative metabolite N acetylglutamate

The intracellular abundance of N-acetylglutamate was found to increase over time, a trend also seen in the nitrogen fixing bacteria *Zymomonas mobilis* [184]. This metabolite is not only the first metabolite in the biosynthesis of arginine but also has been shown to have plant growth promoting effects in root nodule symbiosis [177,194]. Having been shown to induce changes in the plant roots such as hair branching and tip swelling and prolonged elongation of root hairs in white clover but not in legumes [177]. The intracellular concentration of N-acetylglutamate continues to increase despite a decrease in arginine abundance between the 12.5- to 24-hour timepoints. This could suggest that this is an additional plant growth promoting metabolite that would benefit the plants were *K. oxytoca* present in the rhizosphere, along with providing fixed nitrogen in the form of ammonium.

5.4.5 Possible stress protective metabolite decreases as cells transition to diazotrophy before increase during fixation

5-aminolevulinate varied significantly across time in low nitrogen conditions. Following the same trend as malate and Ac-CoA, with the greatest abundance at the 24-hour timepoint. The intracellular abundance of this metabolite was also increased in the △*amtB* mutant at the 12.5-hour timepoint and at similar levels to the WT at 24-hours. Their abundances were also comparable to cells in a nitrogen replete condition, suggesting that this metabolite is produced as a result of nitrogen fixation.

This metabolite is produced from glutamate via glutamyl-tRNA and glutamate-1-semialdehyde using the HemA and HemL [195] and again has been shown to increase plant growth and productivity [196,197]. Application of 5-aminolevulinate was found to increase shoot and root weight as well as improve the accumulation of nitrogen in the leaves and roots [197]. Perhaps uptake of fixed nitrogen could be affected by soil concentration of 5-aminolevulinate. Additionally, 5-aminolevulinate has been shown to mediate chlorophyll production, therefore this is a metabolite that has multiple effects on the plant and is worthy of further consideration. [198].

5.4.6 Limitations and future work

While the work in this chapter benefits from covering a wider range of timepoints than other similar studies [39,184], it lacks the ability to track metabolite changes in in fine detail. Future work could now focus in on one section of diazotrophic growth for example the time from fixation onset to a stable fixing phenotype. This could allow for insights into the maintenance of nitrogen fixation across time for free living diazotrophic bacteria. Information on these later stages of nitrogen fixation could also provide insights into how cells balance the energetic cost of fixation with survival and growth and how this might be used in biotechnological approaches.

As already discussed in Chapter 4, the LC-MS method used in this chapter was also meant to be used to analyse the supernatants collected in the previous chapter. This should still be a priority as it is important to know whether the changes in intracellular concentrations of plant growth promoting metabolites, such as N-acetylglutamate and 5-aminolevulinate, results in increased abundance in the bacteria's environment. Additionally, the corresponding cells and supernatant should be analysed simultaneously to allow for the interaction between bacteria and environment to be investigated. That was the original intention here, but unfortunately, we were not able to do so.

Finally, it is difficult to know how transferable insights gained from batch culture are to 'real world scenarios', excepting that bacterial cells must undergo a range of transitions in nitrogen stress as N containing compounds ebb and flow in the biosphere and in communities of microbes. Moreover there is a growth regime that has been shown to directly improve plant growth, the nitrogen fixing biofilm [40,66,199]. This presents a novel mode of nitrogen fixation which is still understudied and could have a potentially dramatic impact on how we approach free-living bacteria and plant mutualism. Therefore, gaining an understanding of how the metabolism of this growth phenotype differs from batch culture, in addition to metabolic changes brought on by fixation, would be a rewarding avenue of research.

5.5 Conclusion

These analyses are among the first to show distinct metabolite compositions based on nitrogen status and fixation. These differences are due to changes in global metabolism resulting from both nitrogen status and the different stages of diazotrophic growth. Furthermore, cells that had been fixing for a prolonged period appear to recover some normal cell function which could facilitate growth. Although the cells remain in a state of nitrogen stress despite fixing atmospheric nitrogen. Perhaps the most promising finding is that metabolites which have been shown to promote plant growth were found to be more abundant in fixing cells. These could be examples of metabolites, other than ammonium that may be taken up by the plant roots contributing to mutualism between the bacteria and plant.

Chapter 6 A multi-omic investigation of nitrogen fixing biofilms

6.1 Introduction

Biofilms are typically characterised by a community of cells, either a single species or consortia, that exist in a self-produced matrix of extracellular polymeric substances (EPS) and adhere to a surface or sometimes each other [200,201]. Notably, bacteria exhibiting this growth phenotype are reported to be in a completely different physiological state compared to free-living bacteria cultured in liquid medium [201]. This is thought to be due to the intracellular interactions that occur due to cell surface-surface contact as well as interactions with the encapsulating matrix [202,203]. Often these cell-cell interactions are coordinated allowing for cells to differentiate between 'roles' and occupy a range of ecological niches within the biofilm, often linked to the spatial organisation of the biofilm. Therefore, while some similarities do carry over, biofilms tend to possess properties distinct from free-living bacterial cultures with assumptions and predictive that need testing.

Biofilms have been shown to be extremely wide spread, having been found colonising medical equipment, waste water and most importantly for this thesis soil [67,204,205].Notably, biofilms have been shown to be promising candidates for biotechnological application due to their enhanced ability to deal with environmental stress, differentiated genetic regulation and metabolic heterogeneity [204,206]. For example, biofilms have been shown to act as water purifiers, degrade waste and produce chemicals including biofuels [206]. Their resilience and adaptability, as well as their applicability in cutting edge synthetic biological techniques, makes them prime candidates to help tackle some of the world's most pressing scientific issues.

Indeed, since their recent discovery, nitrogen fixing biofilms have received much research interest [29,52,66,199,207–209]. Importantly, nitrogen fixing biofilms are able to fix nitrogen in air something of note given the extreme oxygen sensitivity of the nitrogenase enzyme as well as the need for a low oxygen tension for nif gene expression. As such biofilms have the potential to increase crop growth and durability through metabolic exchange in a wide range of environmental conditions [40,52,209]. These benefits can be inferred to the plant due to the biofilms ability to facilitate root colonisation, in part through protection from plant defences via EPS production [209–211]. Termed EPS mediated epiphytic colonisation, this growth has been shown to facilitate root colonisation by symbiotic and free living species alike and could represent a viable method of plant root-free living bacteria association and mutualism [211,212]. A recent 2017 review listed 10 bacterial species capable of forming nitrogen fixing biofilms, *K. oxytoca* was not on this list and there does not appear to be any published papers on

nitrogen fixing *K. oxytoca* biofilms. Therefore, the work carried out in this chapter represents a first in the known reporting of *K. oxytoca* biofilms being able to fix nitrogen in air. While, not all the requisite tests have not been carried out to confirm these are indeed biofilms, research is being carried out by others in this laboratory to do so. Although in its early stages, microscopy images appear to show layer differentiation in *K. oxytoca* biofilms (H. Wang, Unpublished). Furthermore, the fact that closely related species *Klebsiella pneumoniae* and *E. coli* are known to form biofilms [213,214], gives confidence to the label. Thus, these colonies will be referred to as biofilms throughout this chapter and in subsequent references.

Metabolomics has proven to be a useful tool when studying biofilms, their metabolism and interaction with the environments. As with free-living bacteria, some metabolites will play a key role in gene expression and enzymatic activity having a key role in global metabolism regulation. For example, metabolomics has been used to investigate how plants genetically modified to increase production of biofilm promoting compounds impacted root colonisation and biofilm formation [52]. As with batch culture experiments, there are few, if any, metabolomic studies on free-living nitrogen fixing biofilms. Given the reported metabolic differences between free-living and biofilm growth phenotypes it is important we study biofilms independently of batch culture. However, the batch culture results from Chapter 5, will serve as useful basis to see what metabolic changes could be ubiquitous to nitrogen fixation in *K. oxytoca* and what might be specific to one or other growth phenotype.

Another tool that has proven valuable in understanding nitrogen fixing biofilms, is analysis of the transcriptome [66,199]. This has allowed for the regulatory networks that underpin biofilm nitrogen fixation to begin to be understood. In *Pseudomonas stutzeri* it is proposed that RpoN, RpoS, Gac and Rsm all play a role in biofilm EPS production and RpoN alone initiating NifA expression which goes on to trigger the remaining Nif genes forming the nitrogenase complex. A study that combines these tools in some way could better our understanding of nitrogen fixing *K. oxytoca* biofilms.

6.1.1 Aims and objectives:

The aim of this chapter is to carry out exploratory multi-omics analysis on *K. oxytoca* biofilms under fixing and non-fixing conditions. These data will be used to explore how the metabolic profiles differ between treatments, using targeted LCMS and transcriptome analysis.

Objectives:

- To investigate the metabolic profiles of *K. oxytoca* cells across a 6-day period, under fixing and non-fixing conditions using targeted LCMS analysis developed in Chapter 3.
- To investigate spatiotemporal differences in metabolite composition within nitrogen fixing biofilms, using TOF-SIMs, to elucidate fixation location.
- Carry out exploratory transcriptome analysis via RNA-sequencing on fixing and non-fixing biofilms, to understand broad changes in gene expression.
- To analyse the integrated data of the cells using multivariate and univariate statistical methods.

6.2 Methods

6.2.1 Bacteria strain, mutants, and culturing

The parent strain and mutants of this strain used in the experiments carried out in this chapter can be seen in the table below.

Table 6.1 List of bacterial strain and mutants used in the experiments.

Strain name	Gene knockout	Summary
Wild type (WT) M5a1	None	Has a normal growth phenotype and is able to fix nitrogen
△*nifLA*	△*nifLA*	Unable to fix nitrogen, no diazotrophic growth
△*nifH*	△*nifH*	Unable to fix nitrogen, no diazotrophic growth
△*amtB*	△*amtB*	Able to fix nitrogen and exhibits diazotrophic growth; AmtB is necessary to regulate ammonium levels in the cell through action of the AmtB ammonia transporter.

Biofilms were grown on solid NFDM medium solidified with 1.5 % agarose and supplemented with either 1 mM (low nitrogen) or 20 mM (high nitrogen) NH_4Cl.

Culturing for acetylene reduction assay

For the ARA assay biofilms were grown as outlined in Chapter 2.4.2. In brief, 10 uL of the washed *K. oxytoca* overnight culture diluted to 0.1 OD_{600} was spotted onto sterilised polycarbonate disks placed onto the media. The disks have been shown to not impede growth or nutrient uptake (Buck laboratory, unpublished) and aids in biofilm transference. This growth condition was chosen as it resulted in a sufficient biomass to produce a reliable signal which when tested using the ARA assay. Biofilms were then grown in sealable jars, incubated at 25°C for the duration of the experiment, with 3 biofilms grown per jar.

Culturing for metabolomic and RNA-Seq analysis

Biofilms were cultured as outlined in Chapter 2.4.2. In brief, 10 uL of the washed *K. oxytoca* overnight culture diluted to 0.1 OD_{600} was spotted onto sterilised polycarbonate disks placed onto the medium in square petri dishes (120 x 120 x 17 mm, Greiner Bio-One). Again, biofilms were incubated at 25°C for the duration of the experiment. This growth condition was chosen as it was consistent with the growth condition for the ARA and produced the highest biomass and given the sensitivity of RNA sequencing and the metabolomic assays used in this chapter, biofilm biomass needed to be maximised.

Culturing for TOF-SIMS analysis

Larger biofilms cultures were unsuitable for analysis via TOF-SIMS and therefore single colony biofilms were grown as outlined in Chapter 2. In brief, overnight cultures supplemented with 20 mM heavy labelled ^{15}N NH_4Cl (CAS number: 39466-62-1 Sigma Aldrich, Gillingham UK) were further diluted from 0.1 OD_{600} to approximately 20 cells per mL. For the 50 mL square plates, 50 uL of this culture was spread using a sterile L-shaped spreader. Plates were incubated at 25°C for up to 7 days. Note, Biofilms were grown on NFDM agarose medium supplemented with heavy labelled ^{15}N NH_4Cl at 1 mM, this allowed differentiation between fixed atmospheric ^{14}N nitrogen and ^{15}N nitrogen from the medium being incorporated into the cells metabolites. Initial growth was supported by ^{15}N NH_4Cl, and subsequent growth by atmospheric ^{14}N nitrogen.

6.2.2 Acetylene Reduction Assay

A preliminary test was carried out to see if biofilms were able to fix nitrogen after 7 days of incubation. ARA was carried out as outlined in Chapter 2, to summarise 1 mL of pure acetylene was injected into the headspace of jars and left to incubate for 6 hours. Post incubation, the headspace was sampled and analysed via GC and the acetylene and ethylene peaks were detected, integrated and analysed. Post sampling, the lids were released, and normal airflow resumed removing any acetylene and ethylene from the headspace. The results of this preliminary study indicated that WT biofilms grown on the low 1mM NH_4Cl medium began fixing nitrogen after 2 days and were still fixing nitrogen after 7 days. Additionally, WT biofilms grown on high 10mM NH_4Cl nitrogen medium were not fixing atmospheric dinitrogen gas after 7 days.

The WT strain and two mutants were chosen for study via time course experiment; WT, △*amtB* (positive control), △*nifLA* (negative control). The WT and mutants derived from this parent strain were cultured for six days on low nitrogen medium and sampled every day from day two to six. The results from this experiment would allow for timepoint selection in future experiments. Three biofilms were grown per jar and 3 jars were used per treatment therefore n = 3. Due to following the same biofilms

through the course of this experiment, biomass or CFU was not measured. As such, results are presented in percentage of fixed ethylene. This was calculated by integrating both the acetylene and ethylene peaks, summing these, and calculating the percentage contribution of the ethylene peak.

6.2.3 Metabolomic analysis of biofilms using targeted LCMS

Results from the ARA time course experiment indicated that WT *K. oxytoca* was still fixing nitrogen after 6 days, and that the fixation capacity could be increasing across this time. This information coupled with the preliminary data meant both the high and low nitrogen treatments could both be sampled across a six-day period, allowing for a direct comparison between high and low nitrogen metabolite profiles at all timepoints.

It was decided that WT biofilms grown on both high and low nitrogen medium would be sampled on days two, four and six owing to the trade-off between the number of samples per treatment and number of treatments. As with the batch culture experiments in Chapter 5, a $\triangle$*amtB* positive control grown on low nitrogen was included at the four-day timepoint only. This meant a total of 7 treatments and 7 replicates were sampled per treatment.

Sampling for metabolomic analysis

As with batch culture, sacrificial sampling was implemented in this experiment. Given the structure of a biofilm with the EPS and the nitrogenase's sensitivity to oxygen it was not thought possible to sample 'part' of a biofilm. The sampling method used was the same outlined Chapter 3 and used successfully in Chapter 4 with one small modification. Biofilms were added directly into methanol, chloroform water solution using the polycarbonate discs, allowing for instantaneous quenching/extraction.

LCMS analysis

The LC-MS carried out in this chapter implements the AccQ-Tag RPLC and ion pairing chromatography (IPC) methods outlined in Chapter 3. These methods cover a range of metabolites and were selected as the optimum pairing for a targeted assessment of the metabolome [129]. In total 108 metabolites were detected with 102 of these deemed to be of sufficient quality for analysis.

LCMS data analysis

Peaks were integrated in Skyline (MaCoss Lab software, US) and normalised to biomass using OD_{600}. Data were analysed using in house Python and RStudio scripts. Where applicable, univariate statistics, e.g. T-Test/Mann-Whitney U test were corrected for FDR using the Benjamini-Hochberg method. Pathway analysis was carried out on metabolites with significantly different fold-changes between

treatments using MetaboAnalystR Package [174,175]. Data was normalised using the probabilistic quotient method [113], due to difficulties assessing both biomass and sampling biofilms efficiently.

6.2.4 Spatial-temporal analysis of biofilms using TOF-SIMS

To investigate possible spatial metabolic changes across layers in the biofilm TOF-SIMS analysis was carried out. The same two-, four- and six-day timepoints were selected for WT grown on low nitrogen media, as in the metabolomic profiling experiment. A △*nifLA* mutant was used as a non-fixing control grown on low nitrogen media, while minimising possible differences due to high and low nitrogen. Two replicates were used for each treatment. Biofilms grown under these conditions were approximately 1800 µm in diameter, 25 µm in height and spherical in shape.

TOF-SIMS analysis was carried out by Sarah Fearn from Imperial College London. In short, samples were analysed with a TOF-SIMS 5 secondary-ion mass spectrometer (IONTOF GmbH, Münster, Germany). Biofilms were lifted off the medium using the polycarbonate disc and placed onto an aluminium cold block (at -80°C) to stop metabolism, freeze dried and lyophilized before being analysed. Analysis was carried out with a 10-keV Ar1249 GCIB with a current of approximately 4 nA for controlled sputtering of the sample over a 300- by 300- µm area. Data were collected over a 150- by 150-µm area in HCBM (high-current bunched mode) with a 25-keV Arg^+ analysis beam over an m/z range of 0 to 800. Semi-targeted data (i.e., known ions) was collected in positive mode and the following calibration peaks were used; CH_3^+, $C_3H_7^+$, $C_4H_8^+$, $C_9H_9^+$, $C_4H_7O^+$, and $C_9H_{17}^+$, N_2^+. Spectra were normalised to total ion intensity for a given spectra to allow for comparisons over time and peaks were integrated using a binning approach on the Ion-TOF software.

6.2.5 Transcriptome analysis using RNA-seq

Biofilms were grown as outlined above and based on the results of the acetylene reduction assay sampled at five days. This sampling regime was chosen to maximise WT biomass under nitrogen limited conditions while not allowing for fixation capacity to have plateaued. This research was exploratory in nature and the number of samples that could be analysed was limited. As such, the treatments chose were WT biofilms grown on high and low nitrogen medium and a negative non-fixing control of △*nifH* also grown on low nitrogen media. Two reps of each treatment were collected. Biofilms grown under these conditions were approximately 10-15 mm in diameter, 5 mm in height and spherical in shape.

RNA-Seq sampling

Biofilms were sampled and analysed in the same manner as described by Waite *et al.,* [39]. Samples were fixed with a 1/10 volume of 5% phenol/95% ethanol (v/v), harvested by centrifugation and stored on dry ice. Samples were sent to Vertis Biotechnologie AG (Freising, Germany) where the following was performed: RNA extraction, library preparation, quality control and next-generation sequencing. To extract total RNA from the biofilm samples, they were treated with lysozyme for 15 minutes. The total RNA was then isolated using the mirVana microRNA isolation kit (Ambion), after which contaminant DNA was removed using DNase and ribosomal RNA was depleted using the Ribo-Zero rRNA Kit for bacteria (Illumina). Fragmentation of the primary RNA transcripts was carried out using ultrasound (four pulses of 30 seconds each at 4^{o}C) and ligation of the 3' oligonucleotide adapters was carried out.

cDNA synthesis of the first-strand was performed using the Moloney Murine leukaemia virus reverse transcriptase and was primed using the 3' adapter. Post purification of the first-strand cDNA, a 5' Illumina Truseq sequencing adapter (Illumina) was ligated to the 3' end of the antisense (or second-strand) cDNA, this adapter was inclusive of a 8 nt barcode sequence. PCR amplification of the of the double stranded cDNA was employed to approximately 10-20 ng/µL (12-13 thermocycles) using a high-fidelity polymerase. The amplified cDNA was purified and pooled equimolarly before being size-fractionated using the Agencourt AMPure XP kit (Beckman Coulter Genomics) to 200-500 bp. The pooled sample was then analysed using capillary electrophoresis to ensure library preparation and fragment size selection had been successful and was of sufficient quality for sequencing.

The pooled cDNA was sequenced on a NextSeq 500 system (Illumina) using a read length of 75 nt and a depth of 100 million reads. Reads were trimmed of adapter sequences and duplicates were removed prior to alignment to the *K. oxytoca* M5a1 genome (GenBank WGS accession JAFHKG010000000; BioSample SAMN17288411). Data was provided by Vertis Biotechnologie in the form of raw reads (fastq), binary alignment map (bam) files, and normalised gene counts (reads per kilobase per million, RPKM).

RNA-Seq data analysis

Data was analysed using the R/Bioconductor statistical package DESeq2, which allowed for analysis of differential gene expression using Wald hypothesis testing, and p-value adjustment using a False Discovery Rate threshold of 0.05 (Benjamini and Hochberg). This analysis was used to investigate Log_2

fold changes for differentially expressed genes between treatments using an adjusted p-value of <0.05. Furthermore, normalised gene counts were used in further univariate and multivariate analysis.

6.3 Results

6.3.1 Percentage ethylene fixed changes over time for both WT and △*amtB*

A trend of increasing ethylene fixation was observed across time for WT biofilms grown on low nitrogen media, suggesting a trend of increased nitrogen fixation capacity (Fig. 6.1). Additionally, biomass was observed to be increasing across the same period, though this was not measured. On day two an average of 0.059% of ethylene was fixed. This is significantly lower compared to 16.01% ethylene fixed on day six ($p = 0.029$, Kruskal-Wallace Test). The △*amtB* mutant was found to have fixed more ethylene at day two compared to WT with an average of 1.31% percentage ethylene in the headspace. However, while the △*amtB* mutant also fixed more acetylene across time, the rate of increase slows after day three, with fixation appearing to plateau between days five and six. At day six an average of 13.90% ethylene was fixed by △*amtB* which is lower than the WT, indicating the WT is fixing more nitrogen at this timepoint. In contrast no ethylene was detected in the headspace for the △*nifLA* mutant at any timepoint meaning that nitrogen fixation was not occurring.

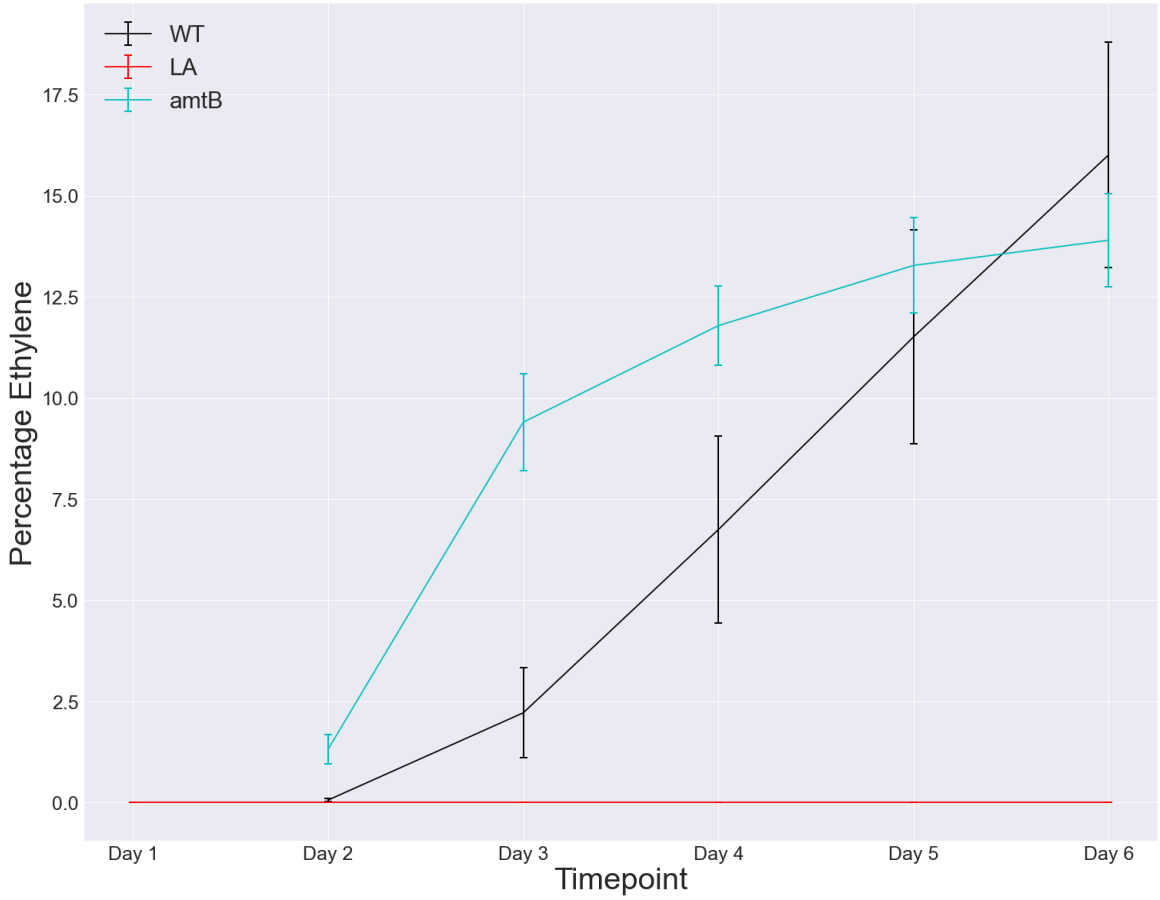


Figure 6.1 Comparison of percentage ethylene fixed by biofilms grown on low nitrogen media, where the WT strain is represented in black, the △nifLA mutant, which is unable to fix nitrogen, is represented in red and the △amtB mutant, which is able to fix nitrogen is represented in light blue. Lines represent the group mean and error bars are the SEM, n= 3.

6.3.2 Multivariate analysis of metabolomic data showed clear differences in the metabolite composition between all treatment groups – timepoint, nitrogen status and strain

The PCA scores plot in Figure 6.2 shows that the metabolite profiles between the high nitrogen, non-fixing, treatments are distinct from both the low nitrogen, fixing, treatments and each other. With the grouping between the high nitrogen day two and day four treatments being closer compared to the day 6 high nitrogen treatment. This shows that the biofilms grown on high nitrogen are still undergoing metabolic changes across time possibly indicating continued growth across this period. The two-day WT biofilm on low nitrogen appears to be metabolically distinct, possibly down to the limited nitrogen fixation occurring at this timepoint.

The three remaining low nitrogen treatments cluster together but the samples still display distinct treatment grouping. This means that while the metabolite profiles of these nitrogen fixing biofilms share similarities, due to fixation, they remain distinct from each other across time and between the WT and △*amtB* mutant . The grouping of the WT and △*amtB* biofilms grown on low nitrogen for four days is the closest, meaning that the △*amtB* mutant is metabolically similar to the WT at its equivalent timepoint. This differs from results in batch culture which found the metabolite profiles of the △*amtB* mutant more comparable to later timepoints in the WT. PC 1 appears to separate by high and low nitrogen treatment groups, whereas PC2 looks to separate by fixing and non-fixing metabolic phenotypes. Combined these results provide evidence of clear metabolic changes due to nitrogen fixation and age of the *K. oytoca* biofilms.

