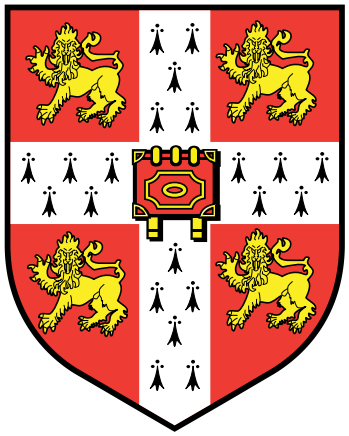
Bioremediation of waste by the bacterium *Rhodopseudomonas palustris*



John Romaine Delmar Hervey

Department of Biochemistry
University of Cambridge

A dissertation submitted for the degree of
Doctor of Philosophy

September 2018
Trinity Hall

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

Abstract

Bioremediation of waste by the bacterium *Rhodopseudomonas palustris*

John Romaine Delmar Hervey

Rhodopseudomonas palustris is a photosynthetic bacterium capable of metabolising a broad range of substrates. It is a highly robust microorganism with a high tolerance for toxic conditions, making it an ideal candidate for the remediation of waste materials by metabolism. Our societies produce large amounts of waste materials, which require processing and safe disposal, but could instead be used as resources to produce fuel and energy or sustainably recycled into new materials. *R. palustris* can be used to process waste glycerol (a byproduct of biodiesel manufacture), and naturally produces hydrogen gas (a clean-burning fuel) as a byproduct of this process. *R. palustris* could potentially be used to bioremediate other waste materials and could be engineered to manufacture other useful products from this process.

This thesis reports discoveries regarding the growth of *R. palustris* on previously untested substrates and the engineering of *R. palustris* to improve its utilisation of substrates and add additional metabolic pathways for the sustainable production of valuable compounds. Firstly, new substrates including urea are identified as supporting growth of *R. palustris*, suggesting that it could be used to process nitrogen-rich agricultural wastes. Nitrogen runoff and pollution pose a significant environmental problem, and safe disposal represents a significant cost to industry. Using *R. palustris* to bioremediate these wastes, converting them to more valuable materials, would help to offset the costs of waste disposal as well as providing a sustainable source of desirable compounds. Secondly, a genetic manipulation system developed to allow the expression of heterologous genes in *R. palustris* is presented, and newly engineered strains of *R. palustris* are characterised that are capable of producing the industrially significant compound cyanophycin when *R. palustris* is grown on glycerol and urea. Finally, strategies are presented for the future development of *R. palustris* as a bioremediation system and cell-factory for sustainable biotechnology.

Dedication

This thesis is dedicated to my parents, who brought me into the world, fed, clothed, and educated me, despite the best efforts of my younger self to thwart many of these activities, and explain to them that I really thought I knew best already. I am fortunate, and indeed privileged, to have had the most supportive and inspiring upbringing anybody could ask for. It is said that it is possible to choose your friends, but not your family; and whilst this cliché certainly contains some truth, I am very grateful to be able to say that my family are both great friends and family.

Acknowledgements

Thanks must go first to the Waste Environmental Education Research Trust, who funded this project. Without their support I could not have undertaken this research, or attended the University of Cambridge. I am truly grateful that such organisations exist, who recognise the great importance of our environment, and promote and support environmental research and education that can help to create a better world for everybody.

Equally importantly, I need to thank my supervisor Professor Christopher Howe, whose vision was responsible for the inception and execution of this project. He has ably put up with me for four years, offering invaluable guidance, expertise, and support throughout, despite my propensity for missing deadlines. I am proud to have been his student.

Thanks also go to Professor John Dennis, who also played a crucial role in the genesis and development of the project. His help and advice have been most gratefully received.

I have been privileged to have worked with some extraordinarily talented and caring people. The Howe lab have been welcoming, supportive, and made life at a work a great pleasure; my thanks go to all members of the lab during my time there.

I have received technical support for my project from a great many people, but in particular I need to mention Dr Karin Muller, for the electron microscopy work, Dr Peter Sharratt for amino acid analysis, and Ms Shilo Dickens and Dr Nataliya Scott from the Department of Biochemistry Sequencing facility.

A great many friends and family have supported, helped, encouraged and entertained me while I was undertaking this work. You are all loved and remembered, and I appreciate everything you have done for me.

One of the happiest aspects of my time at Cambridge has been meeting my partner Greg. I can't thank him enough for his unwavering support, friendship and love throughout. His patience with me while I was writing was incredible, and all the more astonishing given that he was writing his own thesis during much of the same time.

I must also thank my undergraduate supervisor Professor Julian Eaton-Rye, Dr Simon Jackson, and all of the wonderful people I worked with in the Eaton-Rye lab. Working with them inspired and encouraged me to pursue a scientific career, and set me on the path to where I am now.

Finally I want to thank everybody who has been involved in my education over the years. I can't possibly thank you all here, so to those who are not named, please know that you

have not been forgotten and you are very much appreciated. I particularly want to thank my former high school teachers, whose enthusiasm and communication of their subjects was so inspiring. Our teachers are often undervalued (and underpaid) and too often go unthanked and unacknowledged, when they should be venerated. To Dr Treena Blythe, Mr Graeme Bloomfield, Ms Karen Barks, Mr Max Riley, Mrs Kathy Sherwood, Ms Leisa McCauley, Mr Charles Newton, and all of the other wonderful teachers who have taught me, my sincere thanks.

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA analysis of variance 47, 58, 59

BDMA benzyldimethylamine 47

BLAST Basic Local Alignment Search Tool 61, 98, 122, 127

bp base pairs 29, 64, 67, 76–79, 84, 94, 128, 138, 139

CDW cell dry weight 44, 45, 133, 141, 142, 155, 159, 160, 163

DNase deoxyribonuclease 42

DTT dithiothreitol 88

ECL enhanced chemiluminescence 43

FAME fatty acid methyl ester 63

FPKM fragments per kilobase of transcript per million mapped reads 94, 104, 110–113, 115–120

GCMS gas chromatography mass spectrometry 72, 82, 86

HSD honest significant difference 47, 58

LB lysogeny broth 40

LED light emitting diode 52

LH2 peripheral light-harvesting complex 51, 56

MDS multi-dimensional scaling 94, 104, 110

MEC microbial electrohydrogenesis cell 32, 33

MFC microbial fuel cell 26, 29, 33, 35, 49, 126

MNA methyl-5-norbornene-2,3-dicarbocylic anhydride 47

MSTFA N-Methyl-N-(trimethylsilyl) trifluoroacetamide 44

NSA nonenyl succinic anhydride 47

OD optical density 40, 41, 51, 59, 80, 85, 89, 93, 141, 148

PABA para-aminobenzoic acid (4-aminobenzoic acid) 39

PAGE poly-acrylamide gel electrophoresis 42, 133, 140, 141, 144, 155, 158, 159

PBST phosphate buffered saline, Tween-20 43

PCA principal component analysis 94, 104

PCB printed circuit board 52

PCR polymerase chain reaction 40–42, 64, 67, 72, 76–78, 82, 84, 88, 128, 133, 136, 138, 160

PHB polyhydroxybutyrate 35

PNSB purple non-sulphur bacteria 26, 27, 33, 91

PSB purple sulphur bacteria 26, 27

RIN RNA integrity number 19, 96, 97

RIPA radioimmunoprecipitation assay buffer 88

RNase ribonuclease 42

RT reverse transcriptase 78, 79, 139

RT-PCR reverse transcription polymerase chain reaction 40, 72, 75, 78, 79, 133, 135, 136, 139

SDS sodium dodecyl-sulphate 42, 88, 93, 133, 140, 141, 144, 155, 158, 159

SEM standard error of the mean 54, 57, 58, 80, 85, 95, 148

TBE Tris borate EDTA 40

TCA tricarboxylic acid 120, 122

TE Tris EDTA 93

TEM transmission electron microscopy 136, 141, 159

WT wild type 64, 68, 72, 76–87, 89, 95, 135, 136, 138–148, 155, 158–160, 164

YP yeast, peptone 41

1 Introduction

1.1. Background

Global climate change and increases in population pressure and energy usage coupled with a dwindling supply of fossil fuels make the development of clean and sustainable energy technologies a pressing issue. Our societies produce large amounts of waste material, which require processing and safe disposal, but could instead be used as a resource to produce fuel and energy. One important area of research is in the use of micro-organisms to process this waste material. The idea of harnessing micro-organisms for the production of energy, primarily through the production of biodiesel and bio-ethanol fuels, is not new, but although many of these systems show great promise, they are not without their limitations. The conversion of biomass into biofuels consumes large amounts of raw agricultural materials, which may result in competition with the production of food and has a limited efficiency due to the energy requirements of production, harvesting and fuel extraction (Hill *et al.*, 2006). The use of algae for the production of biofuels has also been investigated in detail. Algal systems have some advantages over crop based biofuel approaches, particularly as systems such as algal raceways do not require the use of arable land and thereby alleviate the problem of competition with food production. Current systems, however, are still some way from being viable at large scales, owing to inefficiencies caused by the energy inputs required for production and fuel extraction, as well as the sizeable areas of land required (Scott *et al.*, 2010).

It is increasingly clear then, that no one system is likely to be sufficient to solve the problem of the need for renewable energy and that a variety of potential solutions need to be employed and further researched. Improving the utilisation and efficiency of use of current resources, in particular the utilisation of waste materials as a potential resource, could contribute to these solutions. In 2012 the UK sent approximately 19,733,000 tonnes of municipal waste to landfill, of which 10,293,000 tonnes were categorised as biodegradable (*UK Statistics on Waste, 2010 to 2012.* 2015). This is an example of a poorly tapped resource that could instead be used to produce energy. Furthermore, waste management is a costly and energy consuming process; using waste as a resource could help to offset these costs and instead make waste management a constructive, energy producing process.

Some biodegradable waste is already processed by micro-organisms, in a process called anaerobic digestion. This process uses acidogenic and methanogenic bacteria to degrade the organic components of the waste, producing bio-gas (principally methane) which is burnt for fuel, and digestate waste (Y. Chen *et al.*, 2008). Methane is an extremely potent greenhouse gas, meaning that any loss into the atmosphere significantly contributes to global climate change and the greenhouse effect. This problem is not much improved when, like all carbon based fuels, methane is combusted for energy, releasing carbon dioxide; also a greenhouse gas.

The digestate produced consists mostly of ligneous and cellulosic material that could not be digested by the anaerobic bacteria, as well as any plastics and other non-biodegradable material (Y. Chen *et al.*, 2008). Anaerobic digesters suffer from problems with reactor stability, due to varying populations of micro-organisms present in waste materials, and the production of inhibitory compounds, chiefly ammonia (Kayhanian, 1999; Dupla *et al.*, 2004; Y. Chen *et al.*, 2008). The end product, digestate, is also a potential resource that could be used in further bio-conversion systems.

One possible route to improving waste utilisation is through bio-conversion by specialised micro-organisms, with unusual natural, or engineered, metabolic pathways that can degrade conventionally intractable wastes, or further process partially digested waste (such as digestate), producing fuels and other valuable commodities from these materials. Of particular note is the ability of some bacteria to produce hydrogen gas, as a by-product from a process known as photofermentation (Gest and Kamen, 1949). Hydrogen gas is of great interest as fuel, as it is clean-burning, producing only water as the product of combustion, and does not release CO_2 or other greenhouse gases (except water vapour, which is quickly dispersed as rain). These bacteria could therefore be grown using waste materials as substrate, to produce bio-hydrogen, as a sustainable, clean, future fuel.

1.2. *Rhodopseudomonas palustris*

Rhodopseudomonas palustris is a photosynthetic PNSB (purple non-sulphur bacterium), in the alphaproteobacteria family, first classified by Molisch in 1907 (Molisch, 1907), and later described in more detail by van Niel in 1944 (Niel, 1944). Unlike PSB (purple sulphur bacteria), PNSB (purple non-sulphur bacteria) do not use hydrogen sulphide as an electron donor for photosynthesis. Neither group, however, uses water as their electron source, and since water splitting does not occur, photosynthesis in purple bacteria is anoxygenic and proceeds *via* a cyclic electron transfer system (Scheuring *et al.*, 2006).

R. palustris is found in a diverse range of habitats, principally pond waters and sediments (Oda and Wanders *et al.*, 2002) (its Latin name *"palustris"* referring to marshes), but also animal waste lagoons (M. K. Kim *et al.*, 2004; Sharma *et al.*, 2015), mixed culture MFC (microbial fuel cell) systems (Xing *et al.*, 2008; Venkidusamy and Megharaj, 2016),

Japanese rice paddies (Harada *et al.*, 2003), straw (Roper and Ladha, 1995), tannery effluents (Merugu *et al.*, 2010), rice noodle factory wastewater (Vikineswary *et al.*, 1997), alkali lakewater (Dönmez *et al.*, 1999), coastal sediments and earthworm droppings (Larimer *et al.*, 2004).

R. palustris has been renamed and reclassified several times (J. Gibson *et al.*, 1979; Pfennig and Trüper, 1983; Fowler *et al.*, 1984; Imhoff and Trüper *et al.*, 1984; Stackebrandt *et al.*, 1984; Woese and Stackebrandt and Weisburg *et al.*, 1984; Woese and Weisburg and Paster *et al.*, 1984; Oyaizu and Woese, 1985; Woese and Stackebrandt and Macke *et al.*, 1985; Woese and Weisburg and Hahn *et al.*, 1985; Hiraishi and Santos *et al.*, 1992; Kawasaki *et al.*, 1993; Ludwig *et al.*, 1995; Yurkov and Beatty, 1998; Imhoff and Caumette, 2004; McInerney *et al.*, 2008; Gribaldo and Brochier, 2009), and this can sometimes make make it unclear as to exactly which species is being discussed, particularly in older literature. Generally, and more recently, PNSB are accepted as belonging to the alpha (Rhizobiales & Rhodobacterales orders) and beta (Rhodocyclales & Burkholderiales orders) proteobacteria classes, with *R. palustris* falling in the alphaproteobacteria class, whilst the PSB belong exclusively to the gamma-proteobacteria. Interestingly, the alpha and beta proteobacteria classes contain both photosynthetic and non-photosynthetic bacteria, meaning that the PNSB are more closely related to some non-photosynthetic bacteria, such as the Nitrobacter genus, than they are to some photosynthetic bacteria such as the PSB (Nagashima *et al.*, 1997; Gupta, 2000; Achenbach *et al.*, 2001; Kunisawa, 2001; Bryant and Frigaard, 2006; Wu *et al.*, 2009; Gupta, 2012). A phylogenetic tree of the proteobacteria, reproduced from Gupta (2000), is shown in Figure 1.1.

Genome sequencing, along with the wide habitat diversity, suggests that *R. palustris* has great metabolic diversity (Larimer *et al.*, 2004), which makes it an excellent candidate for bio-technology research into the bio-conversion of waste products. It has been suggested that *R. palustris* has the ability to grow *via* four distinct growth modes; photoheterotrophic and photoautotrophic growth under anaerobic conditions, and chemoheterotrophic and chemoautotrophic growth under aerobic conditions (Puskás *et al.*, 2000; Ho Lee *et al.*, 2002; Larimer *et al.*, 2004). Figure 1.2, reproduced from Larimer *et al.* (2004), shows a diagram of the different types of *R. palustris* metabolism.

R. palustris is a hydrogen producing bacterium, and there has been considerable interest in its development for use in biotechnological systems, especially those involved in bio-hydrogen production (Fißler *et al.*, 1994; Ho Lee *et al.*, 2002; Oh *et al.*, 2002; Gosse *et al.*, 2007; Carlozzi, 2009; Carlozzi and Lambardi, 2009; Sabourin-Provost and Hallenbeck, 2009; Carlozzi and Pintucci *et al.*, 2010; Huang *et al.*, 2010; Yang and Lee, 2011; Adessi and McKinlay *et al.*, 2012; Y.-T. Chen *et al.*, 2012; Pott *et al.*, 2013; McKinlay *et al.*, 2014; Pott *et al.*, 2014; Lazaro *et al.*, 2017). In addition to its hydrogen producing capabilities, the metabolic diversity of *R. palustris* is coupled with great resilience to toxic compounds, enabling it to survive, and thrive, in a great range of conditions. Because of this,

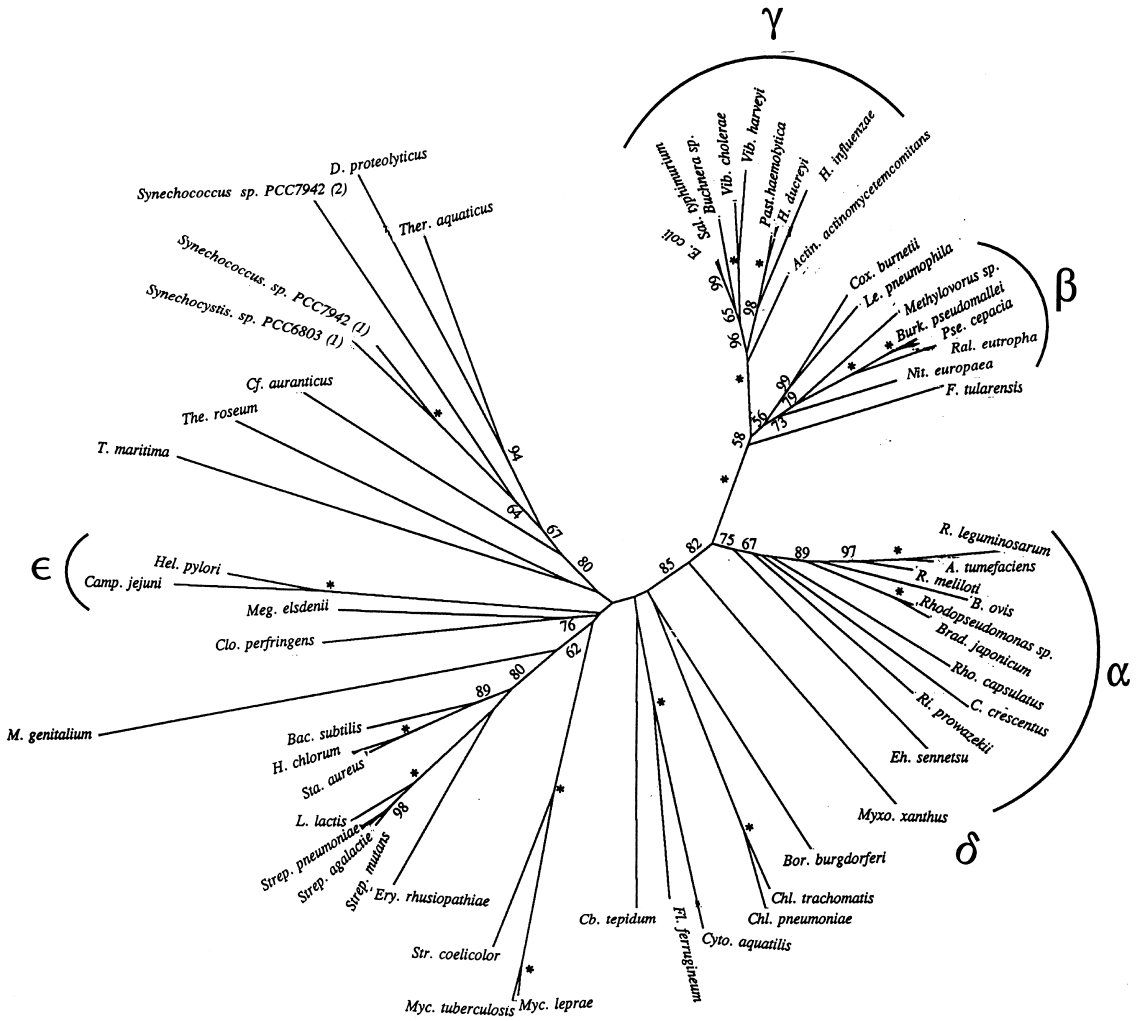


Figure 1.1. – Phylogenetic tree of the proteobacteria. This diagram shows the relationship between the different species of proteobacteria. Reproduced from Gupta, 2000, "An unrooted neighbor-joining distance tree based on Hsp70 sequences. The tree based on 362 aligned amino acids was constructed by standard procedures…Bootstrap scores >50 % are indicated on the nodes. An asterisk on a node indicates a 100 % bootstrap score."

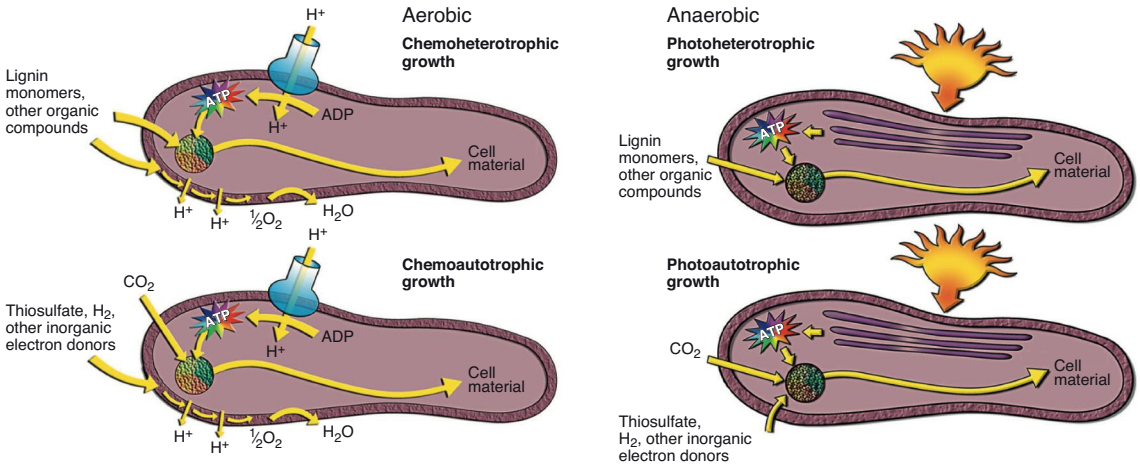


Figure 1.2. – Growth modes of *R. palustris*. This diagram shows the different growth modes of *R. palustris*. Reproduced from Larimer *et al.*, 2004, "Schematic representations of the four types of metabolism that support its growth are shown. The multicolored circle in each cell represents the enzymatic reactions of central metabolism."

R. palustris has great biotechnological potential, and could be used in to bio-remediate a great range of materials, from vegetable wastes (Adessi and McKinlay *et al.*, 2012) and animal wastes/human sewage (Sharma *et al.*, 2015), to uranium (Llorens *et al.*, 2012) and ruthenium (Colica *et al.*, 2012). Other uses of *R. palustris* have included (but are not limited to) wastewater treatment (P. Xu *et al.*, 1996), and the reduction of selenite to elemental selenium, for human health supplementation (B. Li *et al.*, 2014).

There are many strains of *R. palustris* being used in modern bio-technology research. The most widely studied of these is probably CGA009 (the strain used in this study), which had its complete genome sequenced in 2004 (Larimer *et al.*, 2004). The *R. palustris* genome has 4,836 genes and consists of a circular 5,459,213 bp (base pairs) chromosome and a 8,427 bp plasmid.

Other strains of particular note are DX-1 and TIE-1. *R. palustris* DX-1 is notable due to its electrogenic activity, having the ability to generate high power densities in MFC systems (Xing *et al.*, 2008), whilst TIE-1 is known to actually consume electricity, in a process involving iron oxidation (Jiao *et al.*, 2005; Bose *et al.*, 2014). The capacity to grow *R. palustris* in a MFC yields possibilities of producing not only hydrogen gas as a product, but also direct electrical power generation (Xing *et al.*, 2008; Pant *et al.*, 2010; Venkidusamy and Megharaj, 2016).

1.3. Waste materials

1.3.1. Glycerol

Previous work has shown that *R. palustris* is able to metabolise glycerol, producing hydrogen as a by-product (Sabourin-Provost and Hallenbeck, 2009; Pott *et al.*, 2013; 2014). Glycerol is a by-product from the production of biodiesel, via the transesterification of plant triacylglycerides. Glycerol is now produced in large quantities as a waste product from this process (Johnson and Taconi, 2007), and as such, represents an economically suitable substrate for bio-remediation to fuels and other products. The glycerol produced however, is generally caustic and highly contaminated with toxic compounds from the conversion reaction. Production of hydrogen from crude glycerol by *R. palustris* was first demonstrated by Sabourin-Provost and Hallenbeck (2009), but substantial increases in hydrogen yield were achieved by Robert Pott using relatively simple pre-treatment steps (Pott *et al.*, 2013; 2014). Utilising the waste product increases the overall energy benefit of biodiesel manufacture. Using *R. palustris* to degrade the waste glycerol, producing useful fuel is an example of waste re-purposed, and a more efficient utilisation of resources.

1.3.2. Other waste materials

Organic acids, including but not limited to acetate, succinate, benzoates & chlorinated benzoates, coumarate, trans-cinnamic acid, tributyl phosphate, phenanthrene, and cyclohexane carboxylate are some of the more commonly reported substrates that support growth of *R. palustris* (Hegeman, 1967; Harwood and J. Gibson, 1988; Madigan and Gest, 1988; M.-K. Kim and Harwood, 1991; Fißler *et al.*, 1994; Perrotta and Harwood, 1994; Küver *et al.*, 1995; Egland *et al.*, 2001; Oda and De Vries *et al.*, 2001; Berne *et al.*, 2005; C. Pan *et al.*, 2008; Zhao *et al.*, 2011; McKinlay *et al.*, 2014), in addition to some of the complex traditional substrates used in microbiology such as yeast extract.

In order to identify additional waste sources that might be suitable for bio-conversion by *R. palustris*, its growth on different sources of organic carbon will be studied. Individual compounds will be tested, so as to give specific insights into the metabolic processes involved. In addition to carbon, nitrogen is an essential part of any growth substrate, and an investigation of different nitrogen sources in growth substrates for *R. palustris* will be important when identifying suitable wastes for use as substrates. It has also been shown that *R. palustris* can be adapted to grow on different substrates, by culturing it in the presence of those substrates in a directed evolution strategy (Oda and De Vries *et al.*, 2001), raising the possibility of using this strategy to achieve degradation of waste materials.

By studying its growth on different individual compounds, an overview of the metabolism of *R. palustris* may be gleaned. This information can then be used to select waste sources that contain those compounds, and which therefore might be suitable for growth of

R. palustris. Real waste materials will be complex mixtures of different substrates, and may contain toxic contaminants, competing micro-organisms or other inhibitors of growth. Therefore, it will be important to test these more complex substrates for *R. palustris* growth in subsequent experiments, so that growth in this context can be properly characterised and optimised to develop a real-world bio-conversion system, where waste material is processed, and a useful output is produced.

Where growth on some of the individual compounds is identified, it will be important to understand how the metabolism of these compounds is regulated. This will be useful both to optimise growth conditions to achieve high bio-conversion rates, and to provide information that can be used to engineer novel, and specifically regulated, metabolic pathways. To this end, transcriptome sequencing of RNA extracted from cultures grown on different substrates will be used to look for changes in gene regulation and possibly identify novel genes or metabolic pathways.

1.4. Metabolic engineering

In addition to the production of bio-hydrogen and bio-energy from substrates such as waste glycerol, *R. palustris* could also be engineered, either to produce other fuels or high-value products and petro-chemical alternatives, or to expand and improve its utilisation of different substrates, and improve product yields from these substrates.

1.4.1. Genetic tools for *R. palustris*

Genetic manipulation has been used in *R. palustris* to improve its production of hydrogen, particularly involving attempts alter the regulation of nitrogenase, one of the key enzymes responsible for the production of hydrogen (Oda and Samanta *et al.*, 2005; Rey and Oda *et al.*, 2006; Heiniger *et al.*, 2012; Gordon and McKinlay, 2014). Transposon mutagenesis has been used to study the essential genome (Elder *et al.*, 1993; Pechter *et al.*, 2016). A cloning vector based on the native *R. palustris* cryptic plasmid was developed for the insertion of heterologous sequences (Inui *et al.*, 2000), and other plasmid based systems have been used to direct the production of n-butanol (Doud *et al.*, 2017) and squalene (W. Xu *et al.*, 2016). Genes from *R. palustris* have also been used in other organisms, in order to introduce new functionalities (J.-J. Pan *et al.*, 2015).

Development of a stable heterologous expression system in *R. palustris* based on the integration of trans-genes into the genome would allow the stable expression of trans-genes without the use of constant selection, which would be more suitable for industrial systems. The design of an heterologous expression system integrated into the genome will require the identification of a suitable neutral insert site, and selection of promoters and terminators to control the expression of inserted heterologous genes.

1.4.2. Pathway engineering

Given the availability of inexpensive waste glycerol, and its current under-utilisation (Johnson and Taconi, 2007), it makes an excellent choice of substrate for bio-remediation by *R. palustris*. However, to make the production of hydrogen, or other products, from glycerol economically feasible, the efficiency needs to be increased, to give better yields and higher conversion rates (Basak and Das, 2006). By engineering of the metabolic pathways, including glycerol specific transporters, this project will attempt to increase the utilisation of glycerol by *R. palustris*.

Secondly, whilst hydrogen is a very useful product, it does not account for a great proportion of the mass of wastes, since most biomass and other waste materials are primarily carbon and nitrogen based. In order to utilise the waste substrate fully, other valuable compounds, in addition to hydrogen gas, should be produced by *R. palustris*. Selection of suitable products will depend on the waste material used as substrate, but in general, will need to be of high value, easily purified from the bacterial culture, produced by relatively simple metabolic pathways that can easily be introduced to *R. palustris*, and ideally not be a product that is also natively consumed by *R. palustris*.

1.5. Hydrogen production in other microorganisms

Four principal routes exist for hydrogen production by microorganisms: dark fermentation by anaerobic bacteria, biophotolysis by green algae and cyanobacteria, photofermentation by photosynthetic bacteria (such as *R. palustris*), and microbial electrolysis using hybrid MEC (microbial electrohydrogenesis cell) devices (Hallenbeck *et al.*, 2009; Hallenbeck, 2011). Two types of enzyme are ultimately responsible for the production of hydrogen in these pathways, namely hydrogenase and nitrogenase. Hydrogenase catalyses the reversible reaction $2\,H^+ + 2\,e^- \rightleftharpoons H_2$ (Vignais and Colbeau *et al.*, 1985; Vignais and Toussaint, 1994; Vignais and Colbeau, 2004), while nitrogenase catalyses the fixation of nitrogen gas (N_2) to ammonia with the concomitant conversion of protons to hydrogen gas (H_2) as a byproduct (Vignais and Colbeau *et al.*, 1985).

A thorough review of the various benefits and challenges of the different microbial hydrogen production methods is given in Hallenbeck *et al.* (2009) and Hallenbeck (2011). To cover some of the most relevant points though; there are some appealing advantages to using photofermentation with an organism like *R. palustris* for the production of hydrogen. Although hydrogen production by dark fermentation (commonly using species from the *Enterobacter* and *Clostridium* genera) is significantly faster than photofermentation, and does not require illumination and the associated technical challenges, incomplete substrate conversion and the production of unwanted byproducts tends to lead to low overall yields and low substrate to product conversion efficiency. Some studies have suggested a

two-stage process utilising an initial dark fermentation step for rapid hydrogen production, followed by a photofermentation step using the unused substrates and unwanted products from the previous dark fermentation as inputs to produce greater yields of hydrogen (Asada *et al.*, 2006; Hallenbeck *et al.*, 2009). This, however, significantly increases the cost and complexity of the system, making it less suitable for industrial scale applications.

The primary limitation of biophotolysis based hydrogen production, using green photosynthetic organisms, is due to oxygen. These organisms are grown under aerobic conditions, in addition to the oxygen they themselves produce as part of the water-splitting reaction of photosynthesis. Due to their metallo-reaction centres, the hydrogen producing enzymes hydrogenase and nitrogenase are both very sensitive to oxygen, and this sensitivity is one of the primary problems associated with biological hydrogen production. Photofermentation sidesteps this problem, since it occurs under anaerobic conditions using anoxygenic photosynthesis, removing the obstacle of inhibition by oxygen.

Finally, microbial electrolysis uses a variation of a MFC, termed a MEC, to degrade organic materials, producing electrons that pass to an electron sink (the anode), and are then used at the cathode to drive hydrogen production. An external power supply is also required, to produce sufficient voltage to drive hydrogen production. In addition to the need for an external power source, these devices are complex, often require expensive precious metals for cathodes, and hydrogen output can be variable.

Aside from *R. palustris*, some of the most commonly studied PNSB for photofermentative hydrogen production are *Rhodobacter capsulatus*, *Rhodobacter sphaeroides*, and *Rhodospirillum rubrum*. *R. palustris* is of particular interest as, almost uniquely, it possesses three different nitrogenase variants (Larimer *et al.*, 2004; Oda and Samanta *et al.*, 2005). Most nitrogen fixing bacteria use a nitrogenase which has a molybdenum-iron cofactor (Eady, 1996; Dixon and Kahn, 2004). Most bacteria also have the capability to produce another nitrogenase with a different metal cofactor, usually iron-iron, but in some cases vanadium-iron. *R. palustris* is unusual in its ability to produce all three variants, whereas *Rb. capsulatus*, and *Rp. rubrum* only encode Mo and Fe variants, and *Rb. sphaeroides* only has the Mo-nitrogenase (Eady, 1996; Masepohl and Hallenbeck, 2010; Masepohl, 2017).

The three nitrogenase variants all catalyse the same fundamental reaction, however, the vanadium-iron and iron-iron cofactor alternatives have been shown to be capable of producing greater amounts of hydrogen than the molybdenum-iron alternative (Miller and Eady, 1988; Schneider *et al.*, 1997; Oda and Samanta *et al.*, 2005). The ability of *R. palustris* to express all three alternatives makes it particularly attractive candidate for studies involving microbial hydrogen production. The regulation over which nitrogenase isozyme is expressed is complex, but has been shown to be dependent not only on metal availability but also the presence of fixed nitrogen. *R. palustris* has also been shown to be less sensitive to the presence of transition metals (especially molybdenum) than other bacteria when attempting to express the alternative V and Fe nitrogenases (Oda and Samanta *et al.*,

2005).

As long as it is supplied with sufficient ATP, nitrogenase continues to function even in the absence of dinitrogen substrate, and this has been shown to result in an increase in hydrogen production (Sasikala *et al.*, 1990; Hallenbeck, 2011). Because of this, it is desirable to produce hydrogen under non-nitrogen-fixing conditions, to improve yield and production rate. It is therefore necessary to provide the cells with a source of fixed nitrogen, however, nitrogenase expression is inhibited by fixed nitrogen (ammonia in particular). A source of fixed nitrogen that does not inhibit nitrogenase is therefore required. Glutamate (supplied as monosodium glutamate) works for this purpose (Jones and Monty, 1979), but is too expensive for use at industrial scale. The ammonia sensitivity of nitrogenase expression is under active study in *R. palustris* and closely related bacteria, and a number of mutant lines have been developed with reduced sensitivity (H. Fu *et al.*, 1990; J. Liang *et al.*, 1991; Michel-Reydellet and Kaminski, 1999; Masepohl and Drepper *et al.*, 2002; Drepper and Groß *et al.*, 2003; Dixon and Kahn, 2004; Oda and Samanta *et al.*, 2005; Adessi and McKinlay *et al.*, 2012; Heiniger *et al.*, 2012).

Although the hydrogenase enzyme is in theory capable of producing hydrogen more efficiently than nitrogenase, it has been found that the majority of the hydrogen produced by bacteria comes from nitrogenase (Vignais and Colbeau *et al.*, 1985; Fißler *et al.*, 1994; Vignais and Toussaint, 1994; Vignais and Colbeau, 2004). Unlike nitrogenase, the reaction catalysed by hydrogenase is bidirectional, and hydrogenase generally operates the reverse reaction, as an uptake hydrogenase that reclaims the hydrogen released as the byproduct of N_2 fixation by nitrogenase (Fißler *et al.*, 1994; Koku *et al.*, 2002; Rey and Oda *et al.*, 2006). Hydrogenase may thus be responsible for a decrease in hydrogen production, rather than an increase. The *R. palustris* strain CGA009 (the strain to be used in this study) has a mutation in the *hupV* gene, (a component in hydrogen sensing) resulting in a deficiency in hydrogen uptake (Rey and Oda *et al.*, 2006). This again increases its utility as a hydrogen producing strain.

1.6. Considerations for cultivation at industrial scale

There are number of challenges and considerations that will be important when designing a biotechnological system that can be scaled up for industrial applications. As mentioned previously, any genetic alterations to the organism used should be stable over time, and not require constant selective pressure using expensive inputs such as antibiotics. In fact, all inputs should be kept as minimal as possible, so as to reduce costs and resource usage. This means that the growth medium should be kept as minimal, simple, and cheap as possible. Other input costs may include energy for mechanical stirring, aeration, heating (or cooling), and lighting. Further considerations include the availability, composition, toxicity, homogeneity, and physical characteristics (such as viscosity) of feed substrates,

material transport costs, water availability, and residual waste disposal. Sterilising the feed substrates and growth medium at large scales is likely to be unfeasibly energy intensive and expensive, and therefore not practical. As such, any microorganisms used in the process, will need to be insensitive to, or able to out-compete, any microorganisms present in the feed substrates and growth medium. Any products produced by the system will need to be extracted, separated, collected, and possibly purified.

A system utilising *R. palustris* offers a number of attractive features that help to address some of the issues of industrial scale-up. Unlike many photosynthetic organisms, *R. palustris* can grow in anaerobic conditions, meaning that the need for artificial aeration using mechanical stirring or gas bubbling is greatly lessened (stirring may still be required to maintain homogeneity, to allow light penetration, and to prevent biofilm formation, depending on the system used), thus reducing energy costs. Furthermore, although growth may be anaerobic, *R. palustris* is not an obligate anaerobe, and can easily tolerate a micro-aerobic environment, resulting in no need for expensive and complex anaerobic growth and handling chambers and facilities. Additionally, *R. palustris* is robust, and tolerates toxic contaminants well, reducing the need to pre-treat or purify waste feed substrates. The ubiquity of *R. palustris* in a diverse range of environments is testament to its success, and suggests that competes well with other commonly found microorganisms. Where waste glycerol, or other toxic materials are used as substrates, it is likely that robust organisms such as *R. palustris* are likely to have a selective advantage, particularly if the growth medium/substrate is inoculated with a high cell-density, allowing the culture to establish itself quickly and out-compete any other microorganisms that are present. As a result, it should be possible to establish stable *R. palustris* cultures using these substrates, without the need for expensive and impractical substrate sterilisation. *R. palustris* is found in a diverse range of environments, and tolerates a wide range of temperatures and light conditions, again making it suitable for use in large-scale industrial systems, as few inputs will be required to bring the extant ambient conditions of any location into a range suitable for growth and production of *R. palustris*.

The principal output of interest produced by *R. palustris* is hydrogen gas, but other outputs are also possible, including electrical power generation *via* MFC or biophotovoltaic devices, and metabolic storage polymers such as PHB (polyhydroxybutyrate). Other products may be produced by engineering pathways for the production of compounds not already made by *R. palustris*, and pathway engineering may also be used to optimise and increase the yields of desired products. In the case of hydrogen gas, no extraction, separation, or purification is required, since the gas is naturally released from the culture (hydrogen has low solubility in water and partitions rapidly to the gas phase) and can be directly collected from gas vents or valves at the top of the reactor vessel without terminating the culture. Analysis has also shown that the hydrogen gas released by *R. palustris* is of very high purity (~97 mol% as reported by Pott *et al.* (2013)). For the production of compounds

and polymers, an extraction process will be needed, which will necessitate the termination of the culture once the substrate has been fully metabolised. Such an extraction process will consume energy and resources, but the cost of this (as well as the energy and costs of growing the culture initially) could potentially be offset by the collection of hydrogen or electricity while the culture is growing.

Despite the manifold benefits discussed above, probably the most notable advantage of *R. palustris* in a biotechnological context is its great metabolic diversity. The many metabolic pathways and modes of growth of *R. palustris* provide the ability to degrade a broad range of substrates and complex waste mixtures, and offer numerous possibilities for elucidating novel pathways and enzymatic reactions. Even if it is not ultimately selected for use in an industrial system, a study of *R. palustris* in industrial applications could lead to discoveries that could be transferred or applied to other organisms. To date, *R. palustris* has not been widely used at industrial scales, and many of the advantages discussed above are theoretical or speculative, and thus will require rigorous testing before industrial scale systems can be developed.

1.7. Objectives

The goals of this project are:

- To determine what waste materials might support the growth of *R. palustris*: Different sources of carbon and nitrogen will trialled in minimal medium, to determine waste profiles suitable for bio-remediation by *R. palustris*.

  This will be achieved using small-scale growth experiments where *R. palustris* is cultivated in small, lab-scale cultures using minimal medium supplemented with different sources of carbon and nitrogen. Cultures will be grown under optimal conditions for photofermentation, using continuous lighting to support maximal growth so as to rapidly detect substrates that support growth whilst also supporting the production of hydrogen gas.

- To determine if *R. palustris* can be engineered for improved growth or hydrogen production on waste substrates, in particular, glycerol: A stable heterologous expression system will be developed that allows the unmarked integration of trans-genes into the *R. palustris* genome. This expression system will be tested by expressing glycerol specific transporters with the goal of improving glycerol uptake by *R. palustris*. This will provide a proof-of-concept for the expression system, which can later be used to introduce other pathways to *R. palustris*, that have been identified as being of significant interest or benefit.

  The expression system will be designed to provide constitutive expression of transgenes under all growth conditions at a high level. Constitutive expression across all

growth modes makes the system more flexible for future applications, and a relatively high expression level should help to produce high yields and easily detectable phenotypes. A neutral site will be identified to allow the insertion of trans-genes to the *R. palustris* genome, without greatly perturbing its growth or other characteristics.

- To study the regulation of the metabolism of waste compounds: Where interesting compounds supporting the growth of *R. palustris* are identified, transcriptome sequencing will be used to investigate the metabolic pathways and how they are regulated. This data will be useful for the design of future heterologous expression systems and for metabolic pathway engineering.

  In particular, the regulation of pathways that produce desired outputs, such as the nitrogenase pathway for hydrogen production, and how these pathways are affected by the growth substrate, will be of especial interest. The regulation of pathways that metabolise the growth substrates (such as glycerol), and the expression of the genes under control of the promoter used in the heterologous expression system will also be of interest.

- To engineer *R. palustris* to produce new valuable products: The stable heterologous expression system developed earlier will be used to engineer *R. palustris* to produce a suitable high-value product determined by the composition of wastes identified as suitable for bio-remediation by *R. palustris*.

  This compound should be industrially valuable, and should be made to accumulate to a high level in *R. palustris* cultures. It should have a composition that allows it to be easily converted from the waste substrate, without a requirement for significant additional inputs. It will be extracted from terminal *R. palustris* cultures after the input substrate has been completely consumed. Initial experiments will be done at small-scales to determine if the compound can be produced and extracted, and then scaled up to larger pilot scales to assess the yield.

This will all need to be done whilst keeping industrial applications in mind; so the growth conditions used must be suitable for modification to scale-up, and genetic alterations will need to be stable, without requiring constant selection with antibiotics or other selection agents, which would be prohibitively expensive and inefficient at industrial scales.

Materials and Methods

2

2.1. Materials

2.1.1. Reagents

Except where noted otherwise, chemical reagents were obtained from Sigma-Aldrich.

2.1.2. Bacterial strains

Rhodopseudomonas palustris CGA009 (Larimer *et al.*, 2004) (referred to as *R. palustris* throughout) was used throughout as the wild-type strain, and the strain from which all modified strains were developed.

Escherichia coli DH5α (Taylor *et al.*, 1993) was used for cloning of plasmids and to provide genetic material for the experiments described in Chapter 4.

Anabaena varabilis sp. PCC 7120 was used to provide genetic material for the experiments described in Chapter 6.

2.1.3. Media

Except where otherwise noted, *R. palustris* was grown in a minimal salts medium (MinGU) defined as: 12.5 mM KH_2PO_4, 6.8 mM NaCl, 0.8 mM $MgSO_4$, 0.3 mM $CaCl_2$, 9.7 mM K_2HPO_4, 16.0 µM ferric citrate, 14 µM PABA (para-aminobenzoic acid (4-aminobenzoic acid)), 50 mM glycerol, 10 mM urea, and trace elements (0.51 µM $ZnCl_2$, 0.51 µM $MnCl_2$, 0.97 µM H_3BO_3, 0.84 µM $CoCl_2$, 0.12 µM $CuCl_2$, 0.08 µM $NiCl_2$, 0.17 µM $NaMoO_4$.

To avoid problems with precipitation, K_2HPO_4, ferric citrate, PABA, glycerol, and the trace elements, were autoclaved separately, and individually added to autoclaved base medium.

For growth on plates, 15 g $\cdot$ L^{-1} Bacto Agar (BD) was added to the base media prior to autoclaving.

Where used, antibiotics were used at the following concentrations: 100 µg $\cdot$ mL^{-1} kanamycin sulphate (Sigma Aldrich), and 25 µg $\cdot$ mL^{-1} G418 disulphate (Geneticin, Sigma Aldrich).

In some cases, 20 mM $NaNO_3$, 10 mM $[NH_4]_2SO_4$ or 20 mM monosodium glutamate (hereafter referred to as 'glutamate'), were substituted for urea.

2.2. Methods

2.2.1. Growth of *Rhodopseudomonas palustris*

Growth of *R. palustris* was assessed in liquid cultures by OD (optical density) measurements at 660 nm. Measurements were made using a Thermo Spectronic Helios spectrophotometer.

Measurements of pH were made using a Hanna Instruments pH 211 digital pH meter.

Whole-cell absorbance spectra were determined using a Shimadzu UV-1800 spectrophotometer at a sampling interval of 1 nm, across the range 400 nm to 1000 nm. The effects of light scattering were corrected for by placing a diffuser across the front of the cuvettes. Ten layers of Scotch Magic Tape were used for this purpose (Jackson *et al.*, 2014) (this method was later optimised and improved, and is to be published; see Appendix A for manuscript in preparation).

2.2.2. PCR

High fidelity PCR (polymerase chain reaction) (used for cloning) was performed using ThermoFisher HiFi Phusion DNA polymerase, according to the manufacturers directions. The 'High GC' buffer and DMSO additive were both used, at the recommended concentrations.

PCR reactions for genotyping and RT-PCR (reverse transcription polymerase chain reaction) was performed using Promega GoTaq Flexi G2 DNA polymerase, according to the manufacturers directions.

Primers were supplied by Sigma-Aldrich, and primer annealing temperatures were calculated using the online tools provided by the manufacturer of each polymerase.

PCR products were visualised by gel electrophoresis using 1 % (w/v) agarose gels buffered with TBE (Tris borate EDTA), or using Qiagen QIAxcel capillary gel electrophoresis.

2.2.3. Plasmid construction

Fragment sequences to be incorporated into plasmids were amplified by PCR, using primers and template sequences as described in the text. Plasmid assembly was performed using the NEBuilder HiFi DNA Assembly kit from New England Biolabs, as described in the kit protocol.

Escherichia coli DH5α was used to host generated plasmids and was grown in LB (lysogeny broth). Competent cells and heat-shock transformation were performed according

to the Inoue method (Inoue *et al.*, 1990).

2.2.4. Transformation of *Rhodopseudomonas palustris*

R. palustris was grown to mid exponential phase (OD_{660} 0.5 to 1.0) in 50 mL Falcon tubes in minimal medium (MinGU) with constant shaking and illumination. The cells were then centrifuged at 4800 RCF for 15 minutes at 4 °C. The cells were washed by resuspending in 50 mL cold 10 % (v/v) glycerol to remove salts, and then centrifuged again at 4800 RCF for 15 minutes at 4 °C. The wash step was repeated twice. The cells were finally resuspended in 5 mL 10 % (v/v) glycerol.

100 µL of cell suspension was added to an electroporation cuvette with 5 µL of plasmid. The cells were electroporated at 2 kV, 600 Ω, 25 µF, and the time curve recorded.

1 mL of YP (yeast, peptone) medium was added to cuvette, and then transferred to a 2 mL microcentrifuge tube, which was then topped up to 2 mL with YP. The tube was incubated at 30 °C for 3 hours to facilitate cell recovery.

For selection, 50 µL of the cell suspension was plated onto MinGU plates supplemented with kanamycin. These were incubated at room temperature, with direct exposure to sunlight, for 1 to 2 weeks. Individual colonies were picked and streaked onto MinGU plates supplemented with kanamycin and/or G418. The streaked colonies were allowed to grow for 1 to 2 weeks and then genotyped by PCR as described in the text.

Where a second crossover was required, successful primary crossover colonies were used to inoculate 2 mL YP medium and incubated overnight at 30 °C with illumination. 200 µL of cell suspension was plated on MinGU supplemented with 5 % (w/v) sucrose. The plates were incubated at room temperature, with direct exposure to sunlight, for 1 to 2 weeks. Colonies were picked and then patched onto both MinGU plates supplemented with kanamycin and MinGU plates supplemented with 5 % (w/v) sucrose. Colonies that were sensitive to kanamycin, but resistant to sucrose were genotyped by PCR as described in the text. Successful markerless colonies were streaked onto MinGU plates, without selection.

2.2.5. RNA extraction from *Rhodopseudomonas palustris*

R. palustris cultures were grown to mid-log phase (OD_{660} 0.6 to 0.8), and a 10 mL sample taken. The growth medium was removed by centrifugation, and the cell pellet snap-frozen in liquid nitrogen.

The cell pellet was resuspended in 1 mL Ambion TRIzol, and ~200 µL autoclaved and dried acid-washed glass beads (425 µm to 600 µm diameter) were added to each sample. The samples were then vortexed 5 × 30 seconds with 30 seconds cooling on ice in-between, before adding 200 µL chloroform.

The samples were vortexed briefly and then centrifuged for 2 minutes at 20 000 RCF and 4 °C. The upper phase was transferred to a new tube and a second extraction was

performed with 500 µL chloroform. The upper phase was separated to a new tube, and mixed with 500 µL ice-cold isopropanol. The sample was chilled at −20 °C for 20 minutes and then the precipitated RNA was pelleted by centrifugation for 15 minutes at 20 000 RCF and 4 °C.

The supernatant was removed, the pellet washed twice with ice-cold ethanol, carefully dried at room temperature (to evaporate residual ethanol), and finally resuspended in 100 µL RNase (ribonuclease)-free water.

A DNase (deoxyribonuclease) digestion, to remove genomic DNA was performed using a TurboDNA-free kit from Ambion Invitrogen, according to the manufacturers directions. The RNA was then re-purified using a Qiagen RNEasy kit, according to the manufacturers directions.

2.2.6. RT-PCR

cDNA was prepared from RNA extracts using a SuperScript IV VILO Master Mix with ezDNase Enzyme kit, according to the supplied directions.

PCR was then performed (using Promega GoTaq Flexi G2 DNA polymerase) using appropriate primers (as described in the text) and the prepared cDNA as template, at 5 % (v/v) of the total reaction volume.

2.2.7. Protein extraction and Western Blotting

50 mL *R. palustris* culture was centrifuged to remove supernatant and frozen overnight. TThe cells were resuspended in 1.8 mL Lysis Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 1 × Sigma*FAST*™ Protease Inhibitor Cocktail EDTA-free S8830. pH adjusted to 8.0). Glass beads were added before bead-beating for 5 minutes followed by cooling on ice. The suspension was then transferred to 50 mL Falcon tubes, topped up to a total volume of 20 mL with cold Lysis Buffer, before sonicating for 4 × 30 seconds with 30 seconds cooling on ice in-between.

1.8 mL of the cell suspension was placed in 2 mL screw cap tubes with glass beads and shaken on a bead-beater for 5 minutes. 50 µL of the cell lysate was mixed 1:1 with 50 µL 5 × SDS-PAGE Laemmli sample loading buffer (Laemmli, 1970). The samples were boiled for 10 minutes, and then 15 µL loaded on BioRad TGX 4 % to 15 % (w/v) Mini-Protean gels. 15 µL of Thermo Scientific PageRuler Prestained marker ladder was used in a separate lane on each gel. Gels were run at room temperature for 10 minutes at 100 V followed by 20 minutes at 200 V.

When needed, gels were stained for 1 hour with Expedeon InstantBlue™ coomassie stain, destained overnight with water, and preserved in destain solution (20 % (v/v) methanol, 10 % (v/v) glacial acetic acid, in water).

For Western blotting, gels (unstained) were transferred to Invitrogen iBlot™ Nitrocellulose membrane kits and proteins dry-transferred using the Invitrogen iBlot™ device, Program 3, 7 minutes. Gels were stained as above post-transfer, to check for protein transfer.

Membranes were blocked overnight at 4 °C in 5 % (w/v) skim-milk powder dissolved in PBST buffer (13.7 mM NaCl, 0.27 mM KCl, 1 mM Na_2HPO_4, 0.2 mM KH_2HPO_4, 0.2 % (v/v) Tween 20, adjusted to pH 7.2 with 1 M NaOH). Primary antibody (Abcam Anti-Myc tag antibody ab9106) was added to drained membranes at 1:2000 in PBST with 1 % (w/v) skim-milk powder, and incubated at room temperature for 1 hour. Membranes were washed 5 times for 5 minutes each with cold PBST. Secondary antibody (GE Healthcare Amersham™ ECL Anti-rabbit IgG Horseradish Peroxidase-Linked Species-Specific Whole Antibody from donkey NA934) was added to drained membranes at 1:2000 in PBST with 1 % (w/v) skim-milk powder, and incubated at room temperature for 1 hour. Membranes were washed again 5 times for 5 minutes each with cold PBST.

Membranes were visualised by exposing to 1 mL of GE Healthcare Amersham™ ECL Western Blotting Detection Reagents for 1 minute. Excess reagent was blotted away, and the membranes visualised in Syngene G:Box, exposed at 5 × 3 minute (cumulative) intervals on the chemiluminescent setting. An overlay with the prestained protein marker was created by capturing a photograph of the membrane using the Syngene G:Box with the door open (under white light), and displaying this with the chemiluminescent image.

2.2.8. Glycerol quantification

Spectrophotometric determination of glycerol

A colorimetric glycerol assay, based on the Malaprade reaction.

Reagent A: 1.6 M acetic acid

Reagent B: 4 M ammonium acetate

Reagent C: 4 M acetylacetone solution, prepared by mixing 200 µL acetylacetone with 5 mL Reagent A

Reagent D: 4 mM sodium periodate solution, prepared by mixing 21 mg sodium meta periodate 5 mL Reagent B

Samples of cell cultures were taken and centrifuged to pellet the cells. 2 mL of the supernatant was mixed with 1.2 mL Reagent D and briefly vortexed. Next 1.2 mL Reagent C was added, the sample briefly vortexed, and then incubated at 70 °C for 1 minute before immediately being cooled in an ice-water bath. The absorbance was measured at 410 nm and the glycerol concentration determined by comparison to a standard curve.

GCMS determination of glycerol

1 mL samples of cell cultures were taken and centrifuged to pellet the cells. 10 µL of the supernatant was mixed with 10 µL of a 50 mM 2,3-butanediol internal standard solution

and evaporated to dryness in a rotary concentrator.

50 µL of 20 mg · mL^{-1} methoxyamine hydrochloride in (anhydrous) pyridine was added to each sample, the tubes were sealed and then incubated at 70 °C for 1 hour. 50 µL of MSTFA (N-Methyl-N-(trimethylsilyl) trifluoroacetamide) was added to each sample, and incubated at 70 °C for 30 minutes.

Samples were injected using split injection mode (split ratio 1:10) onto a 30 m SGE BPX5 (0.25 mm internal diameter, 0.25 µm film thickness) 5 % phenyl polysilphenylene-siloxane column installed in a Shimadzu QP2010-Ultra single quadrupole GCMS. The injection temperature was set at 230 °C, transfer line at 290 °C, and the ion-source at 250 °C. The initial column oven temperature was 70 °C and then increased to 300 °C at a rate of 10 °C per minute, and held at a final temperature of 300 °C for 10 minutes.

Mass spectra acquisition was performed in scan mode between 50 Da and 500 Da. Glycerol was quantified using Shimadzu GCMSsolution software by the internal standard addition method and a glycerol standard curve.

2.2.9. Cyanophycin purification and characterisation

Small-scale cyanophycin extraction

50 mL *R. palustris* culture was centrifuged to remove supernatant and frozen overnight. The cells were lysed as above, 10 mL of the cell suspension was taken, Triton X-100 was added to a final concentration of 0.1 % (v/v), and the cell suspension mixed overnight at 4 °C on a rotary shaker.

The cell suspension was centrifuged for 30 minutes at 30 000 RCF and the supernatant discarded. The pellet was washed 3 times with distilled water, and centrifuged in-between each wash as above. The pellet was then resuspended in 100 mL 0.1 M HCl, mixed thoroughly, and then centrifuged for 30 minutes at 30 000 RCF, retaining the supernatant. The supernatant was then neutralised using 1 M NaOH, chilled to 4 °C, and then centrifuged for 30 minutes at 30 000 RCF to collect the precipitated cyanophycin. The pellet was then washed with distilled water. The final cyanophycin pellet was resuspended in 2 mL 0.1 M HCl and stored at −20 °C.

Large-scale cyanophycin extraction

5 L *R. palustris* culture was centrifuged to remove supernatant, frozen overnight, and lyophilised to dryness in a freeze-drier. The CDW (cell dry weight) was measured gravimetrically, and then the cells were resuspended in 200 mL water and sonicated for 10 one minute intervals, with one minute cooling on ice in between. Triton X-100 was added to a final concentration of 0.1 % (v/v) and the cell suspension mixed overnight at 4 °C on a rotary shaker.

The cell suspension was centrifuged for 30 minutes at 30 000 RCF and the supernatant discarded. The pellet was washed 5 times with distilled water, and centrifuged in-between each wash as above. The pellet was then resuspended in 100 mL 0.1 M HCl, mixed thoroughly, and then centrifuged for 30 minutes at 30 000 RCF, retaining the supernatant. The supernatant was then neutralised using 10 M NaOH, chilled to 4 °C, and then centrifuged for 30 minutes at 30 000 RCF to collect the precipitated cyanophycin. The pellet was then washed with distilled water. The cyanophycin pellet was the re-extracted, to further purify, using 20 mL 0.1 M HCl, washed, neutralised, and precipitated as above. The final cyanophycin pellet was frozen, lyophilised, and then weighed, to calculate cyanophycin as a percentage of CDW.

Cyanophycin measurement by Sakaguchi assay

A modified Sakaguchi reaction, as described by Allen (1984), was used to determine the non-hydrolysed arginine content of cyanophycin extracts.

Reagent A: 30 mg KI in 100 mL H_2O.

Reagent B: To 100 mL 5 M KOH, 2 g of potassium sodium tartrate ($4H_2O$) is added, followed by 100 mg of 2,4-dichloro–1-naphthol. Then 180 mL of absolute ethanol is added followed by x mL NaOCl (4 % to 6 % (v/v)), made up to 50 mL with H_2O.

Reagent C: $2x$ mL of commercial NaOCl, which had an approximate concentration of 4 % to 6 % (v/v) NaOCl, was made up to 50 mL with H_2O.

To determine the volume of x:

- Make up Reagent B, but without potassium sodium tartrate or NaOCl
- Make a dilution series of NaOCl in H_2O; 0 µL to 100 µL NaOCl to a total volume of 1 mL
- Add 3 mL Reagent B, but without potassium sodium tartrate or NaOCl
- Mix and then let stand for 30 minutes
- Determine the A_{400}; whichever concentration gives highest absorbance is best
- The volume of NaOCl giving the highest absorbance is the amount needed for 3 mL Reagent B
- x is the volume needed for 280 mL Reagent B

Protocol:

- The samples were made up to 0.5 mL with 0.1 M HCl
- A standard curve of 0 µg · mL^{-1} to 1000 µg · mL^{-1} arginine in 0.5 mL with 0.1 M HCl was prepared

- A blank of 0.5 mL with 0.1 M HCl was prepared
- To each of the samples/standards/blank 0.5 mL Reagent A was added
- To each of the samples/standards/blank 1.5 mL mL Reagent B was added
- Samples were incubated for 1 hour at room temperature
- To each of the samples/standards/blank 0.5 mL Reagent C was added
- Samples were incubated for 10 minutes and then the absorbance at 520 nm recorded

Cyanophycin measurement by ion-exchange chromatography

Amino acid hydrolysis and ion-exchange chromatography was performed by Dr Peter Sharratt from the University of Cambridge Department of Biochemistry Centre for Proteomics.

Gas phase hydrolysis was used to hydrolyse the samples so that the free amino acids could be quantified. The samples were spiked with a known amount of an internal standard (norleucine) in 0.1 M HCl, and evaporated to dryness in a centrifugal evaporator. The dried residue was placed in a hydrolysis vial (Waters, WAT007363) to which was added hydrochloric acid (BDH, Aristar diluted 1:1 with water; 0.5ml) containing phenol (saturated solution in water; 7.5ul) followed by dodecanethiol (Sigma; 68ul). The vial was evacuated and then flushed with argon and evacuated several times. After a final evacuation it was closed off and placed in an oven at 115 degrees C for 22h. The vial was then removed from the oven, allowed to cool, and opened. The tube was then placed in a dessicator under vacuum over sodium hydroxide for 40 minutes.

After the residual acid was removed, the sample was resuspended in sodium citrate loading buffer, pH 2.2, and filtered through a 0.2 µm filter. An aliquot of the filtrate was loaded on a Biochrom 30 instrument, and chromatography performed with ion-exchange resin, eluting with a series of buffers over a pH range of 3.2 to 6.45. Peak detection was achieved by post-column mixing of the eluate with ninhydrin at 135 °C and measuring the absorbance at 570 nm and 440 nm. Quantification was performed using Chromeleon software and pre-prepared standard curves of each amino acid, using the internal standard method.

2.2.10. Electron microscopy

Fixing, sectioning, and preparation of sample grids for electron microscopy was performed by Dr Karin Muller at the Cambridge Advanced Imaging Centre.

Cells were harvested from 1 week old culture grown in minimal medium (MinGU) in 50 mL Falcon tubes under illumination. 1 mL of culture was centrifuged for 5 minutes at 5000 RCF and washed twice with cold 0.4 % (w/v) NaCl. The supernatant was removed and the cells were then fixed for 1 hour at 4 °C in 0.4 % (v/v) glutaraldehyde, 0.05 M sodium

cacodylate buffer at pH 7.4. The cells were then centrifuged again and the buffer exchanged for 2 % (v/v) glutaraldehyde, 0.05 M sodium cacodylate, 2 mM $CaCl_2$ buffer at pH 7.4 for 2 hours at room temperature on a rotary mixer.

The cells were washed 5 times in deionised water and then osmicated at 4 °C overnight in 1 % (w/v) osmium tetroxide, 1.5 potassium ferricyanide. Following this, they were washed again 5 times in deionised water and then treated with 0.1 % (w/v) thiocarbohydrazide for 20 minutes in the dark, at room temperature. Next, they were washed again 5 times in deionised water and then treated with 2 % (w/v) osmium tetroxide for 1 hour at room temperature, washed 5 times in deionised water, and treated with 2 % (w/v) uranyl acetate, 0.05 % (w/v) maleate buffer at pH 5.2 overnight at 4 °C.

The cells were washed 5 times in deionised water and then dehydrated using an ethanol series consisting of 3 washes of 5 minutes each in each of 50 %, 70 %, 95 % and 100 % (v/v) ethanol, followed by 100 % (v/v) anhydrous acetone and then 100 % (v/v) anhydrous acetonitrile. They were then put into a mixture of 1:1 (volume:volume) Quetol resin (without BDMA) : anhydrous acetonitrile and incubated overnight with vial caps opened, before being mixed with 100 % (v/v) Quetol resin (without BDMA) and incubated over the weekend on a rotary mixer.

For the next 5 days the cells were exchanged into fresh 100 % (v/v) Quetol resin (with BDMA) each day and incubated on the rotary mixer. On the final day, the cells were centrifuged to produce a compact pellet, the supernatant removed, and the pellet cured in an oven at 60 °C over the weekend.

Thin sections (60 nm to 90 nm) were cut using Leica Ultracut UCT ultramicrotome, collected on bare 300 mesh copper grids and were not post-stained. Grids were viewed using an FEI Tecnai G2 in bright field mode operated at 200 keV using a 20 µm objective aperture to improve contrast.

> Quetol resin mix: 12 g Quetol 651, 15.7 g NSA (nonenyl succinic anhydride), 5.7 g MNA (methyl-5-norbornene-2,3-dicarbocylic anhydride), 0.5 g BDMA (benzyldimethylamine). All resin components were from TAAB.

2.3. Analytical

2.3.1. Statistical analysis

Data analysis was performed using the statistical software R, using the 'tidyverse' packages (Hadley, 2016). One-way ANOVA (analysis of variance) with Tukey's HSD (honest significant difference) post-hoc test (Tukey, 1949) was used to determine significant differences between groups, and Dunnett's post-hoc test (Dunnett, 1955; 1964) was used to determine significant differences from a control group.

Growth on waste materials

3

3.1. Introduction

Rhodopseudomonas palustris, as described by Molisch (1907) and Niel (1944), has been isolated from a diverse range of different environments including; various soils (Harada *et al.*, 2003; Madigan and Martinko *et al.*, 2008; Wong *et al.*, 2014), animal waste ponds (M. K. Kim *et al.*, 2004; Sharma *et al.*, 2015), activated sludge from waste-water treatment facilities (Hiraishi and Kitamura, 1984), aquatic sediments (Oda and Wanders *et al.*, 2002) and, alkali lake-water (Dönmez *et al.*, 1999). Concomitant with this range of environments, *R. palustris* is known to metabolise a diverse range of different compounds, including acetate and butyrate (C.-Y. Chen *et al.*, 2008), malate (Fißler *et al.*, 1994), formate (Yoch and Lindstrom, 1967), lactate, pyruvate, fumarate, succinate, and benzoate (Niel, 1944; Rolls and Lindstrom, 1967), glycerol (Sabourin-Provost and Hallenbeck, 2009; D. Ghosh *et al.*, 2012; Pott *et al.*, 2013), and a range of aromatic organic acids (Hegeman, 1967; Harwood and J. Gibson, 1988; Fißler *et al.*, 1994; Perrotta and Harwood, 1994; Woude *et al.*, 1994; Küver *et al.*, 1995; Oda and De Vries *et al.*, 2001; Oda and Wanders *et al.*, 2002; Austin *et al.*, 2015).

This metabolic diversity makes *R. palustris* an interesting organism to study from the point-of-view of waste bio-remediation. *R. palustris* could be grown on waste materials, converting unwanted, or even toxic, compounds into useful, or harmless, materials. Since *R. palustris* produces hydrogen gas as by-product of photofermentation (Uemura *et al.*, 1961; Knobloch *et al.*, 1971; Fißler *et al.*, 1994; Pott *et al.*, 2013; 2014), and may also be grown in a MFC to produce electricity (Xing *et al.*, 2008; Venkidusamy and Megharaj, 2016), it could be used to convert waste into alternative fuels and energy. Additionally, *R. palustris* could be engineered to express other metabolic pathways, to convert other compounds found in waste materials to other products (see Chapter 6).

To identify different waste substrates that *R. palustris* might be able to degrade, as a source of organic carbon, simple molecules, not previously tested in *R. palustris*, were tested to help to elucidate what kinds of compounds *R. palustris* could metabolise. This would provide insight into which compounds in more complex waste streams might be degradable and suggest profiles for potential waste streams to try.

3.2. Experimental

3.2.1. Growth with different nitrogen sources

In order to test different carbon sources as substrates for the growth of *R. palustris*, it was necessary to exclude any other sources of organic carbon in the growth media, to ensure that any growth was due to the carbon source under test. Previously *R. palustris* had been grown with glutamate as the nitrogen source, however, since glutamate is also a source of organic carbon, it was necessary to replace it with an alternative nitrogen source. Although *R. palustris* is capable of nitrogen fixation, for these experiments it was grown under anaerobic conditions (some air was present, but only as minimal head-space in the sealed growth tubes) to maximise growth and hydrogen production, and atmospheric nitrogen (N_2) is therefore limited. Sodium nitrate ($NaNO_3$), ammonium sulphate ($[NH_4]_2SO_4$) and urea were therefore trialled as possible alternatives to glutamate. Urea does contain fixed carbon, which could act as a contaminating carbon source, but it is processed by the enzyme urease upon entry to the cell, and the carbon is released as inorganic carbon dioxide (CO_2) (Blakeley and Zerner, 1984), thus making it still a viable alternative nitrogen source to glutamate.

In these experiments *R. palustris* was grown in 50 mL transparent, screw cap, plastic tubes with minimal head-space (air) in a New Brunswick, shaking incubator at 30 °C, illuminated by Sylvania GRO-Lux F15W/GRO - T8 photosynthesis rated fluorescent light tubes. The cultures were inoculated from liquid wild-type cultures that were spun down and washed twice with minimal medium to remove any contaminating substrates, glycerol, or toxic by-products. The concentrations of each nitrogen source were adjusted to give the same stoichiometric ratios of nitrogen atoms in each compound; 10 mM urea, 20 mM $NaNO_3$, 10 mM $[NH_4]_2SO_4$, and 20 mM monosodium glutamate.

