

Role and regulation of chemotaxis and
motility in the *Rhizobium*-legume
symbiosis



Samuel Aroney
Brasenose
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Hilary 2021

Soli Deo gloria

Abstract

Fertilisers are used to provide crop plants with essential nutrients such as nitrogen but with a high energy cost and detrimental environmental effects. The symbiosis of legumes with rhizobia exhibits none of these problems since the bacteria provide fixed nitrogen directly to the plant. Further understanding of the initiation of symbiosis may enable improvements to the farming of crop legumes and enable the same processes to be mimicked in engineered systems of other crops. Although motility and chemotaxis are known to be crucial for competitive plant root colonisation (the first stage of symbiosis), little is known about their importance after this stage. In this thesis, I aimed to describe the importance of chemotaxis and motility in sequential stages of symbiosis up to the generation of root nodules filled with nitrogen-fixing rhizobia. To this end, the chemotaxis and motility systems of the model rhizobia *Rhizobium leguminosarum* biovar *viciae* 3841 (Rlv3841) were characterised, in both free-living and plant-associated states.

Rlv3841 motility is driven by flagella rotation modulated by two chemotaxis systems, Che1 and Che2, that respond to signals sensed by chemoreceptors. In this work we found that Rlv3841 has a diverse set of chemoreceptors that enable the bacteria to sense and respond to a variety of chemical gradients, by both externally sensing in the periplasm and internally sensing metabolic state. We also found a novel system that halts the bacteria in response to high concentrations of TCA cycle intermediates independent of the chemotaxis systems. An inorganic phosphate uptake system (PstSCAB) and a putative methyltransferase (encoded by RL3464) were found to be critical to the response, although their downstream effectors were not identified. Finally, we characterised the importance of the chemotaxis and flagella systems of Rlv3841 at sequential stages of the symbiosis. This involved marking bacteria with luminescence, fluorescence and enzymatic activity and assaying the bacteria by optical imaging, flow cytometry and staining. We found that there are several critical stages in which chemotaxis and motility are essential and that these occur after the plant root has been colonised by the rhizobia. In particular, we found that Che1 and flagella are important for movement around the root and to root hair tips and that Che2 is involved in infection thread formation and bacteroid differentiation. Overall, this work improves our understanding of the chemotaxis and motility systems of Rlv3841 and their importance across the whole *Rhizobium*-legume symbiosis cycle.

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Philip Poole. You have been a fantastic mentor and I am extremely grateful for the hours that you sunk into my project. I would also like to thank my second supervisor, Professor Gail Preston, for your insightful comments and also for your support throughout the DTP program. Thirdly, I would like to thank my Postdoc lead, Dr Carmen Sanchez-Canizares. Your help has been invaluable and your guidance has been greatly appreciated.

In addition, the whole Poole lab group has been incredibly welcoming and supportive. I especially want to thank Dr Andrzej Tkacz for your bioinformatic assistance. I will miss our coding discussions. My thanks also goes to Helen and Lida, without whom none of this work would have been possible.

In addition to the people, this work would not have been possible without funding. I would like to thank the Clarendon Fund for their funding support together with the Oxford-Australia Scholarship Fund, Brasenose College and the BBSRC via the Interdisciplinary Biosciences DTP. I also would like to thank Brasenose College and the Department of Plant Sciences for the communities that they both provided.

To all the people at St. Ebbe's, thank you for having made us feel very welcome in Oxford. You have all been fantastic and I greatly enjoyed our discussions, even though I tended to derail them. Thanks specifically to Rick and Sarah, David and Abigail, Sean, Jin and Jae, LaRae and Matt, Jordan and Kristin, Kate, Tuti, Peter and Dan. I cherish your friendships and I am sure that they will continue no matter where in the world we each end up.

I also want to thank the people without whom I never would have made it here. To my parents, thank you Mum and Dad for your patience and encouragement to make me the man that I am today. To my siblings, thank you Shell, Nat and Joe for moulding and honing me into a much sharper wit than I would otherwise be. To my parents in law, thank you Julie and Darren for your support and encouragement to finish quickly so I could bring your daughter back home.

Finally, to my darling wife Katie, thanks so much for always being a comfort to me, distracting me when I was stressed and encouraging me when I was at the end of my tether. This thesis would not have been started or finished without your support.

Contents

Abstract	v
Acknowledgements	vii
List of Figures	xv
List of Abbreviations	xvii
1 Introduction	1
1.1 Nitrogen Cycle	1
1.2 <i>Rhizobium</i> -legume symbiosis	2
1.2.1 Soil and the rhizosphere	2
1.2.2 Colonisation of root surface and root hairs	3
1.2.3 Entering the plant via infection threads	4
1.2.4 The symbiotic core: the nodule	5
1.3 Importance of chemotaxis and motility in symbiosis	6
1.3.1 Screening for gene significance in symbiosis	8
1.4 Motility in Gram-negative bacteria	8
1.4.1 The <i>Escherichia coli</i> motility model	9
1.4.2 The <i>Sinorhizobium meliloti</i> motility model	9
1.4.3 The <i>Rhizobium leguminosarum</i> biovar <i>viciae</i> motility model	11
1.5 Chemotaxis in symbiotic bacteria	12
1.6 Regulation of motility and chemotaxis	18
1.7 Energy taxis in bacteria	18
1.8 Chemotaxis and motility interaction with the phosphotransferase system (PTS) in Rlv3841	20
1.9 Aerotaxis and the electron transport chain	22
1.10 Research aims	23
2 General materials and methodology	27
2.1 Bacterial strains, plasmids and primers	27
2.1.1 Strains	27
2.1.2 Plasmids	32

2.1.3	Primers	34
2.2	Chemicals and kits	37
2.3	General molecular biology techniques	37
2.3.1	DNA isolation	37
2.3.2	DNA concentration	38
2.3.3	PCR amplification	38
2.3.4	DNA gel electrophoresis	38
2.3.5	General DNA sequencing	38
2.4	Cloning	39
2.4.1	Restriction digests and DNA ligation	39
2.4.2	InFusion HD cloning	39
2.4.3	NEBuilder Hi-Fidelity cloning	39
2.4.4	Transformation	39
2.4.5	Tri-parental conjugation from <i>E. coli</i> into Rlv3841	40
2.4.6	Di-parental conjugation from <i>E. coli</i> into Rlv3841	40
2.5	Mutagenesis techniques	41
2.5.1	Tn5 transduction	41
2.5.2	Double cross-over mutagenesis by pK18mobSacB	42
2.6	Media	42
2.6.1	Media recipes	43
2.6.2	Antibiotic working concentrations	43
2.7	Bacterial growth conditions and assays	44
2.7.1	Bacterial growth conditions	44
2.7.2	Generation time of rhizobial strains	44
2.7.3	Semi-solid agar swimming assay	44
2.7.4	Single cell tracking	45
2.7.5	Flagella staining	46
2.8	Legume growth conditions	47
2.8.1	Germination of <i>Pisum sativum</i>	47
2.8.2	Germination of <i>Vicia villosa</i>	47
2.8.3	<i>Pisum sativum</i> growth conditions	47
2.8.4	<i>Vicia villosa</i> growth conditions	48
2.8.5	Recipe for Nitrogen-free rooting solution	48
2.8.6	Plant stain for enzyme activity	48
2.9	Computational analysis	49
2.9.1	<i>In silico</i> cloning	49
2.9.2	Bioinformatic analysis	49
2.9.3	Data-handling and statistical analysis	50

3	A model of sensory taxis in <i>Rhizobium leguminosarum</i>	51
3.1	Introduction	51
3.2	Starvation response in Rlv3841 vs Sme2011	52
3.3	Isolation and characterisation of motility mutants	54
3.4	Motility of chemotaxis-gutted mutants to form a negative control for the absence of chemotaxis	55
3.5	Chemotaxis-gutted mutant still responds to succinate	59
3.6	The chemoreceptors of <i>Rhizobium leguminosarum</i> biovar <i>viciae</i> 3841	63
3.6.1	MCP heptad classes	64
3.6.2	Methylation	65
3.7	Rlv3841 chemoreceptor classification	67
3.7.1	Short-chain carboxylate sensing	68
3.7.2	Oxygen, redox or light sensors	68
3.7.3	Other haem binding proteins	69
3.7.4	Quaternary ammonium compound sensing	70
3.7.5	Plant related sensors	71
3.7.6	Putative organic acid sensing	72
3.7.7	Four-helix bundle sensor domains	73
3.7.8	Dual Cache sensor domains	73
3.7.9	Single Cache sensor domains	75
3.7.10	CHASE sensor domain	75
3.7.11	Che2-linked relatives	76
3.7.12	>200 bp unknown sensor domains	76
3.7.13	<200 bp unknown sensor domains	77
3.7.14	No sensor domains	79
3.8	Testing of internal chemoreceptors as candidates for starvation response	79
3.9	Discussion	86
4	Regulation of motility under starvation conditions	91
4.1	Introduction	91
4.2	Characterisation of chemotaxis-independent but metabolism-based taxis	93
4.3	The TCA-cycle intermediate halting effect is specific to <i>R. leguminosarum</i>	96
4.4	Testing of PTS as a candidate for the halting response	96
4.5	Mutant isolation screening for TCA-intermediate swimming control	101
4.5.1	Swimming responses of the two classes of mutation	108
4.6	Discussion	114
4.6.1	The phosphate transporter PstSCAB	118

4.6.2	Methyltransferase	120
4.7	Methods	123
4.7.1	Random mutagenesis screening	123
4.7.2	Whole genome sequencing	123
4.7.3	SNP analysis	123
4.7.4	Agar washing	124
5	Chemotaxis and motility in symbiosis	125
5.1	Introduction	125
5.2	Analysis of chemotaxis and motility genes in insertion sequencing data	126
5.3	Generation of marked strains to monitor rhizobia across different symbiotic stages	129
5.4	Chemotaxis mutants but not motility mutants are impaired in initial attachment	130
5.5	Motility and chemotaxis are important for root colonisation	132
5.6	Chemotaxis is involved in symbiosis initiation	134
5.7	Motility and chemotaxis are important for nodulation	137
5.8	Chemotaxis is important for nodule development	139
5.9	The <i>che2</i> cluster is important for nodule maturation	141
5.10	Discussion	142
5.10.1	The <i>che1</i> cluster, in coordination with flagella, appears to have a role in attraction to both the root surface and to root hair tips	143
5.10.2	The <i>che2</i> cluster has a role in nodule maturation	148
5.11	Methods	150
5.11.1	INSeq assay	150
5.11.2	Attachment assay	151
5.11.3	Colonisation assay	151
5.11.4	Nodule competition assay	152
5.11.5	Nodule development assay	152
5.11.6	Nodule harvest and bacteroid isolation	153
5.11.7	Flow cytometer measurement	153
6	Conclusions and future perspectives	155
6.1	Impact of the 2020 Pandemic	155
6.2	General overview	155
6.3	The multiple sensory systems of Rlv3841	156
6.4	TCA cycle intermediate halting effect	158
6.5	Chemotaxis and motility in symbiosis	161
6.5.1	The Che1 cluster is important in early stages of symbiosis .	162
6.5.2	The Che2 cluster is important for nodule development . . .	163
6.6	Future directions	164

<i>Contents</i>	<i>xiii</i>
A INSeq results	175
References	181

List of Figures

1.1	Rhizobium-legume symbiosis cycle	3
1.2	Symbiotic infection process	4
1.3	Escherichia coli motility model	10
1.4	Rlv3841 motility model	12
1.5	The main motility clusters of Rlv3841	13
1.6	The T4 Tad pili operon structure of Rlv3841	13
1.7	Typical chemoreceptor topology	16
1.8	*Escherichia coli* Phosphotransferase system	20
1.9	Rlv3841 Phosphotransferase system	21
1.10	The energy metabolism and respiratory chain of nodulating symbionts	24
2.1	Bacterium swimming tracking	46
3.1	Wildtype rhizobia swimming plates	53
3.2	Motility mutant swimming plates	56
3.3	Chemotactic mutant swimming plates	57
3.4	Photos of chemotactic mutant swimming plates	58
3.5	Wildtype and chemotaxis mutant swimming tracks	61
3.6	Maximum likelihood phylogenetic tree of chemoreceptors	65
3.7	McpN domain structure	68
3.8	McpQ and McpY2 domain structure	69
3.9	IcpA and HemAT domain structure	71
3.10	McpX domain structure	72
3.11	McpT domain structure	72
3.12	McpB domain structure	73
3.13	McpC, McpG and McpL domain structure	74
3.14	McpM and McpO domain structure	74
3.15	McpJ and McpS domain structure	75
3.16	McpK domain structure	76
3.17	McrA, McrB and McrC domain structure	77
3.18	McpI, McpP and McpZ domain structure	78
3.19	McpD, McpE and McpR domain structure	78

3.20	RL1386 and RL1447 domain structure	79
3.21	Carbon source testing of internal chemoreceptor mutants	81
3.22	Nitrogen source testing of internal chemoreceptor mutants	82
3.23	Light testing of internal chemoreceptor mutants	83
3.24	Oxygen environment testing of internal chemoreceptor mutants	85
4.1	Swimming plate gradients	95
4.2	Glucose swimming tracks	97
4.3	Pyruvate swimming tracks	98
4.4	Succinate swimming tracks	99
4.5	Broad-range bacterial succinate responses	100
4.6	manX mutant plates	101
4.7	ptsN mutant plates	102
4.8	Papillation isolate mutations	106
4.9	Papillation isolates growth rate	107
4.10	Rlv3841 reacts to nutrient concentrations	108
4.11	Papillation isolates on swimming plates	111
4.12	Papillation isolate swimming tracks	112
4.13	Papillation isolate complementation on swimming plates	114
4.14	Photos of papillation isolate complementation	115
4.15	Papillation isolate complementation swimming tracks	116
4.16	PhoU regulation of PstSCAB transporter and PhoRB two-component system	119
4.17	Neighbour-joining consensus tree of RL3464 methyltransferase related sequences	122
5.1	Rlv3841 chemotaxis and motility gene requirement at different sym- biotic stages	127
5.2	Attachment phenotypes for chemotaxis and motility mutants	131
5.3	Cell counts from pea root colonisation experiments	133
5.4	Colonisation assays in competition against the wildtype	135
5.5	Nodulation dynamics assay	136
5.6	Competition assays on nodulation for chemotaxis and motility mutants	138
5.7	Nodulation maturity	140
5.8	Undifferentiated bacteria and bacteroid counts per nodule	142
5.9	Overview of experiments	143
5.10	Flow cytometer examples	154
6.1	Overview of chemotaxis and motility in symbiosis	162

List of Abbreviations

A	Adenine
ABC	ATP-binding cassette
Ala	Alanine
Amp	Ampicillin
ANOVA	Analysis of variance
Arg	Arginine
BLAST	Basic local alignment search tool
BNF	Biological nitrogen fixation
bp	Base pairs
C	Cytosine
c-di-GMP	Cyclic di-GMP
cps	Counts per second
Cys	Cysteine
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
dpi	Days post inoculation
DWA	Distilled water agar
EPS	Exopolysaccharide
ETC	Electron transport chain
FAD	Flavin adenine dinucleotide
FP2	<i>Azospirillum brasilense</i> FP2
G	Guanine
GFP	Green fluorescent protein
Glu	Glutamate
Gln	Glutamine
Gly	Glycine
Gm	Gentamicin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
HMM	Hidden Markov model
INSeq	Insertion sequencing
Km	Kanamycin
Kos	<i>Kosakonia radicincitans</i> DSM 16656
LB	Luria-Bertani broth
Leu	Leucine

Magenta-GlcA	5-bromo-6-chloro-3-indolyl- β -D-glucuronide
MCP	Methyl-accepting chemotaxis protein
MOPS	3-(N-morpholino)propanesulfonic acid
NAD	Nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
Neo	Neomycin
Nit	Nitrofurantoin
nt	Nucleotides
OD	Optical density
ORS571	<i>Azorhizobium caulinodans</i> ORS571
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Pfam	Protein family
Pro	Proline
PTS	Phosphotransferase system
r	Resistant
RCR2011	<i>Sinorhizobium meliloti</i> RCR2011
Rlp4292	<i>Rhizobium leguminosarum</i> biovar <i>phaseoli</i> 4292
Rlv3841	<i>Rhizobium leguminosarum</i> biovar <i>viciae</i> 3841
RlvVF39	<i>Rhizobium leguminosarum</i> biovar <i>viciae</i> VF39
RNA	Ribonucleic acid
SBW25	<i>Pseudomonas fluorescens</i> SBW25
SD	Standard deviation
Ser	Serine
Sme1021	<i>Sinorhizobium meliloti</i> 1021
Sme2011	<i>Sinorhizobium meliloti</i> Rm2011 str3
Spec	Spectinomycin
Str	Streptomycin
T	Thymine
TCA	Tricarboxylic acid
Tet	Tetracycline
Thr	Threonine
TLP	Transducer-like protein
TY	Tryptone yeast
UMS	Universal minimal salts
wt	Wildtype
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside

Introduction

1.1 Nitrogen Cycle

Nitrogen is an essential element for plant growth and reproduction, indeed for all life. Plants typically get nitrogen from ammonium and nitrate in the soil, which are in turn sourced from decaying organic matter or nitrogen fixation (Jackson, Schimel and Firestone 1989). Sourcing nitrogen becomes a problem in a crop context: even small, diversely planted farms drain the nitrogen from the soil, let alone the fast-growing, nutrient-exhausting monocultures of the contemporary farm. The modern solution is to apply fertilisers to replenish the nitrogen. However, most fertiliser nitrogen does not reach the crop plants, instead leaching into surrounding water-systems and causing eutrophication and other environmental problems (Cheng 2008; Schlesinger 2009; Smil 2011). In addition, nitrogen fertiliser is energetically expensive, accounting for 1% of the worlds energy supply and requiring 6 times as much energy as phosphorous or potassium fertilisers (Santi, Bogusz and Franche 2013; Oldroyd and Dixon 2014; Dent and Cocking 2017).