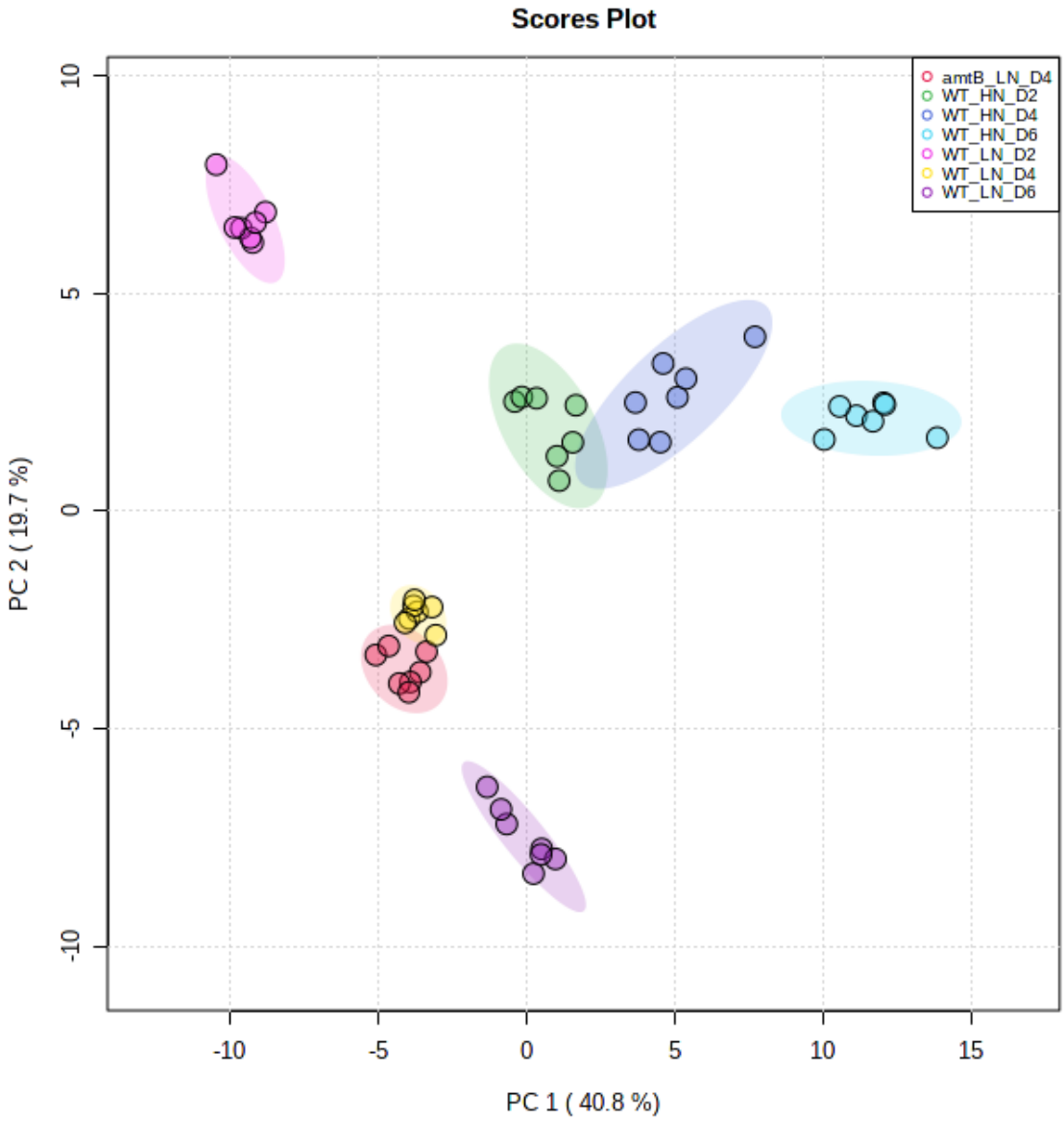


Figure 6.2 PCA scores plot comparing the metabolic profiles of WT and ⊿amtB mutant biofilms grown on either low or high nitrogen medium for 2, 4 and 6 days. Each dot represents a sample and the shaded rings represent 95% confidence intervals, n = 7.

6.3.3 Key nitrogen signalling metabolites show significant differences due to time, nitrogen status and strain

90 of the 103 metabolites analysed were shown to be significantly different across time for the WT low nitrogen treatments using Kruskal-Wallis test. Glutamate and glutamine, the key nitrogen status signalling metabolites were including in this number (Glutamate adjusted p = 0.000107, glutamine adjusted p = 0.0178). The abundances of these metabolites appear to behave independently of one and other, with glutamate decreasing over time and glutamine peaking at day 4 (Fig. 6.3 A & B). Despite the decreasing abundance across time, glutamate was more abundant than glutamine at all timepoints.

As one might expect, significant differences in abundance were also seen between the high and low treatments between comparable timepoints (Fig. 6.3 C & D). Glutamate was found to be significantly higher in the low nitrogen treatment compared to the high nitrogen at the day 2 (adjusted p = 0.00132 Mann-Whitney U test with FDR correction for all comparisons), day 4 (adjusted p = 4.98E-08) and day 6 (adjusted p = 0.00019027) timepoints. In contrast glutamine was significantly lower in the low nitrogen treatment at the day 2 timepoint (adjusted p = 0.00268), not significantly different at the day 4 timepoint (adjusted p = 0.32093248) and significantly more abundant in the low nitrogen treatment at the day 6 timepoint (adjusted p = 0.00044). The biggest difference in abundance is seen in glutamate abundance at the day 6 timepoint, where the high nitrogen condition has approximately half of the abundance of the low nitrogen treatment (Fig. 6.3 C).

A comparison of the WT and △*amtB* biofilms grown on low nitrogen medium after four days shows that glutamate and glutamine behave differently. Glutamate was found to be more abundance in the WT condition compare to △*amtB* (adjusted p = 0.0029). Whereas, glutamine abundance did not differ significantly between WT and △*amtB* mutant (adjusted p = 0.373, Fig. 6.3 C & D). Further consideration of changes in Glutamate levels is given below.

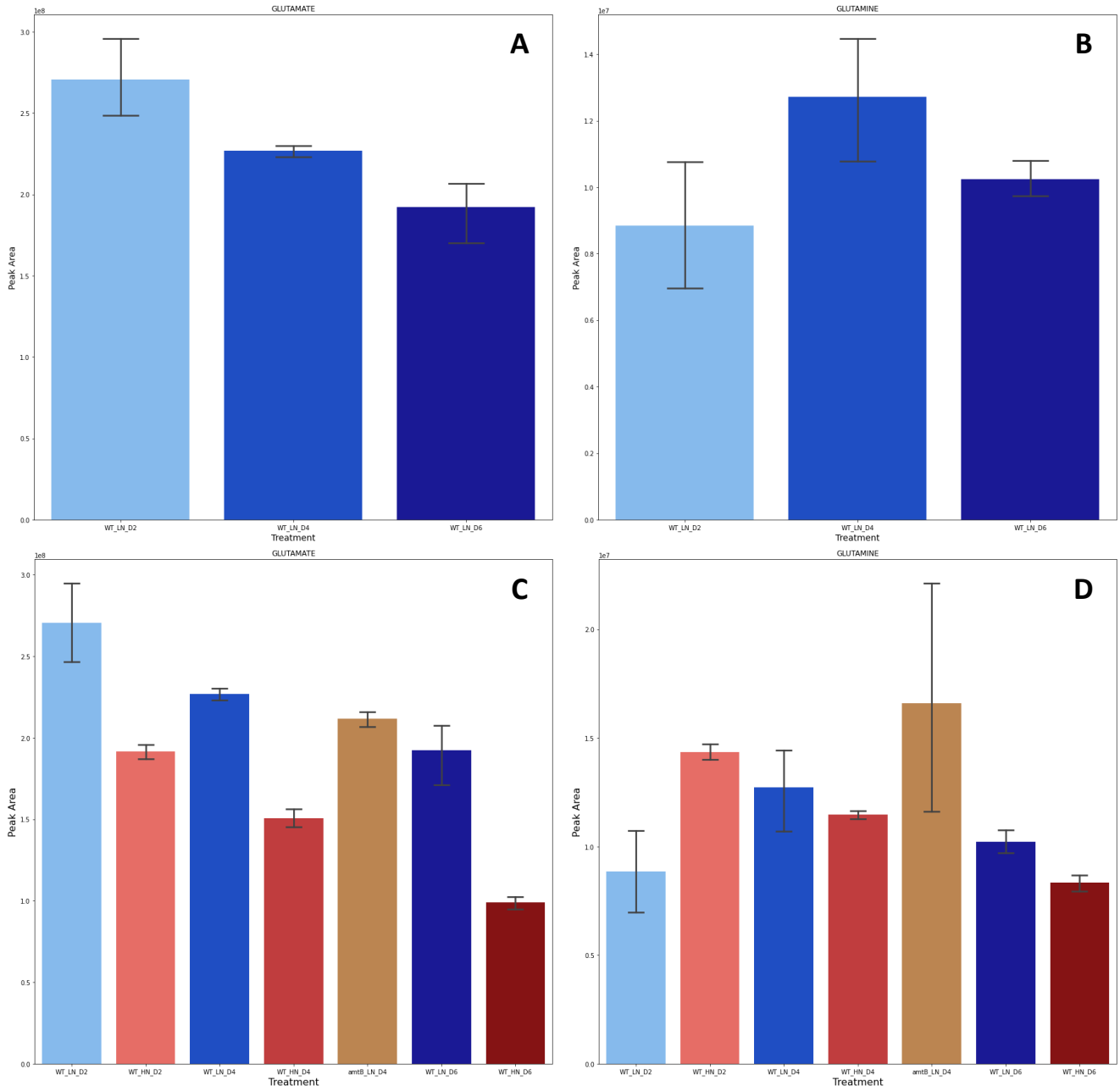


Figure 6.3 Comparison on metabolite abundance; A-B for all timepoints of the low nitrogen condition for (A) glutamate and (B) glutamine. C-D for all treatments and timepoints for (C) glutamate and (D) glutamine, n = 7.

6.3.4 Biofilm amino acid abundance varies across time when grown under fixing conditions

All bar one of the of the remaining amino acids were found to vary significantly overtime via Kruskal-Wallis test (FDR corrected p = <0.05). The exception is leucine (adjusted p = 0.133), which remains relatively unchanged across time. As the batch cultures from Chapter 5, showed changes in amino acid abundance can be grouped. Of the seven most abundant amino acids, three were found to significantly decrease over time glycine, glutamate and phenylalanine and four were found to increase alanine, valine, arginine and tryptophan (Fig. 6.4 A). Given that glutamate is involved in nitrogen assimilation and can be metabolised into many different amino acids, for example alanine and asparagine these dynamics are not unexpected. Changes in the less abundance amino acids can be seen in Figure 6.4 B. In some instances, such as tryptophan, an inverse trend in abundance change compared to batch culture was observed. Abundance of tryptophan was found to increase in biofilms over time but decreases in batch culture. Overall, that the amino acid abundances change overtime, is strong evidence of metabolic changes occurring in nitrogen fixing biofilms.

Although not an amino acid ammonium is also considered here. As discussed earlier in Chapter 1, ammonium assimilation, will strongly impact the abundance of glutamine and glutamate, as well as the other amino acids. Under nitrogen deplete conditions, cells will swap from ammonium assimilation via the GHD pathway to assimilation via the GS-GOGAT pathway, for which the final products of this pathway are 2 molecules of glutamate [26]. These biofilms were under nitrogen stress from early in their growth, meaning ammonium assimilation via the GS-GOGAT pathway. As such, it makes biological sense that the abundance of ammonium was found to significantly decrease overtime, and that this aligns with changes in glutamate abundance in biofilms.

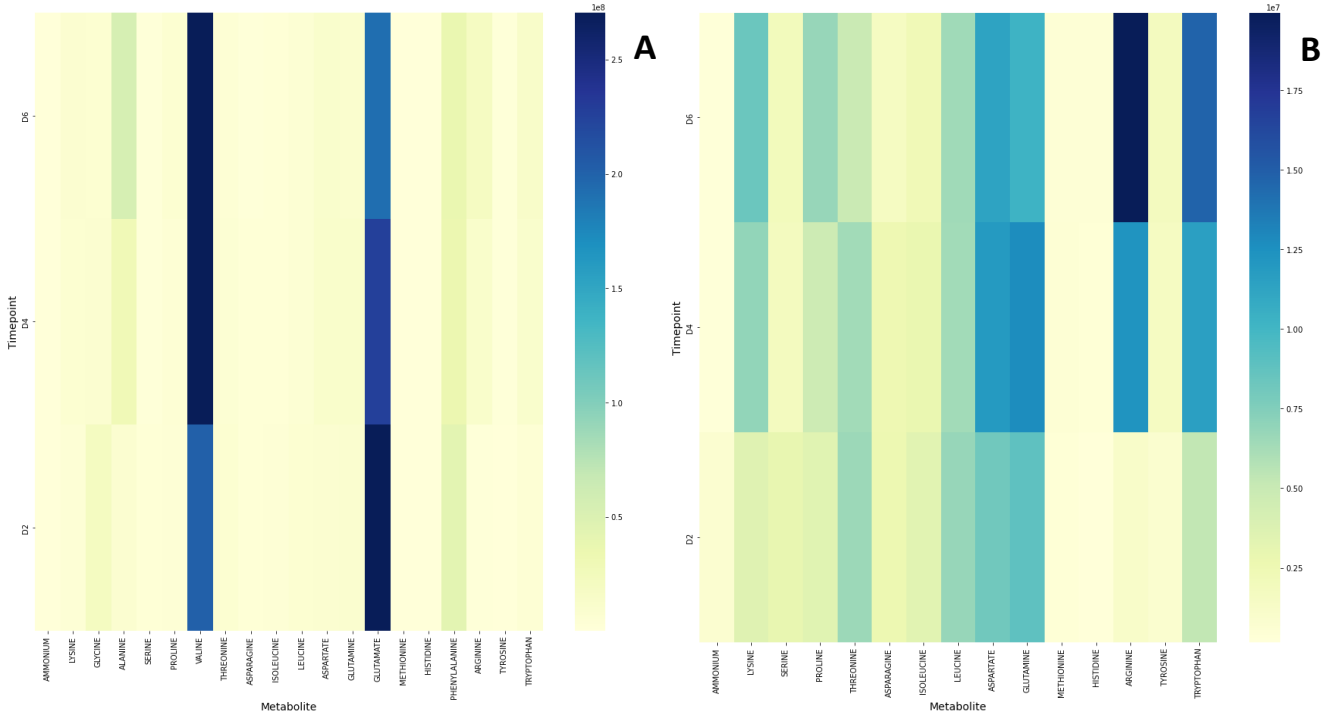


Figure 6.4 Comparison of biofilm amino acid abundance across time of wild-type K. oxytoca grown on low nitrogen media, for (A) all detected amino acids and (B) with the 5 most abundant amino acids removed. Heat map density representative of average abundance at a given timepoint. All amino acids, except leucine were found to undergo significant changes in abundance across time (Kruskal-Wallis test, n = 7, p < 0.05).

6.3.5 Univariate analysis reveals differences in metabolite composition between treatments

As with the cell analysis of batch cultures in Chapter 5, there are clear differences in metabolite profiles between treatments, for both time, nitrogen status and between the WT and △*amtB* mutant. The differences between specific treatments will be discussed along with any resulting metabolite pathways that are increased or decreased as a result of these differences. As in chapter 5, the relatively small sample size per treatment means that the pathway analysis should be considered descriptive. Differences have been calculated using the Mann-Whitney U test and corrected for FDR using the Benjamini-Hochberg method.

Comparison of WT low nitrogen 2-day to 4-day biofilms

When comparing these treatments an even spread of metabolites is observed between those that are increased or decreased in the day 4 samples compared to the day 2 samples (Fig. 6.5). With 79 out of 102 metabolites being significantly different between these treatments. Many of the most significantly different metabolites are more abundant in the day 4 samples. In fact, the four most significantly different metabolites 5-aminopentanoate (adjusted p = 1.57E-09), GABA (adjusted p = 3.37E-09), CDP (adjusted p = 1.50E-08) and homoserine (adjusted p = 1.50E-08) all increase in abundance between days 2 to day 4. Other metabolites of interest also significantly increased in day 4 samples were the Ac-CoA and several amino acids.

No pathways were found to be significantly decreased in day 4 samples relative to day 2 samples. However, purine metabolism was significantly increased in 4-day old biofilms compared to 2-day biofilms (8/73 metabolites, adjusted p = 0.0362, impact = 0.1264). The metabolites contributing to this pathway include GDP, ATP, ADP, GMP and xanthine.

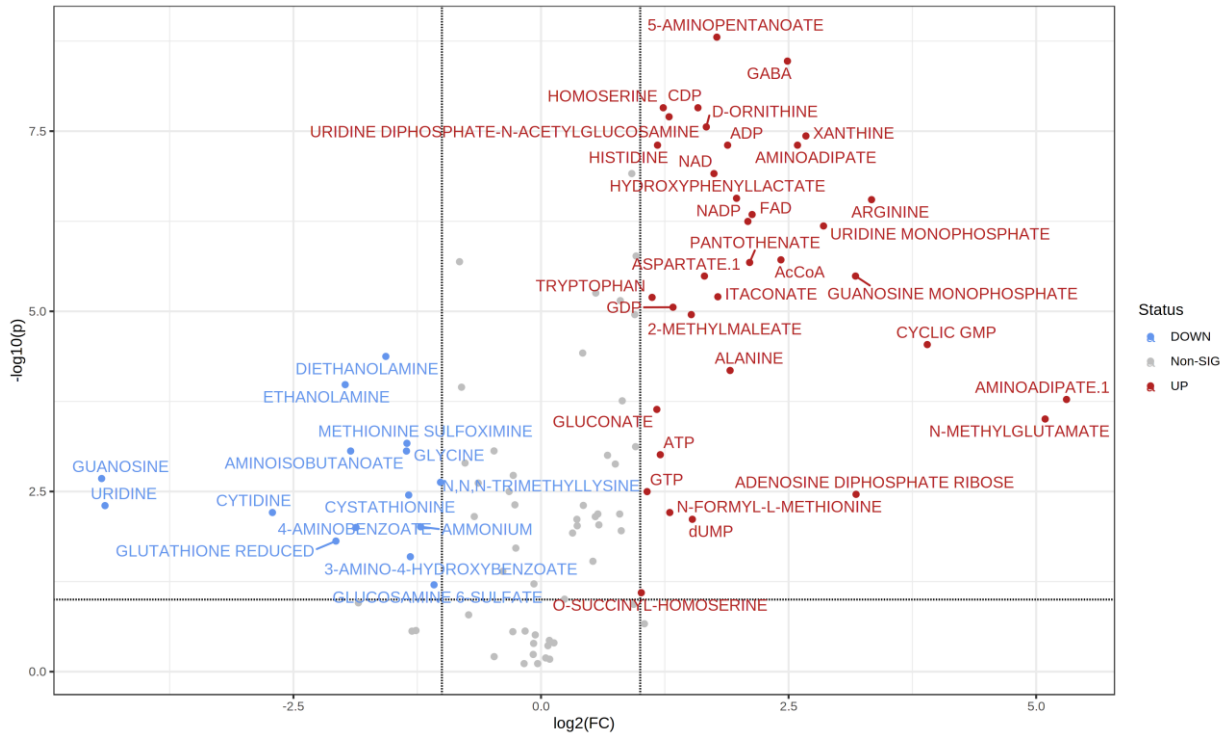


Figure 6.5 Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 4-day biofilm samples compared to WTL 2-day biofilms. Metabolites significantly less abundant at the 4-day timepoint in blue and those significantly more abundant at the 4-day timepoint compared to 2-day timepoint in red. Metabolites that were not different between treatments in grey.

Comparison of WT low nitrogen 2-day to 6-day biofilms

In comparing these treatments, most metabolites are found in lower abundance in the day 2 biofilms compared the day 6 biofilms (Fig. 6.6). Indeed, all the most significantly different metabolites are at lower abundance in the day 2 samples, indicating an increase of these metabolite abundance across time. Overall, 82 metabolites were significantly different and the four most significantly different were 5-aminopentanoate (adjusted p = 2.65E-08), D-ornithine (adjusted p = 2.67E-08), ophthalmate (adjusted p = 2.65E-08) and lysine (adjusted p = 3.66E-08). As seen in Figure 6.4, lysine is one of a number of amino acids that significantly increase over time between the day 2 and day 6 timepoints. Unfortunately, lysine did not pass QC for the analysis in Chapter 5, so no comparisons can be made.

No pathways were significantly decreased in the day 2 biofilms compared to the day 6 biofilms post FDR correction. However, the lysine degradation and biosynthesis pathways were significantly decreased pre-FDR correction. This could suggest that lysine has some role in biofilm nitrogen fixation and maintenance during maturation. No pathways were significantly increased in the day 2 biofilms compared to the day 6 biofilms.

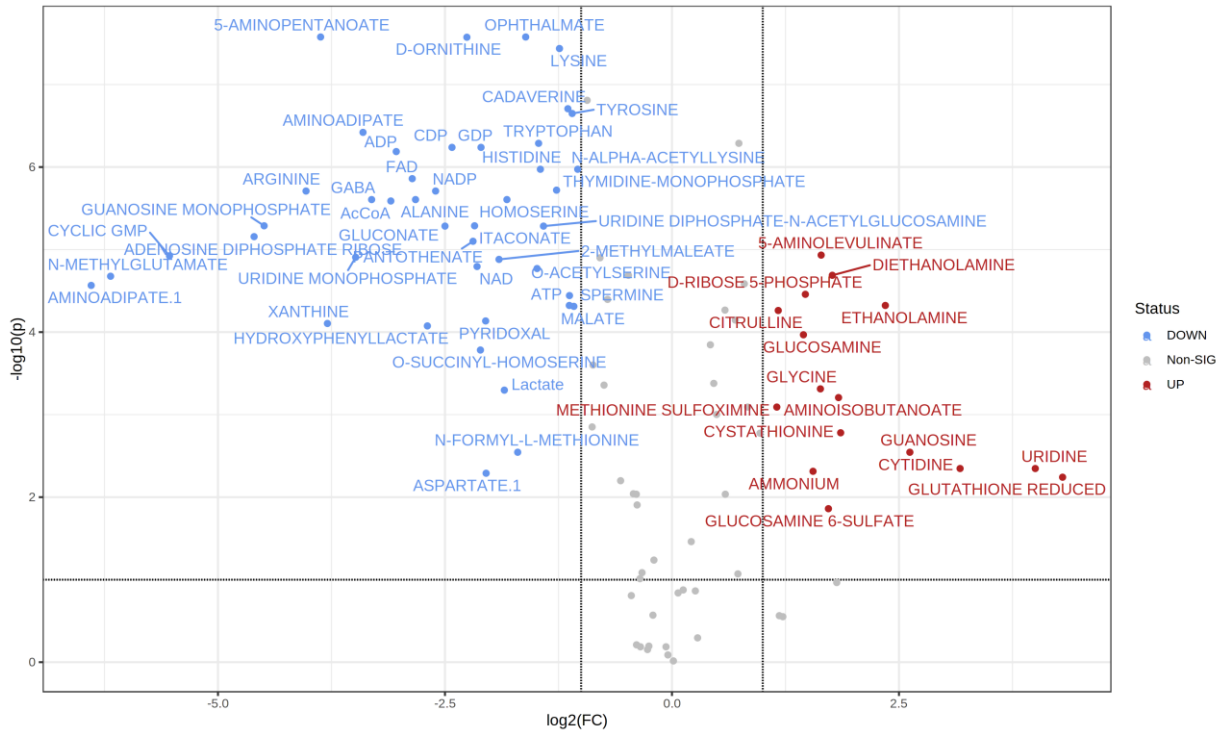


Figure 6.6 Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 2-day biofilm samples compared to WTL 6-day biofilms. Metabolites significantly less abundant at the 2-day timepoint in blue and those significantly more abundant at the 2-day timepoint compared to 6-day timepoint in red. Metabolites that were not different between treatments in grey.

Comparison of WT low nitrogen 4-day to 6-day biofilms

There were fewer differences seen between the day 4 and day 6 biofilms grown on low nitrogen media, compared to those between day 2 and 4, and 2 and 6 biofilms (only 62 differences seen overall). Only three metabolites were found to be significantly more abundant at the day 4 timepoint compared to the day 6 timepoint, meaning a decrease across time, one is noteworthy: 5-aminolevulinate (Fig. 6.7). This metabolite was found to possibly be linked to nitrogen fixation in batch culture. With that said, the 4 most significantly different metabolites are all lower in abundance at day 4 compared to day 6. These are 5-aminopentanoate (adjusted p = 3.86E-08), cadaverine (adjusted p = 1.90E-06), ADP (adjusted p = 3.81E-06) and thymidine monophosphate (adjusted p = 4.49E-06).

The results of the pathway analysis showed that only one pathway was significantly increased or decreased. Purine metabolism was found to be significantly decreased in day 4 biofilms compared to day 6 biofilms (p = 0.0135, 6/73 metabolites, impact 0.101). This could be in response to a change in nitrogen stress between days 4 to 6.

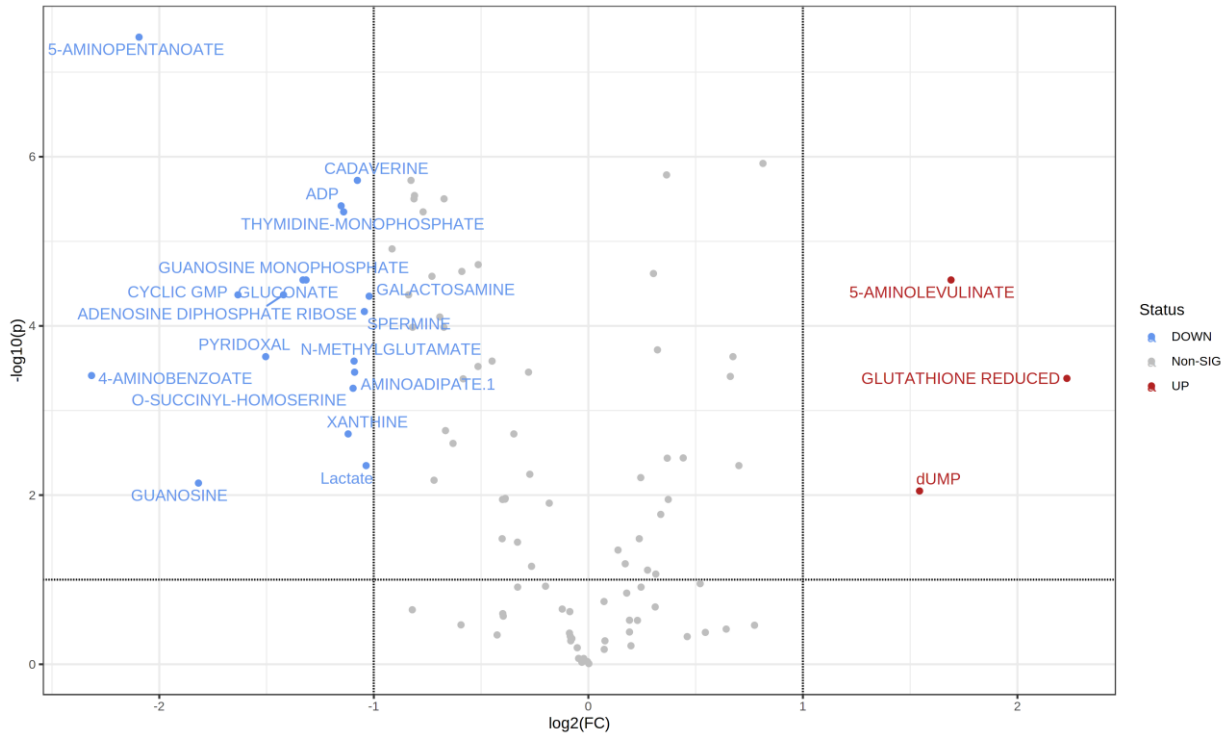


Figure 6.7 Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 4-day biofilm samples compared to WTL 6-day biofilms. Metabolites significantly less abundant at the 4-day timepoint in blue and those significantly more abundant at the 4-day timepoint compared to 6-day timepoint in red. Metabolites that were not different between treatments in grey.