Photoautotrophy

Since urea releases CO_2 when metabolised by urease, it was necessary to determine whether the CO_2 released contributes to the growth of the cell *via* one of the carbon-fixation pathways. In order to test the hypothesis that urea was not providing carbon to support the growth of *R. palustris*, a series of photoautotrophic growth experiments were carried out. *R. palustris* was grown in minimal medium with either urea or $NaHCO_3$ as an inorganic carbon source (both are metabolised to release CO_2). Whilst urea is utilised as both a source of carbon and nitrogen, $NaHCO_3$ needed to be supplemented with an inorganic nitrogen source; in this case $[NH_4]_2SO_4$ was used. Growth with these two different inorganic carbon sources was compared to growth on the same carbon sources supplemented with sodium thiosulphate ($Na_2S_2O_3$); an electron donor, that provides reducing equivalents to the photosynthetic electron transport chain. A culture with $[NH_4]_2SO_4$ as nitrogen

source, but no carbon source was used as a negative control, and a culture supplemented with glycerol and urea was used as a photoheterotrophic, positive control.

The growth experiments were carried out using a custom-built light source that provided light at the following wavelengths: 390 nm, 505 nm, 590 nm, 660 nm, 830 nm and 860 nm. Each light source was calibrated to provide 70 µmol · photons · m^{-2} · s^{-1} illumination. These lights were arranged in a hexagonal pattern as shown in Figure 3.1. This lighting system was used, as the LH2 (peripheral light-harvesting complex) in *R. palustris* has characteristic light absorption peaks (Brotosudarmo *et al.*, 2011) that may not be covered by the standard photosynthesis rated lighting provided in the incubator used in the earlier experiments, specifically in the infra-red spectrum. Figure 3.2 shows the emission spectra of the standard incubator lights compared to the lights in the custom-built system described here. The whole-cell absorbance spectra of *R. palustris* is plotted as an overlay, to show how the characteristic light absorption peaks align with the light emission spectra. Data for a third light is also displayed, as this was used in later experiments (section 3.2.2).

Cultures were grown in 1 L borosilicate glass bottles that covered the hexagonal light panels of the custom-built system, and were stirred by magnetic stir bars. The temperature of the cultures using the custom light source was not directly regulated, but the temperature of the room they were placed in was recorded over a 24 h period and remained within 24 °C to 27 °C. Either 10 mM urea or 10 mM $NaHCO_3$ were provided as an inorganic carbon source (urea is organic, but releases its carbon as CO_2 when metabolised by urease). Where used, sodium thiosulphate was provided as an artificial electron donor, at a concentration of 1.2 mM.

3.2.2. Growth with different carbon sources

In order to grow *R. palustris* with many different carbon sources, with sufficient replicates; these growth experiments were carried out in 96-well (300 µL per well) microplates on an orbital shaker. Growth was measured using a plate reader in absorbance mode. 3 µL of 5 g · L^{-1} substrate was added to each well, which were then inoculated with 300 µL of *R. palustris* culture with a starting OD at 660 nm of 0.2.

Where glycerol was added as a carbon source, it was used at a final concentration of 50 mM. Other carbon sources were added to a final concentration of 5 g · L^{-1}. This concentration was chosen as it is similar in solute concentration to a 50 mM glycerol solution. A concentration in g · L^{-1} was used instead of a molar concentration, as the molecular weight of some of the carbon sources is not known or was not easily measurable.

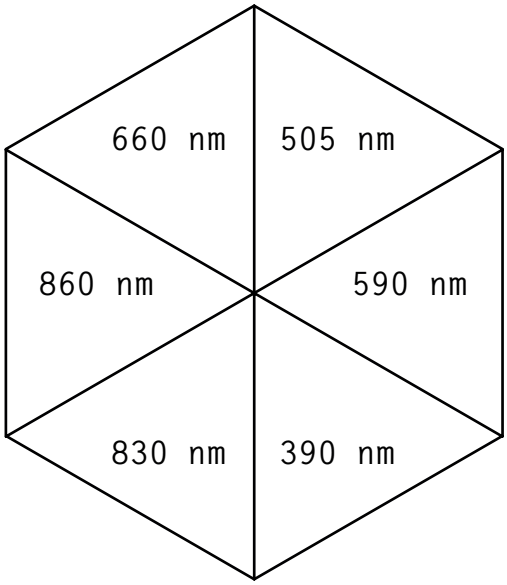


Figure 3.1. – Custom-built light source. This shows a birds-eye view of the arrangement of the different wavelengths of light used to illuminate 1 L bottles of *R. palustris* cultures. This apparatus was constructed by Dr. Paolo Bombelli and Dr. Ruth Laing, and consisted of a PCB (printed circuit board) with six hexagonal lighting areas. Each segment of each hexagonal lighting area was fitted with 12 light emitting diodes (LED) of the wavelengths shown in the diagram. The PCB was protected by transparent poly-carbonate covers, with holes cut over each LED to ensure that the poly-carbonate did not filter any of the light wavelengths. The LEDs were connected to an external adjustable power supply. The whole apparatus was placed over a six-position magnetic stir-plate, so that each lighting area was aligned with a stir-plate position underneath. 1 L bottles of culture are placed on top of each lighting area, so that they are illuminated from beneath, and stirred by the magnetic stir-plate underneath the lighting panel.

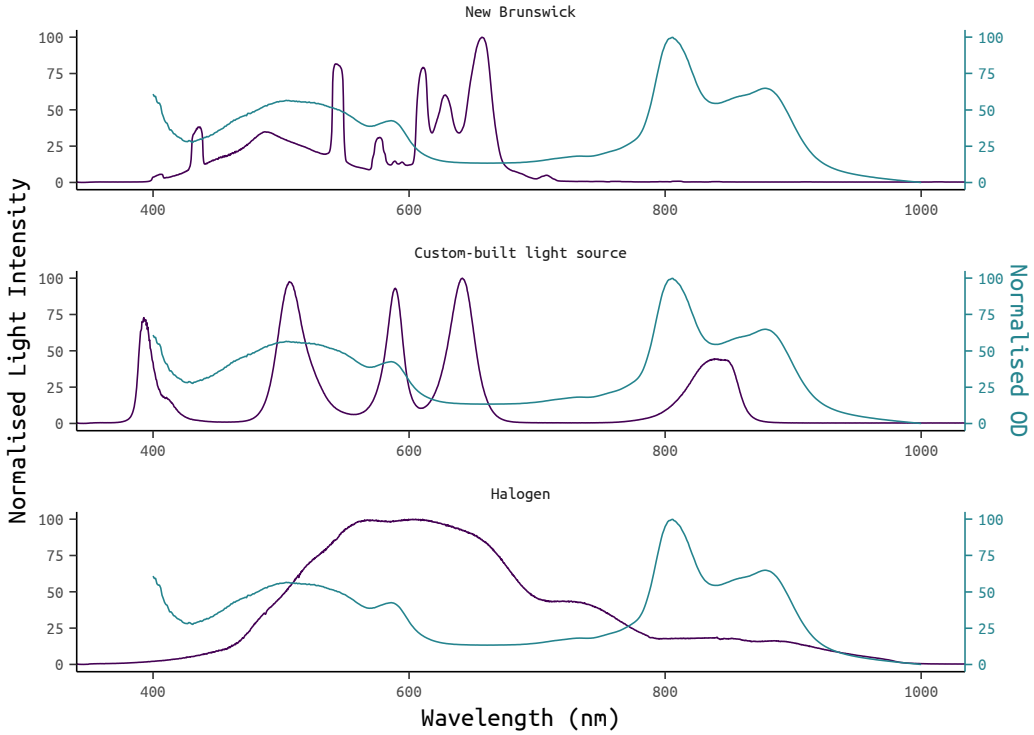


Figure 3.2. – Emission spectra of different light sources. This shows the emission spectra of the standard incubator lights (New Brunswick fitted with Sylvania GRO-Lux F15W/GRO - T8 photosynthesis rated fluorescent light tubes) compared to the lights in the custom-built light system and to standard halogen bulbs. The whole-cell absorbance spectra of *R. palustris* is plotted as an overlay, to show how the characteristic light absorption peaks align with the light emission spectra.

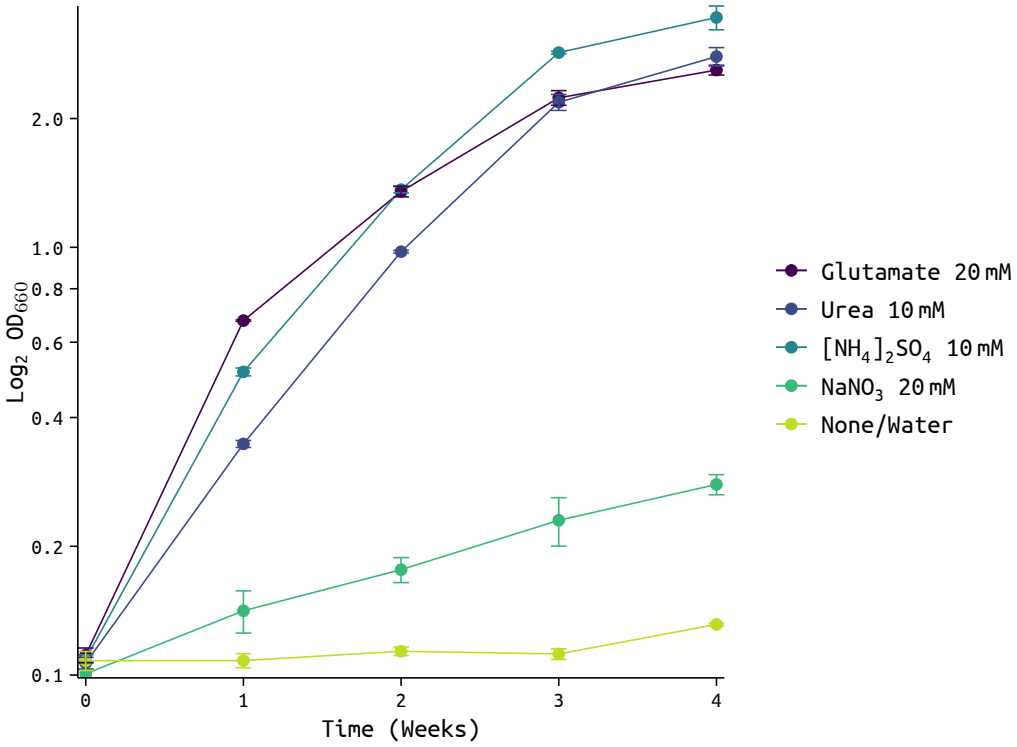


Figure 3.3. – Growth of *R. palustris* with different nitrogen sources. This shows the photoheterotrophic growth of *R. palustris* in minimal medium supplemented with glycerol and illuminated, when different nitrogen sources were added. Error bars show the SEM (standard error of the mean) of three biological replicates. Water was the no-nitrogen, negative control.

3.3. Results

3.3.1. Growth with different nitrogen sources

Figure 3.3 shows the results of an experiment to test the growth of *R. palustris* using different sources of nitrogen, when grown photoheterotrophically with glycerol and illumination in minimal medium. The growth of *R. palustris* when using either $[NH_4]_2SO_4$ or urea as the nitrogen source was similar to growth using glutamate. Growth using $NaNO_3$ was only marginally faster than the no-nitrogen negative control. This indicates that both $[NH_4]_2SO_4$ and urea are viable alternatives to glutamate as nitrogen sources but that $NaNO_3$ would be unsuitable. Possibly the poor growth on $NaNO_3$ was due to limited uptake to the cell, or inefficient conversion to usable nitrogen compounds, however, this remains speculative.

Since the disassociation of $[NH_4]_2SO_4$ in water leads to the formation of the free-proton generating ammonium ion, the pH of the medium containing this nitrogen source was measured, to ensure that the addition of $[NH_4]_2SO_4$ did not overcome its buffering capability. *R. palustris* is grown at a biological pH of 6.7. The pH of the non-inoculated medium did not change upon addition of $[NH_4]_2SO_4$. To determine if there were any changes in

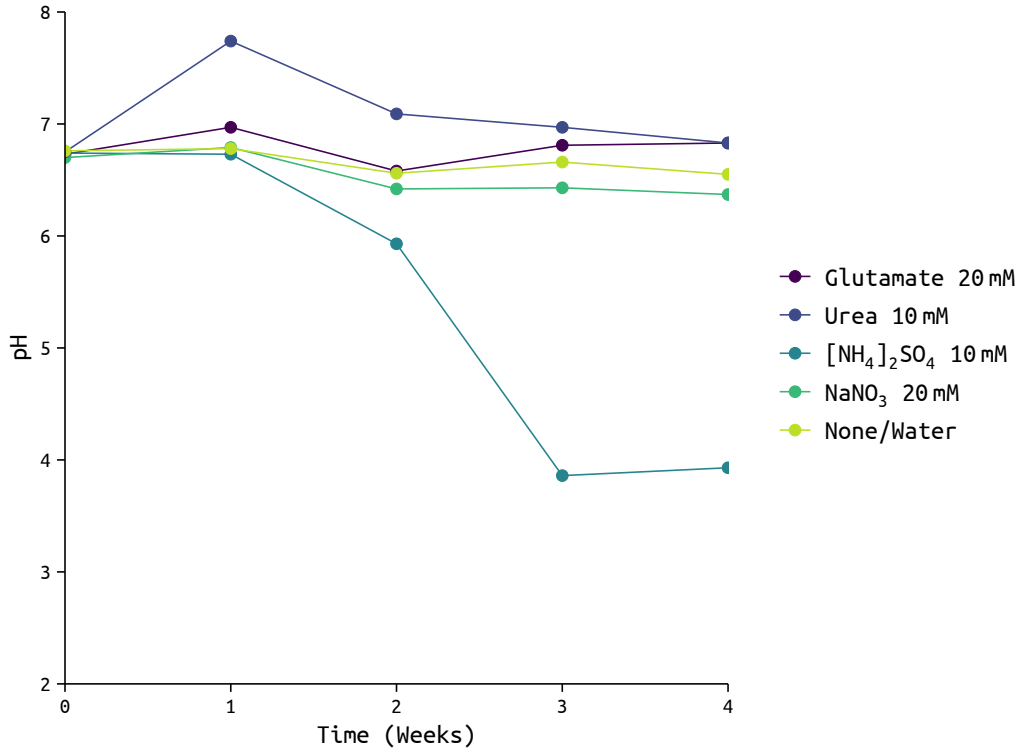


Figure 3.4. – Culture pH of *R. palustris* grown with different sources of nitrogen. This shows the change in culture pH when *R. palustris* was grown in minimal medium supplemented with glycerol and illuminated, when different nitrogen sources were added. Water was the no-nitrogen, negative control. Due to volume limitations the pH of three pooled biological replicates was measured.

pH with the organism present, the pH of *R. palustris* cultures was measured over time, with each of the different nitrogen sources. Figure 3.4 shows the pH of the cultures grown with the different nitrogen sources. Although the $[NH_4]_2SO_4$ did not immediately have any affect on the pH of the medium, the pH of the culture when grown with this nitrogen source dropped dramatically after *circa* three weeks. Initially it was thought that this could have been due to ammonia gas evaporating from the culture over time, however, when the pH of a sterile sample of the medium was tested after three weeks and six weeks, no drop in pH was observed. This suggested that the drop in pH is related to the metabolism of *R. palustris* when grown in the presence of $[NH_4]_2SO_4$.

Along with the change in pH, a distinct change in the colour of the culture was observed when it was grown with $[NH_4]_2SO_4$, compared to cultures grown with the other nitrogen sources. An example of this is shown in Figure 3.5, which shows the sample grown with $[NH_4]_2SO_4$ having a paler, pinker colour than the samples grown with glutamate, which are a darker purple colour.

Whole-cell absorption spectroscopy was carried out to characterise this colour difference and determine what might be causing it (Figure 3.6). The cultures in Figure 3.5 were photographed after five weeks of growth, but unfortunately the whole-cell spectra (Figure

Figure 3.5. – Colour change of *R. palustris* grown with different sources of nitrogen. This shows the change in colour observed when *R. palustris* was grown in minimal medium supplemented with glycerol and illuminated, when different nitrogen sources were added. The nitrogen sources of the samples from left to right: H_2O (no-nitrogen, negative control), $NaNO_3$, glutamate, $[NH_4]_2SO_4$, urea.

3.6) were not measured for a further two weeks after this (seven weeks of growth, total), and the $[NH_4]_2SO_4$ sample appeared to have died, as evidenced by it having turned grey. This may explain why the spectra of the samples grown with $[NH_4]_2SO_4$ matched the absorbance peaks of the no-nitrogen, negative, H_2O control so closely. The samples grown with urea, glutamate and $NaNO_3$ gave very similar spectra, except that the sample grown with glutamate had slightly higher carotenoid and chlorophyll *a* peaks around 420 nm. The decrease in pH may have caused the change in colour of the sample grown with $[NH_4]_2SO_4$, by altering the pigment profile. *R. palustris* has been shown to express different a LH2, containing different ratios of bacteriochlorophyll and carotenoids and different absorption spectra, when grown grown under different conditions, such as varying light intensity (Evans *et al.*, 1990; Brotosudarmo *et al.*, 2011). Possibly, the change in pH could also have caused a change in the LH2 expressed, explaining the change in colour observed. Repeating the whole-cell absorption spectra, or isolating and characterising the LH2 complexes might help to clarify this.

Photoautotrophic growth

Growth of *R. palustris* was observed with both urea, and $NaHCO_3$ as inorganic carbon sources, but only when thiosulphate was present as an artificial electron donor (Figure 3.7). This suggests that *R. palustris* is capable of photoautotrophic growth, and that the CO_2 released by urea catabolism can be re-fixed to provide organic carbon to the cell, provided that an electron donor (sodium thiosulphate in this case) is added to provide reducing equivalents to the photosynthetic electron transport chain. The cultures did not reach a

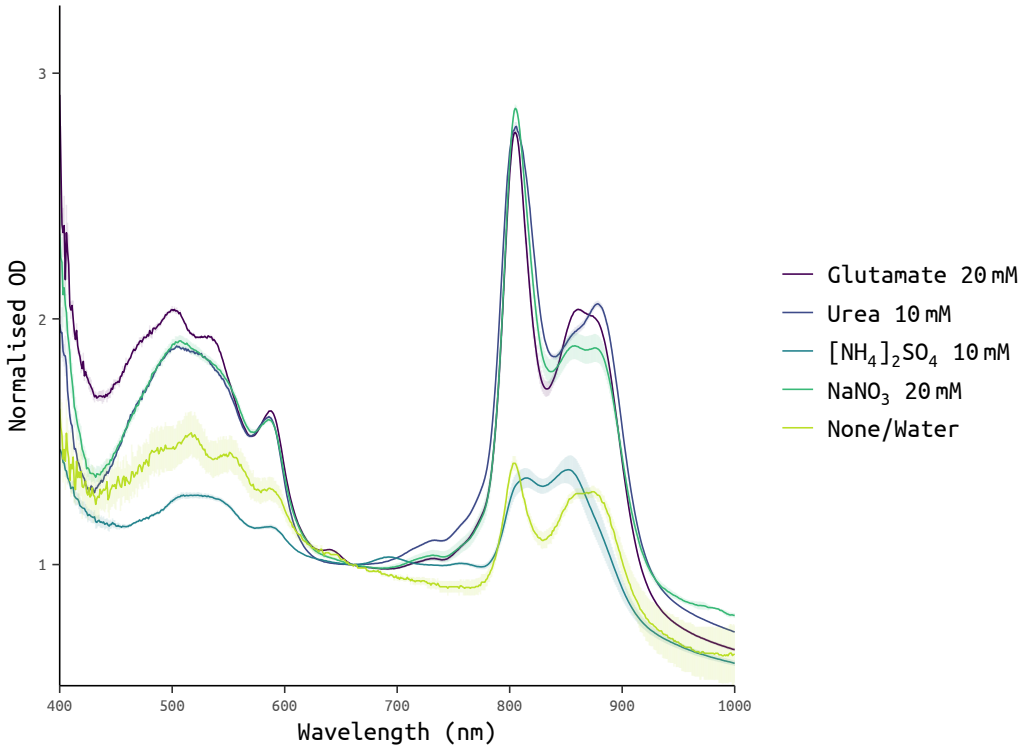


Figure 3.6. – Whole cell absorption spectra of *R. palustris* grown with different sources of nitrogen. This shows the change in absorption spectrum when *R. palustris* was grown in minimal medium supplemented with glycerol and illuminated, when different nitrogen sources were added. These spectra are the mean of three biological replicates, normalised to the absorbance at 660 nm. Water was the no-nitrogen, negative control. The thickness of the lines indicates the SEM.

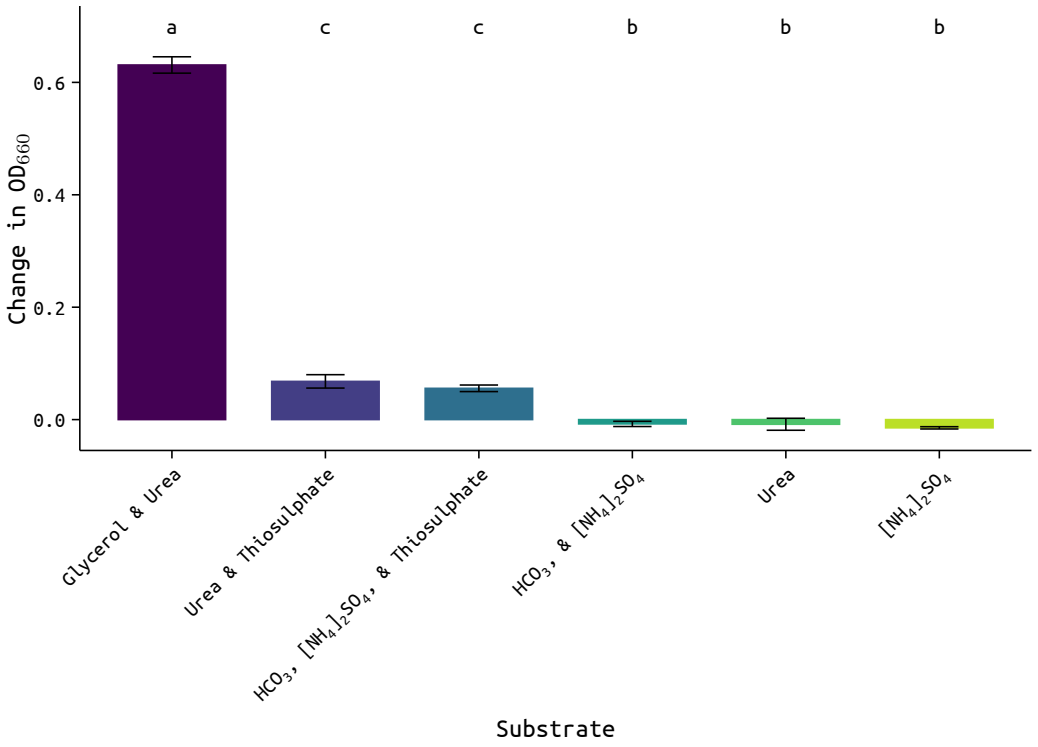


Figure 3.7. – Photoautotrophic growth of *R. palustris*. This shows the growth of *R. palustris* when grown under illumination in minimal media supplemented with an inorganic carbon source; urea or sodium bicarbonate ($NaHCO_3$), and the presence/absence of an electron donor, sodium thiosulphate ($Na_2S_2O_3$). Urea supplies both carbon and nitrogen, bicarbonate is supplemented with ammonia ($[NH_4]_2SO_4$) to provide nitrogen. A culture with only ammonia (to provide nitrogen) but no carbon source acts as the negative control, while the culture supplemented with glycerol and urea acts as the positive control. Cultures were measured after 10 days growth. Error bars show the SEM of three biological replicates. A one-way ANOVA was carried out with a Tukey HSD post-hoc test, with the significance level set to $p < 0.05$. The post-hoc test groupings are annotated above each column in the graph.

high density, compared to the glycerol supplemented positive control, possibly indicating that photoautotrophic growth was significantly slower than photoheterotrophic growth.

These results indicate that whilst urea did provide carbon to the cell, it did not support rapid growth, and growth only occurred when the requirement for an electron donor was met.

3.3.2. Growth with different carbon sources

To identify different waste substrates that *R. palustris* might be able to degrade, as a source of organic carbon, simple molecules were tested to help to elucidate what kinds of compounds *R. palustris* could metabolise. This would provide insight into which compounds in more complex waste streams might be degradable and suggest profiles for potential waste streams to try. *R. palustris* was grown in minimal medium with urea as the ni-

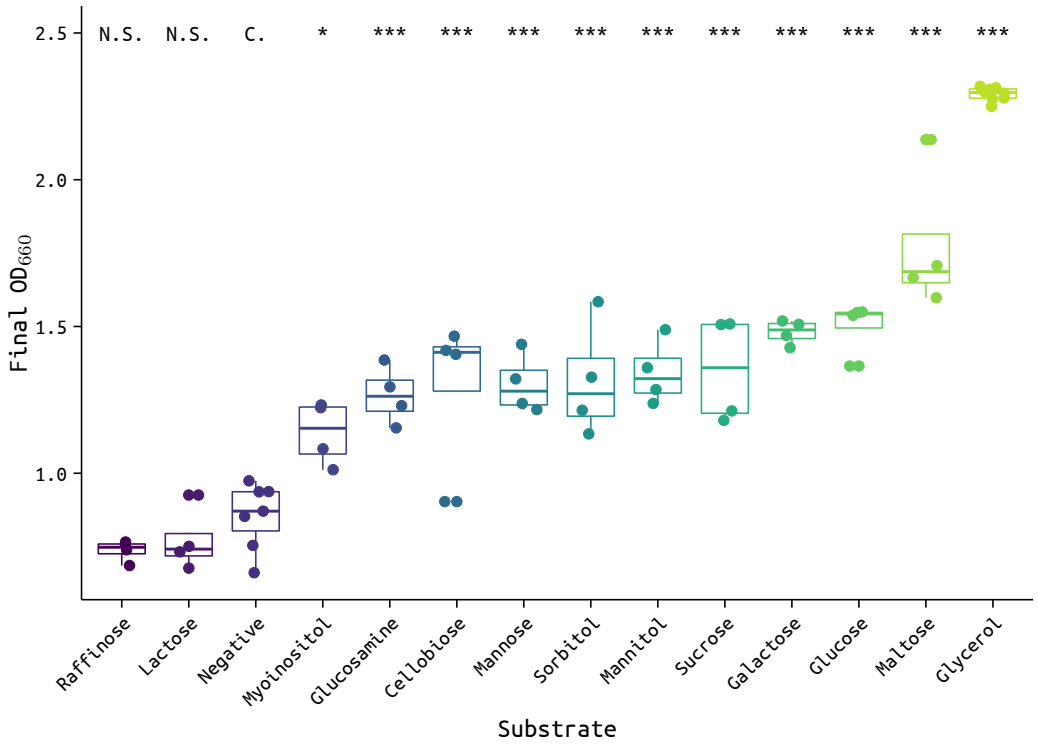


Figure 3.8. – Growth of *R. palustris* with different carbon sources. *R. palustris* was growth in minimal medium, supplemented with urea, and different carbon sources, as indicated. The cultures were all started at an OD_{660} of 0.2, and the final OD_{660} reached after 2 weeks growth is shown. A one-way ANOVA was carried out with Dunnett's post-hoc test, with the significance level set to $p < 0.05$, and the control group set to the negative (no carbon) control. The post-hoc test significance stars are annotated above each column in the graph (N.S. : Not Significant, C : Control, * : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$).

trogen source and illumination to permit photoheterotrophic growth. Different organic compounds were added to each sample and the change in OD_{660} was measured. Figure 3.8 shows the growth of *R. palustris* on each of these different organic compounds. Of all of the compounds tested, only raffinose and lactose showed no significant difference from the negative (no carbon) control. All of the other carbon sources showed a significant increase in growth, with glycerol giving the highest growth increase. A subsequent experiment also demonstrated that xylose could also be metabolised (data not shown).

3.4. Discussion

3.4.1. Growth with different nitrogen sources

In previous studies, *R. palustris* has generally been grown using rich-medium (such as yeast extract), and where minimal medium has been used, ammonium or glutamate is generally provided as the nitrogen source, except in cases where nitrogen is fixed from the atmosphere or supplied N_2 gas (Niel, 1944; Ornston and Stanier, 1966; Hegeman, 1967;

M.-K. Kim and Harwood, 1991; R. Ghosh *et al.*, 1994; C. Pan *et al.*, 2008; Pott *et al.*, 2013). The results presented show that urea, which is not believed to have been previously trialled in *R. palustris*, can be used as a source of nitrogen for the growth of *R. palustris*, and that with an appropriate electron donor, the CO_2 liberated from the reaction with urease can be fixed as organic carbon for the cell. Urea is therefore a suitable replacement for glutamate as the nitrogen source in the minimal medium, when testing different carbon sources for photoheterotrophic growth. In this case, however, it will be necessary to consider that some carbon sources could act as an electron donor and that therefore the growth observed could be due to the fixing of the CO_2 from urea rather than, or in addition to, the carbon from the carbon source under test.

Although ammonium sulphate did support growth of *R. palustris*, when it was supplied as the source of nitrogen, the drop in pH observed in the culture makes ammonium sulphate an unsuitable nitrogen source when *R. palustris* is grown for longer than three weeks.

Urea is also interesting as a waste source itself, as it can be found in high concentrations in agricultural waste, along with ammonia. Based on the nitrogen source testing therefore, *R. palustris* might be suitable for treating agricultural waste or run-off. Agricultural waste contains large amounts of organic carbon, and also potential electron donors. This means that *R. palustris* could grow photoheterotrophically or photoautotrophically on this material, but it is likely that both processes would occur. Growth would also be influenced by biological conditions such as pH or other toxic substances that may be present. As with glycerol (Pott *et al.*, 2014), some pre-treatment might be necessary before *R. palustris* could be grown on agricultural waste.

R. palustris could also be used to pre-treat material before it is used in anaerobic digestion systems. This could be helpful, as it might reduce the amount of inhibitory ammonia produced in the anaerobic digester. The ammonia is produced from the degradation of proteins and urea (Kayhanian, 1999; Y. Chen *et al.*, 2008), but pre-treating the waste by digestion with *R. palustris* could reduce the amounts of these ammonia precursors.

Previous work has shown that *R. palustris* can be used in bio-remediation strategies to process toxic compounds into less toxic forms (Zhao *et al.*, 2011; Colica *et al.*, 2012; Llorens *et al.*, 2012; B. Li *et al.*, 2014). This could suggest a possible role for *R. palustris* in the pre-treatment of waste materials, such as anaerobic digestion feedstocks, to improve the biocompatability of these wastes for use in traditional anaerobic digesters.

Photoautotrophic growth

The evidence gathered provides evidence for photoautotrophic growth of *R. palustris*, albeit in the presence of an artificial electron donor (Figure 3.7). This is in agreement with other studies mentioning photoautotrophic growth of *R. palustris* (Rolls and Lindstrom, 1967; Knobloch *et al.*, 1971; Jiao *et al.*, 2005; Rey and Oda *et al.*, 2006; VerBerkmoes *et al.*, 2006; Huang *et al.*, 2010; Gordon and McKinlay, 2014). The requirement for an

electron donor is not surprising, as there may be other compounds found in the natural environment of *R. palustris* that can perform this function. Many sulphur compounds are known to be able to donate electrons to the photosynthetic electron transport chain in other purple photosynthetic bacteria (Trüper and U. Fischer, 1982), as are hydrogen (H_2) and some organic acids (Wang *et al.*, 1993).

3.4.2. Growth with different carbon sources

The results presented show that *R. palustris* is capable of metabolising a variety of simple compounds, including sugar alcohols, monosaccharides, disaccharides, and glycerol. This fits well with the broad range of habitats that *R. palustris* is found in (1.2), reinforces the view of *R. palustris* as being highly metabolically diverse, and strengthens the case for its use in waste remediation. Almost all of the compounds tested can be found in plant materials; the polymer of galactose, galactan, is found in hemicellulose, cellobiose is a sugar released during the degradation of cellulose, and maltose is released during the hydrolysis of starch. The ability of *R. palustris* to metabolise these compounds suggests that it could be used in the degradation of cellulosic plant wastes and agricultural wastes.

Given that *R. palustris* was able to metabolise both galactose and glucose, it is curious that it was unable to metabolise lactose, which is a disaccharide formed of galactose and glucose monomers joined by a β-1-4 glycosidic linkage. Lactose is normally hydrolysed by the enzyme β-galactosidase, but a BLAST (Basic Local Alignment Search Tool) search of the *E. coli* gene against the *R. palustris* genome produced no significant matches, suggesting that *R. palustris* lacks this enzyme, and thus the ability to metabolise lactose. *R. palustris* was also unable to metabolise raffinose, a trisaccharide consisting of galactose, glucose and fructose monomers linked by 1-6 and 1-2 glycosidic linkages, respectively. Raffinose is normally broken down to galactose and sucrose by the enzyme α-galactosidase, and a BLAST search also failed to find a homologue for this enzyme in *R. palustris*.

3.5. Conclusions

The evidence presented in this study reinforces the view of *R. palustris* extreme metabolic diversity. Urea was identified as a suitable nitrogen source, as well as a possible source of inorganic carbon, and a potential waste material for remediation by *R. palustris*. The ability of *R. palustris* to degrade simple compounds, including sugar alcohols, monosaccharides, and disaccharides, which may be present in agricultural waste materials is demonstrated. These findings strengthen the case for the use of *R. palustris* in the bio-remediation of waste materials, and provide evidence to suggest that agricultural waste in particular may be a suitable candidate for such a strategy.

Future research could focus on broadening the range of known substrates for growth of

R. palustris, in particular those associated with ligno-cellulosic material. Purified cellulose and lignin should be trialled, and examples of real waste materials should be tested.

Increasing glycerol and urea metabolism

4

4.1. Introduction

Due to the rising prices of crude oil, severe environmental concerns (especially regarding global climate change), and political instability, the manufacture of biofuels from biomass has become increasingly popular. The two most common biofuels are bioethanol, typically produced from corn grain, and biodiesel, typically produced from soybean (Hill *et al.*, 2006).

Biodiesel is produced by the transesterification of acylglycerides with methanol and a catalyst (typically NaOH or KOH), yielding fatty acid methyl esters (FAME) (biodiesel) and waste glycerol (Figure 4.1). The waste glycerol produced is generally extremely caustic, due to the catalyst, and often contaminated with fatty acids and soaps due to the unwanted partial hydrolysis of acylglycerides to fatty acids, and the subsequent saponification of the fatty acids. This makes the by-product toxic, and difficult to process into usable products.

There used to be a profitable market for glycerol, however, due to the increase in supply from increased biodiesel manufacture, there is now a glut in the market and the price of waste glycerol has dropped, potentially making it a cheap chemical feedstock (Hill *et al.*, 2006; Johnson and Taconi, 2007).

R. palustris has been shown to have the ability to metabolise glycerol, including the toxic and contaminated waste glycerol produced by biodiesel manufacture (Pott *et al.*, 2013; 2014). Hydrogen gas can be produced as a by-product of this metabolic process, thus yielding another useful, sustainably produced, and clean-burning fuel (H_2O is the by-product of H_2 combustion, rather than CO_2).

Considering this ability, it was reasoned that it might be possible to increase the yields of hydrogen gas, or other products, from *R. palustris* by increasing the transport of glycerol into the cell. One way of achieving this goal would be to express glycerol specific transporters in *R. palustris*. Two genes of interest were identified for the project; *glpF*, a glycerol transporter of the aquaporin family, identified in *E. coli* (D. Fu and Libson and Miercke *et al.*, 2000; D. Fu and Libson and Stroud, 2002), and *aqpZ* (RPA2485) an aquaporin endogenous to *R. palustris* (Simmons *et al.*, 2011).

GlpF is known to be a glycerol specific transporter (D. Fu and Libson and Miercke *et al.*,

2000; D. Fu and Libson and Stroud, 2002; Jensen *et al.*, 2003; Oliva *et al.*, 2010) and if it could be expressed in *R. palustris* this might facilitate glycerol transport, whereas the native *R. palustris* aquaporin AqpZ is not known to carry glycerol (Simmons *et al.*, 2011). An analysis of the *R. palustris* AqpZ by Simmons *et al.* (2011) found it to be homologous to the *Rattus norvegicus* AqpZ1 discussed by Beitz *et al.* (2006). In the study by Beitz *et al.* (2006), different mutations of AqpZ1 and the effects of these on its selectivity were investigated. In particular, they found that two amino acid substitutions, F56A and H180A caused a widening of the channel in AqpZ1 and allowed the transport of glycerol and urea. Simmons *et al.* (2011) identified the corresponding residues in the *R. palustris* AqpZ sequence to be F44 and H176; Figure 4.2 shows an alignment of the protein sequences.

Based on this information, it was reasoned that making the same amino acid substitutions to the homologous residues in the *R. palustris* AqpZ protein might also confer increased ability to transport glycerol and urea. This was deemed to be particularly advantageous, since earlier work (Chapter 3) showed that *R. palustris* can also metabolise urea, as well as glycerol, and therefore increasing the transport of urea might also confer a growth increase. The proposed edits to the *R. palustris* AqpZ are shown in Figure 4.2, aligned with the native *R. palustris* AqpZ protein and both the native and mutated *Rattus norvegicus* AqpZ1.

4.2. Experimental

4.2.1. Construction of heterologous lines

Three strains of *R. palustris* were engineered, two expressing the heterologous protein GlpF from *E. coli*, one with a C-terminal cMyc-6xHis double tag, and one without. In the third strain, the endogenous gene for the aquaporin AqpZ was mutated by two amino acid substitutions (F44A & H174A) and the addition of a C-terminal cMyc-6xHis double tag. These strains were assayed for their glycerol consumption relative to WT (wild type) (Chapter 2.2.8).

GlpF

The coding sequence of the *glpF* gene was amplified by colony PCR from *Escherichia coli* DH5α, and expressed under the control of the endogenous *R. palustris* citrate synthase promoter (a 150 bp sequence directly upstream of the citrate synthase start codon) and terminator (a 94 bp sequence directly downstream of the citrate synthase stop codon). The primers used to amplify these sequences are given in Table 4.1. The citrate synthase promoter was picked as it has high constitutive expression in *R. palustris* (McKinlay *et al.*, 2014). The construct was inserted so as to replace RPA0768, a transposase and putative

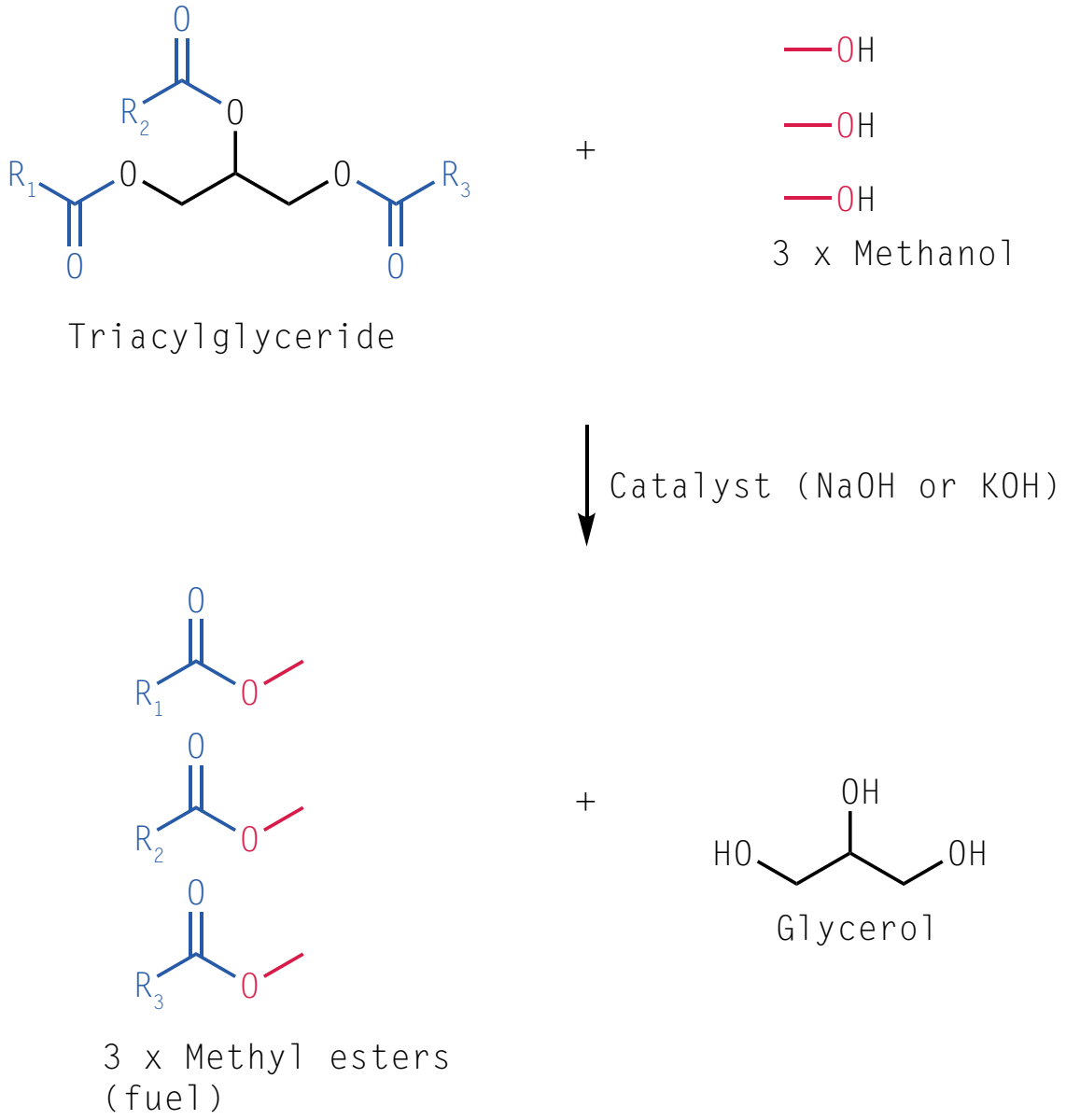


Figure 4.1. – Production of biofuel by transesterification. Diagram showing the reaction for creating biofuel from material such as vegetable oils (triacylglycerides) and the glycerol that is produced as a by-product.

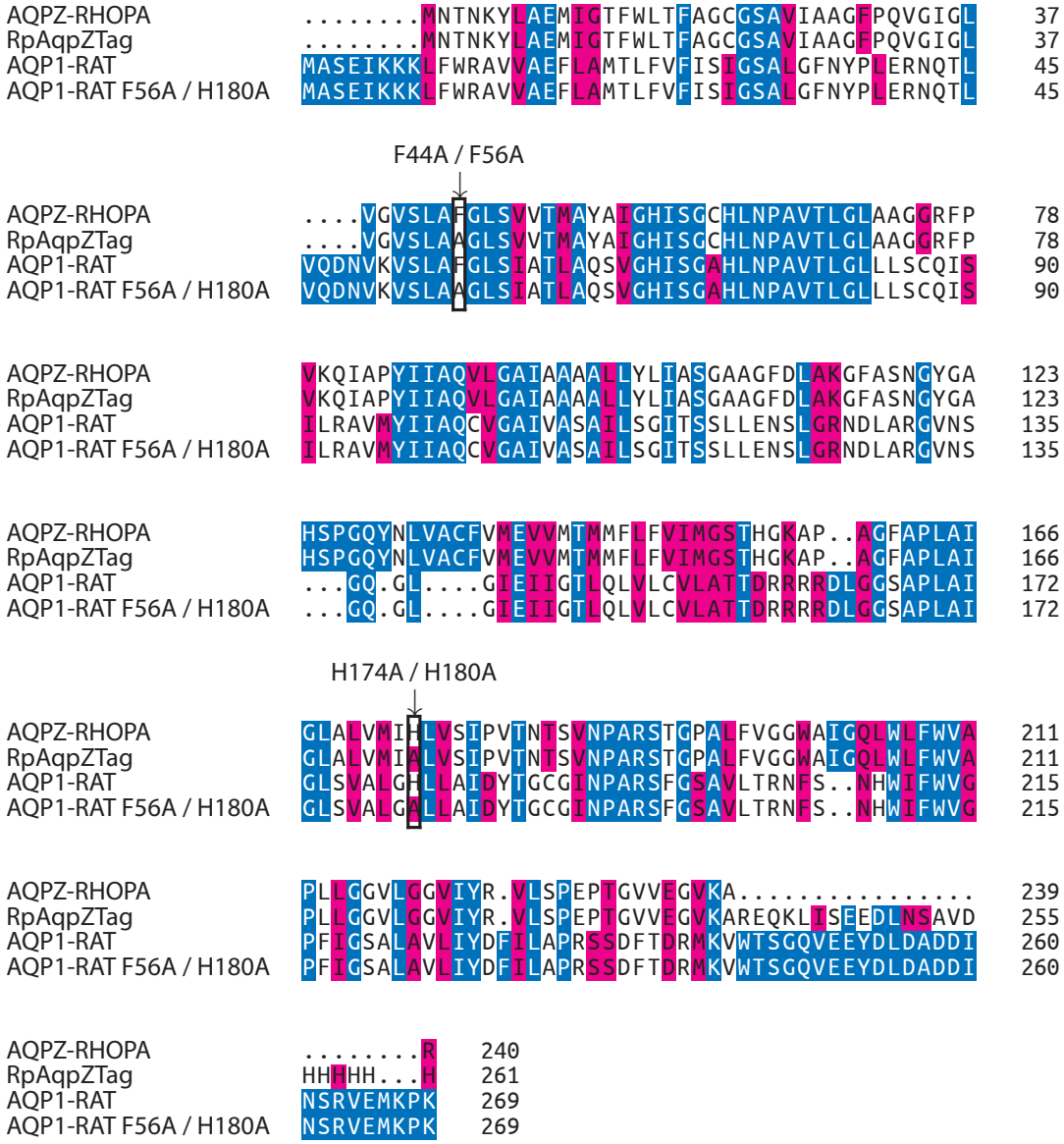


Figure 4.2. – Alignment of AqpZ protein sequences. Alignment of AqpZ protein sequences from *R. palustris* RPA2485 (AQPZ-RHOPA) and *Rattus norvegicus* AqpZ1 (AQP1-RAT), along with the mutated *Rattus norvegicus* AqpZ1 created by Beitz *et al.* (2006) (AQP1-RAT F56A / H180A), and the proposed RpAqpZ::Tag sequence (RpAqpZTag) with the F44A / H174A mutations. Identical sequence regions are shaded in blue, while highly similar regions are shaded in fuchsia. The alignment was performed using T-Coffee software and the shading calculated by the LaTeX package TeXshade.

neutral site (this loci is hereafter referred to as "T1"). The fragments were all assembled into the pKmobSacB backbone vector (Schäfer *et al.*, 1994).

AqpZ

The mutated *aqpZ* gene was artificially synthesised by GeneArt and then assembled into the pKmobSacB backbone vector (Schäfer *et al.*, 1994) with 1000 bp flanking regions. The flanking regions were amplified from the 1000 bp sequence directly upstream and downstream of the native *aqpZ* coding sequence. The primers used to amplify these sequences are given in Table 4.1. Figure 4.4 shows the synthesised mutant sequence aligned with the native sequence. No additional promoter or terminator sequences were inserted, as the mutated gene was inserted so as to replace the native *aqpZ* gene (RPA2485), leaving its native promoter and terminator sequences intact.

Plasmid construction

The plasmids were assembled by Gibson-assembly cloning (D. G. Gibson *et al.*, 2009; D. G. Gibson, 2011), using the NEBuilder HiFi DNA Assembly kit by New England Biolabs (Chapter 2.2.3). Each of the fragments was amplified by PCR (Chapter 2.2.2) using the primers shown in Table 4.1 and bacterial colonies for the template material. The cMyc-6xHis double tag for RpGlpF::T1::Tag was synthesised as part of a 94 bp primer (Table 4.1). The vector backbone was pKmobSacB (Schäfer *et al.*, 1994). Maps of the plasmids and insert sites are shown in Figures 4.5 & 4.6. The sequence of the constructed plasmids was confirmed by restriction digestion and capillary Sanger sequencing of the assembled plasmid insert (data not shown).

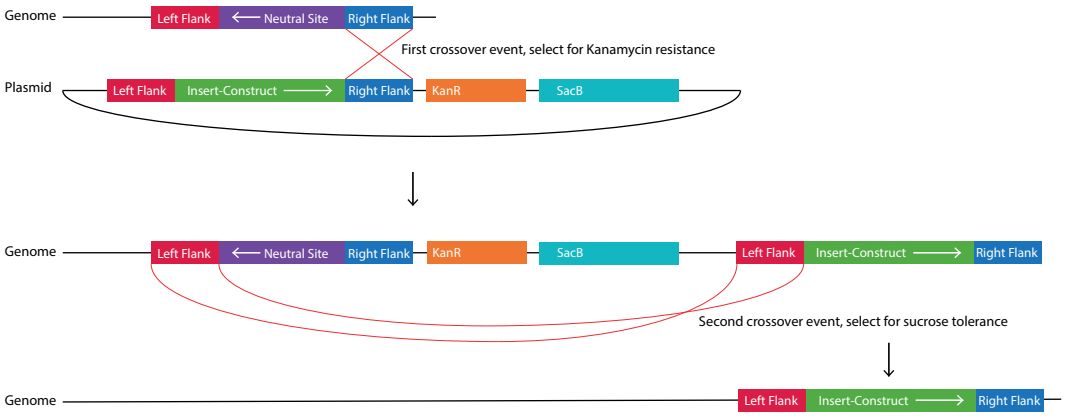


Figure 4.3. – *R. palustris* transformation strategy. Diagram showing the transformation process for creating new *R. palustris* strains from pkMobSacB based plasmids. A single crossover event, between either the left or right flanks (right pictured), is selected for using kanamycin as the selection agent, which results in the introduction of the whole plasmid, next to the insertion locus. A second crossover event (if performed) is selected for using sucrose, and results in either an unmarked 'knock-in' where the insert is left in place and the remainder of the plasmid and the WT insertion locus is excised (pictured), or a reversion to WT.

Table 4.1. – Primers used to make RpGlpF and RpAqpZ plasmids. The gene specific binding regions are shown in uppercase, whilst the overhanging region used to create the overlaps for Gibson assembly are shown in lowercase.

Primer Name	Fragment amplified	Binding template	Primer sequence 3' - 5'	Length (nt)	Overlaps with	Used to make plasmid(s)
T1L-F	T1 Left Flank	*R. palustris* genome	tgcatgcctgcaggtcgactctagaCTCTCACCGGACTTGCCAATC	46	pKmobSacB	RpGlpF::T1, RpGlpF::T1::Tag
T1L-R	T1 Left Flank	*R. palustris* genome	cccgggttttCCGGTCGATCAGGCCTTC	28	Citrate synthase promoter	RpGlpF::T1, RpGlpF::T1::Tag
Prom-F-T1	Citrate synthase promoter	*R. palustris* genome	gatcgaccggAAAACCCGGGAAACGGCT	28	T1 Left Flank	RpGlpF::T1, RpGlpF::T1::Tag
Prom-R-Eco	Citrate synthase promoter	*R. palustris* genome	tttgactcatCGTGTGGTCCCCGATGTT	28	*E. coli* GlpF	RpGlpF::T1, RpGlpF::T1::Tag
EcoGlpF-F	*E. coli* GlpF	*E. coli* genome	ggaccacacgATGAGTCAAACATCAACCTTG	31	Citrate synthase promoter	RpGlpF::T1, RpGlpF::T1::Tag
EcoGlpF-R-NT	*E. coli* GlpF	*E. coli* genome	gggtcgacgaTTACAGCGAAGCTTTTTG	28	Citrate synthase terminator	RpGlpF::T1
EcoGlpF-R-T	*E. coli* GlpF	*E. coli* genome	ctgagatgagtttttgttcCAGCGAAGCTTTTTGTTC	37	Citrate synthase terminator	RpGlpF::T1::Tag

Table 4.1 continued from previous page

Primer Name	Fragment amplified	Binding template	Primer sequence 3' - 5'	Length (nt)	Overlaps with	Used to make plasmid(s)
Term-F-Eco-NT	Citrate synthase terminator	*R. palustris* genome	ttcgctgtaaTCGTCGACCCGGCTCGAT	28	*E. coli* GlpF	RpGlpF::T1
Term-F-Eco-T	Citrate synthase terminator	*R. palustris* genome	agcttcgctg gaacaaaaactcatctcagaagaggatctga atagcgccgtcgaccatcatcatcatcattaa TCGTCGACCCGGCTCGAT	94	*E. coli* GlpF	RpGlpF::T1::Tag
Term-R-T1	Citrate synthase terminator	*R. palustris* genome	gctgacagtgGTTTAGCGGAACGGCTCCG	29	T1 Right Flank	RpGlpF::T1, RpGlpF::T1::Tag
T1R-F	T1 Right Flank	*R. palustris* genome	tccgctaaacCACTGTCAGCCGATTACC	28	Citrate synthase terminator	RpGlpF::T1, RpGlpF::T1::Tag
T1R-R	T1 Right Flank	*R. palustris* genome	attcgagctcggtacccggggatccTTTCTCCCCCGGAAAATG	43	pKmobSacB	RpGlpF::T1 ,RpGlpF::T1::Tag
ALeft-F	AQ Left Flank	*R. palustris* genome	tgcatgcctgcaggtcgactctagaGCACGGTTTGGCCGGCAT	43	pKmobSacB	RpAqpZ
ALeft-R	AQ Left Flank	*R. palustris* genome	tcgtattcatGTGTGTTTCCTCGATCGTCACG	32	F44A H14A AqpZ Tag	RpAqpZ
AqpZTag-F	F44A H174A AqpZ Tag	Synthesised AqpZ	ggaaacacacATGAATACGAACAAGTATCTCG	32	AQ Left Flank	RpAqpZ

Table 4.1 continued from previous page

Primer Name	Fragment amplified	Binding template	Primer sequence 3' - 5'	Length (nt)	Overlaps with	Used to make plasmid(s)
AqpZTag-R	F44A H174A AqpZ Tag	Synthesised AqpZ	gacctcgcggCTAATGATGATGATGATGATG	31	AQ Right Flank	RpAqpZ
ARight-F	AQ Right Flank	*R. palustris* genome	tcatcattagCCGCGAGGTCTGCGCCGC	28	F44A H174A AqpZ Tag	RpAqpZ
ARight-R	AQ Right Flank	*R. palustris* genome	attcgagctcggtacccggggatcCATGCTCAGCGTGTTCGTGCG	45	pKmobSacB	RpAqpZ

Strain construction

The plasmid constructs were introduced to *R. palustris* by electroporation (Chapter 2.2.4). Each plasmid contained two 'flanking regions' (upstream and downstream) which were homologous to the genomic region either side of the insert locus. The antibiotic kanamycin was used to select for a single crossover event, with either one of the flanking regions, which incorporates the entire plasmid into the genome (Figure 4.3).

The genotypes of developed strains were checked by PCR using an external primer pair that bound on either side of the insert site, and in later experiments, a second PCR was performed using an internal forward primer specific to the inserted gene, and the same external reverse primer as before, that bound outside the insert site (Figures 4.5 & 4.6 and Table 4.2).

Expression of *glpF* gene

The presence of transcripts of the heterologous gene was tested by RT-PCR (Chapter 2.2.6), using primers internal to the *glpF* gene (Table 4.2).

Expression of GlpF protein

Protein expression was assayed by Western blot, using the tagged lines and antibodies to c-Myc and poly-His (Chapter 2.2.7).

4.2.2. Assessment of glycerol consumption

Glycerol consumption was initially measured only with the RpGlpF::T1 strain and WT, at the end of a growth curve, using a spectrophotometric method based on the periodate oxidation of glycerol to formaldehyde (Malaprade reaction) and its subsequent derivatisation to 3,5-diacetyl-1,4-dihydrolutidine and measurement at 410 nm (Chapter 2.2.8). Later, after a new set of strains were created, the glycerol consumption of both RpGlpF::T1 strains, the RpAqpZ::Tag strain, and the WT, were measured over time during a growth curve, using a GCMS (gas chromatography mass spectrometry) method (Chapter 2.2.8).

4.3. Results

4.3.1. Construction of heterologous lines

Genotyping

After the transformation, approximately 100 colonies ($\pm \sim 50$) were obtained for each transformed line, 24 RpGlpF::T1, and 27 RpGlpF::T1::Tag colonies were selected for genotyping by PCR. The reaction was carried out using a pair of primers that bound to the

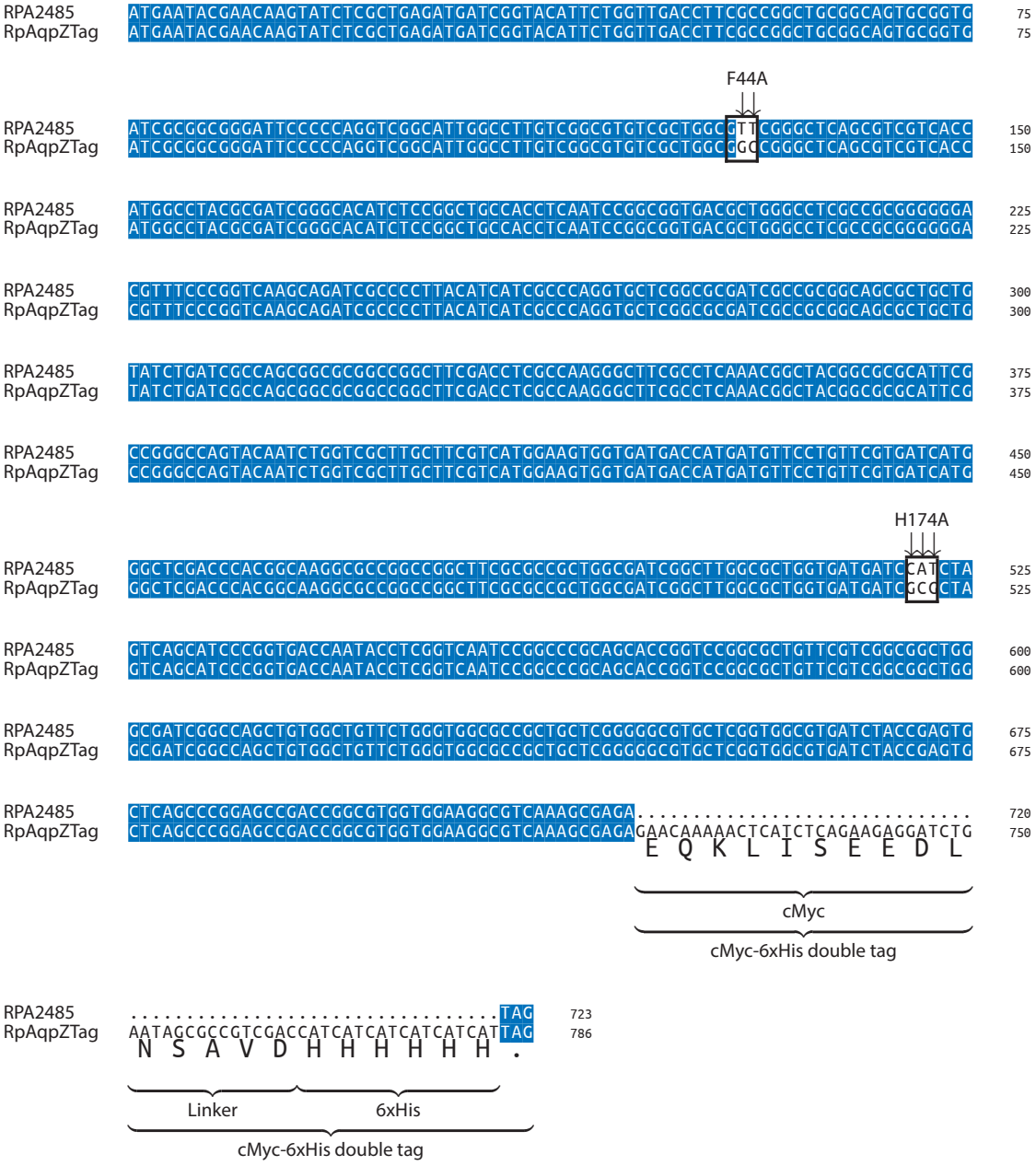
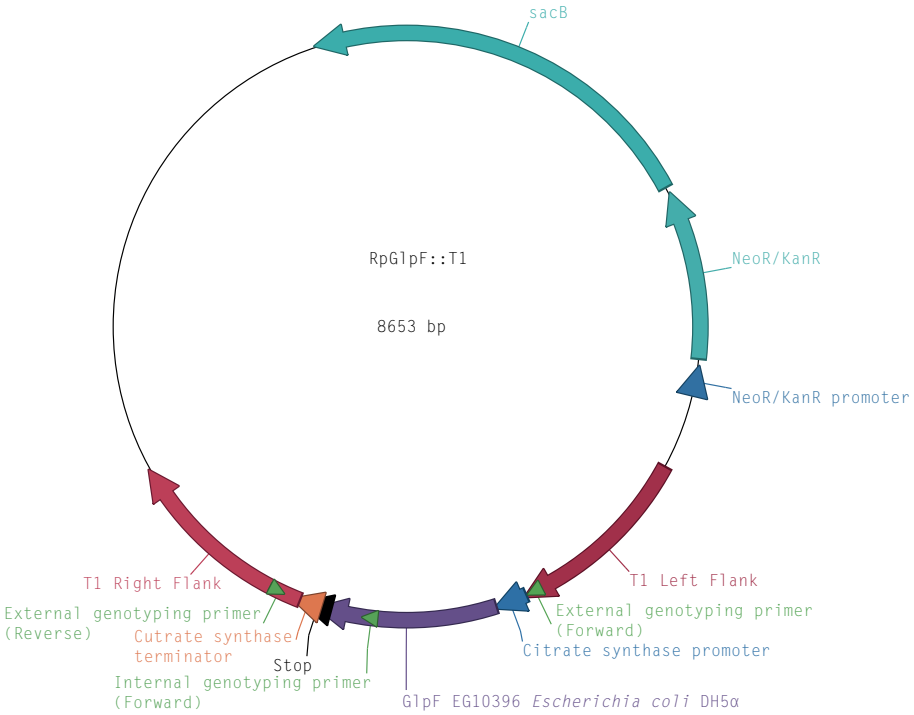
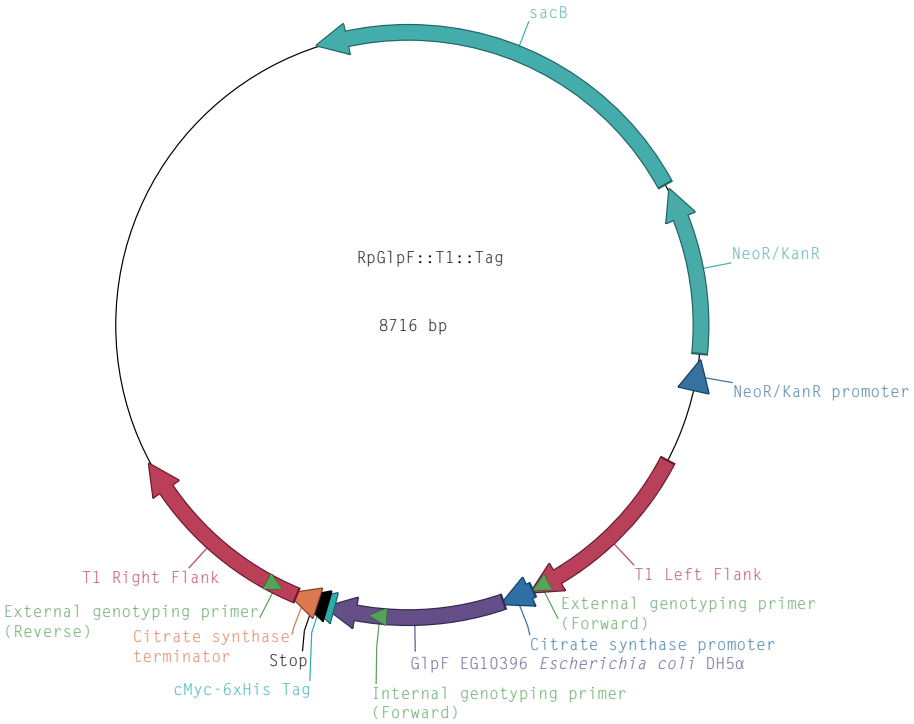


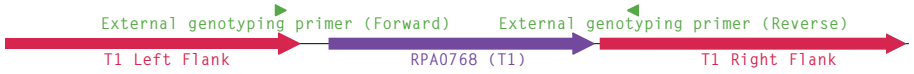
Figure 4.4. – Alignment of synthesised AqpZ DNA sequence. Alignment of AqpZ DNA sequences from *R. palustris* RPA2485 and the synthesised RpAqpZ::Tag sequence (RpAqpZTag) with the F44A / H174A mutations. Identical sequence regions are shaded in blue. The alignment was performed using T-Coffee software and the shading calculated by the LaTeX package TeXshade.



(a) Plasmid map of RpGlpF::T1

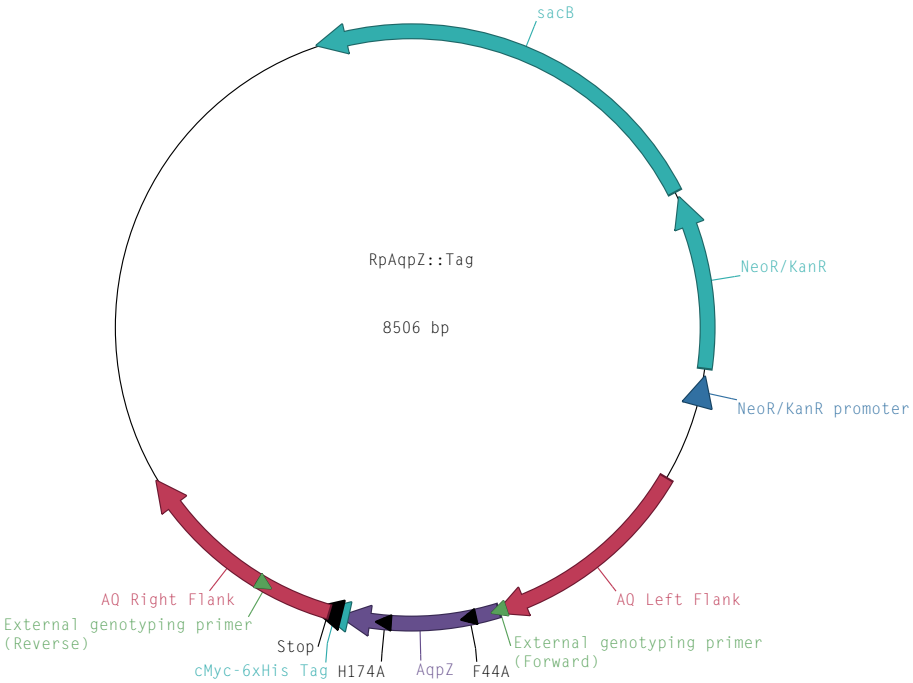


(b) Plasmid map of RpGlpF::T1::Tag



(c) T1 (RPA0768) neutral insertion site

Figure 4.5. – Plasmid maps and genomic context of the T1 insertion site for the RpGlpF::T1 strains.



(a) Plasmid map of RpAqpZ::Tag

(b) Native AqpZ (RPA2485) gene

Figure 4.6. – Plasmid map and genomic context of the AqpZ insertion site for the RpAqpZ::Tag strains.

Table 4.2. – Primers used to genotype and check expression of RpGlpF::T1 and RpAqpZ::Tag strains

Primer Name	Primer sequence 3' - 5'	Length (nt)
External genotyping primer (Forward) - T1	CGGCAGCAGTTCTACGAGAT	20
External genotyping primer (Reverse) - T1	CTTCTCCACTGTCGCGATGA	20
Internal genotyping primer (Forward) - GlpF	TATGGGCCCATTGACAGGTT	20
External genotyping primer (Forward) - AqpZ	CCGTGACGATCGAGGAAACA	20
External genotyping primer (Reverse) - AqpZ	CCACAAGGAAGAGCTGCTGA	20
GlpF RT-PCR Forward	TATGGGCCCATTGACAGGTT	20
GlpF RT-PCR Reverse	CCAACGATAGGGCCGAAAAG	20
Tag primer (Forward)	AGAGGATCTGAATAGCGCCG	20
Tag primer (Reverse)	ACGGCGCTATTCAGATCCTC	20

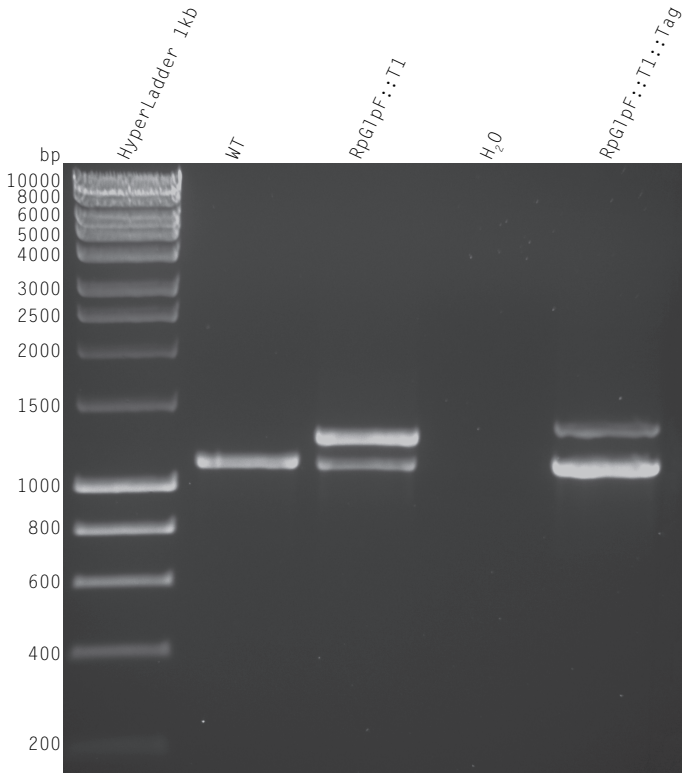


Figure 4.7. – Genotyping of RpGlpF::T1 strains, first crossover. This gel shows representative results of the PCRs used to genotype the putative RpGlpF::T1 strains after the first crossover event. Primers external to the insert were used to test the genotype; all lanes (except the H_2O control) showed a strong band of the same size as in the WT (1095 bp), and the RpGlpF::T1 strains showed an additional, larger band from the inserted plasmid (1268 bp RpGlpF::T1, and 1331 bp RpGlpF::T1::Tag).