An ancient solution—crop rotation—has none of the above problems. The reinvigoration of the soil by the planting of leguminous plants, was noted as early as Theophrastus (370-285 B.C.) (Theophrastus and Hort 2015). We now know that legume plants reinvigorate the soil by the addition of fixed nitrogen obtained through

symbiosis with a group of bacteria called rhizobia. Rhizobia can fix atmospheric nitrogen and perform this task in return for carbon and a safe niche within the legume. Despite atmospheric nitrogen being an almost unlimited resource (78% of the atmosphere is dinitrogen gas), this fixation process is unique to microbes and is known as Biological Nitrogen Fixation (BNF). BNF accounts for an estimated 40-50% of nitrogen in the biosphere (Schlesinger 2009; Terpolilli, Hood and Poole 2012).

Although the symbiosis allows legumes to survive in low-nitrogen conditions and provides a competitive advantage in highly biodiverse environments, it is not found in non-legume crops (Oldroyd *et al.* 2011; Geddes *et al.* 2015). Efforts have therefore been made to engineer cereal crops with BNF, either through developing bacterial symbioses or transferring the fixation machinery to plant organelles (Geddes *et al.* 2015). This thesis aims to further our understanding of the initiation of the *Rhizobium*-legume symbiosis to allow mimicry in engineered systems.

1.2 *Rhizobium*-legume symbiosis

1.2.1 Soil and the rhizosphere

Soil is a highly competitive environment for microbial life, with as many as 10^4 bacterial species and 10^9 bacterial cells per gram (Poole, Ramachandran and Terpolilli 2018). The soil surrounding plant roots, the rhizosphere, has even higher levels of colonisation. The rhizosphere microbes compete for the 20% of plant photosynthate that is exuded into the soil and, in return, the microbes can increase micro- and macro-nutrient availability (Brencic and Winans 2005; Richardson *et al.* 2009). It is from these colonised microbes that legumes select endophytic symbionts to begin the symbiosis cycle (Figure 1.1). To fulfil this function, the symbiotic rhizobia must have two distinct traits: they must be independently competitive colonisers as free-living cells in the soil and they must be efficient nitrogen fixers as symbiotic cells, called bacteroids, inside the plant. Since these roles have very different objectives, rhizobia tend to have different metabolic modes and regulatory pathways, first as free-living status in the soil and, secondly, as

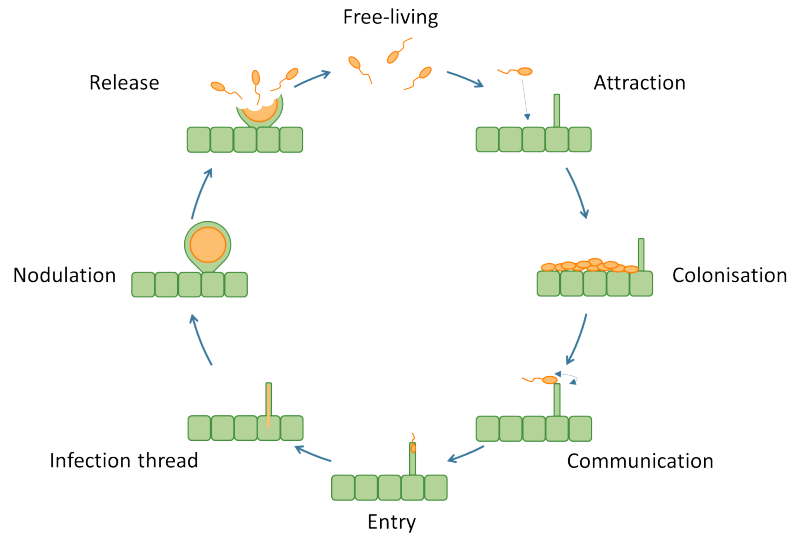


Figure 1.1: *Rhizobium*-legume symbiosis cycle. Rhizobia are initially living in the soil (free-living), they then move towards the plant root surface through chemotaxis (attraction), where they colonise (colonisation). After reaching the root hair tips, the rhizobia engage in a molecular dialogue with the plant (communication) which, if successful, enables their entrance into the root hair (entry). The plant responds to bacterial entry by forming infection threads that grow and ramify into plant cortical cells. Bacteria move down the infection threads and are eventually endocytosed by plant cells where they differentiate into nitrogen-fixing bacteroids (nodulation). Eventually, the nodule undergoes senescence, releasing the undifferentiated nodule bacteria back into the soil (release).

bacteroids within the plant. This differentiation occurs through the complex process by which symbiosis is initiated.

1.2.2 Colonisation of root surface and root hairs

The symbiosis is first established by the migration of rhizobia through the soil towards the host plant roots. This movement is controlled by chemotaxis and appears to be host-agnostic, driving movement to all plant exudates tested (Gaworzewska and Carlile 1982). The process also drives higher colonisation in areas of high exudation, such as branching points and new growth (Brencic and Winans 2005). After arriving at the root, the initial stage of attachment is characterised by non-specific, reversible binding of bacterial surface polysaccharides by plant lectins (Rodriguez-Navarro, Dardanelli and Ruiz-Sainz 2007; Berne *et al.* 2015; Wheatley

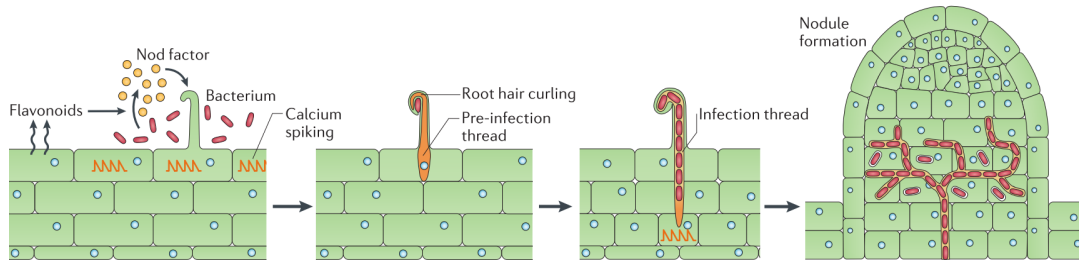


Figure 1.2: Symbiotic infection process. The plant root releases flavonoids which induce the production of nodulation factors (Nod factor) in symbiotic rhizobia. This leads to curling of the root hair to trap attached rhizobia and the formation of the infection thread tube. The bacteria then fill the infection thread by growth and division. Finally, after the rhizobia reach the inner cortex cells, they enter inside the plant cells and differentiate into bacteroids to begin fixing nitrogen. These plant cells expand and replicate to form the nodule. Reproduced from Oldroyd (2013).

and Poole 2018). Some degree of specificity comes with the interactions in the second stage of attachment, with the bacteria producing cellulose fibrils for strong, irreversible binding as a biofilm (Rodriguez-Navarro, Dardanelli and Ruiz-Sainz 2007; Wheatley and Poole 2018). The main source of host-specificity comes from the interactions in late colonisation, at root hair caps of bacterial biofilms (Gage 2004; Downie 2010). Here, the legume produces flavonoids that induce the production of signalling molecules (Nod factors) by the symbiotic bacteria (Figure 1.2) (Fisher and Long 1992; Oldroyd 2013; Poole, Ramachandran and Terpolilli 2018). It is this molecular dialogue that induces the later stages of the symbiosis.

1.2.3 Entering the plant via infection threads

The later stages of symbiosis occur deep within the plant root, so the bacteria must first enter inside. The production of Nod factors mentioned above induces the plant root hairs to curl around the rhizobial biofilm, trapping the bacteria and forcing them inside the root (Figure 1.2) (Prell and Poole 2006; Oldroyd and Downie 2008). Once inside, the Nod factors have the additional effect of dampening the plant immune defences to allow the bacteria to survive within the root (Zipfel and Oldroyd 2017). The rhizobia then progress further into the plant root via structures called infection threads, tubular extrusions formed by invagination of plant cell

walls that provide an avenue through each plant cell (Prell and Poole 2006; Jones *et al.* 2007). The bacteria movement through the infection threads is mainly due to continual growth and division, filling the available space (Gage 2004). However, the filling of the infection threads is not continuous, with bacteria reaching the tip of the thread before the tube is filled (Fournier *et al.* 2008). This suggests that some other form of motility is involved, although the form that this takes is currently unknown. The infection threads become more highly branched as they progress deeper into the root, eventually reaching inner cortical cells that have been prepared by the plant to receive the rhizobia (Monahan-Giovanelli, Pinedo and Gage 2006).

1.2.4 The symbiotic core: the nodule

In response to the rhizobial infection, the plant expands inner root cells to contain the symbionts within budded membranes called symbiosomes (Poole, Ramachandran and Terpolilli 2018). The expansion collectively forms bulbous growths on the root, called nodules. Within these nodules, the bacteria differentiate into bacteroids—bacteria metabolically transformed to perform high levels of nitrogen-fixation. The bacteroid differentiation occurs following two different strategies according to the host plant. In determinate nodules (typical of tropical legumes, such as beans), the bacteroids have increased levels of nitrogen fixation machinery, but remain similar in size and viability (Mergaert *et al.* 2006). In indeterminate nodules (typical of temperate legumes, such as peas), the bacteroids change more drastically, leading to a terminal differentiation by modifying their cell walls to grow much larger and branched and becoming unable to divide (Mergaert *et al.* 2006; Prell *et al.* 2009). In addition, indeterminate nodules have a meristem with undifferentiated bacterial cells that continues to grow and produce more plant cells to renew the same rhizobial infection. In both cases, the nodule forms an extremely low-oxygen environment, using high levels of leghaemoglobin to bind free oxygen and prevent it from inactivating the nitrogenase (Haag *et al.* 2013). This enables the bacteroids to continually produce ammonium by nitrogen fixation, being supplied with nutrients from the plant (Prell and Poole 2006; Oldroyd *et al.* 2011). At the completion

of the symbiosis, the bacteria are released back into the rhizosphere, with the clonal growth in the infection threads providing a substantial population boost (Oldroyd and Downie 2008).

Rhizobium leguminosarum biovar *viciae* 3841 (Rlv3841) is an α -proteobacteria that is a well-established model for symbiotic nitrogen fixation, nodulating a number of legume species, particularly pea (*Pisum sativum*), vetches (various *Vicia*), broad bean (*Vicia faba*), lentils (various *Lens*), grass peas and sweet peas (various *Lathyrus*) (Johnston and Beringer 1975; Poole *et al.* 1994). This bacterium is the reference strain in our laboratory and will constitute the focus of this research project.

1.3 Importance of chemotaxis and motility in symbiosis

Motility and chemotaxis are key processes in the establishment of a successful symbiosis. Indeed, most symbionts of eukaryotes have flagellar and chemotaxis capabilities (Raina *et al.* 2019). Both are also needed for competitive nodulation (Catlow, Glenn and Dilworth 1990; Yost, Rochepeau and Hynes 1998; Miller *et al.* 2007), and plant root surface colonization (Scharf, Hynes and Alexandre 2016). In the first stage of competition as free-living cells in the soil, motility and chemotaxis are critical due to the highly heterogeneous nature of this environment. The heterogeneity is due to point sources of nutrients exuded from eukaryotes and is stabilised by low levels of soil mixing. In such environments a more motile bacteria can outgrow a metabolically superior species (Lauffenburger, Aris and Keller 1982).

Motility and chemotaxis are also crucial in the second stage, root surface colonization. A bacterium can achieve a growth advantage by arriving and establishing itself at the root surface before a competitive environment has formed. In addition, flagella-based swarming motility enables movement along the surface of the root for more effective colonisation (Simons *et al.* 1996; Gao *et al.* 2016). Indeed, motility and chemotaxis mutants of the soil bacterium *Bacillus subtilis* display greatly reduced colonisation on *Arabidopsis thaliana* roots after 4 hours and on tomato roots after 2 weeks (Allard-massicotte *et al.* 2016; Gao *et al.* 2016). In *Rhizobium*

leguminosarum biovar *viciae* VF39SM, deletion of the chemotaxis receptors *mcpB* or *mcpC* resulted in a significant reduction of nodulation competitiveness of pea plants, although the ability to nodulate was not affected (Yost, Rochepeau and Hynes 1998). Importantly, VF39SM *mcpC* mutants do not have a competitiveness disadvantage in symbiosis with other plants, indicating the plant specificity of chemotaxis (Scharf, Hynes and Alexandre 2016). In the case of *Sinorhizobium meliloti*, a symbiont that nodulates alfalfa (*Medicago sativa*), several chemoreceptors have been found to target specific compounds in the host seed exudates (Webb, Helm and Scharf 2016; Webb *et al.* 2017a, 2017b). Once inside the plant root, *S. meliloti* motility genes still appear to be expressed in the infection thread channel but are completely down-regulated in the nodules (Yost 1998). Interestingly, in other eukaryote-microbe symbioses, similar channels are generated to select for and guide their symbionts, via chemotaxis to generated gradients, to the symbiosis space (e.g. the slime cavities of Hornwort and the ducts of the bobtail squid) (Raina *et al.* 2019). However, once inside the nodule, motility and chemotaxis are unnecessary for the now sessile bacteroids. This is also observed in Rlv3841, with down-regulation of chemoreceptors during bacteroid differentiation to undetectable amounts in bacteroids (Yost *et al.* 2004; Karunakaran *et al.* 2009; Tambalo *et al.* 2010b). This regulation appears unrelated to the low oxygen conditions and organic acid availability inside the nodule (Yost *et al.* 2004; Tambalo *et al.* 2010b; Scharf, Hynes and Alexandre 2016).

The importance of chemotaxis to rhizobia is also revealed by the count of chemoreceptor genes. Analysis of 264 completed bacterial genomes gave an average chemoreceptor gene count of 14, which was found to be increased in species with high metabolic diversity, low stability of habitat and those with interactions with other living species (Lacal *et al.* 2010b). The category with the highest number of chemoreceptors included several plant-interacting bacteria, including *Rhizobium leguminosarum*, Rlv3841 having 27 candidate genes, almost double the average count. In contrast, *Escherichia coli*, a human gut microbe, has only 5 chemoreceptor genes.

1.3.1 Screening for gene significance in symbiosis

Insertion sequencing (INSeq) is a genome-wide screening technique able to identify the relative importance of genes and intergenic regions in a given bacterial growth condition (Perry and Yost 2014; Perry, Akter and Yost 2016). A mariner-based transposon vector is used to generate a library of bacterial mutants, each with a single transposon insertion site. This library is then passed through various environmental conditions, sequencing the strains that survive. The relative abundance of a mutant with a particular transposon insertion before and after the screen provides the relative fitness of the mutant with that particular region disrupted.

In a laboratory collaborative project, we applied this technique to the various stages in Rlv3841 symbiosis with pea plant (Wheatley *et al.* 2020). The chemotaxis and motility genes formed two interesting patterns. Strains that had insertions disrupting *che1* cluster and flagellar genes had reduced fitness in nodules but not in the rhizosphere or root surface environments. The lack of importance in the rhizosphere and root surface is probably due to the experimental conditions flooding the environments with a large number of cells. There is thus another required stage after root colonisation—either movement to root hair tips or down infection threads. Whereas those with insertions disrupting *che2* cluster genes had increased fitness in undifferentiated cells recovered from nodules, but not for bacteroids. This may be due to either an increase in the number of cells reaching the nodule (meaning *che2* genes disrupt this process) or a reduction in the number of cells differentiating into bacteroids (meaning *che2* genes are involved in this process). Further interpretations of these results are discussed in Chapter 5.