Comparison of WT low nitrogen 2-day biofilms to WT high nitrogen 2-day biofilms

Many metabolites were found to be significantly differentially abundant between the day 2 high and low nitrogen biofilms. In total there were 80 metabolites that were significantly different between these treatments, not accounting for differences in fold change. The fold change analysis shows most metabolites were in lower abundance in the low nitrogen treatment at this timepoint (Fig. 6.8). It is notable that the amino acid glycine is significantly more abundant in the low nitrogen treatment compared to the high nitrogen treatment. However, the four most significantly different metabolites are in lower abundance in the low nitrogen treatment; arginine (adjusted p = 1.06E-07), tyrosine (adjusted p = 1.73E-07), ophthalmate (adjusted p = 3.73E-07) and GABA (adjusted p = 1.63E-06).

Purine metabolism was significantly reduced in the low nitrogen 2-day biofilms compared to 4-day biofilms (p = 0.0347, 9/73 metabolites, impact 0.232). Pre-FDR correction, pyrimidine metabolism was also reduced in low nitrogen biofilms compared to high nitrogen biofilms at the day 2 timepoint. No pathways were found to be increased in the low nitrogen treatment compared to the high nitrogen at the day 2 timepoint.

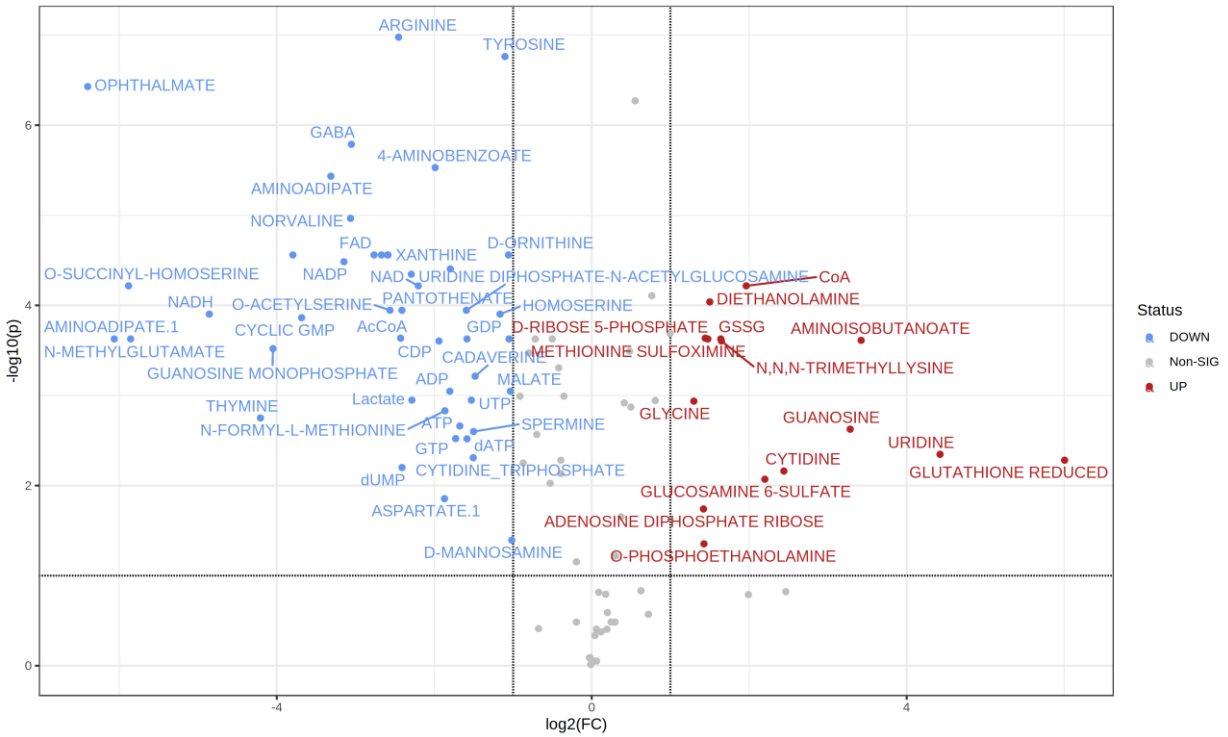


Figure 6.8 Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 2-day biofilm samples compared to WTH 2-day biofilms. Metabolites significantly less abundant at the WTL 2-day timepoint in blue and those significantly more abundant at the WTL 2-day timepoint compared to WTH 2-day timepoint in red. Metabolites that were not different between treatments in grey.

Comparison of WT low nitrogen 4-day biofilms to WT high nitrogen 4-day biofilms

Of the 76 significantly different metabolites between biofilms grown on high and low nitrogen for 4-days only 41 were found to have a fold change greater than one (Fig. 6.9). Most metabolites were in lower abundance in the lower nitrogen treatment, with three of the four most significantly different being found in lower abundance; Ophthalmate (adjusted p = 7.62E-08), ethanolamine (adjusted p = 9.76E-07) and cyclic GMP (adjusted p = 9.44E-07). Whereas glutathione disulphide (GSSG) was high significantly more abundant in the low nitrogen treatment (adjusted p = 3.82E-07). As discussed earlier glutamate was also highly significantly different between these treatments, however the fold change was not > 1.

Like the comparison between high and low nitrogen treatments for the day 2 timepoint, purine metabolism was significantly decreased in low nitrogen biofilms compared to high nitrogen growth (p = 0.0114, 8/73 metabolites, 0.25 impact). Which further suggests changes in this pathway due to the nitrogen status of the biofilm. No pathways were significantly upregulated in the biofilms grown on low nitrogen medium compared to those grown on high nitrogen at the day 4 timepoint.

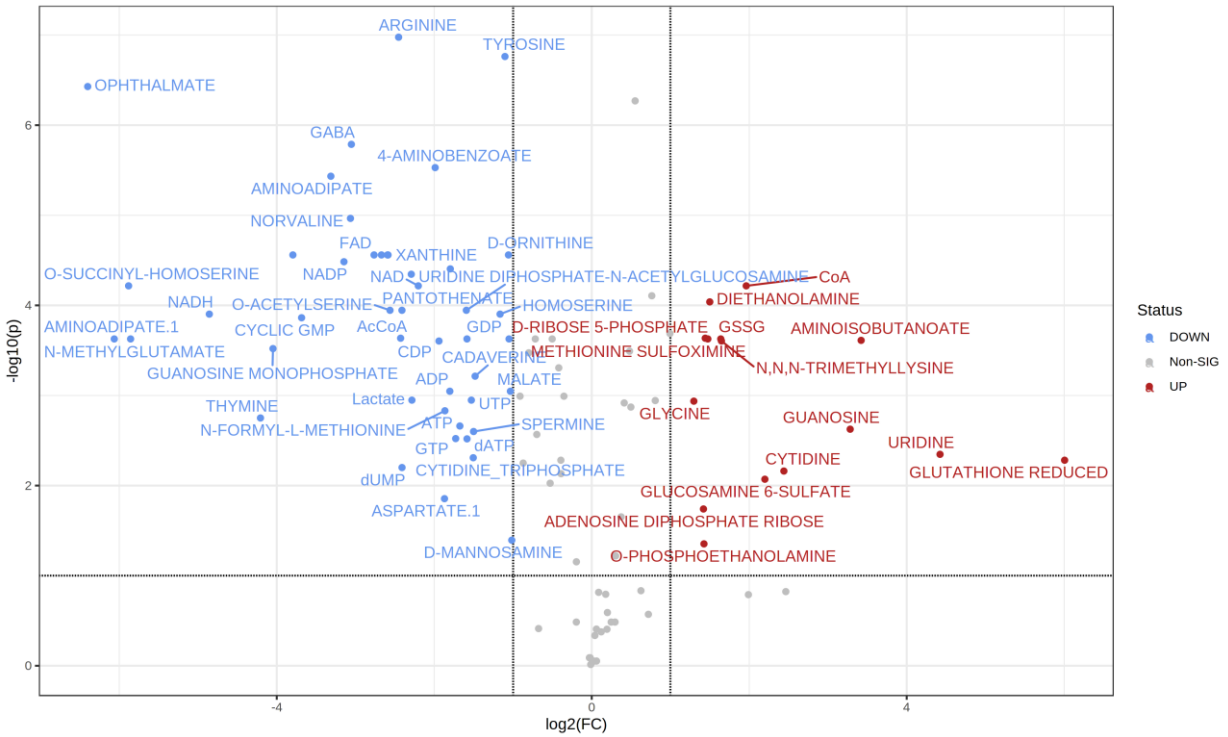


Figure 6.9 Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 4-day biofilm samples compared to WTH 4-day biofilms. Metabolites significantly less abundant at the WTL 4-day timepoint in blue and those significantly more abundant at the WTL 4-day timepoint compared to WTH 4-day timepoint in red. Metabolites that were not different between treatments in grey.

Comparison of WT low nitrogen 6-day biofilms to WT high nitrogen 6-day biofilms

Nearly all of the 102 metabolites analysed differed significantly between the high and low nitrogen treatments at the 6 day timepoint (91 metabolites where p <=0.05). Of these only 47 had a fold change great than one (Fig. 6.10), with most of these metabolites being of lower abundance in the low nitrogen treatment compared to the high nitrogen treatment. Of the four most significantly different metabolites three were in lower abundance in the low nitrogen treatment; 3-amino-4-hydroxybenzoate (adjusted p = 5.55E-08), glucosamine 6-sulphate (adjusted p = 1.19E-07) and ethanolamine (adjusted p = 1.19E-07), and one was at higher abundance 1 – methyl-L—histidine (adjusted p = 2.46E-07).

Once again, purine metabolism was significantly reduced in the low nitrogen treatment compared to the high nitrogen timepoint (p = 0.011, 8/73 metabolites, impact = 0.268). Which suggests a difference in metabolite profile and metabolism that persist across time between the high and low nitrogen treatments. However, unlike earlier timepoints one pathway was found to be significantly upregulated in the low nitrogen treatment at the day 6 timepoint, arginine and proline metabolism (p = 0.0029, 3/29 metabolites, impact = 0.11). This is due to both arginine and proline being significantly more abundant in the low nitrogen treatment after 6 days.

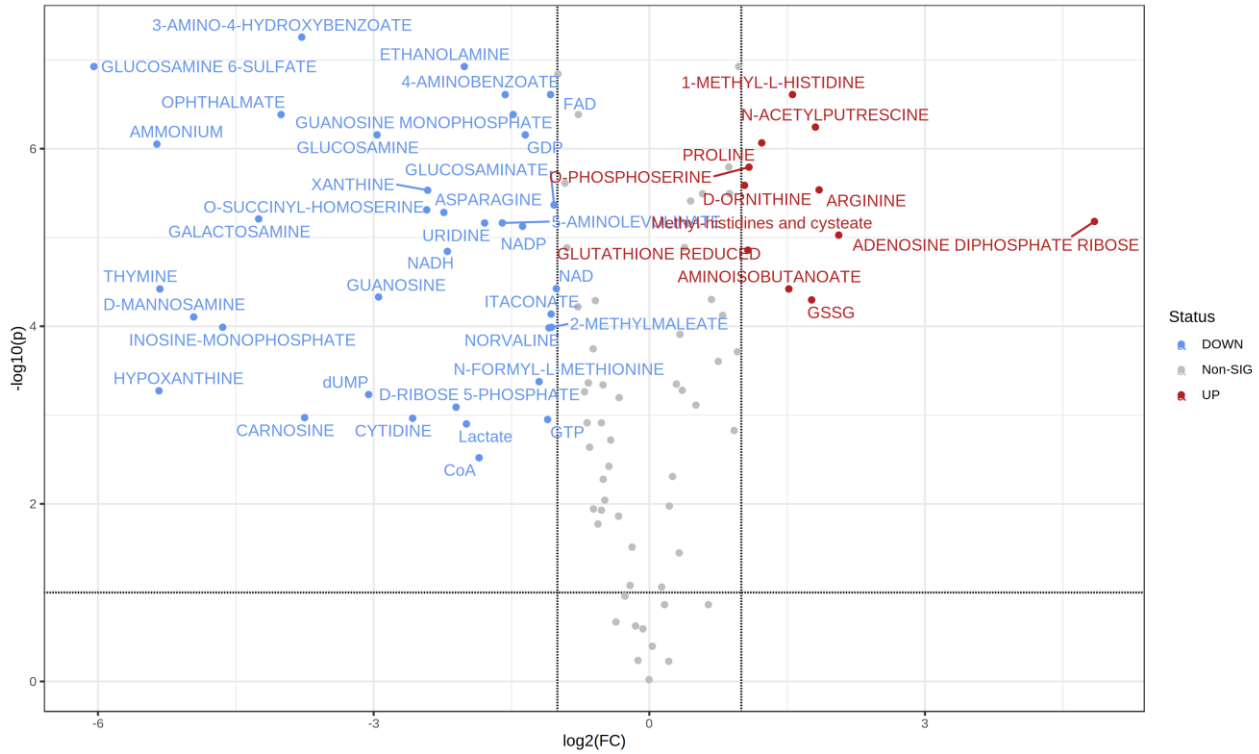


Figure 6.10 Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 6-day biofilm samples compared to WTH 6-day biofilms. Metabolites significantly less abundant at the WTL 6-day timepoint in blue and those significantly more abundant at the WTL 6-day timepoint compared to WTH 6-day timepoint in red. Metabolites that were not different between treatments in grey.

Comparison of WT low nitrogen 4-day biofilms to △amtB low nitrogen 4-day biofilms

As the multivariate analysis would suggest there are fewer metabolites significantly different abundances between these the WT and mutant at the day 4 timepoints compared to between other treatments (Fig. 6.11). Only 18 of the 55 significantly different metabolites had a fold change great than one, with a relatively even distribution between increase and decrease in abundance in the WT compared to △*amtB*. The four most significantly different were Ophhalmate (adjusted p = 8.89E-08), GABA (adjusted p = 1.53E-07), N-acetyl putrescine (adjusted p = 1.53E-07) and Ac-CoA (adjusted p = 1.53E-07). No pathways were found to be significantly up or decreased between these treatments. In the context of the rest of the results this suggests that these two treatments were similar in metabolite composition and metabolism.

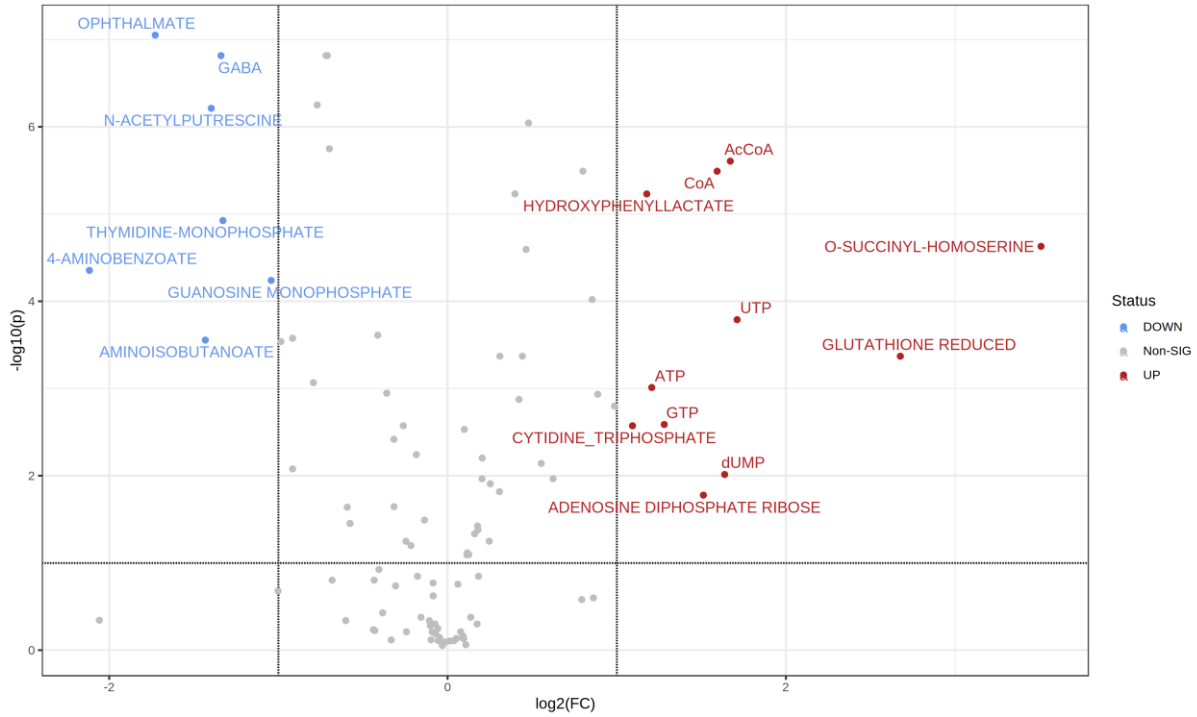


Figure 6.11 Volcano plot of univariate analysis to determine metabolites with significant differences in WTL 4-day biofilm samples compared to △amtBL 4-day biofilms. Metabolites significantly less abundant at the WTL 4-day timepoint in blue and those significantly more abundant at the WTL 4-day timepoint compared to △amtBL 4-day timepoint in red. Metabolites that were not different between treatments in grey.

6.3.6 Potential nitrogen fixation associated metabolites correlate with fixation capacity in biofilms

As discussed in Chapter 5 metabolites such as Ac-CoA, 5-aminolevulinate, malate and N-acetyl-glutamate have been shown to be associated with nitrogen fixation both in the literature and by work carried out in this thesis [176,177,197]. Some of these findings appear to carry across to nitrogen fixing biofilms grown on low nitrogen media, whereas others do not. Ac-CoA, malate and N-acetyl-glutamate all follow a similar trend to batch culture growth whereas 5-aminolevulinate does not.

Figure 6.12 B, shows that the intracellular abundance of 5-aminolevulinate in biofilms follows a different pattern to that seen in batch culture. In batch culture the abundance of this metabolite increased with nitrogen fixation (Chapter 5.3.5), however in biofilms the abundance of this metabolite is relatively high and unchanged between days 2 and 4 before decreasing sharply at day 6 (Fig. 6.12 B). Biofilms at the day 6 timepoint were shown to be fixing more nitrogen compared to the day 2 and 4 timepoints (Fig. 6.1). As with batch culture, the abundance of 5-aminolevulinate was to found to be amongst the most changed across time in the low nitrogen treatment (adjusted p = 3.84E-10). N-acetyl-glutamate also did not follow the exact same trend as seen in batch culture. In batch culture N-acetyl-glutamate increased across time, though in biofilm its abundance increased from day 2 to day 4 before appearing to plateau (Fig. 6.12 D, adjusted p = 3.67E-06).

In contrast, Ac-CoA and malate followed a similar change in abundance to that observed for nitrogen fixing *K. oxytoca* in batch culture. Both Ac-CoA and malate increased across time (Fig. 6.12 A & C) and with fixation capacity with these changes being significant across time (adjusted p = 1.67E-12 and p = 1.93E-06 respectively).

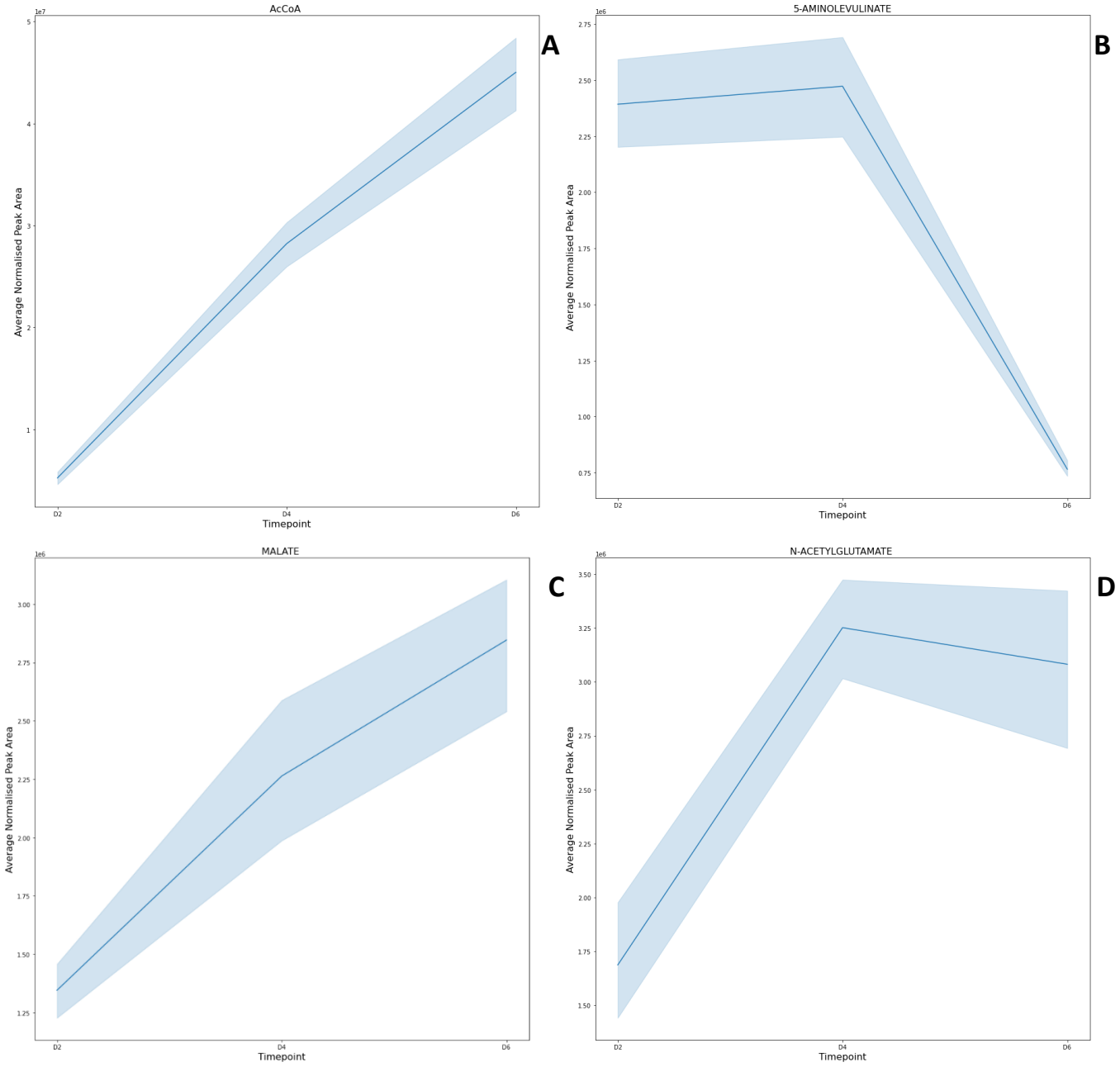


Figure 6.12 Changes in intracellular metabolite abundance over time for K. oxytoca biofilms grown under low nitrogen conditions for (A) Ac-CoA, (B) 5-aminolevulinate, (C) Malate and (D) N-acetylglutamate. Lines represent group mean and shaded areas denote 95% confidence bands, n = 7.

6.3.7 Lysine biosynthesis and degradation metabolites are amongst most significantly changed across time

When considering the metabolites that were most significantly changed across the three low nitrogen timepoints, several relating to lysine biosynthesis and metabolism were in the top 10 most significantly different. Indeed, as discussed earlier these pathways were significantly upregulated in day 6 biofilms grown on low nitrogen compared to day 2 biofilms grown on low nitrogen pre-FDR correction. The two most significantly changed metabolites were 5-aminopentanoate (adjusted p = 5.59E-21), which is involved in lysine degradation and aminoadipate (adjusted p = 1.28E-15) which is involved in lysine biosynthesis pathway. Both metabolites increased in abundance across the three timepoints (Fig. 6.13 A & B).

Ac-CoA discussed above is also involved in lysine degradation, increasing over time, and serving as a link between this pathway and the citrate cycle [215]. Other metabolites that were found to change significantly overtime and are involved in lysine biosynthesis are aspartate (adjusted p = 5.93E-08), homoserine (adjusted p = 7.93E-12) and lysine (adjusted p = 5.40E-12). Overall, these metabolites all increased across all timepoints (Fig. 6.13 C, D & E), however aspartate does not increase between day 4 and day 6 with these timepoints showing no significant difference.

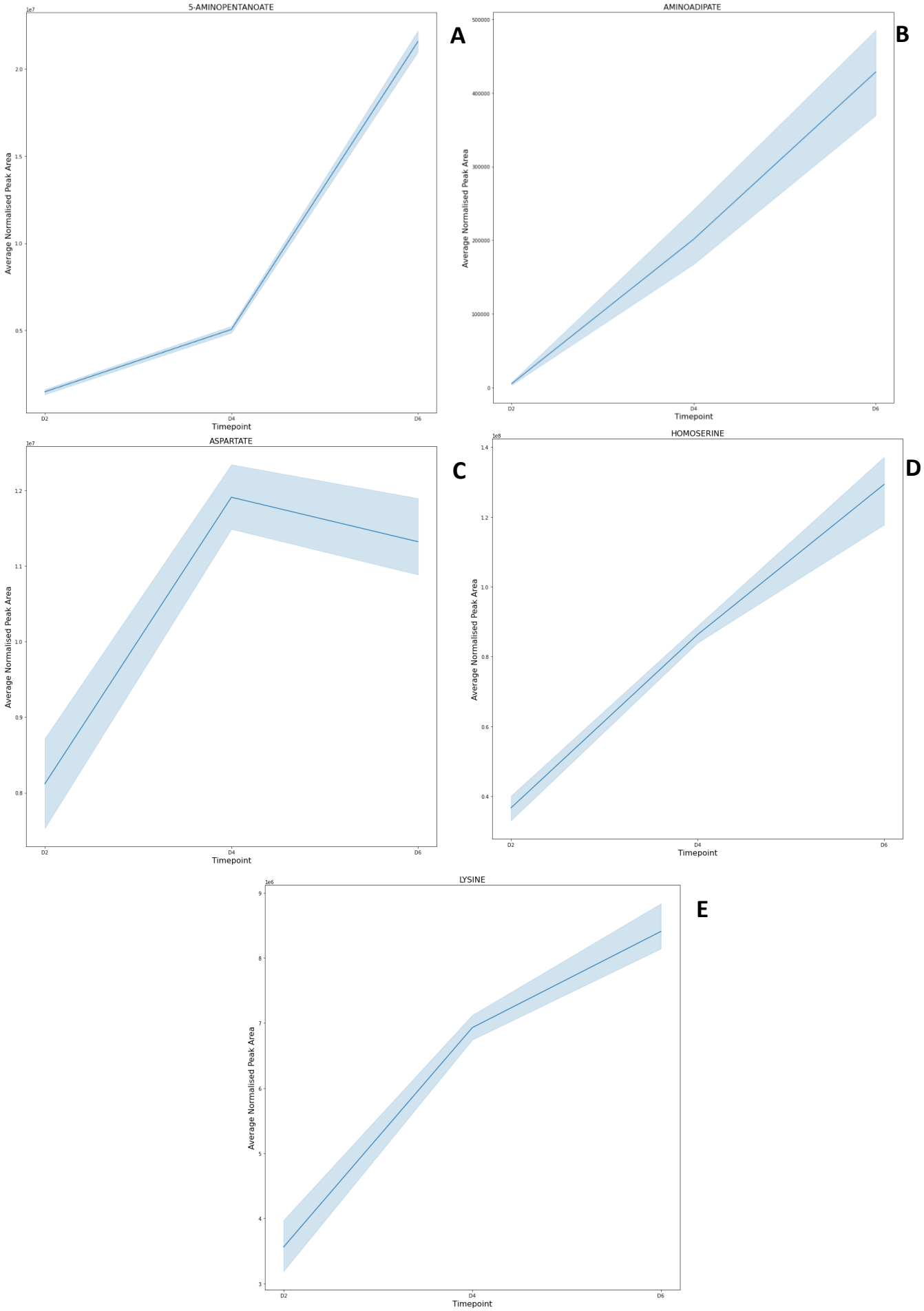


Figure 6.13 Changes in intracellular metabolite abundance over time for K. oxytoca biofilms grown under low nitrogen conditions, for (A) 5-aminopentanoate, (B) Aminoadipate, (C) Aspartate, (D) Homoserine and (E) Lysine. Lines represent group mean and shaded areas denote 95% confidence bands, n = 7.