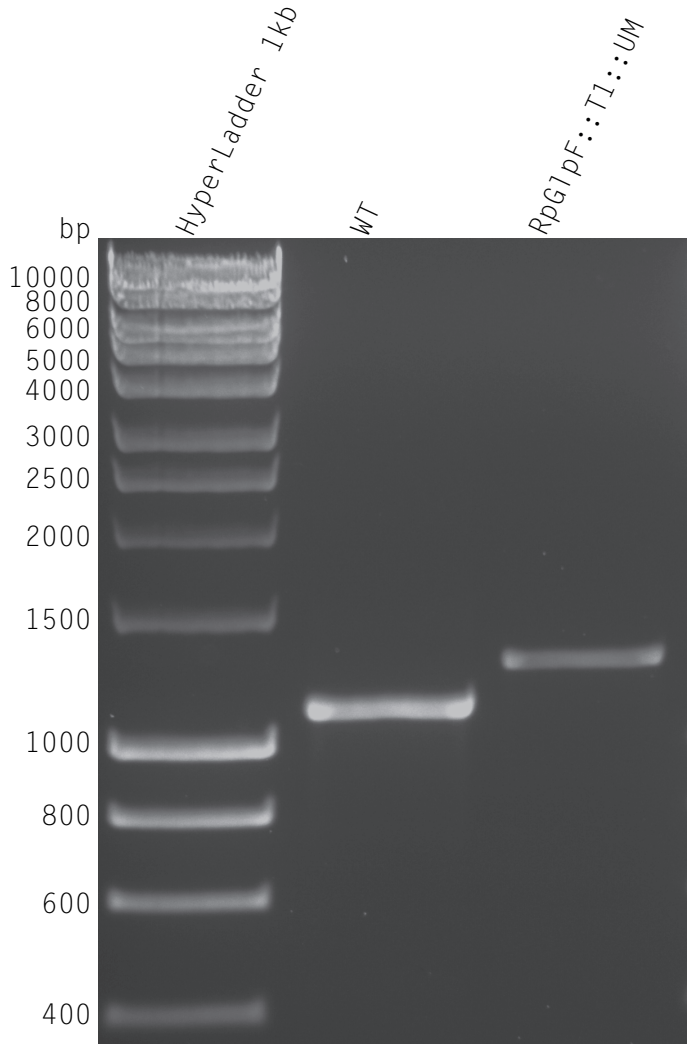


Figure 4.8. – Genotyping of RpGlpF::T1::UM strain, second crossover. This gel shows a representative result of the PCRs used to genotype the RpGlpF::T1::UM strain after the second crossover event. Primers external to the insert were used for genotyping. A single band of the expected size (1095 bp) was observed in the WT, and a single, larger band of the expected size for the insert (1268 bp) was observed in the RpGlpF::T1::UM strain, indicating the presence of the insert and excision of the WT locus.

flanking regions either side of the insertion locus. Two products were expected; one from the WT gene, and one from the inserted gene. Representative results of successful transformants are shown in Figure 4.7. Of the 24 RpGlpF::T1 colonies screened, only 5 produced both a WT and an insert band, and of the 27 RpGlpF::T1::Tag colonies, only 2.

It is not clear why the bands appear at different intensities in Figure 4.7, but it could be related to PCR amplification bias.

All of the successful primary crossover colonies were taken on for secondary crossover (using sucrose sensitivity selection), to generate unmarked strains. Approximately 300 colonies ($\pm \sim 100$) were obtained for each strain, and 25 of each were genotyped by PCR. For this test three results were possible; a single band at the size expected for the WT product (1095 bp), indicating reversion to WT, two bands (one WT and one insert), indicating no crossover event, or a single band at the size expected for the insert (1268 bp RpGlpF::T1, and 1331 bp RpGlpF::T1::Tag), indicating a successful secondary crossover event. Of all of the colonies screened, in all strains, only one RpGlpF::T1 colony produced a successful secondary crossover, yielding the RpGlpF::T1::UM strain. Figure 4.8 shows a representative result of a successful secondary crossover.

Expression of *glpF* gene

The presence of *glpF* transcripts in *R. palustris* was tested by RT-PCR, using primers internal to the *glpF* gene (Table 4.2). A positive result, indicating gene expression, was obtained for the transformed strains (Figure 4.9), indicating that the inserted genes were transcribed. A control reaction, where reverse transcriptase was omitted (-RT), gave no bands (or only very faint ones), showing that the PCR product was due to cDNA template (from RNA transcript) and not from contaminating genomic DNA.

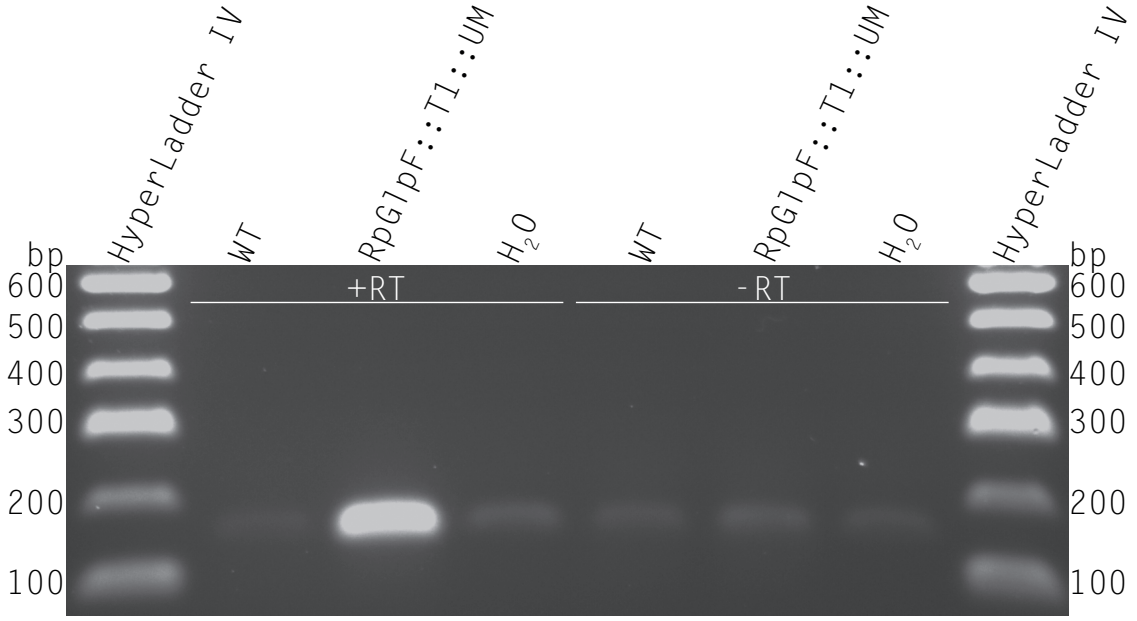
4.3.2. Assessment of glycerol consumption

Growth

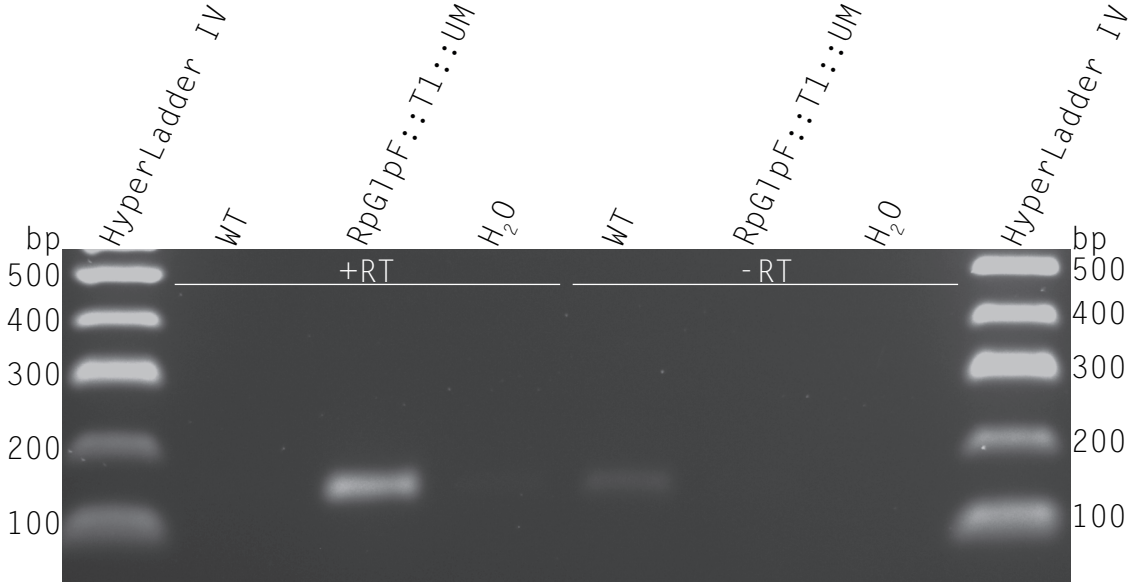
No significant differences in growth were observed between the RpGlpF::T1::UM strain and the WT strain (Figure 4.10).

Glycerol

A the end of the growth curve (Figure 4.10), both the RpGlpF::T1::UM culture, and the WT culture were assayed (using the spectrophotometric glycerol assay - Chapter 2.2.8) to determine the concentration of glycerol remaining in the medium, and thus determine whether one culture had utilised more substrate than the other. Both cultures started at 50 mM glycerol (Chapter 2.1.3). After 54 days of growth, the WT culture had ~0.23 mM



(a) Sample 1



(b) Sample 2

Figure 4.9. – RT-PCR to detect *glpF* gene expression in RpGlpF::T1::UM strain. These gels show the products of the RT-PCR used to detect expression of *glpF* in the RpGlpF::T1::UM strain. A strong positive band of the expected size (150 bp) was observed in the RpGlpF::T1::UM strain but not in the WT or H_2O controls. Only weak, background, genomic DNA signals were detected in the negative RT controls.

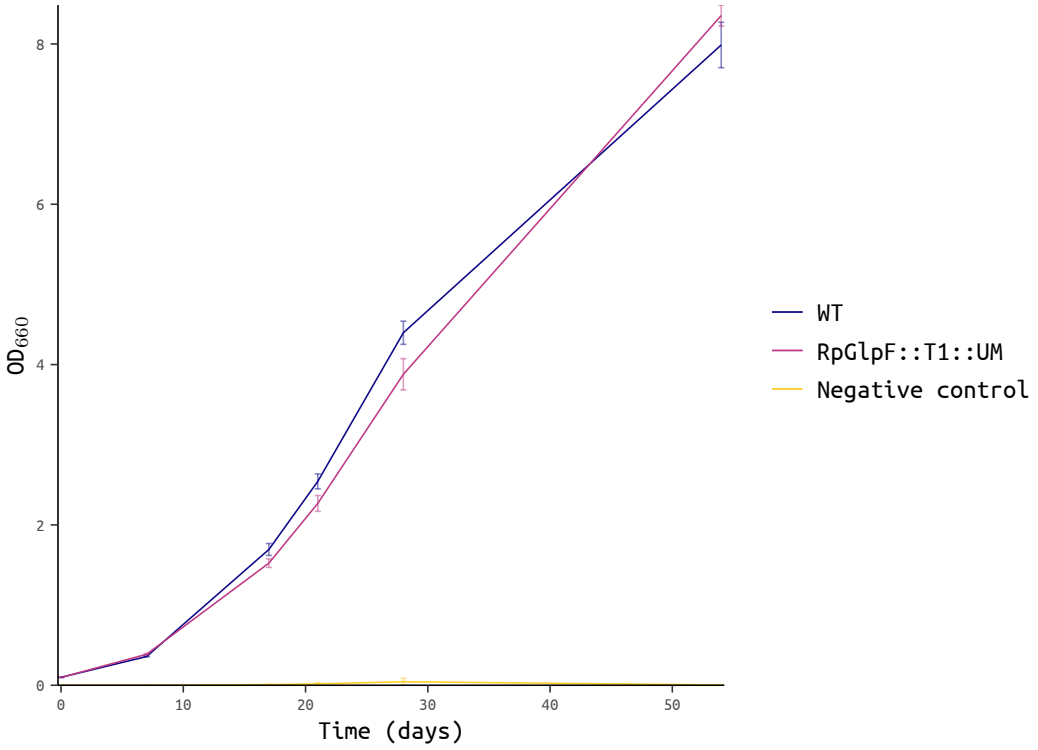
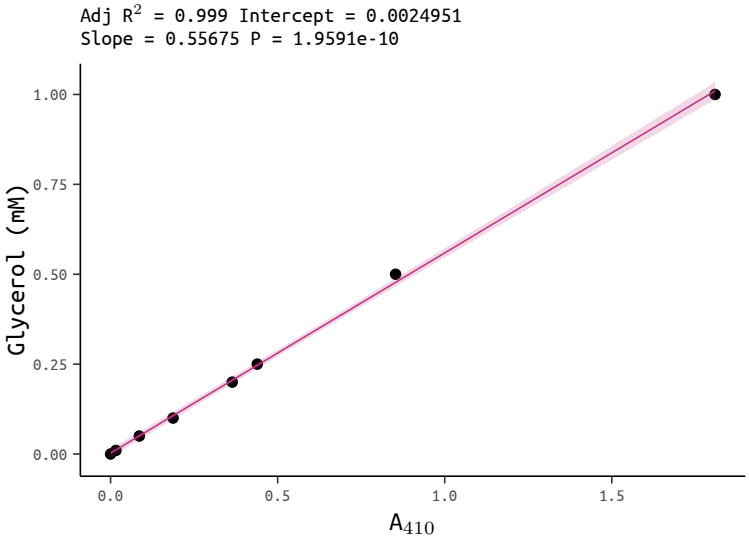
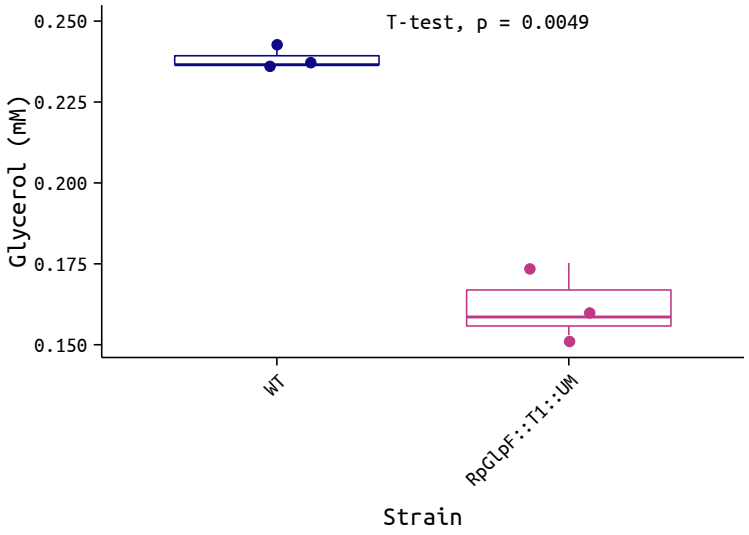


Figure 4.10. – Growth of RpGlpF::T1 strain. The growth of the RpGlpF::T1::UM strain and WT was measured by OD_{660}. The mean of three biological replicates was plotted, with the error bars showing the SEM. The negative control was non-inoculated medium. Samples with OD_{660} absorbances greater than 1 were diluted before measurement.



(a) Glycerol standard curve



(b) Glycerol consumption of WT and RpGlpF::T1::UM strains

Figure 4.11. – Glycerol consumption of WT and RpGlpF::T1::UM strains. (a): Standard curve for the spectrophotometric measurement of glycerol, based on the Malaprade reaction. The absorbance of the product formed at different glycerol concentrations was plotted against its absorbance at 410 nm. (b): Concentration of glycerol, as determined from the spectrophotometric method and standard curve, remaining in WT and RpGlpF::T1::UM cultures after 54 days of growth.

glycerol, whilst the RpGlpF::T1 culture had ~0.16 mM glycerol remaining (Figure 4.11), and this difference passed a Student's T-test for significance at the $p < 0.05$ level.

Expression of GlpF protein

GlpF protein expression was tested by Western blot, using the anti c-Myc and anti poly-His antibodies for the cMyc-6xHis double tag in the tagged RpGlpF::T1::Tag strain. A strong band was observed in the tagged RpGlpF::T1 strain, but not in WT strain negative control. However a slightly weaker band was also observed in the untagged RpGlpF::T1::UM strain. The bands were observed at ~130 kDa, approximately four times larger than the expected GlpF protein size (GlpF: 29.78 kDa, GlpF-Tag: 32.28 kDa). The same results were obtained with both the anti c-Myc and the anti poly-His antibodies (Figure 4.12). To determine whether the untagged RpGlpF::T1::UM strain had been contaminated with a tagged strain, PCR was used to re-check the genotype, and an additional PCR reaction was performed to detect the presence of the tag. Neither reaction indicated that any contamination had occurred, and the presence of the tag could not be detected (data not shown).

New Strains; GlpF, AqpZ

At this point, the RpGlpF::T1 and RpGlpF::T1::Tag strains were remade, to rule out any possibility of contamination that was not detected by PCR, and the RpAqpZ::Tag strain was developed for the first time. Secondary crossovers were not performed with these strains, so all were marked strains. Figure 4.13 shows representative results of the genotyping tests of these new strains.

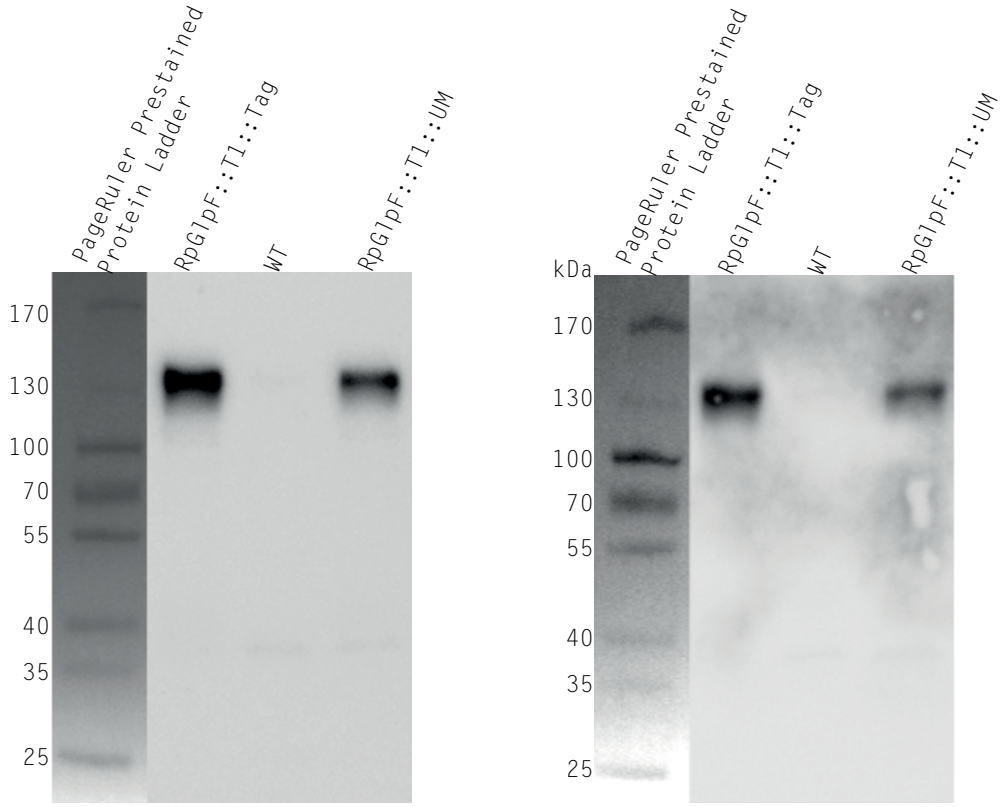
No significant differences in growth were observed between any of the transformed strains and the WT strain (Figure 4.14).

All four cultures were assayed for glycerol content (using the GCMS glycerol assay - Chapter 2.2.8) during the growth curve. No significant differences in the decline in glycerol concentration was observed between any of the strains (Figure 4.15).

GlpF and AqpZ protein expression was tested by Western blot, using the anti c-Myc antibody to the cMyc-6xHis double tag in the tagged RpGlpF::T1::Tag and RpAqpZ::Tag strains. Unexpectedly, a positive signal was observed only in the untagged RpGlpF::T1 negative control (Figure 4.16). The positive band was observed at ~60 kDa which is approximately double the expected size for either GlpF or AqpZ (GlpF: 29.78 kDa, GlpF-Tag: 32.28 kDa, AqpZ: 24.51 kDa, AqpZ-Tag: 26.86 kDa).

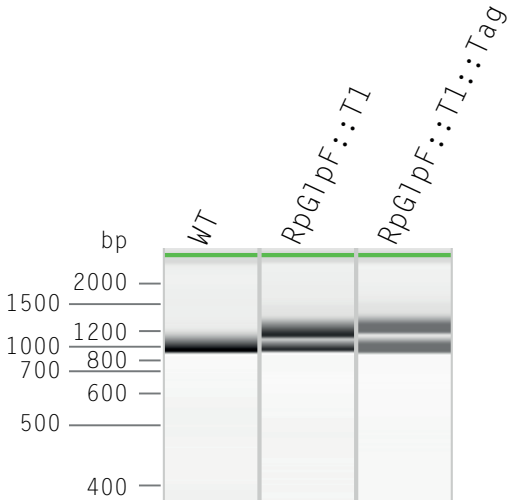
The Western blot was repeated, to ensure that the sample order had not been mixed up, and control blots were performed using only primary, only secondary, and no antibody, to check for cross-reactivity; these were all negative (data not shown).

As before, to determine whether the RpGlpF::T1 strain had been contaminated with a tagged strain, PCR was used to re-check the genotype, and an additional PCR reaction was

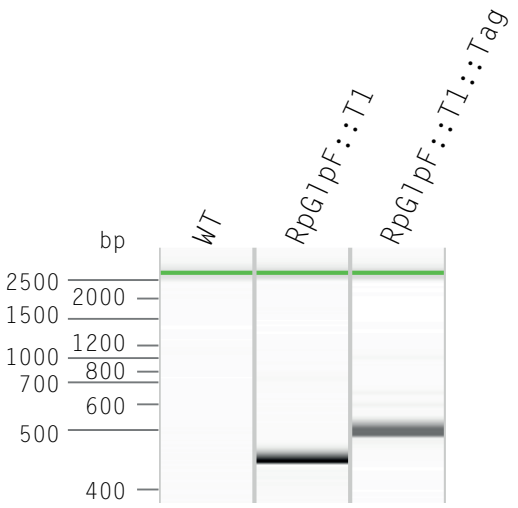


(a) Anti-His **(b)** Anti-Myc

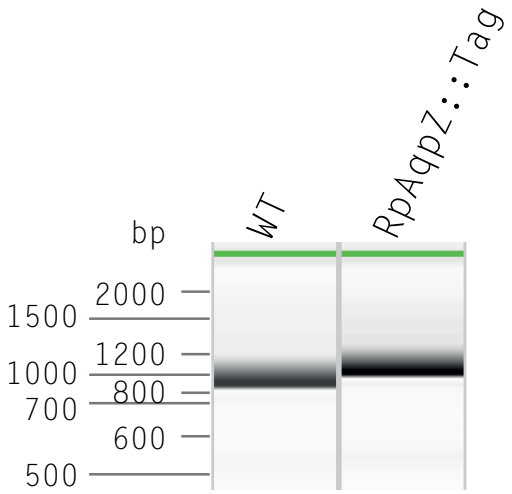
Figure 4.12. – Western blots to detect expression of tagged GlpF protein in RpGlpF::T1 strains. These blots show the signals detected by anti-His and anti-Myc antibodies to the cMyc-6xHis tag in the RpGlpF strains. With both antibodies, a strong positive band was observed in the RpGlpF::T1::Tag strain but not in the WT control. A slightly weaker signal was also observed in the untagged RpGlpF::T1 strain.



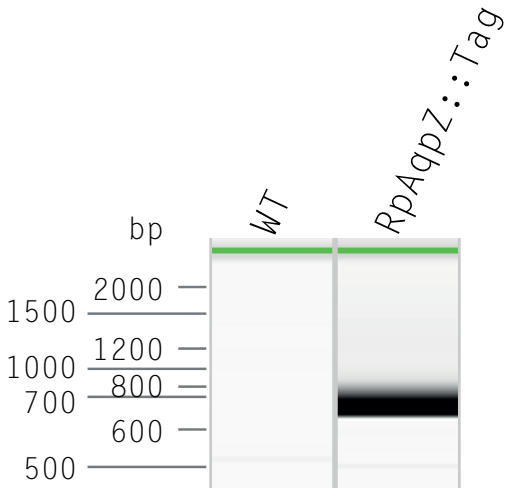
(a) Genotype of new RpGlpF::T1 strains



(b) Check of new RpGlpF::T1 strains genotype with internal primer



(c) Genotype of new RpAqpZ::Tag strain



(d) Check of new RpAqpZ::Tag strain genotype with internal primer

Figure 4.13. – Genotyping of new RpGlpF::T1 and RpAqpZ::Tag strains. These virtual gels, from QIAxcel capillary gel electrophoresis, show representative results of the PCRs used to genotype the RpGlpF and RpAqpZ strains. **(a)** For the RpGlpF::T1 strains, primers external to the insert were used to test the genotype; all lanes showed a strong band of the same size as in the WT (1095 bp), and the RpGlpF::T1 strains showed an additional, larger band from the inserted plasmid (1268 bp RpGlpF::T1, and 1331 bp RpGlpF::T1::Tag). **(b)** The genotype of the RpGlpF::T1 strains was checked with a second PCR using an internal forward primer, specific to the inserted gene, and the same external reverse primer. This gave a product of the expected size (474 bp for RpGlpF::T1, and 537 bp for RpGlpF::T1::Tag) only from the RpGlpF::T1 strains. **(c)** For the RpAqpZ::Tag strain, primers external to the insert were used to test the genotype; both lanes showed a strong band of the same size as in the WT (1065 bp), and the RpAqpZ::Tag strain showed an additional, larger band from the inserted plasmid (1128 bp). **(d)** The genotype of the RpAqpZ::Tag strain was checked with a second PCR using the same external forward primer, and a reverse primer specific to the tag. This gave a product of the expected size (784 bp) only from the RpAqpZ::Tag strain.

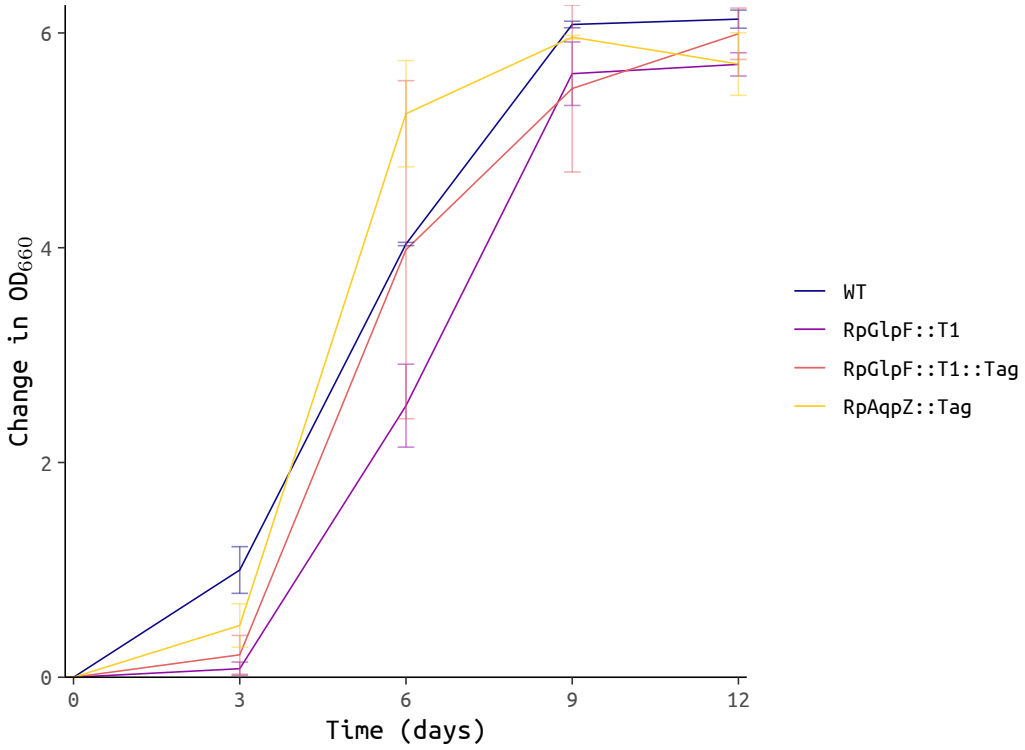


Figure 4.14. – Growth of RpAqpZ::Tag and RpGlpF::T1 strains. This shows the growth of the RpAqpZ::Tag, RpGlpF::T1 strains, and WT, as measured by change in OD_{660} over time. The mean of three biological replicates was plotted, with the error bars showing the SEM.

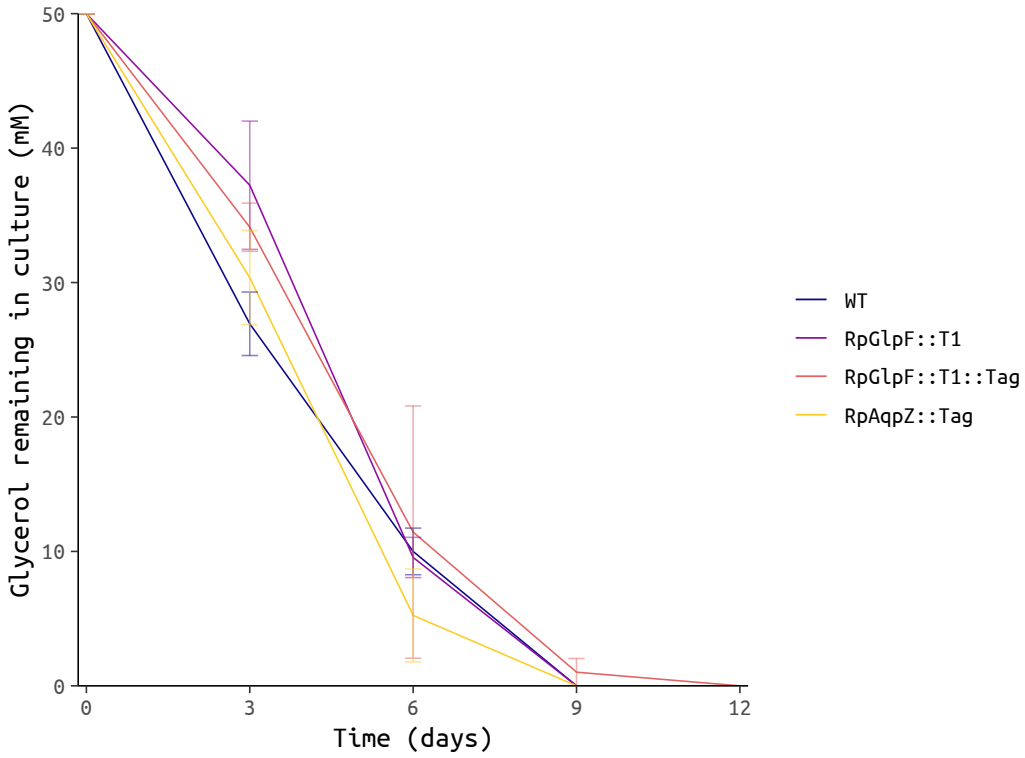


Figure 4.15. – Glycerol consumption of RpAqpZ::Tag and RpGlpF::T1 strains. Concentration of glycerol, as determined by GCMS, remaining in WT, RpGlpF::T1, RpGlpF::T1::Tag & RpAqpZ::Tag cultures over time.

performed to detect the presence of the tag. Neither reaction indicated that any contamination had occurred, and the presence of the tag could not be detected (data not shown).

4.4. Discussion

4.4.1. Construction of heterologous lines

The results presented show that heterologous genes may be incorporated into the *R. palustris* genome, and that both marked and marker-less strains may be created; in this case the RpGlpF::T1::Tag and RpGlpF::T1::UM strains, respectively (Figures 4.7 & 4.8). Evidence is also presented which demonstrates that the method can be used to edit native genes in *R. palustris*, in this case the RpAqpZ::Tag strain (Figure 4.13 (c), (d)).

It was also demonstrated that the transcripts from these heterologous genes are expressed (Figure 4.9), from the inserted native citrate synthase promoter system.

It is not clear whether the heterologous genes were producing functional protein; Western blotting did produce positive signal from the tagged protein line, which was absent in the WT strain negative control. However, a signal was also detected from the untagged heterologous RpGlpF::T1::UM strain, which casts some doubt over the specificity of the detected signal (Figure 4.12). This effect was observed with both the anti-cMyc and the

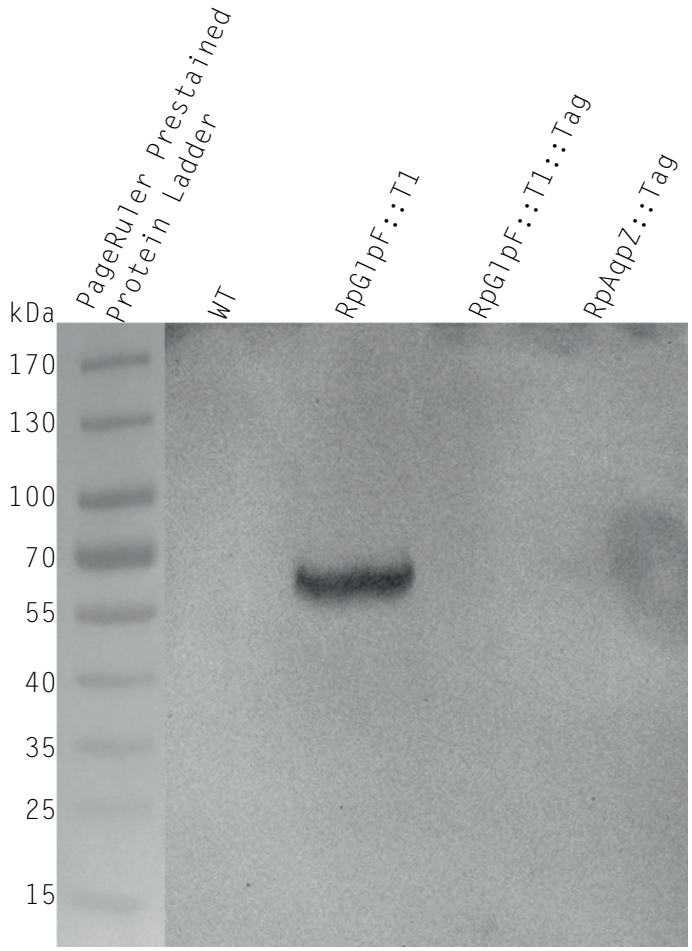


Figure 4.16. – Western blot to detect expression of tagged GlpF and AqpZ proteins in RpGlpF::T1 and RpAqpZ::Tag strains. This blot shows the signals detected by anti-Myc antibody to the cMyc-6xHis tag in the RpAqpZ::Tag and RpGlpF::T1 strains. A strong positive band was observed only in the RpGlpF::T1 untagged strain but not in the WT control or the two tagged RpAqpZ::Tag & RpGlpF::T1::Tag strains.

anti-poly-His antibodies, and it is unclear what could have caused the positive signal from the untagged strain. Cross-contamination of the untagged strain with the tagged strain was ruled out by a PCR reaction that did not detect the presence of the tag.

The signal detected in the Western blot corresponded to a molecular weight approximately four times larger than expected for GlpF protein. GlpF is known to form as tetramers in membranes however (D. Fu and Libson and Miercke *et al.*, 2000; Jensen *et al.*, 2003), and this could potentially account for the larger size of the observed band in the Western blot, although the protein cell lysates were denatured in the presence of SDS (sodium dodecyl-sulphate) and DTT (dithiothreitol) (Chapter 2.2.7) which should have completely denatured the protein and reduced any disulfide bonds, allowing the protein to run on the gel as a monomer. DTT is sensitive to oxidation from the air and is thus relatively unstable, and GlpF does contain cysteine residues, so it is possible that non-native disulfide bonds could have formed between the monomers when the protein was denatured, if the DTT was not fresh enough. Some proteins are also reported to form aggregates that are stable even under the conditions used for sample preparation (Sagné *et al.*, 1996; Rath *et al.*, 2009), and both tetrameric and dimeric GlpF have previously been observed under SDS denaturing conditions (Borgnia and Agre, 2001; Veerappan *et al.*, 2011; Klein *et al.*, 2015).

When the Western blot was repeated with fresh RpGlpF::T1, RpGlpF::T1::Tag, and RpAqpZ::Tag strains, only the untagged RpGlpF::T1 strain produced a signal, but this time, at the expected size for a dimer instead of a tetramer (Figure 4.16). The Western blot was repeated, to ensure that the sample order had not been mixed up, and control blots were performed using only primary, only secondary, and no antibody, to check for cross-reactivity; and these were all negative. Additionally, in this experiment only the anti-Myc antibody produced a signal, the anti-poly-His antibody was negative. There is no obvious explanation for this result. The anti-cMyc antibody was used successfully with the same expression system later, in another project where another heterologous protein was expressed in *R. palustris* (Chapter 6).

The difference in the molecular weight corresponding to the band observed in the two Western blot experiments could potentially be accounted for due to a difference in the protein extraction. In the second experiment the cell lysate was extracted in RIPA (radioimmunoprecipitation assay buffer) buffer, which contains the detergents sodium deoxycholate and Triton X-100 in addition to SDS, in order to improve the solubilisation of membrane-bound proteins. These detergents may have helped to disrupt the GlpF homotetramer, and resulted in the protein appearing as a dimer in the gel/blot. This is similar to previous studies which demonstrated that different detergent conditions can have significant effects on GlpF oligomerisation (Borgnia and Agre, 2001; Veerappan *et al.*, 2011; Klein *et al.*, 2015). Since only the untagged strain produced a signal however, it is far from certain that band observed actually corresponds to GlpF, and may instead be an artifact of some kind.

There were also no phenotypic differences observed between the WT and the transformed strains (see below), and therefore no indirect evidence of heterologous protein function.

4.4.2. Assessment of glycerol consumption

In the first glycerol consumption assay, at the end of the 54 day growth curve comparing WT and RpGlpF::T1, there did appear to be a difference in the glycerol consumption between the two cultures, with the RpGlpF::T1 culture appearing to have consumed more glycerol than the WT strain (Figure 4.11), although the final OD_{660} of the cultures were essentially the same (Figure 4.10). The difference observed, however, was very small (~0.07 mM), and by this point there was very little glycerol left in either culture (~0.16 mM to 0.23 mM). Although the difference appeared to be statistically significant ($p = 0.0049$), given that both cultures started at 50 mM glycerol, this small a difference may not be biologically relevant or significant. There were no detectable differences either in growth or glycerol consumption between the WT and the transformed strains in the subsequent assays (Figures 4.10, 4.14, & 4.15).

The other major substrate present in the medium was urea, and given that no differences in growth were observed, it seems unlikely that urea transport was significantly affected, although this was not measured directly.

This evidence suggests that substrate (specifically glycerol, and to some extent urea) transport into the cell is not the rate-limiting step when considering cell growth or consumption of these substrates. Since the glycerol was not used at a greater rate in the transformed strains, it is unlikely that the yield of hydrogen gas (or other metabolites) would be any higher in the transformed strains, though again, this was not measured directly. It is possible that both glycerol and urea are likely transported by other unidentified, or non-specific transporters, or that there is sufficient passive diffusion through the cell membrane for the expression of a transporter not to have an effect.

Alternatively, it is possible that the transformed strains did not express functional heterologous proteins (see above) and that this is the reason for the lack of any observable difference between these strains and the WT. If protein was produced from the heterologous transcripts, it is also possible that it may not have been properly localised to the cell membrane, and therefore was not functional. Differences in codon usage between *E. coli* and *R. palustris* could also potentially account for a low expression level.

4.5. Conclusions

The evidence presented show that manipulation of the *R. palustris* genome is possible; gene insertion, interruption of endogenous genes, and gene editing are all possible. Expression

of transcripts from heterologous genes is also possible, though as yet there is no conclusive evidence of functional protein expression from these transcripts.

Given the anomalous results observed in the western blots presented here, it will be important to make further studies of the expression of proteins from heterologous gene transcripts in *R. palustris*. This will be crucial for the future development of heterologous expression systems in *R. palustris*, to ensure that functional proteins can be expressed.

Finally, transport of glycerol into the cell did not appear to be rate-limiting, and any further attempts to engineer greater hydrogen or metabolite yields from *R. palustris* should focus on downstream metabolic pathways.

Regulation of urea metabolism

5

5.1. Introduction

PNSB, such as *R. palustris*, are of significant biotechnological interest due to their ability to produce hydrogen gas, usually as a by-product from nitrogen fixation (Gest and Kamen, 1949; Fißler *et al.*, 1994; Koku *et al.*, 2002; Ludden and Roberts, 2002; Basak and Das, 2006; Rey and Heiniger *et al.*, 2007; Carlozzi and Pintucci *et al.*, 2010; Pott *et al.*, 2013). Hydrogen gas is of great interest as a clean-burning fuel (water is the only by-product of hydrogen combustion) and biological production of hydrogen could therefore represent a sustainable source of clean fuel/energy.

Hydrogen metabolism in *R. palustris* occurs *via* two enzymes: nitrogenase and hydrogenase (Vignais and Colbeau *et al.*, 1985). Nitrogenase is the enzyme responsible for the fixation of N_2 to ammonia, producing H_2 as a by-product (Vignais and Colbeau *et al.*, 1985). Uptake hydrogenase consumes hydrogen, oxidising it to produce electrons which are then transferred to cytochromes and ferredoxins and subsequently used as reducing equivalents for CO_2 fixation (Vignais and Toussaint, 1994; Vignais and Colbeau, 2004).

R. palustris strain CGA009 (the strain used throughout this study) is known to be deficient in uptake hydrogenase activity, due to a frameshift mutation in the *hupV* gene, which is a component in hydrogen sensing (Rey and Oda *et al.*, 2006). This affects its ability to recycle hydrogen lost through nitrogen fixation, potentially leading to greater hydrogen yields.

Both nitrogenase and hydrogenase are sensitive to oxygen, the presence of which severely limits H_2 production. In oxygenic photosynthetic organisms such as algae and cyanobacteria, this presents a particular problem, since oxygen is released by the photosynthetic process, thus making efficient biotechnological hydrogen production difficult in these organisms. In PNSB like *R. palustris*, however, photosynthesis is anoxygenic, so there is no oxygen produced and nitrogenase is not inhibited, meaning that these bacteria are particularly suitable for bio-hydrogen production. Moreover, these bacteria are capable of growing anaerobically, so there is also very little exogenous oxygen present.

R. palustris possesses genes for three different subtypes of nitrogenase, which differ principally in their metal cofactors (Oda and Samanta *et al.*, 2005). The most commonly

used nitrogenase is the molybdenum binding nitrogenase, encoded by the *nif* operon, followed by the vanadium and iron binding nitrogenases encoded by the *vnf* and *anf* operons respectively (Oda and Samanta *et al.*, 2005). The expression of these different nitrogenase isozymes is governed primarily by metal availability, as well as in response to increasing nitrogen starvation (Oda and Samanta *et al.*, 2005). Although the molybdenum nitrogenase variant is most common, the vanadium and iron cofactor variants have been associated with high hydrogen yields in other organisms (Miller and Eady, 1988; Schneider *et al.*, 1997; Oda and Samanta *et al.*, 2005). *R. palustris* ability to produce all three variants makes it a particularly interesting candidate for bio-hydrogen production.

Nitrogenase is known to function even in the absence of nitrogen gas, provided it is supplied with sufficient ATP, and this has also been associated with increased hydrogen yields (Sasikala *et al.*, 1990; Hallenbeck, 2011). For optimum hydrogen production, cells would ideally be grown anaerobically and in the absence of N_2, which would then mean that a fixed nitrogen source would be required for cell growth. However, in addition to oxygen sensitivity, both the expression and activity of nitrogenase is also repressed by fixed nitrogen (Jones and Monty, 1979); in particular, both the transcription and activity of nitrogenase is limited by the presence of ammonia. This makes sense, given that it would not be bio-energetically favourable for the organism to expend energy to fix nitrogen, when fixed nitrogen is already available. This poses a problem for biotechnological hydrogen production, particularly when trying to use waste materials (such as agricultural wastes) as substrate, since the growth medium is unlikely to be free from fixed nitrogen, especially ammonium, which is common in waste substrates such as agricultural wastes and biomass.

Nitrogenase expression is dependent on NtrB/C, a two component system that regulates the transcription of numerous genes involved in nitrogen metabolism, including nitrogenase (Masepohl and Kaiser *et al.*, 2001; Drepper and Wiethaus *et al.*, 2006; Heiniger *et al.*, 2012). A study by Heiniger *et al.* (2012) identified two additional proteins that are responsible for the ammonia dependent suppression of nitrogenase activity; DraT2 and GlnK2 which act to post-translationally repress nitrogenase activity. DraT2 is responsible for the post-translational modification of nitrogenase, suppressing its activity, but first requires activation by GlnK2 (Heiniger *et al.*, 2012).

Urea is another compound commonly found in agricultural wastes, and having discovered that *R. palustris* has the ability to metabolise it (Chapter 3), I wanted to investigate how urea metabolism is regulated in *R. palustris*, particularly with respect to how it might affect hydrogen production. Urea is believed to be metabolised by the enzyme urease, releasing ammonia and carbamate, which subsequently decomposes to carbonic acid and additional ammonia (Blakeley and Zerner, 1984; Mobley *et al.*, 1995). It is not clear however, whether the intra-cellular release of ammonia would result in either transcriptional or post-translational suppression of nitrogenase, as would be the case for exogenous ammonia. A previous study of *Rhodobacter capsulatus*, a closely related species, found

that urea metabolism was regulated by the NtrB/C system, which is also responsible for the regulation of ammonium metabolism, but that unlike ammonia, urea did not cause transcriptional repression of nitrogenase by NtrC (Masepohl and Kaiser *et al.*, 2001).

If urea does not inhibit nitrogenase transcription or activity in *R. palustris*, it could make urea a useful nitrogen source for the growth of *R. palustris* for bio-hydrogen production. Additionally, it could help to inform how nitrogen metabolism is regulated, which could have implications for future studies that might seek to decrease the ammonia sensitivity of nitrogenase.

In order to investigate the effects of nitrogen source on the regulation of nitrogen metabolism in *R. palustris*, cultures were grown in the presence of either $[NH_4]_2SO_4$ or urea, and differences in transcript levels determined by RNA sequencing. Additionally, given the suitability of urea as a substrate for growth of *R. palustris*, an analysis of the effects of growth on urea on transcript levels could be informative for future research. For example, it could be used to identify and select promoter systems that lead to high expression levels in the presence of urea, and which could be used in future heterologous expression systems.

In addition to determining the effects of urea on nitrogenase expression, the genes from central metabolism will also be briefly investigated. This will be to check for constitutive expression and validate the use of the citrate synthase promoter in the homologous expression system developed in Chapter 4.

5.2. Experimental

5.2.1. Growth of *R. palustris* cultures

R. palustris was grown in 50 mL cultures in minimal medium (Section 2.1.3) with either 10 mM urea or 10 mM $[NH_4]_2SO_4$ as the nitrogen source. The cultures were grown in a shaking, illuminated incubator for 12 days, and the OD_{660} recorded each day. Samples were taken for RNA extraction after three days. Four biological replicates of each sample were prepared.

5.2.2. RNA extraction

The cultures were centrifuged briefly to remove the growth medium, and the cell pellet re-suspended in 50 µL 10 % SDS in TE (Tris EDTA) buffer, pH 8.0, and heated for 15 minutes at 60 °C on a ThermoMixer at 1400 rpm. 1 mL of TRIzol was added to each and mixed for 15 minutes on a rotary mixer, before subsequent purification using the Direct-zol RNA Kit from Zymo Research. The RNA was diluted 1:50 for analysis by Agilent Bioanalyser.

5.2.3. RNA sequencing

1 µg total RNA from each sample was sent to the University of Cambridge Department of Biochemistry Sequencing Facility, where sequencing libraries were prepared using Illumina RiboZero Bacteria rRNA depletion and Illumina TruSeq Stranded mRNA Library kits.

All eight libraries were sequenced on the Illumina NextSeq500 platform by multiplexing on the High Output 75 cycle kit to give ~400 million 75 bp single-ended reads.

Analysis of read quality was performed using the FastQC software program (Simon Andrews, Babraham Bioinformatics Institute). Reads were trimmed to remove low quality bases and sequencing adapters using TrimGalore! software (Felix Krueger, Babraham Bioinformatics Institute).

5.2.4. Read alignment and data analysis

Reads were aligned to the GenBank GCA_000195775.1_ASM19577v1 *Rhodopseudomonas palustris* CGA009 reference genome using the software package HISAT2 (D. Kim *et al.*, 2015) and assembled and quantified using StringTie and Ballgown (Frazee *et al.*, 2015; M. Pertea and G. M. Pertea *et al.*, 2015; M. Pertea and D. Kim *et al.*, 2016).

Statistical methods

StringTie produced transcript and gene abundance estimates as FPKM (fragments per kilobase of transcript per million mapped reads) values, that had been normalised according to library size. Ballgown was used to perform Log_2 transformation to stabilise variance and then fit linear models to estimate differential gene expression. Transcripts with little change across all samples were filtered out to increase power for differential expression analysis and reduce noise, by removing all transcripts that had a variance of less than one across all eight libraries. The number of transcripts per gene was assessed by mapping transcript IDs to gene IDs.

The distance between each of the samples and replicates was assessed by Pearson correlation (Pearson, 1895), MDS (multi-dimensional scaling) (Kruskal, 1964), and PCA (principal component analysis) (Pearson, 1901) of the Log_2 FPKM values.

The mean Log_2 FPKM values for each of the two sample types ($[NH_4]_2SO_4$ and Urea) were plotted against each other, and the significantly differentially expressed genes indicated, to show the overall spread of the data and the relative distribution of the significantly differentially expressed genes. An MA plot (Dudoit *et al.*, 2002), an application of a Bland-Altman plot (Bland and Altman, 1986), was constructed to show the relationship between the intensity (mean Log_2 FPKM) and the difference between the two samples (Log_2 fold change).

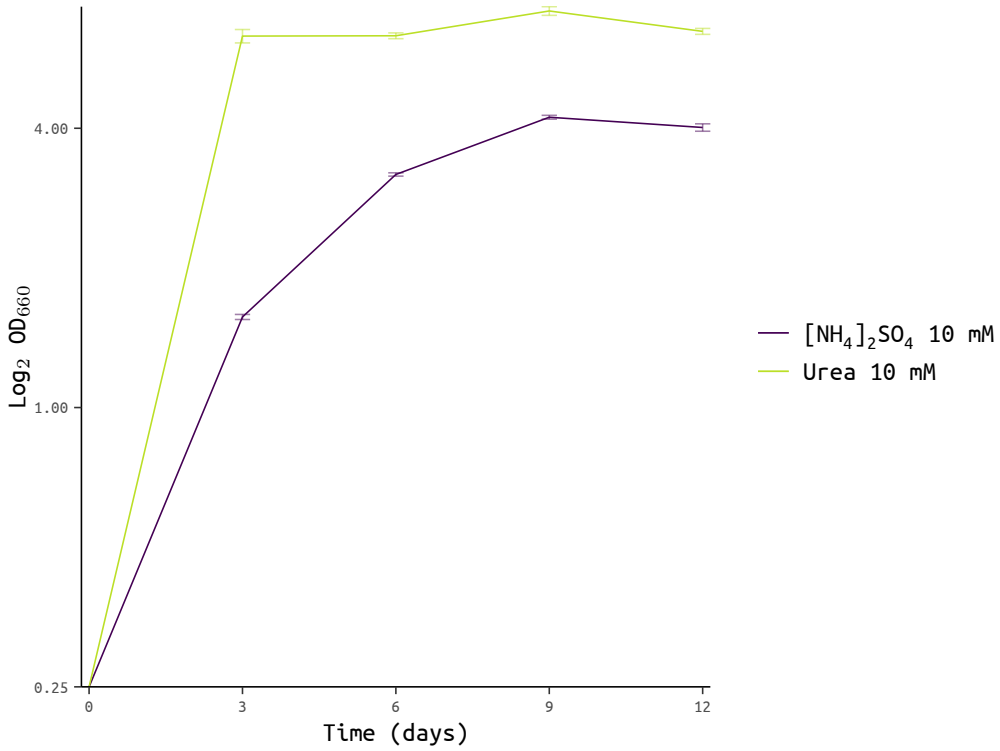


Figure 5.1. – Growth of *R. palustris* with Urea and $[NH_4]_2SO_4$. This shows the growth of WT *R. palustris* with either 10 mM $[NH_4]_2SO_4$ or 10 mM Urea as the nitrogen source. The mean of three biological replicates is plotted, with the error bars showing the SEM. Samples were harvested at Day 3 for RNA extraction.

A heatmap of the top 100 most significantly (lowest adjusted p values) differentially expressed genes was constructed by clustering the genes according to their Euclidean distances and the replicates/samples according to their means. The rows were scaled by Z-score.

5.3. Results

5.3.1. Growth of *R. palustris* for RNA samples

Figure 5.1 shows the growth of the *R. palustris* cultures using either 10 mM urea or 10 mM $[NH_4]_2SO_4$ as the nitrogen source. The cultures given urea grew faster than the cultures given $[NH_4]_2SO_4$ as the nitrogen source, achieving stationary phase within 3 days, compared to ~6 days for the cultures given urea as the nitrogen source.

Additionally, the $[NH_4]_2SO_4$ supplemented cultures displayed a colour change, compared to the urea supplemented cultures, gradually developing a grey hue, and the pH at the end of the growth curve was found to have dropped from ~7 to ~5. Similar observations were noted in previous experiments (Chapter 3.3.1).

Table 5.1. – RNA extraction yields and RIN. Total RNA yield was calculated from the concentrations of samples diluted 1:50, reported by the Bioanalyser software. Where RIN could not be calculated, it is indicated by 'NA'.

Sample	Yield (µg)	RIN
$[NH_4]_2SO_4$ 1	1.2	NA
$[NH_4]_2SO_4$ 2	1.3	NA
$[NH_4]_2SO_4$ 3	1.9	8.9
$[NH_4]_2SO_4$ 4	1.6	NA
Urea 1	5.6	8.0
Urea 2	5.8	NA
Urea 3	3.2	7.7
Urea 4	5.1	7.5

5.3.2. RNA extraction

RNA was extracted from samples taken from the third day of the growth curve. 0.3 mL of culture was taken for the extractions from the $[NH_4]_2SO_4$ supplemented cultures and 1 mL of culture was taken for the extractions from the urea supplemented cultures, due to the different densities of the cultures. Table 5.1 shows the total RNA yields and RIN (RNA integrity number) obtained from each sample, and Figure 5.2 shows the Bioanalyser electropherograms.

The minimum RIN observed was 7.5, and the maximum 8.9. The RIN of some of the samples could not be automatically calculated by the Bioanalyser software (NA) (Table 5.1). However, comparing the electropherogram traces (Figure 5.2) between these, and those that were automatically calculated showed very similar profiles and thus these samples were deemed to be of suitable quality for library preparation.

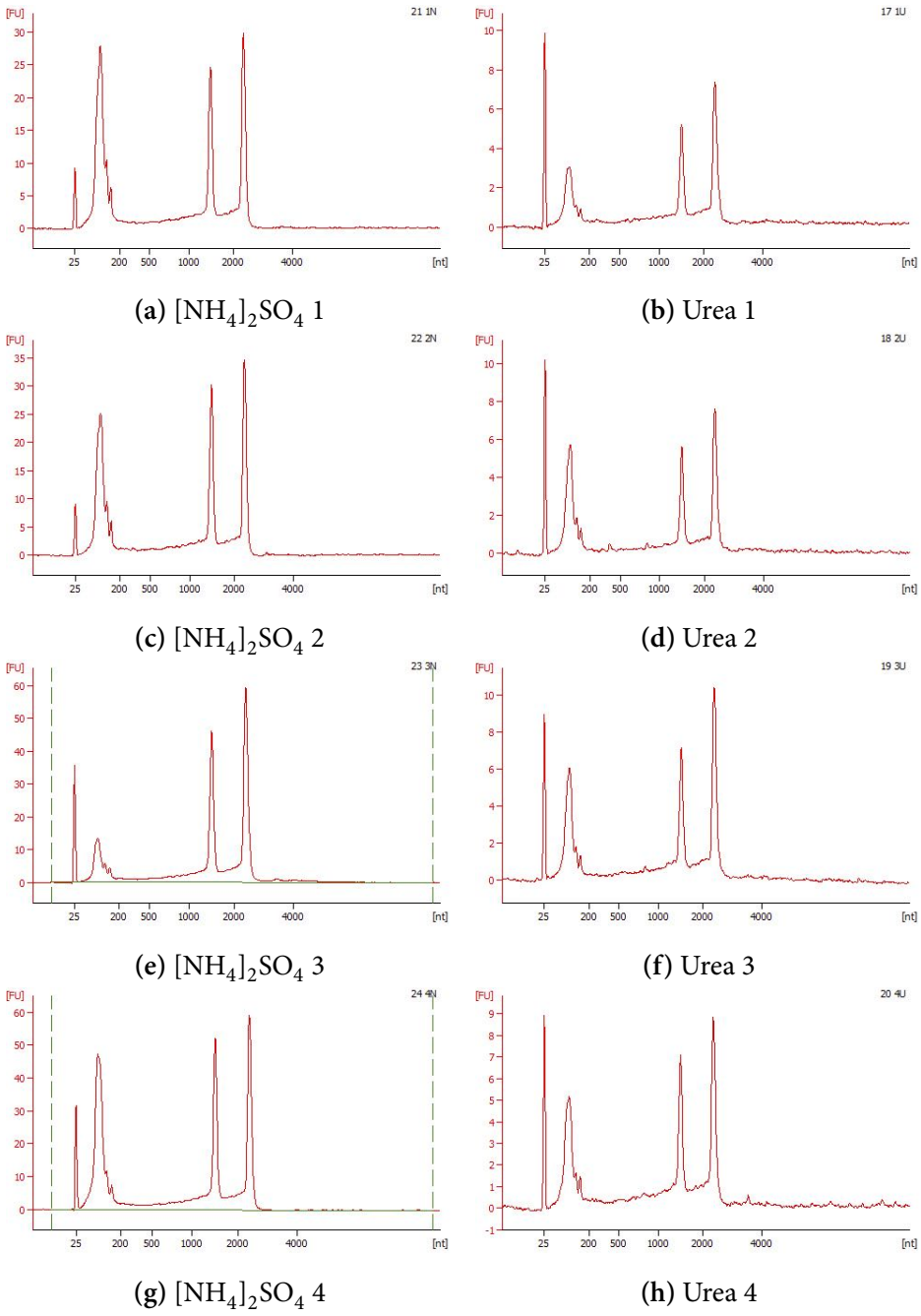
5.3.3. Sequencing quality control and diagnostic plots

Read Quality

Approximately 400 million total reads were expected from the Illumina sequencing reaction, or about 50 million reads per library ($400/8 = 50$). However, the actual number of reads obtained per library was only ~25 to 30 million (Table 5.2).

The only FastQC metrics that the libraries failed on were the 'Per sequence GC content' and 'Per base sequence content' tests. Figures 5.3 and 5.4 show the per sequence GC content and per base sequence content respectively for untrimmed reads, and Figures 5.5 and 5.6 show the data for trimmed reads.

In all eight libraries, two peaks were observed in the GC content distributions, one at ~60 % to 70 % which is expected for *R. palustris*, and another peak at ~10 % (Figure 5.3). The 'Per base sequence content' tests showed a strong positive bias towards thymine, suggesting an over-abundance of this nucleotide in a significant proportion of the reads (Figure 5.4). Read trimming had very little effect on either of these tests (Figures 5.5 & 5.6).



(a) $[NH_4]_2SO_4$ 1 **(b)** Urea 1

(c) $[NH_4]_2SO_4$ 2 **(d)** Urea 2

(e) $[NH_4]_2SO_4$ 3 **(f)** Urea 3

(g) $[NH_4]_2SO_4$ 4 **(h)** Urea 4

Figure 5.2. – Bioanalyser electropherograms of extracted RNA samples. These electropherograms, from Bioanalyser micro-chip RNA-Pico electrophoresis, show the chromatograms of the extracted RNA samples. The assay performed was the Prokaryote Total RNA Pico. The two sharp peaks at ~1542 nt and ~2906 nt represent the 16S and 23S ribosomal RNAs respectively; the ratio between these peaks is important when calculating the RIN.

These results taken together indicate a high level of poly-T rich reads in the libraries.

Table 5.2 shows the number of reads from each sample that aligned to the genome. The alignment rate was quite low, only ~25 to 30 million reads per library mapped to the genome (Table 5.2).

After the alignment, the mapped and unmapped reads were extracted from the .bam alignment files using Samtools (H. Li *et al.*, 2009) and FastQC analysis was performed on the mapped and unmapped reads independently. Figures 5.7 and 5.8 show the per sequence GC content and per base sequence content respectively for the unmapped reads, and Figures 5.9 and 5.10 show the data for the mapped reads.

The 'Per sequence GC content' and 'Per base sequence content' tests of the mapped and unmapped reads, show that the unmapped reads had a low GC content (peak centred ~10 %, Figure 5.7) and strong positive thymine bias (Figure 5.8), suggesting that these reads consisted primarily of poly-T rich sequences. The mapped reads showed a normal GC distribution centred at 60 % to 70 %, and a normal per base sequence content, with slightly elevated guanine and cytosine levels, as expected in a GC rich organism like *R. palustris* (Figures 5.9 & 5.10).

The extracted unmapped reads appear to consist of primarily poly-T and poly-AT sequences. BLAST searches of these reads against the non-redundant RefSeq database did not produce any significant matches.

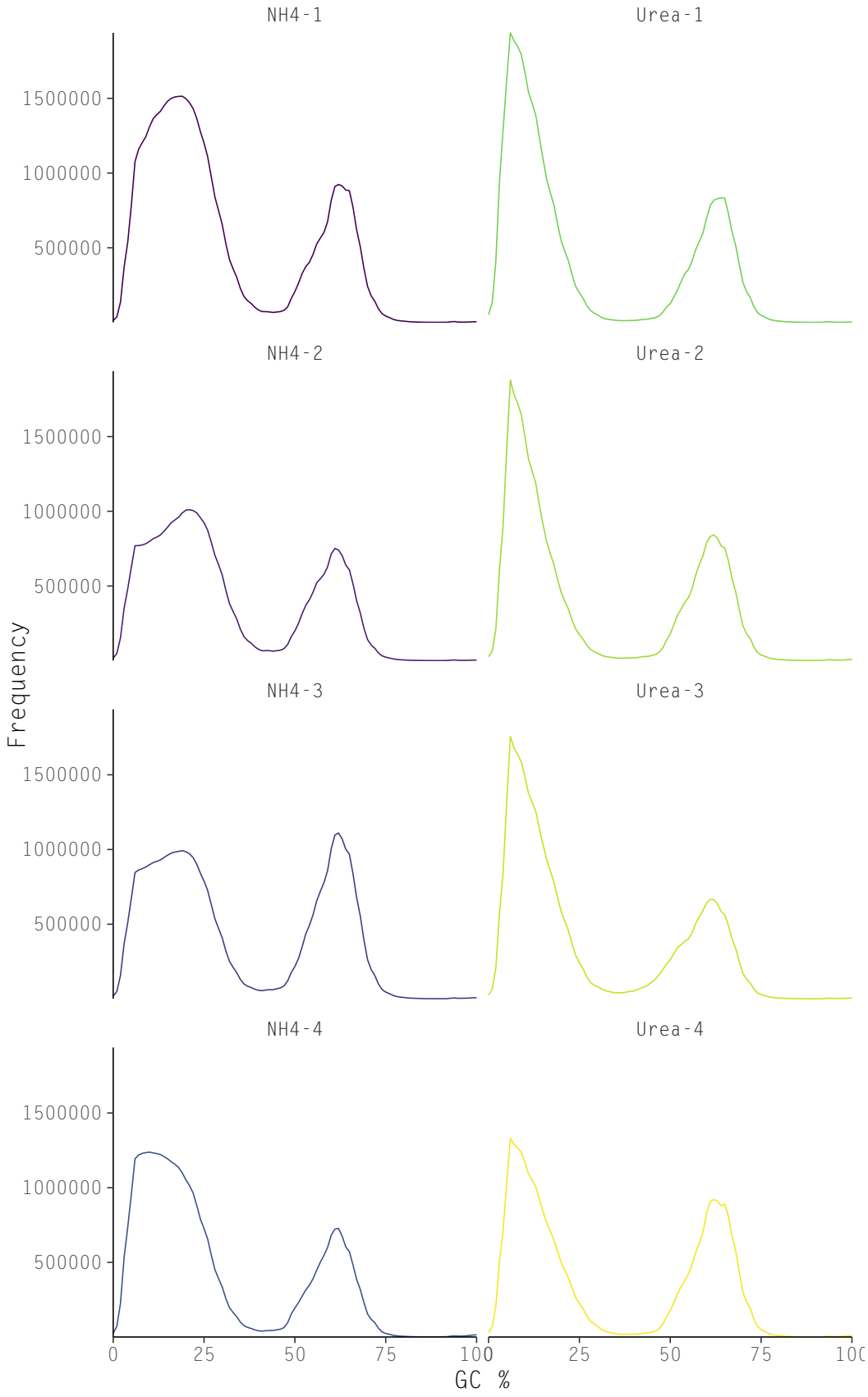
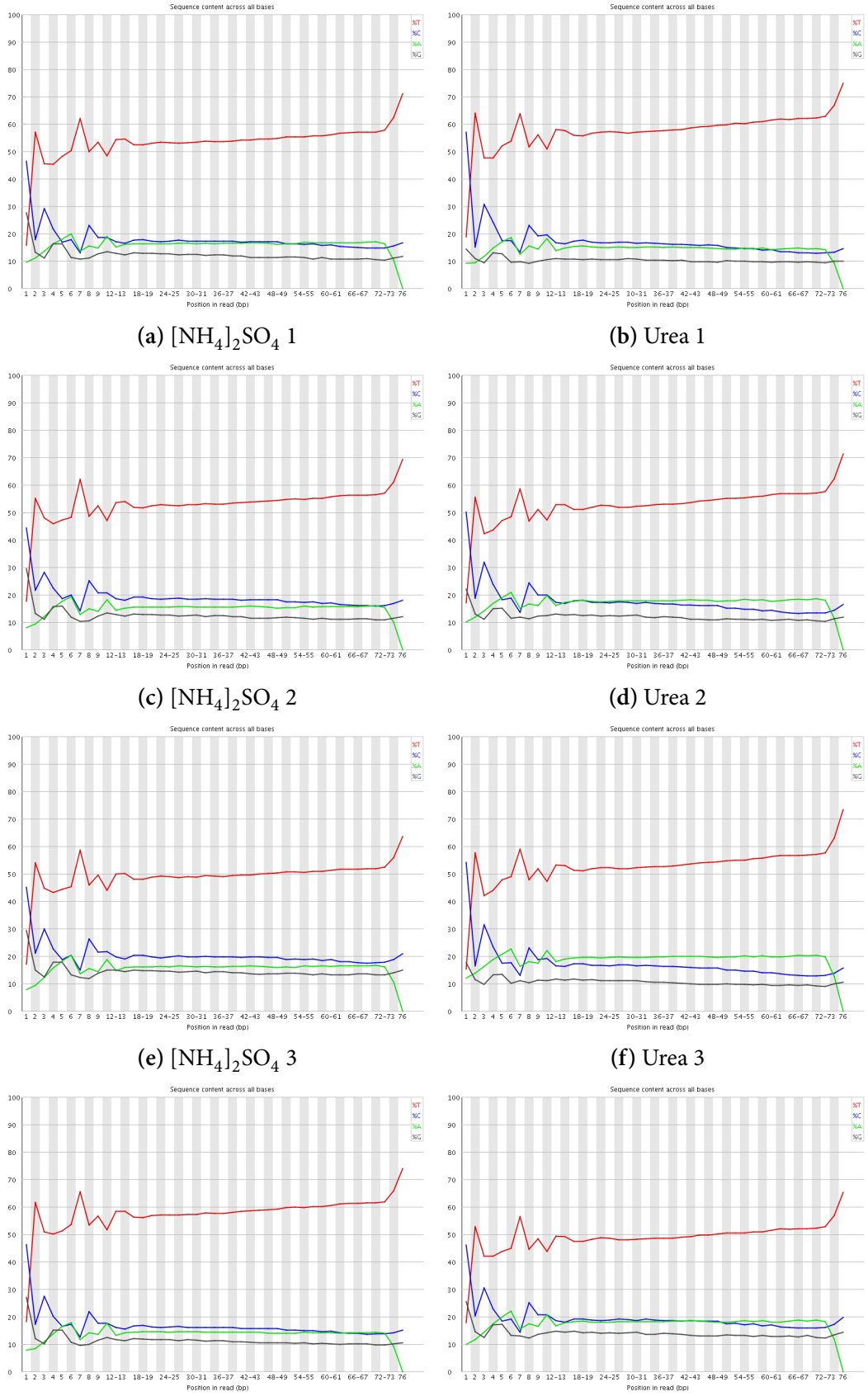


Figure 5.3. – Per sequence GC content of untrimmed reads. Distribution of the GC % of all of the reads in each library.