1.4 Motility in Gram-negative bacteria

There are several strategies bacteria use to navigate their environment. The fastest movement in liquid media is swimming via flagella (Bernab  u-Roda *et al.* 2015). Flagella are rigid helical structures, anchored to the cell wall and rotated by a protein motor to produce thrust (Macnab and Aizawa 1984). Among Gram-negative

bacteria, *Escherichia coli* is the best understood model of flagella-based swimming. Also belonging to this group, *S. meliloti* constitutes a very well-studied example in the case of nitrogen-symbiotic bacteria.

1.4.1 The *Escherichia coli* motility model

E. coli have many flagella that protrude from various points around the cell body (i.e. peritrichous flagellation) but are aimed in the same direction to form an integrated bundle. Each flagellum consists of a basal body, hook and filament. The basal body is made up of a central hollow rod surrounded by anchoring stacks of rings (Wang *et al.* 2012). At the base of the basal body is the motor, where torque is produced via the proton-motive force (Armitage and Schmitt 1997; Kojima and Blair 2001). This motor is further subdivided into the rotor, the proteins that rotate with the flagella, and the stator, the proteins that drive that rotation. The second structure, the hook, is a highly curved helix, a flexible coupling that connects the central rod to the filament. The filament is a left-handed helix consisting of a single protein, flagellin. When rotated in a counter clockwise (CCW) manner, the flagella provide a powerful forward thrust (Figure 1.3). However, if the rotary direction is flipped to a clockwise (CW) direction, a polymorphic change is induced that causes a right-angled bend in the filament to propagate along the flagella. This compels the bundle of flagella to separate, causing a period of rotational movement due to the low impact of inertia at microbial scales. This process is called a tumble and allows bacteria to change their direction of movement (Figure 1.3) (Macnab and Aizawa 1984; Armitage and Schmitt 1997; Berg 2003; Stock, Namba and Lee 2012; Wang *et al.* 2012). The ‘run and tumble’ mechanism produces movement called a biased random walk, the rate of tumbles being random but biased by the chemotaxis system, as described below (Section 1.5).

1.4.2 The *Sinorhizobium meliloti* motility model

Sinorhizobium meliloti is another well-studied Gram-negative swimming bacteria. Although its flagella are built in a similar fashion to those in *E. coli* and in similar

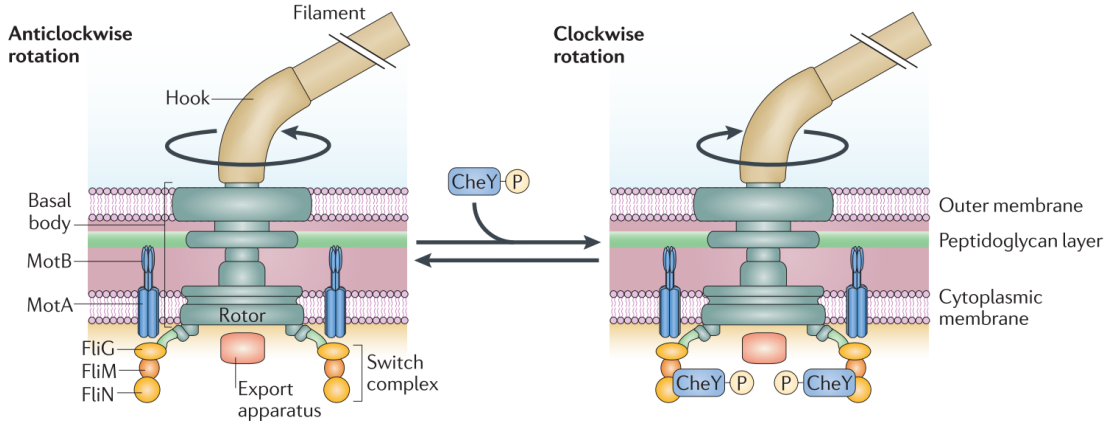


Figure 1.3: Escherichia coli motility model. In *E. coli*, there is an equilibrium between flagella anticlockwise and clockwise rotation modes. In anticlockwise mode, the bacteria swims straight and the motor turns freely. In clockwise rotation, the bacteria stalls and changes bearing. This mode is induced by binding of phosphorylated CheY to the switch complex. Reproduced from Porter, Wadhams and Armitage (2011).

numbers, there are several important differences. The main difference is the shape of the filament; *S. meliloti*'s filament is formed by right-handed helices and as such forward thrust is produced by CW rotation (Armitage and Schmitt 1997; Sourjik and Schmitt 1998). It also has increased rigidity, likely due to the use of several distinct flagellin proteins. Half of the flagella mass consists of the FlaA protein which forms heterodimers with three other proteins: FlaB, FlaC and FlaD (Scharf *et al.* 2001). The rigid *S. meliloti* flagella may be adapted for use in the relatively viscous medium of soil, compared to the native habitat of *E. coli*, the human gut (Armitage and Schmitt 1997; Attmannspacher, Scharf and Schmitt 2005).

In addition, the structure is unable to switch the rotary direction of the motor, instead modulating the speed of rotation. This is controlled through an extra stator protein, MotC, that binds MotB in the periplasm and may thus regulate the speed of the motor. An additional protein, MotE, acts to stabilize MotC and may target the protein to the flagella motor (Eggenhofer, Haslbeck and Scharf 2004).

Despite being unable to cause a tumble in the same way as *E. coli*, *S. meliloti* can still tumble. It does this by asynchronously modulating the rotational speed of its flagella, causing its bundle of flagella to disassociate, causing a tumble (Armitage and Schmitt 1997; Attmannspacher, Scharf and Schmitt 2005). The ability to

modulate the speed of its flagella also enables the bacteria to increase cell velocity in addition to reducing tumbling in response to high attractant concentration (Meier, Muschler and Scharf 2007).

1.4.3 The *Rhizobium leguminosarum* biovar *viciae* motility model

In the case of the target of this work, *Rhizobium leguminosarum*, there are still many open questions. Motility in *R. leguminosarum* appears to be closely related to that in *S. meliloti*, containing all the extra accessory genes (see Figure 1.4 and 1.5). Indeed, the bacteria have been shown to use modulation in flagella rotational speed to cause directional changes in addition to being unable to change the direction of rotation (Miller *et al.* 2007). However, *R. leguminosarum* appear to only have 1-2 subpolar flagella with a smooth surface, lacking helical perturbations found in *S. meliloti* (Tambalo *et al.* 2010a). The protein structure of the *R. leguminosarum* filament is closely related to *S. meliloti*, again consisting of heterologous pairs of a main flagellin (FlaA) and multiple secondary flagellin (FlaB, FlaC, FlaE and FlaH) (Scharf *et al.* 2001; Tambalo *et al.* 2010a).

The mechanism that *R. leguminosarum* uses to change direction through modulation of the rotary speed of a single or pair of flagella is currently unknown. Despite this, this rhizobial species does appear to tumble, approximately every 3 seconds under homogeneous conditions while swimming at approximately 38 $\mu\text{m/s}$ (Miller *et al.* 2007). As with many other bacteria, the rate of these tumbles is driven by the chemotaxis system.

In addition to flagella, Rlv3841 encodes a type IV pilus (Krehenbrink and Downie 2008). The Rlv3841 pili cluster displays high homology to the *tad* systems of *Sinorhizobium meliloti* and *Caulobacter crescentus* (See 1.6) (Tomich, Planet and Figurski 2007; Clock *et al.* 2008; Imam *et al.* 2011). Pili drive twitching motility across surfaces by extending, adhering and retracting to pull the bacterium forward (Mattick 2002).

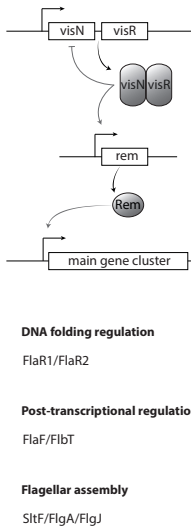


Figure 1.4: Rlv3841 motility model. The Rlv3841 flagella provides forward thrust through clockwise rotation through the exploitation of proton flux by the motor proteins. The basal body and type III secretion system are formed initially and provide the machinery to generate the flagella. A tumble occurs with reduced binding of chemoattractants, that causes increased amounts of CheY-P. These bind motor proteins to reduce the rotary speed of the flagella, leading to a tumble. Motor proteins are coloured red; flagellar proteins, brown; basal body proteins, blue; chemotaxis proteins, cluster 1 in orange and cluster 2 in olive; regulatory proteins, grey; and type III secretion system proteins, green. Gene function assumed to match *Sinorhizobium meliloti* where unknown in *R. leguminosarum*.

1.5 Chemotaxis in symbiotic bacteria

Chemotaxis involves regulating motility through variations in the concentration of metabolically-relevant chemicals. The chemicals involved can be either attractants (i.e. compounds which benefit the bacteria) or repellents (those compounds with negative effects).

Bacterial size is insufficient to sense concentration gradients spatially at swimming speeds so, instead, concentrations are compared temporally (Porter, Wadhams and Armitage 2011). Bacteria bias their movements towards high concentrations of attractants by reducing their tumbles and away from low concentrations of attractants and high concentrations of repellents by increasing their tumbles.

As with motility, the chemotactic systems in *R. leguminosarum* appear to match that of *S. meliloti*, having a similar array of genes. The main differences are the

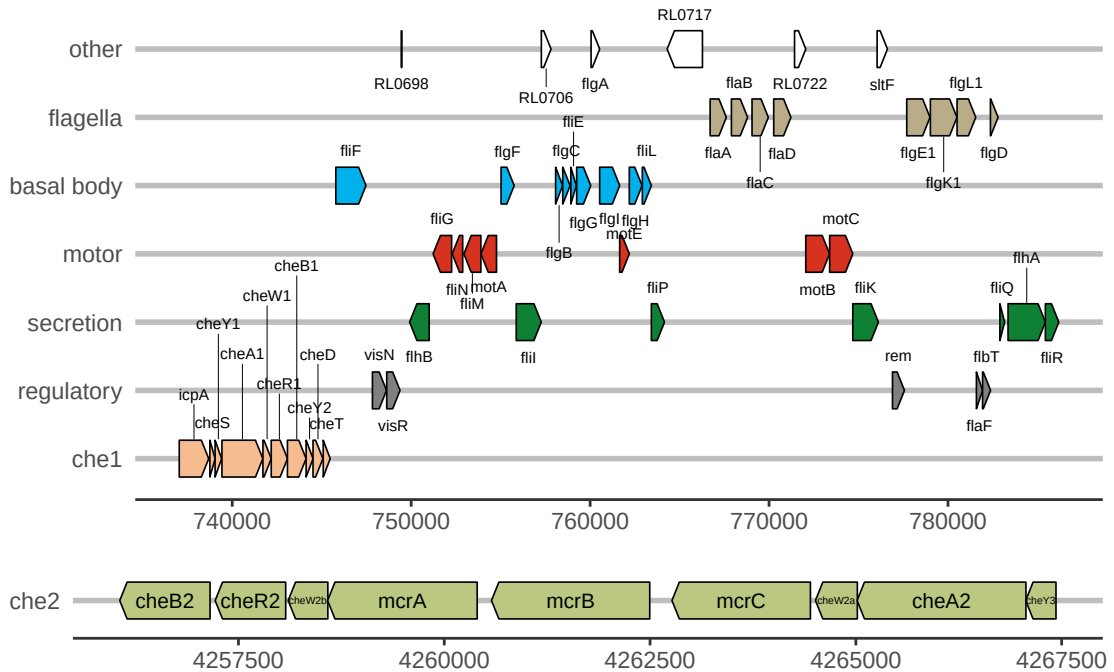


Figure 1.5: The main motility clusters of Rlv3841. The flagella and chemotaxis systems of Rlv3841 are present in only two clusters larger than three genes. The first (RL0685-736) contains genes encoding flagella, basal body, motor, type III secretion system, regulatory and Che1 proteins. The second (RL4028-36) contains genes encoding the Che2 proteins.

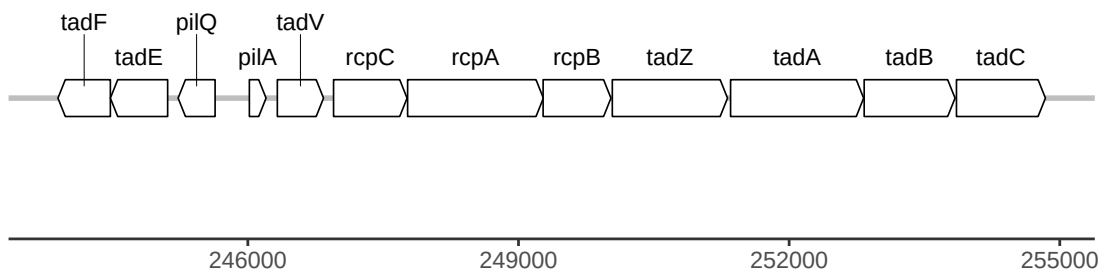


Figure 1.6: The T4 Tad pili operon structure of Rlv3841. Rlv3841 encodes a Type IV *tad* system pilus. The cluster (RL0208-19) contains genes encoding pili secretion, assembly and structural proteins.

presence of a duplicated set of chemotaxis genes and particularly in Rlv3841, several additional chemoreceptors. In the following section, I will describe those elements discovered in *S. meliloti* and those that have been confirmed in *R. leguminosarum* (See Figure 1.4) (Sourjik and Schmitt 1998; Miller *et al.* 2007; Tambalo *et al.* 2010a, 2010b).

The main regulatory system in bacterial chemotaxis is a two-component system, consisting of a histidine kinase (component one) that first receives the signal which it then transmits to the response regulator (component two). In chemotaxis, the histidine kinase is CheA, which phosphorylates the response regulator CheY (see Figure 1.4). In *S. meliloti*, there are two copies of CheY, CheY2 propagating the signal and CheY1 acting as a phosphate sink, increasing the rate at which CheY2 returns to its unphosphorylated state (Sourjik and Schmitt 1998). However, instead of a positive signal causing phosphorylation of the response regulator, as in most two-component systems, a negative signal to CheA increases its autophosphorylation, leading to increased amounts of phosphorylated CheY2 (CheY2-P). This is due to the activity of CheY2-P, which binds the motor protein FliM to induce reduction in rotary speed, inducing a tumble (Sourjik and Schmitt 1998). Since tumbling reduces the chance the bacteria will continue moving in its previous direction, CheY2 is activated when that direction is unfavourable. This biases the bacterial random walk, causing it to avoid environments that activate the phosphorylation cascade.

In addition to phosphorylation, CheY can also be acetylated. Binding of FliM by CheY actually occurs in three stages: firstly, CheY binds to the FliM_N site causing CheY potentiation and allowing binding of CheY-P to the second site on FliN causing brief CW episodes and allowing binding of CheY-Ac to the third site FliM_M for extended CW rotation (Fraiberg *et al.* 2015; Afanзар *et al.* 2021). The acetylation-requiring stage appears to be another regulatory gate that controls adaptation timing. CheY acetylation is enzymatically controlled, with acetylation and deacetylation by acetyl-CoA synthetase and further deacetylation by the sirtuin CobB and phosphate acetyltransferase (Baron, Afanзар and Eisenbach 2017; Baron

and Eisenbach 2017). Additionally, CheY can be autoacetylated from acetyl-CoA or acetyl phosphate as regulatory links to central metabolism.

In addition to the CheY proteins, CheA can phosphorylate another chemotaxis protein, CheB. Phosphorylated CheB demethylates chemoreceptors at an increased rate, increasing their sensitivity by reducing their ability to induce CheA autophosphorylation (Rice and Dahlquist 1991; Armitage and Schmitt 1997; Porter, Wadhams and Armitage 2011). If the cell experiences a persistent negative stimulus, the ability of the appropriate chemoreceptors to induce tumbling is reduced, adjusting their sensitivity to the lower concentration of the stimulus.

Another chemotaxis protein, CheR, methylates the chemoreceptors. This reverses the adjustment of CheB and is always active (Rice and Dahlquist 1991; Armitage and Schmitt 1997). Continued demethylation of a chemoreceptor by phosphorylated CheB is required to keep the chemoreceptor in a hyper-sensitive state. So, if the cell experiences a persistent positive stimulus, CheA has reduced autophosphorylation leading to reduced CheB-P and increased methylation of chemoreceptors. This adjusts the receptors to the new, higher concentration of effector.

Combinations of these simple systems form more complex emergent features. These include accumulation (i.e. balance between two antagonistic receptors can allow cells to target specific optimal concentrations) and log sensing (signal adaptation allows constant movement in an exponential gradient) (Colin and Sourjik 2017).