The results discussed so far are from analysis carried out on whole biofilms, so any heterogeneity in the biofilm is not revealed, and which could of course be important but single cell metabolomics is at its infancy, and so we have to accept the ensemble averaging from bulk cells is what's possible currently. Below, my work using TOF-SIMS begins to provide evidence for the spatial separation of metabolites in N fixing biofilms.

6.3.8 'Crossover' point indicates fixation occurs in the centre of biofilms

To look at how fixation occurs with the biofilm, samples were grown on NFDM medium supplemented with heavy-labelled ^{15}N ammonium. This allowed for differentiation between nitrogen taken up from the medium and fixed nitrogen which would be the more abundant ^{14}N isotope. Semi-targeted mass spectrometry data was collected using TOF-SIMS, to investigate distribution of fixed nitrogen within the biofilm. This technique involves a process called 'sputtering' where layers of a sample are ablated between collection of MS data meaning that a series of MS spectra are collected starting from the outside moving towards the inside centre of a sample. Figure 6.14 depicts a WT biofilm, imaged using the TOF-SIMS camera, these images are visualised in real time and are used to determine the best place to sputter into the sample. Using the data collected, one can follow how a specific peak intensity changes between spectra, informing how the abundance of an ion changes moving deeper into the sample.

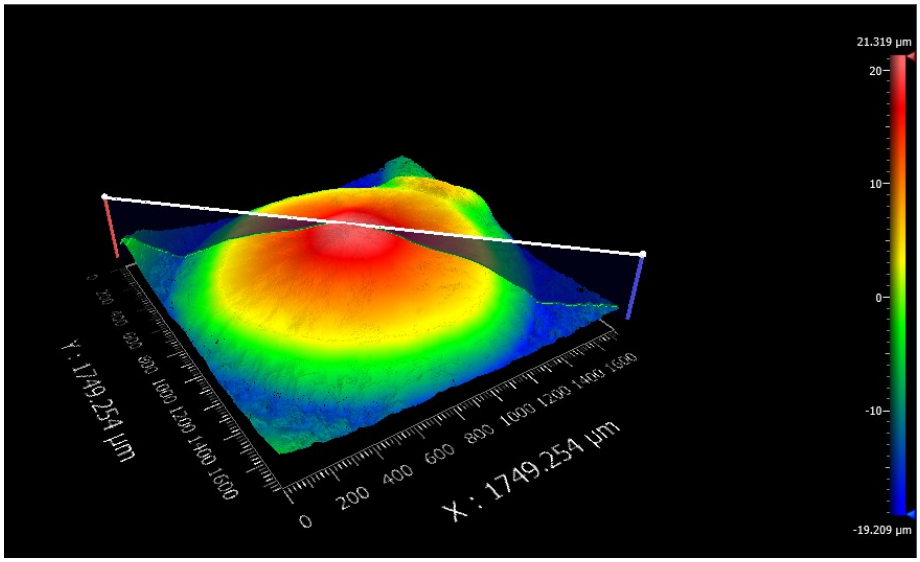


Figure 6.14 Whole biofilm imaged using camera attached to TOF-SIMS instrument. Coloured in relation to height with red representing the highest point and blue the lowest.

The data collected using this technique, allowed us to track how the abundance of ^{14}N and ^{15}N containing ions change in relation to biofilm depth. In total 53 ions were detected, followed and measured through the biofilm. Figure 6.15 shows how the abundance of the $^{14}NH_4$ and $^{15}NH_4$ ions change throughout the biofilm at two-, four- and six- day timepoints. A day 4 △*nifLA* non-fixing control was included to confirm that ^{14}N nitrogen does not get integrated into the biofilm by a means other than fixation. While not directly comparable between samples, a larger sputter time represents a datapoint from deeper in the sample and the △*nifLA* biofilm was sampled to a depth approximately equal to or greater than that of the WT biofilm samples. The data clearly confirms that no fixation occurred for the △*nifLA* negative control and more importantly that ^{14}N nitrogen does not accumulate in the biofilms nor get integrated into the EPS matrix of living cells via a means other than fixation, as only $^{15}NH_4$ was detected in the sample (Fig 6.15. D).

In contrast, a clear pattern of ^{14}N assimilation via fixation was observed for the WT biofilms (Fig. 6.15 A-C). An increase in abundance of the $^{14}NH_4$ ion across sputter time, coupled with a corresponding decrease in $^{15}NH_4$ indicates that fixed atmospheric nitrogen being the sole nitrogen source deeper in the biofilm. Two things are noteworthy about this pattern. First, is that this crossover event happens at a shallower depth for the day 2 sample compared to the day 4 and 6 samples, likely because day 2 samples are smaller and less developed than the later timepoints. The cross over event happens at a similar depth for the day 4 and day 6 timepoints.

The second point is that the differentiation between the $^{14}NH_4$ and $^{15}NH_4$ ions is more pronounced overtime. At the day 2 timepoint the difference in abundance between the two ions is small and in fact the $^{15}NH_4$ ion remains at a greater abundance than the $^{14}NH_4$ ion even when fixed nitrogen begins to be assimilated. However, by the day 6 timepoint, the difference in ion abundance is more clearly pronounced, with the $^{14}NH_4$ ion abundance being much higher than that of the $^{15}NH_4$ ion, suggesting at this depth ^{14}N nitrogen from the atmosphere is predominantly being used.

Overall, these points led us to conclude that there are three ‘layers’ to the biofilm. Moving from the outer layer inwards, the first layer is a ‘non-fixing’ layer which uses nitrogen only from the media. The second layer is a ‘transition’ layer which uses both medium nitrogen and fixed nitrogen, this layer includes the ‘crossover point’ – where fixed nitrogen becomes more abundant than nitrogen from the medium. The final layer in a ‘fixing’ layer where fixed nitrogen is the predominant nitrogen being used. While the trends determining these layers is seen nearly all the nitrogen containing ions measured, the assignment of labels will be done using the NH_4 only to avoid bias (e.g. 1 out of 53 ions). For the

purposes of discussion throughout this chapter and the rest of this thesis these layers will be referred to by these terms.

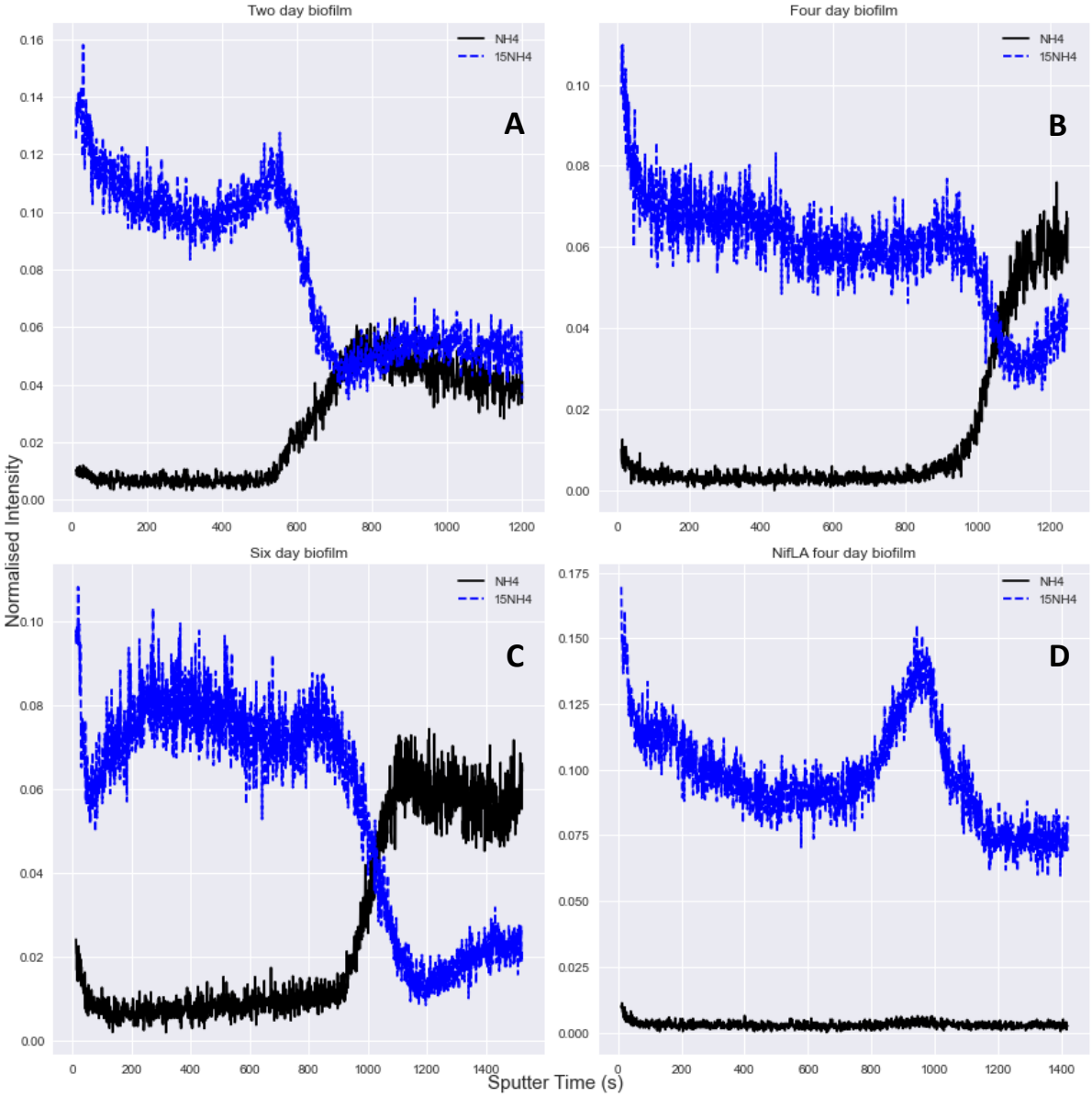


Figure 6.15 The abundance of the $^{15}NH_4$ and $^{14}NH_4$ ions in relation to sputter time, a proxy for sampling depth into the biofilm, for a single WT biofilm sample grown on low nitrogen medium at (A) Two days, (B) Four days (C) Six days. (D) A non-fixing ⍙nifLA control sampled at four days. Ion intensity has been normalised to total ion intensity to allow for comparisons through time, n = 1.

6.3.9 Multivariate analysis of TOF-SIMS data reveals spatiotemporal differences between 'layers' in nitrogen fixing biofilms

The PCA scores plot below (Fig. 6.3.10), shows that the metabolic profiles of biofilms group based on timepoint and 'layers' as outlined above. All three layers can be distinguished by timepoint and grouping follows the order of the timepoints, i.e., the day four timepoint groups fall between day 2 and day 6 timepoint groups for all layers. The fixing and non-fixing layers group separately indicating metabolic profiles distinct from each other. The transition layer for day 2, 4 and 6 timepoints show clear grouping although the day two transition is more differentiated compared to the day 4 and 6 timepoints.

Interestingly, the grouping of the non-fixing layers for the day 4 and 6 timepoints overlaps, although the day 4 non-fixing layer does not have tight grouping. Additionally, while visibly distinct there is a small amount of overlap between the transition layers for these timepoints. However, the grouping of the fixing layers for the day 4 and 6 timepoints is more distinct from one another suggesting greater metabolic differences between the fixing layer compared to the non-fixing and transition layers.

The △*nifLA* non-fixing control overlaps with both the day 2 non-fixing layer and day 4 non-fixing layer, and is close to overlapping with the day 6 non-fixing layer. Which provides further evidence of no fixation occurring in this layer.

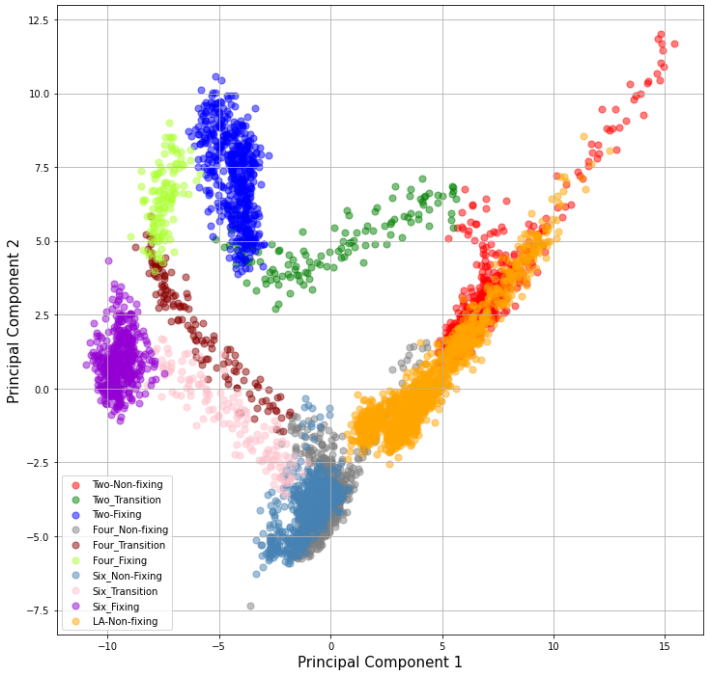


Figure 6.16 PCA scores plot comparing how metabolic profiles of WT and ⊿nifLA biofilms grown on low for 2, 4 and 6 days change spatially within the biofilm. Each dot represents an MS spectrum from a given point within the biofilm, n = 2.

6.3.10 Multivariate analysis of exploratory RNA-Seq data shows clear differences in the transcriptome based on nitrogen status and fixation

Global gene expression associated with nitrogen fixation in biofilms was assessed by comparing the transcriptomes of WT biofilms grown on high and low nitrogen and of a non-fixing control △*nifH* also grown on low nitrogen. Given the small sample size and subsequent reduced statistical power, the analysis of this dataset will be exploratory in nature focusing on the overall transcriptome profiles and genes exhibiting large fold changes.

Considering first the overall differences in transcriptome profile. The PCA scores plot in Figure 6.17, shows that the transcriptomes of WT nitrogen fixing biofilms are distinct from the transcriptomes of both non-fixing WT biofilms grown on high nitrogen and non-fixing △*nifH* biofilms grown on low nitrogen. All three treatments showed clear grouping, with the greatest separation observed between the △*nifH* biofilms and the WT biofilms, along PC1. This is likely because the △*nifH* biofilm was under extreme nitrogen stress compared to the other two treatments. PC2 appears to separate based on fixation with the fixing WT biofilm being grouped away from the other treatments along this component. Overall, these results indicate that both nitrogen stress and nitrogen fixation have a large effect on transcriptome composition and therefore on gene transcription.

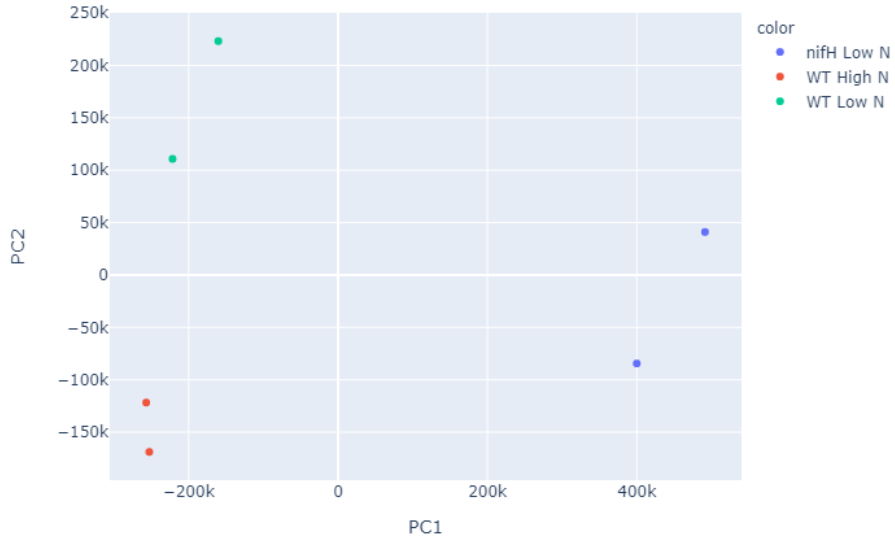


Figure 6.17 PCA scores plot comparing the transcriptome profiles of WT and ⊿nifH biofilms grown on either low or high nitrogen medium for 5 days. Each dot represents a sample, n =2.

6.3.11 Genes relating to nitrogen stress and nitrogen fixation are regulated in low nitrogen conditions

When comparing the transcriptomes of WT fixing biofilms grown on low nitrogen to WT non-fixing biofilms grown on high nitrogen 1228/5279 genes were differentially expressed (Fig. 6.18 A, 514 down-regulated and 714 up-regulated, in nitrogen fixing biofilms adjusted p = <0.05). Given that glucose is not thought to be limited and oxygen exposure is constant between these treatment groups, nitrogen stress and fixation is thought be driving these global changes in metabolism and growth transcription. All of the *nif* genes were found amongst the top 50 up-regulated genes for nitrogen fixing biofilms, with *nifK* and *nifD* being in the top 10 most significantly upregulated genes. However, the most upregulated gene in nitrogen fixing biofilms compared to non-fixing WT biofilms was norW, which is involved in the detoxification of nitrous oxide. This gene was also highly upregulated in the △*nifH* biofilms which suggests it is up-regulated in response to nitrogen stress, rather than fixation. This could be due to nitrogen scavenging a diverse range of nitrogen sources and their metabolic products. In addition to the *nif* genes many genes related to nitrogen stress that are NtrC-dependant were also upregulated including *amtB* (in the top 10 most upregulated, responsible for active transport of ammonium into cells), *glnA* (responsible for GS pathway), *glnK* (PII-type signaling protein) and the *rut* gene (pyrimidine scavenging) (Fig. 6.18 & Fig 6.19). In contrast to the up-regulation of *glnA*, *ghd* is significantly decreased by more than two-fold in fixing biofilms. This highlights the limited role glutamate dehydrogenase (GHD), plays in nitrogen assimilation during grown in nitrogen deplete conditions.

Earlier work carried out in this laboratory [39], putatively assigned additional operons to the NtrC regulon which were found to be upregulated in nitrogen fixing conditions. These are related to nitrogen scavenging and metabolism of nitrate/nitrite (KO: NrtABC-NirB-NasA) and urea (KO: UreABC-UrtABCDE) and purine scavenging (KO: HpxABHpxO-HpxWXYZ). All three were found to be significantly highly upregulated in this study of biofilms.

When comparing the transcriptomes of WT fixing biofilm with Δ*nifH* non-fixing biofilms on low nitrogen 2009/5297 genes were differentially expressed (Fig. 6.18 B, 987 up-regulated, 1022 in nitrogen fixing biofilms, adjusted p = <0.05). Whilst the Δ*nifH* mutant is unable to fix nitrogen due to its inability to produce the *nifH* homodimer it does retain the ability to express the other *nif* genes. Therefore, this mutant was chosen as it allows for a comparison of gene expression, including of the *nif* genes under nitrogen stress when fixation doesn't occur. *nifA* was among the 10 most down-regulated genes in the WT fixing biofilm compared to the Δ*nifH* biofilm grown on low nitrogen. Other

nif genes including *nifD and nifK* were also significantly different between these treatments and saw a large fold change increase in Δ*nifH* biofilms (Fig. 6.18 & Fig. 6.19). Furthermore, many of the nitrogen stress related genes discussed above, including *glnA*, *glnK, amtB* and *rut*, were significantly upregulated in Δ*nifH* biofilms meaning that these biofilms were under greater nitrogen stress compared to fixing WT biofilms and that fixation alleviates some degree of nitrogen stress, while not stopping the expression of nitrogen fixation genes.

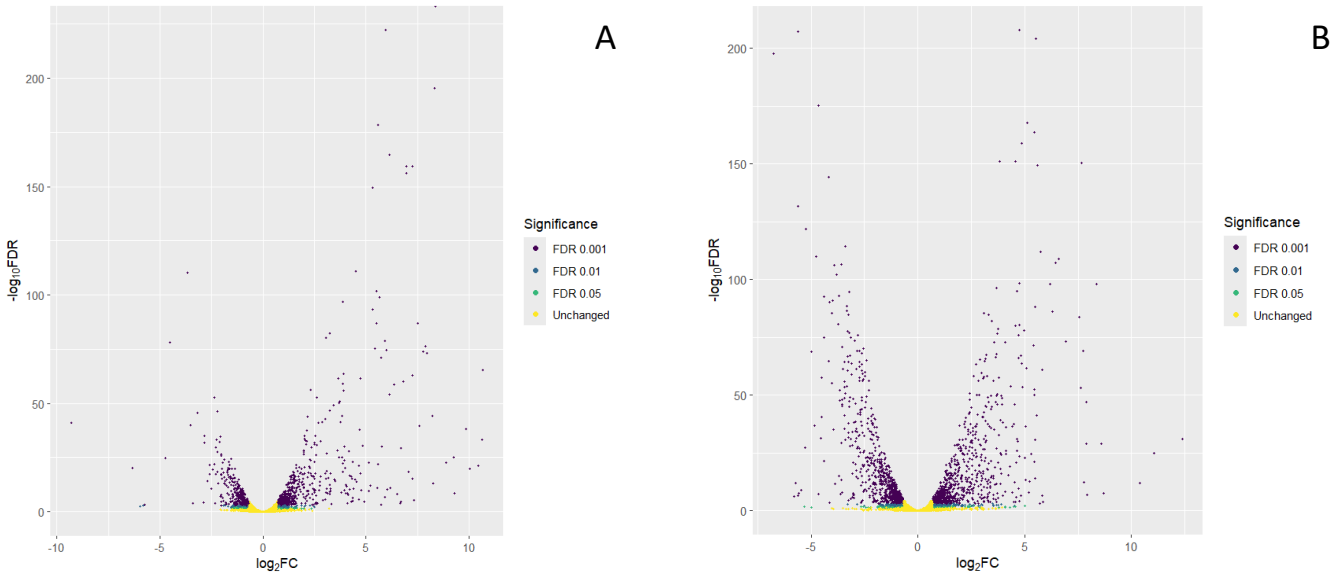


Figure 6.18 Volcano plot showing differential gene expression with significance between (A) Differentially expressed genes in WT fixing biofilms grown on low nitrogen compared to WT non-fixing biofilms grown on high nitrogen. (B) Differentially expressed genes in WT fixing biofilms grown on low nitrogen compared to ΔnifH non-fixing biofilms grown on low nitrogen.

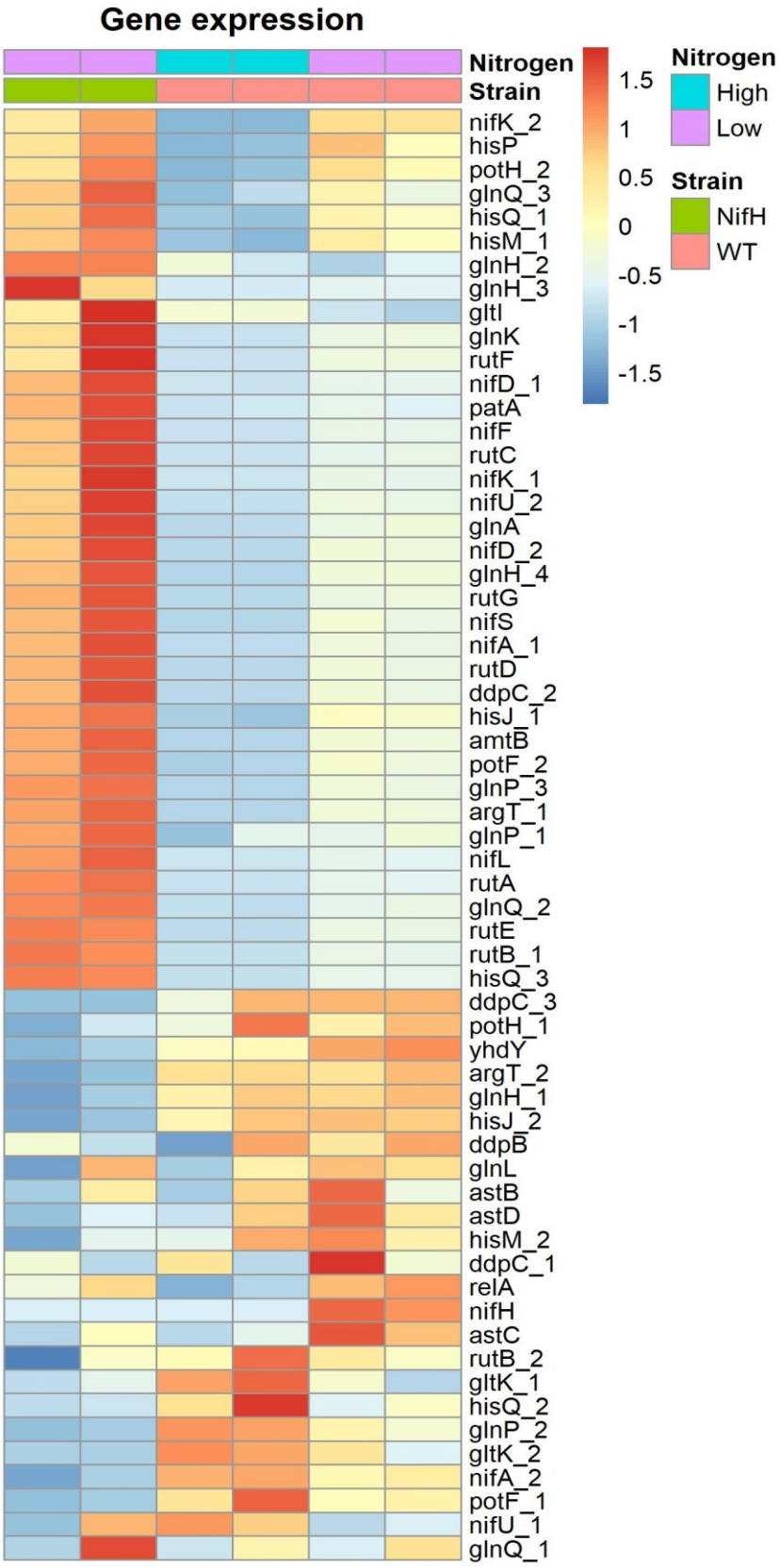


Figure 6.19 Heatmap of the log2 fold change relative to mean RNA-seq read count of nif genes and NtrC-dependent genes associated with the bacterial nitrogen stress response for K. oxytoca biofilms.