(a) $[NH_4]_2SO_4$ 1 **(b)** Urea 1

(c) $[NH_4]_2SO_4$ 2 **(d)** Urea 2

(e) $[NH_4]_2SO_4$ 3 **(f)** Urea 3

(g) $[NH_4]_2SO_4$ 4 **(h)** Urea 4

Figure 5.4. – Per base sequence content of untrimmed reads. Proportion of each nucleotide in all reads in each library, at every position in the read. Guanine (G) is shown in grey, cytosine (C) in blue, adenine (A) in green, and thymine (T) in red.

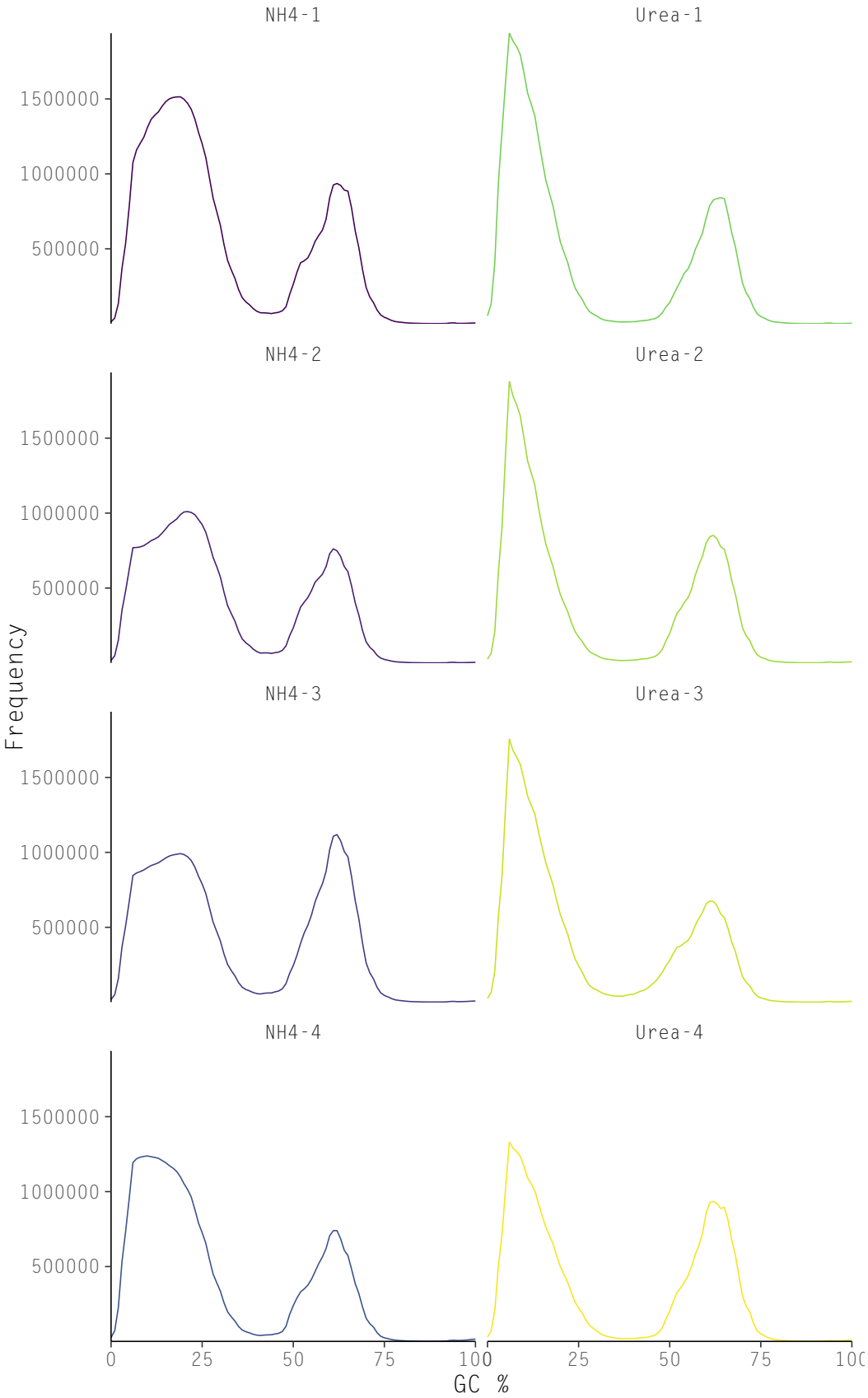
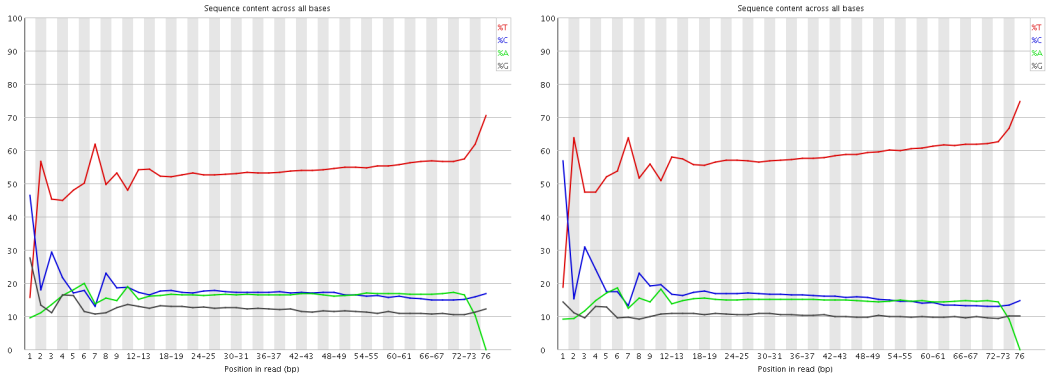
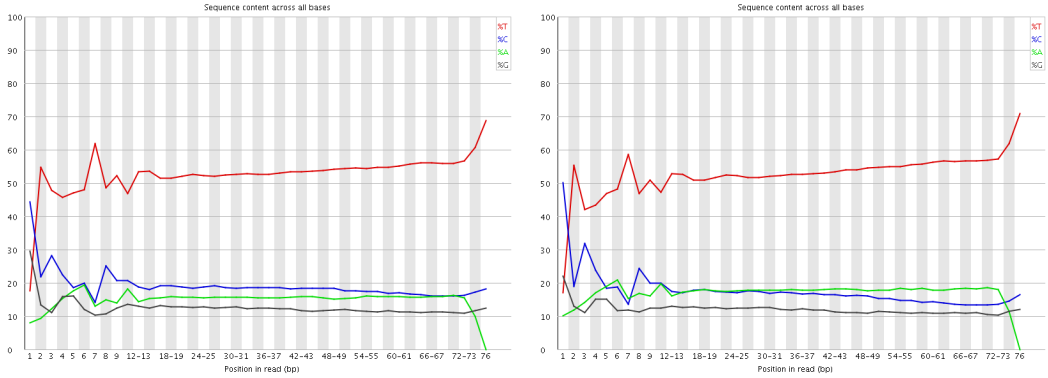


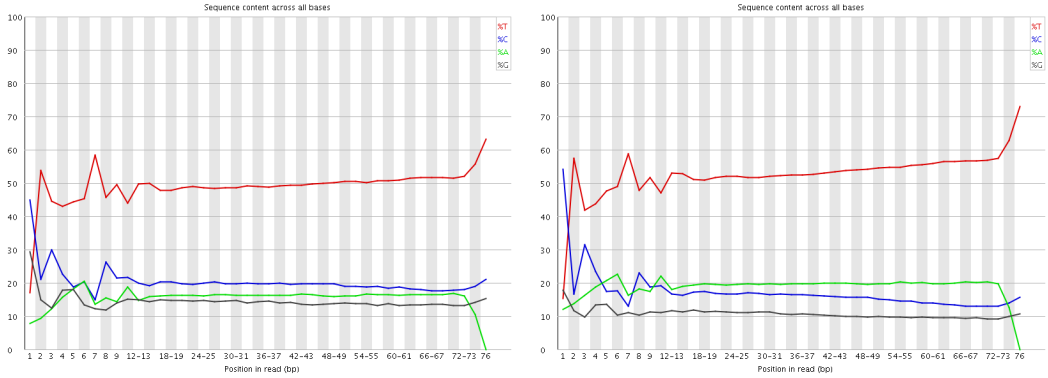
Figure 5.5. – Per sequence GC content of trimmed reads. Distribution of the GC % of all of the reads in each library, after trimming low quality sequences and sequencing adapters with TrimGalore! software.



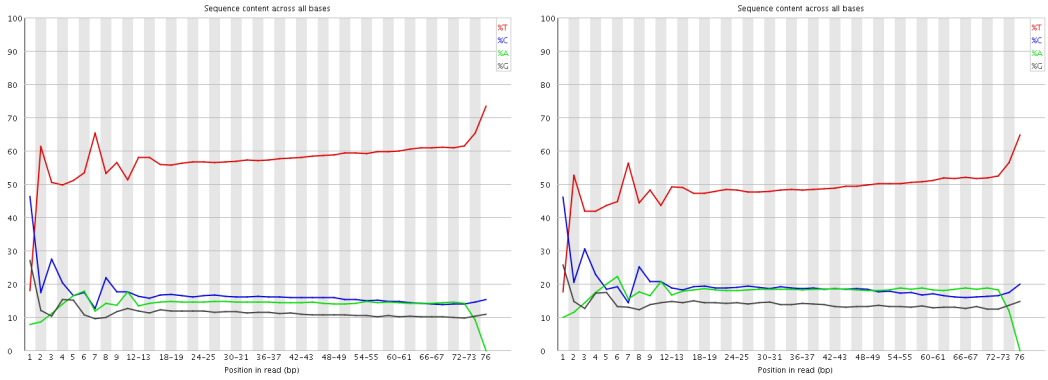
(a) $[NH_4]_2SO_4$ 1 **(b)** Urea 1



(c) $[NH_4]_2SO_4$ 2 **(d)** Urea 2



(e) $[NH_4]_2SO_4$ 3 **(f)** Urea 3



(g) $[NH_4]_2SO_4$ 4 **(h)** Urea 4

Figure 5.6. – Per base sequence content of trimmed reads. Proportion of each nucleotide in all reads in each library, after trimming low quality sequences and sequencing adapters with TrimGalore! software, at every position in the read. Guanine (G) is shown in grey, cytosine (C) in blue, adenine (A) in green, and thymine (T) in red.

Table 5.2. – RNA sequencing read alignment rates. Total number of reads, and proportion of reads that mapped to the *R. palustris* genome using HISAT2, for each library.

Library	Total reads	Reads mapped once	Reads mapped more than once	Unmapped reads	Percentage reads mapped once	Percentage reads mapped more than once	Percentage unmapped reads	Total alignment rate (%)
$[NH_4]_2SO_4$ 1	37,440,173	8,811,682	419,405	28,209,086	23.54	1.12	75.34	24.66
$[NH_4]_2SO_4$ 2	27,461,303	7,226,148	290,652	19,944,503	26.31	1.06	72.63	27.37
$[NH_4]_2SO_4$ 3	30,276,121	10,658,809	235,608	19,381,704	35.21	0.78	64.02	35.98
$[NH_4]_2SO_4$ 4	29,159,224	6,832,238	324,565	22,002,421	23.43	1.11	75.46	24.54
Urea 1	28,950,671	8,207,051	140,400	20,603,220	28.35	0.48	71.17	28.83
Urea 2	26,568,965	8,448,714	180,664	17,939,587	31.80	0.68	67.52	32.48
Urea 3	26,679,217	6,424,275	146,171	20,108,771	24.08	0.55	75.37	24.63
Urea 4	25,058,244	9,303,849	246,382	15,508,013	37.13	0.98	61.89	38.11

Read alignment and quantification

After aligning to the genome, 7,475 transcripts were assembled, which mapped to 6,484 genes. 5,849 genes had only one transcript mapped, and 635 had more than one (Figure 5.11(a)). The majority of transcripts were 400 to 2000 base pairs long (Figure 5.11(b)).

Figure 5.12 shows the distribution of FPKM values for each sample after Log_2 transformation. The variance, median, and distribution were similar across all samples, showing a normal distribution.

Figure 5.13 shows the results of MDS analysis of the Pearson correlation of the Log_2 FPKM values of all samples, and Figure 5.14 shows a heatmap of the distance between each sample based on the Pearson correlation of the Log_2 FPKM values, and clustered according to the Euclidean distance. Figure 5.15 shows a PCA of the Log_2 FPKM values of all of the samples. These plots show that the four biological replicates of each sample clustered closely together, and fell into two distinct groups corresponding to the two sample types ($[NH_4]_2SO_4$ and urea). The majority of the variation in the samples was explained by the sample type, rather than variation between replicates. There was also more variation between the $[NH_4]_2SO_4$ replicates than there was between the urea replicates.

5.3.4. Differential gene expression

535 significantly differentially expressed genes were identified, using an adjusted $p < 0.05$, and, of these 284 had a Log_2 fold change > 2 (unknown transcripts were not considered). Figure 5.16 shows the distribution of the Log_2 fold change values of the significantly differentially expressed genes.

Figure 5.17 shows the correlation between the mean Log_2 FPKM values of both sample groups, with the significantly differentially expressed genes indicated. Figure 5.18 is an MA plot showing the relationship between the Log_2 fold change and the mean Log_2 FPKM value of all genes, with the significantly differentially expressed genes indicated.

A heatmap highlighting the top 100 most significantly differentially expressed genes (lowest p values) is shown in Figure 5.19. A complete list of the 284 significantly differentially expressed genes with adjusted $p < 0.05$, and Log_2 fold change > 2 is given in Appendix B.

A PCA of the Log_2 FPKM values of just the significantly differentially expressed genes is shown in Figure 5.20, showing that almost all (93.8 %) of the variation in these values is captured in the principal component 1. The two sample types cluster tightly together at either end of this axis, suggesting that most of the variation is explained by sample type.

The top 10 most differentially expressed genes (greatest Log_2 fold change) were *nifD*, *fixB*, *nifH4*, RPA4610, *nifX*, *fixC*, *nifK*, *fixA*, *pucBd*, and *nifE*. The *nif* genes are all related to nitrogenase, the *fix* genes are electron transfer flavoproteins (associated with nitrogen fixation), *pucBd* is a subunit of a light-harvesting protein, and RPA4610 is an uncharac-

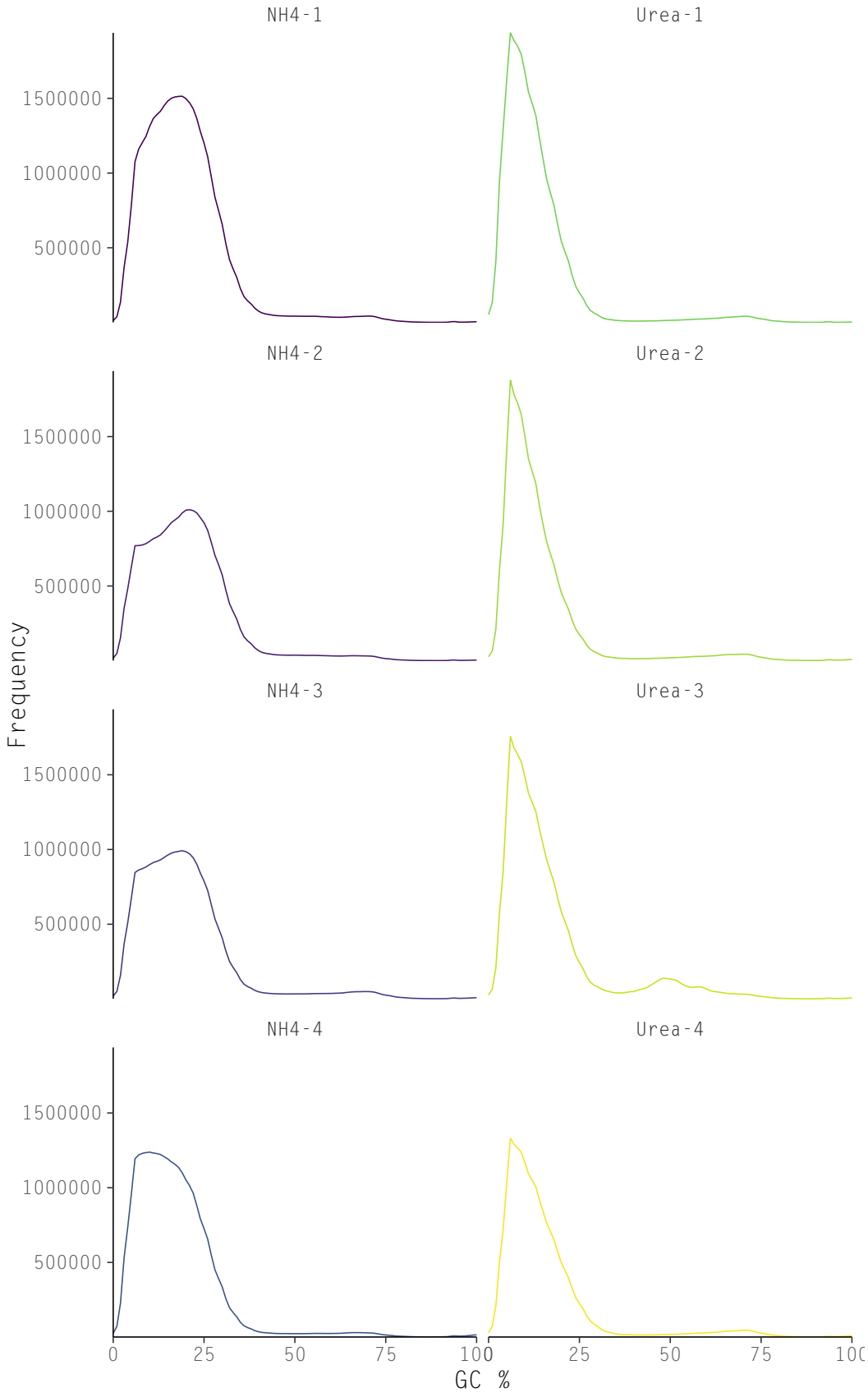
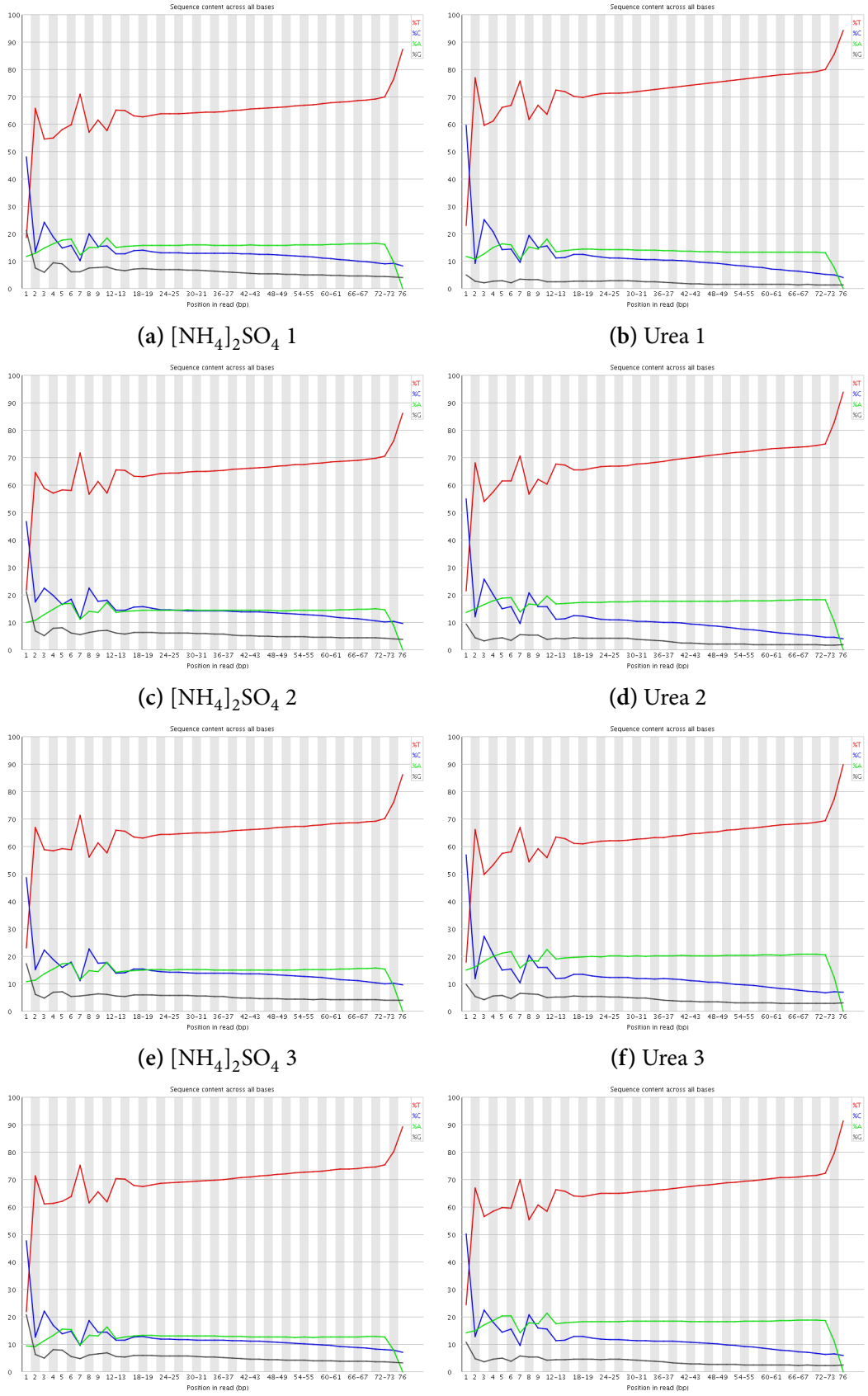


Figure 5.7. – Per sequence GC content of unmapped reads. Distribution of the GC % of only the unmapped reads in each library, that did not align to the *R. palustris* genome.



(a) $[NH_4]_2SO_4$ 1 **(b)** Urea 1

(c) $[NH_4]_2SO_4$ 2 **(d)** Urea 2

(e) $[NH_4]_2SO_4$ 3 **(f)** Urea 3

(g) $[NH_4]_2SO_4$ 4 **(h)** Urea 4

Figure 5.8. – Per base sequence content of unmapped reads. Proportion of each nucleotide in only unmapped reads in each library, at every position in the read. Guanine (G) is shown in grey, cytosine (C) in blue, adenine (A) in green, and thymine (T) in red.

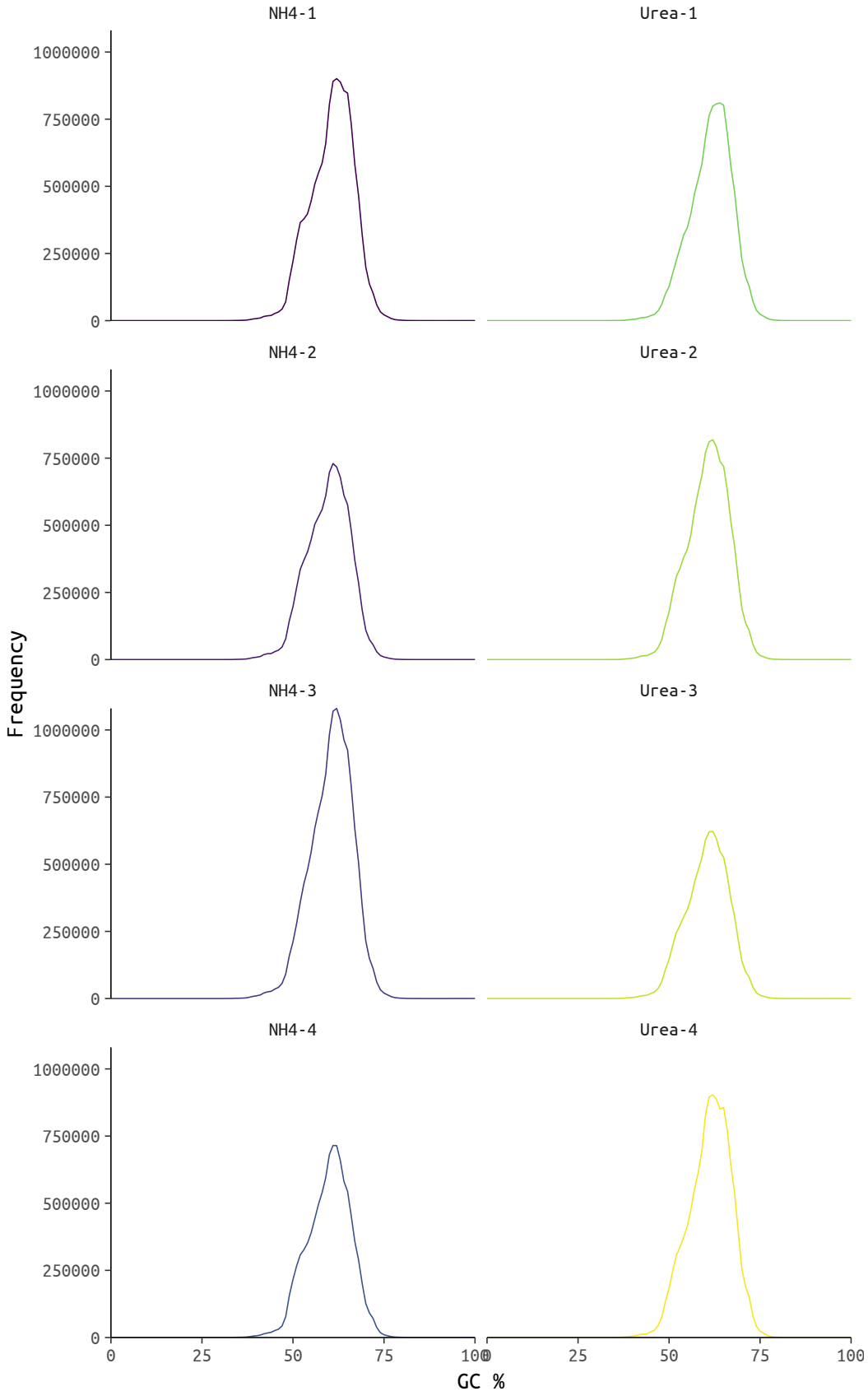
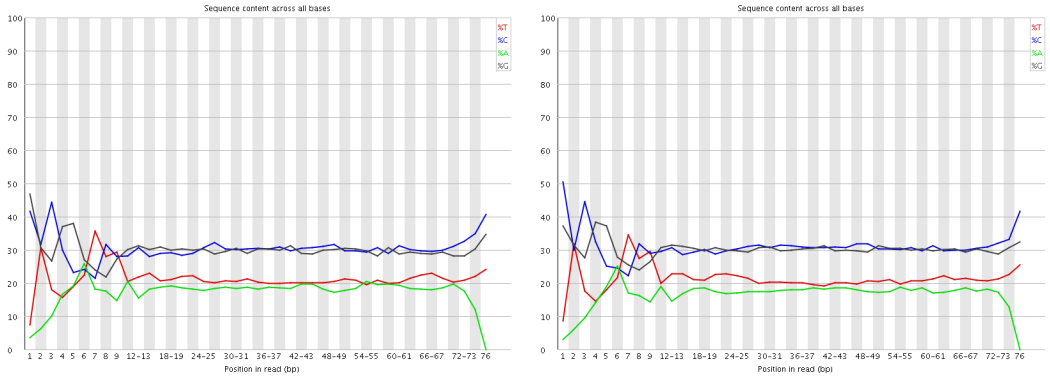
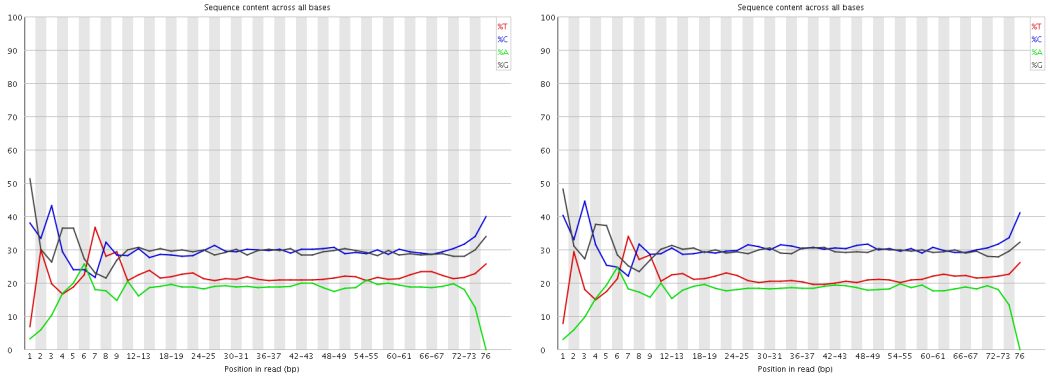


Figure 5.9. – Per sequence GC content of mapped reads. Distribution of the GC % of only the mapped reads in each library, that aligned to the *R. palustris* genome.



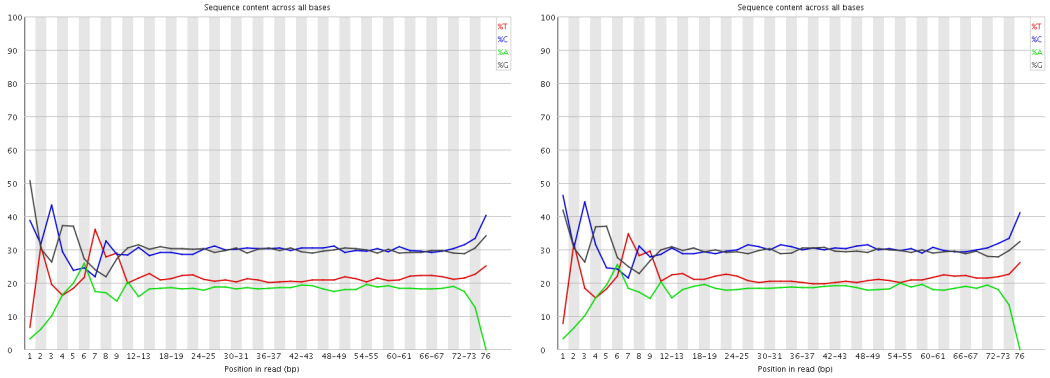
(a) $[NH_4]_2SO_4$ 1

(b) Urea 1



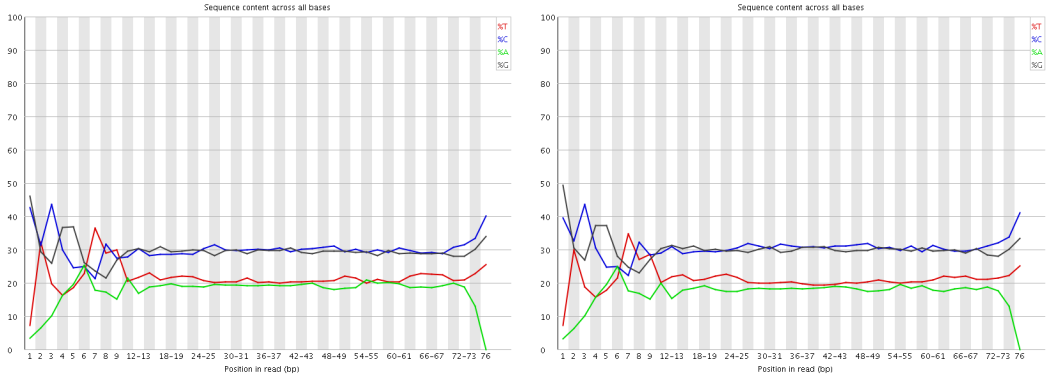
(c) $[NH_4]_2SO_4$ 2

(d) Urea 2



(e) $[NH_4]_2SO_4$ 3

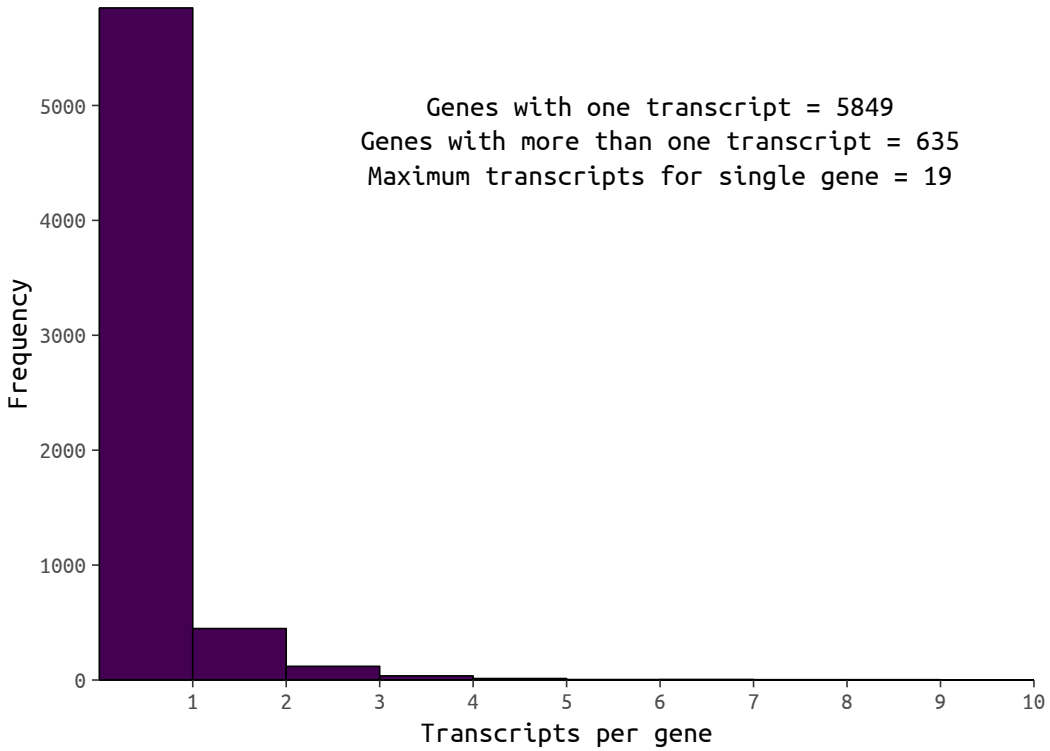
(f) Urea 3



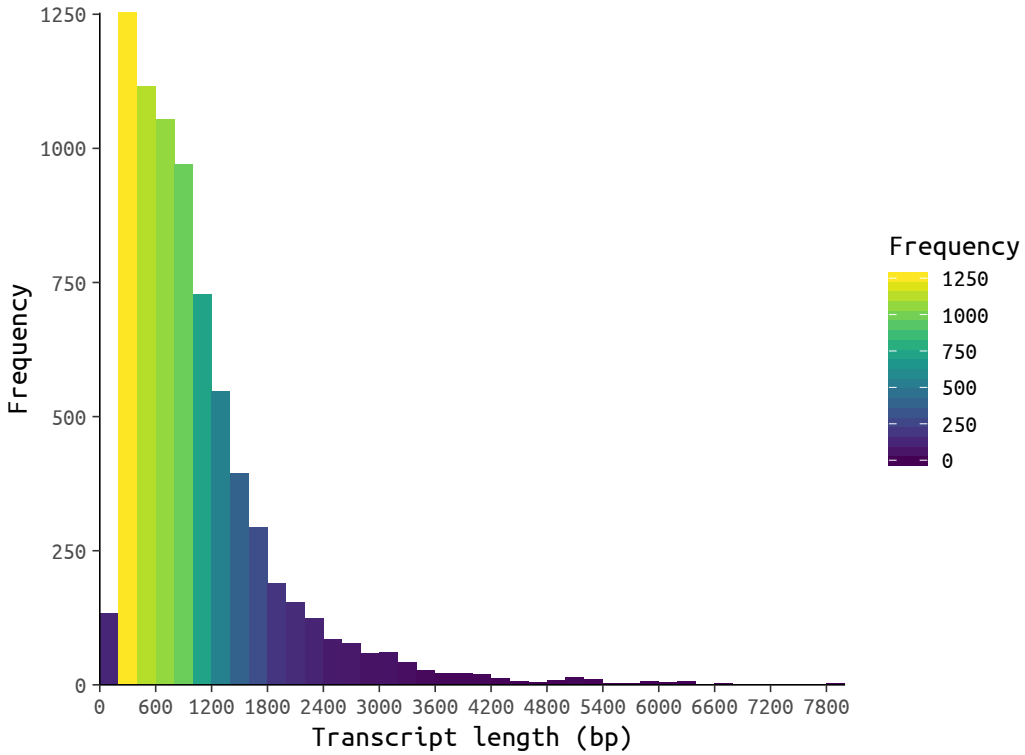
(g) $[NH_4]_2SO_4$ 4

(h) Urea 4

Figure 5.10. – Per base sequence content of mapped reads. Proportion of each nucleotide in only mapped reads in each library, at every position in the read. Guanine (G) is shown in grey, cytosine (C) in blue, adenine (A) in green, and thymine (T) in red.



(a) Number of transcripts per gene



(b) Distribution of transcript lengths

Figure 5.11. – Assembly of transcripts. (a): Number of transcripts, assembled by StringTie to the *R. palustris* genome, per gene (b): Distribution of transcript lengths, assembled by StringTie to the *R. palustris* genome.

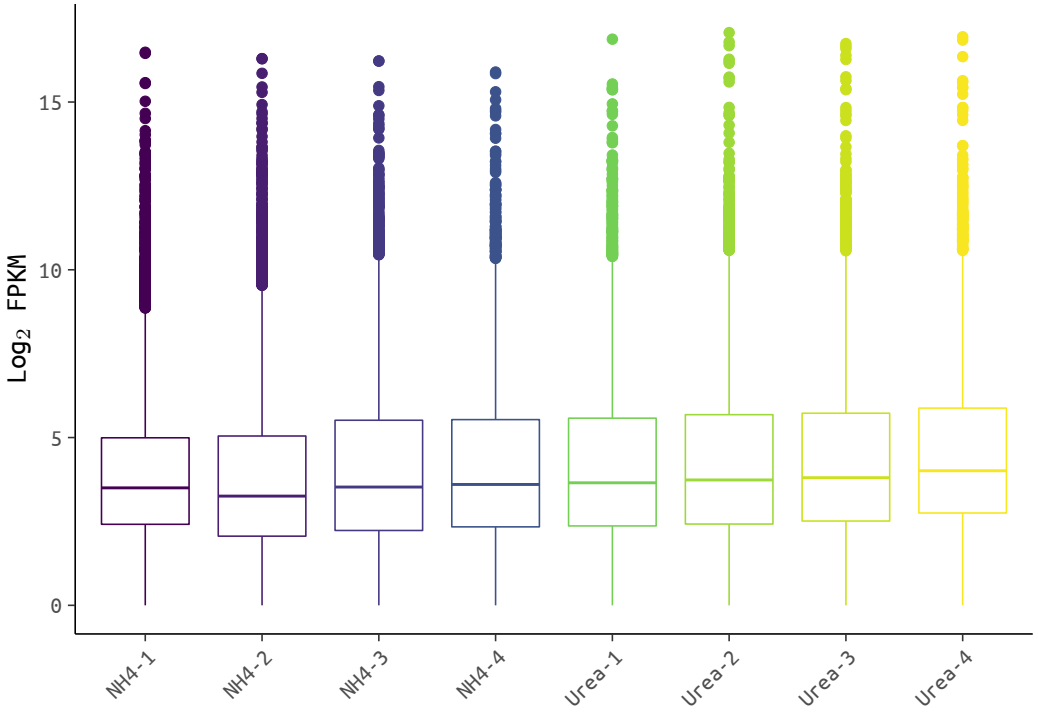


Figure 5.12. – Distribution of Log_2 FPKM values for all libraries.

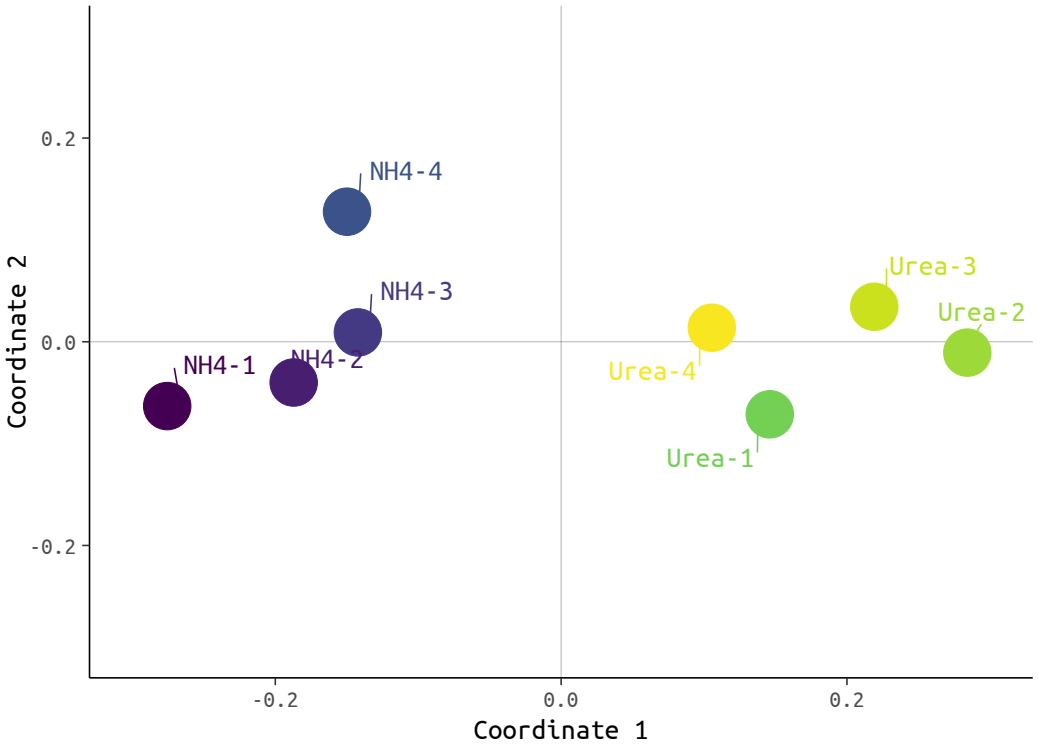


Figure 5.13. – MDS plot. Mean distance between all samples calculated by Pearson correlation of Log_2 FPKM values between all samples and MDS. Biological replicates of each sample type cluster together in two distinct groups.

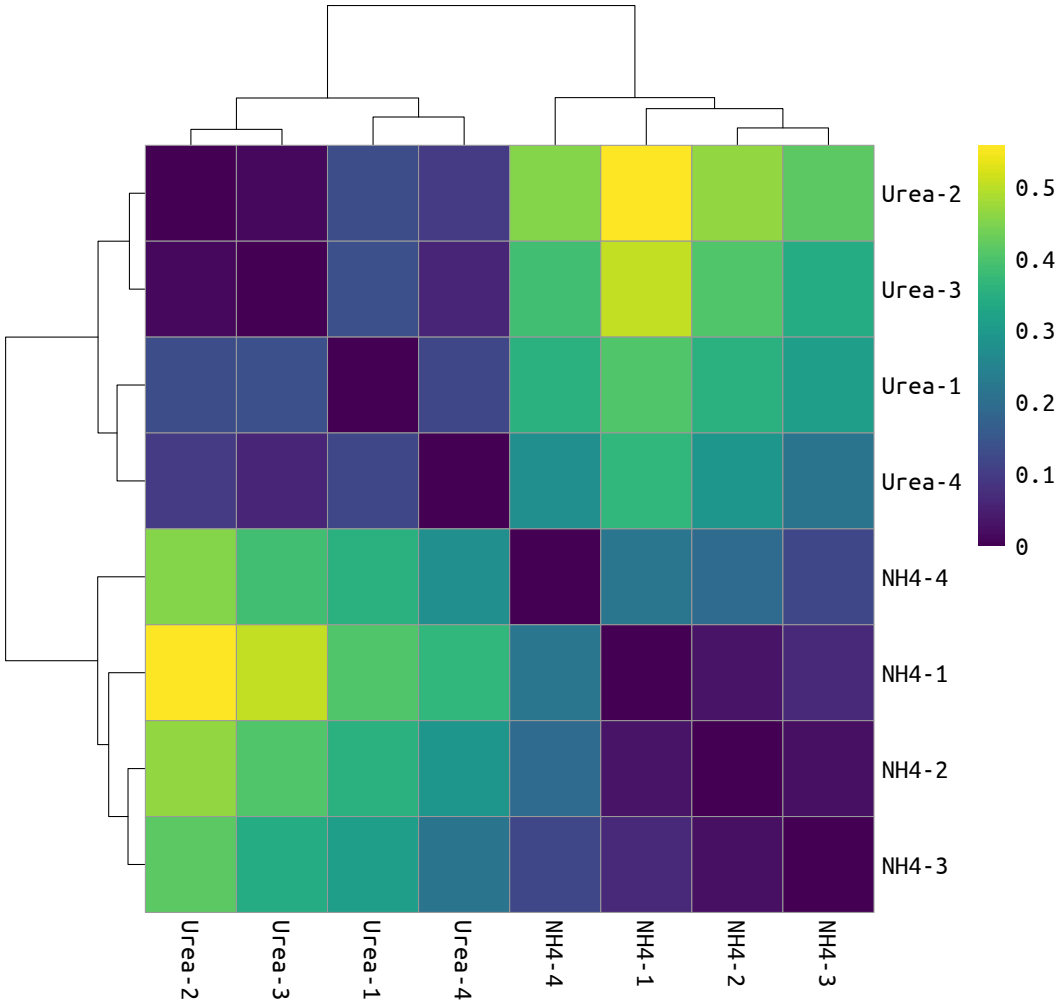


Figure 5.14. – Heatmap of mean sample distances. Mean distance between all samples calculated by Pearson correlation of Log_2 FPKM values between all samples, and clustered by Euclidean distance. This shows the similarity between the biological replicates and dissimilarity between the different sample types.

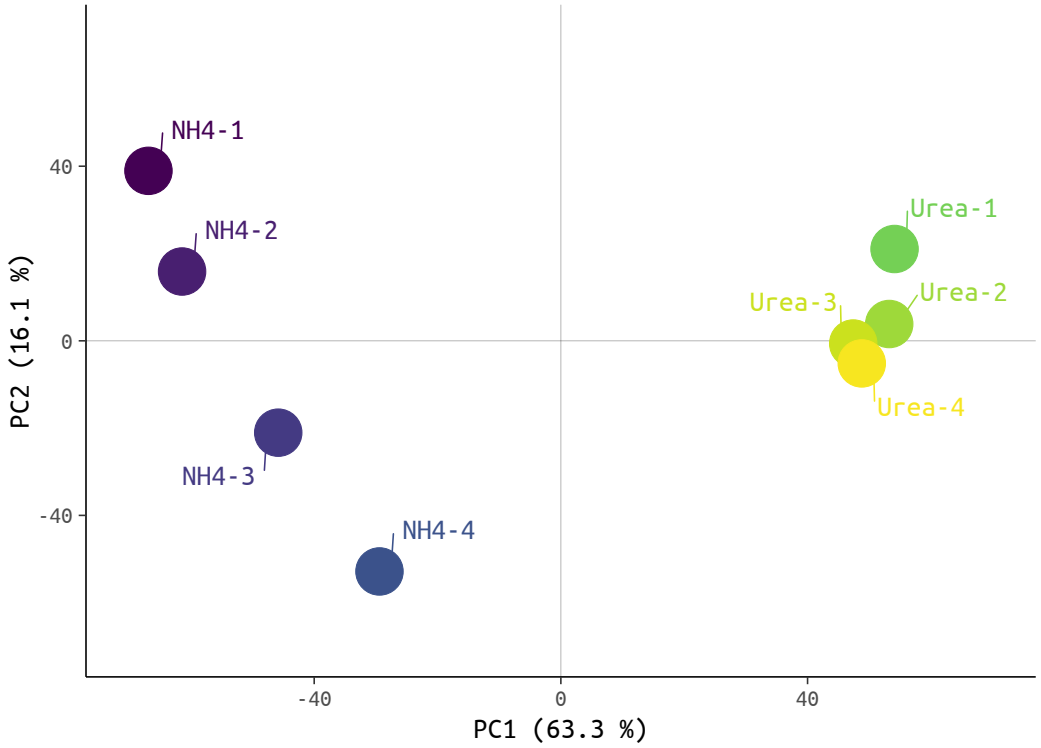


Figure 5.15. – PCA plot of Log_2 FPKM values. Principal component analysis of Log_2 FPKM expression values of all genes. The percentage of variation in the data, captured by each principal component is indicated on the axes.

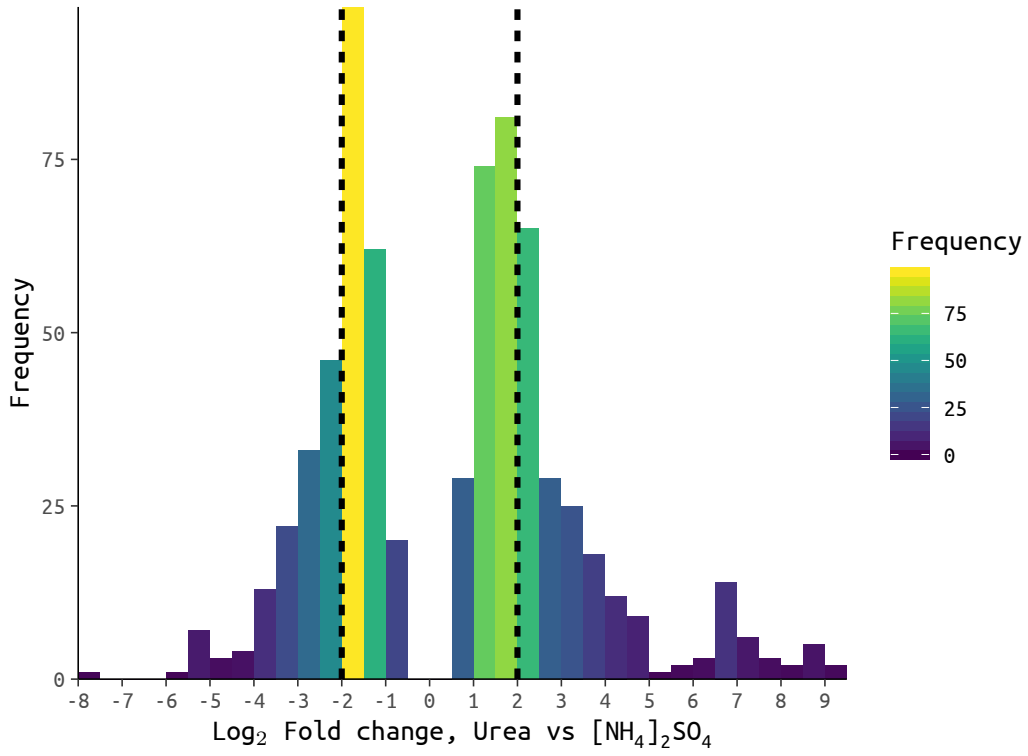


Figure 5.16. – Distribution of Log_2 fold changes. Frequencies of Log_2 fold change values of significantly differentially expressed genes. The dotted lines show the Log_2 fold change of $|2|$ cut-off used when selecting differentially regulated genes.

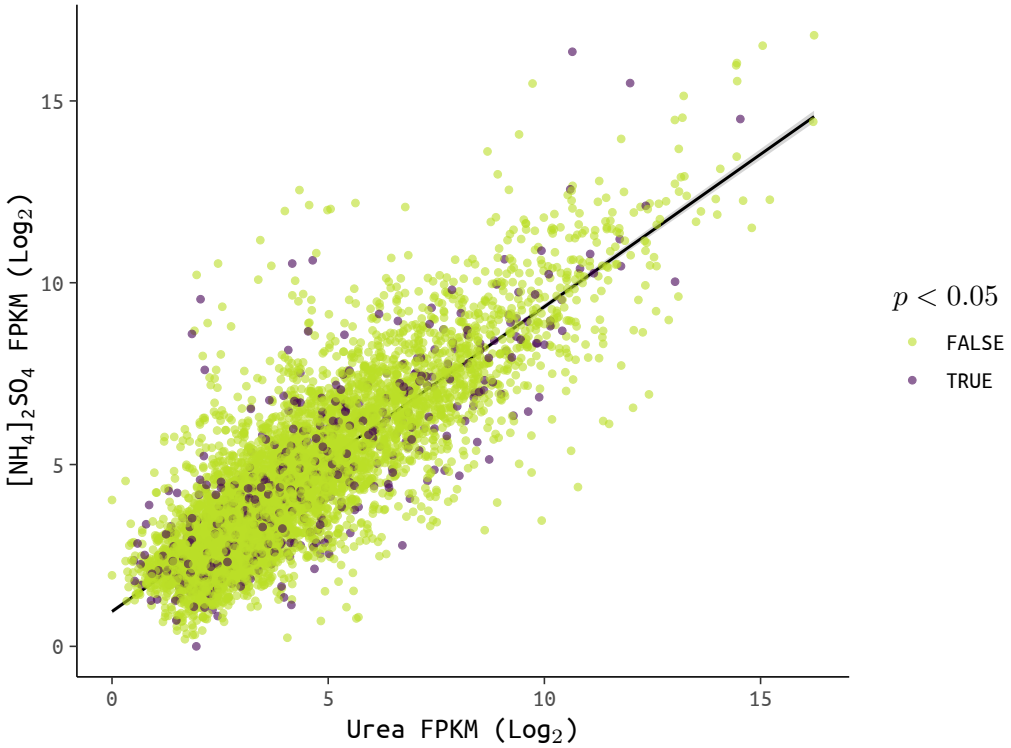


Figure 5.17. – Mean Log_2 FPKM urea *vs.* $[NH_4]_2SO_4$. Mean FPKM gene expression values of the two samples. Significantly differentially regulated genes are highlighted in purple.

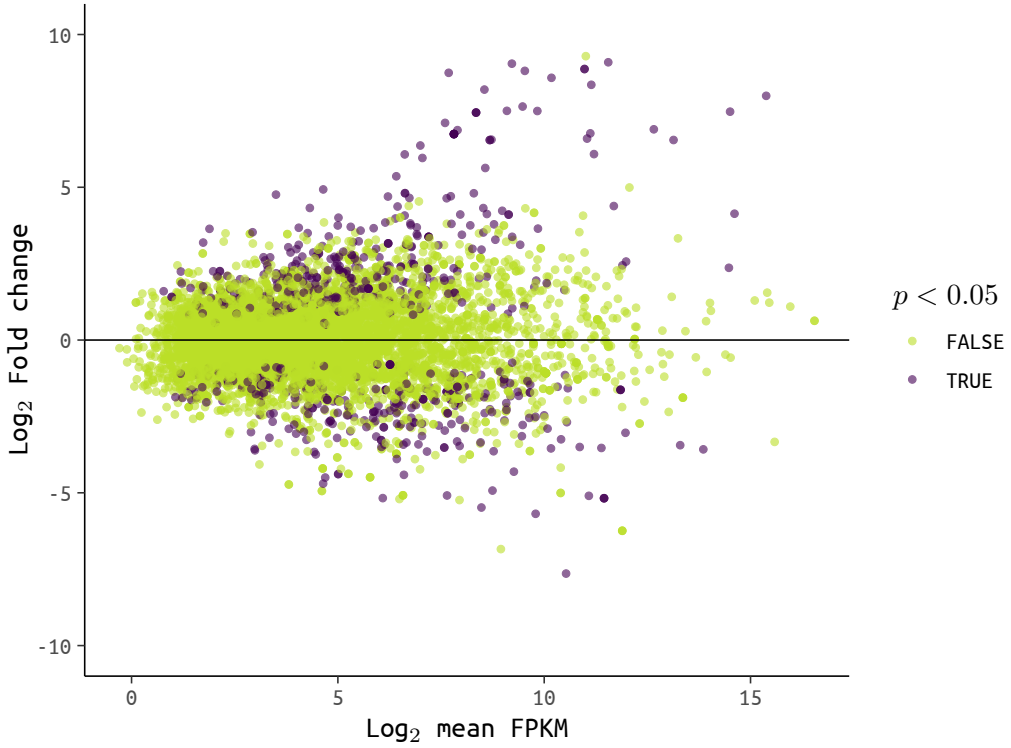


Figure 5.18. – MA plot. Log_2 fold change of all genes plotted against the mean Log_2 FPKM gene expression value. Significantly differentially expressed genes are highlighted in purple.

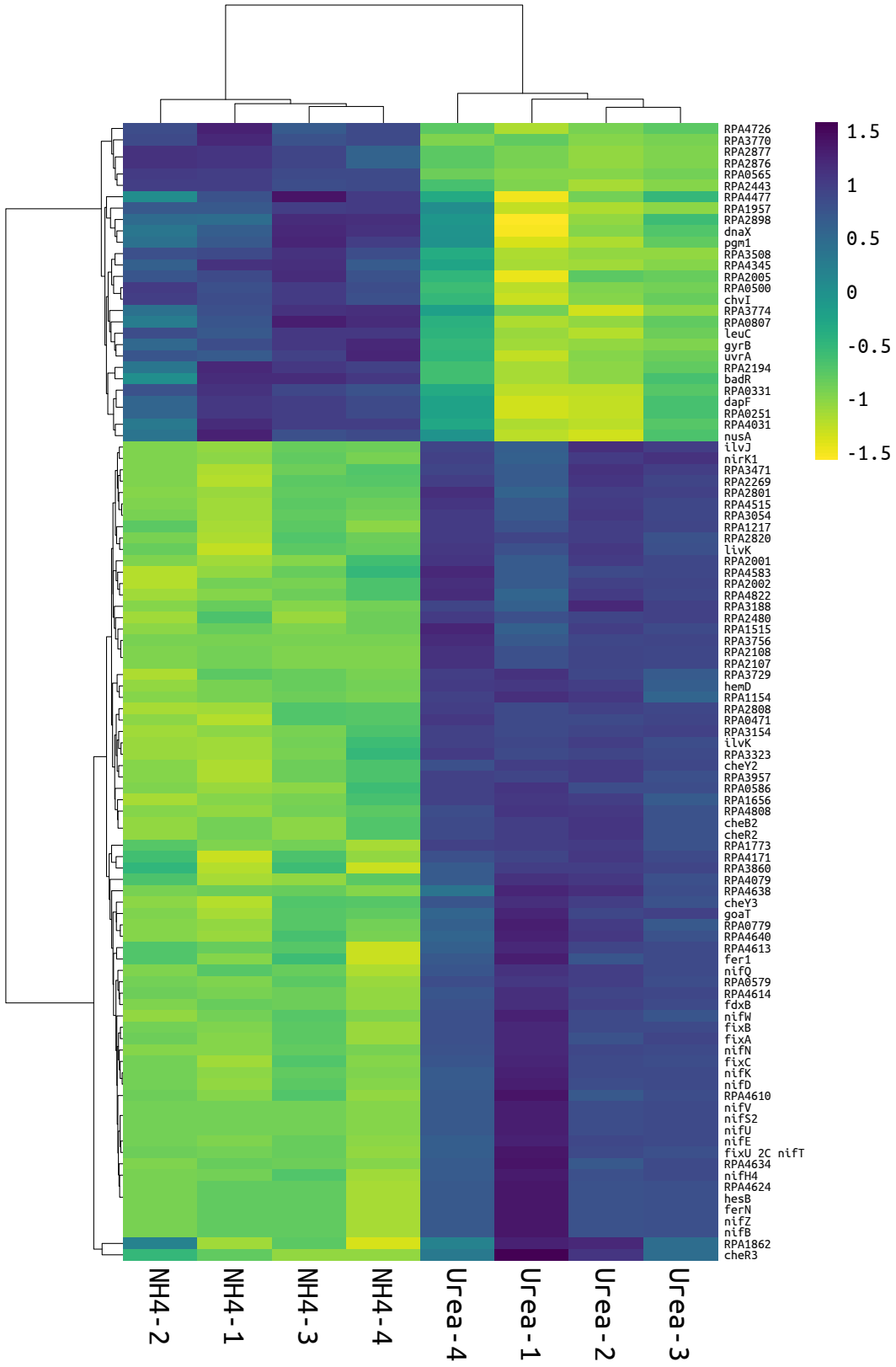


Figure 5.19. – Heatmap of top 100 most significantly differentially expressed genes. Top 100 most significantly differential expressed genes, according to lowest adjusted p-value, with adjusted $p < 0.05$, and Log_2 fold change > 2. Genes (rows), and samples (columns) are clustered according to Euclidean distance and mean. Rows are scaled according to Z-score, such that up-regulated genes (urea, relative to $[NH_4]_2SO_4$) appear in darker (purple) colours and down-regulated genes appear in lighter colours (green/yellow).

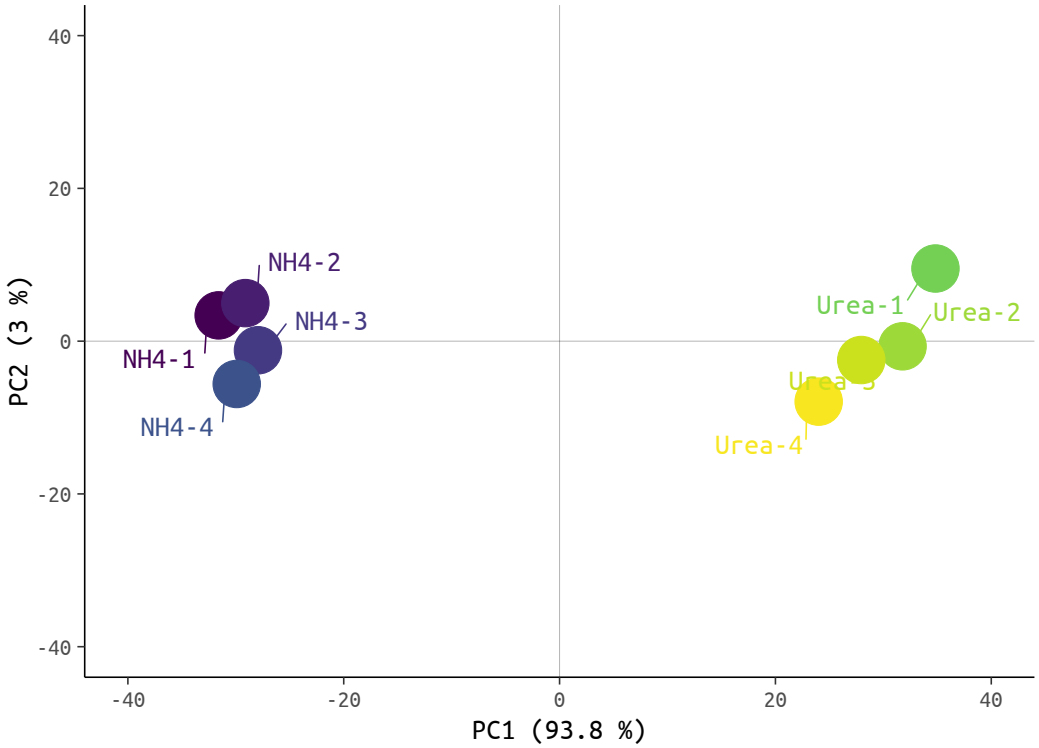


Figure 5.20. – PCA plot of Log_2 FPKM values of significantly differentially expressed genes. Principal component analysis of Log_2 FPKM expression values of only significantly differentially expressed genes. The percentage of variation in the data, captured by each principal component is indicated on the axes.

terised protein, but may belong to the HesB/YadR/YfhF family, which are also associated with nitrogen fixation.

5.3.5. Nitrogen metabolism genes

Nitrogenases

Figure 5.21 shows the Log_2 FPKM gene expression values of the genes on the three nitrogenase variant operons. All but one of the genes on the molybdenum nitrogenase operon are significantly up-regulated in the cultures grown with urea, compared to those grown with $[NH_4]_2SO_4$. The vanadium nitrogenase genes do not appear to be up-regulated, and no significant differential expression is seen, except for *vnfN*, and *vnfX* (both involved in nitrogenase cofactor synthesis), which are specifically down-regulated in the cultures grown with $[NH_4]_2SO_4$. There was no difference observed in the expression of the *anf* iron nitrogenase genes.

Nitrogenase regulation

Figure 5.22 shows the Log_2 FPKM gene expression values of the genes related to the regulation of nitrogen metabolism, including the *ntr* operon which encodes the NtrB/C two-component system, and *draT2* and *glnK2* which encode the nitrogenase post-transcriptional regulators DraT2 and GlnK2. No significant differential gene expression was observed for these genes between the cultures grown with urea, compared to those grown with $[NH_4]_2SO_4$.

Urea metabolism

Figure 5.23 shows the Log_2 FPKM gene expression values of the genes involved in the transport and metabolism of urea. No significant differential gene expression was observed between the cultures grown with urea, compared to those grown with $[NH_4]_2SO_4$, except for *ureG*, which was slightly up-regulated in the cultures grown with $[NH_4]_2SO_4$.

5.3.6. Central metabolism

Figure 5.24 shows the Log_2 FPKM gene expression values of a selection of genes from the citric acid cycle, including citrate synthase (*cisY*), isocitrate lyase (*aceA*), isocitrate dehydrogenase kinase/phosphatase (*aceK*), aconitate hydratase (*acnA*), fumarate hydratase (*fumA*), isocitrate dehydrogenase (*idh*), malate dehydrogenase (*mdh*), succinate dehydrogenase (*sdh*), 2-oxoglutarate dehydrogenase (*sucA, sucB*), and succinyl-CoA synthetase (*sucD, sucC*). No significant differential gene expression was observed for these genes between the cultures grown with urea, compared to those grown with $[NH_4]_2SO_4$.

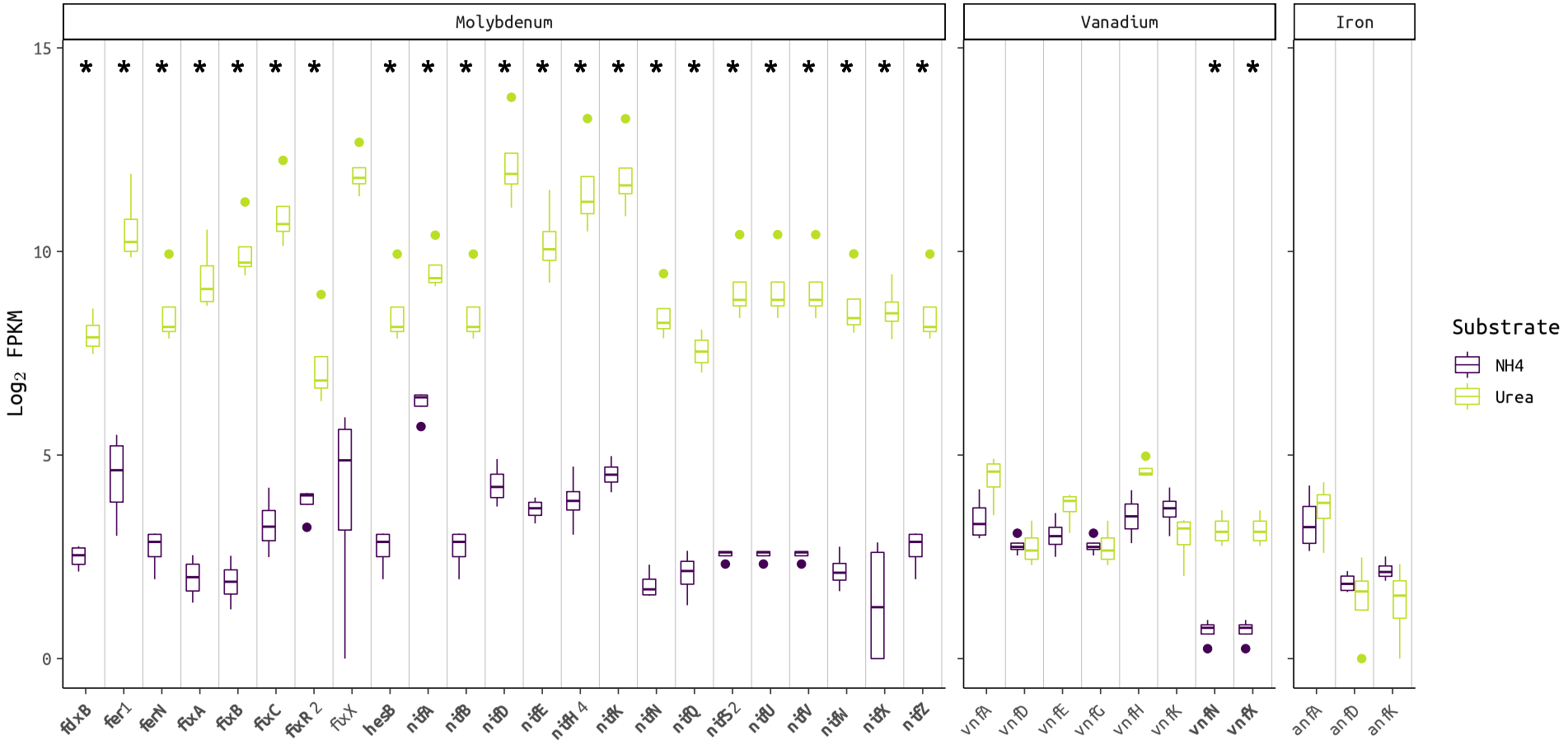


Figure 5.21. – Gene expression values of nitrogenase genes. Log_2 FPKM gene expression values of the known nitrogenase genes, showing the molybdenum, vanadium, and iron variants. Significant differential gene expression between the two nitrogen source conditions for each gene is indicated by a *, and the gene name highlighted in bold.