However, since different chemoreceptors bind different effectors, how does the signal adaptation separate conflicting signals such as in an environment where one receptor has a positive and another a negative signal? To answer this question, we must first discuss chemoreceptor structures and how they are arranged. Chemoreceptors (also known as methyl-accepting chemotaxis proteins or MCPs) are most often transmembrane proteins with a substrate binding domain that projects into the periplasm with transmembrane domains on either side. The cytoplasmic section of the proteins consists of the section where methylation occurs, a signalling region and a flexible bundle that bends to form a hairpin turn where the section is repeated, inverted to form a ‘U’ shape (See Figure 1.7) (Alexander and Zhulin

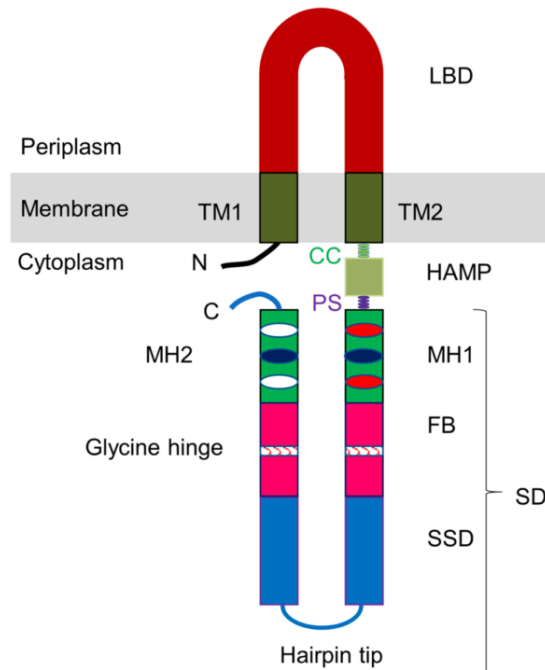


Figure 1.7: Typical chemoreceptor topology. The typical membrane bound chemoreceptor has two transmembrane helices (TM) flanking a ligand binding domain (LBD), such that the LBD senses in the periplasm. The sensing signal is transmitted via a HAMP domain (histidine kinase, adenylyl cyclase, methyl-accepting chemotaxis protein and phosphatase region) to the signalling subdomain (SSD), where the main chemotaxis machinery binds. The sensitivity of the receptor is modified by methylation and amidation of the methylation helices (MH). Additional domains including control cable (CC), phase stutter (PS) and flexible bundle (FB) act to preserve the structure. Reproduced from Salah Ud-Din and Roujeinikova (2017).

2007; Meier, Muschler and Scharf 2007; Salah Ud-Din and Roujeinikova 2017). Additionally, the carboxy-terminus of the protein often consists of a pentapeptide ‘NWETF’ motif (present in 4 out of the 8 MCPs in *S. meliloti* and 9 out of the 25 MCPs in Rlv3841), which provides a binding platform for CheR and CheB (Wu *et al.* 1996; Yost *et al.* 2004; Zatakia *et al.* 2017). The chemoreceptors are then clustered at the cell poles as ‘trimers of dimers.’ This arrangement is stabilised by the binding of CheA via the structural protein CheW to form hexagonal arrays of receptors on the inner membrane (Meier and Scharf 2009; Porter, Wadhams and Armitage 2011; Jones and Armitage 2015).

To summarise, chemoreceptors are generally positioned near the CheA protein that they activate and contain a binding motif for CheB and CheR near the methylation regions. Thus, it is likely that for each chemoreceptor, the phosphorylation state

of CheB proteins is related to the signals that the chemoreceptor sent to its bound CheA. The methylation feedback is effectively contained and therefore independent.

In addition to transmembrane chemoreceptors, both *S. meliloti* Rm2011 and Rlv3841 have cytoplasmic receptors—chemoreceptors without transmembrane domains. *S. meliloti* has IcpA, a transducer-like protein (TLP) that lacks both the transmembrane domains and conserved glutamine and glutamate residues that serve as methyl-accepting sites (Meier and Scharf 2009). It has been postulated that this protein senses the metabolic state of the cell (Meier, Muschler and Scharf 2007). Rlv3841 has a closely related protein that has been annotated as HemAT.

HemAT is a receptor initially described in *Bacillus subtilis* that binds haem and appears to act as an oxygen sensor. It has a similar domain structure to *S. meliloti*'s IcpA, except that it requires specific tyrosine and histidine residues to bind haem (Hou *et al.* 2001; Zhang, Olson and Phillips 2005). These residues are missing in Rlv3841's HemAT. In addition, the chemoreceptor McpH matches the domain structure of HemAT but includes the necessary residues to bind haem. It has thus been proposed that HemAT be reannotated as IcpA and McpH as HemAT (Miller 2007).

Rlv3841 has a copy of each of these proteins in its main chemotaxis cluster (designated Che1). However, it also has the following genes in a separate cluster (designated Che2): *cheA2*, *cheY3*, *cheB2* and *cheR2*, each of which is an extra copy of a *che1* gene. The Che1 cluster has been shown to have a major role in chemotaxis and to be required for competitive nodulation (Miller *et al.* 2007). The Che2 cluster, on the other hand, does not have a clear purpose besides having a small influence on swimming bias (Miller *et al.* 2007). Rlv3841 is not the only bacterium to have multiple sets of chemotaxis proteins. In *Rhodobacter sphaeroides*, when the active set of flagellar genes are deleted, suppressor mutations can activate a normally inactive second set of flagellar genes. Flagella produced by these genes are controlled by distinct chemotaxis proteins from separate operons and produce distinct chemotactic profiles (Del Campo *et al.* 2007). It is possible the *che2*

cluster in Rlv3841 can be activated under specific environmental conditions that cannot be replicated easily in vitro.

1.6 Regulation of motility and chemotaxis

Both *S. meliloti* and Rlv3841 have the same transcriptional activation system for flagellar and chemotaxis genes (See Figure 1.4). The class I genes consist of the heterodimer VisNR and Rem which positively regulate the class II genes (flagellar assembly and motor genes). The class II genes in turn activate class III genes (flagellin and chemotaxis genes) (Sourjik *et al.* 2000; McCarter 2006; Rotter *et al.* 2006; Tambalo *et al.* 2010b). The class system ensures that each gene is active only when its product is useful. In Rlv3841, VisNR and Rem are expressed throughout exponential phase but expression drops sharply during stationary phase, although it is still sufficient for continued motility (McCarter 2006; Tambalo *et al.* 2010b).

The only motility and chemotaxis genes in Rlv3841 that are independent of VisNR are scattered chemoreceptors and, notably, the *che2* cluster (Scharf, Hynes and Alexandre 2016). This possibly allows the bacteria to target the expression of *che2* genes, but without the expression of motility genes. *S. meliloti* respond to starvation conditions by increasing chemotactic movement through unknown mechanisms while in parallel down-regulating flagellar genes (Wei and Bauer 1998). Thus, the regulation of the chemotactic movement is independent of the regulation of the flagellar genes. Since the regulation of *che2* in Rlv3841 is independent, the cluster may be involved in a similar response in Rlv3841. The cells appear to make a final push to find a more favourable environment before motility is ceased to conserve energy.

1.7 Energy taxis in bacteria

Although chemotaxis allows bacteria to target a wide range of effector compounds, what it does not allow is dynamic targeting towards metabolically necessary compounds. To this end bacteria have evolved parallel systems to chemotaxis,

often labelled energy taxis. The simplest form of such systems is the use of cytoplasmic receptors that monitor internal concentrations of critical metabolites (Porter, Wadhams and Armitage 2011). This enables the bacteria to induce tumbling (rejecting the current environment) in response to low cellular energy levels. Energy taxis also spans aerotaxis, phototaxis, redox taxis, taxis to alternative electron acceptors, and chemotaxis to carbon sources (Taylor, Zhulin and Johnson 1999). It even overrides chemotaxis in low-energy cells (Taylor, Zhulin and Johnson 1999).

In *E. coli*, sugar energy taxes involve interactions between CheA and the phosphoenolpyruvate: carbohydrate phosphotransferase system (PTS). PTS is a system for transporting and phosphorylating carbohydrates (Figure 1.8). The phosphate is initially transferred from phosphoenolpyruvate (PEP) by enzyme I (EI) and is subsequently passed via HPr and enzyme IIA (EIIA) to enzyme IIB (EIIB). EIIB-P is responsible for phosphorylating the carbohydrates as they enter through the transporter (EIIC) (Adler and Epstein 1974; Lengeler *et al.* 1981; Postma, Lengeler and Jacobson 1993). There are multiple sets of EIIA, EIIB and EIIC for each carbohydrate, though there is only a single copy of EI and HPr genes. Thus, the phosphorylation state of EI and HPr acts as an overall indicator of sugar uptake (Postma, Lengeler and Jacobson 1993). Non-phosphorylated EI (indicative of increased amount of sugar uptake) inhibits CheA autophosphorylation, leading to increased smooth swimming (Lux *et al.* 1995; Rowsell *et al.* 1995; Taylor, Zhulin and Johnson 1999). Interestingly, the PTS systems of various *Vibrio* species have also been found to control motility, through EIIA interacting with flagella formation (Visick *et al.* 2007; Park *et al.* 2016). *E. coli* also has an additional paralogous system called Nitrogen-PTS (PTS^{Ntr}). The system involves a similar group of enzymes: EI^{Ntr} (encoded by *ptsP*), NPr (*ptsO/npr*) and EIIA^{Ntr} (*ptsN*) found in the *rpoN* region (Rabus *et al.* 1999; Kim *et al.* 2011; Jahn *et al.* 2013). PTS^{Ntr} is involved in the regulation of nitrogen metabolism in the cell.

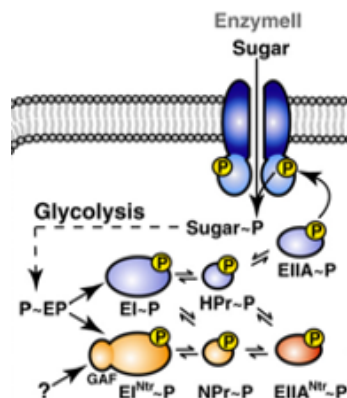


Figure 1.8: *Escherichia coli* Phosphotransferase system. The phosphate is initially transferred from phosphoenolpyruvate (PEP) to either EI or EI^{Ntr}, which interchange it with HPr or NPr. HPr and NPr can transfer the phosphate to either EIa^{Ntr} or EIa. EIa transfers the phosphate to EIb which phosphorylates sugars as they are transporter into the cell via EIc. Through glycolysis, the metabolism of sugars, more PEP is generated. Modified from Pflüger-Grau and Görke (2010).

1.8 Chemotaxis and motility interaction with the phosphotransferase system (PTS) in Rlv3841

The PTS in Rlv3841 is present in a greatly reduced form compared to *E. coli*. There is only a single sugar-PTS component remaining, ManX (equivalent to EIa enzyme). In addition, there are three PTS^{Ntr} components: PtsP (equivalent to EI^{Ntr} in *E. coli*), NPr (as above) and PtsN (EIa^{Ntr} in *E. coli*). The phosphate transfer runs in parallel to that found in *E. coli*. PtsP initially takes the phosphate from PEP, which it transfers to NPr. From here, the path forks, providing the phosphate to either PtsN or ManX. The phosphorylated and non-phosphorylated versions of these enzymes have different roles in controlling cell metabolism, nutrient and water uptake and potentially motility (See Figure 1.9) (Prell *et al.* 2012; Untiet *et al.* 2013). Unlike the sugar PTS, there are no membrane transport components in PTS^{Ntr}, indicating that this system is only sensing the internal metabolic status. Although less well studied, it is the PTS^{Ntr} that is most widely distributed in bacteria suggesting that it has a similar role in a broad spectrum of microorganisms (Sánchez-Cañizares *et al.* 2020).

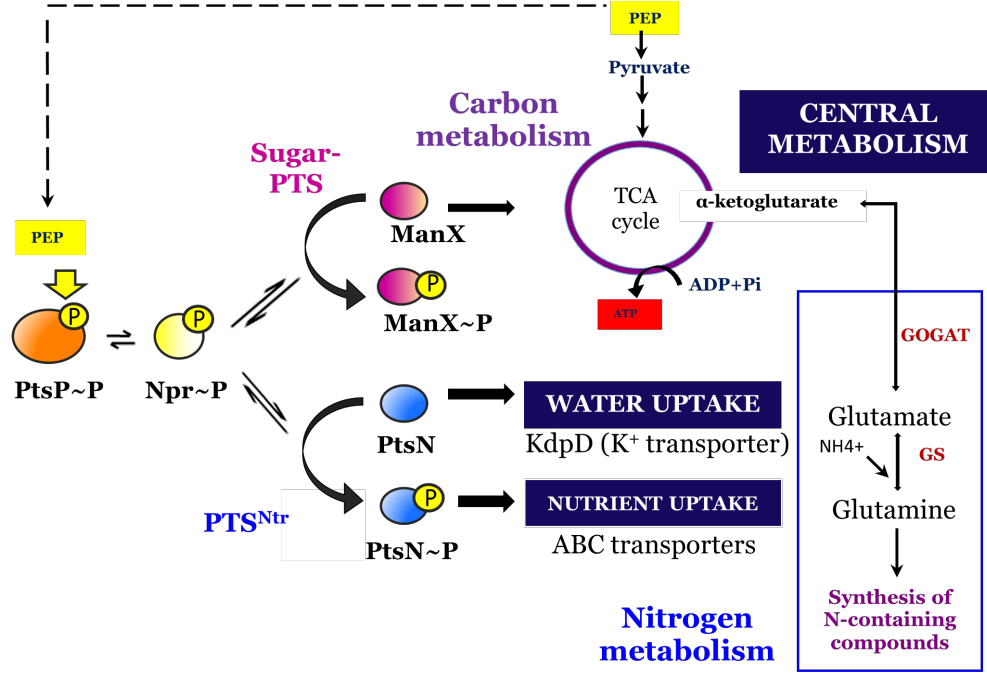


Figure 1.9: Rlv3841 Phosphotransferase system. The phosphate is initially transferred from phosphoenolpyruvate (PEP) to PtsP, which interchanges it with NPr. NPr can transfer the phosphate to either ManX or PtsN. ManX is involved in carbon metabolism, with the non-phosphorylated version activating the tricarboxylic acid (TCA) cycle, resulting in increased synthesis of N-containing compounds. PtsN is involved in nitrogen metabolism, the non-phosphorylated version influencing water via the potassium transporter and the phosphorylated version increasing nutrient concentration via the ATP-binding cassette (ABC) transporters. Thus, phosphorylated ManX and PtsN are related to increased carbon/nitrogen ratio and non-phosphorylation with reduced carbon/nitrogen ratio (Prell *et al.* 2012; Untiet *et al.* 2013; S’anchez-Cañizares *et al.* 2020).

In general, it appears that PtsN regulates nitrogen metabolism and ManX regulates carbon metabolism. Thus, the control of carbon and nitrogen consumption are linked—indeed since growth requires an ample supply of both elements, an increased supply of carbon should ideally be linked to increasing the uptake of nitrogen and vice-versa. Since these conditions are mutually exclusive, it could be expected that the phosphorylated version is active for one protein and the non-phosphorylated version for the other. That indeed appears to be the case, with PtsN-P activating ABC transporters and non-phosphorylated ManX activating the TCA cycle (See Figure 1.9) (S’anchez-Cañizares *et al.* 2020). Because PTS controls both the carbon and nitrogen metabolic activity in the cell, this may be the

system that causes the increase in chemotaxis during carbon and nitrogen starvation conditions. In addition, preliminary unpublished work by Carmen Sánchez-Cañizares from our laboratory has revealed that the PTS enzyme II, PtsN, binds to flagellar filament proteins. This may perhaps be the mechanism through which the cell responds to starvation of nitrogen by increasing motility.

1.9 Aerotaxis and the electron transport chain

Another common form of energy taxis is aerotaxis: movement towards an optimal oxygen concentration. There are multiple strategies that bacteria use to sense oxygen, including directly through sensing haem oxygen binding (e.g. DcrA of *Desulfovibrio vulgaris*) and indirectly through flavin adenine dinucleotide (FAD) binding to sense redox potential (e.g. Aer of *E. coli*) (Fu, Wall and Voordouw 1994; Bibikov *et al.* 1997; Greer-Phillips *et al.* 2003). *E. coli*'s Aer is purported to sense changes in the redox state of FAD, which can be reduced partially to FADH or completely to FADH₂ by the electron transport chain (ETC) (Edwards, Johnson and Taylor 2006). Thus Aer can sense any chemical that alters flux through the ETC, such as oxygen acting as a reductant sink and oxidisable substrates such as organic acids (Rebbapragada *et al.* 1997; Repik *et al.* 2000; Greer-Phillips *et al.* 2003).

In addition to taxis receptors, *E. coli* has sensors of oxygen concentration that signal for transcriptional changes. The ArcB sensor kinase autophosphorylates and phosphorylates the global transcriptional regulator ArcA. This phosphorylation is inhibited by oxidised ubiquinone but not by its reduced form (Malpica *et al.* 2004; Nitzschke and Bettenbrock 2018). Again, since oxygen acts as a reductant sink for the ETC, high concentrations of oxygen lead to higher amounts of oxidised ubiquinone and inhibition of ArcBA, whereas low concentrations lead to higher amounts of reduced ubiquinone and ArcBA activation (Nochino, Toya and Shimizu 2020). A similar system exists in *Rhodobacter*, with the sensor kinase RegB, inhibited by ETC activity, phosphorylating the response regulator RegA to regulate a variety of systems, including aerobic and anaerobic respiration and aerotaxis (Elsen *et al.* 2004; Willett and Crosson 2017).