6.4 Discussion

The work carried out in this chapter has applied a multi-omic approach to understand changes in metabolism resulting from nitrogen fixation in *K. oxytoca* biofilms. The biofilm growth phenotype has been shown to be a successful means of plant associative bacteria to colonise plant roots in the rhizosphere [212]. Therefore, robust data that enables understanding of changes in biofilm metabolism would greatly benefit our understanding of this process in nitrogen fixing bacteria. As with earlier batch culture experiments, this study used ammonium run out to induce a reliable nitrogen fixation response in *K. oxytoca* biofilms. The approach allowed for the changes in biofilm metabolism to be tracked overtime and for the same run-out culturing technique to be used for all analytical methods. *K. oxytoca* biofilms were sampled from 2 - 6 days for and analysed using metabolomic, TOF-SIMS and RNA-sequencing technologies which showed distinct metabolic and transcriptome profiles to be shown between fixing and non-fixing cells. While there are few papers on nitrogen fixing biofilms, and none focused on metabolomics, the overall results in this chapter are consistent with changes seen in both nitrogen starvation response for *K. oxytoca* and similar species [39,47,181]. As both the metabolomic and transcriptome data show a trend of an Ntr controlled genes being activated in response to nitrogen starvation. Additionally, the joint response to nitrogen fixation is similar to that seen in nitrogen fixing cells grown in batch culture, yet there are some growth phenotype related differences.

6.4.1 Increased △*amtB* fixation capacity is surpassed by WT over time

Research carried out in this thesis and in the literature has shown that a △*amtB* knockout mutant is able to fix more nitrogen compared to WT cells for a range of nitrogen fixing species [51,60,107,153]. In batch culture these results have been found to be consistent over time, with a study of *A. vinelandii* finding higher levels of ammonium in the supernatant 98hrs into culturing for the △*amtB* knockout compared to WT [60]. However, the difference between the WT and △*amtB* mutant decreases overtime and is nearly equal at the 98-hour timepoint. Indeed, in batch culture of *K. oxytoca* carried out in Chapter 3 the △*amtB* mutant had fixed more nitrogen at the 10-and-24-hour timepoints, when assayed using ARA. Further, the results in Chapter 5 show greater fixation by ARA at the 12.5-hour timepoint and a metabolic profile comparable to WT cells at 24 hours for △*amtB* cells sampled at 12.5-hours. This is thought to be due to the cells being under greater nitrogen stress compared to WT due to the △*amtB* mutants impaired ability to actively import ammonium back into the cell [51,60]. It is likely that this is a mechanism to maximise the amount of fixed nitrogen that the fixing cells can retain and so helps regulate nitrogen stress and therefore fixation in cells. In fact, *amtB* is one of a number of nitrogen scavenging genes up-regulated in nitrogen stressed cells in both *E. coli* [181] and in *K. oxytoca* known to be fixing nitrogen [39].

The results in this chapter go some way to support these earlier findings in nitrogen fixing biofilms, with the △*amtB* mutant shown to be fixing more than the WT from days 2 to 4 when assayed using ARA. However, at day 5 the WT and △*amtB* mutant are found to be reducing a similar amount of acetylene and by day 6 the WT is reducing more acetylene than the △*amtB* mutant. Although the trend between the two shows that the △*amtB* mutant was levelling off in terms of fixation capacity and the WT was still increasing. Additionally, the metabolic profile of day 4 △*amtB* biofilm is most similar to the day 4 WT biofilm, suggesting that the biofilms are behaving in a similar way metabolically (discussed further below). Therefore, together this is some of the first evidence that given sufficient time the WT strain could outperform the △*amtB* mutant, possibly due to a fitness cost associated with the knockout [60]. For instance, after 4-days the △*amtB* mutant could be reaching the limit of its nitrogen resources from fixed nitrogen and therefore is slowed in its ability to produce more cells or nitrogenase enzyme. In contrast, while small, the advantage offered by the WT's ability to import ammonium back into the cell means that when the WT 'catches up' to the △*amtB* It could have slightly more resources available. These marginal gains are worth considering in biotechnological applications of nitrogen fixing biofilms using △*amtB* related genes in complex soil environments to seek an optimisation of mutants.

6.4.2 Distinct metabolic phenotypes seen between fixing and non-fixing *K. oxytoca* biofilms

Overall, the 'omic' data relating to free living nitrogen fixing biofilms is limited [199], however there is robust data relating to medically relevant biofilms that show clear metabolic differences due to virulence and strain [204]. Additionally, clear time-dependant metabolic differences were found between fixing and non-fixing *K. oxytoca* in batch culture. Therefore, it is unsurprising that we find clear differences in metabolite composition between nitrogen fixing biofilms due to age suggesting that fixation within biofilms is a dynamic process that changes overtime in relation to environmental cues. As with batch culture cells, fixation did not return to biofilms to a nitrogen replete state, with metabolic profiles between fixing biofilms being distinct from non-fixing biofilms on high nitrogen medium for all timepoints.

The grouping of the treatments also tracks with timepoint for both fixing and non-fixing WT biofilms, e.g. the day 4 timepoint falls between the day 2 and 6 timepoints. This could indicate changes in biofilm composition that is common between nitrogen treatments and is due to biofilm aging/maturation rather than nitrogen stress. Indeed, studies into the relationship between biofilm age and antimicrobial tolerance, older more mature biofilms were found to be less susceptible, with this

increased defence occurring over a period of several days [216]. Therefore, it is possible that nitrogen fixing biofilms could take some time to reach maturation, possibly longer than 6 days and future work should carry out more long-term investigation into this topic.

Unlike in batch culture the △*amtB* positive control was not a distinct 'super fixer', with its metabolite composition being very similar to the that of the WT at the same timepoint. One might have expected a metabolic profile comparable to that of 6-day old biofilms were the assumptions of batch culture to be carried forward. Perhaps this serves as an example of how assumptions between growth phenotypes cannot be carried forward, as outlined by other comparisons of batch and biofilm for the same species [217].

6.4.3 Biofilm amino acid and ammonium abundance changes over time under nitrogen fixing conditions

Given that the biofilms sampled in this work were already under nitrogen stress and fixing atmospheric nitrogen, changes in metabolite abundance across time are likely primarily due to nitrogen fixation rather than transitioning from nitrogen replete to depleted conditions. It is therefore difficult to compare to studies that investigate changes in metabolite composition as a cell transitions into nitrogen stressed state (As was done in Chapter 5). However, the results in this Chapter largely agree with changes in amino acid abundance resulting from nitrogen fixation published in the literature [39,184].

Changes in abundance were seen for two of the key nitrogen signalling molecules glutamine and glutamate. Glutamine abundance remained much lower than glutamate abundance across all three timepoints, indicating nitrogen stress, which is consistent with previous results for nitrogen fixing *K. oxytoca* [39] and *Zymomonas mobilis* [184]. As in Chapter 5, glutamine abundance changed over time, with intracellular biofilm glutamine abundance increasing between days 2 and 4, before decreasing slightly between day 4 and 6. Evidence of the changing nature of nitrogen fixing biofilms, it could suggest that the biofilms have not reached maturation after 6 days. This is further supported by the fact that glutamate abundance decreased significantly over time. These findings differ from batch culture results in Chapter 5.3 and from results previously published research in batch culture regarding glutamate abundance under prolonged nitrogen stress [178,183,185]. However, these results agree with previously published work in this laboratory that found glutamate levels decreased overtime as *K. oxytoca* transitions to diazotrophy [39].

Of the remaining amino acids analysed, all bar isoleucine were significantly changed overtime with 11 generally decreasing overtime and 6 generally increasing overtime. It is difficult to compare these changes in abundance to previous work because most studies focus on early transition [39,184]. While some changes in abundance are similar to published data and data from Chapter 5, e.g. a decrease in methionine and isoleucine abundance, there are some disagreements that suggest an overall different amino acid composition [39]. For example, alanine increases significantly overtime, a result not seen in the transition to diazoptrophy in *K. oxytoca* [39] or *Z. mobilis* [184].

The lysine biosynthesis and degradation pathways were up-regulated in 6 day old fixing biofilms compared to 2 day old fixing biofilms. This is due to several of the metabolites involved in these pathways increasing over time and could speak to lysine playing an important role in nitrogen fixing biofilms. One way this could be the case is through its role in cell wall synthesis. Amino acids also play a role in cell wall synthesis and repair being a constituent part of the bacterial peptidoglycan [218]. The three amino acids that make up with pentapeptide chain of the peptidoglycan monomer are alanine, glutamine and lysine [219]. Glutamine abundance was increased from day 2 to day 4 and remained high to day 6 additionally, alanine and lysine increased across all three timepoints. While amino acids have many roles in the cell that all three metabolites were increased with biofilm maturation, a trend not seen in batch culture, suggests that up-regulation of cell wall synthesis and repair could be biofilm specific. If so, this could be due to the cell-cell contact experienced in biofilms compared to batch culture [200].

6.4.4 Spatiotemporal differences observed for nitrogen fixing biofilms

One of the ways biofilms are characterised is by their heterogeneity and layers [67,200]. A typical way that layers organise themselves in biofilms is along an oxygen gradient, with the cells on the outside respiring aerobically, consuming oxygen and creating an anaerobic inner layer that respires anaerobically [200,220,221]. Given that nitrogen fixation requires microaerobic or anaerobic conditions [26] and that nitrogen fixing biofilms can fix nitrogen in air [66], we believed nitrogen fixation to be occurring in the centre of the biofilm. Previous research has shown that bacteria growing aerobically in outer layers of the biofilm are able to remove oxygen from the air efficiently enough to allow anaerobic growth to occur within the inner layers [220]. Therefore, for nitrogen fixing bacteria it is possible the air that reaches the cells at the centre of the biofilm is sufficiently depleted of oxygen to allow for nitrogenase formation and function. One way to investigate this is to use TOF-SIMS a mass spectroscopy imaging technique which gives spatial resolution of metabolites and has been successfully used in biofilms previously [222].

Biofilms were grown on heavy labelled $^{15}N\ NH_4$, meaning that any ^{14}N nitrogen in the sample had to have come from fixed atmospheric nitrogen. This allowed for the ratio $^{15}N/^{14}N$ of nitrogen containing ions to be traced from the outside of the biofilm inwards. The results show a clear trend across all nitrogen containing ions where the contain ^{15}N on the outside and ^{14}N deeper into the biofilm with a clear crossover point. This crossover happens over a relatively short distance in the scheme of a biofilm, suggesting that cells in the outer layer are removing oxygen but not nitrogen from the air, efficiently enough to create anaerobic conditions. This is in line with previous research suggesting that biofilm respiration can change from aerobic to anaerobic over a few micrometres [212,220]. It is proposed that the cells on the outer layer are able to consume oxygen faster than is can diffuse through the biofilm, resulting in anaerobic layers [200]. These results strongly suggests that the cells on the inside of the biofilm are fixing nitrogen and those on the outside are not, suggesting at minimum two layers. Arguably a third layer should also be considered as it links the two layers but is distinct from both metabolically.

In addition to the differentiation of layers within a nitrogen fixing biofilm, the fixing and transition layers appear to vary with age, which agrees with the targeted metabolic profiling of whole biofilms. That two separate analytical techniques indicate the same phenomena is further evidence of continued metabolic changes over time in nitrogen fixing biofilms. One would suggest that any future biotechnological approaches concerning nitrogen fixing biofilms target the inner fixing layer at an appropriate stage in biofilm maturation.

6.4.5 Distinct transcriptome phenotypes seen between fixing and non-fixing K. oxytoca biofilms

When comparing the transcriptome between fixing biofilms and non-fixing biofilms grown in nitrogen replete conditions approximately 20% of the genes were differentially expressed. This is lower than the 50% of genes differentially expressed when comparing the transcriptomes under the same conditions in batch culture of *K. oxytoca* [39]. Similarly, to the results in batch culture though is that 14 of the 25 canonical NtrC/Nac-regulated operons found differentially expressed in *E. coli* via microarray were also upregulated in this study. Further building on earlier work in this lab, the putatively asigned NtrC-dependant operons e.g. urea/nitrate/nitrite/purine metabolism found to be upregulated in batch culture were also significantly upregulated here.

Given the role *K. oxytoca* has played in biotechnological approach to the nitrogen crisis [19,30,46], it is surprising that its NtrC dependant regulon has not been fully characterised. Especially given NtrC's role in nitrogen stress and fixation. While there is a high degree of similarity between the genomes of *E. coli* and *K. pneumoniae* there are nearly 800 unique genes, which could correspond to novel metabolic pathways such as nutrient scavenging, utilization and synthesis [223].

6.4.6 Nitrogen stress persists in nitrogen fixation in biofilms

Collectively the results from the metabolomic and transcriptomic analysis show nitrogen stress continues even after prolonged periods of nitrogen fixation. The metabolic profiles of fixing biofilms remain distinct from non-fixing biofilms after 6-days, with key nitrogen status signalling molecules such as glutamine remaining low compared glutamate indicating prolonged ammonium assimilation via the GS-GOGAT pathway [47,181,183]. Furthermore, both the *nif* genes and the nitrogen stress mediated Ntr genes, including *amtB* and *GS* are among the most differentially expressed between these groups after 5-days. These results support earlier findings which have indicated that nitrogen stress persists during fixation, only abating if cells are returned to nitrogen replete conditions [26,39,184]. Specifically, nitrogen stress is required for continued expression of GS and subsequent expression of the *nif* genes through the NtrC controlled regulatory cascade. However, comparison of the non-fixing △*nifH* mutant's transcriptome to the WT's suggest that fixation is managing nitrogen stress to some degree. The transcription levels of nitrogen scavenging genes such as *amtB* and *rut* as well as the *nif* genes themselves are significantly increased in the △*nifH* mutant compared to the WT.

6.4.7 Metabolites increased by fixation in batch culture show a similar trend in biofilms

TCA cycle metabolites show similar trend to fixing cells in batch culture

As discussed in Chapter 5, malate and Ac-CoA were both found to be increased in nitrogen fixing cells in batch cultures, with both metabolites are involved in the TCA cycle among others. The results here show that both metabolites also increased across time with biofilm maturation. In fact, the abundance of Ac-CoA increased approximately fivefold between day 2 and 6 samples. In addition to the high energetic demands of nitrogen fixation, the production of EPS also requires energy to produce [224]. Therefore, the altered TCA cycle metabolites over time, where both nitrogen fixation and biofilm maturation occur is not unexpected.

Not only has carbon metabolism, and the Crc gene specifically, been shown to be a regulator of nitrogen fixation, in *Pseudomonas stutzeri* a nitrogen fixing bacteria [190], this bacteria has also been shown to be able to form nitrogen fixing biofilms [199]. Research into the closely related species

Pseudomonas aeruginosa found that Crc plays a role in regulating biofilm formation as well [225]. O'Toole *et al.,* found that the Crc gene was necessary for biofilm formation in *Pseudomonas aeruginosa*, and go on to suggest that mechanism is via its role in the transduction of nutritional signals that go on to initiate biofilm formation [225]. Indeed, when comparing the proposed regulatory cascades of both nitrogen fixation and biofilm formation the RpoN regulon is implicated in both, and the Crc genes falls within this regulon. Additionally, several studies have shown than biofilm formation is sensitive to environmental cues [52,203,206,209,216], as is nitrogen fixation, so it is possible that these could be regulated in by the same gene in *K. oxytoca.* Additionally, the difference in Ac-CoA in low nitrogen treatments compared to high nitrogen treatments decreases over time indicating that cell carbon metabolism could be becoming more similar to unstressed cells over time.

N-Acetylglutamate increased between two-and four-day old biofilms

N-acetylglutamate was seen to be increasing between the two- and four-day biofilms a time where other experiments have shown an increase in nitrogen fixing capacity via ARA. These results are in agreement with those from batch culture for *K. oxytoca* (Chapter 5) and for *Z. mobilis* [184]. However, the abundance remains relatively similar between the day 4 and day 6 timepoints. This could suggest metabolic changes relating to broader biofilm metabolism rather than strictly fixation occurring between these timepoints as metabolites such as amino acids and Ac-CoA continue to increase during this period.

N-Acetylglutamate has been shown to increase root hair and tip growth meaning that were *K. oxytoca* biofilms to excrete this metabolite into its surrounding environment its beneficial effects would be local to the roots [177]. Additionally, as covered in more detail above, plant associative bacteria that form biofilms are known to have success in colonising the root and forming mutualistic relationships [212]. Thus, biofilm formation in *K. oxytoca* seems the most likely method of plant associative interaction, or root colonisation. As such, that N-acetylglutamate is found in high abundance in nitrogen fixing biofilms provides further evidence that this could be a metabolite exchanged between plant and bacteria.

5-Aminolevulinate differentially abundant across time in nitrogen fixing biofilms

5-aminolevulinate was another metabolite with potential plant growth promoting properties found to increase with fixation in batch culture. This metabolite has been shown to have multiple effects on plant growth, including increasing shoot and root weight as well as mediating chlorophyll production [196,197]. While abundance of this metabolite remains relatively high over all three timepoints, when compared to all detected metabolites, its abundance was found to change significantly overtime.

Abundance was high and comparably similar between the day 2 and day 4 timepoints before decreasing by approximately 70% between days 4 and 6, enough to be a key differentiating metabolite between days 2 and 6 of biofilm growth.

This metabolite is produced from glutamate via glutamyl-tRNA and glutamate-1-semialdehyde using the HemA and HemL [195]. In *Enterobacteriaceae* the *Hem* genes are controlled by the iron sensitive Fur and it's orthologs, where in the presence of high levels of ion Fur binds DNA and the transcription of the downstream genes is repressed [226–229]. Therefore, one would expect that in the presence of increased iron levels within *K. oxytoca* cells the production of 5-aminolevulinate to be decreased. Data in this chapter strongly suggests the fixation increases between the day 4 and day 6 timepoints and to do this cells would have to produce more nitrogenase. Given that the nitrogenase in *K. oxytoca* is comprised of the reductase Fe protein and the catalytic MoFe protein [19] (discussed in more detail in Chapter 1) cells would have to scavenge more iron for the environment, potentially increasing intracellular abundance. Thus, the decrease in 5-aminolevulinate between days 4 and 6 could be due to an increase in nitrogenase production.

6.4.8 Limitations and future work

The work in this chapter benefits from a multi-omic approach which together provides a coherent message of nitrogen fixation in *K. oxytoca* biofilms. However, each analysis is carried out on separate biofilms from different experiments. Ideally, one would have carried out metabolomics, mass-spec imagining and RNA-seq transcriptomics on one biofilm, with multiple replicates. This would allow for direct links and comparisons to be made between the transcriptomic message and the metabolite composition within the biofilm due to nitrogen fixation and biofilm maturation. This approach has been has been used in other multi-omic experiments involving bacteria [230,231] including those found in the rhizosphere [39,232]. While the combination of these analyses from separate samples yields important insights, the changes and fluxes captured using multi-omics on one sample would provide the most useful information for future biotechnological applications [30]. Additionally, *nif* protein stoichiometry has been well characterised in batch culture [39,69]. Future work should include proteomic analysis of biofilms to investigate whether the *nif* protein levels remain constant between the two growth states.

As discussed with the batch culture experiments in Chapter 5, it's important to understand how if and how intracellular metabolic changes result in metabolite exchanges with the cells environment. While the biofilms adhesion to a surface, e.g. plant roots, results in easier exchange of metabolites between the environment and bacteria it is imperative that we are able to track these changes. Biofilms this is not quite as simple as with batch culture where one can easily analyse the supernatant. However, it is

possible having been done in water biofilms models, tracking the breakdown of medicinal compounds [233]. In the context of these experiments, one would hope to understand if the increases in intracellular abundance of metabolites such as 5-aminolevulinate and N-acetylglutamate result in increased abundance in the environment and more importantly in plant roots.

Lastly, biofilms present perhaps the best means of plant associative bacteria colonising plant roots and forming mutualistic relationships [211,212], with a closely related species of *K. oxytoca, Klebsiella SGM 81,* shown to colonise plant roots [197]. Therefore, biofilm-plant root interactions present a useful model to study the changes in cell metabolism resulting from nitrogen stress, fixation and biofilm formation in concert with the metabolite exchange between plant roots and bacteria. This approach would allow investigation into whether the insights gained through study of the biofilm in isolation translate to 'real world scenarios'. Given the mechanical lysis involved in the sampling techniques used throughout this thesis, these methods would need minimal adaptation to be used for whole plant root/biofilm samples; given that plant matter can require more robust methods of extraction compared to bacteria [234]. Thus, future work should apply the 'omic' technologies used in this chapter to *K. oxytoca* biofilms in mutualism with plant roots.

6.5 Conclusion

The work carried out in this chapter is among the first to use a multi-omic approach to investigate nitrogen fixation in biofilms. The insights gained show that in *K. oxytoca* nitrogen fixation appears to occur in the centre of biofilms, where the cells are protected from oxygen, and that the metabolic profile of these biofilms changes over time. Some of these changes can be attributed to biofilm maturation but also there are changes caused by nitrogen stress and fixation that change over time, as the metabolic profiles of fixing and non-fixing biofilms remain distinct over time. The biofilms capacity for nitrogen fixation increases overtime, which suggests that fixed nitrogen is sufficient to support both cell growth and the production of more nitrogenase enzyme. Finally, plant growth promoting metabolites that were found to increase in abundance with fixation in batch culture, also increased in abundance with fixation in biofilms. This gives hope that these metabolites could also be exchanged with plant roots along with fixed nitrogen, falling into a class of metabolites that may either support growth or help facilitate the mutualism.

Chapter 7 Concluding remarks and future perspectives

The work in thesis was carried out to investigate the global metabolic changes associated with nitrogen fixation in the free living, plant-associative, rhizobacteria *Klebsiella oxytoca*, to gain insights that could be used in addressing the nitrogen crisis. These insights can enable strategies to manipulate biological fixation, to the benefit of agriculture initially in laboratory settings, and later in small scale field work too. My approach was to characterise different nitrogen fixing states of a model bacterium to help build a foundation for future work. The work carried out in Chapter 3 tests and builds the tools that form the primary form of analysis for the rest of the work in this thesis. As with all metabolomic experiments sampling, both quenching of metabolism and extraction of metabolites, is a fundamental part of the experimental process. After comparison of several sampling methods one, the methanol/chloroform/water sampling, was found to be most suited to the experimental protocol involved with nitrogen fixation and *K. oxytoca.*

Post sampling there are several metabolomics approaches that can be implemented, which fall broadly into two camps; untargeted and targeted metabolomics. In Chapter 3, work was carried out that contributed to the development of a targeted metabolomic assay using UPLC-MS which combined two chromatography methods. Combined these cover a range of metabolite classes including almost 300 metabolites but do omit some biologically important groups such as the thiols, which future work should address. Thiols could for example be important in the redox state of cells, as linked here to the conditions needed for nitrogenase gene expression and nitrogenase activity in dinitrogen reduction for ammonia formation.

The sampling and analytical techniques from Chapter 3 were used with great success to show clear differences in the metabolic profiles resulting from nitrogen fixation in WT *K. oxytoca* in batch culture and biofilm growth (Chapters 5 and 6). These differences were due to many metabolites but the abundance of the amino acids in particular was found to differ between the non-fixing and fixing cells as well as across time. Additionally, several plant growth promoting metabolites were found to increase overtime with nitrogen fixation capacity in both growth phenotypes, suggesting the production of these metabolites could be linked to fixation. This is a clear benefit of implementing a targeted approach with known metabolites, compared to an untargeted approach. Biologically interesting metabolites can tracked through time and their potential function investigated. Unfortunately, these UPLC-MS methods could not be used to investigate changes in the environment brought about by fixation in this thesis primarily due to ion-suppression caused by salts in the media.

Future work should address this by developing low salt medium that can facilitate nitrogen fixation or by removing the salts from the samples prior to analysis.

However, global profiling of the supernatant proved possible using NMR in Chapter 4. While the results of this analysis did not reveal clear overall changes in the supernatant resulting from nitrogen fixation it did reveal the accumulation of Butan-2,3-diol occurred in in fixing cells. While this metabolite appears to be produced by *K. oxytoca* generally during growth in nitrogen replete conditions, that it is still produced during fixation provides proof of plant growth promoting metabolites being produced and released by fixing cells. Butan-2,3-diol has plant growth promoting properties by providing drought resistance to plants and is one of a selection of plant growth promoting metabolites discussed in this thesis. However, the only one known to be released into the environment.

Some of the most promising results stem from the work carried out on *K. oxytoca biofilms*. The work here provides clear evidence that *K. oxytoca* biofilms are able to fix nitrogen in air and elucidated some of the structure and mechanisms that facilitate fixation. Nitrogen fixation was shown to occur in the centre of the biofilms leading to the proposal that an outer non-fixing layer is protecting the oxygen sensitive nitrogenase deeper within the biofilm. Although, the full regulatory mechanism is not known it would appear that the cells in the outer layer remove oxygen from the air sufficiently to induce nitrogen fixation while allowing dinitrogen gas to reach the fixing cells. The metabolite profiles of these layers were shown to change overtime, which agrees with whole biofilm metabolic profiling that also showed age related differences. The potential implications of these findings establishing that a highly oxygen sensitive process can be established in air through using a particular mode of bacterial growth will be discussed below.

7.1 Challenges and future perspectives of metabolomic approaches to bacteria model systems

As discussed earlier in Chapter 1.5.3 there are pros and cons to both targeted and untargeted metabolomics. Indeed, one of the pros of targeted analysis is illustrated in the results of this thesis, in that known metabolites can detected in samples and tracked through time. Allowing for any changes in metabolite abundance due to metabolic changes to be measured and investigated. Immediate future work concerning nitrogen fixation should include using methods like those used in this thesis to investigate whether the changes observed *K. oxytoca* are seen in other nitrogen fixing species of interest such as *Azotobacter vinelandii* and *Psuedomonas stutzeri.* Furthermore, all three species have been shown to form biofilms, with only *K. oxytoca* and *P. stutzeri* are known to fix nitrogen in this

growth phenotype [199,235]. Detailed investigation using targeted metabolomic methods would allow for changes in metabolic phenotype common to all three species to be uncovered, and provide potentially novel mechanisms of adaptation to biofilm dependent fixation for biotechnological approaches to explore.

The LC-MS methods developed as part of Chapter 3 are a useful tool, and as shown in this thesis can provide insights into metabolic changes. By design the metabolites chosen to include cover a broad range of metabolite classes and subsequent metabolic pathways. This means that there are cases where not all the metabolites in a given pathway of interest may be measured, resulting in the gaps within the pathway and changes in metabolite being missed. This was done intentionally, with the idea that this broad approach could be used instead of untargeted methods, while still allowing an overall impression of metabolic differences between treatments to be understood. Therefore, in an ideal situation a 'two tier' approach would be adopted whereby the overall changes in metabolic profiles were detected using approaches like those in Chapter 3. Follow up analysis would then be carried out using targeted methods including many, hopefully all, of the metabolites from a pathway of interest. For example, in the context of the results of this thesis, the initial assessment using a broad targeted method has found that 5-aminolevulinate increases in abundance in nitrogen fixing cells and is of interest because of its potential plant growth promoting properties. To fully understand this in the context of the cell and the environment a second experiment where the only the metabolites involved in this pathway are measured both intracellularly and, in the environment (in this context the metabolites would be glutamate, glutamyl-tRNA, glutamine-l-semialdehyde, and 5-aminolevulinate). The caveat here is that metabolites in a given pathway may have sufficiently different chemical properties that prohibit them being measured in a single targeted method. Additionally, multiple analytical methods may need to be employed to capture information from all the metabolites, e.g. LC-MS, GC-MS, NMR, which would make the follow up assay a costly endeavour. However, from a systems perspective the 'two tiered' approach would allow for maximum information to be gained from a system and therefore provide the best possible chance of a biotechnical solution to be successful.