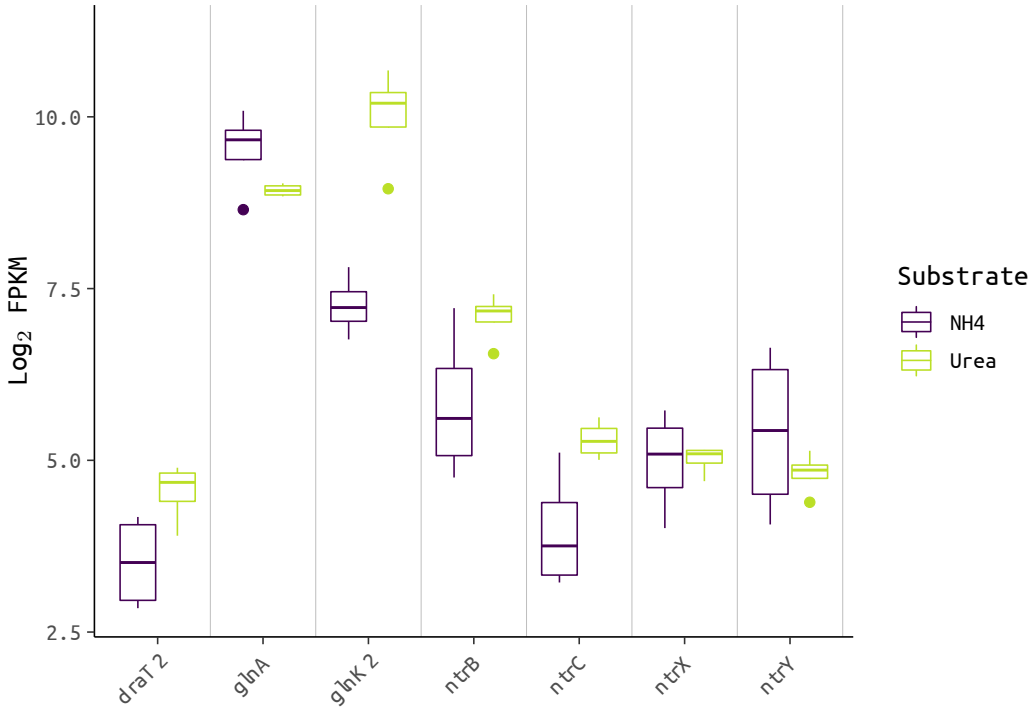


Figure 5.22. – Gene expression values of genes involved in nitrogenase regulation. Log_2 FPKM gene expression values of a selection of genes involved in the transcriptional and translational regulation of nitrogenase. None of these genes were shown to be significantly differentially expressed between the two nitrogen source conditions.

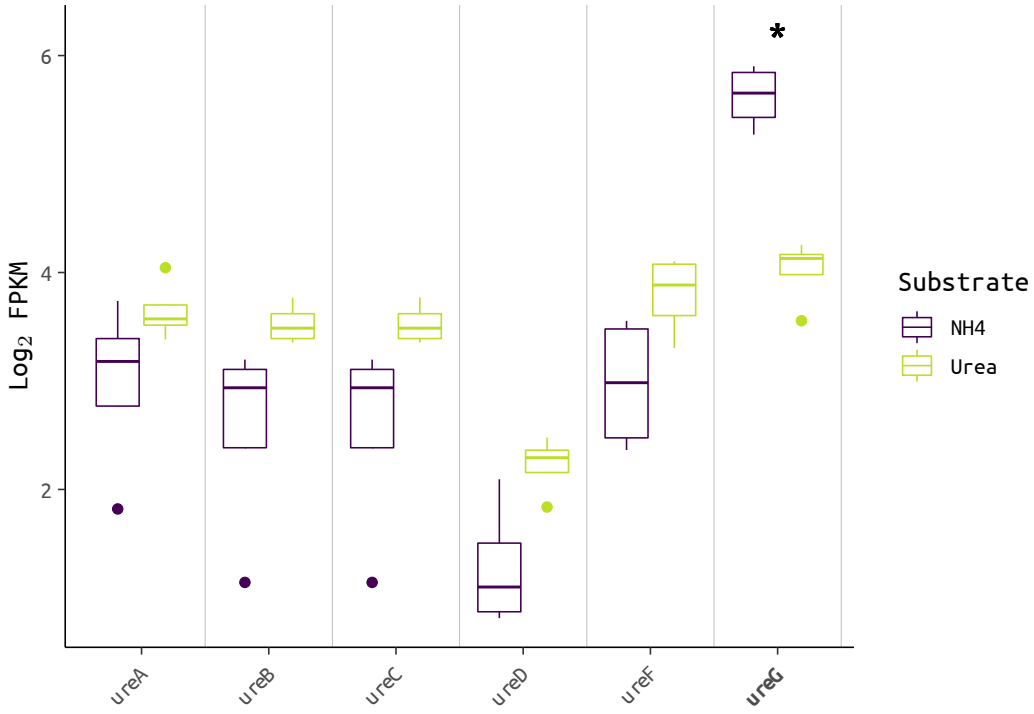


Figure 5.23. – Gene expression values of urea genes. Log_2 FPKM gene expression values of a selection of genes involved in the transport and metabolism of urea. Significant differential gene expression between the two nitrogen source conditions for each gene is indicated by a *, and the gene name highlighted in bold.

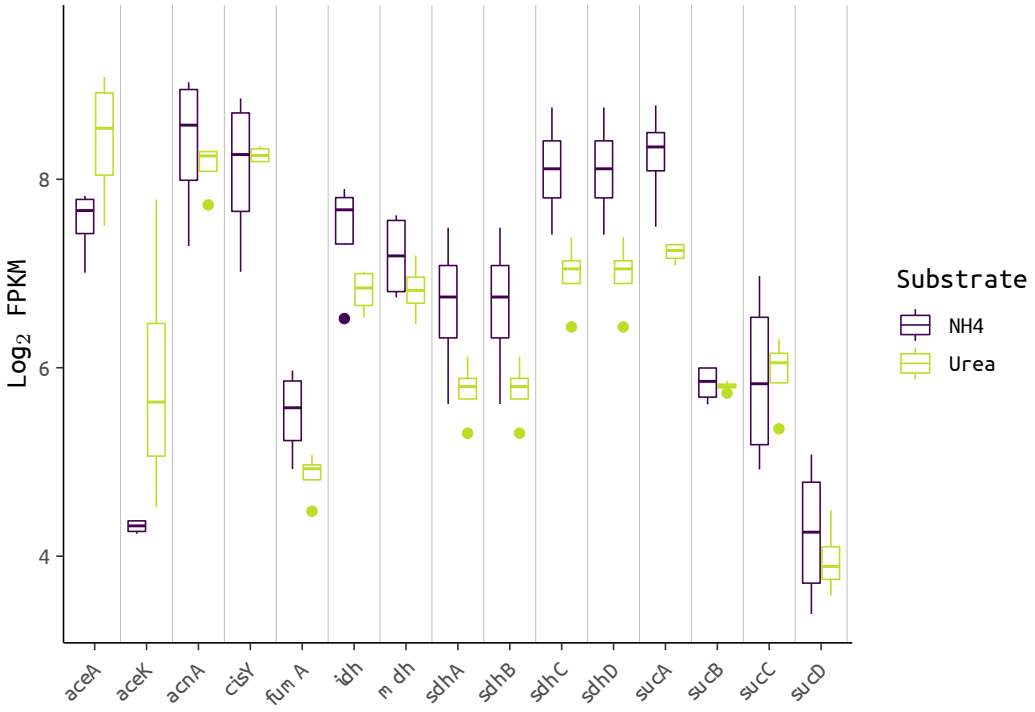


Figure 5.24. – Gene expression values of central metabolism genes. Log_2 FPKM gene expression values of a selection of genes involved in the TCA (tricarboxylic acid) cycle of central metabolism. None of these genes were shown to be significantly differentially expressed between the two nitrogen source conditions.

5.4. Discussion

The results presented show that there is substantial differential gene expression between *R. palustris* cultures grown with urea as the nitrogen source, as compared to using $[NH_4]_2SO_4$ as the nitrogen source.

RNA transcriptome sequencing resulted in transcripts being mapped to 6,484 genes. The *R. palustris* genome contains only 4,836 predicted genes (Larimer *et al.*, 2004), suggesting that some of the genes identified in this transcriptome may not be unique, or that novel genes/transcripts may have been assembled.

After unknown transcripts/genes were removed, 535 significantly differentially expressed genes were identified, using an adjusted $p < 0.05$, and of these, 284 had a $|Log_2$ fold change$| > 2$.

The genes with the greatest up-regulation, in cultures grown in the presence of urea, were associated with nitrogen fixation, and included one uncharacterised protein (RPA4610). These genes should be studied in greater detail so that their regulatory regions can be characterised, and used for future heterologous expression systems. This would allow the engineering of *R. palustris* strains that express heterologous proteins in response to the growth substrate.

Due to time constraints, only the nitrogenase, nitrogen regulation, and urease gene families are analysed here.

Possibly the most striking result of the experiment is the up-regulation of the molybdenum nitrogenase (*nif*) in the cultures grown with urea (Figure 5.21). This suggests that unlike ammonia, urea does not cause transcriptional repression of nitrogenase, and that therefore the presence of urea may not be inhibitory to hydrogen production, provided that post-translational regulatory systems have not suppressed its activity. This suggests a possible role for urea as a suitable substrate for biotechnological bio-hydrogen production.

It seems unlikely that the vanadium or iron nitrogenase variants were expressed at a high level in either condition, and no significant differential gene expression was observed, except for two genes (*vnfN* and *vnfX*) from the vanadium nitrogenase (Figure 5.21). *vnfN* and *vnfX* are involved in vanadium nitrogenase cofactor biosynthesis (Rubio and Ludden, 2005) and were significantly down-regulated in the cultures grown in the presence of ammonia. This suggests that they could be affected by the same regulatory mechanism that inhibits the molybdenum nitrogenase in the presence of ammonium.

Nitrogenase expression is dependent on the two component system NtrB/C, and although there was no detected change in *ntr* expression between the urea and $[NH_4]_2SO_4$ grown cells (Figure 5.22), nitrogenase genes were transcriptionally down-regulated in the $[NH_4]_2SO_4$ grown cells, suggesting that NtrC was not active. In contrast, nitrogenase transcription was up-regulated for cells grown with urea suggesting NtrC was active and was not affected by the ammonia released from urea metabolism in these samples.

The proteins DraT2 and GlnK2 are responsible for the ammonia induced suppression of nitrogenase activity; and act to post-translationally repress nitrogenase activity (Heiniger *et al.*, 2012). After activation by GlnK2, DraT2 causes the post-translational modification of nitrogenase, inhibiting its activity (Heiniger *et al.*, 2012). In this study, no change in the expression of *draT2* was detected (Figure 5.22), however, there was a small increase in *glnK2* expression in the urea grown cells (although this does not meet the threshold for statistical significance, its adjusted p value = 0.09, Figure 5.22). It is not clear whether nitrogenase was affected by DraT2 in the urea grown cells, as this would not have resulted in a change in transcript levels, and nitrogenase activity was not directly assayed.

Another interesting is result is the lack of any change in the transcript levels of the *ure* urease genes, except for *ureG* which was up-regulated in the $[NH_4]_2SO_4$ grown cells (Figure 5.23). Urease transcription in *Rhodobacter capsulatus* is NtrC dependent, and is required for urea utilisation (Masepohl and Kaiser *et al.*, 2001; Drepper and Wiethaus *et al.*, 2006). Assuming that the same is true in *R. palustris*, it is unclear as to why there was no change in urease transcript levels, since urea was the only major nitrogen source available, and NtrC was likely to have been activated.

If NtrC was active in cells grown with urea, but not in cells grown with ammonia, there must be some mechanism to explain why exogenous ammonia prevented activation, when intracellular ammonia released by urea catabolism did not. One speculative explanation (similar to one briefly mentioned by Masepohl and Kaiser *et al.* (2001)) could be that the consumption of ammonia by GlnA (glutamine synthetase) occurs at a greater rate than ammonia release from urease, leading to insufficient intracellular concentrations of ammonia for the activation of NtrC. No significant differential gene expression was observed for *glnA* however (Figure 5.22).

One potential confounding factor in this study, is that the cultures grown with either urea or $[NH_4]_2SO_4$ did not grow equally well, meaning that the cultures were at different growth stages when they were harvested. The clear change in nitrogen metabolism genes however, suggests that the responses observed were due to the change in nitrogen source, rather than a difference in growth stage. Care will be needed when interpreting changes in expression of other non-nitrogen related genes however, to determine whether the growth difference was a factor.

It is also not known why such a low proportion of the reads in each library mapped to the *R. palustris* genome (Table 5.2). BLAST searches of the unmapped reads against the ReSeq database failed to identify any significant matches, and this combined with the highly repetitive structure of the reads, which consisted of primarily poly-T and poly-AT sequences, may suggest that these sequences are artefacts produced during the ribosomal RNA depletion step, or during the library preparation.

Finally, an analysis of key genes in the TCA cycle showed no significant differential gene expression between either of the nitrogen source conditions (Figure 5.24). Citrate syn-

thase *cis*$\mathcal{Y}$ was expressed at a high level, and combined with the constitutive expression this helps to validate its promoter as a suitable choice for the heterologous expression system developed in Chapter 4.

5.5. Conclusions

This study establishes that urea is a suitable source of nitrogen for the production of biohydrogen by $\mathcal{R}$. *palustris*, as unlike ammonia it does not result in suppression of hydrogen producing nitrogenases.

A large number of differentially expressed genes are identified in the study, and additional analysis of the transcriptome dataset (including an analysis of any novel transcripts/genes) will be needed in the future, to gain a better understanding of the complete nitrogen regulatory pathways in $\mathcal{R}$. *palustris* grown with these substrates.

Further investigations of the regulatory pathway that allows nitrogenase expression in the presence of urea may be required, possibly including enzymatic and mutagenic studies of the enzymes described in the speculative mechanism described above. Additionally, since only the molybdenum nitrogenase variant appeared to be expressed, studies could be undertaken to preferentially express the vanadium or iron nitrogenase isozymes, and determine what effect this might have on hydrogen yields.

Expression levels of core central metabolism genes were tested and the citrate synthase promoter was validated as being suitable for the heterologous expression system developed in Chapter 4 based on its constitutive, high expression level.

Finally, genes that are highly up-regulated in the presence of urea have been identified, and future studies of their regulatory regions could be used to select promoters for the controlled expression of heterologous genes in $\mathcal{R}$. *palustris*.

6 Production of cyanophycin

6.1. Introduction

Having earlier determined that *R. palustris* could metabolise both ammonia and urea (Chapter 3), I decided to test whether *R. palustris* could be used to de-nitrify agricultural wastes using a model system.

Agricultural wastes, such as animal waste slurry and surface-water run-off from fertilised land, contain high levels of nitrogen, primarily in the form of ammonia and organic nitrogen such as urea (Ndegwa *et al.*, 2008; Hristov *et al.*, 2011). These wastes can pose environmental hazards through eutrophication; nitrogen run-off into waterways can cause toxic algal blooms that poison them and have severe detrimental effects on aquatic life and ecosystems, as well as rendering the water unsuitable for human use without extensive (and expensive) pre-treatment (Hanley, 1990; Burt and Haycock, 1991; Carpenter *et al.*, 1998; Giles, 2005). Additionally, volatilisation of ammonia, from these wastes, into the atmosphere poses a health and environmental risk as a non-point pollutant that contributes to fine-particulate formation in the atmosphere, nitrate contamination of surface and ground waters, and possibly adding to the global warming effect *via* the formation of nitrous oxide (Burt and Haycock, 1991; Carpenter *et al.*, 1998; Galloway and Aber *et al.*, 2003; Galloway and Dentener *et al.*, 2004; Ndegwa *et al.*, 2008; Chadwick *et al.*, 2011; Hristov *et al.*, 2011; Velthof *et al.*, 2014; Holly *et al.*, 2017).

Agricultural wastes such as manure and animal waste slurries can be used as fertilisers, but due to the high water content, storage and spreading can be very expensive, and can only be done at certain times of the year to avoid problems with surface run-off, and to ensure that application is appropriately synchronised with crop plantation and harvest, and weather conditions (Pain *et al.*, 1986; Burt and Haycock, 1991; Webb *et al.*, 2010; Christensen *et al.*, 2013; Holly *et al.*, 2017). Farming has become more intensive and compartmentalised, so that sources of manure/slurry may be produced in quantities that are too great, or in the wrong location for use as fertiliser for crop production (Burt and Haycock, 1991; Oenema *et al.*, 2007). Furthermore, agricultural wastes such as manure and slurry as fertiliser may be less effective than artificial alternatives, and are thus often applied in higher volumes, increasing the risk of leaching and run-off (Burton and Turner,

2003; Christensen *et al.*, 2013). Additionally, if manure or slurries are stored for long periods of time prior to use as fertiliser, they still pose an environmental hazard due to gaseous losses (Oenema *et al.*, 2007; Hristov *et al.*, 2011; Hafner *et al.*, 2013; Browne *et al.*, 2015; Holly *et al.*, 2017). For these reasons, artificial fertilisers such as ammonium nitrate and urea, usually produced *via* the Haber-Bosch process (Haber and Rossignol, 1913), are now often preferred over traditional manure or slurry fertilisers (Pain *et al.*, 1986; Galloway and Aber *et al.*, 2003). Unfortunately this has contributed to the problem of global nitrogen imbalances and nitrogen pollution problems, as now more nitrogen inputs are going into the land and ecosystems than are being taken out through crop and animal harvesting (Hanley, 1990; Burt and Haycock, 1991; Carpenter *et al.*, 1998; Galloway and Aber *et al.*, 2003; Galloway and Dentener *et al.*, 2004; Giles, 2005; Oenema *et al.*, 2007; Hou *et al.*, 2016).

In addition to direct use as fertiliser, manures and slurries can be processed in various ways; anaerobic digestion, composting, drying and separating, and incineration may be used to improve their characteristics as fertilisers and/or for the production of bio-gasses and energy (Perälä and Regina, 2006; Oenema *et al.*, 2007; Yokoyama *et al.*, 2007; Fangueiro *et al.*, 2010; Hamelin *et al.*, 2011; Petersen and Sommer, 2011; Christensen *et al.*, 2013; Aguirre-Villegas *et al.*, 2014; Cavalli *et al.*, 2014; Sun *et al.*, 2014; Browne *et al.*, 2015; Neerackal *et al.*, 2015; Riggio *et al.*, 2015). Although these processes have certain advantages, they still do not remove terminal nitrogen from the system, and nitrogen pollution is still a problem. *R. palustris* could potentially be used to metabolise the nitrogen in agricultural wastes, including manures and slurries. However, this also does not completely remove the nitrogen from the system - the nitrogen would still be present as fixed nitrogen inside the *R. palustris* cells. Growing *R. palustris* on agricultural wastes could have benefits, such as the production of hydrogen gas or bio-electricity (in a MFC device), but still leaves a waste disposal problem: the high nutrient content of the *R. palustris* culture would still need to be safely disposed of, as a large build-up of *R. palustris* would likely also be detrimental to the environment.

Disposing of the *R. palustris* need not pose a problem though, as *R. palustris* has the potential to be used as a 'cell-factory'; used to convert waste into some desirable product. One possible material is cyanophycin. As a storage polymer rich in nitrogen and carbon, it could be used as a sink to absorb waste nutrients.

Cyanophycin (a.k.a. multi-L-arginyl-poly-L-aspartic acid) is a non-ribosomally produced polypeptide, consisting of a poly-aspartic acid backbone with linked arginine side residues, in an equimolar ratio, and its molecular weight varies between 25 kDa to 100 kDa (Simon, 1971; 1973; Simon and Weathers, 1976). A diagram of its structure is shown in Figure 6.1. It was first described at least as early as 1959 by Fritsch (1959), and in 1968 by Lang (1968), and possibly as early as 1897 or 1905 by Alfred Fischer (A. Fischer, 1897; 1905). Its structure was first described by Simon (1971) and Simon and Weathers (1976)

and later by Ziegler and Diener *et al.* (1998). It is produced naturally in most cyanobacteria (Füser and Steinbüchel, 2007), primarily as a nitrogen and carbon storage compound (Simon, 1971; 1973; B. Liang *et al.*, 2014).

Cyanophycin is potentially industrially valuable (Mooibroek *et al.*, 2007), having applications in a variety of areas, principally as a sustainable and biodegradable substitute for some industrial petrochemicals (Joentgen *et al.*, 1998; Schwamborn, 1998; Mooibroek *et al.*, 2007; Solaiman *et al.*, 2011; Aravind *et al.*, 2016). To begin with, cyanophycin can be hydrolysed to arginine and aspartate, and partially hydrolysed to dipeptides, which have uses in foods and pharmaceuticals (Sallam and Steinbüchel, 2010). Additionally, cyanophycin can be partially hydrolysed to reduce the arginine content, leaving poly-aspartic acid (Joentgen *et al.*, 1998). Poly-aspartic is of considerable interest as a replacement for poly-acrylate, which is used in water-softening applications and other industrial processes (Schwamborn, 1998; Neumann *et al.*, 2005).

In addition to the applications described, cyanophycin has some interesting and useful properties. Foremost among these, is its varying solubility. At neutral, or biological, pH cyanophycin is insoluble, and naturally forms granules inside the cells, but at low pH, it becomes soluble (Simon, 1971; 1973). This provides a simple way for cyanophycin to be selectively purified from cells, by washing it out of the cells with acid, and then neutralising the acid to precipitate the cyanophycin (Simon and Weathers, 1976).

Since it is naturally insoluble at neutral/biological pH, and naturally forms granules inside the cells, cyanophycin could be purified relatively easily. Cyanophycin synthesis is catalysed from its substrates, arginine and aspartate, by just one enzyme, cyanophycin synthetase (Ziegler and Diener *et al.*, 1998; Aboulmagd *et al.*, 2000). Cyanophycin synthetase has two catalytic sites which catalyse two parts of the reaction, L-aspartate addition, EC 6.3.2.29, and L-arginine addition, EC 6.3.2.30 (Ziegler and Diener *et al.*, 1998; Aboulmagd *et al.*, 2000; Berg *et al.*, 2000; Aboulmagd *et al.*, 2001; Ziegler and Deutzmann *et al.*, 2002). A diagram of the proposed cyanophycin synthesis mechanism is shown in Figure 6.1. Cyanophycin is degraded by cyanophycinase, and is extremely resistant to other proteases (Simon and Weathers, 1976; Allen, 1984; Obst and Oppermann-Sanio *et al.*, 2002; Obst and Sallam *et al.*, 2004).

BLAST searches of both cyanophycin synthetase and cyanophycinase protein sequences against the *R. palustris* genome indicate that *R. palustris* has no native cyanophycin anabolic or catabolic pathways, suggesting that cyanophycin is not natively produced in *R. palustris*, and would not naturally be degraded. This view is supported by the findings of the *in silico* analysis of cyanophycin synthetases and cyanophycinases in diverse prokaryotes by Füser and Steinbüchel (2007).

These features make cyanophycin synthesis an ideal pathway candidate to engineer into *R. palustris*; only one new enzyme needs to be expressed, since cyanophycin is synthesised by cyanophycin synthase from the substrates aspartate and arginine, both of which are nat-

urally produced by the native metabolism of *R. palustris* (Yen and Gest, 1974; Larimer *et al.*, 2004; Pechter *et al.*, 2016). There are also no natural catabolic pathways, meaning that the cyanophycin should accumulate inside the cell without being degraded by any endogenous enzymes. It was therefore decided to attempt to engineer *R. palustris* to produce cyanophycin, by introduction of a cyanobacterial cyanophycin synthase gene under the control of the *R. palustris* citrate synthase promoter.

6.2. Experimental

6.2.1. Construction of heterologous lines

Four different *R. palustris* lines expressing cyanophycin synthetase were constructed. The *cphA* gene from *Anabaena varabilis* sp. PCC 7120 was expressed in *R. palustris* under control of the native citrate synthase promoter. The gene was inserted at each of two different loci in the *R. palustris* genome, either replacing the transposase RPA0768 (a putative neutral site, hereafter referred to as "T1"), or interrupting genes for polyhydroxybutyrate synthesis, RPA0531 and RPA0532 (this locus hereafter referred to as "Pha"). At each of these loci, the *cphA* gene was expressed both with, and without a C-terminal tag. The tag used was a double tag consisting of c-Myc followed by a hexahistidyl tag.

Plasmid construction

The plasmids were assembled by Gibson-assembly cloning (D. G. Gibson *et al.*, 2009; D. G. Gibson, 2011), using the NEBuilder HiFi DNA Assembly kit by New England Biolabs (Chapter 2.2.3). Each of the fragments to be incorporated into the plasmids was amplified by PCR (Chapter 2.2.2) using the primers shown in Table 6.1 and *R. palustris* or *A. variabilis* bacterial colonies for the template material. The tag, where used, was synthesised as part of a 97 bp primer (Table 6.1). The vector backbone was pKmobSacB (Schäfer *et al.*, 1994). Maps of the plasmids and insert sites are shown in Figure 6.2. Each plasmid contained the *A. variabilis cphA* gene flanked by the *R. palustris* citrate synthase promoter and terminator, and then flanking sequences derived from the T1 or Pha locus as shown in Figure 6.2 (e), (f). The sequence of the constructed plasmids was confirmed by restriction digestion and capillary Sanger sequencing (data not shown).

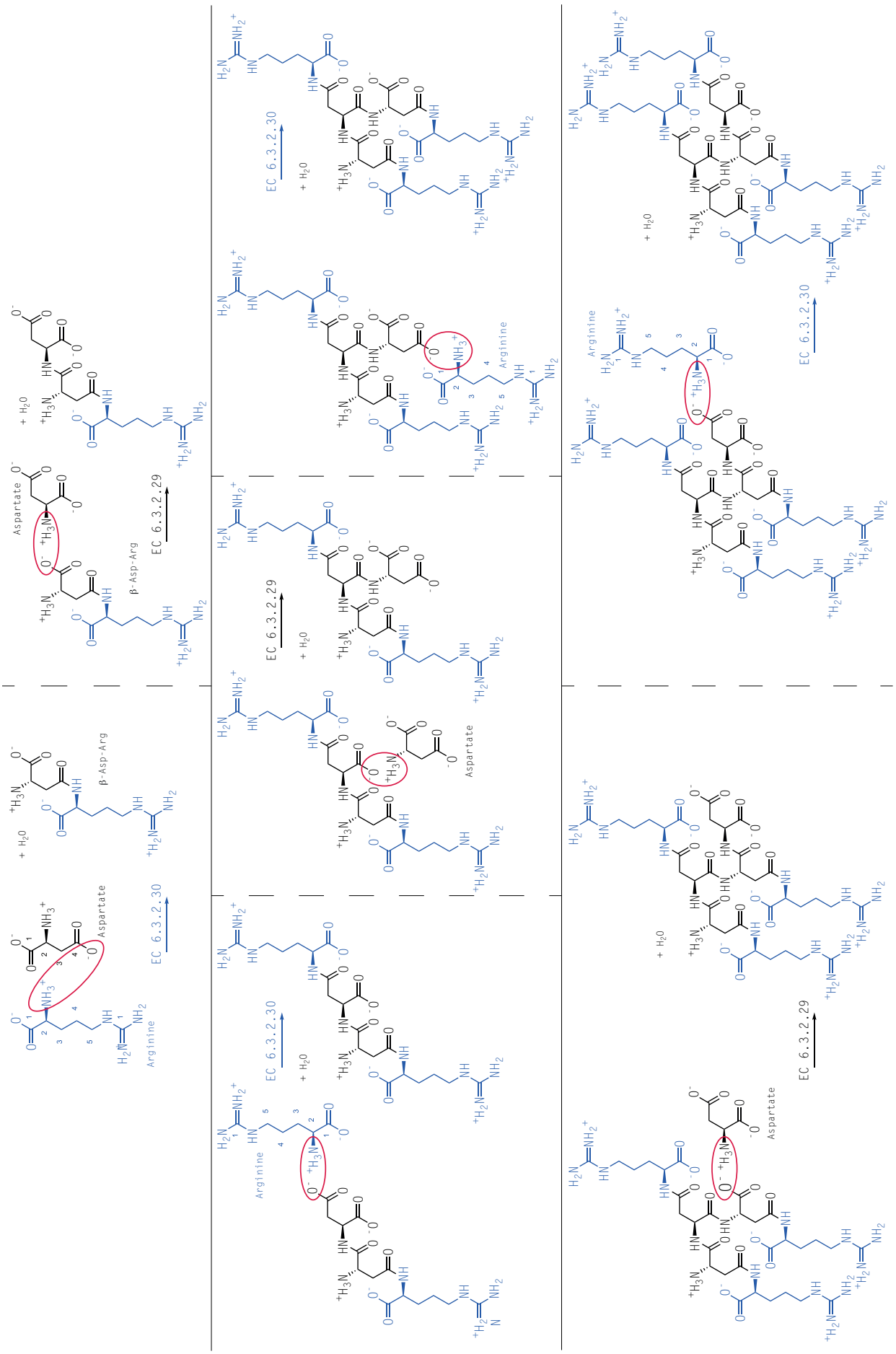


Figure 6.1. – Structure of cyanophycin. Diagram of the synthesis and structure of cyanophycin polymer, its substrates and monomers. Arginine is shown in blue, and (poly)-aspartate is shown in black. Cyanophycin synthesis is catalysed by one enzyme, cyanophycin synthetase, which has two active sites that catalyse the two parts of the reaction; L-aspartate addition by EC 6.3.2.29, and L-arginine addition by EC 6.3.2.30. The stages of cyanophycin synthesis are shown in the different panels, proceeding left to right, top to bottom. The functional groups participating in the reaction are circled in red. In the final panel, a polymer composed of four monomers is displayed, however the real polymer would be much longer than this.

Table 6.1. – Primers used to make RpCphA plasmids. The gene specific binding regions are shown in uppercase, whilst the overhanging region used to create the overlaps for Gibson assembly are shown in lowercase.

Primer Name	Fragment amplified	Binding template	Primer sequence 3' - 5'	Length (nt)	Overlaps with	Used to make plasmid(s)
PHAL-F	Pha Left Flank	*R. palustris* genome	aagcttgcatgcctgcaggtcgactctagaGCGTGATAACCGAACAATC	49	pKmobSacB	RpCphA::Pha, RpCphA::Pha::Tag
PHAL-R	Pha Left Flank	*R. palustris* genome	tttcccgggttttGCGGAGAAATTCAAAAACG	32	Citrate synthase promoter	RpCphA::Pha, RpCphA::Pha::Tag
Prom-F-PHA	Citrate synthase promoter	*R. palustris* genome	tgaatttctccgcAAAACCCGGGAAACGGCT	31	Pha Left Flank	RpCphA::Pha, RpCphA::Pha::Tag, RpCphA::T1, RpCphA::T1::Tag
Prom-R-Ana	Citrate synthase promoter	*R. palustris* genome	tgaggattctcatCGTGTGGTCCCCGATGTT	31	*Anabaena* CphA1	RpCphA::Pha, RpCphA::Pha::Tag, RpCphA::T1, RpCphA::T1::Tag
AnaCphA-F	*Anabaena* CphA1	*Anabaena* genome	cggggaccacacgATGAGAATCCTCAAGATCC	32	Citrate synthase promoter	RpCphA::Pha, RpCphA::Pha::Tag, RpCphA::T1, RpCphA::T1::Tag
AnaCphA-R-NT	*Anabaena* CphA1	*Anabaena* genome	gccgggtcgacgaCTACAGCAAAGTATTAATTACTGAAG	39	Citrate synthase terminator	RpCphA::Pha, RpCphA::T1

Table 6.1 continued from previous page

Primer Name	Fragment amplified	Binding template	Primer sequence 3' - 5'	Length (nt)	Overlaps with	Used to make plasmid(s)
Term-F-Ana-NT	Citrate synthase terminator	*R. palustris* genome	tactttgctgtagTCGTCGACCCGGCTCGAT	31	*Anabaena* CphA1	RpCphA::Pha, RpCphA::T1
Term-R-PHA	Citrate synthase terminator	*R. palustris* genome	tcaacggttaccaGTTTAGCGGAACGGCTCCG	32	Pha Right Flank	RpCphA::Pha, RpCphA::Pha::Tag
PHAR-F	Pha Right Flank	*R. palustris* genome	cgttccgctaaacTGGTAACCGTTGAAGGCATC	33	Citrate synthase terminator	RpCphA::Pha, RpCphA::Pha::Tag
PHAR-R	Pha Right Flank	*R. palustris* genome	tacgaattcgagctcggtacccggggatcCTCGAAGATGATCTGTGCG	48	pKmobSacB	RpCphA::Pha, RpCphA::Pha::Tag
AnaCphA-R-T	*Anabaena* CphA1	*Anabaena* genome	ctgagatgagtttttgttcCAGCAAAGTATTAATTACTGAAGAAG	45	Citrate synthase terminator	RpCphA::Pha::Tag, RpCphA::T1::Tag
Term-F-Ana-T	Citrate synthase terminator	*R. palustris* genome	taatactttgctg gaacaaaaactcatctcagaagaggatctgaa tagcgccgtcgaccatcatcatcatcatcattaa TCGTCGACCCGGCTCGAT	97	*Anabaena* CphA1	RpCphA::Pha::Tag, RpCphA::T1::Tag
T1L-F	T1 Left Flank	*R. palustris* genome	aagcttgcatgcctgcaggtcgactctagaCTCTCACCGGACTTGCCAATC	51	pKmobSacB	RpCphA::T1, RpCphA::T1::Tag
T1L-R	T1 Left Flank	*R. palustris* genome	tttcccgggttttCCGGTCGATCAGGCCTTC	31	Citrate synthase promoter	RpCphA::T1, RpCphA::T1::Tag

Table 6.1 continued from previous page

Primer Name	Fragment amplified	Binding template	Primer sequence 3' - 5'	Length (nt)	Overlaps with	Used to make plasmid(s)
Prom-F-T1	Citrate synthase promoter	*R. palustris* genome	cctgatcgaccggAAAACCCGGGAAACGGCT	31	T1 Left Flank	RpCphA::T1, RpCphA::T1::Tag
Term-R-T1	Citrate synthase terminator	*R. palustris* genome	tcggctgacagtgGTTTAGCGGAACGGCTCCG	32	T1 Right Flank	RpCphA::T1, RpCphA::T1::Tag
T1R-F	T1 Right Flank	*R. palustris* genome	cgttccgctaaacCACTGTCAGCCGATTACC	31	Citrate synthase terminator	RpCphA::T1, RpCphA::T1::Tag
T1R-R	T1 Right Flank	*R. palustris* genome	tacgaattcgagctcggtacccggggatccTTTCTCCCCCGGAAAATG	48	pKmobSacB	RpCphA::T1, RpCphA::T1::Tag

Strain construction

The plasmids were then introduced into *R. palustris* by electroporation (Chapter 2.2.4). The antibiotic kanamycin was used to select for an initial single crossover event incorporating the plasmid into the genome (Figure 6.3). (In general, single crossovers happen much more frequently in *R. palustris* than double crossovers.) The aim was then be to generate a markerless strain by selection of sucrose resistance (requiring loss of sacB.) However, because of time constraints, only the strains from the first crossover event (using kanamycin selection) were created.

The genotypes of strains generated were checked by PCR using two sets of primers. One comprised an external pair that bound on either side of the insert site. The second comprised a forward primer that bound inside the inserted gene and an external reverse primer that was the same as for the previous pair, that bound outside of the insert site (see Figure 6.2 and Table 6.2).

Expression of *cphA* gene

Expression of the heterologous gene at the RNA level was checked by RT-PCR (Chapter 2.2.6), using primers internal to the *cphA* gene (Table 6.2).

Expression of CphA protein

Protein expression was assayed by western Blot, using the tagged lines and antibodies to c-Myc (Chapter 2.2.7).

6.2.2. Purification and characterisation of cyanophycin

Cyanophycin was purified using a method similar to that described by Simon and Weathers (1976) (Chapter 2.2.9), involving cell disruption by ultra-sonication and detergent, followed by solubilisation of cyanophycin in 0.1 M HCl and subsequent neutralisation and precipitation after cell debris had been removed by centrifugation. The extract was then assayed using the modified Sakaguchi reaction, as described by Allen (1984), which is used to determine the concentration of non-hydrolysed arginine (Chapter 2.2.9). The cyanophycin was then further analysed, to assess purity, by assaying the total free amino acid composition, after hydrolysis, using ion-exchange chromatography (Chapter 2.2.9). SDS-PAGE (poly-acrylamide gel electrophoresis) was used to determine the size of the polymer as isolated from cells of *R. palustris*.

Cyanophycin was initially purified from "small-scale" 50 mL cultures for preliminary analyses. This was then scaled up to first 1 L ("mid-scale"), and then later, 5 L ("large-scale") cultures in order to make a gravimetric analysis and express cyanophycin yield as a percentage of CDW. These larger cultures were grown in a greenhouse on the roof of

(a) Plasmid map of RpCphA::T1

(b) Plasmid map of RpCphA::Pha

(c) Plasmid map of RpCphA::T1::Tag

(d) Plasmid map of RpCphA::Pha::Tag

(e) T1 (RPA0768) neutral insertion site

(f) Polyhydroxybutyrate (RPA0531 & RPA0532) insertion site

Figure 6.2. – Plasmid maps and genomic context of the insertion sites for the RpCphA strains.

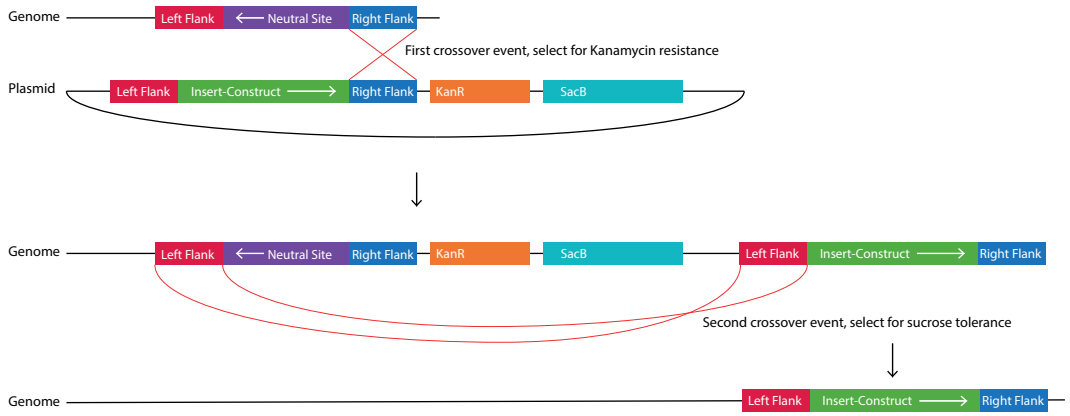


Figure 6.3. – ***R. palustris* transformation strategy.** Diagram showing the transformation process for creating new *R. palustris* strains from pkMobSacB based plasmids. A single crossover event, between either the left or right flanks (right pictured), is selected for using kanamycin as the selection agent, which results in the introduction of the whole plasmid, next to the insertion locus. A second crossover event (if performed) is selected for using sucrose, and results in either an unmarked 'knock-in' where the insert is left in place and the remainder of the plasmid and the WT insertion locus is excised (pictured), or a reversion to WT.

Table 6.2. – Primers used to genotype and check expression of RpCphA strains

Primer Name	Primer sequence 3' - 5'	Length (nt)
External genotyping primer (Forward) - T1	CGGCAGCAGTTCTACGAGAT	20
External genotyping primer (Reverse) - T1	CTTCTCCACTGTCGCGATGA	20
External genotyping primer (Forward) - Pha	GGCAGATCACATTGACGGTG	20
External genotyping primer (Reverse) - Pha	CGTCCAAGAGCACGTCAAAA	20
Internal genotyping primer (Forward)	TGCCGGTGCAGTTATGGATA	20
CphA RT-PCR Forward	ATACACAGATTCCGCCACCA	20
CphA RT-PCR Reverse	TACGCTCTGGGTTAGGCTTT	20

the University of Cambridge Biochemistry Department, with natural lighting and diurnal rhythms. The greenhouse was air-conditioned to maintain a temperature of approximately 25 °C (Figure 6.4).

Electron microscopy

Cyanophycin is insoluble at biological pH, and naturally accumulates as granules inside the cells of organisms that produce it. The RpCphA strains were therefore investigated by TEM (transmission electron microscopy), and compared to the WT strain, to determine if similar cyanophycin granules could be seen in the RpCphA strains.

Initially, a simple negative stain was performed using uranyl acetate, to investigate if there were any externally visible differences between the RpCphA and WT strains. Following this, the cells were fixed in resin and then sectioned using an ultra-microtome, to investigate cross-sections of the cells and look for the intracellular cyanophycin granules (Chapter 2.2.10).

6.3. Results

6.3.1. Construction of heterologous lines

Genotyping

After the transformation, approximately 100 colonies ($\pm \sim 50$) were obtained for each transformed line, and 12 of each were selected for genotyping by two PCR reactions. In each case, one reaction was carried out using a pair of primers that bound to the flanking regions either side of the insertion locus, and another reaction was performed using one primer internal to the inserted gene, and another primer that bound to a region downstream of the insertion loci.

For the first PCR reaction, two products were expected; one from the WT gene, and one from the inserted gene. From the second PCR reaction, a product should only be generated from the transformed strain. Representative results of successful transformants are shown in Figure 6.5 (a),(b) (first PCR) and 6.5 (c),(d) (second PCR). Of the 12 colonies screened, only 1 of each produced a strain that tested positive using both PCR reactions, except for the RpCphA::T1 line, which produced 2.

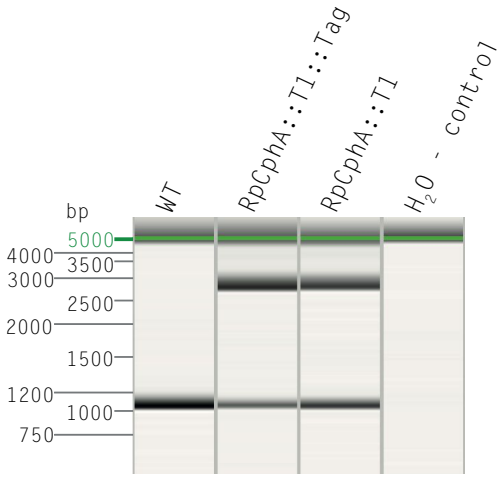
Expression of *cphA* gene

cphA transcript expression in *R. palustris* was tested by RT-PCR, using primers internal to the *cphA* gene. A positive result, indicating gene transcription, was obtained for the transformed strains (Figure 6.6 (a)). A control reaction was carried out where reverse transcriptase was omitted. No bands were seen, indicating that the product amplification

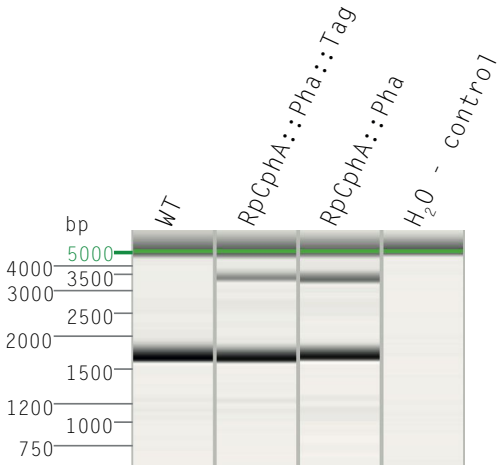
(a) Cultures at start

(b) Cultures at end

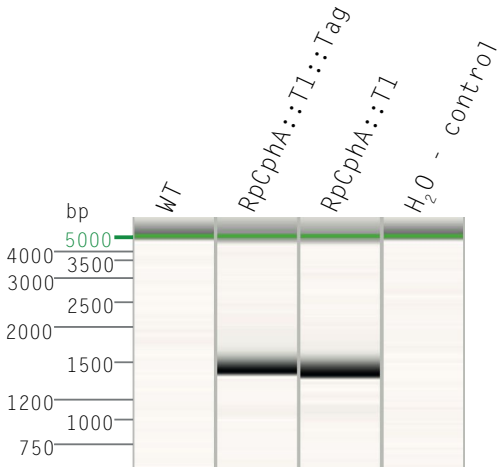
Figure 6.4. – 5 L *R. palustris* cultures for cyanophycin extraction.



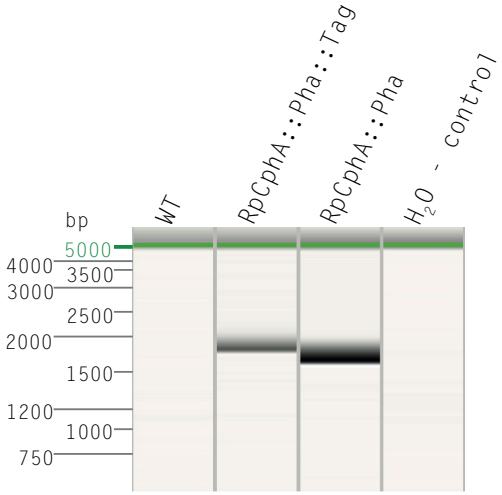
(a) Genotype of RpCphA::T1 strains



(b) Genotype of RpCphA::Pha strains

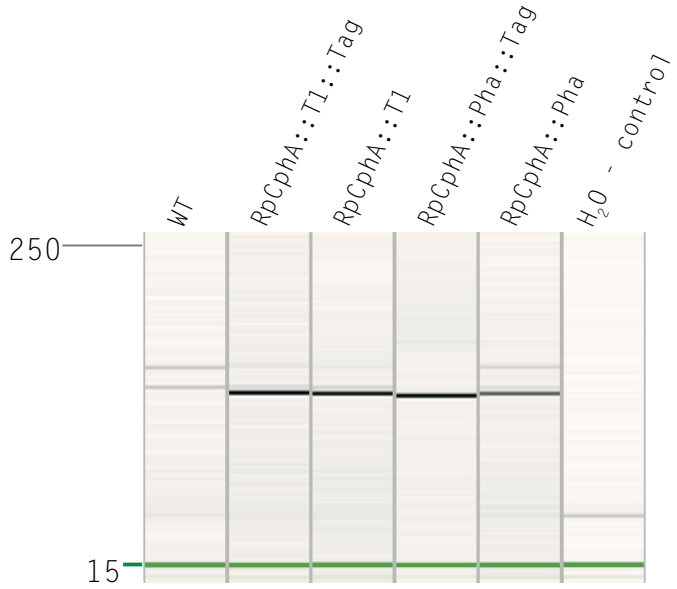


(c) Check of RpCphA::T1 genotype with internal primer

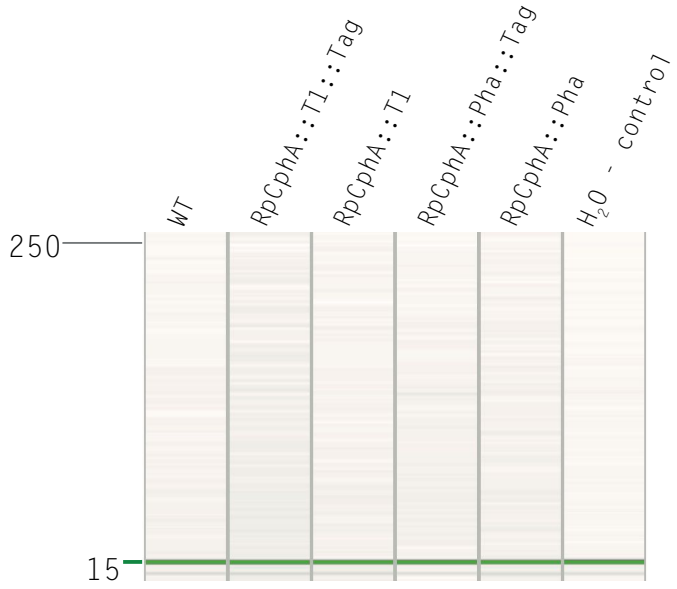


(d) Check of RpCphA::Pha genotype with internal primer

Figure 6.5. – Genotyping of RpCphA strains. These virtual gels, from QIAxcel capillary gel electrophoresis, show representative results of the PCRs used to genotype the RpCphA strains. Where the external primers were used to genotype (panels **a** & **b**), all lanes (except the H_2O control) showed a WT band of the expected size (1095 bp for the T1 locus and 1931 bp for the Pha locus), and the RpCphA strains showed an additional, larger band of the expected size (3191, 3128, 3933, & 3870 bp for RpCphA::T1::Tag, RpCphA::T1, RpCphA::Pha::Tag, & RpCphA::Pha respectively) from the inserted plasmid. Where the internal primers (specific to the insert) were used (panels **c** & **d**) positive signals were only obtained from the RpCphA strains, and were of the expected size (1569, 1506, 1965, & 1902 bp for RpCphA::T1::Tag, RpCphA::T1, RpCphA::Pha::Tag, & RpCphA::Pha respectively).



(a) RT-PCR to detect *cphA* gene expression in RpCphA strains



(b) No-RT negative control for RT-PCR to detect *cphA* gene expression in RpCphA strains

Figure 6.6. – *cphA* gene expression in RpCphA strains. These virtual gels, from QIAxcel capillary gel electrophoresis, show the products of the RT-PCR used to detect expression of *cphA* in the RpCphA strains. Positive bands of the expected size (137 bp) were observed in the RpCphA strains but not in the WT control (a). No signals were detected in the negative RT control (b).

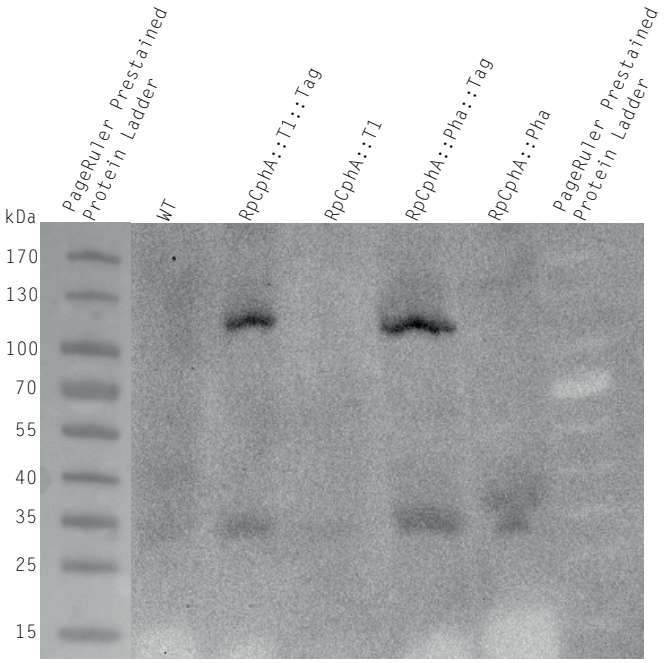


Figure 6.7. – Western blot of CphA expression in *R. palustris* strains. Whole-cell protein lysates separated by SDS-PAGE, transferred to nitrocellulose membrane and detected using anti-c-Myc antibody. Strong positive signals were observed in the two tagged RpCphA strains but not in the WT control or the untagged RpCphA strains.

was due to cDNA template (from RNA transcript) and not from contaminating genomic DNA (Figure 6.6 (b)).

Expression of CphA protein

CphA protein expression was tested by western Blot, using the anti c-Myc antibody to the c-Myc tag in the tagged *R. palustris cphA* transformed strains. A band was observed for both of the tagged strains, at the expected size (100.8 kDa). A faint band was also observed at ~35 kDa, which could be cyanophycin polymer.

6.3.2. Purification and characterisation of cyanophycin

Small-scale extraction

To test if cyanophycin had been successfully produced, 50 mL samples of *R. palustris* cultures were subjected to the cyanophycin extraction protocol and then analysed. This was done with all four strains and the WT. The precipitated cyanophycin extract (solubilised in 0.1 M HCl) was assayed by the Sakaguchi reaction. The RpCphA::T1 strains (tagged and untagged) showed significantly higher arginine concentrations than the WT control, indicating the likely presence of cyanophycin (Figure 6.8). The RpCphA::Pha strains (especially the RpCphA::Pha::Tag strain) showed a slightly greater yield of arginine than the WT control, but this was not statistically significant.

During the western blotting experiments used to detect cyanophycin synthetase (Figure 6.7), large bands were observed for the RpCphA strains in the 15 kDa to 25 kDa range of the SDS-PAGE gels of the whole-cell lysate (Figure 6.9). Running samples of the extracted cyanophycin on similar SDS-PAGE gels showed the same bands in the 15 kDa to 25 kDa range for the RpCphA strains (Figure 6.10), indicating that the bands seen in the whole cell extracts were indeed cyanophycin polymer of that size.

Amino acid analysis by ion-exchange chromatography was used to test the identity and assess the purity of the cyanophycin. The extracts were hydrolysed prior to analysis and then products of the RpCphA extracts compared to those of the WT extracts. An annotated chromatogram of the residual amino acid extracts is shown in Figure 6.11, and the raw chromatograms of all of the extracts is shown in Figure 6.12. Since the samples were loaded to the ion-exchange chromatography according to the arginine concentration calculated by the Sakaguchi assay, and the chromatograms were normalised twice, firstly according to the amount loaded and secondly according to the amount of the internal standard (Norleucine) added; these are shown in Figure 6.13. The figures show that apart from trace amino acids, the RpCphA strains produced an extract that contains large amounts of arginine and aspartate, norleucine (the internal standard), and the residual ammonium (Figures 6.12 & 6.13).

The amounts of each amino acid calculated from the chromatograms are shown in Figure 6.14, having first been normalised according to the volume of the original extraction and the OD_{660} of the culture. This shows that the cyanophycin extracted had an approximate arginine to aspartate ratio of 1:1, a small amount of lysine, and only trace levels of the other amino acids. The two RpCphA::T1 strains both produced a significant amount of cyanophycin, while the RpCphA::Pha::Tag strain appeared to produce a small amount, and the RpCphA::Pha strain only appeared to produce trace amino acids in the extraction.

Negatively stained TEM micrographs showed no significant differences between the WT and the RpCphA strains (Figure 6.15). In sectioned TEM micrographs, dark granules could be seen in all of the strains, including the WT strain. In the RpCphA strains however, some of the bacterial cells showed much larger granules (Figures 6.17, 6.18, 6.19, 6.20), which did not appear to be present in the WT strain (Figure 6.16). It is possible that these might be cyanophycin granules.

Mid-scale extraction

Cyanophycin was extracted from 1 L samples of *R. palustris* culture. Very little cyanophycin was obtained from the 1 L cultures, and little difference was seen, in the percentage yield of the CDW, between the WT strain (reflecting the baseline for the assay) and the RpCphA strains (Table 6.3).

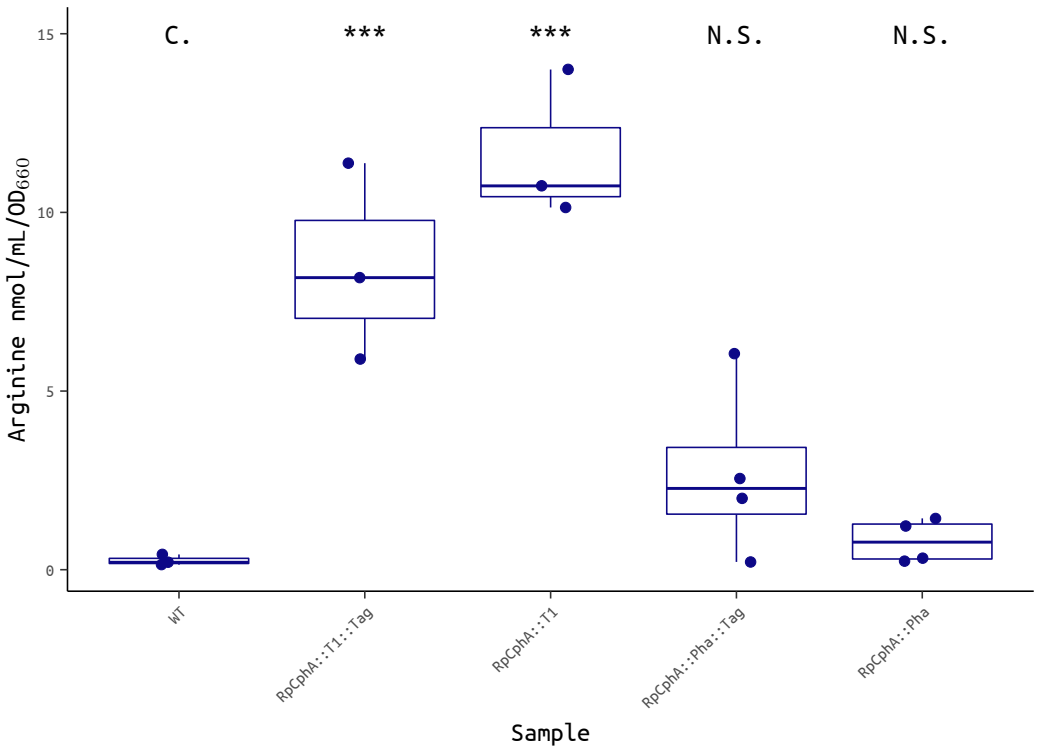


Figure 6.8. – Non-hydrolysed arginine concentration of small-scale cyanophycin extraction. Concentration of non-hydrolysed arginine in small-scale cyanophycin extractions of three biological replicates of the four RpCphA strains and WT control, as determined by the Sakaguchi assay and calculated from an arginine standard curve. A significant increase in arginine is seen in the RpCphA::T1 & RpCphA::T1::Tag strains, relative to the WT control. A slight increase (relative to the WT control) may be present in the RpCphA::Pha & RpCphA::Pha::Tag strains, but this was not statistically significant.

Table 6.3. – Cyanophycin extraction yields from 1 L cultures. Data presented are for a single sample of each strain.

	CDW (g)	Starting OD_{660}	Harvested OD_{660}	Change in OD_{660}	Extracted pellet (g)	Extraction as % CDW
WT	1.32	0.177	2.822	2.645	0.014	1.06
RpCphA::T1::Tag	0.91	0.228	2.496	2.268	0.021	2.32
RpCphA::T1	0.58	0.242	2.117	1.875	0.030	5.15
RpCphA::Pha::Tag	0.16	0.261	0.792	0.531	0.003	1.84
RpCphA::Pha	0.29	0.244	1.252	1.008	0.003	1.03

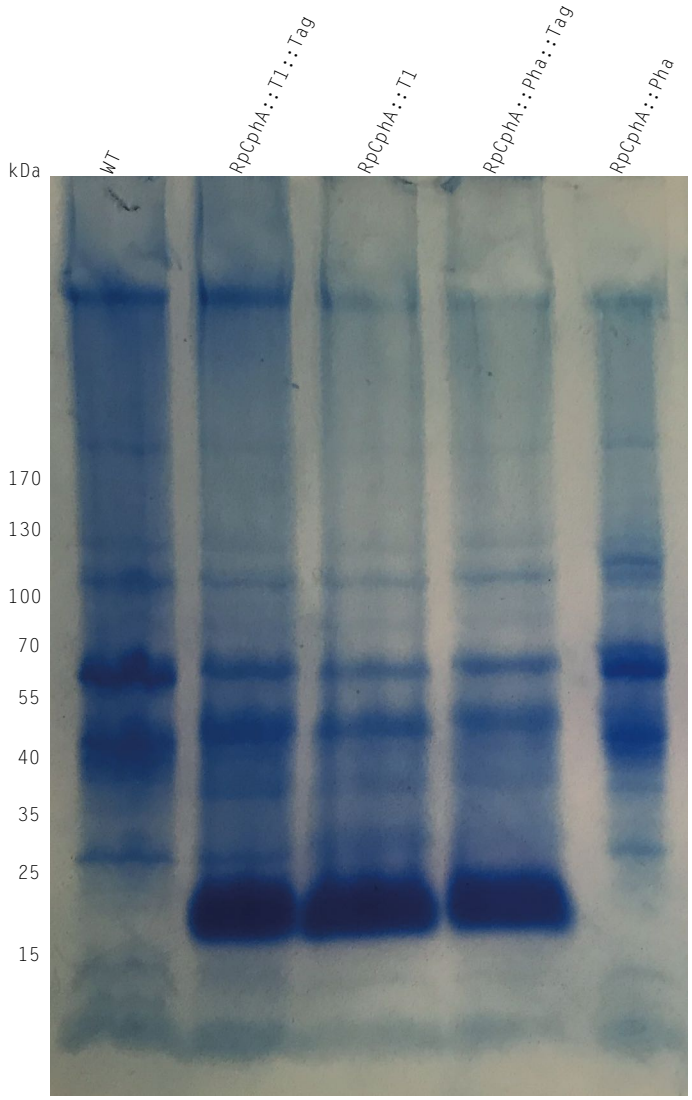


Figure 6.9. – SDS-PAGE of whole-cell lysate from *R. palustris* RpCphA cultures. Denatured protein from whole-cell lysate of the four RpCphA strains and WT control, separated according to size. No significant difference is observed between the RpCphA strains and the WT control in the ~100 kDa region where the introduced cyanophycin synthetase should be seen, however there is a significant band between 15 and 25 kDa in the RpCphA strains, not present in the WT control. This is likely to be cyanophycin polymer.

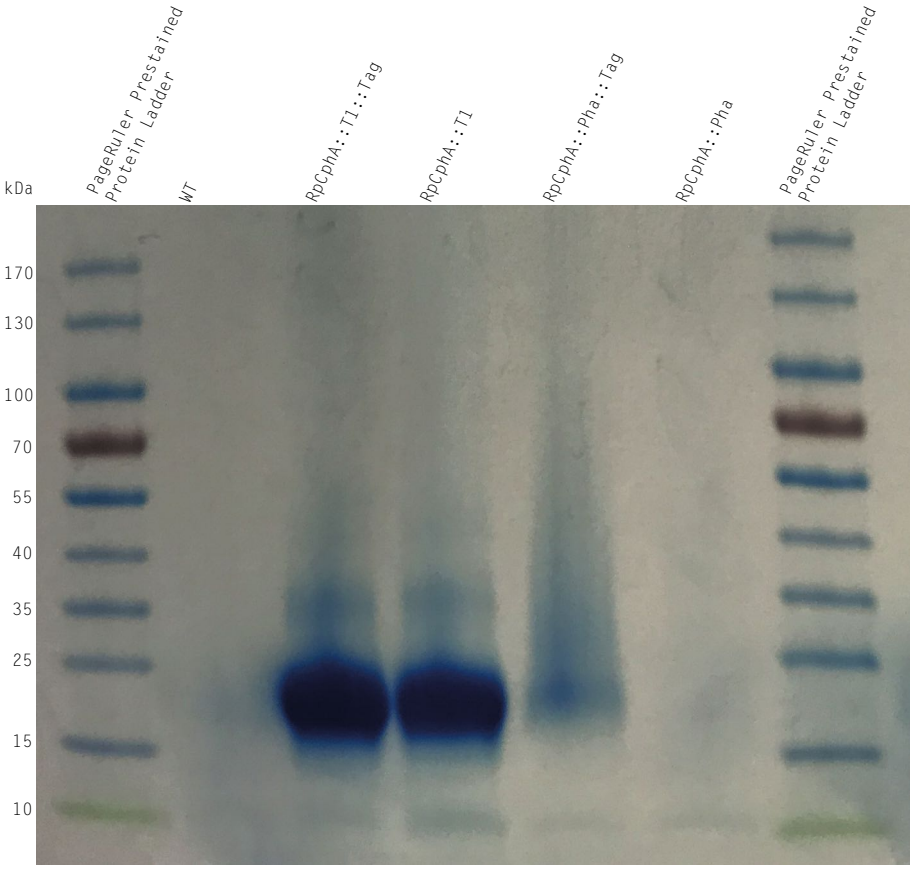


Figure 6.10. – SDS-PAGE **of cyanophycin extracted from *R. palustris* RpCphA cultures.** Small-scale cyanophycin extracts from RpCphA strains and WT control, denatured and run as protein on SDS-PAGE gels. Large bands of cyanophycin polymer are visible in the lanes corresponding to the RpCphA::T1 & RpCphA::T1::Tag strains at the region of the gel representing to 15 and 25 kDa. A faint band is visible in the RpCphA::Pha::Tag strain but not observed in the RpCphA::Pha strain.

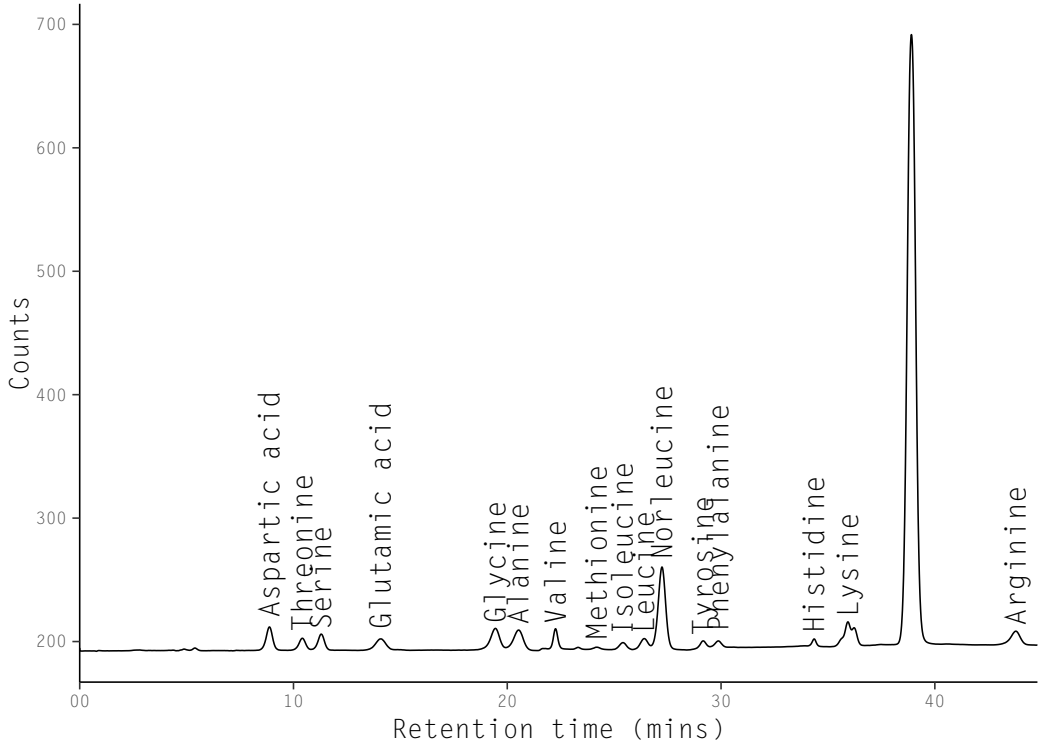


Figure 6.11. – Ion-exchange chromatogram showing residual amino acid composition from small-scale cyanophycin extraction on WT control. Ion-exchange chromatogram of hydrolysed amino acids from a small-scale cyanophycin extraction performed on the WT control. Peaks are annotated with their respective amino acid. The large unlabelled peak is caused by ammonium released during the hydrolysis reaction.