Rlv3841 encodes the ActSR system, homologous to RegBA of *Rhodobacter*, but with low identity (<30%) to ArcBA of *E. coli* (Elsen *et al.* 2004). The ActSR system is also present in *Sinorhizobium meliloti* and *Rhizobium leguminosarum* biovar *viciae* VF39 where it has been found to impact acid tolerance, likely through pH sensing and activation of tolerance genes (Tiwari *et al.* 1996; Priefer *et al.* 2001). In addition, in *Sinorhizobium medicae*, ActSR was found to regulate a variety of systems including detoxification, nitrate assimilation, nitrogen fixation and microaerobic respiration at low pH and microaerobic conditions and to be activated by a reduction in flux through the ETC (Fenner *et al.* 2004). Low pH may also cause a reduction in ETC flux through the increased concentration of protons inhibiting proton pumping by terminal oxidases. (Fenner *et al.* 2004). The system in Rlv3841 is thus likely to also sense flux through the ETC and so could respond to oxygen concentration, low pH and potentially oxidisable metabolites as above.

ActSR is especially interesting due to the changes in the ETC during symbiosis (See Figure 1.10). Since the bacteroid is under microaerobic conditions while being fed with malate, which feeds directly into the TCA cycle, there is a high amount of reductant NADH produced that, together with succinate, reduces quinone at the ETC, producing a strong proton motive force without oxygen as a sink (Udvardi and Poole 2013; Degli Esposti and Martinez Romero 2016). This reductive pressure is relieved mainly by ATP hydrolysis by nitrogenase which enables further ATP synthase activity to stimulate respiration (Haaker *et al.* 1996). ActSR is thus likely active during bacteroid differentiation and could be transcriptionally regulating motility during the transition to sessility. If that is indeed the case, ActSR regulation of motility could be relevant to growth on high levels of oxidisable metabolites, such as organic acids.

1.10 Research aims

Our main aim in this project was to assess the influence of chemotaxis and motility upon the different stages of the *Rhizobium*-legume symbiosis. These systems have been shown previously to be crucial for competitively forming nodule-based

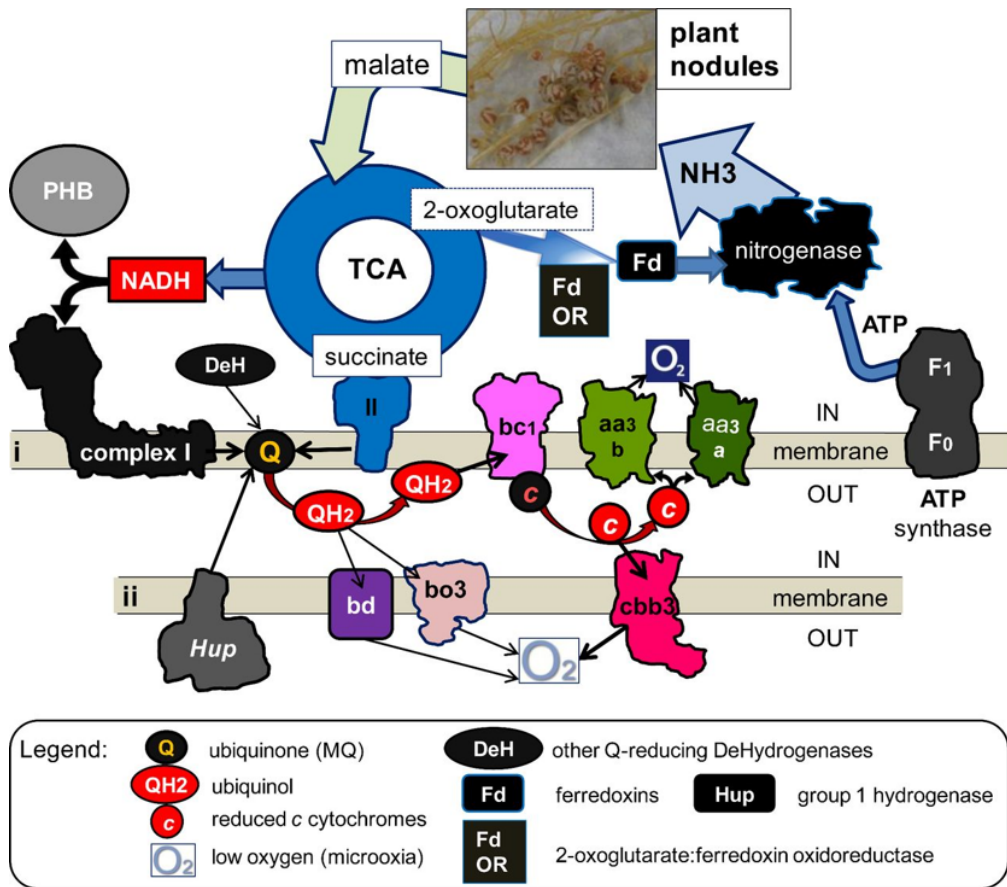


Figure 1.10: The energy metabolism and respiratory chain of nodulating symbionts. Symbiotic rhizobia receive organic acids from the plant in return for fixed nitrogen. These metabolites feed into the TCA cycle, driving NADH production and subsequently increased oxygen consumption through the electron transport chain. This additionally strengthens the proton-motive force by moving protons out through the inner membrane. Reproduced from Degli Esposti and Martinez Romero (2016).

symbiosis with legumes. However, the precise stage in the symbiosis process that causes this difference is currently unknown.

We began this project by characterising various chemotaxis and motility gene mutants of Rlv3841, including non-motile flagella mutants and markerless deletions of both chemotaxis clusters. This also included a bioinformatic analysis of all 27 candidate chemoreceptor genes encoded by Rlv3841 (see Chapter 3). During the characterisation, we discovered a novel swimming behaviour of Rlv3841: the bacterium not only adjusts its velocity in response to nutrient conditions, it can cease swimming altogether in response to high concentrations of TCA-cycle intermediates. To investigate the mechanism/s behind this phenotype, we began by testing the

candidate PTS mutants. When this regulatory system proved not to be important for the TCA-cycle intermediate response, we isolated spontaneous mutants that no longer responded to high levels of TCA-cycle intermediates. The whole genomes of these mutants were sequenced and bioinformatically analysed to detect differences from the wildtype to discover the genes responsible for the effect as described in Chapter 4. The chemotaxis and motility mutants (characterised in Chapter 3) were then compared to the wildtype at various stages of symbiosis, as explained in Chapter 5. The resulting data has provided a more precise description of the importance of chemotaxis and motility during the whole symbiotic cycle. It was found that the Che1 cluster in concert with flagella is important during early stages of symbiosis and that the Che2 cluster is important during nodule development.

General materials and methodology

2.1 Bacterial strains, plasmids and primers

2.1.1 Strains

The bacterial strains used in this thesis are listed in Table 2.1. Each strain was stocked in 15% glycerol and flash-frozen in liquid nitrogen for storage at -80 °C.

Table 2.1: Strains

Strain	Description	Reference
DH5 α	<i>E. coli</i> supE44, hsdR17, recA1 thi-1, Δ lacU169(ϕ 801acZ Δ M15) endA1 gyrA96 relA1	Hanahan (1983)
ST18	<i>E. coli</i> S17 λ pir Δ hemA	Schobert & Thoma (2009)
A1289	Rlv3841 Tn5:: <i>hemAT</i> pooled from Tn5 library	A. Downie
A1296	Rlv3841 Tn5:: <i>mcpQ</i> pooled from Tn5 library	A. Downie
A1300	Rlv3841 Tn5:: <i>mcpY2</i> pooled from Tn5 library	A. Downie
AA047	Rlv3841 <i>ptsN2</i> (markerless mutation) in <i>ptsN1</i> :: Ω Spec background	Sánchez-Cañizares et al. (2020)
D5250	Rlv3841 carrying pIJ11282 (pnptII::lux)	Frederix et al. (2014)
LM100	Rlv3841 Δ <i>che1</i>	Miller et al. (2007)
LM300	Rlv3841 Δ <i>che1,2</i>	Miller et al. (2007)
LM400	Rlv3841 Δ <i>che2cheBII</i>	Miller et al. (2007)
LMB271	Rlv3841 <i>ptsN1</i> :: Ω Spec	Prell et al. (2012)

Table 2.1: (continued)

Strain	Description	Reference
LMB612	Rlv3841 carrying pLMB712 (Rlv3841 pnodA::lux)	Pini et al. (2017)
LMB692	Rlv3841 $\Delta manX$	Sánchez-Cañizares et al. (2020)
OPS1012	Rlv3841 <i>manX</i> H9A (non-phosphorylatable)	Sánchez-Cañizares et al. (2020)
OPS1099	Rlv3841 <i>manX</i> H9D (phosphomimic)	Sánchez-Cañizares et al. (2020)
OPS1101	AA047 <i>ptsN1</i> H66A (non-phosphorylatable)	Sánchez-Cañizares et al. (2020)
OPS1103	AA047 <i>ptsN1</i> H66D (phosphomimic)	Sánchez-Cañizares et al. (2020)
OPS1500	Rlv3841 transduced with Tn5:: <i>hemAT</i> from A1289	This work
OPS1505	Rlv3841 transduced with Tn5:: <i>mcpY2</i> from A1300	This work
OPS1846	LM100 glmS::miniTn7T-aad9-T1-J23104-RBstd-mCherry-DT16-T0	This work
OPS1847	LM100 glmS::miniTn7T-aad9-T1-J23104-RBstd-sfGFP-DT16-T0	This work
OPS1848	LM100 glmS::miniTn7T-aad9-T1-J23104-RBstd-sYFP-DT16-T0	This work
OPS1854	LM300 glmS::miniTn7T-aad9-T1-J23104-RBstd-sYFP-DT16-T0	This work
OPS1855	LM300 glmS::miniTn7T-aad9-T1-J23104-RBstd-mCherry-DT16-T0	This work
OPS1856	LM300 glmS::miniTn7T-aad9-T1-J23104-RBstd-sfGFP-DT16-T0	This work
OPS1858	LM400 glmS::miniTn7T-aad9-T1-J23104-RBstd-mCherry-DT16-T0	This work
OPS1859	LM400 glmS::miniTn7T-aad9-T1-J23104-RBstd-sfGFP-DT16-T0	This work
OPS1860	LM400 glmS::miniTn7T-aad9-T1-J23104-RBstd-sYFP-DT16-T0	This work
OPS1869	Rlv3841 glmS::miniTn7T-aad9-T1-J23104-RBstd-sYFP-DT16-T0	This work
OPS1871	LM100 glmS::miniTn7T-aad9-T1-J23104-RBstd-gusA-DT16-T0	This work
OPS1872	LM300 glmS::miniTn7T-aad9-T1-J23104-RBstd-gusA-DT16-T0	This work
OPS1873	LM400 glmS::miniTn7T-aad9-T1-J23104-RBstd-gusA-DT16-T0	This work
OPS1881	Rlv3841 transduced with Tn5:: <i>mcpQ</i> from A1296	This work

Table 2.1: (*continued*)

Strain	Description	Reference
OPS1896	Rlv3841 glmS::miniTn7T-aad9-T1-J23104-RBstd-gusA-DT16-T0	This work
OPS1897	Rlv3841 glmS::miniTn7T-aad9-T1-J23104-RBstd-mCherry-DT16-T0	This work
OPS1953	Rlv3841 glmS::miniTn7T-aad9-T1-J23104-RBstd-sfGFP-DT16-T0	This work
OPS2085	Rlv3841 glmS::miniTn7T-aad9-T1-J23104-RBstd-celB-DT16-T0	This work
OPS2107	Rlv3841 $\Delta motA$	This work
OPS2108	Rlv3841 $\Delta icpA$	This work
OPS2139	LM100 carrying pLMB712 (Rlv3841 pnodA::lux)	This work
OPS2140	LM300 carrying pLMB712 (Rlv3841 pnodA::lux)	This work
OPS2141	LM400 carrying pLMB712 (Rlv3841 pnodA::lux)	This work
OPS2163	Rlv3841 $\Delta fliG$	This work
OPS2234	Rlv3841 $\Delta flaA$	This work
OPS2236	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2237	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2238	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2288	Rlv3841 $\Delta icpA$ transduced with Tn5::hemAT	This work
OPS2290	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2291	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work

Table 2.1: *(continued)*

Strain	Description	Reference
OPS2292	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2293	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2294	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2295	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2296	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2297	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2298	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2299	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2300	Rlv3841 spontaneous papillation isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2301	Rlv3841 centre halo isolate; isolated from fumarate/ammonium UMS swimming plates	This work
OPS2335	LM100 carrying pIJ11282 (pnptII::lux)	This work
OPS2336	LM300 carrying pIJ11282 (pnptII::lux)	This work
OPS2337	LM400 carrying pIJ11282 (pnptII::lux)	This work

Table 2.1: *(continued)*

Strain	Description	Reference
OPS2338	OPS2107 carrying pIJ11282 (pnptII::lux)	This work
OPS2339	OPS2163 carrying pIJ11282 (pnptII::lux)	This work
OPS2346	OPS2107 glmS::miniTn7T-aad9-T1-J23104-RBstd-gusA-DT16-T0	This work
OPS2442	OPS2107 glmS::miniTn7T-aad9-T1-J23104-RBstd-sYFP-DT16-T0	This work
OPS2443	OPS2107 glmS::miniTn7T-aad9-T1-J23104-RBstd-mCherry-DT16-T0	This work
OPS2445	OPS2163 glmS::miniTn7T-aad9-T1-J23104-RBstd-sYFP-DT16-T0	This work
OPS2446	OPS2163 glmS::miniTn7T-aad9-T1-J23104-RBstd-mCherry-DT16-T0	This work
OPS2447	OPS2163 glmS::miniTn7T-aad9-T1-J23104-RBstd-gusA-DT16-T0	This work
OPS2449	OPS2234 glmS::miniTn7T-aad9-T1-J23104-RBstd-sYFP-DT16-T0	This work
OPS2450	OPS2234 glmS::miniTn7T-aad9-T1-J23104-RBstd-mCherry-DT16-T0	This work
OPS2526	OPS2163 carrying pLMB712 (Rlv3841 pnodA::lux)	This work
OPS2528	OPS2107 carrying pLMB712 (Rlv3841 pnodA::lux)	This work
OPS2558	OPS2290 glmS::miniTn7T-aad9-T1-native-pstSCAB-T0	This work
OPS2559	OPS2292 glmS::miniTn7T-aad9-T1-native-pstSCAB-T0	This work
OPS2560	OPS2297 glmS::miniTn7T-aad9-T1-native-pstSCAB-T0	This work
OPS2561	OPS2300 glmS::miniTn7T-aad9-T1-native-pstSCAB-T0	This work
OPS2562	OPS2237 glmS::miniTn7T-aad9-T1-native-RL3464-T0	This work
OPS2563	OPS2294 glmS::miniTn7T-aad9-T1-native-RL3464-T0	This work
OPS2564	OPS2298 glmS::miniTn7T-aad9-T1-native-RL3464-T0	This work
OPS2565	Rlv3841 glmS::miniTn7T-aad9-T1-J23104-RL3464-DT16-T0	This work
OPS2566	OPS2237 glmS::miniTn7T-aad9-T1-J23104-RL3464-DT16-T0	This work

Table 2.1: (continued)

Strain	Description	Reference
OPS2567	OPS2238 glmS::miniTn7T-aad9-T1-J23104-RL3464-DT16-T0	This work
OPS2568	OPS2294 glmS::miniTn7T-aad9-T1-J23104-RL3464-DT16-T0	This work
OPS2569	OPS2298 glmS::miniTn7T-aad9-T1-J23104-RL3464-DT16-T0	This work
OPS2570	Rlv3841 glmS::miniTn7T-aad9-T1-J23104-RL3464(OPS2294)-DT16-T0	This work
OPS2571	OPS2237 glmS::miniTn7T-aad9-T1-J23104-RL3464(OPS2294)-DT16-T0	This work
OPS2572	OPS2238 glmS::miniTn7T-aad9-T1-J23104-RL3464(OPS2294)-DT16-T0	This work
OPS2573	OPS2294 glmS::miniTn7T-aad9-T1-J23104-RL3464(OPS2294)-DT16-T0	This work
OPS2574	OPS2298 glmS::miniTn7T-aad9-T1-J23104-RL3464(OPS2294)-DT16-T0	This work
OPS2591	OPS2237 glmS::miniTn7T-aad9-T1-T0	This work
OPS2592	OPS2238 glmS::miniTn7T-aad9-T1-T0	This work
OPS2593	OPS2238 glmS::miniTn7T-aad9-T1-native-RL3464-T0	This work
OPS2594	OPS2290 glmS::miniTn7T-aad9-T1-T0	This work
OPS2595	OPS2292 glmS::miniTn7T-aad9-T1-T0	This work
OPS2596	OPS2294 glmS::miniTn7T-aad9-T1-T0	This work
OPS2597	OPS2295 glmS::miniTn7T-aad9-T1-T0	This work
OPS2598	OPS2295 glmS::miniTn7T-aad9-T1-native-pstSCAB-T0	This work
OPS2599	OPS2297 glmS::miniTn7T-aad9-T1-T0	This work
OPS2600	OPS2298 glmS::miniTn7T-aad9-T1-T0	This work
OPS2601	OPS2300 glmS::miniTn7T-aad9-T1-T0	This work
Rlv3841	<i>R. leguminosarum</i> bv. <i>viciae</i> ; <i>Str^R</i> derivative of strain 300	Johnston & Beringer (1975)
RU1235	Rlv3841 carrying pRU583 (Rlv3841 phoA::GFP)	Allaway et al. (2001)
Sme2011	<i>Sinorhizobium meliloti</i> 2011 <i>str3</i> ; <i>Str^R</i> derivative of strain SU47	Boucher et al. (1977)

2.1.2 Plasmids

The plasmids used in this thesis are listed in Table 2.2.