However, untargeted approaches should not be completely dismissed with regards to bacterial global profiling, including the soil and rhizobacteria specifically. In fact, untargeted metabolomics approaches, such as UPLC-MS, are increasingly being used to phenotype a range disturbed systems ranging from cancer to waste water [236,237], which are known to be impacted by bacteria metabolism. Therefore, it could be of use to use untargeted metabolomics to investigate changes in the soil brought on by both the bacteria within in and anthropogenic perturbation. On this note, a recent study by

Withers *et al.,* used untargeted GC-MS to show the metabolite composition of five different soils differed significantly and correlated with various environmental factors such as pH, moisture and salinity [238]. Their work also shows that a metabolomic approach provided better discriminatory power compared to traditional soil quality indicators such as pH, moisture, and nutrient availability. Thus, this study shows the power of untargeted metabolomics in its applicability to gain an unbiased snapshot of what is occurring in a system, how that snapshot relates to other factors and to tell one snapshot from another [238]. Were one to collect untargeted metabolomic data from enough different soil types over a sufficiently long period of time, machine learning approaches could be employed to look for associations between metabolome composition and undesired outcomes. For example, farmers could use metabolome data to better inform crop cycling or targeted nutrient supplementation to the soil. Given the profound anthropogenic changes being caused in the soil, and the broader environmental damage that these in turn cause, accurate soil measurements could be used to improve farming practices and soil quality. Ultimately, this approach could be part of a suite of scientifically informed methods to improve agricultural practices and reduce their environmental impact.

7.2 Biofilms – a biotechnological solution to Nitrogen Crisis?

Biofilms such as those formed by *K. oxytoca* and *P. stutzeri* in air have the potential to be incredibly useful in agricultural practices due to their ability to fix nitrogen in air and reported beneficial effects on plant growth [40,211,212]. Therefore, a detailed and extensive understanding of biofilm formation and nitrogen fixation could allow for the plant growth promoting benefits to be maximised using biotechnological approaches. The results of Chapter 6, show clear metabolic differences between fixing and non-fixing biofilms, including between fixing and non-fixing biofilms both under nitrogen stress. Therefore, there are profound metabolic changes resulting from nitrogen fixation in *K. oxytoca* biofilms which, while further research is needed, provide solid evidence of altered pathways that could be engineered for the benefit of the plant. A simple example could include biofilm root colonisation of a Δ*nifL* mutant, shown to increase ammonium extraction in batch culture of *A. vinelandii* [60,63]. However, as illustrated by the results comparing the Δ*amtB* and WT in Chapter 6, increased fixation by mutants in batch culture does not always carry over to long term differences in biofilm. As such, a potentially fruitful line of future work could be fully elucidating the differences in global gene expression, nitrogen regulation and AmtB and GS activities between batch culture and biofilms for rhizosphere bacteria like *K. oxytoca*. Looking at the post translational modification of GS, GlnK and PII is of interest here in reporting nitrogen status. One further area of interest is to map the flow of glutamine into metabolism when glutamine is derived from fixation, but cells are in different growth modes. Such flux analyses may reveal bottlenecks in macromolecular synthesis or places where amino acids might be potential

nitrogen sources for plant growth. As discussed below, advances in single cell analyses may well provide insights into the cell specialisations in biofilms, revealing information lost in the ensemble averaging many classical analytical methods rely on.

So far, only single species biofilms have been considered and discussed. However, mixed-species biofilms could present another avenue towards bio-fertilisers and plant growth promotion. Reported to be the dominant form in nature, mixed-species biofilms can be both cooperative and antagonistic [209,239]. In cooperative relationships the metabolic capabilities of the respective species are shared within the biofilm and to the environment. This can include increased resistance to host defences, such as toxic root exudates and facilitate out-competing other species [239]. An example of mixed species biofilms that has been shown to have plant growth promoting effects is fungal-bacterial biofilms [208]. This biofilm composition, which included azotobacter species was found to increase the combined dry weight of the shoots, roots and nodules compared to inorganic fertilisers [208]. However, limited 'omic' data has been collected with regards to mixed-species biofilms and how their respective metabolic phenotypes interact. While whole biofilm omic approaches could be used to understand the metabolic exchanges occurring within a biofilm and with its environment, approaches at the level of the cell would provide the most insights. Thus, the advancement and application of single cell transcriptomics and metabolomics to mixed-biofilms would be the most powerful approaches in future work given they are inherently heterogenous. Single cell approaches would allow more detailed insight into the specific factors that mediate nitrogen fixation and other plant growth promoting effects in these biofilms. Single cell transcriptomics is perhaps more developed in terms of it's readiness to be used, having already been used to investigate antimicrobial resistance in *Staphylococcus aureus* [240]. In comparison, single cell metabolomics is well used a phenotyping tool [241], but targeted single cell metabolomics remains difficult and not as accurate as more traditional approaches [242]. As with single species biofilms data concerning altered metabolic states is needed to maximise beneficial phenotypes of the biofilms.

Another consideration is in the application of biofilms to the plants and their roots. Given that the soil contains many different species of bacteria and environmental factors may not always favour your desired species, facilitating successful colonisation of the plant and its roots is non-trivial. An example of successful application of bacteria as a means of pest control was demonstrated by Curtis *et al.,* over a two year period [243]. In this work, two bacterial species were applied to three locations; the soil, the plant crown and the main roots via a system of drip irrigation. The result showed that the bacteria applied in this way worked as well as other commercial options. Other popular methods include seed

coating, seedling priming and soil drench [244]. A recent study comparing these three methods, found gram-specific success, with Gram negative bacteria, like *K. oxytoca*, being shown to have greatest colonisation success with seedling priming and soil drench compared to seed coating [244]. The authors attribute this to the difference in cell wall structure between Gram positive and negative bacteria citing peptidoglycan specifically. Given that their work was not concerned with biofilm, a different approach may need to be considered. As was shown in Chapter 6, some of the metabolites associated with peptidoglycan production were increased in day 6 biofilms. Perhaps, a better understanding the ESP and cell wall structure in biofilms could show that seed coating is the best mechanism of biofilm delivery to plant roots to help establish root colonisation by beneficial bacteria.

7.3 The soil microbiome and metabolic networks

The work carried out in this thesis has focussed on a single species of bacteria found in the rhizosphere. However, many different species of microorganisms can be found coexisting both in the rhizosphere and soil microbiome at large. As discussed throughout this thesis, changes in rhizobacterial metabolism will have an impact on the plant roots. However, depending on their proximity and interaction with the root it is likely that these changes will impact other microorganisms as well [245]. A great example of species interaction in the soil microbiome is would be mixed species biofilms, that occur 'naturally' in the soil [209,212,239]. Microbiome research is currently a popular field of study, particularly in relation to the human gastrointestinal microbiome [246]. Yet the soil microbiome should not be overlooked and is being well researched in relation to improving agricultural practices, crop health and yield, and reducing the environmental impact of farming [247,248].

The work carried out in this thesis contributes to soil microbiome research in a variety of ways. From a methodological point of view, the sampling and analytical methods developed as part of Chapter 3 could both be used, potentially with minor adaptations such as soil solubilisation, to analyse the soil microbiome metabolome. Currently one of the most common approaches to microbiome research is NGS sequencing. Sequencing approaches are employed to investigate bacterial composition and potential microbiome function, but will provide data on microbes in unknown physiological states. For example, shallow shotgun sequencing might reveal that a particular soil microbiome contains the *nif* genes and thus has the capacity to fix nitrogen. However, given the sensitivity of the nitrogenase to oxygen and nitrogen availability it may be that fixation is not occurring. Therefore, metabolomic approaches capture what is going on in the soil at the point of sampling and include both the downstream outputs of that genetic potential and the available resources within a given environment. At this point both targeted and untargeted approaches have their utility. Targeted approaches will give

accurate functional information about what is occurring within the soil and also the abundance of any plant growth promoting metabolites. In comparison, untargeted metabolomics could be used to phenotype the soil and make inferences about soil health and suitability to be farmed. As discussed earlier, this approach was able to differentiate soil quality more clearly than traditional approaches [238], an outcome that will only improve with more data overtime.

While understanding the overall soil microbiome metabolome can give insights into crop growth and soil quality, it is difficult to attribute particular desired effects to a given species. As such, I would argue there remains great utility in understanding the metabolism and metabolic changes that occur within a single species. In doing so, one can consider how to alter metabolism either through engineering or environmental manipulation to produce outcomes that benefit the collective, namely the soil microbiome and the crops. Thus, the results from Chapters 5 and 6 help to understand the role that *K. oxytoca* could play in the soil microbiome. Given that nitrogen is often limiting in agricultural settings, work that aims to understand specific species able to alleviate nitrogen limitation could have a big impact, despite there being other potentially important bacteria in the soil. Additionally, ambitious synthetic approaches that aim to bypass the soil microbiome all together, e.g. inserting *nif* genes into cereal crops [19,42,172], could benefit from metabolic data from single species to maximise impact and efficiency of the *nif* genes in the crop. The nif genes may impose a burden on a non-native host that requires directed adaptation, e.g. of FeS metabolism, accommodating high ATP use, and provision of reductant. Knowing how native hosts accommodate these burdens is important for launching nif genes across phylogenetic barriers.

Single species metabolomic analysis can also be used to maximise the benefits of biofertilizers. Given the potentially thousands of species present in different soil microbiomes, creating effective biofertilizers supplemented with a selected consortia of bacteria is difficult [249]. Understanding the metabolic outputs of bacteria in specific physiological states, over and above their functional potential from NGS, allows for complementary species to paired together. Of course, this is a simplified picture in the sense of once the bacteria are grouped they begin to interact with each other and change the environment around them. However, this initial understanding provides an informed starting point and a basis for effective metabolic models to be developed.

7.4 Concluding remarks

Given the complexity of the rhizosphere and the soil microbiome more broadly, more work is needed to deepen and broaden the knowledge base of bacterial nitrogen metabolism under a range of nitrogen fixing settings to fully realise the array of biotechnological approaches attempting to address the nitrogen crisis. Moreover, until recently our understanding of diazotrophic bacteria had largely been limited to descriptions of the regulatory cascades responsible for *nif* expression and nitrogenase function. Therefore, it is unsurprising that despite the technological success of some synthetic biological approaches, such as refactored *nif* clusters [30,46], they still perform poorly compared to WT species. It is hoped that a comprehensive understanding of cell metabolism under nitrogen fixing conditions, can aid in future endeavours and spark new lines of enquiry. Given the pressing environmental challenges the world now faces, and the role that agricultural practices play in exacerbating climate change, work that addresses this problem head on, such as that in this thesis is imperative and timely.

References:

1. Mogelgaard, K. *Unfinished Business - The Pursuit of Rights And Choices For All.* (UN Publications, S.l., 2019).
2. FAO. *Report of the FAO Regional Meeting on Agricultural Biotechnologies in Sustainable Food Systems and Nutrition in Asia-Pacific, 11-13 September 2017, Kuala Lumpur, Malaysia*. (FAO, Bangkok, Thailand, 2018).
3. Carpenter-Boggs, L., Pikul, J. L., Vigil, M. F. & Riedell, W. E. Soil Nitrogen Mineralization Influenced by Crop Rotation and Nitrogen Fertilization. *Soil Sci. Soc. Am. J.* **64**, 2038–2045 (2000).
4. Zhao, J. *et al.* Does crop rotation yield more in China? A meta-analysis. *Field Crops Res.* **245**, 107659 (2020).
5. Zuber, S. M., Behnke, G. D., Nafziger, E. D. & Villamil, M. B. Carbon and Nitrogen Content of Soil Organic Matter and Microbial Biomass under Long-Term Crop Rotation and Tillage in Illinois, USA. *Agriculture* **8**, 37 (2018).
6. Rockström, J. *et al.* A safe operating space for humanity. *Nature* **461**, 472–475 (2009).
7. Stein, L. Y. & Klotz, M. G. The nitrogen cycle. *Curr. Biol.* **26**, R94–R98 (2016).
8. Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z. & Winiwarter, W. How a century of ammonia synthesis changed the world. *Nat. Geosci.* **1**, 636–639 (2008).
9. Rogers, C. & Oldroyd, G. E. D. Synthetic biology approaches to engineering the nitrogen symbiosis in cereals. *J. Exp. Bot.* **65**, 1939–1946 (2014).
10. Mann, M. E. *et al.* Influence of Anthropogenic Climate Change on Planetary Wave Resonance and Extreme Weather Events. *Sci. Rep.* **7**, 45242 (2017).

11. Houser, M. & Stuart, D. An accelerating treadmill and an overlooked contradiction in industrial agriculture: Climate change and nitrogen fertilizer. *J. Agrar. Change* **20**, 215–237 (2020).
12. Weis, T. The Accelerating Biophysical Contradictions of Industrial Capitalist Agriculture. *J. Agrar. Change* **10**, 315–341 (2010).
13. Clark, B. & Foster, J. B. Ecological Imperialism and the Global Metabolic Rift: Unequal Exchange and the Guano/Nitrates Trade. *Int. J. Comp. Sociol.* **50**, 311–334 (2009).
14. Stokstad, E. Nitrogen crisis threatens Dutch environment—and economy. *Science* vol. 366 (2019).
15. Mueller, N. D. *et al.* Closing yield gaps through nutrient and water management. *Nature* **490**, 254–257 (2012).
16. Pimentel, D. & Pimentel, M. Sustainability of meat-based and plant-based diets and the environment. *Am. J. Clin. Nutr.* **78**, 660S-663S (2003).
17. Breitburg, D. *et al.* Declining oxygen in the global ocean and coastal waters. *Science* **359**, eaam7240 (2018).
18. Chakraborty, S., Tiwari, P. K., Sasmal, S. K., Misra, A. K. & Chattopadhyay, J. Effects of fertilizers used in agricultural fields on algal blooms. *Eur. Phys. J. Spec. Top.* **226**, 2119–2133 (2017).
19. Oldroyd, G. E. & Dixon, R. Biotechnological solutions to the nitrogen problem. *Curr. Opin. Biotechnol.* **26**, 19–24 (2014).
20. Etesami, H. Root nodules of legumes: A suitable ecological niche for isolating non-rhizobial bacteria with biotechnological potential in agriculture. *Curr. Res. Biotechnol.* **4**, 78–86 (2022).

21. Garg, N. & Geetanjali. Symbiotic Nitrogen Fixation in Legume Nodules: Process and Signaling: A Review. in *Sustainable Agriculture* (eds. Lichtfouse, E., Navarrete, M., Debaeke, P., Véronique, S. & Alberola, C.) 519–531 (Springer Netherlands, Dordrecht, 2009).

22. Vitousek, P. M., Menge, D. N. L., Reed, S. C. & Cleveland, C. C. Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20130119–20130119 (2013).

23. Prell, J. & Poole, P. Metabolic changes of rhizobia in legume nodules. *Trends Microbiol.* **14**, 161–168 (2006).

24. Yang, Y., Hu, X.-P. & Ma, B.-G. Construction and simulation of the *Bradyrhizobium diazoefficiens* USDA110 metabolic network: a comparison between free-living and symbiotic states. *Mol. Biosyst.* **13**, 607–620 (2017).

25. Smercina, D. N., Evans, S. E., Friesen, M. L. & Tiemann, L. K. To Fix or Not To Fix: Controls on Free-Living Nitrogen Fixation in the Rhizosphere. *Appl. Environ. Microbiol.* **85** (2019).

26. Dixon, R. & Kahn, D. Genetic regulation of biological nitrogen fixation. *Nat. Rev. Microbiol.* **2**, 621 (2004).

27. Bao, G. *et al.* Genome Sequence of *Klebsiella oxytoca* M5al, a Promising Strain for Nitrogen Fixation and Chemical Production. *Genome Announc.* **1** (2013).

28. Low, Y. M. *et al.* Elucidating the survival and response of carbapenem resistant *Klebsiella pneumoniae* after exposure to imipenem at sub-lethal concentrations. *Pathog. Glob. Health* **112**, 378–386 (2018).

29. Liu, Y. *et al.* Study on Mechanisms of Colonization of Nitrogen-Fixing PGPB, *Klebsiella pneumoniae* NG14 on the Root Surface of Rice and the Formation of Biofilm. *Curr. Microbiol.* **62**, 1113–1122 (2011).

30. Temme, K., Zhao, D. & Voigt, C. A. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proc. Natl. Acad. Sci.* **109**, 7085–7090 (2012).

31. Drulis-Kawa, Z. *et al.* Isolation and characterisation of KP34—a novel ϕKMV-like bacteriophage for *Klebsiella pneumoniae*. *Appl. Microbiol. Biotechnol.* **90**, 1333–1345 (2011).

32. Dieckmann, R. *et al.* Rapid characterisation of *Klebsiella oxytoca* isolates from contaminated liquid hand soap using mass spectrometry, FTIR and Raman spectroscopy. *Faraday Discuss.* **187**, 353–375 (2016).

33. Ma, Y. *et al.* Microbiological characterisation of *Klebsiella pneumoniae* isolates causing bloodstream infections from five tertiary hospitals in Beijing, China. *J. Glob. Antimicrob. Resist.* **12**, 162–166 (2018).

34. Cakmakci, M. L., Evans, H. J. & Seidler, R. J. Characteristics of nitrogen-fixing *Klebsiella oxytoca* isolated from wheat roots. *Plant Soil* **61**, 53–63 (1981).

35. Remya, P. A., Shanthi, M. & Sekar, U. Characterisation of Virulence Genes Associated with Pathogenicity in *Klebsiella pneumoniae*. *Indian J. Med. Microbiol.* **37**, 210–218 (2019).

36. Fujii, T. *et al.* Effect of inoculation with *Klebsiella oxytoca* and *Enterobacter cloacae* on dinitrogen fixation by rice-bacteria associations. *Plant Soil* **103**, 221–226 (1987).

37. Ladha, J. K., Barraquio, W. L. & Watanabe, I. Isolation and identification of nitrogen-fixing *Enterobacter cloacae* and *Klebsiella planticola* associated with rice plants. *Can. J. Microbiol.* **29**, 1301–1308 (1983).

38. Martínez, L., Caballero-Mellado, J., Orozco, J. & Martínez-Romero, E. Diazotrophic bacteria associated with banana (Musa spp.). *Plant Soil* **257**, 35–47 (2003).

39. Waite, C. J. *et al.* Resource Allocation During the Transition to Diazotrophy in *Klebsiella oxytoca*. *Front. Microbiol.* **12**, 2194 (2021).

40. Gang, S., Saraf, M., Waite, C. J., Buck, M. & Schumacher, J. Mutualism between *Klebsiella SGM* 81 and *Dianthus caryophyllus* in modulating root plasticity and rhizospheric bacterial density. *Plant Soil* **424**, 273–288 (2018).

41. Bashir, T. *et al.* Molecular Origins of Transcriptional Heterogeneity in Diazotrophic *Klebsiella oxytoca*. *mSystems* **0** (2022).

42. Curatti, L. & Rubio, L. M. Challenges to develop nitrogen-fixing cereals by direct nif-gene transfer. *Plant Sci.* **225**, 130–137 (2014).

43. Raymond, J., Siefert, J. L., Staples, C. R. & Blankenship, R. E. The Natural History of Nitrogen Fixation. *Mol. Biol. Evol.* **21**, 541–554 (2004).

44. Reed, S. C., Cleveland, C. C. & Townsend, A. R. Functional Ecology of Free-Living Nitrogen Fixation: A Contemporary Perspective. *Annu. Rev. Ecol. Evol. Syst.* **42**, 489–512 (2011).

45. Bellenger, J. P., Xu, Y., Zhang, X., Morel, F. M. M. & Kraepiel, A. M. L. Possible contribution of alternative nitrogenases to nitrogen fixation by asymbiotic N2-fixing bacteria in soils. *Soil Biol. Biochem.* **69**, 413–420 (2014).

46. Smanski, M. J. *et al.* Functional optimization of gene clusters by combinatorial design and assembly. *Nat. Biotechnol.* **32**, 1241–1249 (2014).

47. Reitzer, L. Nitrogen Assimilation and Global Regulation in *Escherichia coli*. *Annu. Rev. Microbiol.* **57**, 155–176 (2003).

48. Reitzer, L. J. *Escherichia Coli and Salmonella: Cellular and Molecular Biology*. (ASM Press, Washington, DC, 1996).

49. Schumacher, J. Native and Synthetic Gene Regulation to Nitrogen Limitation Stress. *Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria* 234–246 (John Wiley & Sons, Ltd, 2016).

50. van Heeswijk, W. C., Westerhoff, H. V. & Boogerd, F. C. Nitrogen Assimilation in *Escherichia coli*: Putting Molecular Data into a Systems Perspective. *Microbiol. Mol. Biol. Rev.* **77**, 628–695 (2013).

51. Coutts, G. Membrane sequestration of the signal transduction protein GlnK by the ammonium transporter AmtB. *EMBO J.* **21**, 536–545 (2002).

52. Yan, D. *et al.* Genetic modification of flavone biosynthesis in rice enhances biofilm formation of soil diazotrophic bacteria and biological nitrogen fixation. *Plant Biotechnol. J.* (2022).

53. Larrea-Álvarez, M. & Purton, S. The Chloroplast of *Chlamydomonas reinhardtii* as a Testbed for Engineering Nitrogen Fixation into Plants. *Int. J. Mol. Sci.* **22**, 8806 (2021).

54. Haskett, T. L. *et al.* Engineered plant control of associative nitrogen fixation. *Proc. Natl. Acad. Sci.* **119** (2022).

55. Wang, B., Barahona, M., Buck, M. & Schumacher, J. Rewiring cell signalling through chimaeric regulatory protein engineering. *Biochem. Soc. Trans.* **41**, 1195–1200 (2013).

56. Hardy, R. W. F., Burns, R. C. & Holsten, R. D. Applications of the acetylene-ethylene assay for measurement of nitrogen fixation. *Soil Biol. Biochem.* **5**, 47–81 (1973).

57. O'Hara, G. W. Nutritional constraints on root nodule bacteria affecting symbiotic nitrogen fixation: a review. *Aust. J. Exp. Agric.* **41**, 417–433 (2001).

58. Dunwell, J. M. Transgenic cereals: Current status and future prospects. *J. Cereal Sci.* **59**, 419–434 (2014).

59. Bueno Batista, M. & Dixon, R. Manipulating nitrogen regulation in diazotrophic bacteria for agronomic benefit. *Biochem. Soc. Trans.* **47**, 603–614 (2019).

60. Barney, B. M., Eberhart, L. J., Ohlert, J. M., Knutson, C. M. & Plunkett, M. H. Gene Deletions Resulting in Increased Nitrogen Release by *Azotobacter vinelandii*: Application of a Novel Nitrogen Biosensor. *Appl. Environ. Microbiol.* **81**, 4316–4328 (2015).

61. Christiansen-Weniger, C. & Van Veen, J. A. NH_4^+-Excreting *Azospirillum brasilense* Mutants Enhance the Nitrogen Supply of a Wheat Host. *Appl. Environ. Microbiol.* **57**, 3006–3012 (1991).

62. Santos, K. F. D. N. *et al.* Wheat colonization by an *Azospirillum brasilense* ammonium-excreting strain reveals upregulation of nitrogenase and superior plant growth promotion. *Plant Soil* **415**, 245–255 (2017).

63. Bali, A. Excretion of Ammonium by a nifL Mutant of *Azotobacter vinelandii* Fixing Nitrogen. *APPL Env. MICROBIOL* **58**, 8 (1992).

64. Poza-Carrión, C., Jiménez-Vicente, E., Navarro-Rodríguez, M., Echavarri-Erasun, C. & Rubio, L. M. Kinetics of nif Gene Expression in a Nitrogen-Fixing Bacterium. *J. Bacteriol.* **196**, 595–603 (2014).

65. Carnahan, J. E., Mortenson, L. E., Mower, H. F. & Castle, J. E. Nitrogen fixation in cell-free extracts of *Clostridium pasteurianum*. *Biochim. Biophys. Acta* **44**, 520–535 (1960).

66. Wang, D., Xu, A., Elmerich, C. & Ma, L. Z. Biofilm formation enables free-living nitrogen-fixing rhizobacteria to fix nitrogen under aerobic conditions. *ISME J.* **11**, 1602–1613 (2017).

67. Branda, S. S., Vik, Å., Friedman, L. & Kolter, R. Biofilms: the matrix revisited. *Trends Microbiol.* **13**, 20–26 (2005).

68. Flemming, H.-C. & Wingender, J. The biofilm matrix. *Nat. Rev. Microbiol.* **8**, 623–633 (2010).

69. Delmotte, N. *et al.* An integrated proteomics and transcriptomics reference data set provides new insights into the *Bradyrhizobium japonicum* bacteroid metabolism in soybean root nodules. *PROTEOMICS* **10**, 1391–1400 (2010).

70. Lardi, M. *et al.* Metabolomic Profiling of *Bradyrhizobium diazoefficiens*-Induced Root Nodules Reveals Both Host Plant-Specific and Developmental Signatures. *Int. J. Mol. Sci.* **17**, 815 (2016).

71. Cesco, S. *et al.* Plant-borne flavonoids released into the rhizosphere: impact on soil bio-activities related to plant nutrition. A review. *Biol. Fertil. Soils* **48**, 123–149 (2012).

72. Brzóska, K., Meczyńska, S. & Kruszewski, M. Iron-sulfur cluster proteins: electron transfer and beyond. *Acta Biochim. Pol.* **53**, 685–691 (2006).

73. Vicente, E. J. & Dean, D. R. Keeping the nitrogen-fixation dream alive. *Proc. Natl. Acad. Sci.* **114**, 3009–3011 (2017).

74. Becana, M., Wienkoop, S. & Matamoros, M. A. Sulfur Transport and Metabolism in Legume Root Nodules. *Front. Plant Sci.* **9**, (2018).

75. Nicholson, J. K., Lindon, J. C. & Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29**, 1181–1189 (1999).

76. Fuchs, S. *et al.* A metabolic signature of long life in *Caenorhabditis elegans*. *BMC Biol.* **8**, 14 (2010).

77. Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G. & Kell, D. B. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.* **22**, 245–252 (2004).

78. Janečková, H. *et al.* Targeted metabolomic analysis of plasma samples for the diagnosis of inherited metabolic disorders. *J. Chromatogr. A* **1226**, 11–17 (2012).

79. Bolten, C. J., Kiefer, P., Letisse, F., Portais, J.-C. & Wittmann, C. Sampling for Metabolome Analysis of Microorganisms. *Anal. Chem.* **79**, 3843–3849 (2007).

80. Raamsdonk, L. M. *et al.* A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* **19**, 45–50 (2001).

81. Aderemi, A. V., Ayeleso, A. O., Oyedapo, O. O. & Mukwevho, E. Metabolomics: A Scoping Review of Its Role as a Tool for Disease Biomarker Discovery in Selected Non-Communicable Diseases. *Metabolites* **11**, 418 (2021).

82. Hamany Djande, C. Y., Pretorius, C., Tugizimana, F., Piater, L. A. & Dubery, I. A. Metabolomics: A Tool for Cultivar Phenotyping and Investigation of Grain Crops. *Agronomy* **10**, 831 (2020).

83. Dunn, W. B. & Ellis, David. I. Metabolomics: Current analytical platforms and methodologies. *TrAC Trends Anal. Chem.* **24**, 285–294 (2005).

84. Lin, C. Y., Wu, H., Tjeerdema, R. S. & Viant, M. R. Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. *Metabolomics* **3**, 55–67 (2007).

85. Mohd Kamal, K. *et al.* Bacterial Metabolomics: Sample Preparation Methods. *Biochem. Res. Int.* **2022**, e9186536 (2022).