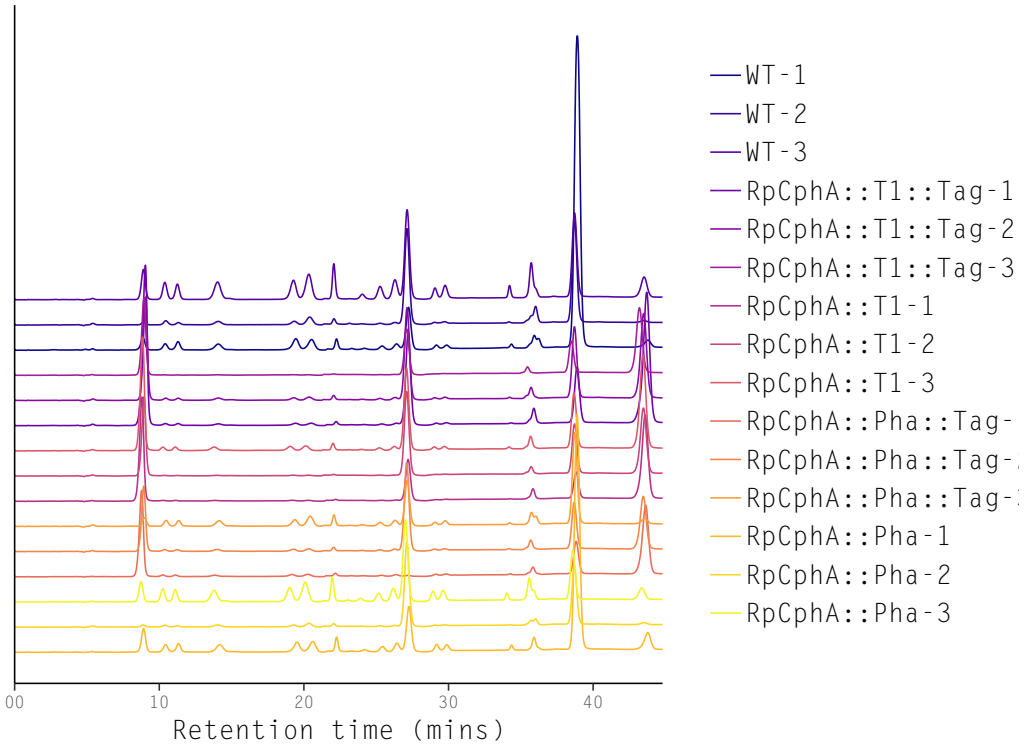


Figure 6.12. – Ion-exchange chromatogram showing total amino acid composition of hydrolysed small-scale cyanophycin extractions. Raw, unprocessed ion-exchange chromatograms of hydrolysed amino acids from small-scale cyanophycin extractions on three biological replicates of the four RpCphA strains and WT control.

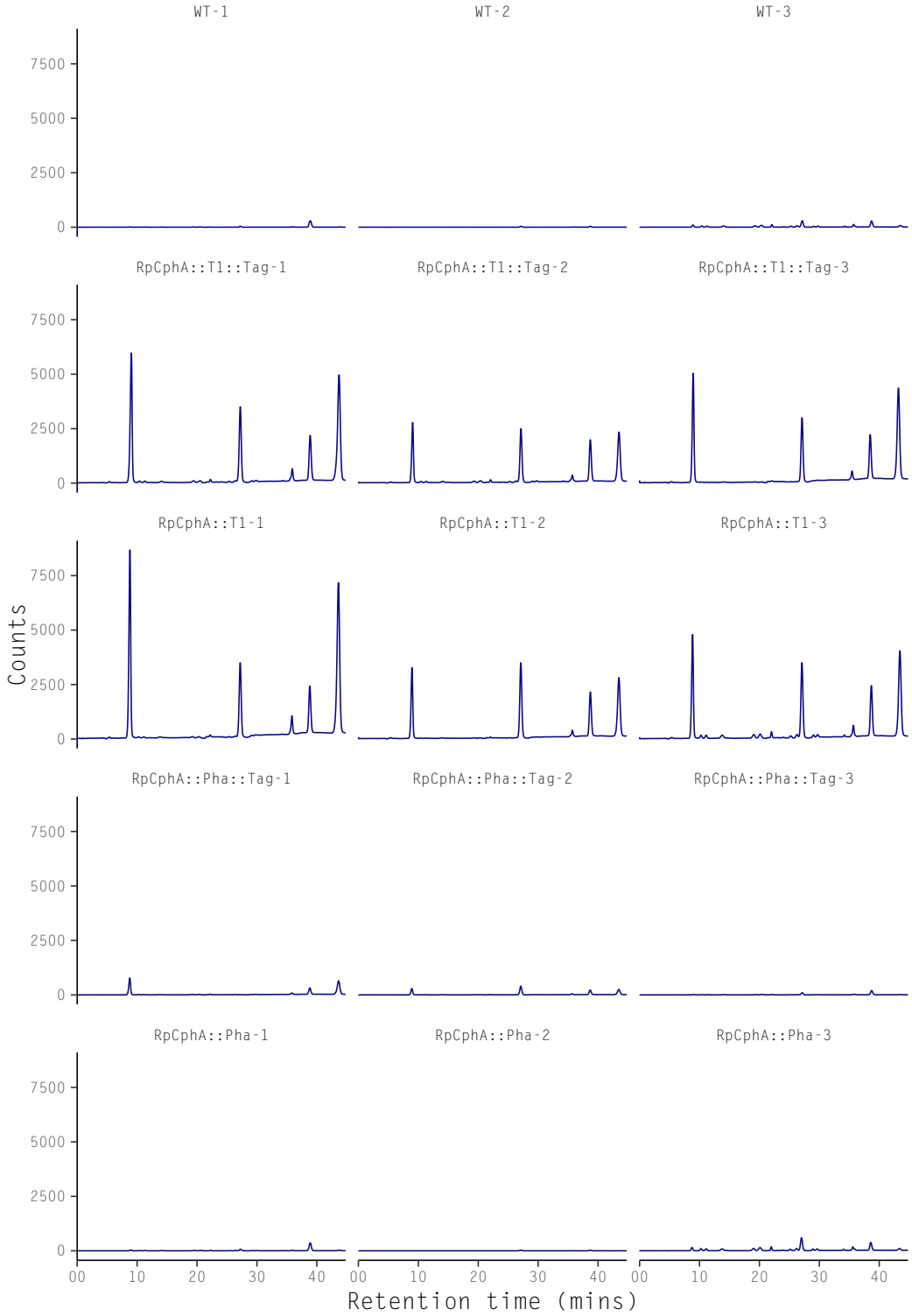


Figure 6.13. – Normalised ion-exchange chromatograms showing total amino acid composition of hydrolysed small-scale cyanophycin extractions. Ion-exchange chromatograms normalised by internal standard addition and amount loaded (dilution factor). Extractions are from three biological replicates of the four RpCphA strains and WT control. Strong peaks are present for aspartate and arginine (the principal components of cyanophycin) at the start and end of the chromatograms, respectively, for the RpCphA::T1 & RpCphA::T1::Tag strains. The other strong peaks are for the internal standard (Norleucine) and the ammonium released by the hydrolysis. A small peak representing lysine is also present.

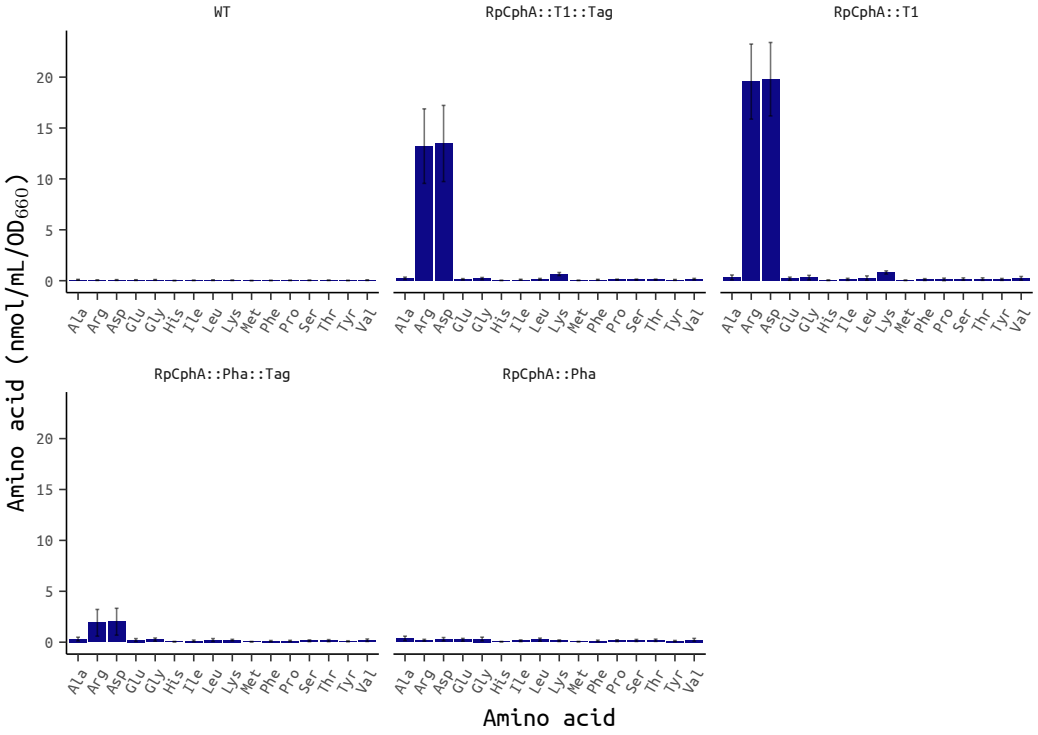
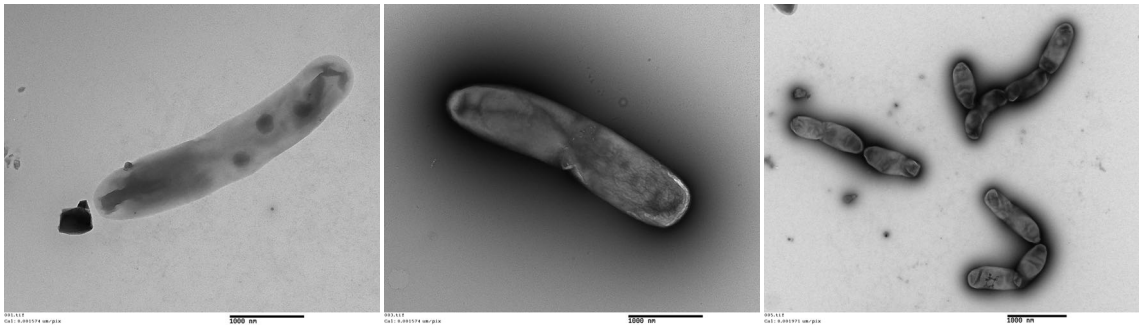
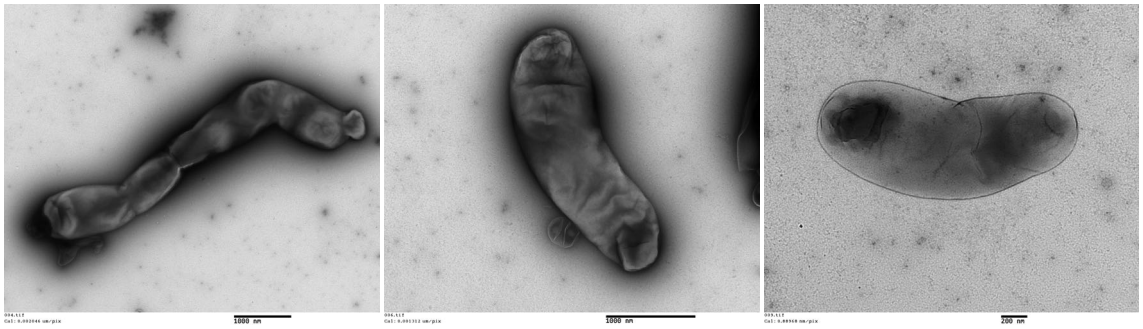


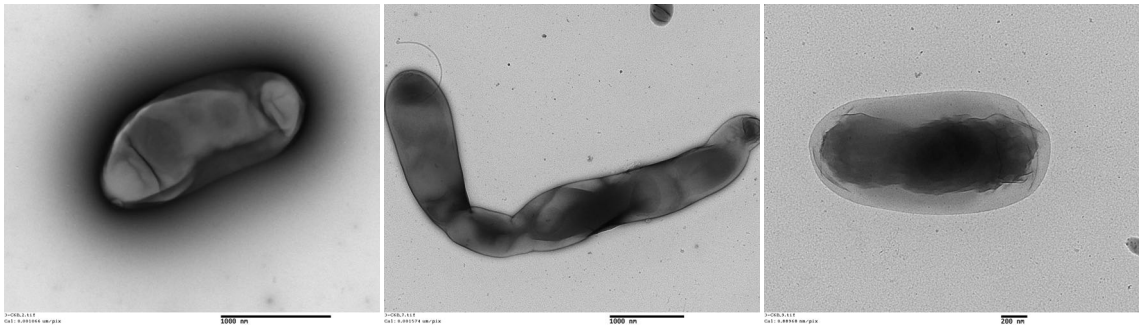
Figure 6.14. – Total amino acid composition of hydrolysed small-scale cyanophycin extraction. Amount of each amino acid present in hydrolysed small-scale cyanophycin extractions from three biological replicates of the four RpCphA strains and WT control. Calculated from ion-exchange chromatograms normalised by internal standard addition, amount loaded (dilution factor), extraction volume, and OD_{660} of original culture. The mean of three biological replicates is shown, and error bars represent the SEM. High levels of aspartate and arginine (the principal components of cyanophycin) are present in the RpCphA::T1 & RpCphA::T1::Tag strains, with other residual amino acids detected at trace levels in all samples. A slight increase in lysine above trace levels is also detected in the RpCphA::T1 & RpCphA::T1::Tag strains, consistent with recombinant cyanophycin production.



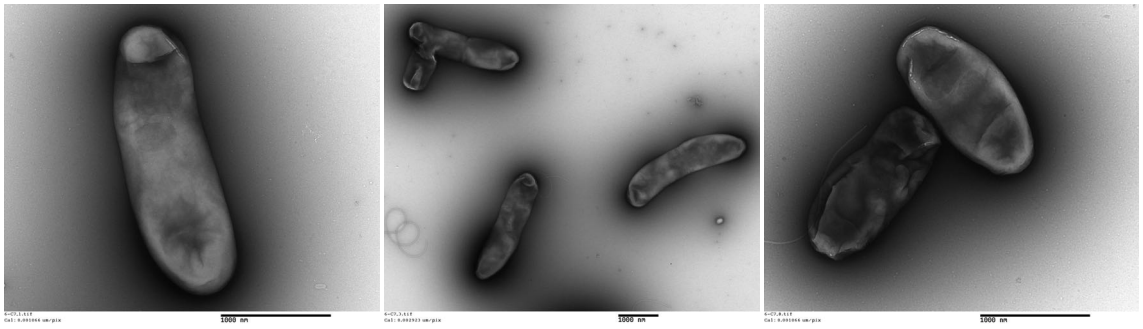
(a) WT



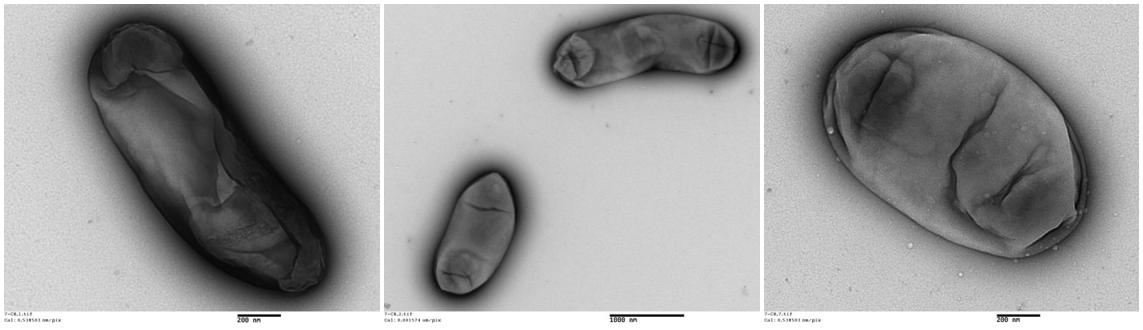
(b) RpCphA::T1::Tag



(c) RpCphA::T1



(d) RpCphA::Pha::Tag



(e) RpCphA::Pha

Figure 6.15. – TEM **negative stained micrographs.** Three different views of each sample are shown.

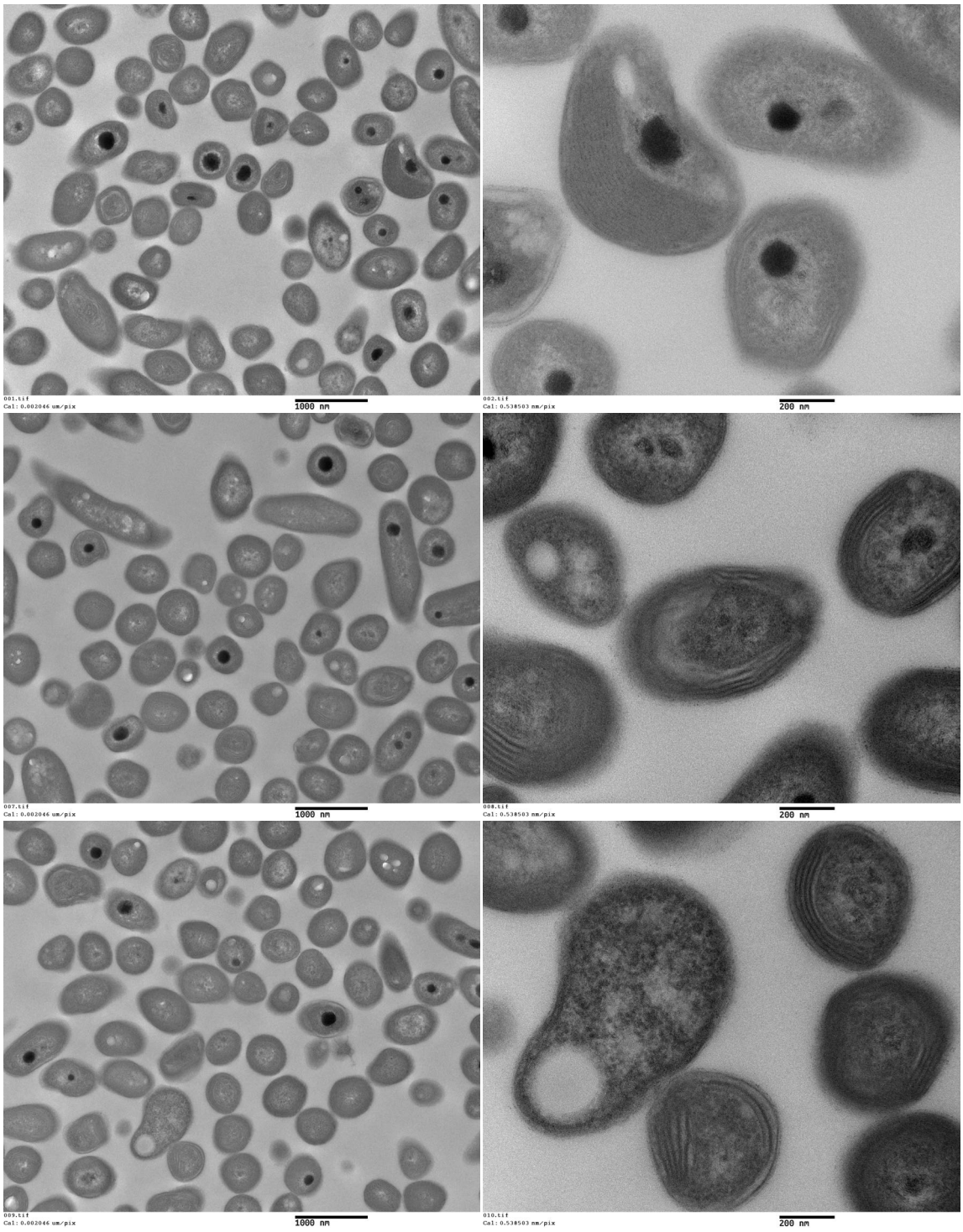


Figure 6.16. – TEM **of sectioned WT.** Three different views of the sample are shown (in column), with a lower magnification image on the left, and a higher magnification on the right.

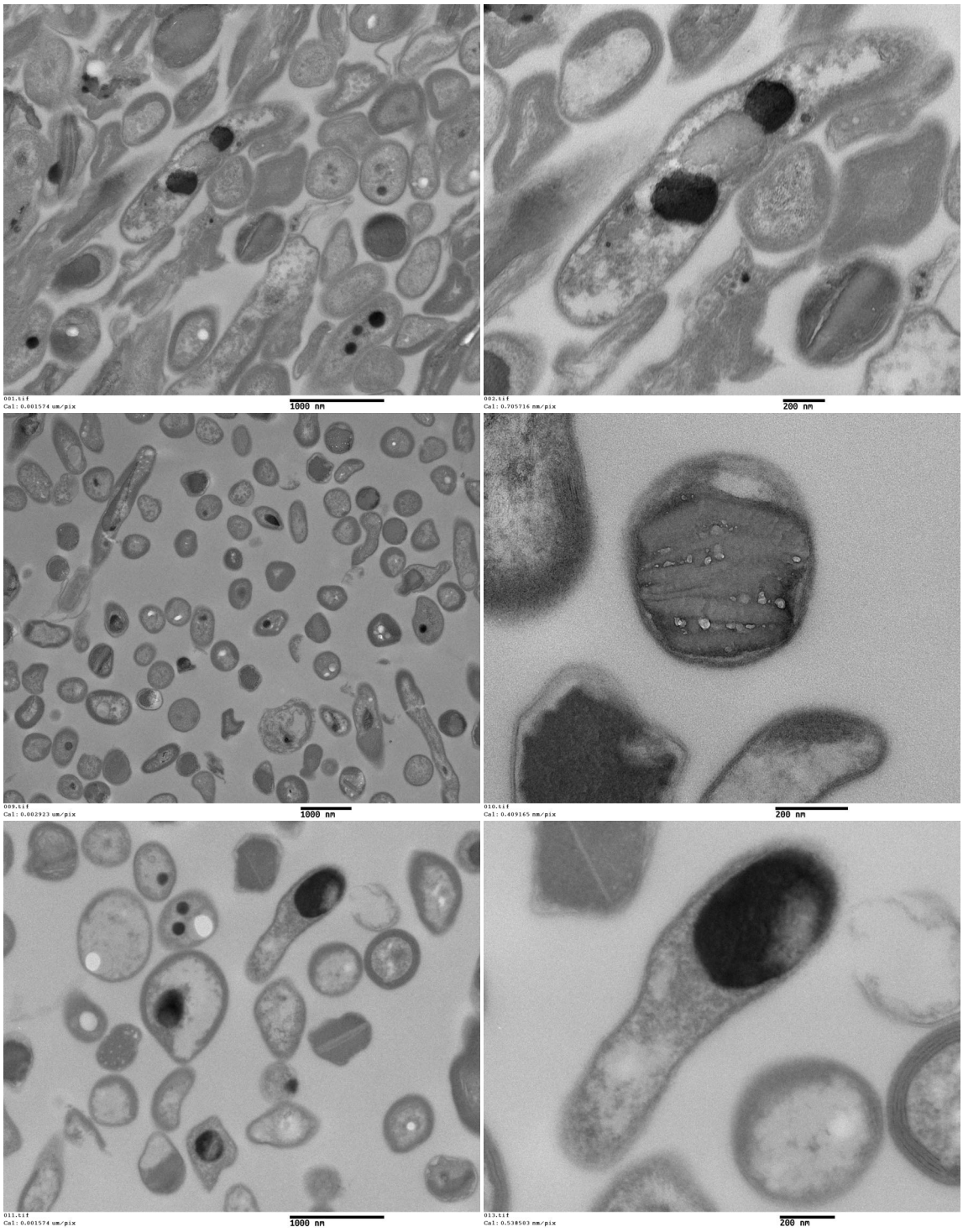


Figure 6.17. – TEM of sectioned RpCphA::T1::Tag. Three different views of the sample are shown (in column), with a lower magnification image on the left, and a higher magnification on the right.

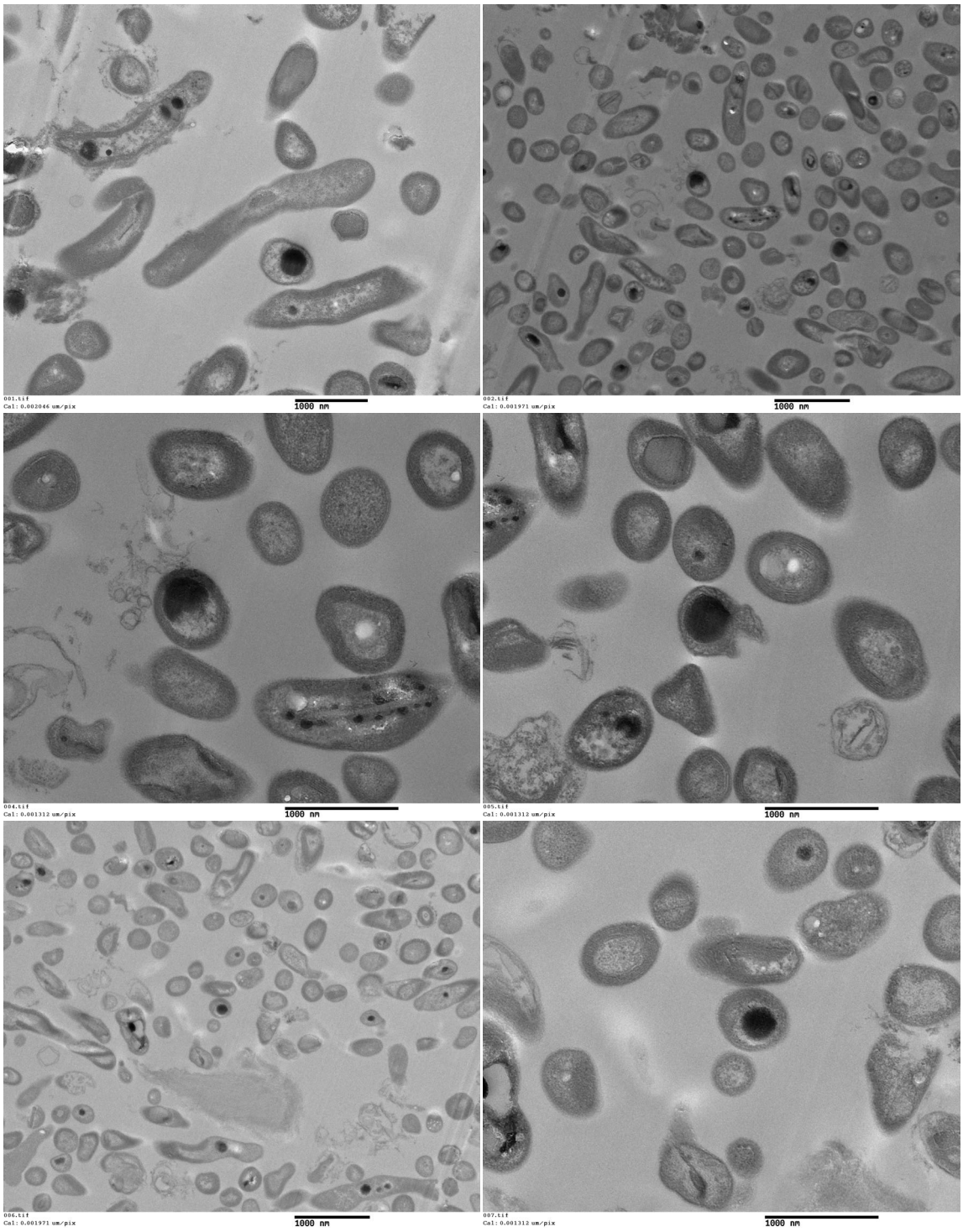


Figure 6.18. – TEM **of sectioned RpCphA::T1.** Three different views of the sample are shown (in column), with a lower magnification image on the left, and a higher magnification on the right.

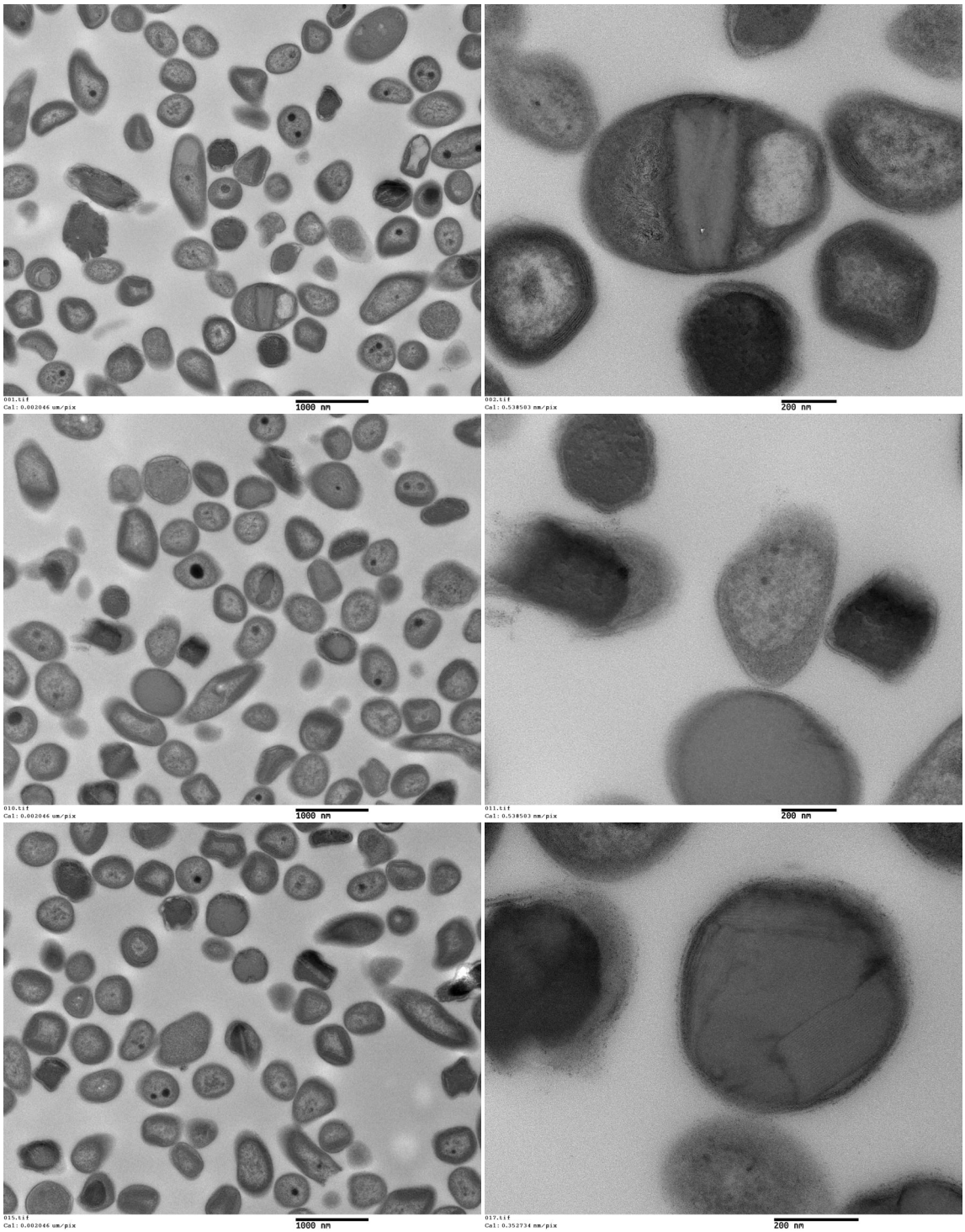


Figure 6.19. – TEM of sectioned RpCphA::Pha::Tag. Three different views of the sample are shown (in column), with a lower magnification image on the left, and a higher magnification on the right.

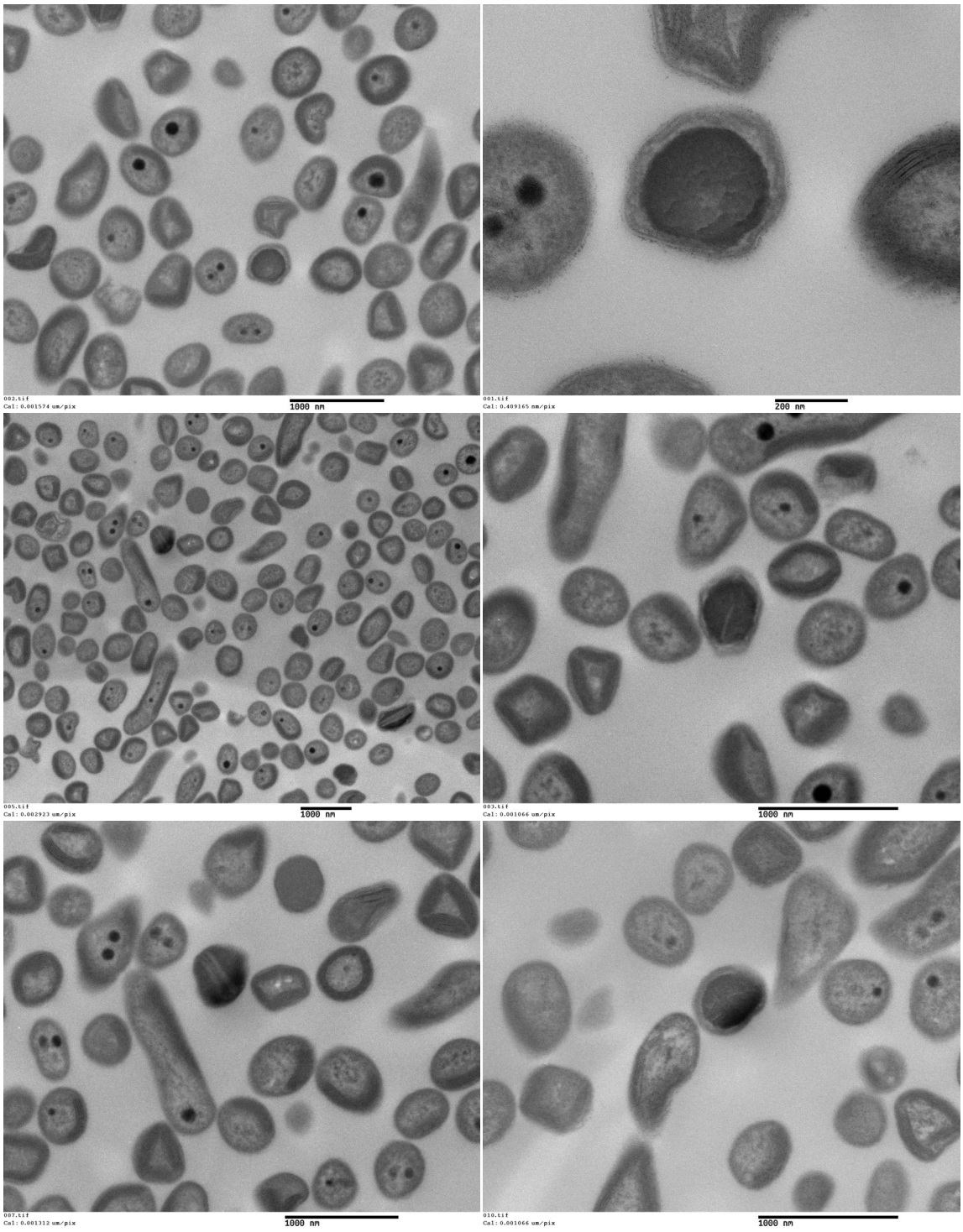


Figure 6.20. – TEM **of sectioned RpCphA::Pha.** Three different views of the sample are shown (in column), with a lower magnification image on the left, and a higher magnification on the right.

Table 6.4. – Cyanophycin extraction yields from 5 L cultures. Data presented are for a single sample of each strain.

	CDW (g)	Starting OD_{660}	Harvested OD_{660}	Change in OD_{660}	Extracted pellet (g)	Extraction as % CDW
WT	10.146	0.282	10.39	10.108	0.034	0.34
RpCphA::T1::Tag	6.365	0.250	7.33	7.080	0.559	8.78
RpCphA::T1	6.440	0.212	8.14	7.928	0.611	9.49
RpCphA::Pha::Tag	4.378	0.079	6.5	6.421	0.579	13.23
RpCphA::Pha	4.576	0.125	6.49	6.365	0.507	11.08

Large-scale extraction

Cyanophycin was extracted from 5 L of *R. palustris* culture. The cyanophycin yields (as a percentage of CDW) from the RpCphA strains were approximately 10 %, whereas only 0.3 % was obtained from the WT control strain, reflecting the baseline for the assay (Table 6.4. This represents an approximate yield of 100 mg · L^{-1}. Photographs of the harvested cultures prior-to and after freeze-drying are shown in Figure 6.21, and photographs of the cultures mid-extraction and the dried cyanophycin product can be seen in Figure 6.22.

6.4. Discussion

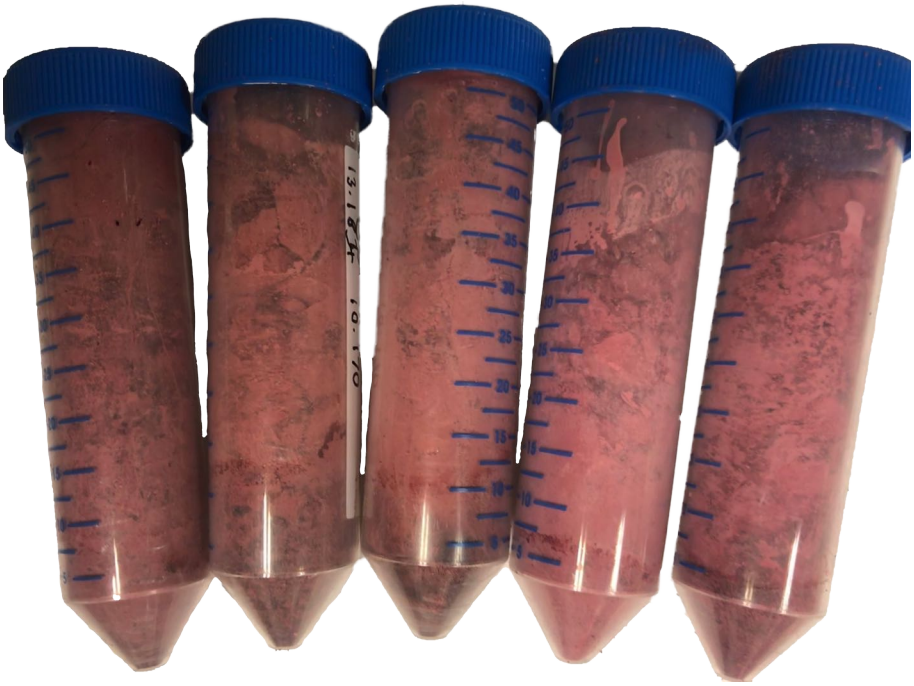
6.4.1. Construction of heterologous lines

The results presented show that *R. palustris* can be engineered to express functional heterologous proteins, with a level of expression sufficient to generate significant amounts of product as a percent of CDW. The results show that the heterologous genes were integrated into the genome (Figure 6.5), that the genes were expressed (Figure 6.6), that the protein was expressed from the transcript (Figure 6.7), and that it was functional (Section 6.3.2).

The citrate synthase promoter was chosen for this system as it was believed that as a gene of central metabolism it would constitutively have a reasonably high level of expression under all conditions. This prediction for citrate synthase expression was supported by data found in the literature, as reported by McKinlay *et al.* (2014) in their transcriptomic data.

It is believed that prior to this report, no study had been made of expression levels of heterologous gene expression in *R. palustris* under any promoter system. Whilst the citrate synthase promoter used in this study did allow expression of the inserted gene and functional protein (Figures 6.6, 6.7, & 6.8), it may be seen from the SDS-PAGE protein gels of the cell-lysate (Figure 6.9) that the inserted gene did not produce protein in quantities sufficient to see as an additional protein band, relative to the WT, at the size that would be expected for cyanophycin synthetase (~100 kDa). This indicates that although the citrate synthase synthase promoter is certainly effective, other promoter systems, that might give higher expression, should certainly be investigated for future heterologous expression

(a) 5 L *R. palustris* cultures ready to be harvested



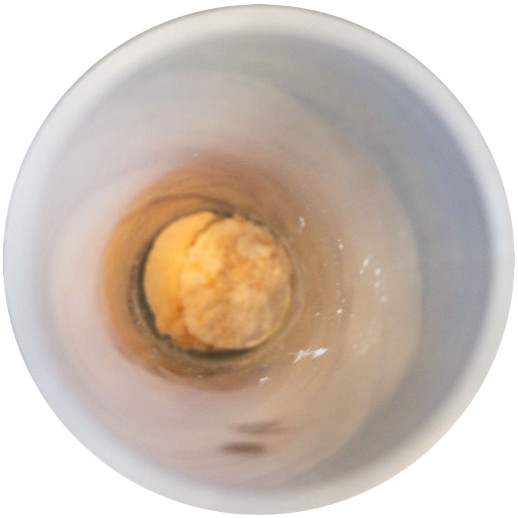
(b) Dried cells from 5 L *R. palustris* cultures

(c) Close-up of dried cells from 5 L *R. palustris* cultures

Figure 6.21. – 5 L *R. palustris* cultures for cyanophycin extraction.

(a) 5 L *R. palustris* cultures mid-extraction

(b) Dried cyanophycin extracted from 5 L *R. palustris* cultures



(c) Close-up of dried cyanophycin extracted from 5 L *R. palustris* cultures

Figure 6.22. – Cyanophycin extracted from 5 L *R. palustris* cultures.

systems in *R. palustris*. There may also be additional advantages to consider for other promoters, such as choosing non-constitutive or inducible promoters, which would allow control over the expression of the product.

In this study, the mutant strains produced still contained the dual marker system, as the plasmid was inserted *via* a single crossover event (selected for using kanamycin) (Figure 6.3). This means that the mutant strains were grown under constant selection, and it is possible that this could have had growth effects relative to the WT strain, which was not exposed to any selection agent. Furthermore, if the strains were to be used in an industrial setting, use of a selectable agent would not be desirable, due to cost and practical considerations. Additionally, it is likely that these strains might be somewhat unstable, as the presence of multiple regions of homology in the insert locus could lead to recombination events (as occurs when performing a second recombination event to produce marker-less strains, see Figure 6.3) which could lead to reversion to WT and loss of the inserted gene (see below). If cells are cultured for long periods of time (as is often the case when they are cultured in order to produce and harvest some product) the selectable agent will be depleted through metabolism and degradation, and thus may become less effective as time progresses. The cultured cells may also start to acquire natural resistance to the agent. In this case kanamycin was the selectable agent, which is an aminoglycoside and works by inhibiting protein synthesis (Recht *et al.*, 1999; Wilson, 2009; Orelle *et al.*, 2013). This means that cells in exponential growth phase, actively producing many proteins, are likely to be more sensitive to it than cells in late log-phase, meaning that older cultures will be less susceptible to it, and may more easily revert to WT. Markerless strains, with removal of the repeated sequences, may have increased stability and not need the selective agent to be present during growth.

6.4.2. Purification and characterisation of cyanophycin

The cyanophycin synthetase enzyme was successfully engineered into *R. palustris* in four different strains, and cyanophycin extractions performed on these strains, along with WT as a negative control. The products of these extractions were analysed and the results strongly indicate that cyanophycin was produced by the recombinant strains.

SDS-PAGE protein gels of the cell-lysate and cyanophycin extract (Figures 6.9 & 6.10) showed that the cyanophycin polymer produced by the mutant strains was between 15 kDa to 25 kDa. This is slightly smaller than that made by native cyanophycin producers (Simon, 1971; 1973; Simon and Weathers, 1976; Allen, 1984; Berg *et al.*, 2000), which tend to produce cyanophycin in the 25 kDa to 100 kDa range, but typical of other recombinant systems (Frey *et al.*, 2002; Voss *et al.*, 2004; Füser and Steinbüchel, 2005; Hai *et al.*, 2006; Steinle and Steinbüchel, 2010; Frommeyer and Steinbüchel, 2013; Tseng *et al.*, 2013; Zhang *et al.*, 2013).

Amino acid analysis of the cyanophycin extracts (Figures 6.12, 6.13, & 6.14) showed that they contained approximately equimolar ratios of arginine and aspartate, as expected for cyanophycin, with the possible addition of a small quantity of lysine, again typical of other recombinant systems (Frommeyer and Steinbüchel, 2013). Other amino acids were detected in the extracts at trace levels. However, these were also seen at similar levels in the WT control, indicating that the cyanophycin purification method does pick up some other contaminant cell materials. These amino acids therefore probably represent contamination, rather than genuine incorporation into cyanophycin.

TEM micrographs were used to try to establish the presence of cyanophycin granules inside the mutant strains. Negative staining with uranyl acetate showed only the outside of the cells, and no difference was visible between the mutant strains and the WT (Figure 6.15). Micrographs of cells that had been fixed and sectioned, to show cross-sections of the cells, showed small, circular, dark granules in all of the strains, including WT (Figures 6.16, 6.17, 6.18, 6.19, & 6.20), which were putatively identified as polyphosphate granules (Adessi and Spini *et al.*, 2016). More interestingly, in the mutant strains, some cells appeared to contain larger, slightly less regular, dark patches (Figures 6.17, 6.18, 6.19, & 6.20), which did not appear to be present in the WT cells (Figure 6.16). It is possible that these dark patches could be evidence of cyanophycin granules, although it is difficult to be certain, as their distribution seems to vary considerably.

Cyanophycin extraction from "large-scale" 5 L cultures showed that the mutant strains gave a cyanophycin yield of approximately 10 % of CDW (Table 6.4). The yield calculated from the extraction from the "mid-scale" 1 L cultures is unlikely to be accurate due to the extremely small amounts of cyanophycin recovered and the difficulties of accurately weighing such small masses (Table 6.3). It is not clear why the "mid-scale" cultures gave such low yields.

In the "small-scale" cyanophycin extractions, only the two RpCphA::T1 strains produced significant amounts of cyanophycin (Figures 6.8 & 6.14) while the RpCphA::Pha:Tag appeared to produce only a very small amount, and the RpCphA::Pha strain did not appear to produce any. This can also be seen in the SDS-PAGE protein gels of the cell-lysate and cyanophycin extract (Figures 6.9 & 6.10). However, in the "large-scale" cyanophycin extractions, all four strains appeared to be producing equivalent quantities of cyanophycin (Table 6.4). The presence or absence of the C-terminal tag also did not appear to have any effect on cyanophycin yield (Table 6.4).

One possible explanation for the discrepancy in cyanophycin production between the small and large cultures of the RpCphA::Pha strains could be to do with reversion to WT, as discussed above. This seems more likely than any difference caused by the different insertion loci, as the yields from the large scale cultures were equivalent. The idea to use the two different insertion loci was based on the thought that replacing the synthesis of polyhydroxybutyrate (also a nutrient storage polymer) with the synthesis of cyanophycin

might lead to higher cyanophycin yields, due to less competition between the two storage polymers for substrates. However, in these experiments only the marked, single crossover strains were created, meaning that the WT locus was not deleted, as this would have required the second crossover (Figure 6.3). Therefore, polyhydroxybutyrate should still have been produced in the RpCphA::Pha strains, just as it would have been in the RpCphA::T1 strains, and as such, the cyanophycin yield of the strains of both insertion loci could reasonably have been expected to remain the same.

The "small-scale" cyanophycin cultures were subcultured many times for many experiments, and this is the likely cause of reversion or partial reversion to WT. Subculturing the "large-scale" 5 L cultures, in order to repeat the "large-scale" cyanophycin extraction, resulted in complete loss of cyanophycin yield (data not shown), suggesting again that subculturing old cultures resulted in reversion to WT. Genotyping samples of these cultures by PCR yielded only very faint bands for the inserted gene, when compared to the positive control reaction, and to the previous genotyping experiments carried out when the strains were first made (data not shown), further supporting the idea that some cells in the cultures had lost the inserted gene.

6.5. Conclusions

Evidence is presented of the successful expression of functional heterologous genes in *R. palustris*, showing that it may successfully be used as a biotechnological "cell-factory" platform for the production of desirable compounds.

Production of cyanophycin, a valuable polymer that could be be used as a sustainable replacement for some petrochemical compounds in industry, was achieved at a yield of approximately 10 % of CDW. Given the ability of *R. palustris* to metabolise both ammonia and urea (Chapter 3), it could potentially be used to produce cyanophycin (or other compounds) from growth on agricultural (or other) wastes.

For future research, other promoter systems should be investigated to try to improve yields, and the strains should be further engineered by performing a second crossover event, to produce marker-less strains that should be more stable. It would be important to test the stability in future work. Generation of marker-less recombinant strains would also facilitate the removal of endogenous pathways, that might compete with the inserted pathways for cellular resources and affect yields.

7 Discussion and Future Work

7.1. Growth on waste substrates

Growth of *R. palustris* on a range of different simple substrates was trialled, and successful growth on a substantial range of these reinforced the idea of the broad metabolic diversity of *R. palustris*. Both photoheterotrophic growth and photoautotrophic growth were demonstrated under anaerobic conditions, and urea was identified as a new nitrogen source and substrate, in addition to several plant derived sugars including cellobiose and xylose. This suggested a possible use for *R. palustris* in the degradation and bioremediation of agricultural wastes, which currently poses severe environmental and economic problems.

Investigations of photoautotrophic growth, the absorption spectra of *R. palustris* and the emission spectra of various light sources showed that infra-red light was important for rapid growth of *R. palustris* in laboratory conditions. Photoautotrophic growth of *R. palustris* was demonstrated when an artificial electron source, thiosulphate, was supplied, and when the light-source contained infra-red wavelengths. Photoautotrophic growth was shown to occur by the fixation of carbon released by the metabolism of urea, providing further evidence of the great metabolic versatility possessed by *R. palustris*, and of urea as an effectively used substrate.

7.2. Increasing glycerol and urea metabolism

A new expression system was developed, which allowed the integration of heterologous genes into the *R. palustris* genome under control of the native citrate synthase promoter and terminator. Both marked and unmarked *R. palustris* mutant strains were developed, and a putative neutral insert site (the transposon RPA0768) was tested, with no adverse effects noted. This is a useful addition to the 'genetic toolbox' of *R. palustris* genetic manipultaion, as it allows heterologous genes to be inserted into the genome, and the construction of stable, marker-less strains.

A heterologous gene, encoding the glycerol transporter GlpF, under the control of a promoter/terminator system from the native *R. palustris* citrate synthase gene was suc-

cessfully introduced. Transcription was demonstrated. Evidence of protein expression or activity from this transcript was inconclusive due to the unexplained positive signal from the untagged protein in the western blot (rendering the positive signal from the tagged protein difficult to interpret reliably). No phenotype was observed. There are a number of possible reasons for this. It could be that the GlpF protein was not expressed. It could also be that the protein was expressed but was not functional. Finally, it could be that the protein was expressed and functional, but that GlpF was not rate limiting.

The sequence encoding the native *R. palustris* AqpZ aquaporin was modified and tagged, and reintroduced into *R. palustris*. However, no change in phenotype was observed and detection of the protein using the inserted tag was unsuccessful.

It is speculated that glycerol transport into *R. palustris* cells is not rate-limiting in terms of substrate consumption or hydrogen yield, under the conditions tested, but this conclusion depends on whether the heterologous GlpF or modified AqpZ proteins were produced and functional. Further work is required to determine if the GlpF and AqpZ proteins were indeed expressed. If glycerol transport is indeed not rate-limiting, this suggests that future efforts to engineer *R. palustris* to improve its utilisation of substrates should focus on downstream elements of metabolism, rather than transport of substrates into the cell (at least in the case of glycerol, and probably urea too).

7.3. Regulation of urea metabolism

A study of differential gene expression between *R. palustris* cultures grown with urea or ammonium sulphate as the nitrogen source found that cultures grown on urea showed much higher levels of transcripts for nitrogenase than cultures grown with ammonia. This result is similar to that observed in *Rhodobacter capsulatus*, a related alphaproteobacteria species (Masepohl and Kaiser *et al.*, 2001), and has interesting implications, from both a practical and a basic biological standpoint.

From an industrial point of view, expression and activity of nitrogenase is highly desirable, since the by-product of nitrogen fixation is hydrogen gas; an efficient and clean-burning fuel. However, nitrogenase activity is easily suppressed by the presence of oxygen or fixed nitrogen, which can severely limit yields when attempting to produce hydrogen from waste materials. The discovery of a fixed nitrogen source, that is common in agricultural waste (a potential substrate), and which does not cause suppression of nitrogenase expression is therefore of significant interest. It will, however, be important to determine if the levels of nitrogenase transcripts are associated with similar levels of active protein, and whether urea activates any post-translational suppression of nitrogenase activity.

From a basic biological point of view, the question arises as to how nitrogenase suppression is avoided when urea is present, since urea releases ammonia intracellularly when it is metabolised. A speculative explanation of this phenomenon (along similar lines to

Masepohl and Kaiser *et al.* (2001)) could be that consumption of ammonia by glutamine synthase could outstrip its production by urease, thus leading to a low intracellular concentration/pool of free ammonia for detection by the transcriptional regulatory systems that affect nitrogenase expression.

Many genes were highly up-regulated when *R. palustris* was grown in the presence of urea, and a more thorough investigation of the transcriptome dataset presented here is needed in order to gain a better understanding of the nitrogen metabolism pathways and regulatory systems.

The discovery here of highly expressed, and potentially inducible, genes will be beneficial for future metabolic engineering of *R. palustris*, as their promoter systems and regulatory networks could be used in the design of future gene editing and heterologous expression systems. It will be important to identify the sequences involved in transcriptional regulation of these genes.

7.4. Production of cyanophycin

The high-value compound cyanophycin, a non-ribosomal polypeptide consisting of arginine and poly-aspartate, was identified as being suitable for production by *R. palustris* grown in a bio-remediation system. This was due to its high nitrogen content, which would allow it to act as a nitrogen sink for the purposes of de-nitrifying nitrogen rich wastes, such as agricultural waste. The pathway for cyanophycin synthesis involved just one enzyme, and was engineered into *R. palustris* using the citrate synthase promoter expression system developed earlier in this study.

Four different strains were developed by inserting both tagged and untagged cyanophycin synthetase genes at each of two different genomic loci. No adverse effects on growth of *R. palustris* were observed as a result of these insertions. The gene was shown to be transcribed in all four strains, and two tagged strains were used to show that protein was produced from the transcripts.

A product was purified from the new strains grown on medium containing glycerol and urea, tested, and identified as cyanophycin, with a characteristic molecular weight of ~30 kDa.

Cultures of the new strains were grown at a 5 L scale under greenhouse conditions, and cyanophycin was extracted, and found to account for ~10 % of the CDW of the 5 L culture, or ~100 mg · L^{-1} of culture.

This demonstration of *R. palustris* growth, with inexpensive glycerol and urea as the substrate, and the high-yield production of a valuable compound, cyanophycin, is a significant proof of concept for the bio-remediation of waste by *R. palustris*. An important feature is that the growth was done under natural lighting, showing that this system can be run with minimal energy inputs.

Another advantage of *R. palustris* based production of cyanophycin, is that it is likely that hydrogen gas would be produced in addition to the cyanophycin. Hydrogen gas was not measured directly in this study, but the transcriptomic work with urea did show that nitrogenase transcription was up-regulated in cultures grown with urea, relative to cultures grown with ammonia. Since the cyanophycin cultures were grown with the same urea and glycerol medium, it is likely that they too expressed nitrogenase. In addition to this, it was noticeable on opening the culture bottles that there had been a build-up of gas. It will be important to assay for hydrogen production in future work.

The yield and characteristics of the extracted cyanophycin were similar to those of other recombinant systems (Hai *et al.*, 2006; Mooibroek *et al.*, 2007; Kroll and Steinle *et al.*, 2009; Steinle and Witthoff *et al.*, 2010; Frommeyer and Steinbüchel, 2013), although higher yields (to a maximum of ~30 %) have been obtained in well developed systems on rich/complex media (Frey *et al.*, 2002; Elbahloul and Steinbüchel, 2006; Kroll and Klinter *et al.*, 2011; Frommeyer and Steinbüchel, 2013; Tseng *et al.*, 2013). There is significant scope for further optimisation of cyanophycin production and extraction from recombinant *R. palustris*, and higher yields should be easily obtainable.

Different promoter systems could be tested, to see if an increase in the expression of cyanophycin synthetase would also result in an increase in cyanophycin yield. The promoters from genes that were found to be highly up-regulated and highly expressed in urea grown cultures might be useful for this approach. Codon optimisation of the cyanophycin synthetase gene might also improve cyanophycin synthetase expression, and other studies have found that cyanophycin synthetase genes from different cyanobacteria result in different yields when used in recombinant systems, and shown that optimisation of the gene/protein sequence can improve yields (Hai *et al.*, 2006; Steinle and Witthoff *et al.*, 2010).

The RpCphA::Pha strains developed in this study were designed to interrupt genes for polyhydroxybutyrate synthesis, but since only the single crossover strains were developed the WT copy of these genes would have remained intact. Generating the unmarked strains from the marked strains, by performing the second crossover, would facilitate the permanent interruption of these genes, resulting in the loss of polyhydroxybutyrate synthesis. This could help to increase cyanophycin by removing polyhydroxybutyrate as a resource sink, making more resources available for cyanophycin synthesis.

A metabolomic analysis of how cyanophycin synthesis affects the nitrogen balance throughout the cell would also be very useful, in order to determine how cyanophycin yields could be improved. This could be used to inform further optimisations of growth conditions or strategies for additional engineering to divert more of the cells resources to cyanophycin production. The effect of different nitrogen sources (and indeed substrates in general) on cyanophycin yield should also be investigated.

So far, all of the experiments described have been performed using optimised minimal

medium and pure substrates. For future use in bio-remediation systems however, trials will need to be made using real waste substrates. Crude glycerol and various agricultural wastes should be used as substrates, so that a real industrial system can be developed. One intriguing possibility that could be investigated is the mixing of crude glycerol with agricultural waste, for use as substrate. In this way, glycerol could provide the carbon source, and the agricultural waste could provide nitrogen (from ammonia and urea) and other nutrients. This could have several advantages; firstly a complete substrate could be developed, with no artificial supplements required, secondly, the agricultural waste might make the crude glycerol less toxic *via* dilution, and thirdly, the toxic nature of the crude glycerol might partially sterilise the agricultural waste, so that there would be fewer competing micro-organisms to compete with the *R. palustris* culture. Finally some interesting work from the Howe lab (unpublished data), has shown that crude glycerol can be used for cell lysis of plant and algal materials. This could indicate that the crude glycerol could be used as a pre-treatment, to extract nutrients from partially digested plant matter in agricultural waste that could then be used to feed bacterial cultures like *R. palustris*.

An alternative application for crude glycerol might even be to use highly concentrated crude glycerol to aid in the lysis of *R. palustris* when trying to extract desirable products such as cyanophycin. *R. palustris* is particularly resistant to lysis (presumably this helps to facilitate its survival in harsh or toxic environments), so efficient extraction of desirable compounds will be an important challenge to address.

Other studies have made several variations and refinements of the cyanophycin extraction protocol, and these should be used to guide an optimisation of the extraction of cyanophycin from *R. palustris* (Steinle and Steinbüchel, 2010). Additionally, in the study presented here, only the insoluble fraction of cyanophycin was extracted. There is also a soluble fraction that was not recovered with the extraction method used here (Füser and Steinbüchel, 2005; Frommeyer and Steinbüchel, 2013; Wiefel and Steinbüchel, 2014). A method should be developed to extract the soluble fraction, and assess what portion of the total cyanophycin yield it represents, so that the most efficient extraction method may be determined.

In this study, the cyanophycin producing strains developed, were marked strains, and were found to be unstable over time. Earlier work however, showed that the same transformation system could be used to produce stable, unmarked strains. The development of stable, unmarked, cyanophycin producing strains, that do not require selection agents, and are stable over long timeframes, should be a priority for future research.

7.5. Final Conclusions

This study has highlighted the potential for the use of *R. palustris* in bio-remediation applications. New substrates have been identified, and waste materials likely to be suitable

for bio-conversion by *R. palustris* proposed. An heterologous gene expression system for *R. palustris*, with insertion of transgenes into the genome, has been developed and demonstrated, further enhancing the capabilities and potential of *R. palustris* as a platform for bio-technology applications and research. The gene expression and regulation of key nitrogen metabolism pathways has been investigated, leading to insights that could be used to increase the range of conditions under which hydrogen may be produced by *R. palustris*, and possibly to increase hydrogen yields. Finally, the production of cyanophycin, a potentially valuable industrial polymer, is demonstrated for the first time in recombinant *R. palustris*, only using minimal medium, natural lighting, and producing a high yield, probably with the simultaneous production of hydrogen gas, a valuable and clean-burning fuel. This represents an attractive proof of principle for the expression of other useful products in R. palustris.

Bibliography

Aboulmagd, E., F. B. Oppermann-Sanio and A. Steinbüchel (2000). Molecular characterization of the cyanophycin synthetase from *Synechocystis* sp. strain PCC6308. *Archives of Microbiology* **174,** 297–306 (cit. on p. 127).

– (2001). Purification of *Synechocystis* sp. strain PCC6308 cyanophycin synthetase and its characterization with respect to substrate and primer specificity. *Applied and environmental microbiology* **67,** 2176–2182 (cit. on p. 127).

Achenbach, L. A., J. Carey and M. T. Madigan (2001). Photosynthetic and phylogenetic primers for detection of anoxygenic phototrophs in natural environments. *Applied and environmental microbiology* **67,** 2922–2926 (cit. on p. 27).

Adessi, A., J. B. McKinlay, C. S. Harwood and R. De Philippis (2012). A *Rhodopseudomonas palustris nifA** mutant produces H_2 from NH_4^+-containing vegetable wastes. *International Journal of Hydrogen Energy* **37,** 15893–15900 (cit. on pp. 27, 29, 34).

Adessi, A., G. Spini, L. Presta, A. Mengoni, C. Viti, L. Giovannetti, R. Fani and R. De Philippis (2016). Draft genome sequence and overview of the purple non sulfur bacterium *Rhodopseudomonas palustris* 42OL. *Standards in Genomic Sciences* **11,** 24 (cit. on p. 159).

Aguirre-Villegas, H., R. A. Larson and D. J. Reinemann (2014). From waste-to-worth: energy, emissions, and nutrient implications of manure processing pathways. *Biofuels, Bioproducts and Biorefining* **8,** 770–793 (cit. on p. 126).

Allen, M. M. (1984). Cyanobacterial Cell Inclusions. *Annual Review of Microbiology* **38,** 1–25 (cit. on pp. 45, 127, 133, 158).

Aravind, J., T. Saranya, G. Sudha and P. Kanmani (2016). A Mini Review on Cyanophycin: Production, Analysis and Its Applications. *Integrated Waste Management in India. Environmental Science and Engineering.* Springer-Verlag, 49–58 (cit. on p. 127).

Asada, Y., M. Tokumoto, Y. Aihara, M. Oku, K. Ishimi, T. Wakayama, J. Miyake, M. Tomiyama and H. Kohno (2006). Hydrogen production by co-cultures of *Lactobacillus* and a photosynthetic bacterium, *Rhodobacter sphaeroides* RV. *International Journal of Hydrogen Energy* **31** (11), 1509–1513 (cit. on p. 33).

Austin, S., W. S. Kontur, A. Ulbrich, J. Z. Oshlag, W. Zhang, A. Higbee, Y. Zhang, J. J. Coon, D. B. Hodge, T. J. Donohue and D. R. Noguera (2015). Metabolism of multiple aromatic compounds in corn stover hydrolysate by *Rhodopseudomonas palustris*. *Environmental Science & Technology* **49** (14), 8914–8922 (cit. on p. 49).

Basak, N. and D. Das (2006). The Prospect of Purple Non-Sulfur (PNS) Photosynthetic Bacteria for Hydrogen Production: The Present State of the Art. *World Journal of Microbiology and Biotechnology* **23,** 31–42 (cit. on pp. 32, 91).

Beitz, E., B. Wu, L. M. Holm, J. E. Schultz and T. Zeuthen (2006). Point mutations in the aromatic/arginine region in aquaporin 1 allow passage of urea, glycerol, ammonia, and protons. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 269–274 (cit. on p. 64).

Berg, H., K. Ziegler, K. Piotukh, K. Baier, W. Lockau and R. Volkmer-Engert (2000). Biosynthesis of the cyanobacterial reserve polymer multi-L-arginyl-poly-L-aspartic acid (cyanophycin). *European Journal of Biochemistry* **267,** 5561–5570 (cit. on pp. 127, 158).

Berne, C., B. Allainmat and D. Garcia (2005). Tributyl phosphate degradation by *Rhodopseudomonas palustris* and other photosynthetic bacteria. *Biotechnology Letters* **27,** 561–566 (cit. on p. 30).

Blakeley, R. L. and B. Zerner (1984). Jack bean urease: the first nickel enzyme. *Journal of Molecular Catalysis* **23,** 263–292 (cit. on pp. 50, 92).

Bland, J. M. and D. G. Altman (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1,** 307–310 (cit. on p. 94).

Borgnia, M. J. and P. Agre (2001). Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 2888–2893 (cit. on p. 88).

Bose, A., E. J. Gardel, C. Vidoudez, E. A. Parra and P. R. Girguis (2014). Electron uptake by iron-oxidizing phototrophic bacteria. *Nature communications* **5,** 3391 (cit. on p. 29).

Brotosudarmo, T. H. P., A. M. Collins, A. Gall, A. W. Roszak, A. T. Gardiner, R. E. Blankenship and R. J. Cogdell (2011). The light intensity under which cells are grown controls the type of peripheral light-harvesting complexes that are assembled in a purple photosynthetic bacterium. *Biochemical Journal* **440,** 51–61 (cit. on pp. 51, 56).

Browne, J. D., S. Gilkinson and J. P. Frost (2015). The effects of storage time and temperature on biogas production from dairy cow slurry. *Biosystems Engineering* **129,** 48–56 (cit. on p. 126).

Bryant, D. A. and N.-U. Frigaard (2006). Prokaryotic photosynthesis and phototrophy illuminated. *Trends in microbiology* **14,** 488–496 (cit. on p. 27).

Burt, T. and N. Haycock (1991). Farming and nitrate pollution. *Geography* **76,** 60–63 (cit. on pp. 125, 126).

Burton, C. H. and C. Turner (2003). Manure management : treatment strategies for sustainable agriculture. Silsoe Research Institute, 451 (cit. on p. 125).

Carlozzi, P. (2009). The effect of irradiance growing on hydrogen photoevolution and on the kinetic growth in *Rhodopseudomonas palustris*, strain 42OL. *International Journal of Hydrogen Energy* **34,** 7949–7958 (cit. on p. 27).

Carlozzi, P. and M. Lambardi (2009). Fed-batch operation for bio-H_2 production by *Rhodopseudomonas palustris* (strain 42OL). *Renewable Energy* **34,** 2577–2584 (cit. on p. 27).

Carlozzi, P., C. Pintucci, R. Piccardi, A. Buccioni, S. Minieri and M. Lambardi (2010). Green energy from *Rhodopseudomonas palustris* grown at low to high irradiance values, under fed-batch operational conditions. *Biotechnology Letters* **32,** 477–481 (cit. on pp. 27, 91).

Carpenter, S. R., N. F. Caraco, D. L. Correll, R. W. Howarth, A. N. Sharpley and V. H. Smith (1998). Nonpoint Pollution of Surface Waters with Phosphorus and Nitrogen. *Ecological Applications* **8,** 559 (cit. on pp. 125, 126).

Cavalli, D., G. Cabassi, L. Borrelli, R. Fuccella, L. Degano, L. Bechini and P. Marino (2014). Nitrogen fertiliser value of digested dairy cow slurry, its liquid and solid fractions, and of dairy cow slurry. *Italian Journal of Agronomy* **9,** 71 (cit. on p. 126).

Chadwick, D., S. G. Sommer, R. Thorman, D. Fangueiro, L. Cardenas, B. Amon and T. H. Misselbrook (2011). Manure management: Implications for greenhouse gas emissions. *Animal Feed Science and Technology* **166-167,** 514–531 (cit. on p. 125).

Chen, C.-Y., W.-B. Lu, C.-H. Liu and J.-S. Chang (2008). Improved phototrophic H_2 production with *Rhodopseudomonas palustris* WP3-5 using acetate and butyrate as dual carbon substrates. *Bioresource technology* **99,** 3609–3616 (cit. on p. 49).

Chen, Y., J. J. Cheng and K. S. Creamer (2008). Inhibition of anaerobic digestion process: a review. *Bioresource technology* **99,** 4044–4064 (cit. on pp. 26, 60).

Chen, Y.-T., S.-C. Wu and C. M. Lee (2012). Relationship between cell growth, hydrogen production and poly-β-hydroxybutyrate (PHB) accumulation by *Rhodopseudomonas palustris* WP3-5. *International Journal of Hydrogen Energy* **37,** 13887–13894 (cit. on p. 27).

Christensen, M. L., K. V. Christensen and S. G. Sommer (2013). Solid-Liquid Separation of Animal Slurry. *Animal Manure Recycling.* John Wiley & Sons, Ltd, 105–130 (cit. on pp. 125, 126).

Colica, G., S. Caparrotta and R. De Philippis (2012). Selective biosorption and recovery of Ruthenium from industrial effluents with *Rhodopseudomonas palustris* strains. *Applied Microbiology and Biotechnology* **95,** 381–387 (cit. on pp. 29, 60).

Dixon, R. and D. Kahn (2004). Genetic regulation of biological nitrogen fixation. (Cit. on pp. 33, 34).