Table 2.2: Plasmids

Plasmid	Description	Reference
pIJ11282	<i>nptII</i> promoter driving <i>luxCDABE</i> ; Amp ^R , Tet ^R	Frederix et al. (2014)
pK18mobsacB	pK18 based mobilisable plasmid for double recombination in <i>Rhizobium</i> ; suicide vector on sucrose; Kan ^R	Schäfer et al. (1994)
pLMB712	pIJ11268 derivative; Rlv3841 <i>nodABC</i> promoter driving <i>luxCDABE</i> ; Amp ^R , Tet ^R	Pini et al. (2017)
pOPS0856	pUC18T-miniTn7T-aad9-T1-J23104-RBstd-mCherry-DT16-T0; Spec ^R	S. Guha
pOPS0857	pUC18T-miniTn7T-aad9-T1-J23104f-RBstd-gusA-DT16-T0; Spec ^R	S. Guha
pOPS0858	pUC18T-miniTn7T-aad9-T1-J23104-RBstd-sfGFP-DT16-T0; Spec ^R	S. Guha
pOPS0860	pUC18T-miniTn7T-aad9-T1-J23104-RBstd-sYFP-DT16-T0; Spec ^R	S. Guha
pOPS1376	pUC18T-miniTn7T-aad9-T1-J123104-RBstd-celB-DT16-T0; Amp ^R , Spec ^R	This work
pOPS1411	pK18mobsacB derivative; containing flanking regions of Rlv3841 <i>icpA</i> ; F with oxp3769+oxp3770, R with oxp3771+oxp3772; Kan ^R	This work
pOPS1412	pK18mobsacB derivative; containing flanking regions of Rlv3841 <i>motA</i> ; F with oxp3773+oxp3774, R with oxp3775+oxp3776; Kan ^R	This work
pOPS1422	pK18mobsacB derivative; containing flanking regions of Rlv3841 <i>fliG</i> ; F with oxp3933+oxp3934, R with oxp3935+oxp3936; Kan ^R	This work
pOPS1423	pK18mobsacB derivative; containing flanking regions of Rlv3841 <i>flaA</i> ; F with oxp3921+oxp3922, R with oxp3923+oxp3924; Kan ^R	This work
pOPS1666	pUC18T-miniTn7T-Sp-T1-native-RL3464-T0; Amp ^R , Spec ^R	This work
pOPS1667	pUC18T-miniTn7T-Sp-T1-native-RL3464(OPS2294)-T0; Amp ^R , Spec ^R	This work
pOPS1676	pUC18T-miniTn7T-Sp-T1-native-pstSCAB-T0; Amp ^R , Spec ^R	This work
pOPS1677	pUC18T-miniTn7T-Sp-T1-J23104-RBstd-RL3464-T0; Amp ^R , Spec ^R	This work

Table 2.2: (continued)

Plasmid	Description	Reference
pOPS1679	pUC18T-miniTn7T-Sp-T1-J23104-RBstd-RL3464(OPS2294)-T0; Amp ^R , Spec ^R	This work
pRK2013	Helper plasmid for mobilisation of non-self-transmissible plasmids; Kan ^R	Figurski & Helinski (1979)

2.1.3 Primers

The primers used in this thesis are listed in Table 2.3. All primers were synthesised by Eurofins MWG Operon.

Table 2.3: Primers

Name	Description	Sequence
M13F	Sense for mapping pK18mobSacB insertions	CACGACGTTGTAAAACGA
M13R	Antisense for mapping pK18mobSacB insertions	GGATAACAATTTACACAGG
oxp2646	Sense for mapping Tn5 insertion into <i>hemAT</i>	GCTTGATGGTCATGGCGTTC
oxp2649	Sense for mapping Tn5 insertion into <i>mcpQ</i>	CTAAGCCATCAAAGCGCGTC
oxp2660	Sense for mapping Tn5 insertion into <i>mcpY2</i>	ATCAACCAGAGCATTGCGGA
oxp3769	Sense for PCR of <i>icpA</i> F 1kb flanking region for cloning into pK18mobsacB	CGAATTCGAGCTCGGTACCCGATA AAGAACAGCAGGATCACATCG
oxp3770	Antisense for PCR of <i>icpA</i> F 1kb flanking region for cloning into pK18mobsacB	ATCCGATAGGCGCCTGTCCCCAAC CACTTTC
oxp3771	Sense for PCR of <i>icpA</i> R 1kb flanking region for cloning into pK18mobsacB	GGGACAGGCGCCTATCGGATGATT GCCTC
oxp3772	Antisense for PCR of <i>icpA</i> R 1kb flanking region for cloning into pK18mobsacB	GTCGACTCTAGAGGATCCCCCGA CTTCAGCATCACCTTC
oxp3773	Sense for PCR of <i>motA</i> F 1kb flanking region for cloning into pK18mobsacB	CGAATTCGAGCTCGGTACCCCGGA AGATCCAGGGAGAG
oxp3774	Antisense for PCR of <i>motA</i> F 1kb flanking region for cloning into pK18mobsacB	TGGGCTGACGATGACTATGAGCAA TGCTTC
oxp3775	Sense for PCR of <i>motA</i> R 1kb flanking region for cloning into pK18mobsacB	TCATAGTCATCGTCAGCCCAATCT GAAC

Table 2.3: (continued)

Name	Description	Sequence
oxp3776	Antisense for PCR of <i>motA</i> R 1kb flanking region for cloning into pK18mobsacB	GTCGACTCTAGAGGATCCCCAAAA CCGCCGGACACTTTTG
oxp3917	Sense for mapping of <i>icpA</i> in-frame mutation	GGAATGAGGAGGAGGGCAATCG
oxp3918	Antisense for mapping of <i>icpA</i> in-frame mutation	CGAGGGAGACGGCAATTCTG
oxp3919	Sense for mapping of <i>motA</i> in-frame mutation	GAAGTCTCGCCGCCAAAGG
oxp3920	Antisense for mapping of <i>motA</i> in-frame mutation	TCATGCTTCAAATCCCGCCC
oxp3921	Sense for PCR of <i>flaA</i> F 1kb flanking region for cloning into pK18mobsacB	CGAATTCGAGCTCGGTACCCATGC GAGGTTGCCGGGAAAAC
oxp3922	Antisense for PCR of <i>flaA</i> F 1kb flanking region for cloning into pK18mobsacB	CCGACCGGTCAAGTCTGTCCCCTG GAAACTAG
oxp3923	Sense for PCR of <i>flaA</i> R 1kb flanking region for cloning into pK18mobsacB	GGACAGACTTGACCGGTCCGCACA AGGTTTC
oxp3924	Antisense for PCR of <i>flaA</i> R 1kb flanking region for cloning into pK18mobsacB	GTCGACTCTAGAGGATCCCCAGTC GCTGAGAGCGGAGAC
oxp3925	Sense for mapping of <i>flaA</i> in-frame mutation	ACGGGTTGTCTTTCGCG
oxp3926	Antisense for mapping of <i>flaA</i> in-frame mutation	AGCGTGAGGATGGTCTGCG
oxp3933	Sense for PCR of <i>fliG</i> F 1kb flanking region for cloning into pK18mobsacB	CGAATTCGAGCTCGGTACCCTTGG CGATCATGCGCCGG
oxp3934	Antisense for PCR of <i>fliG</i> F 1kb flanking region for cloning into pK18mobsacB	TTAGACCATGCACGTCGAGACTGT TGACAAATC
oxp3935	Sense for PCR of <i>fliG</i> R 1kb flanking region for cloning into pK18mobsacB	TCTCGACGTGCATGGTCTAACCTC TCCG
oxp3936	Antisense for PCR of <i>fliG</i> R 1kb flanking region for cloning into pK18mobsacB	GTCGACTCTAGAGGATCCCCACTG CTCAAGACCAAGGTG
oxp3937	Sense for mapping of <i>fliG</i> in-frame mutation	GCGTGATCTTCGGCAAGGG
oxp3938	Antisense for mapping of <i>fliG</i> in-frame mutation	CGAACCGAGCCTTTCAGCC
oxp4055	Sense for sequencing of <i>motA</i> in-frame mutation	GATCCATCGCCGGTTTGTCG

Table 2.3: (continued)

Name	Description	Sequence
oxp4056	Antisense for sequencing of <i>motA</i> in-frame mutation	GCAACTCGCCATTTCCCTCG
oxp4057	Sense for sequencing of <i>icpA</i> in-frame mutation	CGGAGCTTGCGTGACCTTG
oxp4058	Antisense for sequencing of <i>icpA</i> in-frame mutation	TCTCGGCGGACATCAGAACC
oxp4059	Sense for sequencing of <i>flaA</i> in-frame mutation	GCGAGATTGCCCTGGACATG
oxp4060	Antisense for sequencing of <i>flaA</i> in-frame mutation	TGGAGACCCTGGTTGATGCC
oxp4063	Sense for sequencing of <i>fliG</i> in-frame mutation	GCAAGGACCTAAAGCGCGTC
oxp4064	Antisense for sequencing of <i>fliG</i> in-frame mutation	TCGGCAGCGCTTTGAATTCC
oxp4636	Sense for sequencing of RL0543 mutations	CGGGCTTTTTGGGGCTATGG
oxp4637	Antisense for sequencing of RL0543 mutations	AACCGGATAATTGGCGGCG
oxp4638	Sense for sequencing of RL0545 mutations	TGCCTCAACCGCATGAACG
oxp4639	Antisense for sequencing of RL0545 mutations	AACGGGAAGGCTTGGTGTCTG
oxp4640	Sense for sequencing of RL3464 mutations	CGGCGTGTTTCATCGTCCATG
oxp4641	Antisense for sequencing of RL3464 mutations	ATGTTGAACCAGCCGCGTG
oxp4656	Sense for PCR of RL3463-4 with native promoter for cloning into pUC18T-mini-Tn7T-aad9	CATGCATGAGCTCACTAGTGCATC GGTCATCAGCTTGGC
oxp4657	Antisense for PCR of RL3463-4 with native promoter for cloning into pUC18T-mini-Tn7T-aad9	CTTCTCGAGGAATTCCTGCAATGT TGAACCAGCCGCGTG
oxp4657	Antisense for PCR of RL3464 with J23104 promoter for cloning into pUC18T-mini-Tn7T-aad9	CTTCTCGAGGAATTCCTGCAATGT TGAACCAGCCGCGTG
oxp4732	Sense for PCR of RL3464 with J23104 promoter for cloning into pUC18T-mini-Tn7T-aad9	CATGCATGAGCTCACTAGTGTTGA CAGCTAGCTCAGTCCTAGGTATTG TGCTAGCTACTCTAGAATTAAAGA GGAGAAATTAACCAATGACCACCG CTCTAGGC

Table 2.3: (continued)

Name	Description	Sequence
oxp4733	Sense for PCR of RL0543-5 with native promoter for cloning into pUC18T-mini-Tn7T-aad9	GGGTACCCATCATGCATGAGCTCA CTAGTGAGATGATCGGCCCTGACA G
oxp4734	Antisense for PCR of RL0543-5 with native promoter for cloning into pUC18T-mini-Tn7T-aad9	CGGGCCCAAGCTTCTCGAGGAATT CCTGCAAATATGTGTGATGCCAT GG
Tn5	Sense for mapping from either end of the Tn5 transposon	GAACGTTACCATGTTAGGAGGT

2.2 Chemicals and kits

Antibiotics ampicillin (Amp), gentamycin (Gm), kanamycin (Km), neomycin (Neo), nitrofurantoin (Nit), streptomycin (Strep), spectinomycin (Spec) and tetracycline (Tet) sourced from Sigma Aldrich. Plasmid isolation, gel extraction and PCR purification kits were purchased from New England Biolabs. DNeasy Blood & Tissue Kit for genomic DNA isolation were purchased from Qiagen. Restriction enzymes and buffers for molecular biology were purchased from New England Biolabs or Thermo Scientific. Primers were synthesised by Eurofins Genomics. OneTaq, GoTaq, Phusion and q5 PCR kits for PCR amplification were purchased from New England Biolabs or Promega.

2.3 General molecular biology techniques

2.3.1 DNA isolation

Genomic DNA was isolated from Rlv3841 cultures using DNeasy Blood & Tissue kit (Qiagen), following the manufacturer's protocols. Plasmid were isolated from *E. coli* cultures using Plasmid purification kit (New England Biolabs), following the manufacturer's protocols. DNA produced by PCR or restriction digest was purified where necessary using PCR purification or Gel purification kits (New England Biolabs). The DNA was purified following the manufacturer's protocols. In addition, the elution volume was heated to 55 °C for large plasmids.

2.3.2 DNA concentration

DNA concentration was routinely determined using broad-range dsDNA Qubit assay. 2 μ L of sample was analysed as per the manufacturer's protocols.

2.3.3 PCR amplification

OneTaq or GoTaq polymerases were used for colony screening with 15 μ L reactions. Screening of *E. coli* transformants: single colony transferred into the reaction using a pipette tip. A typical OneTaq cycle was 10 minutes of 94 °C denaturation followed by 30 cycles of 94 °C denaturation for 30 seconds, 56-59 °C annealing for 30 seconds and 70 °C extension for one minute per kb, followed by 68 °C extension for 10 minutes. A typical GoTaq cycle was as above, with denaturation at 95 °C and extension at 72 °C.

q5 or Phusion polymerases were used for cloning or sequencing with 50 μ L reactions and 5 pmol of template (typically genomic DNA). Thermocycler conditions were set according to the manufacturer's instructions for polymerases, with T_m determined by Geneious R9. Typical annealing temperatures of 61-65 °C. Typical q5 or Phusion cycles were as above, with denaturation at 98 °C and extension for 30 seconds per kb per cycle.

2.3.4 DNA gel electrophoresis

GeneRule 1kb+ marker ladder (Thermo Scientific) was used for DNA electrophoresis. Electrophoresis was used to identify the size of PCR products and restriction digests. 0.9% agarose (Sigma Aldrich) gel in TAE buffer pH 8.4 (400 mM Tris acetate, 1 mM EDTA) were run at 100 mV for 40 minutes. DNA was stained with Sybr Safe DNA stain (Invitrogen) and visualised with GelDoc EZ System (BioRad).

2.3.5 General DNA sequencing

15 μ L of purified DNA at 50-100 ng/ μ L was mixed with 2 μ L of primer for Sanger sequencing by Eurofins genomics.

2.4 Cloning

2.4.1 Restriction digests and DNA ligation

When necessary, purified DNA was digested with restriction endonucleases in 50 μ L reactions. The enzymes were heat inactivated after digestion. DNA was ligated with T4 DNA ligase in T4 DNA ligase buffer (Thermo Scientific) at RT for 30 minutes.

2.4.2 InFusion HD cloning

InFusion HD cloning was performed using InFusion HD cloning kit (ClonTech). The parts were generated with 15-20 bp overhangs on the primers and were cloned using a thermocycler according to the manufacturer's protocols. This process was tested *in silico* using Geneious R9.

2.4.3 NEBuilder Hi-Fidelity cloning

NEBuilder Hi-Fi cloning was performed using NEB NEBuilder master mix (NEB). Primers to amplify the parts were generated using the online NEBuilder assembly tool (nebuilder.neb.com). The process was additionally tested *in silico* using Geneious R9.