86. Zakhartsev, M., Vielhauer, O., Horn, T., Yang, X. & Reuss, M. Fast sampling for quantitative microbial metabolomics: new aspects on cold methanol quenching: metabolite co-precipitation. *Metabolomics* **11**, 286–301 (2015).

87. Villas-Bôas, S. G. & Bruheim, P. Cold glycerol–saline: The promising quenching solution for accurate intracellular metabolite analysis of microbial cells. *Anal. Biochem.* **370**, 87–97 (2007).

88. Kell, D. B. *et al.* Metabolic footprinting and systems biology: the medium is the message. *Nat. Rev. Microbiol.* **3**, 557–565 (2005).

89. Barrera-Galicia, G. C. *et al.* Metabolic Footprints of *Burkholderia* Sensu Lato Rhizosphere Bacteria Active against Maize *Fusarium* Pathogens. *Microorganisms* **9**, 2061 (2021).

90. Want, E. J., Cravatt, B. F. & Siuzdak, G. The Expanding Role of Mass Spectrometry in Metabolite Profiling and Characterization. *ChemBioChem* **6**, 1941–1951 (2005).

91. Clarke, W. Chapter 1 - Mass spectrometry in the clinical laboratory: determining the need and avoiding pitfalls. in *Mass Spectrometry for the Clinical Laboratory* (eds. Nair, H. & Clarke, W.) 1–15 (Academic Press, San Diego, 2017).

92. Cole, R. B. Some tenets pertaining to electrospray ionization mass spectrometry. *J. Mass Spectrom.* **35**, 763–772 (2000).

93. Schwaiger-Haber, M. *et al.* A Workflow to Perform Targeted Metabolomics at the Untargeted Scale on a Triple Quadrupole Mass Spectrometer. *ACS Meas. Sci. Au* **1**, 35–45 (2021).

94. Bylesjö, M. *et al.* OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* **20**, 341–351 (2006).

95. Yotsukura, S. & Mamitsuka, H. Evaluation of serum-based cancer biomarkers: A brief review from a clinical and computational viewpoint. *Crit. Rev. Oncol. Hematol.* **93**, 103–115 (2015).

96. Zhou, H. *et al.* Serum untargeted lipidomics by UHPLC-ESI-HRMS aids the biomarker discovery of colorectal adenoma. *BMC Cancer* **22**, 314 (2022).

97. Da Cunha, P. A. *et al.* Metabolomic Analysis of Plasma from Breast Cancer Patients Using Ultra-High-Performance Liquid Chromatography Coupled with Mass Spectrometry: An Untargeted Study. *Metabolites* **12**, 447 (2022).

98. Hung, W.-L. & Wang, Y. A Targeted Mass Spectrometry-Based Metabolomics Approach toward the Understanding of Host Responses to Huanglongbing Disease. *J. Agric. Food Chem.* **66**, 10651–10661 (2018).

99. Lardi, M. *et al.* Metabolomics and Transcriptomics Identify Multiple Downstream Targets of *Paraburkholderia phymatum* σ54 During Symbiosis with *Phaseolus vulgaris.* *Int. J. Mol. Sci.* **19**, 1049 (2018).

100. Meena, K. K. *et al.* Abiotic Stress Responses and Microbe-Mediated Mitigation in Plants: The Omics Strategies. *Front. Plant Sci.* **8**, (2017).

101. Du, J., Zhou, J., Xue, J., Song, H. & Yuan, Y. Metabolomic profiling elucidates community dynamics of the *Ketogulonicigenium vulgare–Bacillus megaterium* consortium. *Metabolomics* **8**, 960–973 (2012).

102. Rasmussen, S., Parsons, A. J. & Newman, J. A. Metabolomics analysis of the *Lolium perenne–Neotyphodium lolii* symbiosis: more than just alkaloids? *Phytochem. Rev.* **8**, 535–550 (2009).

103. Sharma, V. *et al.* Metabolomics Intervention Towards Better Understanding of Plant Traits. *Cells* **10**, 346 (2021).

104. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *PNAS* **97**, 6640–6645 (2000).

105. Brown, S. D. & Jun, S. Complete Genome Sequence of *Escherichia coli* NCM3722. *Genome Announc.* **3**, e00879-15 (2015).

106. Zumaquero, A., Macho, A. P., Rufián, J. S. & Beuzón, C. R. Analysis of the role of the type III effector inventory of *Pseudomonas syringae* pv. phaseolicola 1448a in interaction with the plant. *J. Bacteriol.* **192**, 4474–4488 (2010).

107. Moure, V. R. *et al.* The ammonium transporter AmtB and the PII signal transduction protein GlnZ are required to inhibit DraG in *Azospirillum brasilense*. *FEBS J.* **286**, 1214–1229 (2019).

108. Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2**, 2692–2703 (2007).

109. CSM. CSM Matlab Docs. *CSM Matlab Docs.* https://csmsoftware.github.io/docs/impacts/index.html.

110. Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in ^{1}H NMR Metabonomics. *Anal. Chem.* **78**, 4281–4290 (2006).

111. Lopatkin, A. J. *et al.* Bacterial metabolic state more accurately predicts antibiotic lethality than growth rate. *Nat. Microbiol.* **4**, 2109–2117 (2019).

112. Wu, Y. & Li, L. Sample normalization methods in quantitative metabolomics. *J. Chromatogr. A* **1430**, 80–95 (2016).

113. Gika, H. G., Theodoridis, G. A., Wingate, J. E. & Wilson, I. D. Within-Day Reproducibility of an HPLC−MS-Based Method for Metabonomic Analysis: Application to Human Urine. *J. Proteome Res.* **6**, 3291–3303 (2007).

114. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

115. Wold, S., Esbensen, K. & Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **2**, 37–52 (1987).

116. Kohl, P., Crampin, E. J., Quinn, T. A. & Noble, D. Systems Biology: An Approach. *Clin. Pharmacol. Ther.* **88**, 25–33 (2010).

117. Douma, R. D. *et al.* Intracellular metabolite determination in the presence of extracellular abundance: Application to the penicillin biosynthesis pathway in Penicillium chrysogenum. *Biotechnol. Bioeng.* **107**, 105–115 (2010).

118. Ye, D. *et al.* Improved Sample Preparation for Untargeted Metabolomics Profiling of *Escherichia coli*. *Microbiol. Spectr.* **9**, e00625-21 (2021).

119. Rabinowitz, J. D. & Kimball, E. Acidic Acetonitrile for Cellular Metabolome Extraction from *Escherichia coli*. *Anal. Chem.* **79**, 6167–6173 (2007).

120. Villas-Bôas, S. G. Sampling and Sample Preparation. in *Metabolome Analysis* 39–82 (John Wiley & Sons, Ltd, 2006).

121. Schädel, F., David, F. & Franco-Lara, E. Evaluation of cell damage caused by cold sampling and quenching for metabolome analysis. *Appl. Microbiol. Biotechnol.* **92**, 1261–1274 (2011).

122. Pinu, F. R., Villas-Boas, S. G. & Aggio, R. Analysis of Intracellular Metabolites from Microorganisms: Quenching and Extraction Protocols. *Metabolites* **7**, 53 (2017).

123. van der Laan, T. *et al.* High-Throughput Fractionation Coupled to Mass Spectrometry for Improved Quantitation in Metabolomics. *Anal. Chem.* **92**, 14330–14338 (2020).

124. Wang, X. *et al.* Evaluation and optimization of sample preparation methods for metabolic profiling analysis of *Escherichia coli. ELECTROPHORESIS* **36**, 2140–2147 (2015).

125. Park, C., Yun, S., Lee, S. Y., Park, K. & Lee, J. Metabolic Profiling of *Klebsiella oxytoca*: Evaluation of Methods for Extraction of Intracellular Metabolites Using UPLC/Q-TOF-MS. *Appl. Biochem. Biotechnol.* **167**, 425–438 (2012).

126. Michopoulos, F. *et al.* Targeted profiling of polar intracellular metabolites using ion-pair-high performance liquid chromatography and -ultra high-performance liquid chromatography coupled to tandem mass spectrometry: Applications to serum, urine and tissue extracts. *J. Chromatogr. A* **1349**, 60–68 (2014).

127. Gray, N. *et al.* High-Speed Quantitative UPLC-MS Analysis of Multiple Amines in Human Plasma and Serum via Precolumn Derivatization with 6-Aminoquinolyl-N-hydroxysuccinimidyl Carbamate: Application to Acetaminophen-Induced Liver Failure. *Anal. Chem.* **89**, 2478–2487 (2017).

128. Smart, K. F., Aggio, R. B. M., Van Houtte, J. R. & Villas-Bôas, S. G. Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography–mass spectrometry. *Nat. Protoc.* **5**, 1709–1729 (2010).

129. Sagi-Kiss, V. *et al.* Ion-Pairing Chromatography and Amine Derivatization Provide Complementary Approaches for the Targeted LC-MS Analysis of the Polar Metabolome. *J. Proteome Res.* **21**, 1428–1437 (2022).

130. Belle, J. E. L., Harris, N. G., Williams, S. R. & Bhakoo, K. K. A comparison of cell and tissue extraction techniques using high-resolution 1H-NMR spectroscopy. *NMR Biomed.* **15**, 37–44 (2002).

131. Boughton, B. A. *et al.* Comprehensive Profiling and Quantitation of Amine Group Containing Metabolites. *Anal. Chem.* **83**, 7523–7530 (2011).

132. Behrends, V., Ebbels, T. M. D., Williams, H. D. & Bundy, J. G. Time-Resolved Metabolic Footprinting for Nonlinear Modeling of Bacterial Substrate Utilization. *Appl. Environ. Microbiol.* **75**, 2453–2463 (2009).

133. Spura, J. *et al.* A method for enzyme quenching in microbial metabolome analysis successfully applied to gram-positive and gram-negative bacteria and yeast. *Anal. Biochem.* **394**, 192–201 (2009).

134. Behrends, V., Williams, H. D. & Bundy, J. G. Metabolic Footprinting: Extracellular Metabolomic Analysis. in *Pseudomonas Methods and Protocols* (eds. Filloux, A. & Ramos, J.-L.) 281–292 (Springer, New York, NY, 2014). doi:10.1007/978-1-4939-0473-0_23.

135. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **5**, 593–599 (2009).

136. Ortmayr, K., Schwaiger, M., Hann, S. & Koellensperger, G. An integrated metabolomics workflow for the quantification of sulfur pathway intermediates employing thiol protection with N -ethyl maleimide and hydrophilic interaction liquid chromatography tandem mass spectrometry. *Analyst* **140**, 7687–7695 (2015).

137. Ritz, D. & Beckwith, J. Roles of Thiol-Redox Pathways in Bacteria. *Annu. Rev. Microbiol.* **55**, 21–48 (2001).

138. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007).

139. Edwards, R. A. & Schifferli, D. M. Differential regulation of fasA and fasH expression of *Escherichia coli* 987P fimbriae by environmental cues. *Mol. Microbiol.* **25**, 797–809 (1997).

140. Merrick, M. J. & Edwards, R. A. Nitrogen control in bacteria. *Microbiol. Rev.* **59**, 604–622 (1995).

141. Fulks, R. M. & Stadtman, E. R. Regulation of glutamine synthetase, aspartokinase, and total protein turnover in *Klebsiella aerogenes*. *Biochim. Biophys. Acta BBA - Gen. Subj.* **843**, 214–229 (1985).

142. Pezzatti, J. *et al.* Choosing an Optimal Sample Preparation in Caulobacter crescentus for Untargeted Metabolomics Approaches. *Metabolites* **9**, 193 (2019).

143. Taskova, R. M., Zorn, H., Krings, U., Bouws, H. & Berger, R. G. A Comparison of Cell Wall Disruption Techniques for the Isolation of Intracellular Metabolites from Pleurotus and Lepista sp. *Z. Für Naturforschung C* **61**, 347–350 (2006).

144. Chisti, Y. & Moo-Young, M. Disruption of microbial cells for intracellular products. *Enzyme Microb. Technol.* **8**, 194–204 (1986).

145. Harrison, S. T. L. Bacterial cell disruption: A key unit operation in the recovery of intracellular products. *Biotechnol. Adv.* **9**, 217–240 (1991).

146. Shehadul Islam, M., Aryasomayajula, A. & Selvaganapathy, P. R. A Review on Macroscale and Microscale Cell Lysis Methods. *Micromachines* **8**, 83 (2017).

147. Thorfinnsdottir, L. B., García-Calvo, L., Bø, G. H., Bruheim, P. & Røst, L. M. Optimized Fast Filtration-Based Sampling and Extraction Enables Precise and Absolute

Quantification of the *Escherichia coli* Central Carbon Metabolome. *Metabolites* **13**, 150 (2023).

148. Compant, S., Clément, C. & Sessitsch, A. Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. *Soil Biol. Biochem.* **42**, 669–678 (2010).
149. Ryu, M.-H. *et al.* Control of nitrogen fixation in bacteria that associate with cereals. *Nat. Microbiol.* **5**, 314–330 (2020).
150. Hernández, G. *et al.* Global Changes in the Transcript and Metabolic Profiles during Symbiotic Nitrogen Fixation in Phosphorus-Stressed Common Bean Plants. *Plant Physiol.* **151**, 1221–1238 (2009).
151. Magnúsdóttir, S. & Thiele, I. Modeling metabolism of the human gut microbiome. *Curr. Opin. Biotechnol.* **51**, 90–96 (2018).
152. Bergersen, F. J. & Turner, G. L. Bacteroids from soybean root nodules: respiration and N2-fixation in flow-chamber reactions with oxyleghaemoglobin. *Proc. R. Soc. Lond. B Biol. Sci.* **238**, 295–320 (1990).
153. Plunkett, M. H., Knutson, C. M. & Barney, B. M. Key factors affecting ammonium production by an *Azotobacter vinelandii* strain deregulated for biological nitrogen fixation. *Microb. Cell Factories* **19**, 107 (2020).
154. Low, Y. M. *et al.* Genotypic and metabolic approaches towards the segregation of *Klebsiella pneumoniae* strains producing different antibiotic resistant enzymes. *Metabolomics* **13**, 65 (2017).
155. Huang, A. C. *et al.* A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. *Science* **364** (2019).

156. Beatty, P. H. & Good, A. G. Future Prospects for Cereals That Fix Nitrogen. *Science* **333**, 416–417 (2011).

157. Biebl, H., Zeng, A.-P., Menzel, K. & Deckwer, W.-D. Fermentation of glycerol to 1,3-propanediol and 2,3-butanediol by *Klebsiella pneumoniae*. *Appl. Microbiol. Biotechnol.* **50**, 24–29 (1998).

158. Park, K. Y., Seo, S. Y., Oh, B.-R., Seo, J.-W. & Kim, Y. J. 2,3-butanediol Induces Systemic Acquired Resistance in the Plant Immune Response. *J. Plant Biol.* **61**, 424–434 (2018).

159. Rubio, L. M. & Ludden, P. W. Biosynthesis of the Iron-Molybdenum Cofactor of Nitrogenase. *Annu. Rev. Microbiol.* **62**, 93–111 (2008).

160. Sun, J.-L., Zhang, S.-K., Chen, J.-Y. & Han, B.-Z. Metabolic profiling of *Staphylococcus aureus* cultivated under aerobic and anaerobic conditions with 1H NMR-based nontargeted analysis. *Can. J. Microbiol.* **58**, 709–718 (2012).

161. Yan, Z. *et al.* 1H NMR-based metabolomics approach for understanding the fermentation behaviour of *Bacillus licheniformis*. *J. Inst. Brew.* **121**, 425–431 (2015).

162. Presser, K. A., Ratkowsky, D. A. & Ross, T. Modelling the growth rate of *Escherichia coli* as a function of pH and lactic acid concentration. *Appl. Environ. Microbiol.* **63**, 2355–2360 (1997).

163. Syu, M.-J. Biological production of 2,3-butanediol. *Appl. Microbiol. Biotechnol.* **55**, 10–18 (2001).

164. Wu, L. *et al.* Acetoin and 2,3-butanediol from *Bacillus amyloliquefaciens* induce stomatal closure in Arabidopsis thaliana and Nicotiana benthamiana. *J. Exp. Bot.* **69**, 5625–5635 (2018).

165. Yi, H.-S. *et al.* Impact of a Bacterial Volatile 2,3-Butanediol on *Bacillus subtilis* Rhizosphere Robustness. *Front. Microbiol.* **7**, (2016).

166. Castro, J. *et al.* Comparative transcriptomic analysis of *Gardnerella vaginalis* biofilms vs. planktonic cultures using RNA-seq. *Npj Biofilms Microbiomes* **3**, 1–7 (2017).

167. Mattow, J. *et al.* Comparative proteome analysis of culture supernatant proteins from virulent *Mycobacterium tuberculosis* H37Rv and attenuated *M. bovis* BCG Copenhagen. *ELECTROPHORESIS* **24**, 3405–3420 (2003).

168. Dos Santos, P. C. Molecular Biology and Genetic Engineering in Nitrogen Fixation. in *Nitrogen Fixation: Methods and Protocols* (ed. Ribbe, M. W.) 81–92 (Humana Press, Totowa, NJ, 2011).

169. Agtuca, B. J. *et al.* Metabolomic profiling of wild-type and mutant soybean root nodules using laser-ablation electrospray ionization mass spectrometry reveals altered metabolism. *Plant J.* **103**, 1937–1958 (2020).

170. Abdelrahman, M. *et al.* Metabolomics and Transcriptomics in Legumes Under Phosphate Deficiency in Relation to Nitrogen Fixation by Root Nodules. *Front. Plant Sci.* **9**, (2018).

171. Roberts, L. D., Souza, A. L., Gerszten, R. E. & Clish, C. B. Targeted Metabolomics. *Curr. Protoc. Mol. Biol.* **98**, 30.2.1-30.2.24 (2012).

172. Li, Q. & Chen, S. Transfer of Nitrogen Fixation (nif) Genes to Non-diazotrophic Hosts. *ChemBioChem* **21**, 1717–1722 (2020).

173. Fisher, S. H. & Sonenshein, A. L. Control of carbon and nitrogen metabolism in *Bacillus subtilis*. *Annu. Rev. Microbiol.* **45**, 107–135 (1991).

174. Chong, J. & Xia, J. Using MetaboAnalyst 4.0 for Metabolomics Data Analysis, Interpretation, and Integration with Other Omics Data. in *Computational Methods and Data Analysis for Metabolomics* (ed. Li, S.) 337–360 (Springer US, New York, NY, 2020). doi:10.1007/978-1-0716-0239-3_17.

175. Pang, Z. *et al.* MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, W388–W396 (2021).

176. Booth, N. J., Smith, P. M. C., Ramesh, S. A. & Day, D. A. Malate Transport and Metabolism in Nitrogen-Fixing Legume Nodules. *Molecules* **26**, 6876 (2021).

177. Philip-Hollingsworth, S., Hollingsworth, R. I. & Dazzo, F. B. N-Acetylglutamic acid: an extracellular nod signal of *Rhizobium trifolii* ANU843 that induces root hair branching and nodule-like primordia in white clover roots. *J. Biol. Chem.* **266**, 16854–16858 (1991).

178. Schumacher, J. *et al.* Nitrogen and Carbon Status Are Integrated at the Transcriptional Level by the Nitrogen Regulator NtrC In Vivo. *mBio* **4**, e00881-13 (2013).

179. Gosztolai, A. *et al.* GlnK Facilitates the Dynamic Regulation of Bacterial Nitrogen Assimilation. *Biophys. J.* **112**, 2219–2230 (2017).

180. Schreiber, F. *et al.* Phenotypic heterogeneity driven by nutrient limitation promotes growth in fluctuating environments. *Nat. Microbiol.* **1**, 1–7 (2016).

181. Zimmer, D. P. *et al.* Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: Scavenging as a defence against nitrogen limitation. *Proc. Natl. Acad. Sci.* **97**, 14674–14679 (2000).

182. Barney, B. M. *et al.* Transcriptional Analysis of an Ammonium-Excreting Strain of *Azotobacter vinelandii* Deregulated for Nitrogen Fixation. *Appl. Environ. Microbiol.* **83**, (2017).

183. Schmitz, R. A. Internal Glutamine and Glutamate Pools in *Klebsiella pneumoniae* Grown Under Different Conditions of Nitrogen Availability. *Curr. Microbiol.* **41**, 357–362 (2000).

184. Martien, J. I. *et al.* Metabolic Remodelling during Nitrogen Fixation in *Zymomonas mobilis*. *mSystems* **6**, e00987-21 (2021).

185. Brauer, M. J. *et al.* Conservation of the metabolomic response to starvation across two divergent microbes. *Proc. Natl. Acad. Sci.* **103**, 19302–19307 (2006).

186. Ohashi, Y. *et al.* Depiction of metabolome changes in histidine -starved *Escherichia coli* by CE-TOFMS. *Mol. Biosyst.* **4**, 135–147 (2008).

187. Baek, J.-O. *et al.* Expression and characterization of a second L-amino acid deaminase isolated from Proteus mirabilis in *Escherichia coli*. *J. Basic Microbiol.* **51**, 129–135 (2011).

188. Montero-Morán, G. M., Lara-González, S., Álvarez-Añorve, L. I., Plumbridge, J. A. & Calcagno, M. L. On the Multiple Functional Roles of the Active Site Histidine in Catalysis and Allosteric Regulation of *Escherichia coli* Glucosamine 6-Phosphate Deaminase. *Biochemistry* **40**, 10187–10196 (2001).

189. Larrainzar, E. *et al.* Carbon Metabolism and Bacteroid Functioning Are Involved in the Regulation of Nitrogen Fixation in Medicago truncatula Under Drought and Recovery. *Mol. Plant-Microbe Interactions* **22**, 1565–1576 (2009).

190. Lin, M. *et al.* Regulatory Coupling of Nitrogen and Carbon Metabolism in Nitrogen-Fixing *Pseudomonas stutzeri* A1501. in *Biological Nitrogen Fixation* 109–120 (John Wiley & Sons, Ltd, 2015).

191. Cho, J.-H. *et al.* Fermentation and evaluation of *Klebsiella pneumoniae* and *K. oxytoca* on the production of 2,3-butanediol. *Bioprocess Biosyst. Eng.* **35**, 1081–1088 (2012).

192. Ji, X.-J., Huang, H., Li, S., Du, J. & Lian, M. Enhanced 2,3-butanediol production by altering the mixed acid fermentation pathway in *Klebsiella oxytoca*. *Biotechnol. Lett.* **30**, 731–734 (2008).

193. Novak, K., Kutscha, R. & Pflügl, S. Microbial upgrading of acetate into 2,3-butanediol and acetoin by *E. coli* W. *Biotechnol. Biofuels* **13**, 177 (2020).

194. Hession, A. O., Esrey, E. G., Croes, R. A. & Maxwell, C. A. N-Acetylglutamate and N-Acetylaspartate in Soybeans (*Glycine max L.*), Maize (*Zea maize L.*), and Other Foodstuffs. *J. Agric. Food Chem.* **56**, 9121–9126 (2008).

195. McGinnis, S. D. & O'Brian, M. R. The Rhizobial *hemA* Gene Is Required for Symbiosis in Species with Deficient α-Aminolevulinic Acid Uptake Activity. *Plant Physiol.* **108**, 1547–1552 (1995).

196. Korkmaz, A. Effects of Exogenous Application of 5-Aminolevulinic Acid in Crop Plants. in *Abiotic Stress Responses in Plants: Metabolism, Productivity and Sustainability* (eds. Ahmad, P. & Prasad, M. N. V.) 215–234 (Springer, New York, NY, 2012).

197. Kosar, F., Akram, N. A. & Ashraf, M. Exogenously-applied 5-aminolevulinic acid modulates some key physiological characteristics and antioxidative defense system in spring wheat (*Triticum aestivum L.*) seedlings under water stress. *South Afr. J. Bot.* **96**, 71–77 (2015).

198. Von Wettstein, D., Gough, S. & Kannangara, C. Chlorophyll Biosynthesis. *Plant Cell* **7**, 1039–1057 (1995).

199. Shang, L. *et al.* A regulatory network involving Rpo, Gac and Rsm for nitrogen-fixing biofilm formation by *Pseudomonas stutzeri*. *Npj Biofilms Microbiomes* **7**, 1–15 (2021).

200. Flemming, H.-C. *et al.* Biofilms: an emergent form of bacterial life. *Nat. Rev. Microbiol.* **14**, 563–575 (2016).

201. López, D., Vlamakis, H. & Kolter, R. Biofilms. *Cold Spring Harb. Perspect. Biol.* **2** (2010).

202. Konopka, A. What is microbial community ecology? *ISME J.* **3**, 1223–1230 (2009).

203. Bester, E., Kroukamp, O., Wolfaardt, G. M., Boonzaaier, L. & Liss, S. N. Metabolic Differentiation in Biofilms as Indicated by Carbon Dioxide Production Rates. *Appl. Environ. Microbiol.* **76**, 1189–1197 (2010).

204. Høiby, N. *et al.* The clinical impact of bacterial biofilms. *Int. J. Oral Sci.* **3**, 55–65 (2011).

205. *Geomicrobiology*. (CRC Press, Boca Raton, 2008).

206. Halan, B., Buehler, K. & Schmid, A. Biofilms as living catalysts in continuous chemical syntheses. *Trends Biotechnol.* **30**, 453–465 (2012).

207. Meneses, C. H. S. G., Rouws, L. F. M., Simões-Araújo, J. L., Vidal, M. S. & Baldani, J. I. Exopolysaccharide Production Is Required for Biofilm Formation and Plant Colonization by the Nitrogen-Fixing Endophyte *Gluconacetobacter diazotrophicus*. *Mol. Plant. Microbe Interact.* **24**, 1448–1458 (2011).

208. Herath, H. M. L. I., Menikdiwela, K. R., Igalavithana, A. D. & Seneviratne, G. Developed Fungal-Bacterial Biofilms Having Nitrogen Fixers: Universal Biofertilizers for Legumes and Non-Legumes. in *Biological Nitrogen Fixation* (ed. de Bruijn, F. J.) 1041–1046 (John Wiley & Sons, Inc, Hoboken, NJ, USA, 2015).

209. Velmourougane, K., Prasanna, R. & Saxena, A. K. Agriculturally important microbial biofilms: Present status and future prospects. *J. Basic Microbiol.* **57**, 548–573 (2017).

210. Davies, K. G. & Whitbread, R. Factors affecting the colonisation of a root system by fluorescent *Pseudomonads*: The effects of water, temperature and soil microflora. *Plant Soil* **116**, 247–256 (1989).

211. Bogino, P. C., Oliva, M. D. las M., Sorroche, F. G. & Giordano, W. The Role of Bacterial Biofilms and Surface Components in Plant-Bacterial Associations. *Int. J. Mol. Sci.* **14**, 15838–15859 (2013).

212. Ramey, B. E., Koutsoudis, M., Bodman, S. B. von & Fuqua, C. Biofilm formation in plant–microbe associations. *Curr. Opin. Microbiol.* **7**, 602–609 (2004).

213. Wang, G., Zhao, G., Chao, X., Xie, L. & Wang, H. The Characteristic of Virulence, Biofilm and Antibiotic Resistance of *Klebsiella pneumoniae*. *Int. J. Environ. Res. Public. Health* **17**, 6278 (2020).

214. Sharma, G. *et al. Escherichia coli* biofilm: development and therapeutic strategies. *J. Appl. Microbiol.* **121**, 309–319 (2016).

215. Hoover, T. R. *et al.* Identification of the V factor needed for synthesis of the iron-molybdenum cofactor of nitrogenase as homocitrate. *Nature* **329**, 855–857 (1987).