Dönmez, G. Ç., A. Öztürk and L. Çakmakçi (1999). Properties of the *Rhodopseudomonas palustris* Strains Isolated From an Alkaline Lake in Turkey. *Turkish Journal of Biology* **23,** 457–563 (cit. on pp. 27, 49).

Doud, D. F. R., E. C. Holmes, H. Richter, B. Molitor, G. Jander and L. T. Angenent (2017). Metabolic engineering of *Rhodopseudomonas palustris* for the obligate reduction of n-butyrate to n-butanol. *Biotechnology for Biofuels* **10,** 178 (cit. on p. 31).

Drepper, T., S. Groß, A. F. Yakunin, P. C. Hallenbeck, B. Masepohl and W. Klipp (2003). Role of GlnB and GlnK in ammonium control of both nitrogenase systems in the pho-

totrophic bacterium *Rhodobacter capsulatus*. *Microbiology* **149** (8), 2203–2212 (cit. on p. 34).

Drepper, T., J. Wiethaus, D. Giaourakis, S. Gross, B. Schubert, M. Vogt, Y. Wiencek, A. G. McEwan and B. Masepohl (2006). Cross-talk towards the response regulator NtrC controlling nitrogen metabolism in *Rhodobacter capsulatus*. *FEMS microbiology letters* **258,** 250–256 (cit. on pp. 92, 122).

Dudoit, S., Y. H. Yang, M. J. Callow and T. P. Speed (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* **12,** 111–139 (cit. on p. 94).

Dunnett, C. W. (1955). A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *Journal of the American Statistical Association* **50,** 1096–1121 (cit. on p. 47).

– (1964). New Tables for Multiple Comparisons with a Control. *Biometrics* **20,** 482 (cit. on p. 47).

Dupla, M., T. Conte, J. C. Bouvier, N. Bernet and J. P. Steyer (2004). Dynamic evaluation of a fixed bed anaerobic digestion process in response to organic overloads and toxicant shock loads. *Water Science and Technology* **49,** 61–68 (cit. on p. 26).

Eady, R. R. (1996). Structure-function relationships of alternative nitrogenases. *Chemical Reviews* **96** (7), 3013–3030 (cit. on p. 33).

Egland, P. G., J. Gibson and C. S. Harwood (2001). Reductive, Coenzyme A-Mediated Pathway for 3-Chlorobenzoate Degradation in the Phototrophic Bacterium *Rhodopseudomonas palustris*. *Applied and Environmental Microbiology* **67,** 1396–1399 (cit. on p. 30).

Elbahloul, Y. and A. Steinbüchel (2006). Engineering the genotype of *Acinetobacter* sp. strain ADP1 to enhance biosynthesis of cyanophycin. *Applied and environmental microbiology* **72,** 1410–1419 (cit. on p. 164).

Elder, D. J. E., P. Morgan and D. J. Kelly (1993). Transposon Tn5 mutagenesis in *Rhodopseudomonas palustris*. *FEMS Microbiology Letters* **111,** 23–30 (cit. on p. 31).

Evans, M. B., A. M. Hawthornthwaite and R. J. Cogdell (1990). Isolation and characterisation of the different B800--850 light-harvesting complexes from low- and high-light grown cells of *Rhodopseudomonas palustris*, strain 2.1.6. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1016,** 71–76 (cit. on p. 56).

Fangueiro, D., M. Gusmão, J. Grilo, G. Porfírio, E. Vasconcelos and F. Cabral (2010). Proportion, composition and potential N mineralisation of particle size fractions obtained by mechanical separation of animal slurry. *Biosystems Engineering* **106,** 333–337 (cit. on p. 126).

Fischer, A. (1897). Untersuchungen über den Bau der Cyanophyceen und Bakterien. G. Fischer (cit. on p. 126).

– (1905). Die Zelle der Cyanophyceen. Druck von Breitkopf & Härtel (cit. on p. 126).

Fißler, J., C. Schirra, G. W. Kohring and F. Giffhorn (1994). Hydrogen production from aromatic acids by *Rhodopseudomonas palustris*. *Applied Microbiology and Biotechnology* **41,** 395–399 (cit. on pp. 27, 30, 34, 49, 91).

Fowler, V. J., N. Pfennig, W. Schubert and E. Stackebrandt (1984). Towards a phylogeny of phototrophic purple sulfur bacteria – 16S rRNA oligonucleotide cataloguing of 11 species of *Chromatiaceae*. *Archives of Microbiology* **139,** 382–387 (cit. on p. 27).

Frazee, A. C., G. M. Pertea, A. E. Jaffe, B. Langmead, S. L. Salzberg and J. T. Leek (2015). Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature Biotechnology* **33,** 243–246 (cit. on p. 94).

Frey, K. M., F. B. Oppermann-Sanio, H. Schmidt and A. Steinbüchel (2002). Technical-Scale Production of Cyanophycin with Recombinant Strains of *Escherichia coli*. *Applied and Environmental Microbiology* **68,** 3377–3384 (cit. on pp. 158, 164).

Fritsch, F. E. (1959). The Structure and Reproduction of the Algae. **2.** Cambridge University Press, Cambridge, U.K., 787 (cit. on p. 126).

Frommeyer, M. and A. Steinbüchel (2013). Increased lysine content is the main characteristic of the soluble form of the polyamide cyanophycin synthesized by recombinant *Escherichia coli*. *Applied and environmental microbiology* **79,** 4474–4483 (cit. on pp. 158, 159, 164, 165).

Fu, D., A. Libson, L. J. W. Miercke, C. Weitzman, P. Nollert, J. Krucinski and R. M. Stroud (2000). Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290,** 481–486 (cit. on pp. 63, 88).

Fu, D., A. Libson and R. M. Stroud (2002). The structure of GlpF, a glycerol conducting channel. *Novartis Foundation symposium* **245,** 51–61, discussion 61–5, 165–8 (cit. on pp. 63, 64).

Fu, H., R. H. Burris and G. P. Roberts (1990). Reversible ADP-ribosylation is demonstrated to be a regulatory mechanism in prokaryotes by heterologous expression. *Proceedings of the National Academy of Sciences of the United States of America* **87** (5), 1720–1724 (cit. on p. 34).

Füser, G. and A. Steinbüchel (2005). Investigations on the Solubility Behavior of Cyanophycin. Solubility of Cyanophycin in Solutions of Simple Inorganic Salts. *Biomacromolecules* **6,** 1367–1374 (cit. on pp. 158, 165).

– (2007). Analysis of Genome Sequences for Genes of Cyanophycin Metabolism: Identifying Putative Cyanophycin Metabolizing Prokaryotes. *Macromolecular Bioscience* **7,** 278–296 (cit. on p. 127).

Galloway, J. N., J. D. Aber, J. W. Erisman, S. P. Seitzinger, R. W. Howarth, E. B. Cowling and B. J. Cosby (2003). The Nitrogen Cascade. *BioScience* **53,** 341–356 (cit. on pp. 125, 126).

Galloway, J. N., F. J. Dentener, D. G. Capone, E. W. Boyer, R. W. Howarth, S. P. Seitzinger, G. P. Asner, C. C. Cleveland, P. A. Green, E. A. Holland, D. M. Karl, A. F. Michaels, J. H.

Porter, A. R. Townsend and C. J. Vöosmarty (2004). Nitrogen Cycles: Past, Present, and Future. *Biogeochemistry* **70,** 153–226 (cit. on pp. 125, 126).

Gest, H. and M. D. Kamen (1949). Photoproduction of Molecular Hydrogen by *Rhodospirillum rubrum*. *Science* **109,** 558–559 (cit. on pp. 26, 91).

Ghosh, D., A. Tourigny and P. C. Hallenbeck (2012). Near stoichiometric reforming of biodiesel derived crude glycerol to hydrogen by photofermentation. *International Journal of Hydrogen Energy* **37,** 2273–2277 (cit. on p. 49).

Ghosh, R., A. Hardmeyer, I. Thoenen and R. Bachofen (1994). Optimization of the Sistrom Culture Medium for Large-Scale Batch Cultivation of *Rhodospirillum rubrum* under Semiaerobic Conditions with Maximal Yield of Photosynthetic Membranes. *Applied and environmental microbiology* **60,** 1698–1700 (cit. on p. 60).

Gibson, D. G. (2011). Enzymatic Assembly of Overlapping DNA Fragments. *Methods in Enzymology,* 349–361 (cit. on pp. 67, 128).

Gibson, D. G., L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison and H. O. Smith (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343–345 (cit. on pp. 67, 128).

Gibson, J., E. Stackebrandt, L. B. Zablen, R. Gupta and C. R. Woese (1979). A phylogenetic analysis of the purple photosynthetic bacteria. *Current Microbiology* **3,** 59–64 (cit. on p. 27).

Giles, J. (2005). Nitrogen study fertilizes fears of pollution. (Cit. on pp. 125, 126).

Gordon, G. C. and J. B. McKinlay (2014). Calvin cycle mutants of photoheterotrophic purple nonsulfur bacteria fail to grow due to an electron imbalance rather than toxic metabolite accumulation. *Journal of bacteriology* **196,** 1231–1237 (cit. on pp. 31, 60).

Gosse, J. L., B. J. Engel, F. E. Rey, C. S. Harwood, L. E. Scriven and M. C. Flickinger (2007). Hydrogen production by photoreactive nanoporous latex coatings of nongrowing *Rhodopseudomonas palustris* CGA009. *Biotechnology Progress.* **23,** 124–130 (cit. on p. 27).

Gribaldo, S. and C. Brochier (2009). Phylogeny of prokaryotes: does it exist and why should we care? *Research in microbiology* **160,** 513–521 (cit. on p. 27).

Gupta, R. S. (2000). The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiology Reviews* **24,** 367–402 (cit. on p. 27).

– (2012). Origin and spread of photosynthesis based upon conserved sequence features in key bacteriochlorophyll biosynthesis proteins. *Molecular biology and evolution* **29,** 3397–3412 (cit. on p. 27).

Haber, F. and R. L. Rossignol (1913). Über die technische Darstellung von Ammoniak aus den Elementen. *Zeitschrift für Elektrochemie und angewandte physikalische Chemie* **19,** 53–72 (cit. on p. 126).

Hadley, W. (2016). ggplot2 : elegrant graphics for data analysis. Springer-Verlag New York (cit. on p. 47).

Hafner, S. D., F. Montes and C. Alan Rotz (2013). The role of carbon dioxide in emission of ammonia from manure. *Atmospheric Environment* **66,** 63–71 (cit. on p. 126).

Hai, T., K. M. Frey and A. Steinbüchel (2006). Engineered cyanophycin synthetase (CphA) from *Nostoc ellipsosporum* confers enhanced CphA activity and cyanophycin accumulation to *Escherichia coli*. *Applied and environmental microbiology* **72,** 7652–7660 (cit. on pp. 158, 164).

Hallenbeck, P. C. (2011). Microbial paths to renewable hydrogen production. *Biofuels* **2,** 285–302 (cit. on pp. 32, 34, 92).

Hallenbeck, P. C., D. Ghosh, M. T. Skonieczny and V. Yargeau (2009). Microbiological and engineering aspects of biohydrogen production. *Indian Journal of Microbiology* **49,** 48–59 (cit. on pp. 32, 33).

Hamelin, L., M. Wesnæs, H. Wenzel and B. M. Petersen (2011). Environmental Consequences of Future Biogas Technologies Based on Separated Slurry. *Environmental Science & Technology* **45,** 5869–5877 (cit. on p. 126).

Hanley, N. (1990). The economics of nitrate pollution. *European Review of Agricultural Economics* **17,** 129–151 (cit. on pp. 125, 126).

Harada, N., S. Otsuka, M. Nishiyama and S. Matsumoto (2003). Characteristics of phototrophic purple bacteria isolated from a Japanese paddy soil. *Soil Science and Plant Nutrition* **49,** 521–526 (cit. on pp. 27, 49).

Harwood, C. S. and J. Gibson (1988). Anaerobic and aerobic metabolism of diverse aromatic compounds by the photosynthetic bacterium *Rhodopseudomonas palustris*. *Applied and Environmental Microbiology* **54,** 712–717 (cit. on pp. 30, 49).

Hegeman, G. D. (1967). Metabolism of P-Hydroxybenzoate by *Rhodopseudomonas palustris* and Its Regulation. *Archiv für Mikrobiologie* **59,** 143–148 (cit. on pp. 30, 49, 59).

Heiniger, E. K., Y. Oda, S. K. Samanta and C. S. Harwood (2012). How posttranslational modification of nitrogenase is circumvented in *Rhodopseudomonas palustris* strains that produce hydrogen gas constitutively. *Applied and environmental microbiology* **78,** 1023–1032 (cit. on pp. 31, 34, 92, 122).

Hill, J., E. Nelson, D. Tilman, S. Polasky and D. Tiffany (2006). Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 11206–11210 (cit. on pp. 25, 63).

Hiraishi, A. and H. Kitamura (1984). Distribution of phototrophic purple nonsulfur bacteria in activated sludge systems and other aquatic environments. *Nippon Suisan Gakkaishi* **50,** 1929–1937 (cit. on p. 49).

Hiraishi, A., T. S. Santos, J. Sugiyama and K. Komagata (1992). *Rhodopseudomonas rutila* Is a Later Subjective Synonym of *Rhodopseudomonas palustris*. *International Journal of Systematic Bacteriology* **42,** 186–188 (cit. on p. 27).

Ho Lee, T., J. Y. Park and S. Park (2002). Growth of *Rhodopseudomonas palustris* under phototrophic and non-phototrophic conditions and its CO-dependent H_2 production. *Biotechnology Letters* **24,** 91–96 (cit. on p. 27).

Holly, M. A., R. A. Larson, J. M. Powell, M. D. Ruark and H. Aguirre-Villegas (2017). Greenhouse gas and ammonia emissions from digested and separated dairy manure during storage and after land application. *Agriculture, Ecosystems & Environment* **239,** 410–419 (cit. on pp. 125, 126).

Hou, Y., Z. Bai, J. P. Lesschen, I. G. Staritsky, N. Sikirica, L. Ma, G. L. Velthof and O. Oenema (2016). Feed use and nitrogen excretion of livestock in EU-27. *Agriculture, Ecosystems & Environment* **218,** 232–244 (cit. on p. 126).

Hristov, A. N., M. Hanigan, A. Cole, R. Todd, T. A. McAllister, P. M. Ndegwa and A. Rotz (2011). Review: Ammonia emissions from dairy farms and beef feedlots. *Canadian Journal of Animal Science* **91,** 1–35 (cit. on pp. 125, 126).

Huang, J. J., E. K. Heiniger, J. B. McKinlay and C. S. Harwood (2010). Production of hydrogen gas from light and the inorganic electron donor thiosulfate by *Rhodopseudomonas palustris*. *Applied and Environmental Microbiology* **76,** 7717–7722 (cit. on pp. 27, 60).

Imhoff, J. F. and P. Caumette (2004). Recommended standards for the description of new species of anoxygenic phototrophic bacteria. *International journal of systematic and evolutionary microbiology* **54,** 1415–1421 (cit. on p. 27).

Imhoff, J. F., H. G. Trüper and N. Pfennig (1984). Rearrangement of the Species and Genera of the Phototrophic "Purple Nonsulfur Bacteria". *International Journal of Systematic Bacteriology* **34,** 340–343 (cit. on p. 27).

Inoue, H., H. Nojima and H. Okayama (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96,** 23–28 (cit. on p. 41).

Inui, M., J. H. Roh, K. Zahn and H. Yukawa (2000). Sequence analysis of the cryptic plasmid pMG101 from *Rhodopseudomonas palustris* and construction of stable cloning vectors. *Applied and Environmental Microbiology* **66,** 54–63 (cit. on p. 31).

Jackson, S. A., J. R. D. Hervey, A. J. Dale and J. J. Eaton-Rye (2014). Removal of both Ycf48 and Psb27 in *Synechocystis* sp. PCC 6803 disrupts Photosystem II assembly and alters QA− oxidation in the mature complex. *FEBS Letters* **588,** 3751–3760 (cit. on p. 40).

Jensen, M. Ø., E. Tajkhorshid and K. Schulten (2003). Electrostatic tuning of permeation and selectivity in aquaporin water channels. *Biophysical journal* **85,** 2884–2899 (cit. on pp. 64, 88).

Jiao, Y., A. Kappler, L. R. Croal and D. K. Newman (2005). Isolation and characterization of a genetically tractable photoautotrophic Fe(II)-oxidizing bacterium, *Rhodopseudomonas palustris* strain TIE-1. *Applied and environmental microbiology* **71,** 4487–4496 (cit. on pp. 29, 60).

Joentgen, W., T. Groth, A. Steinbüchel, T. Hai and F. B. Oppermann (1998). Polyasparaginic acid homopolymers an copolymers, biotechnical production and use thereof. (Cit. on p. 127).

Johnson, D. T. and K. A. Taconi (2007). The glycerin glut: Options for the value-added conversion of crude glycerol resulting from biodiesel production. *Environmental Progress* **26,** 338–348 (cit. on pp. 30, 32, 63).

Jones, B. L. and K. J. Monty (1979). Glutamine as a feedback inhibitor of the *Rhodopseudomonas sphaeroides* nitrogenase system. *Journal of Bacteriology* **139,** 1007–1013 (cit. on pp. 34, 92).

Kawasaki, H., Y. Hoshino and K. Yamasato (1993). Phylogenetic diversity of phototrophic purple non-sulfur bacteria in the Proteobacteria α group. *FEMS Microbiology Letters* **112,** 61–66 (cit. on p. 27).

Kayhanian, M. (1999). Ammonia Inhibition in High-Solids Biogasification: An Overview and Practical Solutions. *Environmental Technology* **20,** 355–365 (cit. on pp. 26, 60).

Kim, D., B. Langmead and S. L. Salzberg (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12,** 357–360 (cit. on p. 94).

Kim, M.-K. and C. S. Harwood (1991). Regulation of benzoate-CoA ligase in *Rhodopseudomonas palustris*. *FEMS microbiology letters* **83** (2), 199–203 (cit. on pp. 30, 59).

Kim, M. K., K.-M. Choi, C.-R. Yin, K.-Y. Lee, W.-T. Im, J. H. Lim and S.-T. Lee (2004). Odorous swine wastewater treatment by purple non-sulfur bacteria, *Rhodopseudomonas palustris*, isolated from eutrophicated ponds. *Biotechnology Letters* **26,** 819–822 (cit. on pp. 26, 49).

Klein, N., J. Neumann, J. D. O'Neil and D. Schneider (2015). Folding and stability of the aquaglyceroporin GlpF: Implications for human aqua(glycero)porin diseases. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1848,** 622–633 (cit. on p. 88).

Knobloch, K., J. H. Eley and M. I. H. Aleem (1971). Generation of reducing power in bacterial photosynthesis. *Rhodopseudomonas palustris*. *Biochemical and Biophysical Research Communications* **43,** 834–839 (cit. on pp. 49, 60).

Koku, H., I. Eroğlu, U. Gündüz, M. Yücel and L. Türker (2002). Aspects of the metabolism of hydrogen production by *Rhodobacter sphaeroides*. *International Journal of Hydrogen Energy* **27,** 1315–1329 (cit. on pp. 34, 91).

Kroll, J., S. Klinter and A. Steinbüchel (2011). A novel plasmid addiction system for large-scale production of cyanophycin in *Escherichia coli* using mineral salts medium. *Applied Microbiology and Biotechnology* **89,** 593–604 (cit. on p. 164).

Kroll, J., A. Steinle, R. Reichelt, C. Ewering and A. Steinbüchel (2009). Establishment of a novel anabolism-based addiction system with an artificially introduced mevalonate pathway: complete stabilization of plasmids as universal application in white biotechnology. *Metabolic engineering* **11,** 168–177 (cit. on p. 164).

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29,** 1–27 (cit. on p. 94).

Kunisawa, T. (2001). Gene arrangements and phylogeny in the class Proteobacteria. *Journal of theoretical biology* **213,** 9–19 (cit. on p. 27).

Küver, J., Y. Xu and J. Gibson (1995). Metabolism of cyclohexane carboxylic acid by the photosynthetic bacterium *Rhodopseudomonas palustris*. *Archives of Microbiology* **164,** 337–345 (cit. on pp. 30, 49).

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227,** 680–685 (cit. on p. 42).

Lang, N. J. (1968). The Fine Structure of Blue-Green Algae. *Annual Review of Microbiology* **22,** 15–46 (cit. on p. 126).

Larimer, F. W., P. Chain, L. Hauser, J. Lamerdin, S. Malfatti, L. Do, M. L. Land, D. a. Pelletier, J. T. Beatty, A. S. Lang, F. R. Tabita, J. L. Gibson, T. E. Hanson, C. Bobst, J. L. T. Y. Torres, C. M. Peres, F. H. Harrison, J. Gibson and C. S. Harwood (2004). Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. *Nature Biotechnology* **22,** 55–61 (cit. on pp. 27, 29, 33, 39, 121, 128).

Lazaro, C. Z., Z. Y. Hitit and P. C. Hallenbeck (2017). Optimization of the yield of dark microaerobic production of hydrogen from lactate by *Rhodopseudomonas palustris*. *Bioresource Technology* **245,** 123–131 (cit. on p. 27).

Li, B., N. Liu, Y. Li, W. Jing, J. Fan, D. Li, L. Zhang, X. Zhang, Z. Zhang and L. Wang (2014). Reduction of Selenite to Red Elemental Selenium by *Rhodopseudomonas palustris* Strain N. *PLoS ONE* **9.** Ed. by V. Bansal, e95955 (cit. on pp. 29, 60).

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis and R. Durbin (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (cit. on p. 98).

Liang, B., T.-D. Wu, H.-J. Sun, H. Vali, J.-L. Guerquin-Kern, C.-H. Wang and T. Bosak (2014). Cyanophycin mediates the accumulation and storage of fixed carbon in non-heterocystous filamentous cyanobacteria from coniform mats. *PloS one* **9,** e88142 (cit. on p. 127).

Liang, J., G. M. Nielsen, D. P. Lies, R. H. Burris, G. P. Roberts and P. W. Ludden (1991). Mutations in the *draT* and *draG* genes of *Rhodospirillum rubrum* result in loss of regulation of nitrogenase by reversible ADP-ribosylation. *Journal of Bacteriology* **173** (21), 6903–6909 (cit. on p. 34).

Llorens, I., G. Untereiner, D. Jaillard, B. Gouget, V. Chapon and M. Carriere (2012). Uranium Interaction with Two Multi-Resistant Environmental Bacteria: *Cupriavidus metallidurans* CH34 and *Rhodopseudomonas palustris*. *PLoS ONE* **7.** Ed. by V. Bansal, e51783 (cit. on pp. 29, 60).

Ludden, P. W. and G. P. Roberts (2002). Nitrogen fixation by photosynthetic bacteria. *Photosynthesis Research* **73** (1-3), 115–118 (cit. on p. 91).

Ludwig, W., R. Rossellö-Mora, R. Aznar, S. Klugbauer, S. Spring, K. Reetz, C. Beimfohr, E. Brockmann, G. Kirchhof, S. Dorn, M. Bachleitner, N. Klugbauer, N. Springer, D. Lane, R. Nietupsky, M. Weizenegger and K.-h. Schleifer (1995). Comparative Sequence Analysis of 23S rRNA from Proteobacteria. *Systematic and Applied Microbiology* **18,** 164–188 (cit. on p. 27).

Madigan, M. T. and H. Gest (1988). Selective enrichment and isolation of *Rhodopseudomonas palustris* using trans-cinnamic acid as sole carbon source. *FEMS Microbiology Letters* **53,** 53–58 (cit. on p. 30).

Madigan, M. T., J. M. Martinko, P. V. Dunlap and D. P. Clark (2008). Prokaryotic diversity: bacteria the phylogeny of bacteria. *Brock Biology of Microorganisms (12th Edition).* **2,** 1168 (cit. on p. 49).

Masepohl, B. (2017). Regulation of nitrogen fixation in photosynthetic purple nonsulfur bacteria. *Modern Topics in the Phototrophic Prokaryotes: Metabolism, Bioenergetics, and Omics.* Springer-Verlag, 1–25 (cit. on p. 33).

Masepohl, B., T. Drepper, A. Paschen, S. Groß, A. Pawlowski, K. Raabe, K. U. Riedel and W. Klipp (2002). Regulation of nitrogen fixation in the phototrophic purple bacterium *Rhodobacter capsulatus*. *Journal of Molecular Microbiology and Biotechnology.* **4,** 243–248 (cit. on p. 34).

Masepohl, B. and P. C. Hallenbeck (2010). Nitrogen and molybdenum control of nitrogen fixation in the phototrophic bacterium *Rhodobacter capsulatus*. (Cit. on p. 33).

Masepohl, B., B. Kaiser, N. Isakovic, C. L. Richard, R. G. Kranz and W. Klipp (2001). Urea utilization in the phototrophic bacterium *Rhodobacter capsulatus* is regulated by the transcriptional activator NtrC. *Journal of bacteriology* **183,** 637–643 (cit. on pp. 92, 93, 122, 162, 163).

McInerney, J. O., J. A. Cotton and D. Pisani (2008). The prokaryotic tree of life: past, present... and future? *Trends in ecology & evolution* **23,** 276–281 (cit. on p. 27).

McKinlay, J. B., Y. Oda, M. Rühl, A. L. Posto, U. Sauer and C. S. Harwood (2014). Non-growing *Rhodopseudomonas palustris* increases the hydrogen gas yield from acetate by shifting from the glyoxylate shunt to the tricarboxylic acid cycle. *The Journal of biological chemistry* **289,** 1960–1970 (cit. on pp. 27, 30, 64, 155).

Merugu, R., S. Girisham and S. M. Reddy (2010). Production of PHB (Polyhydroxybutyrate) by *Rhodopseudomonas palustris* KU003 under Nitrogen Limitation. *International Journal of Applied Biology and Pharmaceutical Technology* **1,** 847–850 (cit. on p. 27).

Michel-Reydellet, N. and P. A. Kaminski (1999). *Azorhizobium caulinodans* P(II) and GlnK proteins control nitrogen fixation and ammonia assimilation. *Journal of Bacteriology* **181,** 2655–2658 (cit. on p. 34).

Miller, R. W. and R. R. Eady (1988). Molybdenum and vanadium nitrogenases of *Azotobacter chroococcum*. Low temperature favours N_2 reduction by vanadium nitrogenase. *The Biochemical journal* **256** (2), 429–434 (cit. on pp. 33, 92).

Mobley, H. L. T., M. D. Island and R. P. Hausinger (1995). Molecular biology of microbial ureases. *Microbiological reviews* **59,** 451–480 (cit. on p. 92).

Molisch, H. (1907). Die purpurbakterien : nach neuen untersuchungen; eine mikrobiologische studie. Jena. Verlag von Gustav Fischer., 1–95 (cit. on pp. 26, 49).

Mooibroek, H., N. Oosterhuis, M. Giuseppin, M. Toonen, H. Franssen, E. Scott, J. Sanders and A. Steinbüchel (2007). Assessment of technological options and economical feasibility for cyanophycin biopolymer and high-value amino acid production. *Applied Microbiology and Biotechnology* **77,** 257–267 (cit. on pp. 127, 164).

Nagashima, K. V. P., A. Hiraishi, K. Shimada and K. Matsuura (1997). Horizontal transfer of genes coding for the photosynthetic reaction centers of purple bacteria. *Journal of Molecular Evolution* **45,** 131–136 (cit. on p. 27).

Ndegwa, P. M., A. N. Hristov, J. Arogo and R. E. Sheffield (2008). A review of ammonia emission mitigation techniques for concentrated animal feeding operations. *Biosystems Engineering* **100,** 453–469 (cit. on p. 125).

Neerackal, G. M., P. M. Ndegwa, H. S. Joo, X. Wang, J. H. Harrison, A. J. Heber, J. Q. Ni and C. Frear (2015). Effects of Anaerobic Digestion and Solids Separation on Ammonia Emissions from Stored and Land Applied Dairy Manure. *Water, Air, & Soil Pollution* **226,** 301 (cit. on p. 126).

Neumann, K., D. P. Stephan, K. Ziegler, M. Hühns, I. Broer, W. Lockau and E. K. Pistorius (2005). Production of cyanophycin, a suitable source for the biodegradable polymer polyaspartate, in transgenic plants. *Plant Biotechnology Journal* **3,** 249–258 (cit. on p. 127).

Niel, C. B. van (1944). The culture, general physiology, morphology, and classification of the non-sulfur purple and brown bacteria. *Bacteriological reviews* **8,** 1–118 (cit. on pp. 26, 49, 59).

Obst, M., F. B. Oppermann-Sanio, H. Luftmann and A. Steinbüchel (2002). Isolation of cyanophycin-degrading bacteria, cloning and characterization of an extracellular cyanophycinase gene (cphE) from *Pseudomonas anguilliseptica* strain BI. The cphE gene from *P. anguilliseptica* BI encodes a cyanophycin hydrolyzing enzyme. *The Journal of biological chemistry* **277,** 25096–25105 (cit. on p. 127).

Obst, M., A. Sallam, H. Luftmann and A. Steinbüchel (2004). Isolation and Characterization of Gram-Positive Cyanophycin-Degrading Bacteria. Kinetic Studies on Cyanophycin Depolymerase Activity in Aerobic Bacteria. *Biomacromolecules* **5,** 153–161 (cit. on p. 127).

Oda, Y., Y. P. De Vries, L. J. Forney and J. C. Gottschal (2001). Acquisition of the ability for *Rhodopseudomonas palustris* to degrade chlorinated benzoic acids as the sole carbon source. *FEMS Microbiology Ecology* **38,** 133–139 (cit. on pp. 30, 49).

Oda, Y., S. K. Samanta, F. E. Rey, L. Wu, X. Liu, T. Yan, J. Zhou and C. S. Harwood (2005). Functional genomic analysis of three nitrogenase isozymes in the photosynthetic bac-

terium *Rhodopseudomonas palustris*. *Journal of bacteriology* **187,** 7784–7794 (cit. on pp. 31, 33, 34, 91, 92).

Oda, Y., W. Wanders, L. A. Huisman, W. G. Meijer, J. C. Gottschal and L. J. Forney (2002). Genotypic and phenotypic diversity within species of purple nonsulfur bacteria isolated from aquatic sediments. *Applied and environmental microbiology* **68,** 3467–3477 (cit. on pp. 26, 49).

Oenema, O., D. Oudendag and G. L. Velthof (2007). Nutrient losses from manure management in the European Union. *Livestock Science* **112,** 261–272 (cit. on pp. 125, 126).

Oh, Y. K., E. H. Seol, E. Y. Lee and S. Park (2002). Fermentative hydrogen production by a new chemoheterotrophic bacterium *Rhodopseudomonas palustris* P4. *International Journal of Hydrogen Energy* **27,** 1373–1379 (cit. on p. 27).

Oliva, R., G. Calamita, J. M. Thornton and M. Pellegrini-Calace (2010). Electrostatics of aquaporin and aquaglyceroporin channels correlates with their transport selectivity. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 4135–4140 (cit. on p. 64).

Orelle, C., S. Carlson, B. Kaushal, M. M. Almutairi, H. Liu, A. Ochabowicz, S. Quan, V. C. Pham, C. L. Squires, B. T. Murphy and A. S. Mankin (2013). Tools for characterizing bacterial protein synthesis inhibitors. *Antimicrobial agents and chemotherapy* **57,** 5994–6004 (cit. on p. 158).

Ornston, L. N. and R. Y. Stanier (1966). The Conversion of Catechol and Protocatechuate to β-Ketoadipate by *Pseudomonas putida*. *Journal of Biological Chemistry* **241,** 3776–3786 (cit. on p. 59).

Oyaizu, H. and C. R. Woese (1985). Phylogenetic Relationships Among the Sulfate Respiring Bacteria, Myxobacteria and Purple Bacteria. *Systematic and Applied Microbiology* **6,** 257–263 (cit. on p. 27).

Pain, B. F., K. A. Smith and C. J. Dyer (1986). Factors affecting the response of cut grass to the nitrogen content of dairy cow slurry. *Agricultural Wastes* **17,** 189–202 (cit. on pp. 125, 126).

Pan, C., Y. Oda, P. K. Lankford, B. Zhang, N. F. Samatova, D. A. Pelletier, C. S. Harwood and R. L. Hettich (2008). Characterization of anaerobic catabolism of p-coumarate in *Rhodopseudomonas palustris* by integrating transcriptomics and quantitative proteomics. *Molecular & cellular proteomics : MCP* **7,** 938–948 (cit. on pp. 30, 60).

Pan, J.-J., J. O. Solbiati, G. Ramamoorthy, B. S. Hillerich, R. D. Seidel, J. E. Cronan, S. C. Almo and C. D. Poulter (2015). Biosynthesis of Squalene from Farnesyl Diphosphate in Bacteria: Three Steps Catalyzed by Three Enzymes. *ACS Central Science* **1,** 77–82 (cit. on p. 31).

Pant, D., G. Van Bogaert, L. Diels and K. Vanbroekhoven (2010). A review of the substrates used in microbial fuel cells (MFCs) for sustainable energy production. *Bioresource technology* **101,** 1533–1543 (cit. on p. 29).

Pearson, K. (1895). Note on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London* **58** (347-352), 240–242 (cit. on p. 94).

– (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6* (cit. on p. 94).

Pechter, K. B., L. Gallagher, H. Pyles, C. S. Manoil and C. S. Harwood (2016). Essential genome of the metabolically versatile alphaproteobacterium *Rhodopseudomonas palustris*. *Journal of bacteriology* **198** (5), 867–876 (cit. on pp. 31, 128).

Perälä, P. and K. Regina (2006). The effect of cow slurry fermentation and application technique on greenhouse gas and ammonia emissions from a grass field. *12th Ramiran International Conference, "Technology for Recycling of Manure and Organic Residues in a Whole Farm Perspective"*. **2,** 245–247 (cit. on p. 126).

Perrotta, J. A. and C. S. Harwood (1994). Anaerobic Metabolism of Cyclohex-1-Ene-1-Carboxylate, a Proposed Intermediate of Benzoate Degradation, by *Rhodopseudomonas palustris*. *Applied and environmental microbiology* **60,** 1775–1782 (cit. on pp. 30, 49).

Pertea, M., D. Kim, G. M. Pertea, J. T. Leek and S. L. Salzberg (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* **11,** 1650–1667 (cit. on p. 94).

Pertea, M., G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell and S. L. Salzberg (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33,** 290–295 (cit. on p. 94).

Petersen, S. O. and S. G. Sommer (2011). Ammonia and nitrous oxide interactions: Roles of manure organic matter management. *Animal Feed Science and Technology* **166,** 503–513 (cit. on p. 126).

Pfennig, N. and H. G. Trüper (1983). Taxonomy of phototrophic green and purple bacteria: A review. *Annales de l'Institut Pasteur / Microbiologie* **134,** 9–20 (cit. on p. 27).

Pott, R. W. M., C. J. Howe and J. S. Dennis (2013). Photofermentation of crude glycerol from biodiesel using *Rhodopseudomonas palustris*: Comparison with organic acids and the identification of inhibitory compounds. *Bioresource Technology* **130,** 725–730 (cit. on pp. 27, 30, 35, 49, 60, 63, 91).

– (2014). The purification of crude glycerol derived from biodiesel manufacture and its use as a substrate by *Rhodopseudomonas palustris* to produce hydrogen. *Bioresource Technology* **152,** 464–470 (cit. on pp. 27, 30, 49, 60, 63).

Puskás, L. G., M. Inui, K. Zahn and H. Yukawa (2000). A periplasmic, α-type carbonic anhydrase from *Rhodopseudomonas palustris* is essential for bicarbonate uptake. *Microbiology* **146,** 2957–2966 (cit. on p. 27).

Rath, A., M. Glibowicka, V. G. Nadeau, G. Chen and C. M. Deber (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 1760–1765 (cit. on p. 88).

Recht, M. I., S. Douthwaite and J. D. Puglisi (1999). Basis for prokaryotic specificity of action of aminoglycoside antibiotics. *The EMBO journal* **18,** 3133–3138 (cit. on p. 158).

Rey, F. E., E. K. Heiniger and C. S. Harwood (2007). Redirection of Metabolism for Biological Hydrogen Production. *Applied and Environmental Microbiology* **73,** 1665–1671 (cit. on p. 91).

Rey, F. E., Y. Oda and C. S. Harwood (2006). Regulation of uptake hydrogenase and effects of hydrogen utilization on gene expression in *Rhodopseudomonas palustris*. *Journal of bacteriology* **188,** 6143–6152 (cit. on pp. 31, 34, 60, 91).

Riggio, V., E. Comino and M. Rosso (2015). Energy production from anaerobic co-digestion processing of cow slurry, olive pomace and apple pulp. *Renewable Energy* **83,** 1043–1049 (cit. on p. 126).

Rolls, J. P. and E. S. Lindstrom (1967). Effect of Thiosulfate on the Photosynthetic Growth of *Rhodopseudomonas palustris*. *Journal of Bacteriology* **94,** 860–866 (cit. on pp. 49, 60).

Roper, M. M. and J. K. Ladha (1995). Biological N_2 fixation by heterotrophic and phototrophic bacteria in association with straw. *Plant and Soil* **174,** 211–224 (cit. on p. 27).

Rubio, L. M. and P. W. Ludden (2005). Maturation of nitrogenase: A biochemical puzzle. *Journal of Bacteriology* **187** (2), 405–414 (cit. on p. 121).

Sabourin-Provost, G. and P. C. Hallenbeck (2009). High yield conversion of a crude glycerol fraction from biodiesel production to hydrogen by photofermentation. *Bioresource technology* **100,** 3513–3517 (cit. on pp. 27, 30, 49).

Sagné, C., M. F. Isambert, J. P. Henry and B. Gasnier (1996). SDS-resistant aggregation of membrane proteins: application to the purification of the vesicular monoamine transporter. *Biochemical Journal* **316,** 825–831 (cit. on p. 88).

Sallam, A. and A. Steinbüchel (2010). Dipeptides in nutrition and therapy: cyanophycin-derived dipeptides as natural alternatives and their biotechnological production. *Applied Microbiology and Biotechnology* **87,** 815–828 (cit. on p. 127).

Sasikala, K., C. V. Ramana, P. Raghuveer Rao and M. Subrahmanyam (1990). Effect of gas phase on the photoproduction of hydrogen and substrate conversion efficiency in the photosynthetic bacterium *Rhodobacter sphaeroides* O.U. 001. *International Journal of Hydrogen Energy* **15** (11), 795–797 (cit. on pp. 34, 92).

Schäfer, A., A. Tauch, W. Jäger, J. Kalinowski, G. Thierbach and A. Pühler (1994). Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* **145,** 69–73 (cit. on pp. 67, 128).

Scheuring, S., R. P. Gonçalves, V. Prima and J. N. Sturgis (2006). The Photosynthetic Apparatus of *Rhodopseudomonas palustris*: Structures and Organization. *Journal of Molecular Biology* **358,** 83–96 (cit. on p. 26).

Schneider, K., U. Gollan, M. Dröttboom, S. Selsemeier-Voigt and A. Müller (1997). Comparative biochemical characterization of the iron-only nitrogenase and the molybdenum

nitrogenase from *Rhodobacter capsulatus*. *European Journal of Biochemistry* **244** (3), 789–800 (cit. on pp. 33, 92).

Schwamborn, M. (1998). Chemical synthesis of polyaspartates: a biodegradable alternative to currently used polycarboxylate homo- and copolymers. *Polymer Degradation and Stability* **59,** 39–45 (cit. on p. 127).

Scott, S. A., M. P. Davey, J. S. Dennis, I. Horst, C. J. Howe, D. J. Lea-Smith and A. G. Smith (2010). Biodiesel from algae: challenges and prospects. *Current opinion in biotechnology* **21,** 277–286 (cit. on p. 25).

Sharma, N., K. C. Doerner, P. C. Alok and M. Choudhary (2015). Skatole remediation potential of *Rhodopseudomonas palustris* WKU-KDNS3 isolated from an animal waste lagoon. *Letters in Applied Microbiology* **60,** 298–306 (cit. on pp. 26, 29, 49).

Simmons, S. S., R. D. Isokpehi, S. D. Brown, D. L. Mcallister, C. C. Hall, W. M. Mcduffy, T. L. Medley, U. K. Udensi, R. V. Rajnarayanan, W. K. Ayensu and H. H. Cohly (2011). Functional Annotation Analytics of *Rhodopseudomonas palustris* Genomes. *Bioinformatics and Biology Insights* **5,** 115–129 (cit. on pp. 63, 64).

Simon, R. D. (1971). Cyanophycin Granules from the Blue-Green Alga *Anabaena cylindrica*: A Reserve Material Consisting of Copolymers of Aspartic Acid and Arginine. *Proceedings of the National Academy of Sciences of the United States of America* **68,** 265–267 (cit. on pp. 126, 127, 158).

– (1973). The effect of chloramphenicol on the production of cyanophycin granule polypeptide in the blue-green alga *Anabaena cylindrica*. *Archiv für Mikrobiologie* **92,** 115–122 (cit. on pp. 126, 127, 158).

Simon, R. D. and P. Weathers (1976). Determination of the structure of the novel polypeptide containing aspartic acid and arginine which is found in cyanobacteria. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **420,** 165–176 (cit. on pp. 126, 127, 133, 158).

Solaiman, D. K. Y. Y., R. A. Garcia, R. D. Ashby, G. J. Piazza and A. Steinbüchel (2011). Rendered-protein hydrolysates for microbial synthesis of cyanophycin biopolymer. *New Biotechnology* **28,** 552–558 (cit. on p. 127).

Stackebrandt, E., V. J. Fowler, W. Schubert and J. F. Imhoff (1984). Towards a phylogeny of phototrophic purple sulfur bacteria – the genus *Ectothiorhodospira*. *Archives of Microbiology* **137,** 366–370 (cit. on p. 27).

Steinle, A. and A. Steinbüchel (2010). Establishment of a simple and effective isolation method for cyanophycin from recombinant *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology* **85,** 1393–1399 (cit. on pp. 158, 165).

Steinle, A., S. Witthoff, J. P. Krause and A. Steinbüchel (2010). Establishment of cyanophycin biosynthesis in *Pichia pastoris* and optimization by use of engineered cyanophycin synthetases. *Applied and environmental microbiology* **76,** 1062–1070 (cit. on p. 164).

Sun, F., J. H. Harrison, P. M. Ndegwa and K. Johnson (2014). Effect of Manure Treatment on Ammonia and Greenhouse Gases Emissions Following Surface Application. *Water, Air, & Soil Pollution* **225,** 1923 (cit. on p. 126).

Taylor, R. G., D. C. Walker and R. R. Mclnnes (1993). *E.coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Research* **21,** 1677–1678 (cit. on p. 39).

Trüper, H. G. and U. Fischer (1982). Anaerobic Oxidation of Sulphur Compounds as Electron Donors for Bacterial Photosynthesis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **298,** 529–542 (cit. on p. 61).

Tseng, W.-C., T.-Y. Fang, K.-C. Chang and C.-L. Pan (2013). Expression of *Synechocystis* sp. PCC6803 cyanophycin synthetase in *Lactococcus lactis* nisin-controlled gene expression system (NICE) and cyanophycin production. *Biochemical Engineering Journal* **78,** 114–119 (cit. on pp. 158, 164).

Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics* **5,** 99 (cit. on p. 47).

Uemura, T., K. Suzuki, K. Nagano and S. Morita (1961). Comparative studies on growth, respiration, photosynthesis and pigment content in *Rhodopseudomonas palustris*. *Plant & Cell Physiology* **2,** 451–461 (cit. on p. 49).

UK Statistics on Waste, 2010 to 2012. (2015). Tech. rep., 1–18 (cit. on p. 25).

Veerappan, A., F. Cymer, N. Klein and D. Schneider (2011). The tetrameric α-helical membrane protein GlpF unfolds via a dimeric folding intermediate. *Biochemistry* **50** (47), 10223–10230 (cit. on p. 88).

Velthof, G. L., J. P. Lesschen, J. Webb, S. Pietrzak, Z. Miatkowski, M. Pinto, J. Kros and O. Oenema (2014). The impact of the Nitrates Directive on nitrogen emissions from agriculture in the EU-27 during 2000–2008. *Science of The Total Environment* **468-469,** 1225–1233 (cit. on p. 125).

Venkidusamy, K. and M. Megharaj (2016). A Novel Electrophototrophic Bacterium *Rhodopseudomonas palustris* Strain RP2, Exhibits Hydrocarbonoclastic Potential in Anaerobic Environments. *Frontiers in microbiology* **7,** 1071 (cit. on pp. 26, 29, 49).

VerBerkmoes, N. C., M. B. Shah, P. K. Lankford, D. A. Pelletier, M. B. Strader, D. L. Tabb, W. H. McDonald, J. W. Barton, G. B. Hurst, L. Hauser, B. H. Davison, J. T. Beatty, C. S. Harwood, F. R. Tabita, R. L. Hettich and F. W. Larimer (2006). Determination and Comparison of the Baseline Proteomes of the Versatile Microbe *Rhodopseudomonas palustris* under Its Major Metabolic States. *Journal of Proteome Research* **5,** 287–298 (cit. on p. 60).

Vignais, P. M. and A. Colbeau (2004). Molecular biology of microbial hydrogenases. *Current Issues in Molecular Biology* **6** (2), 159–188 (cit. on pp. 32, 34, 91).

Vignais, P. M., A. Colbeau, J. C. Willison and Y. Jouanneau (1985). Hydrogenase, nitrogenase, and hydrogen metabolism in the photosynthetic bacteria. *Advances in microbial physiology* **26,** 155–234 (cit. on pp. 32, 34, 91).

Vignais, P. M. and B. Toussaint (1994). Molecular biology of membrane-bound H_2 uptake hydrogenases. *Archives of microbiology* **161,** 1–10 (cit. on pp. 32, 34, 91).

Vikineswary, S., K. Getha, S. Maheswary, V. C. Chong, I. Shaliza and C. A. Sastry (1997). Growth of *Rhodopseudomonas palustris* strain B1 in sago starch processing wastewater. *Studies in Environmental Science.* **66,** 335–348 (cit. on p. 27).

Voss, I., S. C. Diniz, E. Aboulmagd and A. Steinbüchel (2004). Identification of the *Anabaena* sp. Strain PCC7120 Cyanophycin Synthetase as Suitable Enzyme for Production of Cyanophycin in Gram-Negative Bacteria Like *Pseudomonas putida* and *Ralstonia eutropha*. *Biomacromolecules* **5,** 1588–1595 (cit. on p. 158).

Wang, X., H. V. Modak and F. R. Tabita (1993). Photolithoautotrophic growth and control of CO_2 fixation in *Rhodobacter sphaeroides* and *Rhodospirillum rubrum* in the absence of ribulose bisphosphate carboxylase-oxygenase. *Journal of Bacteriology* **175,** 7109–7114 (cit. on p. 61).

Webb, J., B. F. Pain, S. Bittman and J. Morgan (2010). The impacts of manure application methods on emissions of ammonia, nitrous oxide and on crop response—A review. *Agriculture, Ecosystems & Environment* **137,** 39–46 (cit. on p. 125).

Wiefel, L. and A. Steinbüchel (2014). Solubility behavior of cyanophycin depending on lysine content. *Applied and environmental microbiology* **80,** 1091–1096 (cit. on p. 165).

Wilson, D. N. (2009). The A-Z of bacterial translation inhibitors. *Critical reviews in biochemistry and molecular biology* **44,** 393–433 (cit. on p. 158).

Woese, C. R., E. Stackebrandt, T. J. Macke and G. E. Fox (1985). A Phylogenetic Definition of the Major Eubacterial Taxa. *Systematic and Applied Microbiology* **6,** 143–151 (cit. on p. 27).

Woese, C. R., E. Stackebrandt, W. G. Weisburg, B. J. Paster, M. T. Madigan, V. J. Fowler, C. M. Hahn, P. Blanz, R. Gupta, K. H. Nealson and G. E. Fox (1984). The phylogeny of purple bacteria: The alpha subdivision. *Systematic and Applied Microbiology* **5,** 315–326 (cit. on p. 27).

Woese, C. R., W. G. Weisburg, C. M. Hahn, B. J. Paster, L. B. Zablen, B. J. Lewis, T. J. Macke, W. Ludwig and E. Stackebrandt (1985). The Phylogeny of Purple Bacteria: The Gamma Subdivision. *Systematic and Applied Microbiology* **6,** 25–33 (cit. on p. 27).

Woese, C. R., W. G. Weisburg, B. J. Paster, C. M. Hahn, R. S. Tanner, N. R. Krieg, H. P. Koops, H. Harms and E. Stackebrandt (1984). The phylogeny of purple bacteria: The beta subdivision. *Systematic and Applied Microbiology* **5,** 327–336 (cit. on p. 27).

Wong, W.-T., C.-H. Tseng, S.-H. Hsu, H.-S. Lur, C.-W. Mo, C.-N. Huang, S.-C. Hsu, K.-T. Lee and C.-T. Liu (2014). Promoting effects of a single *Rhodopseudomonas palustris* inoculant on plant growth by *Brassica rapa chinensis* under low fertilizer input. *Microbes and environments* **29,** 303–313 (cit. on p. 49).

Woude, B. J. van der, M. de Boer, N. M. J. van der Put, F. M. van der Geld, R. A. Prins and J. C. Gottschal (1994). Anaerobic degradation of halogenated benzoic acids by photoheterotrophic bacteria. *FEMS Microbiology Letters* **119,** 199–207 (cit. on p. 49).

Wu, D., P. Hugenholtz, K. Mavromatis, R. Pukall, E. Dalin, N. N. Ivanova, V. Kunin, L. Goodwin, M. Wu, B. J. Tindall, S. D. Hooper, A. Pati, A. Lykidis, S. Spring, I. J. Anderson, P. D'haeseleer, A. Zemla, M. Singer, A. Lapidus, M. Nolan, A. Copeland, C. Han, F. Chen, J.-F. Cheng, S. Lucas, C. Kerfeld, E. Lang, S. Gronow, P. Chain, D. Bruce, E. M. Rubin, N. C. Kyrpides, H.-P. Klenk and J. A. Eisen (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462,** 1056–1060 (cit. on p. 27).

Xing, D., Y. Zuo, S. Cheng, J. M. Regan and B. E. Logan (2008). Electricity Generation by *Rhodopseudomonas palustris* DX-1. *Environmental Science & Technology* **42,** 4146–4151 (cit. on pp. 26, 29, 49).

Xu, P., X. M. Qian, Y. X. Wang and Y. B. Xu (1996). Modelling for waste water treatment by *Rhodopseudomonas palustris* Y6 immobilized on fibre in a columnar bioreactor. *Applied microbiology and biotechnology* **44,** 676–82 (cit. on p. 29).

Xu, W., C. Chai, L. Shao, J. Yao and Y. Wang (2016). Metabolic engineering of *Rhodopseudomonas palustris* for squalene production. *Journal of industrial microbiology & biotechnology* **43** (5), 719–725 (cit. on p. 31).

Yang, C. F. and C. M. Lee (2011). Enhancement of photohydrogen production using phbC deficient mutant *Rhodopseudomonas palustris* strain M23. *Bioresource Technology* **102,** 5418–5424 (cit. on p. 27).

Yen, H.-c. and H. Gest (1974). Regulation of biosynthesis of aspartate family amino acids in the photosynthetic bacterium Rhodopseudomonas palustris. *Archives of Microbiology* **101,** 187–210 (cit. on p. 128).

Yoch, D. C. and E. S. Lindstrom (1967). Photosynthetic conversion of formate and CO_2 to glutamate by *Rhodopseudomonas palustris*. *Biochemical and Biophysical Research Communications* **28,** 65–69 (cit. on p. 49).

Yokoyama, H., M. Waki, N. Moriya, T. Yasuda, Y. Tanaka and K. Haga (2007). Effect of fermentation temperature on hydrogen production from cow waste slurry by using anaerobic microflora within the slurry. *Applied Microbiology and Biotechnology* **74,** 474–483 (cit. on p. 126).

Yurkov, V. V. and J. T. Beatty (1998). Aerobic anoxygenic phototrophic bacteria. *Microbiology and molecular biology reviews : MMBR* **62,** 695–724 (cit. on p. 27).

Zhang, Y., A. Kumar, P. V. Vadlani and S. Narayanan (2013). Production of nitrogen-based platform chemical: cyanophycin biosynthesis using recombinant *Escherichia coli* and renewable media substitutes. *Journal of Chemical Technology & Biotechnology* **88,** 1321–1327 (cit. on p. 158).

Zhao, L., C. Zhao, D. Han, S. Yang, S. Chen and C.-P. Yu (2011). Anaerobic utilization of phenanthrene by *Rhodopseudomonas palustris*. *Biotechnology Letters* **33,** 2135–2140 (cit. on pp. 30, 60).

Ziegler, K., R. Deutzmann and W. Lockau (2002). Cyanophycin synthetase-like enzymes of non-cyanobacterial eubacteria: characterization of the polymer produced by a recombinant synthetase of *Desulfitobacterium hafniense*. *Zeitschrift fur Naturforschung. C, Journal of biosciences* **57,** 522–529 (cit. on p. 127).

Ziegler, K., A. Diener, C. Herpin, R. Richter, R. Deutzmann and W. Lockau (1998). Molecular characterization of cyanophycin synthetase, the enzyme catalyzing the biosynthesis of the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartate (cyanophycin). *European journal of biochemistry* **254,** 154–159 (cit. on p. 127).

A dual compartment cuvette system for correcting scattering in whole-cell absorbance spectroscopy of photosynthetic microorganisms

A

A.1. Authors

John R. D. Hervey[1], Paolo Bombelli[1], David J. Lea-Smith[1,2], Alan K. Hulme[3], Nathan R. Hulme[3], Robert Keighley[4], Christopher J. Howe[1]*

* Corresponding Author

1. Department of Biochemistry, University of Cambridge, Hopkins Building, Downing Site, Tennis Court Road, Cambridge, CB2 1QW, United Kingdom.

2. School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom

3. Starna Scientific Ltd

4. Shimadzu Corporation

Corresponding author address: ch26@cam.ac.uk; Department of Biochemistry, University of Cambridge, Downing Site, Tennis Court Road, Cambridge, CB2 1QW, UK.

A.2. Abstract

Absorption spectroscopy is widely used to determine the absorption and transmission spectra of chromophores in solution, in addition to suspensions of particles, including micro-organisms. Light scattering, caused by photons deflected from part or all of the cells or other particles in suspension, results in distortions to the absorption spectra, lost information and poor resolution. A spectrophotometer with an integrating sphere may be used to alleviate this problem. However, this instrument is not commonly found in biology laboratories due to the high cost. Here, we describe a novel, rapid, and inexpensive technique that minimises the effect of light scattering when performing whole-cell spectroscopy. This method involves use of a custom made dual compartment cuvette containing

titanium dioxide in one chamber as a scattering agent. Measurements were conducted of a range of different photosynthetic micro-organisms of varying cell size and morphology, including cyanobacteria, microalgae and a purple non-sulphur bacterium. A concentration of 1 mg · mL^{-1} titanium dioxide, measured in a spectrophotometer with a slit width of 5 nm, produced spectra for cyanobacteria and microalgae similar (1 % to 4 % difference) to those obtained using an integrating sphere. The spectrum > 520 nm was similar when measuring the purple non-sulphur bacterium. This system produced superior results compared to those obtained using a recently reported method, the application of the diffusing agent, Scotch™ Magic tape, to the side of the cuvette. The protocol can be completed in an equivalent period of time to standard whole-cell absorbance spectroscopy techniques, and is, in principle, suitable for any dual-beam spectrophotometer.

A.3. Key words

Absorption spectroscopy, cyanobacteria, microalgae, purple non-sulphur bacteria, light scattering, dual compartment cuvette, chromophores, whole cell spectra.

A.4. Introduction

Optical absorption spectroscopy is a technique used to determine how solutions and particles in suspension interact with specific wavelengths of light. It is widely applied to a range of biological and chemical suspensions and is especially useful in determining the pigment composition of photosynthetic organisms (Merzlyak and Chivkunova *et al.*, 2008). However, the apparent absorbance at a given wavelength of any particular suspension depends not only on the actual absorbance but also on the scattering of light by particles and structures present in the suspension (Castanho *et al.*, 1997; Latimer and Eubanks, 1962; Merzlyak and Chivkunova *et al.*, 2008; Merzlyak and Naqvi, 2000; Naqvi *et al.*, 2004; Prado *et al.*, 1996; Twersky, 1970). The scattering effect varies depending on the specific wavelength examined, and the size, shape, and structure of the individual particles in suspension (Latimer and Eubanks, 1962; Merzlyak and Chivkunova *et al.*, 2008; Twersky, 1970). This distorts the absorption measurements, as a portion of the signal recorded will be due to the variable scattering effect, rather than the absorption of light (Latimer and Eubanks, 1962; Merzlyak and Chivkunova *et al.*, 2008; Merzlyak and Naqvi, 2000; Naqvi *et al.*, 2004; Twersky, 1970). This problem has been widely studied, and various methods have been proposed to correct for the light scattered (Jackson *et al.*, 2014; Latimer and Eubanks, 1962; Merzlyak and Naqvi, 2000; Naqvi *et al.*, 2004; Shibata *et al.*, 1954; Smith *et al.*, 1957). Ideally, an integrating sphere is used, in order to collect all of the scattered light at the detector (Latimer and Eubanks, 1962; Merzlyak and Naqvi, 2000). Spectrophotometers incorporating integrating sphere detectors are specialised and expensive pieces

of equipment, and too costly for the majority of research laboratories. Therefore other techniques to correct scattering have been developed. Commonly, these utilise a diffuser on the front of the cuvette to scatter uniformly the beam coming from the light source (Jackson *et al.*, 2014; Shibata *et al.*, 1954; Smith *et al.*, 1957). Scattering introduced by the diffuser is much greater than the scattering from the sample, so the scattering from the sample becomes negligible and is no longer apparent in the recorded spectra. In 1954, Shibata and colleagues (Shibata *et al.*, 1954) proposed the use of filter paper dipped in paraffin wax as a diffuser and in 1957, Smith and colleagues refined the method by exchanging the waxed paper for opalescent glass (Smith *et al.*, 1957). More recently, opaque tape such as Scotch™ Magic tape has been used as a diffuser (Jackson *et al.*, 2014). However, many of these diffusing agents are not easily standardised, so results may vary from experiment to experiment, and from lab to lab; this makes reporting and reproduction of results difficult.

In this study, we outline a new diffuser system, for the correction of scattering in biological suspensions, that is easily standardised, inexpensive, and produces accurate and reproducible results comparable to data obtained using an integrating sphere. This technique utilises custom built dual-compartment cuvettes, manufactured by Starna Scientific, in which scattering is corrected via a suspension of titanium dioxide. These cuvettes can be used in standard dual beam spectrophotometers, making this technique accessible to the majority of biological laboratories. Figure A.1 depicts the details of the dual-compartment cuvettes and its correct use in a dual-beam spectrophotometer. Titanium dioxide is used, as it is widely available and inexpensive, and forms an opaque, scattering suspension in water. Importantly, titanium dioxide can be standardised with respect to particle size and concentration, thus making it suitable as an easily standardised and reproducible diffuser. We demonstrate that this system is superior to using Scotch™ Magic tape as the diffuser, producing results comparable to data collected using an integrating sphere detector.

A.5. Materials and methods

A.5.1. Culture medium and growth conditions

Synechocystis sp. PCC 6803 and *Synechocystis* sp. PCC 6803 Olive strains were cultured in BG11 medium at 30 °C. *Chlorella vulgaris* CCAP 211/52, *Chlamydomonas reinhardtii* CC1021 and *Chlamydomonas reinhardtii* CW15 strains were cultured in standard TP medium at 30 °C. *Synechococcus* sp. PCC 7002 and *Dunaliella salina* CCAP 19/12 were grown in artificial salt water medium (ASW) at 25 °C. *Rhodopseudomonas palustris* CGA009 was grown in a minimal salts medium, supplemented with 10 mM urea and 50 mM glycerol. All of these strains were cultured in an illuminated incubator with continuous shaking (125 rpm) and light (~50 µmol · photons · m^{-2} · s^{-1}). *Escherichia coli* DH5α was grown in lysogeny broth at 37 °C with continuous shaking.

A.5.2. Spectrophotometric methods

One set of spectra was recorded in a Shimadzu UV-2600 spectrophotometer, with a slit width of 5 nm. Reference spectra were recorded using the same device with an add-on ISR-2600Plus two-detector integrating sphere, with a slit width of 5 nm. A second set of data were collected using a Shimadzu UV-1800 spectrophotometer, which has a slit width of 1 nm. These data were compared to reference spectra collected in the Shimadzu UV-2600 with the add-on ISR-2600Plus two-detector integrating sphere, with the slit width set to 1 nm. All experiments were conducted at room temperature.

A.5.3. Dual compartment cuvette measurements with titanium dioxide

Two quartz cuvettes were used in the dual-beam spectrophotometer. In these customized cuvettes, the chamber is divided into two compartments, each with an optical path of 5 mm (Figure A.1). The cuvettes were orientated so that the investigation beam passed through both chambers. A baseline reading was taken with blank growth medium in one chamber of both cuvettes (i.e., the sample and the reference cuvettes) and a premixed suspension of titanium dioxide (Sigma Aldrich, Titanium IV oxide, anatase, -325 mesh) in water in the other chamber (Figure A.1). The titanium dioxide suspension was placed in the chamber closer to the light source, such that the beam passed through the suspension before passing through the sample. The spectrophotometer was zeroed with this setup by recording a baseline correction. The cuvette is inverted several times each time a new sample is measured, to ensure that the titanium dioxide remains in suspension and does not settle. The concentrations of titanium dioxide used were 0.1 mg · mL^{-1}, 0.2 mg · mL^{-1}, 0.5 mg · mL^{-1} and 1 mg · mL^{-1}. To record a spectrum, the titanium dioxide suspension was left in place in both cuvettes, but the blank medium in the sample cuvette was replaced with the sample (e.g., microalgal cell suspension). For the reference spectra recorded using the integrating sphere detector, the titanium dioxide suspension was replaced with water. This way the path length of the sample was the same as that used for the other samples. Three technical replicate measurements were taken for each sample.

A.5.4. Single compartment cuvette measurements with Scotch™ Magic tape

Two standard, single-compartment, quartz cuvettes were used in the dual-beam spectrophotometer. These cuvettes had 0, 1, 5, or 10 layers of Scotch™ Magic tape applied to the side of the cuvette that was closer to the light source, such that the investigation beam passed through the tape before passing through the sample. Adding more than 10 layers resulted in a cuvette unable to fit in the spectrophotometer. Prior to analysis of samples,

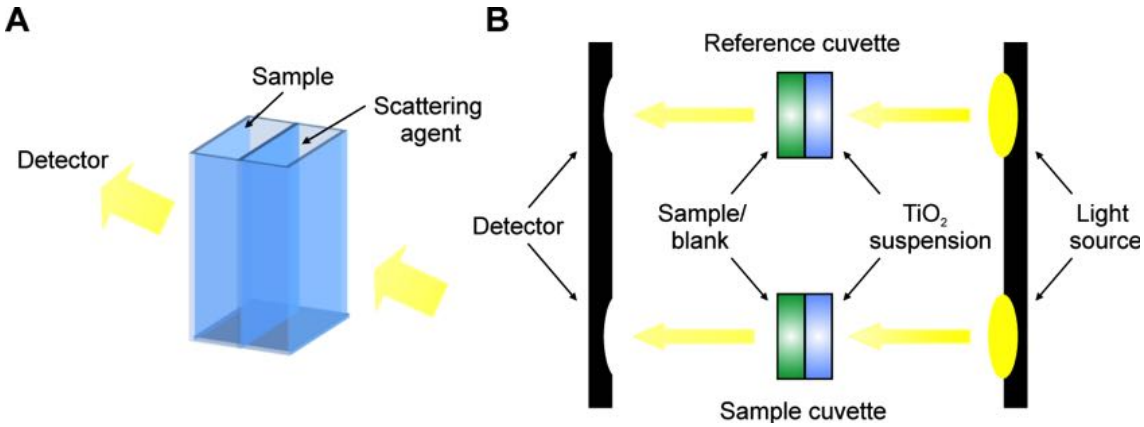


Figure A.1. – Design and correct use of the custom, two chamber cuvette. (A) Schematic of the device detailing the compartments for inclusion of a sample and a scattering agent; (B) Schematic detailing correct use of the cuvette in a dual-beam spectrophotometer.

Table A.1. – Species examined in this study. Size refers to the diameter in spherical cells and length/width in ovoid and rod shaped cells.

Species	Shape	Size (µm)	Reference
Synechocystis sp. PCC 6803	Spherical	2.02 - 2.06	(Lea-Smith and Bombelli *et al.*, 2014)
Synechocystis sp. PCC 6803 Olive	Spherical	1.82	(Lea-Smith and Bombelli *et al.*, 2014)
Synechococcus sp. PCC 7002	Ovoid	2.30 / 1.61	(Lea-Smith and Ortiz-Suarez *et al.*, 2016)
Chlorella vulgaris CCAP 211/52	Spherical	3	(Grooth *et al.*, 1985)
Dunaliella salina CCAP 19/12	Ovoid	5 - 10	(Preetha *et al.*, 2012)
Chlamydomonas reinhardtii CC1021	Spherical	10	(Ratcliff *et al.*, 2013)
Chlamydomonas reinhardtii CW15	Spherical	10	(Davies and Plaskitt, 1971; Ratcliff *et al.*, 2013)
Rhodopseudomonas palustris CGA009	Rod	2 / 0.5	(Hayashi *et al.*, 1982; Niel, 1944)
Escherichia coli DH5α	Rod	2 / 0.5	(Bronk *et al.*, 1992)

the spectrophotometer was zeroed, and a baseline recorded, with blank media in both cuvettes. To record a spectrum, the blank medium in the sample cuvette was replaced with the sample (e.g., microalgal cell suspension). For the reference spectra recorded using the integrating sphere detector, the tape was removed from the cuvettes. Three technical replicate measurements were taken for each sample.

A.6. Results

In order to determine the absorption profile of a range of photosynthetic organisms we analysed two cyanobacterial species (*Synechocystis* and *Synechococcus*), a non-sulphur purple bacterium (*R. palustris*) and three eukaryotic microalgal species (*Ch. vulgaris*, *C. reinhardtii*, *D. salina*). The absorption profile of the non-photosynthetic bacterium *E. coli* was also examined, so that comparisons could be made to a sample that did not contain any chromophores. Species were selected based on differences in size and shape, which are outlined in Table A.1. In addition, a *Synechocystis* mutant strain (Olive) lacking the phycocyanin portion (λ maximum = 625 nm) of the light harvesting phycobilisome complex and whose cells are smaller than wild-type [11], and a cell wall deficient *C. reinhardtii* strain (CW15) [12], were also tested.

A.6.1. Analysis of the absorption spectra of strains using the dual compartment cuvette with titanium dioxide

We first analysed the absorption profile of all the strains using a spectrophotometer with an integrating sphere detector and a slit width of 5 nm. The same samples were then examined using the dual chamber cuvette system with no titanium dioxide or different concentrations between 0.1 1mgmL. The raw data (Figure A.2) were analysed in the region 400 nm to 750 nm, except for *R. palustris* where the range was from 400 nm to 900 nm. Each spectrum was normalised by dividing every point forming the curve by the maximum value of absorbance measured in the considered interval. By doing that, to the maximum value of absorbance in each spectrum was given the value of "1". This normalization permits a direct comparison of spectra derived from samples with various cell densities (Figure A.3).