2.4.4 Transformation

Cloned plasmids were transformed into chemically competent *E. coli* DH5 α cells generated in the laboratory. A DH5 α culture was incubated in 500 mL LB for 3 hours at 37 °C. Once the culture reached an optical density of OD₆₀₀ 0.3-0.4, it was incubated on ice 10 minutes. Then, the cells were pelleted at 4000 rpm at 4 °C for 10 minutes. The pellet was resuspended in 10 mL ice cold 0.1 M CaCl₂ 15% glycerol (v/v) and incubated on ice for 15-30 minutes. The suspension was then pelleted at 4000 rpm 4 °C for 10 minutes. The pellet was resuspended in 20 mL ice cold 0.1 M CaCl₂ 15% glycerol (v/v) and incubated at 4 °C for 30 minutes. Aliquots (50-200 μ L) were dispensed into chilled Eppendorf tubes and snap frozen in liquid nitrogen. Competent cells were stored at -80 °C.

For transformation, *E. coli* was thawed on ice and combined at a ratio of 50 μL per 2-10 μL of DNA mix. The cells were then heat-shocked at 42 °C for 30 seconds and returned to ice for 2 minutes. The cells were then diluted to 1 mL in SOC media (Melford) and shaken at 37 °C for 1 hour. 100 μL was plated and the remained was pelleted at 10000 rpm for 1 minute and resuspended in 100 μL and was also spread-plated on LB media with appropriate antibiotics and X-gal where necessary. Plates were incubated overnight at 37 °C.

2.4.5 Tri-parental conjugation from *E. coli* into Rlv3841

Tri-parental mating was used to transfer plasmids from their *E. coli* stock, via an *E. coli* strain containing the conjugation plasmid pRK2013, into various Rlv3841 strains. The rhizobia were grown for 3 days on TY slopes and washed in 3 mL of TY. The *E. coli* were grown in 10 mL of LB liquid culture overnight, then subcultured 1:100 and grown to exponential phase. These were all pelleted at 4500 rpm for 10 minutes and washed thrice with TY (at 6000 rpm for 5 minutes for later pelleting stages). 300 μL each of the *E. coli* and 400 μL of rhizobia were then mixed and pelleted at 6000 rpm for 5 minutes. This pellet was then resuspended in 20 μL of TY, spotted on plain TY media and incubated overnight at 28 °C. A streak from the droplet was spread on TY media with appropriate antibiotics and left to grow 3-4 days at 28 °C. Nitrofurantoin at 20 $\mu\text{g} \cdot \text{mL}^{-1}$ was used to select against *E. coli*.

2.4.6 Di-parental conjugation from *E. coli* into Rlv3841

Alternatively, di-parental mating was also used with *E. coli* ST18, containing both the conjugation genes and the plasmid to be transferred. This strain requires a supplement of 50 $\mu\text{g} \cdot \text{mL}^{-1}$ of 5-aminolevulinic acid (ALA) to be added to the media. The rhizobia were grown in 10 mL of TY liquid culture for 3 days. The *E. coli* were grown in 10 mL of LB liquid culture overnight, then subcultured 1:100 and grown to exponential phase. These were all pelleted at 4500 rpm for 10 minutes and washed twice with TY (first 1 mL, then 400 μL). 100 μL each of the *E. coli* and rhizobia were mixed and spotted onto TY + ALA. 100 μL of each was also

spotted individually for negative controls. These droplets were incubated overnight at 28 °C after drying. The droplets were then resuspended in 200 μ L of TY media and serially diluted ten-fold three times. 100 μ L of each dilution was then spread onto selective TY media, without supplemented ALA.

2.5 Mutagenesis techniques

2.5.1 Tn5 transduction

This technique was used to remove possible extraneous mutations in Tn5 marked strains. Firstly, the donor strain was grown on a TY slope and washed in 3 mL sterile H₂O. 100 μ L of this was mixed with a serial dilution (10^{-2} - 10^{-6}) of the phage RL38 (Buchanan-Wollaston 1979). Controls with no bacteria or no phage were also made. These mixtures were added to 4 mL of molten 50% TY/ 50% TY agar (at 42 °C) and poured into a petri dish. The dishes were then incubated for 2-3 days at 28 °C. At this stage, a plate was chosen at the point just before complete confluence and eluted by adding 10 mL of H₂O and rocking for 2 hours. The solution (called phage prep or phage lysate) was millipore filtered and combined with 2-4 drops of chloroform for storage at 4 °C.

After this, the recipient strain was grown on TY slopes for 2-3 days. The slopes were washed with 3 mL TY. 200 μ L of this was mixed with 1 μ L of a three times ten-fold serial dilution of the phage lysate. Again, mixtures with no phage or non bacteria were used as negative controls. These mixtures were incubated at 28 °C for 1 hour then 100 μ L of each was spread on TY plates with a selective concentration of kanamycin (Section 2.6.2). After incubating the plates at 28 °C for 3-5 days, early single colonies were picked and spread on selective TY plates. These colonies were screened using colony PCR with an inward facing primer specific for the target insertion and the general outward-facing Tn5 primer. The amplified region of successful colonies were sequenced using these same primers (Eurofins MWG Operon). Positive mutants were subsequently stocked at -80 °C.

2.5.2 Double cross-over mutagenesis by pK18mobSacB

This technique was used to isolate stable mutants. The plasmid pK18mobSacB was first digested with SmaI. Primers were designed to amplify regions of approximately 1000 bp precisely flanking the DNA region of interest. These were designed as per Section 2.4.3 to work with NEBuilder Hi-Fidelity cloning. The flanking regions were then amplified using the primers with q5 polymerase and cloned into pK18mobSacB with NEBuilder. The product of this reaction was transformed into chemically competent DH5 α , and plated on LB containing a selective concentration of kanamycin and X-gal to allow blue-white selection (Section 2.6.2). Correct colonies (white) were confirmed using PCR mapping and subsequent sequencing (Eurofins MWG Operon). M13F and M13R primers were used for the mapping and sequencing reactions. Once a plasmid was confirmed to be correct, it was transformed into *E. coli* ST18 to enable di-parental mating.

Next, the plasmid was conjugated into Rlv3841 (Section 2.4.6). Conjugants were selected on TY with a selective concentration of kanamycin. Provided there was no growth on control plates, two positive colonies were inoculated into 10 mL of non-selective TY media. After grown to stationary phase, the cultures were serial diluted ten-fold four times. 100 μ L of each dilution was then spread onto TY media supplemented with 10% (v/v) sucrose. 20-40 colonies were then screened using colony PCR with the outer primers for the flanking regions above. A successful deletion produced a truncated PCR product relative to the wild type. The colonies were also passaged onto TY and TY with kanamycin. A successful deletion was positive for growth on TY but not on TY with kanamycin. The amplified region of two colonies that were successful under both tests were sequenced using newly designed inward facing primers (Eurofins MWG Operon). The mutants were subsequently stocked at -80 °C.

2.6 Media

2.6.1 Media recipes

Tryptone-yeast (TY) media was used as a rich-media growth source for *Rhizobium* strains. This consisted of: $3 \text{ g} \cdot \text{L}^{-1}$ yeast extract, $5 \text{ g} \cdot \text{L}^{-1}$ tryptone and $1.3 \text{ g} \cdot \text{L}^{-1}$ CaCl_2 . $17.5 \text{ g} \cdot \text{L}^{-1}$ agar was added prior to autoclaving for solid medium.

Universal minimal salts (UMS) media was used as a minimal-media growth source for *Rhizobium* strains. This consisted of: $0.5 \text{ g} \cdot \text{L}^{-1}$ $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, $0.2 \text{ g} \cdot \text{L}^{-1}$ NaCl , $4.19 \text{ g} \cdot \text{L}^{-1}$ MOPS, 0.5 mM K_2HPO_4 and 1 mL of trace elements ($0.375 \text{ g} \cdot \text{L}^{-1}$ $\text{EDTA} - \text{Na}_2$, $0.16 \text{ g} \cdot \text{L}^{-1}$ $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, $0.2 \text{ g} \cdot \text{L}^{-1}$ NaMoO_4 , $0.25 \text{ g} \cdot \text{L}^{-1}$ H_3BO_3 , $0.2 \text{ g} \cdot \text{L}^{-1}$ $\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$, $0.02 \text{ g} \cdot \text{L}^{-1}$ $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ and $1 \text{ g} \cdot \text{L}^{-1}$ $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$). The pH of the mixture was adjusted to 7 prior to autoclaving. After autoclaving, the media was supplemented with 1 mL of calcium stock solution ($75 \text{ g} \cdot \text{L}^{-1}$ $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$), 1 mL of iron stock solution ($12 \text{ g} \cdot \text{L}^{-1}$ $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ dissolved in 1 M HCl) and 1 mL of vitamin stock solution ($1 \text{ g} \cdot \text{L}^{-1}$ thiamine hydrochloride, $2 \text{ g} \cdot \text{L}^{-1}$ D-pantothenic acid calcium salt, $100 \text{ mg} \cdot \text{L}^{-1}$ biotin). Prior to use, the media was also supplemented with carbon and nitrogen sources. These included glucose, succinate, fumarate, malate, acetate, arabinose and pyruvate carbon sources and ammonium chloride, glutamate and glutamine nitrogen sources.

Luria-Bertani broth (LB) media was used as a rich-media growth source for *E. coli* strains. This consisted of: $10 \text{ g} \cdot \text{L}^{-1}$ tryptone, $5 \text{ g} \cdot \text{L}^{-1}$ yeast extract and $5 \text{ g} \cdot \text{L}^{-1}$ NaCl . $14 \text{ g} \cdot \text{L}^{-1}$ agar was added prior to autoclaving for solid medium.

Antibiotics were added to media in concentrations described in Section 2.6.2. Since Rlv3841 is a spontaneous streptomycin resistant strain, streptomycin at $500 \mu\text{g} \cdot \text{mL}^{-1}$ was used routinely for selection of Rlv3841.

In addition, X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) at $40 \mu\text{g} \cdot \text{mL}^{-1}$ was added to media for blue-white screening with *E. coli*.

2.6.2 Antibiotic working concentrations

Working concentrations of antibiotics used in this work. a: higher concentration used for pK19mobSacB integration selection. b: lower concentration used for liquid cultures.

Antibiotic	Antibiotic resuspended in	<i>R. leguminosarum</i> concentration	<i>E. coli</i> concentration
Ampicillin (Amp)	Water	100 $\mu\text{g} \cdot \text{mL}^{-1}$	100 $\mu\text{g} \cdot \text{mL}^{-1}$
Gentamycin (Gm)	HEPES 10 mM, pH 7.0	20 $\mu\text{g} \cdot \text{mL}^{-1}$	10 $\mu\text{g} \cdot \text{mL}^{-1}$
Kanamycin (Km)	Water	40/160 ^a $\mu\text{g} \cdot \text{mL}^{-1}$	20 $\mu\text{g} \cdot \text{mL}^{-1}$
Neomycin (Neo)	HEPES 10 mM, pH 7.0	80 $\mu\text{g} \cdot \text{mL}^{-1}$	
Nitrofurantoin (Nit)	DMF		5 $\mu\text{g} \cdot \text{mL}^{-1}$
Streptomycin (Strep)	HEPES 10 mM, pH 7.0	500 $\mu\text{g} \cdot \text{mL}^{-1}$	25 $\mu\text{g} \cdot \text{mL}^{-1}$
Spectinomycin (Spec)	HEPES 10 mM, pH 7.0	100 $\mu\text{g} \cdot \text{mL}^{-1}$	50 $\mu\text{g} \cdot \text{mL}^{-1}$
Tetracycline (Tet)	Methanol	5/1 ^b $\mu\text{g} \cdot \text{mL}^{-1}$	10 $\mu\text{g} \cdot \text{mL}^{-1}$

2.7 Bacterial growth conditions and assays

2.7.1 Bacterial growth conditions

All rhizobia grown at 28 °C in TY or UMS media. For liquid cultures, aeration was achieved through shaking at 200-225 rpm at 28°C. *E. coli* strains were grown at 37 °C in LB media. Liquid cultures were aerated by shaking at 150 rpm in 37 °C.

2.7.2 Generation time of rhizobial strains

To assess the growth rate of a strain, 24-well plates were inoculated at $\text{OD}_{600} = 0.01$ in 20 mM succinate and 10 mM NH_4Cl supplemented UMS media. The growth was determined by the increase in optical density over time as measured by a PlateReader.

2.7.3 Semi-solid agar swimming assay

To assess the swimming ability of strains under various media conditions, they were grown on 0.22% agar UMS media plates supplemented with various metabolisable nitrogen and carbon sources. These carbon and nitrogen compounds were all metabolites and therefore there is the potential for metabolism to generate tactic gradients within the plate. The inoculant was grown on TY slopes for 3 days

before being resuspended in 3-4 mL UMS and normalised to an optical density of $OD_{600} = 1.8$. 2 μ L of the culture was then inoculated into the centre of each plate, forming a column through the agar. The plates were incubated at 28 °C for 4-7 days. The diameter of the bacterial halo was then measured to the nearest mm. These halos are formed as a result of both motility and metabolism, with both increased swimming and growth producing larger halos. In addition, the 4-7 day timescale allows changes due to gene expression, potentially changing the presence or number of flagella and taxis proteins. The overall diameter is a summation of all of the above effects and represents semi-solid propagation of the bacteria.

2.7.4 Single cell tracking

To assess the distribution of swimming patterns within strain populations, individual cells were tracked. The inoculant was grown on TY slopes for 3 days before being resuspended in 3-4 mL UMS and diluted to an optical density of $OD_{600} = 0.1$ in 10 mL of UMS supplemented with 10 mM glucose and 10 mM NH_4Cl . These cultures were incubated overnight at 28 °C. Next, the cultures were diluted again to an optical density of $OD_{600} = 0.1$ in 10 mL of UMS supplemented with the tested carbon and nitrogen sources. These cultures were incubated for 3 hours at 28 °C, and staggered to keep the incubation time constant. Each culture was then drawn into a rectangular capillary, sealed at both ends with vacuum grease, and placed on a microscope slide. The cells were then visualised on a Meiji microscope with 40x lens and recorded using an Infinity lite camera (1440 x 1080 pixels at 10 fps). Four technical replicate images were taken along the length of the capillary, at the top surface of the capillary. The images were then run through an image analysis pipeline (github.com/AroneyS/Capillary-tracking). This pipeline produced individual tracks for each bacteria, calculating the overall swimming speed and the tumble rate (See Figure 2.1).

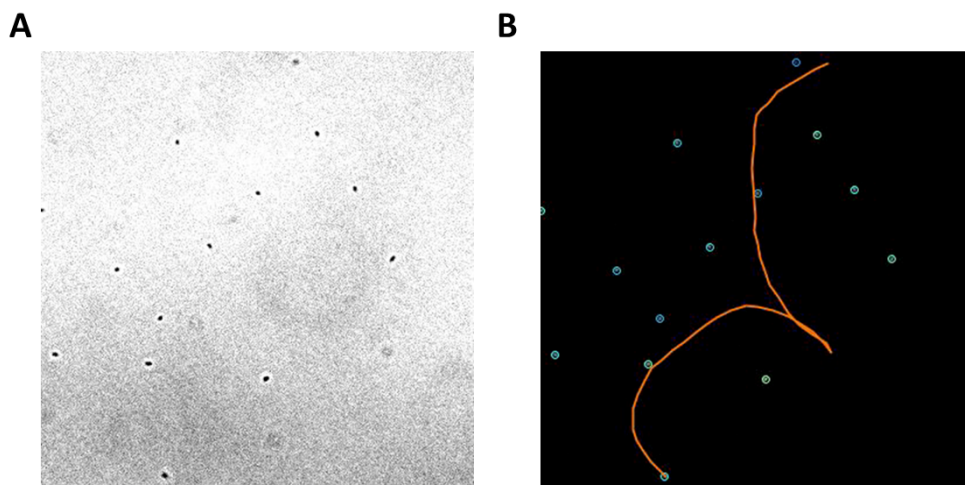


Figure 2.1: Bacterium swimming tracking. Each raw image (**A**) was processed to enable automated identification of individual bacteria (spots) that were then combined into tracks (line) as in **B**.

2.7.5 Flagella staining

Due to disruptions caused by the 2020 pandemic, the flagella staining experiments were unable to be completed.

These experiments would have determined the presence of flagella in the non-motile mutant strains (to confirm presence in the *motA* mutant and absence in the *fliG* mutant) and under high TCA-cycle intermediate conditions (to determine if the halting effect was due to lack of flagella).

The flagella staining solution consisted of $0.2 \text{ g} \cdot \text{mL}^{-1}$ Tannic acid solution in 5% phenol combined 1:1 with a saturated $\text{KAl}(\text{SO}_4)_2 \cdot 12\text{H}_2\text{O}$ solution and then together combined 10:1 with saturated crystal violet stain in ethanol ($0.12 \text{ g} \cdot \text{mL}^{-1}$) (Kodaka *et al.* 1982a; Heimbrook, Wang and Campbell 1989a).