216. Stewart, P. S. Antimicrobial Tolerance in Biofilms. *Microbiol. Spectr.* **3**, 3.3.07 (2015).

217. Beebout, C. J. *et al.* Respiratory Heterogeneity Shapes Biofilm Formation and Host Colonization in Uropathogenic *Escherichia coli*. *mBio* **10**, e02400-18 (2019).

218. Idrees, M., Mohammad, A. R., Karodia, N. & Rahman, A. Multimodal Role of Amino Acids in Microbial Control and Drug Development. *Antibiotics* **9**, 330 (2020).

219. Ghosh, A. S., Chowdhury, C. & Nelson, D. E. Physiological functions of D-alanine carboxypeptidases in *Escherichia coli*. *Trends Microbiol.* **16**, 309–317 (2008).

220. von Ohle, C. *et al.* Real-Time Microsensor Measurement of Local Metabolic Activities in Ex Vivo Dental Biofilms Exposed to Sucrose and Treated with Chlorhexidine. *Appl. Environ. Microbiol.* **76**, 2326–2334 (2010).

221. Kragh, K. N. *et al.* Role of Multicellular Aggregates in Biofilm Formation. *mBio* **7**, e00237-16 (2016).

222. Davies, S. K. *et al.* Visualizing Antimicrobials in Bacterial Biofilms: Three-Dimensional Biochemical Imaging Using TOF-SIMS. *mSphere* **2** (2017).

223. Liao, Y.-C. *et al.* An Experimentally Validated Genome-Scale Metabolic Reconstruction of *Klebsiella pneumoniae* MGH 78578, *i* YL1228. *J. Bacteriol.* **193**, 1710–1717 (2011).

224. Wingender, J., Neu, T. R. & Flemming, H.-C. What are Bacterial Extracellular Polymeric Substances? in *Microbial Extracellular Polymeric Substances* (eds. Wingender, J., Neu, T. R. & Flemming, H.-C.) 1–19 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1999).

225. O'toole *et al.* The Global Carbon Metabolism Regulator Crc Is a Component of a Signal Transduction Pathway Required for Biofilm Development by *Pseudomonas aeruginosa*. *Journal of Bacteriology*. **182-2**, 425-431 (2000).

226. Stojiljkovic, I., Bäumler, A. J. & Hantke, K. Fur Regulon in Gram-negative Bacteria: Identification and Characterization of New Iron-regulated *Escherichia coli* Genes by a Fur Titration Assay. *J. Mol. Biol.* **236**, 531–545 (1994).

227. Panina, E. M., Mironov, A. A. & Gelfand, M. S. Comparative analysis of FUR regulons in gamma-proteobacteria. *Nucleic Acids Res.* **29**, 5195–5206 (2001).

228. Qi, Z., Hamza, I. & O'Brian, M. R. Heme is an effector molecule for iron-dependent degradation of the bacterial iron response regulator (Irr) protein. *Proc. Natl. Acad. Sci.* **96**, 13056–13061 (1999).

229. Gollop, R. & Avissar, Y. J. Control of Heme Biosynthesis in Rhizobium SP. in *Advances in Nitrogen Fixation Research: Proceedings of the 5th International Symposium on Nitrogen Fixation, Noordwijkerhout, The Netherlands, August 28 – September 3, 1983* (eds. Veeger, C. & Newton, W. E.) 256–256 (Springer Netherlands, Dordrecht, 1984).

230. Zhu, B. *et al.* Multi-omics analysis of niche specificity provides new insights into ecological adaptation in bacteria. *ISME J.* **10**, 2072–2075 (2016).

231. O'Donnell, S. T., Ross, R. P. & Stanton, C. The Progress of Multi-Omics Technologies: Determining Function in Lactic Acid Bacteria Using a Systems Level Approach. *Front. Microbiol.* **10**, (2020).

232. Yamazaki, S. *et al.* Field multi-omics analysis reveals a close association between bacterial communities and mineral properties in the soybean rhizosphere. *Sci. Rep.* **11**, 8878 (2021).

233. Winkler, M., Lawrence, J. R. & Neu, T. R. Selective degradation of ibuprofen and clofibric acid in two model river biofilm systems. *Water Res.* **35**, 3197–3205 (2001).

234. Kim, H. K. & Verpoorte, R. Sample preparation for plant metabolomics. *Phytochem. Anal.* **21**, 4–13 (2010).

235. Altaf, Mohd. M. & Ahmad, I. In vitro and In vivo biofilm formation by *Azotobacter* isolates and its relevance to rhizosphere colonization. *Rhizosphere* **3**, 138–142 (2017).

236. Gika, H., Virgiliou, C., Theodoridis, G., Plumb, R. S. & Wilson, I. D. Untargeted LC/MS-based metabolic phenotyping (metabonomics/metabolomics): The state of the art. *J. Chromatogr. B* **1117**, 136–147 (2019).

237. Thistlethwaite, L. R. *et al.* Clinical diagnosis of metabolic disorders using untargeted metabolomic profiling and disease-specific networks learned from profiling data. *Sci. Rep.* **12**, 6556 (2022).

238. Withers, E., Hill, P. W., Chadwick, D. R. & Jones, D. L. Use of untargeted metabolomics for assessing soil quality and microbial function. *Soil Biol. Biochem.* **143**, 107758 (2020).

239. Elias, S. & Banin, E. Multi-species biofilms: living with friendly neighbors. *FEMS Microbiol. Rev.* **36**, 990–1004 (2012).

240. Peyrusson, F. *et al.* Intracellular *Staphylococcus aureus* persists upon antibiotic exposure. *Nat. Commun.* **11**, 2200 (2020).

241. Guo, S., Zhang, C. & Le, A. The limitless applications of single-cell metabolomics. *Curr. Opin. Biotechnol.* **71**, 115–122 (2021).

242. Emara, S. *et al.* Single-Cell Metabolomics. in *Metabolomics: From Fundamentals to Clinical Applications* (ed. Sussulini, A.) 323–343 (Springer International Publishing, Cham, 2017).

243. De Curtis, F., Lima, G., Vitullo, D. & De Cicco, V. Biocontrol of *Rhizoctonia solani* and *Sclerotium rolfsii* on tomato by delivering antagonistic bacteria through a drip irrigation system. *Crop Prot.* **29**, 663–670 (2010).

244. Chai, Y. N., Futrell, S. & Schachtman, D. P. Assessment of Bacterial Inoculant Delivery Methods for Cereal Crops. *Front. Microbiol.* **13**, (2022).

245. Jansson, J. K. & Hofmockel, K. S. The soil microbiome—from metagenomics to metaphenomics. *Curr. Opin. Microbiol.* **43**, 162–168 (2018).

246. Barko, P. c., McMichael, M. a., Swanson, K. s. & Williams, D. a. The Gastrointestinal Microbiome: A Review. *J. Vet. Intern. Med.* **32**, 9–25 (2018).

247. Banerjee, S. & van der Heijden, M. G. A. Soil microbiomes and one health. *Nat. Rev. Microbiol.* **21**, 6–20 (2023).

248. Compant, S., Samad, A., Faist, H. & Sessitsch, A. A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. *J. Adv. Res.* **19**, 29–37 (2019).

249. Mahanty, T. *et al.* Biofertilizers: a potential approach for sustainable agriculture development. *Environ. Sci. Pollut. Res.* **24**, 3315–3335 (2017).

Appendices

Appendix 1. Settings used for liquid chromatography-mass spectrometry experiments.

	AccQ-Tag RPLC	RPLC POS	IPC	RPLC NEG
LC unit	ACQUITY Binary Solvent Manager (Waters Corporation) + CTC Pal autosampler	UFLX XR (Shimadzu)	ACQUITY Binary Solvent Manager (Waters Corporation) + CTC Pal autosampler	ACQUITY Binary Solvent Manager (Waters Corporation) + CTC Pal autosampler
Mobile phase A	0.1 % Formic acid water	0.1 % Formic acid water	10mM of tributylamine + 15mM of acetic acid + acetylacetone (0.02% v/v) in water	0.1 % Formic acid water
Mobile phase B	0.1 % Formic acid Acetonitrile	0.1 % Formic acid Acetonitrile	80% methanol and 20% isopropanol	0.1 % Formic acid Acetonitrile
Flow rate	0.6 ml/min	0.4 ml/min	0.4 ml/min	0.4 ml/min
Column temperature	50 ℃	45 ℃	45 ℃	45 ℃
Column	Waters HSS T3 (1.8 µm, 2.1 x 150 mm)	Waters HSS T3 (1.8 µm, 2.1 x 100 mm)	Waters HSS T3 (1.8 µm, 2.1 x 100 mm)	Waters HSS T3 (1.8 µm, 2.1 x 100 mm)
Injection volume	5 µl	5 µl	5 µl	5 µl
Gradient	0.1 to 12 min from 1% solvent B to 28% solvent B; at a flow rate of 0.6 mL/min. Post acquisition wash and recovery steps took 2.65 min in total and went up to 100% B, and 1 min at initial conditions for equilibration.	0min 1%; 0.5 min, 1% B; 10 min, 55% B; 11 min, 100% B; 12 min, 100% B; 12.5 min, 1% B; 14.5 min, 1% B.	0-0.5 min: 0% B; 0.5-6 min: 0% B-5%B; 4-6min: 5%B; 6-6.5 min: 5-20% B; 6.5-8.5 min: 20% B; 8.5-14min: 20-55%B; 14-15 min: 55-100% B; 15-17min: 100%B; 17-18min 100-0% B. 3 min at initial conditions before next injection cycle starts.	0min: 99% A; 0.5min: 99%A; 10 min: 45%A; 11min: 0%A; 12 min: 0%A; 12.5min: 99%A; 14.5 min: 99%A. Linear gradient between all times
MS type	Waters XEVO TQ-S (Waters Corporation)	QTRAP 5500 MS/MS system [Applied Biosystem (CA, USA)]	Waters XEVO TQ-S (Waters Corporation)	Waters XEVO TQ-S (Waters Corporation)
Ion mode	POS	POS	NEG	NEG
Source parameter s	capillary voltage, 1.5 kV; source temperature, 150°C; cone voltage: 40 V;	De-clustering, entrance and collision cell exit potentials were set to 30 V, 10 V and 10 V, respectively.	capillary voltage, 1.5 kV; source temperature, 150°C; cone voltage: 40 V;	capillary voltage, 2 kV; source temperature, 150°C; cone voltage: 40 V;

Additional source settings:	desolvation temperature 600°C; desolvation gas flow, 1000 L/h; cone gas flow, 150 L/h; nebuliser gas 7.0 bar; collision gas, 0.15 mL/min.	Curtain gas, 40 psi; collision gas, medium; ionspray voltage, 5500 V; temperature, 550°C; ion source gas 1 and 2, 40 and 60 psi, respectively.	desolvation temperature 600°C; desolvation gas flow, 1000 L/h; cone gas flow, 150 L/h; nebuliser gas 7.0 bar; collision gas, 0.15 mL/min.	desolvation temperature 300°C; desolvation gas flow, 1000 L/h; cone gas flow, 150 L/h; nebuliser gas 7.0 bar; collision gas, 0.15 mL/min.

Appendix 2. Example python script for EDA and basic plotting

```
#!/usr/bin/env python

# Import packages
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
import math
import statsmodels.api as sm
from statsmodels.formula.api import ols
from tabulate import tabulate
import scipy.stats as stats
from scipy.stats import ttest_ind
import os
import plotly.express as px


# Carry out T-test with option to print results and export dataframe to csv
def MRC_Ttest(test_list, dataframe, variable, group1_name, group2_name, comparison, csv_name,
to_print = False, export = False):
    list_name = []
    for i in test_list:
        t_test_results = {}
        group1 = dataframe.loc[(dataframe[variable] == group1_name), i]
        group2 = dataframe.loc[(dataframe[variable] == group2_name), i]
        test = ttest_ind(group1, group2, equal_var = False)
        t_test_results[i] = test
        tempdf = pd.DataFrame.from_dict(t_test_results, orient = 'Index')
        tempdf.columns = ['t-statistic','p_val']
        list_name.append(tempdf)
        if to_print == True:
            print(i)
            print(test)
            print('')

    final_dfname = pd.concat(list_name).reset_index()
    final_dfname = final_dfname.rename(columns = {'index':'Metabolite'})
    final_dfname['comparison'] = comparison
    #print(final_dfname.head())

    if export == True:
        final_dfname.to_csv(csv_name, index = False)

        return final_dfname
    else:
        return final_dfname
```

```
def Ttest_plots(df1, df2, comp_name, signif_val = 0.001, order = [], to_save = False ):
   plot_list = df1.loc[df1['p_val'] <= signif_val, 'Metabolite'].to_list()
   for i in range(len(plot_list)):
      fig, bar = plt.subplots(figsize = (15,15))
      bar = sns.barplot(data = df2, x = 'Sample', y =df2[plot_list[i]],capsize=.2, order =order, palette =
palette)
      bar.set_ylabel('Average Normalised Peak Area', fontsize=16)
      bar.set_xlabel('Treatment', fontsize= 16)
      bar.set_title(plot_list[i])
      save_title = plot_list[i] +'_' + comp_name +'_combined_biofilm.png'
      output_path = os.path.join(wind_SP_2+save_title)

      if to_save == True:
         fig.savefig(output_path, bbox_inches='tight')

#Set paths
#Windows path
wind_path = 'Combined_MFC_data.csv'
wind_path_T = 'Combined_MFC_data.csv'


# Import data
df = pd.read_csv(wind_path)
dft = pd.read_csv(wind_path_T)

# Rename Day column to bring in-line with batch culture
df = df.rename(columns = {'Day':'Timepoint'})
df = df.fillna(0)

#Check df
df.head(15)

#Set Order and palette
order = ['WT_LN_D2', 'WT_HN_D2', 'WT_LN_D4', 'WT_HN_D4', 'amtB_LN_D4',
'WT_LN_D6','WT_HN_D6']
palette = {'WT_LN_D2':'#75bbfd',
        'WT_HN_D2':'#fc5a50',
        'WT_LN_D4':'#0343df',
        'WT_HN_D4':'tab:red',
        'amtB_LN_D4':'peru',
        'WT_LN_D6':'#0504aa',
        'WT_HN_D6':'#980002'}

# Test plot exploratory data analysis
fig, bar = plt.subplots(figsize = (15,15))
bar = sns.barplot(data = df, x = 'Sample', y ='GDP',capsize=.2, order = order, palette = palette)
bar.set_ylabel('Peak area', fontsize=14)
bar.set_xlabel('Type', fontsize=14)
bar.set_xticklabels(df.Sample, fontsize=8, rotation=90);

# Create new dataframes for analysis
```

```
# WT L – compare 3 timepoints -- WTLdf one-way anova
WTLdf = df.loc[((df['Strain'] == 'WT') & (df['Nitrogen'] == 'LN'))]

# WT 2 days H&L -- WTtwodf
WTtwodf = df.loc[((df['Strain'] == 'WT') & (df['Timepoint'] == 'D2'))]

# WT 4 days H&L -- WTfourdf
WTfourdf = df.loc[((df['Strain'] == 'WT') & (df['Timepoint'] == 'D4'))]

# WT 6 days H&L -- WTsixdf
WTsixdf = df.loc[((df['Strain'] == 'WT') & (df['Timepoint'] == 'D6'))]

# WT 2-way anova WT nitrogen status and time -- WTcombidf
WTcombidf = df.loc[(df['Strain'] == 'WT')]

# WT and amtB at 4 days -- straindf
straindf = df.loc[((df['Nitrogen'] == 'LN') & (df['Timepoint'] == 'D4'))]

# Compare WT low day 2 vs day 4 -- WTL2v4
WTL2v4 = df.loc[((df['Nitrogen'] == 'LN') & (df['Strain'] == 'WT') & (df['Timepoint'] !='D6'))]

# Compare WT low day 2 vs day 6 -- WTL2v6
WTL2v6 = df.loc[((df['Nitrogen'] == 'LN') & (df['Strain'] == 'WT') & (df['Timepoint'] != 'D4'))]

# Compare WT low day 4 vs day 6 -- WTL4v6
WTL4v6 = df.loc[((df['Nitrogen'] == 'LN') & (df['Strain'] == 'WT') & (df['Timepoint'] != 'D2'))]

# Run T-tests #
# Compare WT 2 days H&L -- WTtwodf
final_WTtwodf = MRC_Ttest(test_list = metabolite_list, dataframe = WTtwodf, variable = 'Nitrogen',
group1_name = 'LN', group2_name = 'HN', comparison = 'WTday2', csv_name = 'WTday2.csv',
to_print = False, export = False)

# Compare WT 4 days H&L -- WTfourdf
final_WTfourdf = MRC_Ttest(test_list = metabolite_list, dataframe = WTfourdf, variable = 'Nitrogen',
group1_name = 'LN', group2_name = 'HN', comparison = 'WTday4', csv_name = 'WTday4.csv',
to_print = False, export = False)

# Compare WT 6 days H&L -- WTsixdf
final_WTsixdf = MRC_Ttest(test_list = metabolite_list, dataframe = WTsixdf, variable = 'Nitrogen',
group1_name = 'LN', group2_name = 'HN', comparison = 'WTday6', csv_name = 'WTday6.csv',
to_print = False, export = False)

# Compare WT and amtB at 4 days -- straindf
final_straindf = MRC_Ttest(test_list = metabolite_list, dataframe = straindf, variable = 'Strain',
group1_name = 'WT', group2_name = 'amtB', comparison = 'StrainD4', csv_name = 'strainD4.csv',
to_print = False, export = False)

# Compare WT low day 2 vs day 4 -- WTL2v4
```

```
final_WTL2v4 = MRC_Ttest(test_list = metabolite_list, dataframe = WTL2v4, variable = 'Timepoint',
group1_name = 'D2', group2_name = 'D4', comparison = 'WTL2v4', csv_name = 'WTL2v4.csv',
to_print = False, export = False)

# Compare WT low day 2 vs day 6 -- WTL2v6
final_WTL2v6 = MRC_Ttest(test_list = metabolite_list, dataframe = WTL2v6, variable = 'Timepoint',
group1_name = 'D2', group2_name = 'D6', comparison = 'WTL2v6', csv_name = 'WTL2v6.csv',
to_print = False, export = False)

# Compare WT low day 4 vs day 6 -- WTL4v6
final_WTL4v6 = MRC_Ttest(test_list = metabolite_list, dataframe = WTL4v6, variable = 'Timepoint',
group1_name = 'D4', group2_name = 'D6', comparison = 'WTL4v6', csv_name = 'WTL4v6.csv',
to_print = False, export = False)

# T-test plots #
# Compare WT 2 days H&L -- WTtwodf
Ttest_plots(final_WTtwodf, WTtwodf,'WTtwodf',order =['WT_LN_D2','WT_HN_D2'], to_save = False)

# Compare WT 4 days H&L -- WTfourdf
Ttest_plots(final_WTfourdf, WTfourdf, 'WTfourdf', order =['WT_LN_D4','WT_HN_D4'], to_save =
False)

# Compare WT 6 days H&L -- WTsixdf
Ttest_plots(final_WTsixdf, WTsixdf, 'WTsixdf',order =['WT_LN_D6','WT_HN_D6'], to_save = False)

# Compare WT and amtB at 4 days -- straindf
Ttest_plots(final_straindf, straindf, 'straindf',order =['WT_LN_D4','amtB_LN_D4'], to_save = False)

# Compare WT low day 2 vs day 4 -- WTL2v4
Ttest_plots(final_WTL2v4, WTL2v4, 'WTL2v4',order =['WT_LN_D2','WT_LN_D4'], to_save = False)

# Compare WT low day 2 vs day 6 -- WTL2v6
Ttest_plots(final_WTL2v6, WTL2v6, 'WTL2v6',order =['WT_LN_D2','WT_LN_D6'], to_save = False)

# Compare WT low day 4 vs day 6 -- WTL4v6
Ttest_plots(final_WTL4v6, WTL4v6, 'WTL4v6',order =['WT_LN_D4','WT_LN_D6'], to_save = False)

# Low N barplot
# Note - Create new order if needed
order = ['WT_LN_D2', 'WT_LN_D4', 'WT_LN_D6']
palette = {'WT_LN_D2':'#75bbfd',
        'WT_HN_D2':'#fc5a50',
        'WT_LN_D4':'#0343df',
        'WT_HN_D4':'tab:red',
        'amtB_LN_D4':'peru',
        'WT_LN_D6':'#0504aa',
        'WT_HN_D6':'#980002'}

# Generate plots comparing individual metabolites across treatments
wind_SP_1 = 'save\\path\\here\\'
mac_SP_1 = '/save\\path\\here\\'
```

```
# Make sure to update output path
for i in range(len(metabolite_list)):
   fig, bar = plt.subplots(figsize = (15,15))
   bar = sns.barplot(data = WTLdf, x = 'Sample', y =WTLdf[metabolite_list[i]],capsize=.2, order = order,
palette = palette)
   bar.set_ylabel('Peak Area', fontsize=14)
   bar.set_xlabel('Treatment', fontsize=14)
   bar.set_title(metabolite_list[i])
   save_title = metabolite_list[i] + '_Biofilm_WTL.png'
   output_path = os.path.join(wind_SP_1+save_title)
   fig.savefig(output_path, bbox_inches='tight')

#Read in and plot AA heat map
lowN_heat = pd.read_csv('LowN_heatmap_MFC_data.csv')
lowN_heat.head()

#Define Amino Acid List
AA_list =
['Metabolite','ARGININE','HISTIDINE','AMMONIUM','ASPARTATE','GLUTAMATE','SERINE','THREONINE',
      'ASPARAGINE','GLUTAMINE','GLYCINE','PROLINE','CYSTEINE','ALANINE','VALINE',
      'ISOLEUCINE','LEUCINE','METHIONINE','PHENYLALANINE','TYROSINE','TRYPTOPHAN','LYSINE']

# Drop non-AA cols
df = lowN_heat.drop(columns=[col for col in lowN_heat if col not in AA_list])
df = df.set_index('Metabolite')
df.head()

fig, heat = plt.subplots(figsize = (15,15))
heat = sns.heatmap(df, annot=False,cmap="YlGnBu")
heat.set_ylabel('Timepoint', fontsize=14)
heat.set_xlabel('Metabolite', fontsize=14)
plt.setp(line.get_legend().get_texts(), fontsize='8') # for legend text
plt.setp(line.get_legend().get_title(), fontsize='18') # for legend title
fig.savefig('path\\Biofilm_LowN_all.png', bbox_inches='tight')

# Define AA list with most abundant metabolites removed
short_AA = ['Metabolite','ARGININE','HISTIDINE','AMMONIUM','ASPARTATE','SERINE','THREONINE',
      'ASPARAGINE','GLUTAMINE','PROLINE','CYSTEINE',
      'ISOLEUCINE','LEUCINE','METHIONINE','TYROSINE','TRYPTOPHAN','LYSINE']

# Drop highly abundant-AA cols
df = df.drop(columns=[col for col in df if col not in short_AA])
#df = df.set_index('Metabolite')
df.head()

fig, heat = plt.subplots(figsize = (15,15))
heat = sns.heatmap(df, annot=False,cmap="YlGnBu")
heat.set_ylabel('Timepoint', fontsize=14)
heat.set_xlabel('Metabolite', fontsize=14)
plt.setp(line.get_legend().get_texts(), fontsize='8') # for legend text
```

```
plt.setp(line.get_legend().get_title(), fontsize='18') # for legend title
fig.savefig('path\\Biofilm_LowN_small.png', bbox_inches='tight')

#Check Dataframe
df.head()

# Compare WT L 3 timepoints - WTLdf one-way anova
WTLdf_list = []
for i in metabolite_list:
   model = ols('WTLdf[i] ~ C(Timepoint)', data = WTLdf).fit()
   result = sm.stats.anova_lm(model, type = 2)
   result['Metabolite'] = i
   WTLdf_list.append(result)
   print(i)
   print(tabulate(result, headers = ['component','df','sum_sq','mean_sq','F','p_value','Metabolite']))
   print('')

# Creates dataframe from list of 1way anova dataframes
final_WTLdf = pd.concat(WTLdf_list).reset_index()
final_WTLdf = final_WTLdf.rename(columns = {'index':'component'})
final_WTLdf.head()

# Export Analysis to csv
final_WTLdf.to_csv('WTL_anova_biofilm.csv', index = False)

# WT 2-way anova WT nitrogen status and time -- WTcombidf
WTcombidf_list = []
for i in metabolite_list:
   model = ols('WTcombidf[i] ~ C(Timepoint) + C(Nitrogen) + C(Timepoint):C(Nitrogen)', data =
WTcombidf).fit()
   result = sm.stats.anova_lm(model, type = 2)
   result['Metabolite'] = i
   WTcombidf_list.append(result)
   print(i)
   print(tabulate(result, headers = ['component','df','sum_sq','mean_sq','F','p_value','metabolite']))
   print('')

# Generate dataframe based on list of 2way anova dataframes
final_WTcombidf = pd.concat(WTcombidf_list).reset_index()
final_WTcombidf = final_WTcombidf.rename(columns = {'index':'component'})
final_WTcombidf.head()

# Export Analysis to csv
final_WTcombidf.to_csv('WT_2anova_biofilm.csv', index = False)

## Exploratory multivariate data analysis: PCA
pca = PCA()
components = pca.fit_transform(df[metabolite_list])
labels = {
   str(i): f"PC {i+1} ({var:.1f}%)"
   for i, var in enumerate(pca.explained_variance_ratio_ * 100)
```

```
}

fig = px.scatter_matrix(
    components,
    labels=labels,
    dimensions=range(4),
    color=df["Sample"]
)
fig.update_traces(diagonal_visible=False)
fig.show()
fig.write_image('pcacompare_comb_biofilm.png')

# PCA plotting two components only

X = df[metabolite_list]

pca = PCA(n_components=2)
components = pca.fit_transform(X)
pcadf = pd.DataFrame(components, columns = ['PC1','PC2'])
fig = px.scatter(pcadf, x='PC1', y='PC2', color=df['Sample'])

fig.show()
pca.explained_variance_ratio_
fig.write_image('pca2comp_comb_biofilm.png')

# Two component PCA plot with loadings
X = df[metabolite_list]

pca = PCA(n_components=4)
components = pca.fit_transform(X)
pcadf = pd.DataFrame(components, columns = ['PC1','PC2','PC3','PC$'])

loadings = pca.components_.T * np.sqrt(pca.explained_variance_)

fig = px.scatter(pcadf, x='PC1', y='PC2', color=df['Sample'])

for i, feature in enumerate(metabolite_list):
    fig.add_shape(
        type='line',
        x0=0, y0=0,
        x1=loadings[i, 0],
        y1=loadings[i, 1]
    )
    fig.add_annotation(
        x=loadings[i, 0],
        y=loadings[i, 1],
        ax=0, ay=0,
        xanchor="center",
        yanchor="bottom",
        text=feature,
    )
```

```
fig.show()
fig.write_image('pca2comp_LOAD_comb_biofilm.png')
pca.explained_variance_ratio_

# Heatmap comparing loadings to PCs
df_comp = pd.DataFrame(pca.components_,columns = metabolite_list)
plt.figure(figsize = (20,7))
sns.heatmap(df_comp,cmap = 'coolwarm', annot = False, annot_kws = {'size':12})
plt.yticks([0,1,2,3],['PC1','PC2','PC3','PC4'], fontsize = 10)
plt.tight_layout()
plt.savefig('ComponentHeatmap_comb_biofilm.png', bbox_inches = 'tight')
```