When titanium dioxide at 1 mg $\cdot$ mL^{-1} was used, the spectra of each of the cyanobacterial and microalgal species were similar to the profile obtained using the integrated sphere, in terms of both the magnitude and the overall shape. When titanium dioxide at 0.5 mg $\cdot$ mL^{-1} was used, the spectral profile was similar from 400 nm to 680 nm but diverged between 680 nm to 750 nm. This concentration is therefore unsuitable for most applications since this part of the spectrum includes absorption from chlorophyll *a*. The *R. palustris* profile was similar between the results obtained using the integrated sphere and the dual chamber cuvette system when titanium dioxide was used at 1 mg $\cdot$ mL^{-1}, except below 520 nm and

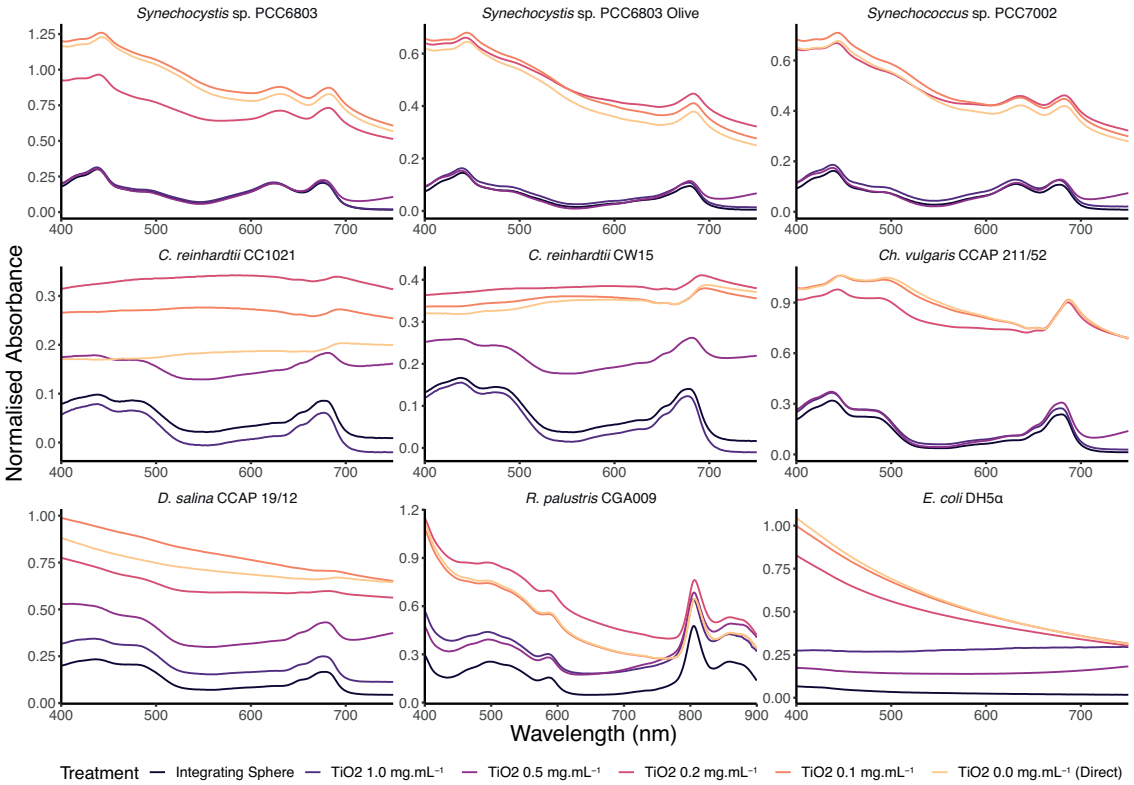


Figure A.2. – Raw data comparison of whole-cell absorbance spectra with the dual compartment cuvette (slit width 5 nm). Samples were analysed using the integrating sphere (black) or in the dual compartment cuvette with 0 (yellow), 0.1 (orange), 0.2 (red), 0.5 (purple) or 1 (blue) $mg \cdot mL^{-1}$ TiO_2. Results are not standardised. The mean of three samples is displayed.

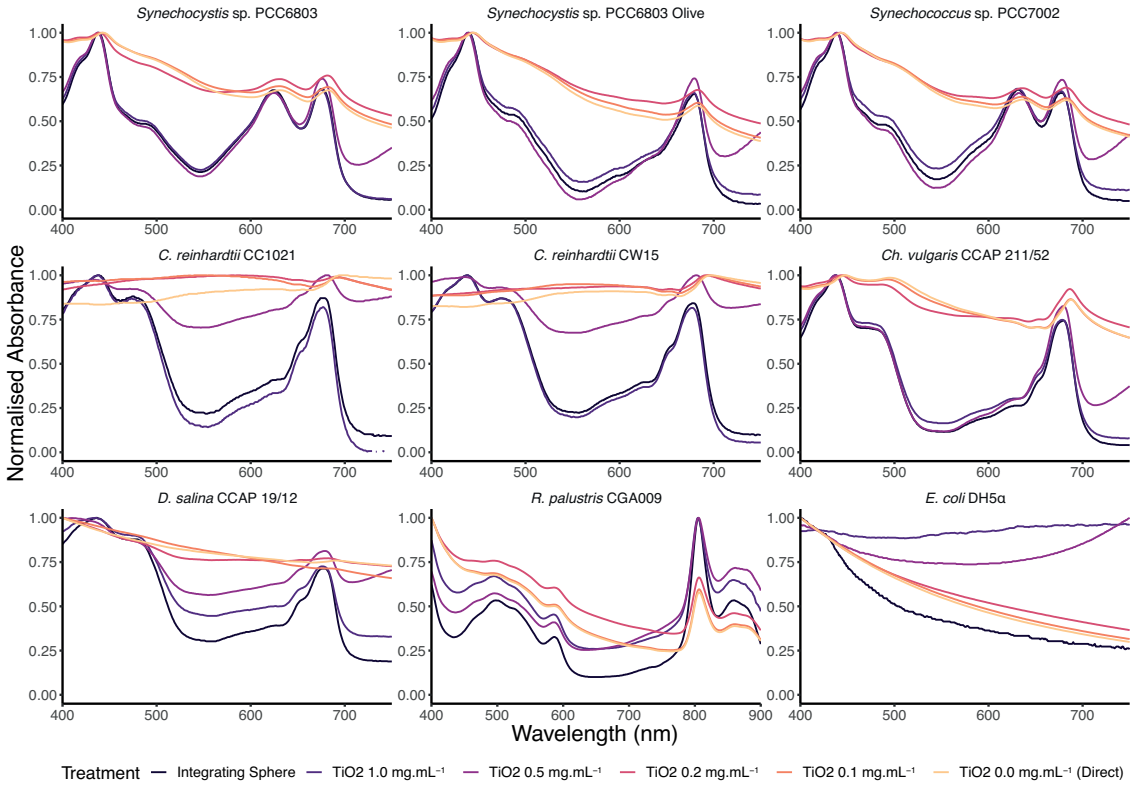


Figure A.3. – Comparison of whole-cell absorbance spectra with the dual compartment cuvette (slit width 5 nm). Samples were analysed using the integrating sphere (black) or in the dual compartment cuvette with 0 (yellow), 0.1 (orange), 0.2 (red), 0.5 (purple) or 1 (blue) mg · mL^{-1} TiO_2. Results are standardised as described in the text. The mean of three samples is displayed.

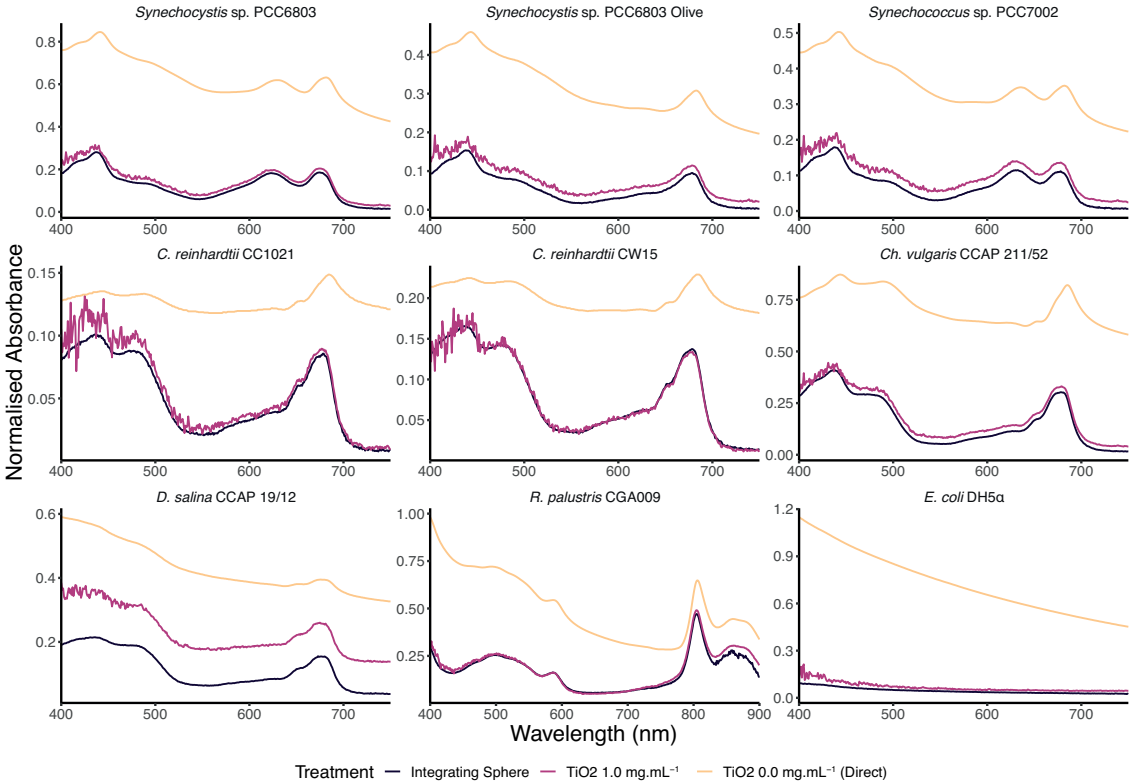


Figure A.4. – Raw data comparison of whole-cell absorbance spectra with the dual compartment cuvette (slit width 1 nm). Samples were analysed using the integrating sphere (black) or in the dual compartment cuvette with 0 (yellow) or 1 (purple) mg · mL^{-1} TiO_2. Results are not standardised. The mean of three samples is displayed.

920 nm. Therefore this technique would be suitable for analysing absorption of the main pigments in this species in the infrared part of the spectrum, specifically bacteriochlorophyll *a* (λ maximum = 803 nm) and b (λ maximum = 860 nm). The absence of absorption peaks between 400 nm to 750 nm in the *E. coli* spectrum precluded a direct comparison.

We then analysed all strains using a spectrophotometer with an integrating sphere detector and a slit width of 1 nm, followed by the dual chamber cuvette system with no titanium dioxide or 1 mg · mL^{-1} (Figure A.4; Figure A.5). The spectral profiles of the cyanobacterial samples, *C. reinhardtii* CW15 and *Ch. vulgaris* were similar between those analysed using the integrating sphere and the dual chamber cuvette system with titanium dioxide at 1 mg · mL^{-1}. However, with the exception of *C. reinhardtii* CW15, variation between replicates was higher between 400 nm to 500 nm compared to samples examined using the spectrophotometer with a slit width of 5 nm. *C. reinhardtii* CC1021, *D. salina* and *R. palustris* displayed greater variation, suggesting that this setup is not suitable for analysis of these strains.

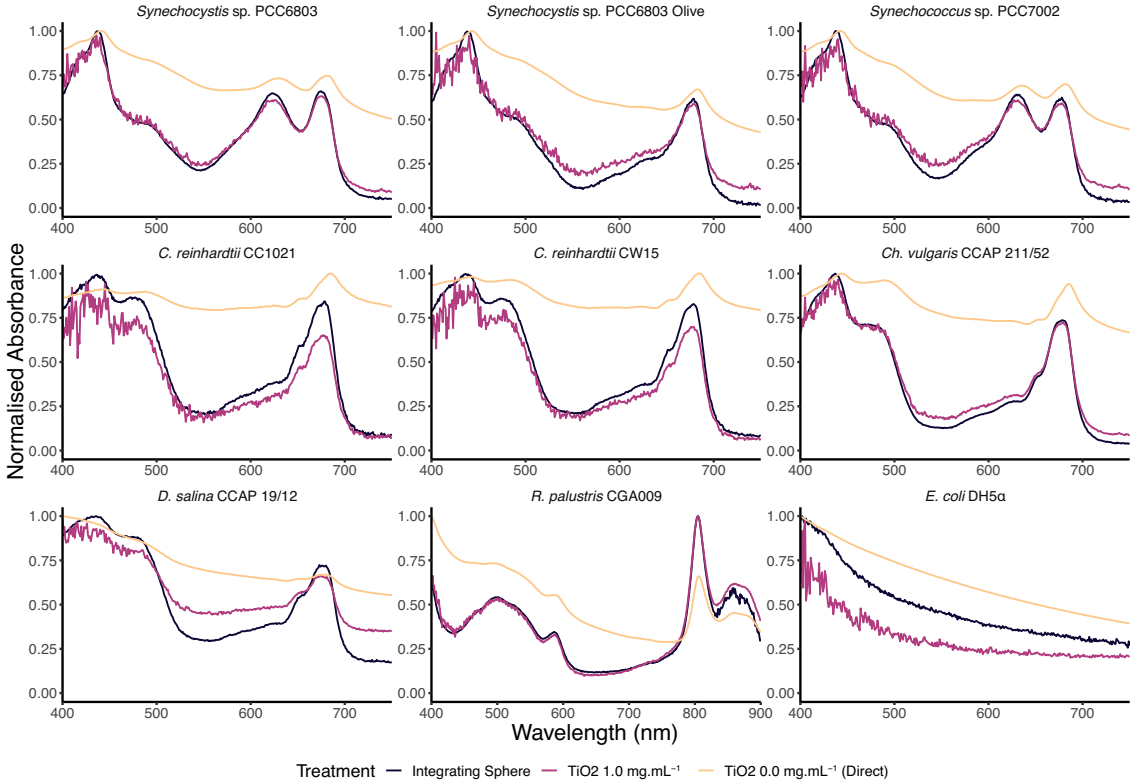


Figure A.5. – Comparison of whole-cell absorbance spectra with the dual compartment cuvette (slit width 1 nm). Samples were analysed using the integrating sphere (black) or in the dual compartment cuvette with 0 (yellow) or 1 (purple) mg · mL^{-1} TiO_2. Results are standardised as described in the text. The mean of three samples is displayed.

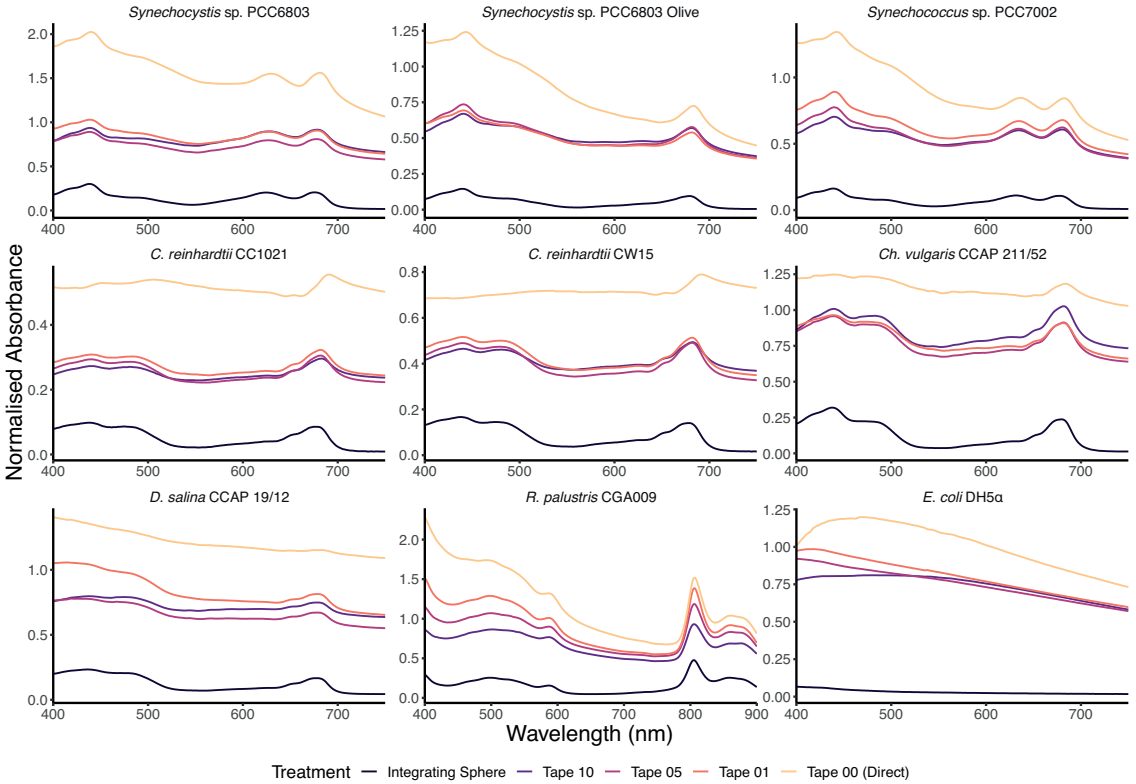


Figure A.6. – Raw data comparison of whole-cell absorbance spectra with Scotch™ Magic tape (slit width 5 nm). Samples were analysed using the integrating sphere (black) or in the single compartment cuvette coated with 0 (yellow), 1 (orange), 5 (purple) or 10 (blue) pieces of Scotch™ Magic tape. Results are not standardised. The mean of three samples is displayed.

A.6.2. Analysis of the absorption spectra of strains using the single compartment cuvette with Scotch™ Magic tape

Next we performed a comparison of this method with one recently reported in the literature, the coating of a single chamber cuvette with Scotch™ Magic tape [10]. The samples were analysed in a spectrophotometer with a slit width of 5 nm (Figure A.6; Figure A.7) or 1 nm (Figure A.8; Figure A.9), with different numbers of layers of Scotch™ Magic tape. With a slit width of 5 nm, adding even one layer of tape markedly changed the profile compared to just analysing the sample using the cuvette only. However, the profile of none of the samples measured with either 1, 5 or 10 layers of tape resembled closely the profile across the spectrum obtained using the integrating sphere. Surprisingly, when using the spectrophotometer with the slit width of 1 nm, there was little difference in the profile of samples between the cuvette only and the cuvette with ten layers of tape, with the exception of *D. salina*.

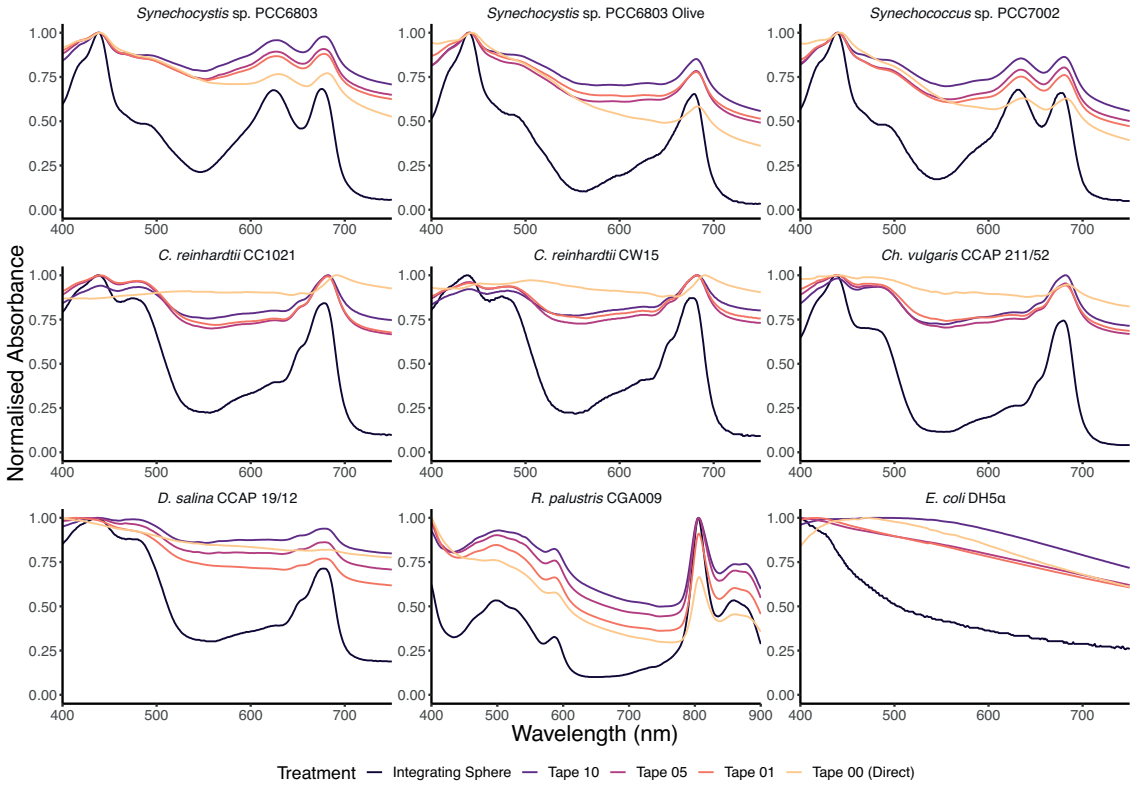


Figure A.7. – Comparison of whole-cell absorbance spectra with Scotch™ Magic tape (slit width 5 nm). Samples were analysed using the integrating sphere (black) or in the single compartment cuvette coated with 0 (yellow), 1 (orange), 5 (purple) or 10 (blue) pieces of Scotch™ Magic tape. Results are standardised as described in the text. The mean of three samples is displayed.

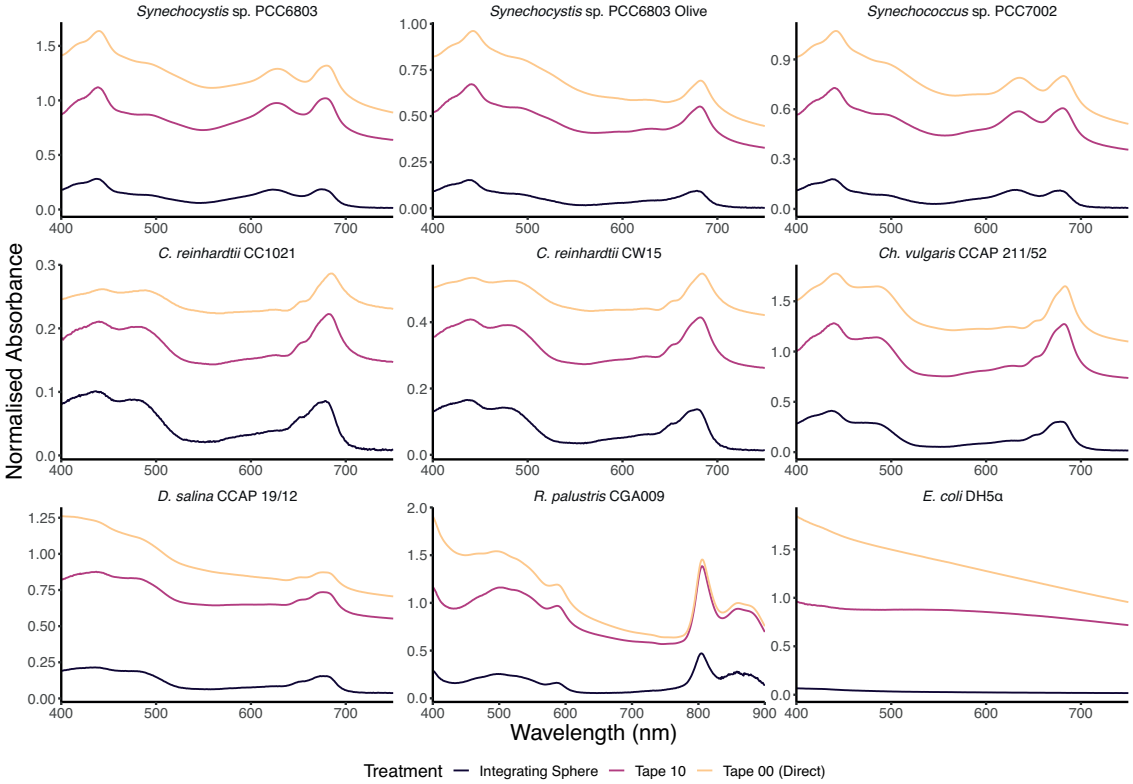


Figure A.8. – Raw data comparison of whole-cell absorbance spectra with Scotch™ Magic tape (slit width 1 nm). Samples were analysed using the integrating sphere (black) or in the single compartment cuvette coated with 0 (yellow) or 10 (purple) pieces of Scotch™ Magic tape. Results are not standardised. The mean of three samples is displayed.

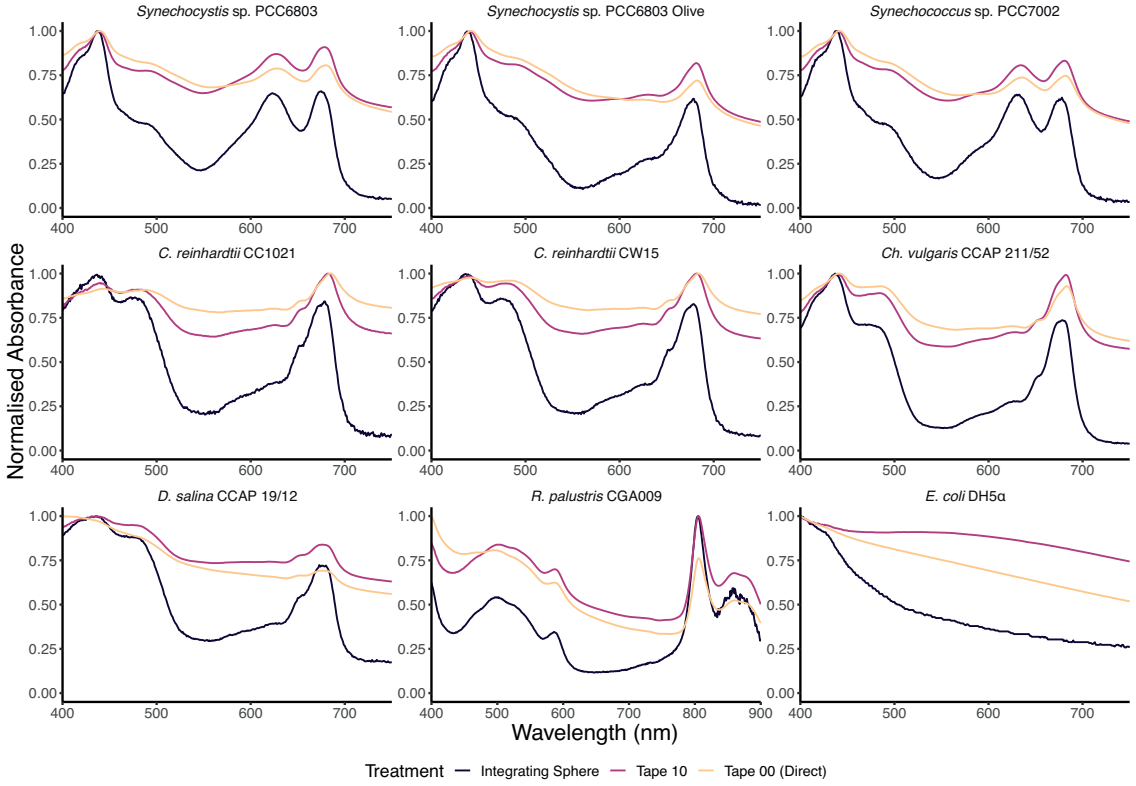


Figure A.9. – Comparison of whole-cell absorbance spectra with Scotch™ Magic tape (slit width 1 nm). Samples were analysed using the integrating sphere (black) or in the single compartment cuvette coated with 0 (yellow) or 10 (purple) pieces of Scotch™ Magic tape. Results are standardised as described in the text. The mean of three samples is displayed.

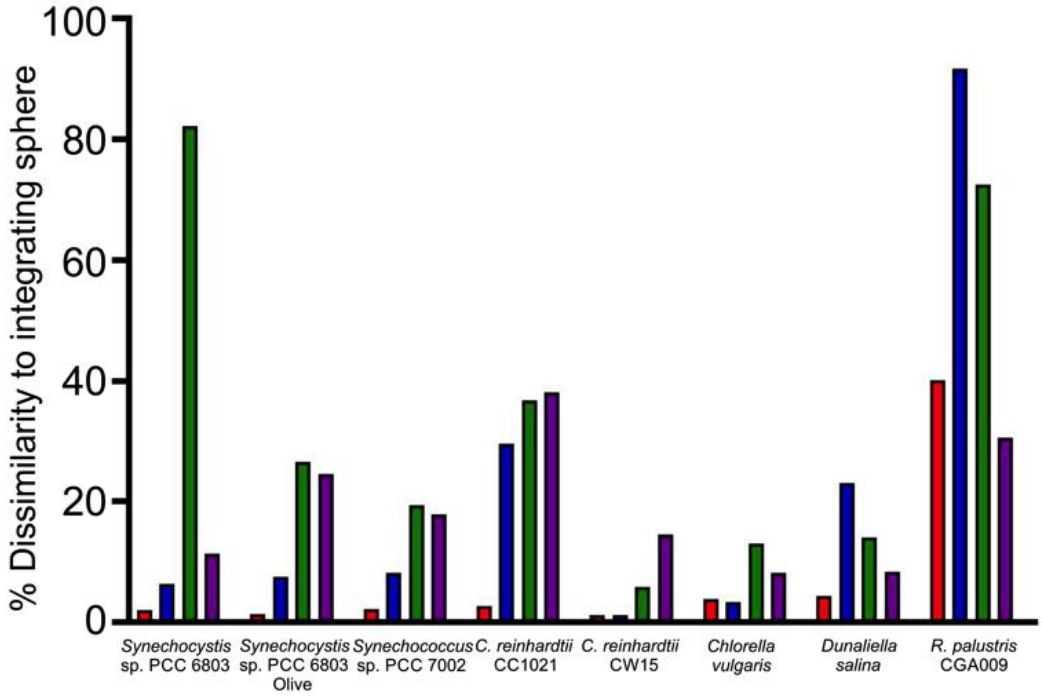


Figure A.10. – Average difference across the spectrum (400 nm to 750 nm (400 nm to 900 nm for *R. palustris*)) between data obtained using the integrating sphere and the dual compartment/TiO_2 and single compartment/tape systems. Shown is the result most comparable to the integrating sphere from experiments conducted using the dual compartment/TiO_2 system with a slit width of 5 nm (red) and 1 nm (blue), and the single compartment/tape system with a slit width of 5 nm (green) and 1 nm (purple).

A.6.3. Analysis of absorbance is optimal using the dual compartment cuvette with 1 mg · mL^{-1} titanium dioxide in a spectrophotometer with a slit width of 5 nm

Finally, we determined the average difference across the spectrum (400 nm to 750 nm for cyanobacteria/microalgae; 400 nm to 900 nm for Rhodopseudomonas) between the different methods compared to the reference spectra collected the with integrating sphere (Figure A.10; Figure A.11). For all species, 1 mg · mL^{-1} titanium dioxide in a dual compartment cuvette in a spectrophotometer with a slit width of 5 nm was the optimal method. Using a spectrophotometer with a slit width of 1 nm resulted in spectra more comparable to the reference than using a single compartment cuvette with layers of tape.

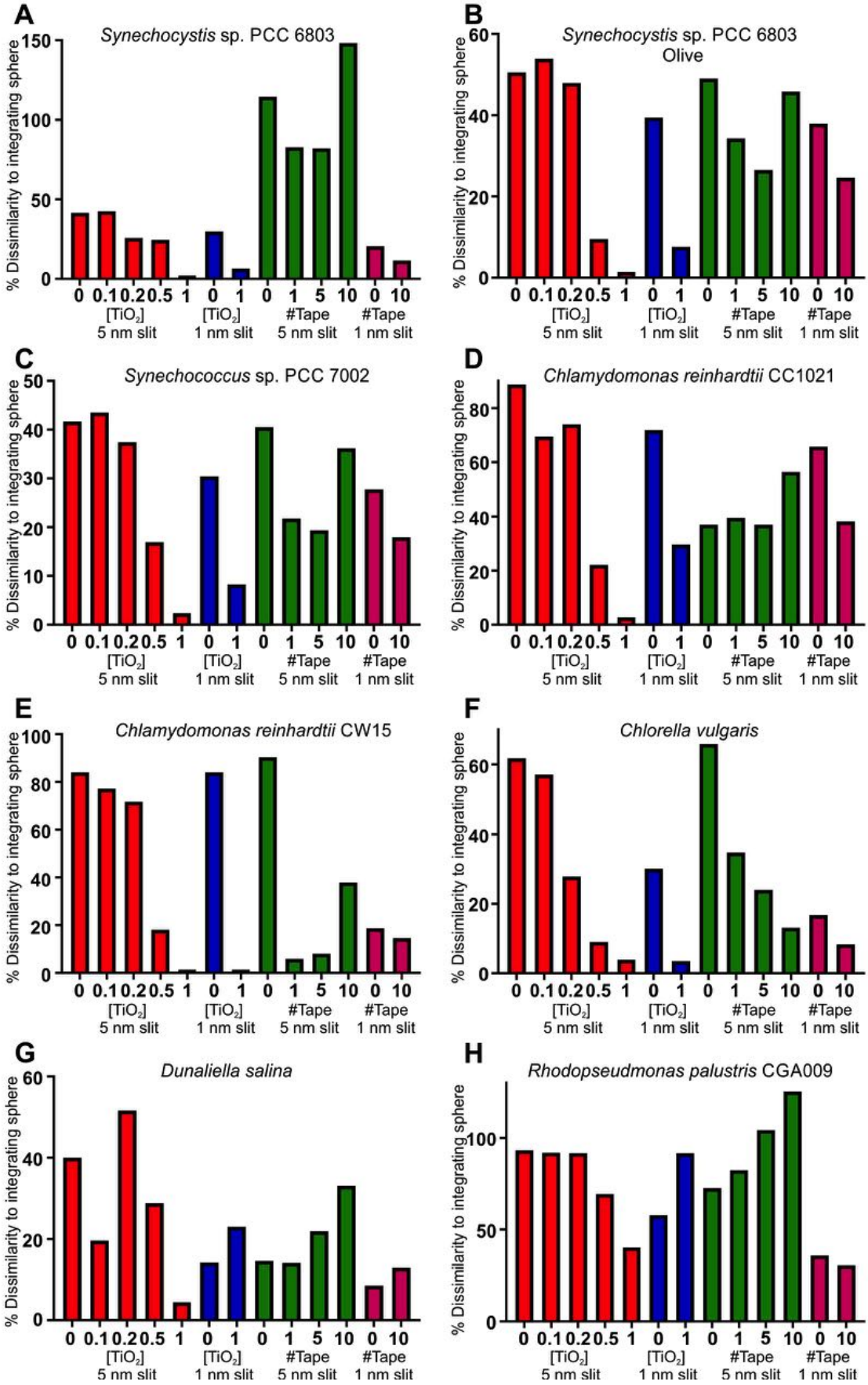


Figure A.11. – Differences compared to results obtaining using the integrating sphere. Average differences between 400 nm to 750 nm (400 nm to 900 nm for *R. palustris*) results obtained using the four methods outlined in this paper compared to the optimal data acquired via the integrating sphere.

A.7. Discussion

In this study, we tested seven microorganisms ranging in diameter from 2 µM to 10 µM using two different scattering agents. Overall, we showed that the two-compartment cuvette system with titanium dioxide can be used to make accurate absorption measurements, unhindered by the effects of light scattering, in standard dual-beam spectrophotometers, without an integrating sphere. This method can easily be replicated between different laboratories, since the concentration and particle size of the titanium dioxide is known and consistent. Increasing the concentration of titanium dioxide brought the measured spectra closer to the reference spectra in every case. Using a titanium dioxide concentration of 1 mg · mL^{-1} gave spectra very similar to those collected using an integrating sphere, suggesting that it was the optimum concentration of titanium dioxide for the diffuser. Increasing the concentration of titanium dioxide further would not be beneficial, since the recorded spectra already matched the reference spectra almost perfectly, and the increased opacity from higher concentrations would only serve to reduce the signal that could be measured.

This method offers several advantages compared to those using waxed paper (Shibata *et al.*, 1954), opalescent glass (Smith *et al.*, 1957), or tape (Jackson *et al.*, 2014) as the scattering diffuser. First, the spectra collected are similar to the reference spectra collected using the integrating sphere. Secondly, as previously mentioned, it is easily standardised and replicated across different laboratories and experiments. Thirdly, the method can be easily modified to change the scattering effect, by changing the concentration of titanium dioxide used. Finally, whilst customised cuvettes are needed, they can be used in any standard dual-beam spectrophotometer without modification, although a device with a slit width of 5 nm is optimal. Adding tape or other diffusers to the side of an existing cuvette add to its size, and can result in problems fitting the cuvette into the spectrophotometer. This method is therefore applicable to whole-cell spectroscopy of organisms and could potentially be applied to other opaque or suspended samples currently analysed using a spectrophotometer with an integrating sphere.

A.8. Acknowledgements

The authors would like to thank Starna Scientific Ltd., who manufactured the custom, dual-compartment cuvettes, and Shimadzu UK Ltd., for generously allowing us to use their instruments and facilities. John Hervey and David Lea-Smith are funded by the Waste Environmental Education Research Trust. Paolo Bombelli is funded by the Leverhulme Trust.

References

Bronk, B. V., W. P. Van De Merwe and M. Stanley (1992). In Vivo Measure of Average Bacterial Cell Size From a Polarized Light Scattering Function. *Cytomtery* **13,** 155–162 (cit. on p. 191).

Castanho, M. A. R. B., N. C. Santos and L. M. S. Loura (1997). Separating the turbidity spectra of vesicles from the absorption spectra of membrane probes and other chromophores. *European Biophysics Journal* **26,** 253–259 (cit. on p. 188).

Davies, D. R. and A. Plaskitt (1971). Genetical and structural analyses of cell-wall formation in *Chlamydomonas reinhardtii*. *Genetical Research* **17,** 33–43 (cit. on p. 191).

Grooth, B. G. le, T. H. Geerken and J. Greve (1985). The cytodisk: A cytometer based upon a new principle of cell alignment. *Cytometry* **6,** 226–233 (cit. on p. 191).

Hayashi, H., M. Miyao and S. Morita (1982). Absorption and Fluorescence Spectra of Light-Harvesting Bacteriochlorophyll-Protein Complexes from *Rhodopseudomonas palustris* in the Near-Infrared Region. *Journal of Biochemistry* **91,** 1017–1027 (cit. on p. 191).

Jackson, S. A., J. R. D. Hervey, A. J. Dale and J. J. Eaton-Rye (2014). Removal of both Ycf48 and Psb27 in *Synechocystis* sp. PCC 6803 disrupts Photosystem II assembly and alters QA– oxidation in the mature complex. *FEBS Letters* **588,** 3751–3760 (cit. on pp. 188, 189, 203).

Latimer, P. and C. A. H. Eubanks (1962). Absorption spectrophotometry of turbid suspensions: a method of correcting for large systematic distortions. *Archives of biochemistry and biophysics* **98,** 274–285 (cit. on p. 188).

Lea-Smith, D. J., P. Bombelli, J. S. Dennis, S. A. Scott, A. G. Smith and C. J. Howe (2014). Phycobilisome-Deficient Strains of *Synechocystis* sp. PCC 6803 Have Reduced Size and Require Carbon-Limiting Conditions to Exhibit Enhanced Productivity. *Plant physiology* **165,** 705–714 (cit. on p. 191).

Lea-Smith, D. J., M. L. Ortiz-Suarez, T. Lenn, D. J. Nürnberg, L. L. Baers, M. P. Davey, L. Parolini, R. G. Huber, C. A. R. Cotton, G. Mastroianni, P. Bombelli, P. Ungerer, T. J. Stevens, A. G. Smith, P. J. Bond, C. W. Mullineaux and C. J. Howe (2016). Hydrocarbons Are Essential for Optimal Cell Size, Division, and Growth of Cyanobacteria. *Plant Physiology* **172,** 1928–1940 (cit. on p. 191).

Merzlyak, M. N., O. B. Chivkunova, I. P. Maslova, K. R. Naqvi, A. E. Solovchenko and G. L. Klyachko-Gurvich (2008). Light absorption and scattering by cell suspensions of some cyanobacteria and microalgae. *Russian Journal of Plant Physiology* **55,** 420–425 (cit. on p. 188).

Merzlyak, M. N. and K. R. Naqvi (2000). On recording the true absorption spectrum and the scattering spectrum of a turbid sample: Application to cell suspensions of the cyanobacterium *Anabaena variabilis*. *Journal of Photochemistry and Photobiology B: Biology* **58,** 123–129 (cit. on p. 188).

Naqvi, K. R., M. N. Merzlyak and T. B. Melø (2004). Absorption and scattering of light by suspensions of cells and subcellular particles: an analysis in terms of Kramers-Kronig relations. *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology* **3,** 132–137 (cit. on p. 188).

Niel, C. B. van (1944). The culture, general physiology, morphology, and classification of the non-sulfur purple and brown bacteria. *Bacteriological reviews* **8,** 1–118 (cit. on p. 191).

Prado, A., C. Puyo, J. Arlucea, F. M. Goni and J. Arechaga (1996). The Turbidity of Cell Nuclei in Suspension: A Complex Case of Light Scattering. *Journal Colloid Interface Science* **177,** 9–13 (cit. on p. 188).

Preetha, K., L. John, C. S. Subin and K. K. Vijayan (2012). Phenotypic and genetic characterization of *Dunaliella* (Chlorophyta) from Indian salinas and their diversity. *Aquatic biosystems* **8,** 27 (cit. on p. 191).

Ratcliff, W. C., M. D. Herron, K. Howell, J. T. Pentz, F. Rosenzweig and M. Travisano (2013). Experimental evolution of an alternating uni- and multicellular life cycle in *Chlamydomonas reinhardtii*. *Nature Communications* **4,** 2742 (cit. on p. 191).

Shibata, K., A. A. Benson and M. Calvin (1954). The absorption spectra of suspensions of living micro-organisms. *Biochimica et Biophysica Acta* **15,** 461–470 (cit. on pp. 188, 189, 203).

Smith, J. H. C., K. Shibata and R. W. Hart (1957). A Spectrophotometer accessory for measuring absorption spectra of light-scattering samples: Spectra of dark-grown albino leaves and of adsorbed chlorophylls. *Archives of Biochemistry and Biophysics* **72,** 457–464 (cit. on pp. 188, 189, 203).

Twersky, V. (1970). Absorption and multiple scattering by biological suspensions. *Journal of the Optical Society of America* **60,** 1084–1093 (cit. on p. 188).

Significantly differentially expressed genes of *R. palustris* cultures grown with Urea vs. $[NH_4]_2SO_4$ as nitrogen source

B

Table B.1. – Significantly differentially expressed genes, adjusted p-value < 0.05, Log_2 Fold change > 2 (Urea relative to $[NH_4]_2SO_4$)

Gene name	Transcript ID	Fold change	p-value	Adjusted p-value	Log2 fold change
RPA4614	MSTRG.6350	181.1164958	0.0000027	0.0055736	7.500774
gyrB	MSTRG.37	0.0931913	0.0000017	0.0055736	-3.423662
RPA0565	MSTRG.863	0.0328749	0.0000076	0.0089731	-4.926868
RPA2108	MSTRG.2687	6.9134872	0.0000068	0.0089731	2.789414
RPA2107	MSTRG.2687	6.9134872	0.0000068	0.0089731	2.789414
RPA0579	MSTRG.884	5.3618557	0.0000129	0.0112048	2.422732
fixB	MSTRG.6340	528.3572268	0.0000399	0.0123797	9.045370
fdxB	MSTRG.6347	62.2064003	0.0000236	0.0123797	5.958991
RPA3756	MSTRG.4952	11.5331683	0.0000256	0.0123797	3.527717
RPA2002	MSTRG.2538	10.4933268	0.0000355	0.0123797	3.391400
RPA1773	MSTRG.2167	8.6678446	0.0000319	0.0123797	3.115673
RPA3154	MSTRG.3922	4.3224059	0.0000340	0.0123797	2.111835
nifE	MSTRG.6354	199.7290100	0.0000439	0.0125828	7.641900
nifS2	MSTRG.6343	174.2330120	0.0000599	0.0125828	7.444874
nifU	MSTRG.6343	174.2330120	0.0000599	0.0125828	7.444874
nifV	MSTRG.6343	174.2330120	0.0000599	0.0125828	7.444874
uvrA	MSTRG.3468	0.0977537	0.0000567	0.0125828	-3.354705
RPA4031	MSTRG.5303	0.1498780	0.0000537	0.0125828	-2.738139
RPA0807	MSTRG.1086	0.2305962	0.0000582	0.0125828	-2.116559
nifK	MSTRG.6355	327.2046703	0.0000793	0.0133294	8.354050
fixA	MSTRG.6341	293.4704085	0.0000863	0.0133294	8.197071
nifN	MSTRG.6352	137.8367293	0.0000855	0.0133294	7.106817
RPA4613	MSTRG.6349	94.2240083	0.0000826	0.0133294	6.558023
nifQ	MSTRG.6346	67.4387063	0.0000876	0.0133294	6.075505
RPA2443	MSTRG.3324	0.0504889	0.0000847	0.0133294	-4.307889
RPA2194	MSTRG.3073	0.1238349	0.0000725	0.0133294	-3.013510

Table B.1. – Significantly differentially expressed genes, adjusted p-value < 0.05, Log_2 Fold change > 2 (Urea relative to $[NH_4]_2SO_4$) *(continued)*

Gene name	Transcript ID	Fold change	p-value	Adjusted p-value	Log2 fold change
dnaX	MSTRG.936	0.2073262	0.0000845	0.0133294	-2.270026
RPA4079	MSTRG.5088	4.2717426	0.0000860	0.0133294	2.094825
RPA4808	MSTRG.6086	12.6403765	0.0000988	0.0139898	3.659968
nifD	MSTRG.6356	545.0400877	0.0001156	0.0141239	9.090219
fixU%2C nifT	MSTRG.6360	49.5425609	0.0001181	0.0141239	5.630597
RPA4515	MSTRG.6233	11.9747788	0.0001303	0.0141239	3.581927
leuC	MSTRG.419	0.0931819	0.0001277	0.0141239	-3.423806
pgm1	MSTRG.515	0.1485070	0.0001221	0.0141239	-2.751397
RPA2480	MSTRG.3378	4.3691921	0.0001268	0.0141239	2.127366
RPA1515	MSTRG.1942	6.0823038	0.0001357	0.0142425	2.604618
RPA3508	MSTRG.4533	0.0469745	0.0001463	0.0146066	-4.411980
RPA3054	MSTRG.3558	12.5239389	0.0001488	0.0146066	3.646616
fixC	MSTRG.6339	383.3931093	0.0001538	0.0148616	8.582681
RPA2808	MSTRG.3455	9.1851321	0.0001701	0.0150732	3.199301
cheB2	MSTRG.2922	8.7564163	0.0001804	0.0150732	3.130340
cheR2	MSTRG.2922	8.7564163	0.0001804	0.0150732	3.130340
RPA1217	MSTRG.1379	6.6069429	0.0001832	0.0150732	2.723983
hemD	MSTRG.5307	4.0000912	0.0001826	0.0150732	2.000033
dapF	MSTRG.432	0.1381627	0.0001891	0.0152221	-2.855560
RPA0251	MSTRG.432	0.1381627	0.0001891	0.0152221	-2.855560
RPA3770	MSTRG.4970	0.0864246	0.0001972	0.0153947	-3.532414
RPA4638	MSTRG.6372	24.9266985	0.0002157	0.0162118	4.639620
cheY2	MSTRG.2921	17.4486946	0.0002311	0.0164137	4.125047
RPA4822	MSTRG.6103	6.5233313	0.0002249	0.0164137	2.705609
RPA2801	MSTRG.3446	5.3693708	0.0002331	0.0164137	2.424753
RPA0500	MSTRG.563	0.1900318	0.0002337	0.0164137	-2.395687
RPA1862	MSTRG.2266	4.7852417	0.0002346	0.0164137	2.258592
nifH4	MSTRG.6357	467.6659077	0.0002641	0.0164969	8.869335
ilvJ	MSTRG.4658	67.8952472	0.0002702	0.0164969	6.085239
RPA3774	MSTRG.4974	0.1918378	0.0002682	0.0164969	-2.382041
nirK1	MSTRG.3889	5.0560096	0.0002750	0.0164969	2.337999
RPA2820	MSTRG.3472	4.2001318	0.0002541	0.0164969	2.070435
RPA4583	MSTRG.6314	4.0817564	0.0002748	0.0164969	2.029190
RPA3957	MSTRG.4919	5.0954025	0.0002883	0.0170403	2.349196
RPA0471	MSTRG.786	5.6726179	0.0003048	0.0173392	2.504015
cheR3	MSTRG.2977	8.2452742	0.0003249	0.0178959	3.043567
chvI	MSTRG.539	0.1708446	0.0003352	0.0178959	-2.549243
nifZ	MSTRG.6361	106.9724665	0.0003409	0.0180447	6.741096

Table B.1. – Significantly differentially expressed genes, adjusted p-value < 0.05, Log_2 Fold change > 2 (Urea relative to $[NH_4]_2SO_4$) *(continued)*

Gene name	Transcript ID	Fold change	p-value	Adjusted p-value	Log2 fold change
nifB	MSTRG.6361	106.9724665	0.0003409	0.0180447	6.741096
ferN	MSTRG.6361	106.9724665	0.0003409	0.0180447	6.741096
hesB	MSTRG.6361	106.9724665	0.0003409	0.0180447	6.741096
RPA4624	MSTRG.6361	106.9724665	0.0003409	0.0180447	6.741096
RPA4345	MSTRG.5363	0.1085388	0.0003671	0.0183437	-3.203717
RPA0779	MSTRG.790	30.4493996	0.0003865	0.0189060	4.928342
badR	MSTRG.1408	0.1089788	0.0003882	0.0189060	-3.197881
ilvK	MSTRG.4657	8.3790595	0.0003982	0.0190863	3.066788
nifW	MSTRG.6342	116.9488674	0.0004022	0.0191262	6.869734
RPA4171	MSTRG.5440	17.5633758	0.0004141	0.0193899	4.134498
RPA2001	MSTRG.2537	13.4452520	0.0004190	0.0193967	3.749025
RPA1957	MSTRG.2419	0.1039325	0.0004229	0.0193967	-3.266282
RPA3729	MSTRG.4665	6.3546519	0.0004238	0.0193967	2.667813
RPA4634	MSTRG.6367	82.5375396	0.0004696	0.0194182	6.366979
RPA3188	MSTRG.3970	27.9731232	0.0004687	0.0194182	4.805969
RPA3860	MSTRG.4795	20.8912674	0.0004559	0.0194182	4.384828
goaT	MSTRG.3217	12.7735383	0.0004619	0.0194182	3.675086
RPA0331	MSTRG.503	0.0855734	0.0004521	0.0194182	-3.546695
RPA1154	MSTRG.1677	8.4374012	0.0004674	0.0194182	3.076799
nusA	MSTRG.739	0.1356441	0.0004768	0.0194182	-2.882102
RPA0586	MSTRG.896	4.1123181	0.0004489	0.0194182	2.039952
RPA4610	MSTRG.6345	448.6113015	0.0004845	0.0194535	8.809322
RPA4726	MSTRG.5986	0.0796984	0.0005143	0.0202290	-3.649306
RPA3471	MSTRG.4352	11.4728458	0.0005121	0.0202290	3.520151
RPA3323	MSTRG.4077	4.9113058	0.0005215	0.0203503	2.296107
RPA2005	MSTRG.2542	0.2140893	0.0005493	0.0210090	-2.223715
fer1	MSTRG.6364	180.4973218	0.0005743	0.0210203	7.495834
RPA2877	MSTRG.3761	0.0276143	0.0005721	0.0210203	-5.178441
RPA2876	MSTRG.3761	0.0276143	0.0005721	0.0210203	-5.178441
RPA4477	MSTRG.6182	0.0277053	0.0005871	0.0210203	-5.173696
RPA4640	MSTRG.6376	26.0889761	0.0005833	0.0210203	4.705368
RPA1656	MSTRG.2950	6.0610552	0.0005809	0.0210203	2.599569
livK	MSTRG.5023	5.0278871	0.0005791	0.0210203	2.329952
RPA2269	MSTRG.3156	13.9539435	0.0006013	0.0214031	3.802601
cheY3	MSTRG.2972	12.8841005	0.0006088	0.0215225	3.687520
RPA2898	MSTRG.3835	0.1861544	0.0006152	0.0215225	-2.425428
nifX	MSTRG.6351	429.2705429	0.0006642	0.0215670	8.745743
RPA1000	MSTRG.1246	41.0478183	0.0006320	0.0215670	5.359234

Table B.1. – Significantly differentially expressed genes, adjusted p-value < 0.05, Log_2 Fold change > 2 (Urea relative to $[NH_4]_2SO_4$) *(continued)*

Gene name	Transcript ID	Fold change	p-value	Adjusted p-value	Log2 fold change
RPA4621	MSTRG.6358	14.0823309	0.0006624	0.0215670	3.815814
RPA0105	MSTRG.81	10.6172726	0.0006539	0.0215670	3.408341
RPA3093	MSTRG.3617	9.7314897	0.0006225	0.0215670	3.282661
RPA1043	MSTRG.1180	0.1436433	0.0006564	0.0215670	-2.799437
RPA1928	MSTRG.2340	93.3653388	0.0006795	0.0217695	6.544815
RPA1927	MSTRG.2340	93.3653388	0.0006795	0.0217695	6.544815
cbbT2	MSTRG.973	0.1388132	0.0006791	0.0217695	-2.848784
RPA3912	MSTRG.4870	5.4426998	0.0006916	0.0219395	2.444323
uvrC	MSTRG.1698	0.2017912	0.0006938	0.0219395	-2.309065
RPA2800	MSTRG.3445	8.5856330	0.0007161	0.0222383	3.101924
RPA4167	MSTRG.5430	4.8700725	0.0007573	0.0223082	2.283943
nifA	MSTRG.6365	13.8023451	0.0007691	0.0225073	3.786841
pucAd	MSTRG.3738	177.8308877	0.0008080	0.0229761	7.474362
RPA3610	MSTRG.4415	27.9187286	0.0008025	0.0229761	4.803161
RPA0516	MSTRG.588	0.0477903	0.0007917	0.0229761	-4.387139
RPA0515	MSTRG.588	0.0477903	0.0007917	0.0229761	-4.387139
RPA3018	MSTRG.3743	7.1275743	0.0008115	0.0229761	2.833411
rphyB	MSTRG.3743	7.1275743	0.0008115	0.0229761	2.833411
RPA1204	MSTRG.1359	6.4988634	0.0008112	0.0229761	2.700187
RPA1410	MSTRG.1824	4.2068546	0.0007975	0.0229761	2.072742
RPA3395	MSTRG.4235	0.0937304	0.0008630	0.0235525	-3.415340
RPA3655	MSTRG.4496	9.3954533	0.0008814	0.0235525	3.231963
trxA	MSTRG.134	0.1455597	0.0008812	0.0235525	-2.780317
proB	MSTRG.227	0.1467784	0.0008785	0.0235525	-2.768288
RPA0196	MSTRG.365	0.1539050	0.0008540	0.0235525	-2.699888
RPA0823	MSTRG.1111	0.1818748	0.0008478	0.0235525	-2.458982
RPA4284	MSTRG.5885	0.1870587	0.0008861	0.0235525	-2.418437
RPA3700	MSTRG.4711	0.2200201	0.0008638	0.0235525	-2.184293
RPA0738	MSTRG.1511	16.8473465	0.0009269	0.0237763	4.074450
RPA3458	MSTRG.4336	11.0349722	0.0009278	0.0237763	3.464011
RPA3407	MSTRG.4251	6.5277550	0.0009208	0.0237763	2.706587
RPA3628	MSTRG.4580	0.1479156	0.0009454	0.0238494	-2.757154
RPA1198	MSTRG.1318	0.1612044	0.0009394	0.0238494	-2.633037
RPA0054	MSTRG.110	0.0837210	0.0009915	0.0242392	-3.578266
RPA0439	MSTRG.740	0.1815962	0.0010235	0.0247225	-2.461194
RPA0803	MSTRG.1077	0.2135563	0.0010294	0.0247657	-2.227312
RPA0802	MSTRG.1077	0.2135563	0.0010294	0.0247657	-2.227312
RPA2720	MSTRG.5793	0.1496907	0.0010508	0.0249322	-2.739944

Table B.1. – Significantly differentially expressed genes, adjusted p-value < 0.05, Log_2 Fold change > 2 (Urea relative to $[NH_4]_2SO_4$) *(continued)*

Gene name	Transcript ID	Fold change	p-value	Adjusted p-value	Log2 fold change
RPA3550	MSTRG.4440	4.9232205	0.0010698	0.0251135	2.299602
hpd	MSTRG.39	0.0050017	0.0010944	0.0251402	-7.643369
RPA1446	MSTRG.2012	12.4595590	0.0010780	0.0251402	3.639181
RPA1280	MSTRG.1913	10.4006982	0.0010885	0.0251402	3.378608
clpB	MSTRG.6129	0.1128683	0.0010997	0.0251639	-3.147287
RPA3011	MSTRG.3736	108.7043437	0.0011460	0.0258801	6.764266
nuoL3	MSTRG.3713	0.0846577	0.0011480	0.0258801	-3.562215
RPA2747	MSTRG.5821	7.7708475	0.0011563	0.0259057	2.958072
RPA4302	MSTRG.5910	16.0508903	0.0011631	0.0259334	4.004581
pucBd	MSTRG.3739	254.7243009	0.0012124	0.0263514	7.992793
clpP	MSTRG.3676	0.1051436	0.0012556	0.0263514	-3.249567
tesA	MSTRG.364	0.1282072	0.0012534	0.0263514	-2.963450
RPA3545	MSTRG.4379	6.7295048	0.0012526	0.0263514	2.750500
RPA4684	MSTRG.6433	6.5416482	0.0012236	0.0263514	2.709654
RPA3218	MSTRG.4013	5.3416649	0.0012748	0.0263514	2.417289
RPA3794	MSTRG.5003	5.2690102	0.0012378	0.0263514	2.397532
RPA2651	MSTRG.5709	4.7874546	0.0012026	0.0263514	2.259259
FarB	MSTRG.587	0.2226613	0.0012608	0.0263514	-2.167077
RPA3209	MSTRG.3998	4.1127271	0.0012394	0.0263514	2.040095
RPA2996	MSTRG.3714	0.0441695	0.0013421	0.0270545	-4.500805
RPA1218	MSTRG.1324	6.5317555	0.0013479	0.0270545	2.707471
RPA2125	MSTRG.2709	4.0718282	0.0013298	0.0270545	2.025677
hslU	MSTRG.473	0.1031353	0.0013871	0.0272868	-3.277390
RPA0729	MSTRG.1501	6.0475550	0.0013928	0.0272868	2.596352
RPA1763	MSTRG.2227	5.3469007	0.0013970	0.0272868	2.418703
RPA4348	MSTRG.5368	0.2256324	0.0013784	0.0272868	-2.147953
RPA2297	MSTRG.3187	17.2022462	0.0014121	0.0272868	4.104525
RPA3392	MSTRG.4231	0.1637037	0.0014042	0.0272868	-2.610841
RPA1238	MSTRG.1857	5.6315255	0.0014174	0.0273038	2.493526
hbaE	MSTRG.1423	25.9272686	0.0014337	0.0273789	4.696398
regR	MSTRG.872	0.1950382	0.0014580	0.0275616	-2.358171
RPA3460	MSTRG.4339	7.4177653	0.0014679	0.0275980	2.890985
RPA3867	MSTRG.4827	4.7579397	0.0014690	0.0275980	2.250337
RPA4639	MSTRG.6374	20.0114461	0.0014871	0.0276110	4.322754
RPA0743	MSTRG.1515	15.7513476	0.0015014	0.0276110	3.977403
coxL%2C cutL	MSTRG.6410	14.3747211	0.0014988	0.0276110	3.845462
RPA1258	MSTRG.1883	13.5283738	0.0014932	0.0276110	3.757916
RPA3733	MSTRG.4669	0.2404553	0.0014755	0.0276110	-2.056159

Table B.1. – Significantly differentially expressed genes, adjusted p-value < 0.05, Log_2 Fold change > 2 (Urea relative to $[NH_4]_2SO_4$) *(continued)*

Gene name	Transcript ID	Fold change	p-value	Adjusted p-value	Log2 fold change
RPA1686	MSTRG.2989	7.2782659	0.0015391	0.0282178	2.863595
RPA2300	MSTRG.3189	7.2823531	0.0015676	0.0284711	2.864405
RPA3994	MSTRG.5247	4.4926157	0.0015948	0.0287212	2.167556
RPA1956	MSTRG.2417	0.1840105	0.0016096	0.0289011	-2.442140
trx	MSTRG.263	0.2080742	0.0016277	0.0291401	-2.264830
RPA2869	MSTRG.3802	8.6842628	0.0016351	0.0291592	3.118403
bchP	MSTRG.1967	5.1772550	0.0016383	0.0291592	2.372187
RPA2098	MSTRG.2609	4.6454241	0.0016546	0.0291927	2.215810
RPA2477	MSTRG.3373	0.2392582	0.0016873	0.0293446	-2.063360
RPA0403	MSTRG.689	0.1340910	0.0017351	0.0299080	-2.898716
hmrR	MSTRG.4109	0.1988714	0.0017486	0.0299080	-2.330092
tRNA-Lys2	MSTRG.5713	0.2428130	0.0017596	0.0299080	-2.042082
RPA0534	MSTRG.615	10.8919074	0.0017858	0.0299456	3.445185
dnaK	MSTRG.507	0.0920337	0.0017975	0.0300141	-3.441694
RPA3014	MSTRG.3740	18.7636286	0.0018370	0.0303036	4.229867
RPA1261	MSTRG.1889	5.6015109	0.0019106	0.0310135	2.485816
RPA4683	MSTRG.6431	96.5555365	0.0019293	0.0311502	6.593287
RPA2997	MSTRG.3715	0.0294164	0.0019640	0.0312134	-5.087236
RPA3509	MSTRG.4534	0.2388798	0.0019609	0.0312134	-2.065644
RPA3511	MSTRG.4534	0.2388798	0.0019609	0.0312134	-2.065644
fixR2	MSTRG.6366	25.1239283	0.0019815	0.0313614	4.650990
RPA3378	MSTRG.4212	13.4836718	0.0019945	0.0314147	3.753141
RPA0229	MSTRG.404	7.8675223	0.0020024	0.0314147	2.975909
RPA2290	MSTRG.3178	4.1989130	0.0020372	0.0317282	2.070016
serA2	MSTRG.5917	0.1002302	0.0020758	0.0322160	-3.318611
RPA0767	MSTRG.648	4.1777961	0.0021002	0.0322830	2.062742
rho	MSTRG.304	0.0904605	0.0021200	0.0323416	-3.466568
RPA0886	MSTRG.1010	6.1918155	0.0021660	0.0328785	2.630362
RPA4322	MSTRG.5935	27.0868262	0.0022152	0.0330495	4.759519
RPA0465	MSTRG.779	13.0372862	0.0022122	0.0330495	3.704572
RPA0861	MSTRG.981	0.0872553	0.0022071	0.0330495	-3.518613
RPA0862	MSTRG.981	0.0872553	0.0022071	0.0330495	-3.518613
RPA1133	MSTRG.1654	4.0236603	0.0022367	0.0332070	2.008509
RPA2199	MSTRG.3084	0.1810021	0.0022635	0.0335227	-2.465922
RPA3597	MSTRG.4397	0.1961421	0.0023359	0.0341705	-2.350029
accA	MSTRG.576	0.2166948	0.0023345	0.0341705	-2.206263
RPA2140	MSTRG.2733	4.1338679	0.0023409	0.0341705	2.047492
RPA1995	MSTRG.2380	0.1585322	0.0023770	0.0344662	-2.657152

Table B.1. – Significantly differentially expressed genes, adjusted p-value < 0.05, Log_2 Fold change > 2 (Urea relative to $[NH_4]_2SO_4$) *(continued)*

Gene name	Transcript ID	Fold change	p-value	Adjusted p-value	Log2 fold change
RPA4068	MSTRG.5069	0.1749338	0.0023782	0.0344662	-2.515119
RPA0580	MSTRG.886	17.2438369	0.0024389	0.0344943	4.108009
cyc	MSTRG.1512	6.6456585	0.0024424	0.0344943	2.732412
RPA3288	MSTRG.4149	4.7636007	0.0024251	0.0344943	2.252053
RPA3031	MSTRG.3528	0.1219230	0.0025389	0.0353850	-3.035958
RPA4823	MSTRG.6104	7.9240426	0.0025306	0.0353850	2.986237
vnfX	MSTRG.1703	4.7661709	0.0025713	0.0356732	2.252831
vnfN	MSTRG.1703	4.7661709	0.0025713	0.0356732	2.252831
RPA2749	MSTRG.5825	4.7489990	0.0025843	0.0356732	2.247623
dapA1	MSTRG.3716	0.0789823	0.0026424	0.0359833	-3.662327
cbbL	MSTRG.2822	0.1120332	0.0026822	0.0363625	-3.158002
RPA3373	MSTRG.4203	0.0291760	0.0027378	0.0367228	-5.099076
uvrB	MSTRG.5327	0.1903467	0.0028147	0.0367228	-2.393299
RPA4807	MSTRG.6084	4.8187751	0.0027877	0.0367228	2.268667
mexF pseudogene	MSTRG.3744	0.2253633	0.0027492	0.0367228	-2.149676
RPA4445	MSTRG.6144	0.0673109	0.0028238	0.0367270	-3.893017
RPA0537	MSTRG.620	0.2366136	0.0028329	0.0367679	-2.079395
RPA4594	MSTRG.6327	4.0534929	0.0029453	0.0376609	2.019166
RPA4692	MSTRG.6448	6.3629411	0.0029776	0.0376806	2.669694
RPA4201	MSTRG.5476	0.0883483	0.0029889	0.0377457	-3.500654
flgF	MSTRG.4804	4.5448310	0.0030014	0.0377469	2.184227
RPA2377	MSTRG.3001	0.1260467	0.0030567	0.0381315	-2.987970
pucAc pseudogene	MSTRG.3734	119.1466609	0.0030968	0.0383444	6.896595
RPA1741	MSTRG.2195	19.9539954	0.0030844	0.0383444	4.318606
cheW2	MSTRG.2920	10.3846718	0.0031290	0.0385553	3.376384
cheA2	MSTRG.2920	10.3846718	0.0031290	0.0385553	3.376384
cheA3	MSTRG.2920	10.3846718	0.0031290	0.0385553	3.376384
RPA4622	MSTRG.6359	20.6765399	0.0031929	0.0391581	4.369923
RPA1160	MSTRG.1685	0.1550547	0.0031985	0.0391581	-2.689151
RPA3595	MSTRG.4396	0.2011366	0.0033051	0.0399172	-2.313752
RPA1278	MSTRG.1910	0.1903431	0.0033475	0.0401102	-2.393326
RPA2124	MSTRG.2707	11.3178762	0.0034604	0.0407518	3.500531
RPA2241	MSTRG.3125	0.1566976	0.0034613	0.0407518	-2.673945
RPA3908	MSTRG.4818	5.2713207	0.0034686	0.0407592	2.398165
RPA4526	MSTRG.6248	0.1901721	0.0034893	0.0409242	-2.394622
RPA1494	MSTRG.2079	5.9315148	0.0035399	0.0413573	2.568401
RPA0458	MSTRG.768	4.9265899	0.0035682	0.0415289	2.300589
RPA0459	MSTRG.768	4.9265899	0.0035682	0.0415289	2.300589

Table B.1. – Significantly differentially expressed genes, adjusted p-value < 0.05, Log_2 Fold change > 2 (Urea relative to $[NH_4]_2SO_4$) *(continued)*

Gene name	Transcript ID	Fold change	p-value	Adjusted p-value	Log2 fold change
hemA	MSTRG.1982	5.5752982	0.0037161	0.0424390	2.479049
RPA1553	MSTRG.1982	5.5752982	0.0037161	0.0424390	2.479049
RPA2543	MSTRG.5575	8.9571324	0.0037933	0.0429757	3.163037
RPA2542	MSTRG.5575	8.9571324	0.0037933	0.0429757	3.163037
RPA2541	MSTRG.5575	8.9571324	0.0037933	0.0429757	3.163037
pucBc	MSTRG.3733	93.4714368	0.0038157	0.0430112	6.546454
RPA4109	MSTRG.5146	0.1335279	0.0038757	0.0433660	-2.904787
RPA0793	MSTRG.1066	0.0746457	0.0039046	0.0436099	-3.743797
RPA1599	MSTRG.2876	0.0980276	0.0039429	0.0437477	-3.350669
RPA1800	MSTRG.2479	0.1742578	0.0039457	0.0437477	-2.520705
RPA0487	MSTRG.468	0.1896747	0.0039366	0.0437477	-2.398401
secA	MSTRG.580	0.2053905	0.0039403	0.0437477	-2.283558
RPA1236	MSTRG.1854	4.5737398	0.0039786	0.0439527	2.193374
RPA2004	MSTRG.2541	0.2287666	0.0039927	0.0440279	-2.128052
flgI	MSTRG.4868	4.6434082	0.0040301	0.0441210	2.215184
RPA3910	MSTRG.4868	4.6434082	0.0040301	0.0441210	2.215184
RPA1716	MSTRG.2095	8.2058790	0.0041066	0.0448777	3.036658
RPA0544	MSTRG.631	0.1601209	0.0041372	0.0449701	-2.642766
RPA0545	MSTRG.631	0.1601209	0.0041372	0.0449701	-2.642766
RPA3928	MSTRG.4892	4.2983279	0.0042454	0.0455313	2.103775
RPA0728	MSTRG.1500	5.3959966	0.0043234	0.0461016	2.431889
RPA1495	MSTRG.2080	5.1494828	0.0043099	0.0461016	2.364428
RPA2588	MSTRG.5639	0.2266695	0.0043246	0.0461016	-2.141338
RPA2132	MSTRG.2718	0.0885675	0.0046716	0.0484429	-3.497079
RPA0850	MSTRG.814	0.1696984	0.0047277	0.0484462	-2.558955
RPA2570	MSTRG.5606	6.5997119	0.0047764	0.0486997	2.722403
RPA0785	MSTRG.798	4.1074649	0.0048436	0.0491821	2.038248
prc	MSTRG.903	0.1398619	0.0048964	0.0493778	-2.837925
RPA4303	MSTRG.5912	0.2485736	0.0049078	0.0493778	-2.008255
RPA4328	MSTRG.5942	7.9120454	0.0049256	0.0494752	2.984051