As above, the bacteria were grown on TY slopes for 3 days before being resuspended in 3-4 mL UMS and diluted to an optical density of $\text{OD}_{600} = 0.1$ in 10 mL of UMS supplemented with 10 mM glucose and 10 mM NH_4Cl . These cultures were incubated overnight at 28°C . Next, the cultures were diluted again to an optical density of $\text{OD}_{600} = 0.1$ in 10 mL of UMS supplemented with the

tested carbon and nitrogen sources. These cultures were incubated for 3 hours at 28 °C, and staggered to keep the incubation time constant. Each culture was then spotted onto a microscope slide and dried. The slide was then flooded with the staining solution for 5 minutes before being washed thoroughly with tap water for 2-3 minutes. After drying, the presence of flagella was then scored by visualising on a microscope at x97 magnification.

2.8 Legume growth conditions

2.8.1 Germination of *Pisum sativum*

P. sativum cv. Aveola seeds were surface sterilised by immersing in 95% ethanol for 60 seconds then washing in sterile H₂O, then immersing in 2% sodium hypochlorite for 5 minutes. Seeds were then repeatedly washed in 400 mL sterile MilliQ H₂O to remove sterilisers. Seeds were germinated for 2 day by embedding in 0.8% water agar and leaving in dark at RT.

2.8.2 Germination of *Vicia villosa*

V. villosa seeds were pored with sandpaper until a change in seed coat colour was observed. The seeds were then surface sterilised by immersing in 2% sodium hypochlorite for 5 minutes. Seeds were then repeatedly washed in 400 mL sterile MilliQ H₂O to remove sterilisers. Seeds were left imbibing for 3-4 hours in a sealed falcon tube of sterile H₂O. Seeds were then germinated by embedding in 0.8% water agar, covering the plates in aluminium foil and left at 4 °C for 2 days, then at RT for a further 2 days.

2.8.3 *Pisum sativum* growth conditions

Pea was typically grown in 500 mL pots or boiling tubes. The pots were filled with either a 50:50 mixture of medium vermiculite and silver sand or only medium vermiculite. Both were supplemented with either 50 mL of 2.57x or 150 mL of 1x nitrogen-free rooting solution respectively (see Section 2.8.5). Pots were autoclaved prior to planting. The pots were planted with germinated seedlings and inoculated

with 1 mL of diluted bacterial cultures. Pots were then covered with cling-film to prevent bacterial contamination and grown in a controlled environment room. After 3-6 days, when the shoot had reached the cling-film, it was cut with a sterile scalpel. Any additional peas added as backups were removed after one week. The plants were watered with 50 mL of sterile H₂O weekly until harvest.

Boiling tubes were 3/4 filled with washed and dried fine vermiculite supplemented with 25 mL of nitrogen-free rooting solution (see Section 2.8.5). The tubes were sealed with foam plugs, then autoclaved prior to planting. Germinated seedlings were planted within the vermiculite and inoculated with 1 mL of diluted bacterial cultures. The tubes were then sealed again with the foam plug and grown in a controlled environment room until harvest.

2.8.4 *Vicia villosa* growth conditions

Vetch was typically grown in 10 cm square petri dishes containing sloped 1% agar FP media (Fahraeus 1957). The plates were planted with five germinated seedlings per plate and inoculated with 50 μ L of diluted bacterial cultures. Plates were then sealed with millipore tape to prevent bacterial contamination and covered in aluminium foil to block light on the roots. The plates were then grown in a tissue culture room until harvest.

2.8.5 Recipe for Nitrogen-free rooting solution

Nitrogen-free rooting solution was used to provide the nutrients for plant growth without a source of nitrogen to incentivise rhizobial symbiosis. This consisted of: 4 mM Na₂HPO₄, 3.7 mM K₂PO₄, 1 mM CaCl₂, 800 μ M MgSO₄, 100 μ M KCl, 35 μ M H₃BO₃, 10 μ M Fe EDTA, 9 μ M MnCl₂, 0.8 μ M ZnCl₂, 0.5 μ M Na₂MoO₄, 0.3 μ M CuSO₄.

2.8.6 Plant stain for enzyme activity

To differentiate nodules colonised by different strains on the same plant, the bacteria were marked with two enzyme genes to enable differential staining. Firstly, the gene

gusA encoding a β -glucuronidase which produces the colour magenta with addition of Magenta-GlcA (5-bromo-6-chloro-3-indolyl- β -D-glucuronide). Magenta-GlcA stock solution was made with 0.2 g of Magenta-GlcA dissolved in 10 mL of dimethylformamide (DMF) and stored, covered in aluminium foil, at -40 °C. Secondly, the gene *celB* encoding a thermostable β -galactosidase which produces the colour blue with addition of X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactoside). X-Gal stock solution was made with 0.2 g of X-Gal dissolved in 10 mL of dimethylformamide (DMF) and stored, covered in aluminium foil, at -40 °C. The plants were stained sequentially, as detailed below, in a phosphate buffer consisting of 7 g/L NaH₂PO₄, 7.2 g/L Na₂HPO₄, 1 mM EDTA (pH 8), 1% Sarkosyl and 1 mL/L Triton.

Firstly, the plant roots were cut after the seed and at the end of the roots after the last nodules. For β -glucuronidase activity, the roots were submerged in a 50 mL Falcon tube containing 40 mL phosphate buffer supplemented with 400 μ L of Magenta-GlcA solution. The tubes were then covered with aluminium foil and incubated at 28 °C overnight, producing magenta nodules. The buffer was then refreshed and the tubes were incubated at 70 °C for 1 hour to destroy endogenous β -galactosidases. For β -galactosidase activity, after cooling, 500 μ L of X-Gal solution was added to each tube. The tubes were then incubated at 37 °C overnight, producing blue nodules. Finally, the roots were rinsed with diluted bleach (1:1 dilution in water) for 10-15 minutes and stored in water at 4 °C.

2.9 Computational analysis

2.9.1 *In silico* cloning

All cloning reactions and genome integrations were designed *in silico* and confirmed by sequence alignment (see Section 2.3.5) using Geneious R10.

2.9.2 Bioinformatic analysis

Global nucleotide and protein alignments were carried out using BLASTn and BLASTp (NCBI), respectively. Local alignments, against references from the

NCBI database, were carried out with Geneious R10 using the Geneious alignment algorithm.

2.9.3 Data-handling and statistical analysis

Data handling, statistical analysis and graph generation was carried out in R (R Core Team 2019). The following packages were used: tidyverse (Wickham *et al.* 2019), cowplot (Wilke 2019), lme4 (Bates *et al.* 2015), lmerTest (Kuznetsova, Brockhoff and Christensen 2017), emmeans (Lenth 2020), ggeffects (Lüdecke 2018), flexmix (Leisch 2004; Grün and Leisch 2007, 2008), agricolae (de Mendiburu 2020), sjPlot (Lüdecke 2020), gridExtra (Auguie 2017), ggsignif (Ahlmann-Eltze 2019) and scales (Wickham and Seidel 2020).

A model of sensory taxis in *Rhizobium leguminosarum*

3.1 Introduction

Post-translational control of flagella-based motility in bacteria occurs through chemotaxis systems. These systems control flagella by biasing the tumbling frequency, causing changes in swimming direction. This control occurs via binding of phosphorylated CheY to the FlhMN switch complex (see Chapter 1). As described in Section 1.5, *Rhizobium leguminosarum* has two different chemotaxis clusters which drive movement of a single flagellum.

The Che1 cluster is the main controller of free swimming behaviour and, while the Che2 cluster has a minor impact on regulation of the Che1 cluster, the role of the Che2 cluster is unknown (Miller *et al.* 2007). These clusters contain the only three *cheY* genes found in the Rlv3841 genome by BLAST. Rlv3841 has a total of 25 chemoreceptors but only McpB and McpC have been characterised (Yost, Rochepeau and Hynes 1998; Yost *et al.* 2004). An *mcpB* mutant had inhibited chemotaxis towards a wide range of compounds, while both mutants had greatly reduced nodule formation in competition against wildtype.

To test the sensory systems of *Rhizobium leguminosarum* and probe their effects on flagella-based motility, a number of mutants of Rlv3841 were characterised

in this work (Table 3.1). Firstly, three mutants in flagellar critical genes were isolated as a baseline for non-motility. Secondly, three motile chemotaxis mutants were obtained from Michael Hynes (Hynes lab, University of Calgary) to test their individual impacts (Miller *et al.* 2007). With all these mutants, we defined the motility parameters and investigated the chemotaxis response of *R. leguminosarum* under various metabolic and growth conditions. In addition, we also searched for potential chemoreceptors and conducted a bioinformatic classification, characterising experimentally those that appeared to sense internally.

Table 3.1: The mutants analysed in Chapter 3.

Strain name	Genes affected	Source
OPS2107	<i>motA</i>	This work
OPS2163	<i>fliG</i>	This work
OPS2234	<i>flaA</i>	This work
LM100	<i>che1</i> cluster	(Miller <i>et al.</i> 2007)
LM400	<i>che2</i> cluster	(Miller <i>et al.</i> 2007)
LM300	<i>che1</i> and <i>che2</i> clusters	(Miller <i>et al.</i> 2007)
OPS1881	<i>mcpQ</i>	This work
OPS1505	<i>mcpY2</i>	This work
OPS1500	<i>hemAT</i>	This work
OPS2108	<i>icpA</i>	This work
OPS2288	<i>hemAT</i> and <i>icpA</i>	This work

3.2 Starvation response in *Rlv3841* vs *Sme2011*

We began the investigation by monitoring wildtype swimming levels in semi-solid agar ‘swimming plates,’ across a wide range of carbon sources in minimal media supplemented with 10 mM ammonium and different carbon sources at low and high concentrations (10-fold difference) (Figure 3.1). These carbon and nitrogen compounds were all metabolites and therefore there is the potential for metabolism to generate tactic gradients within the plate. The wildtype strains chosen were *Rhizobium leguminosarum* biovar *viciae* 3841 (*Rlv3841*) and a well-studied relative, *Sinorhizobium meliloti* Rm2011 (*Sme2011*).

Firstly, *Sme2011* swam further than or equal to *Rlv3841* under all conditions tested. This is likely to be due to *Sme2011* having increased swimming speed,

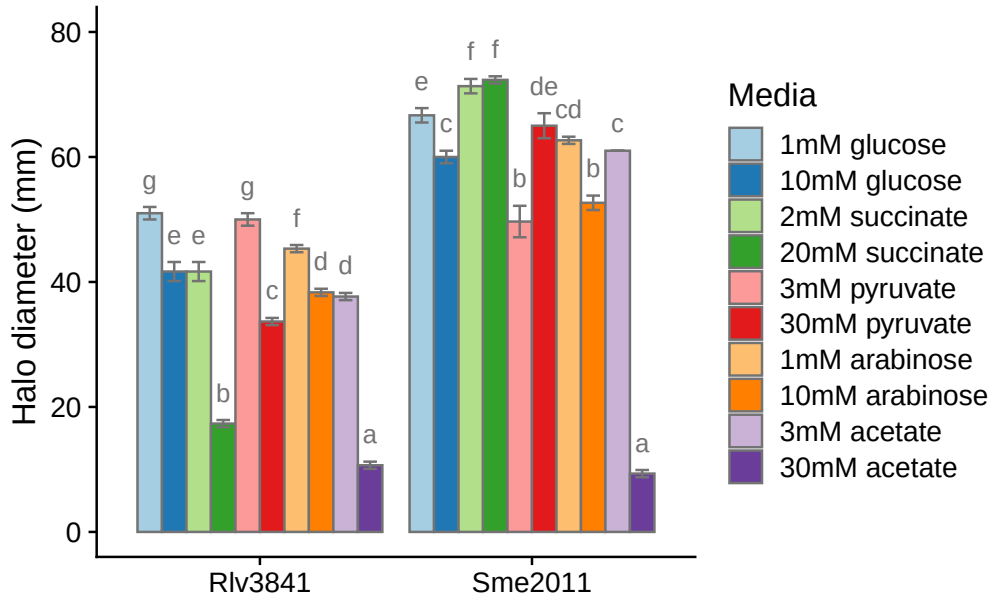


Figure 3.1: Wildtype rhizobia swimming in semi-solid agar plates. Swimming halo diameters for *Rhizobium leguminosarum* biovar *viciae* 3841 (Rlv3841) vs *Sinorhizobium meliloti* Rm2011 (Sme2011). Data was modelled with a Gaussian linear model with the factors strain, carbon source, carbon concentration and their interactions. All factors were found to be significant, $P < 1e-13$, ANOVA. Tukey's Post-hoc test was used to stratify data within strains into groups, where those groups not sharing a letter were significantly different at $\alpha=0.05$. Bars are mean $\pm$ SD, $N=3$.

although this cannot be confirmed in this type of assay. On these plates, an optimum halo diameter is formed at a tumbling rate high enough to efficiently navigate the agar matrix but low enough to allow sufficient swimming (Mohari *et al.* 2015). For a strain and condition, the halo diameter can be considered a measure of discontentment or unfavourability. Meaning that the higher the halo diameter, the less content the bacteria are with their current environment. This arises from the idea that chemotaxis increases the tumble rate under low attractant or high repellent conditions.

Rlv3841 displayed reduced halo diameters at higher carbon concentration across all compounds. Of these, the largest effects were on succinate and acetate, with weak effects on sugars. On inspection of the halos, there appears to be a metabolic effect occurring on acetate. In this carbon source, the halos were very small and faint, indicating a low rate of growth. In comparison, although on high succinate the halos were small, Rlv3841 formed a thick, clumped halo. Sme2011 also displayed

a weak effect under glucose and arabinose and, like Rlv3841, showed a strong effect on acetate. However, there was no change for succinate and a reversed response for pyruvate. In high acetate, Sme2011 also displayed small, faint halos. Interestingly, Sme2011 appears to be blind to succinate, with no change in swimming response. This may be explained by the presence of a chemoreceptor in Rlv3841 that is not present in Sme2011. Indeed, Rlv3841 encodes 25 different receptors, whereas Sme2011 has only 8 (Zatakia *et al.* 2017).

3.3 Isolation and characterisation of motility mutants

To test the phenotypes of flagellar mutants, three different in-frame deletion mutants were isolated from Rlv3841. Firstly, *motA* (RL0703), encoding a critical motor stator protein, was deleted to produce a strain with non-rotating flagella. Secondly, *fliG* (RL0700), encoding a critical motor rotor protein, was deleted to produce a non-flagellated strain. And finally, *flaA* (RL0718), encoding a flagellar filament protein, was deleted to produce a strain unable to form stable flagellar filaments. Plans to confirm the flagellation status of these mutants were not completed due to the 2020 Pandemic lock-downs.

To isolate these mutants, double crossover mutations were made using a pK19-mobsacB strategy, as described in Section 2.5.2. This produced plasmids pOPS1412, pOPS1422 and pOPS1423. After single crossover, integration was mapped using primers *oxp3919+oxp3920*, *oxp3937+oxp3938* and *oxp3925+oxp3926* and double crossover was selected for using 10% sucrose. Deletion of each gene was confirmed by mapping using primers *oxp4055+oxp4056*, *oxp4063+oxp4064* and *oxp4059+oxp4060* and sequencing the amplified fragments. These strains were named OPS2107 (Δ *motA*), OPS2163 (Δ *fliG*) and OPS2234 (Δ *flaA*).

The swimming ability was first tested in swimming plates, across a wide range of carbon sources in minimal media. All of the *motA*, *fliG* and *flaA* mutants displayed significantly reduced swimming diameters across all media conditions (Figure 3.2), confirming that they were all affected in motility. Indeed the expected

non-flagellated and non-rotating flagella producing strains *motA* and *fliG* appeared to form only growth-driven halos, indicating a small role under these conditions for the presumably intact pili system.

3.4 Motility of chemotaxis-gutted mutants to form a negative control for the absence of chemotaxis

To test the sensory systems in *Rhizobium leguminosarum*, three mutants in Rlv3841 were obtained from Michael Hynes' laboratory: LM100 ($\Delta che1$, RL0685-94), LM400 ($\Delta che2$, RL4028-36) and LM300 ($\Delta che1che2$). The main chemotaxis system is *che1*, with only a minor effect on tumbling rate attributed to *che2* (Miller *et al.* 2007). All three strains were found to be motile by testing their swimming ability in minimal media swimming plates with a variety of carbon sources at two different concentrations and also an additional test with a nitrogen-poor (glutamate) and a nitrogen-rich (glutamine) source, with high glucose (10 mM) as carbon source (Figure 3.3).

Wildtype Rlv3841 displayed significant differences between high and low concentrations of glucose, succinate, pyruvate, arabinose and acetate. In addition, we also discovered the same effect in fumarate and malate media. One confounding factor was revealed by the glutamate vs glutamine data. Although there were significant differences between wildtype swimming in glutamate vs glutamine media in earlier experiments, this was not a reproducible difference. This might be explained by the residual nitrogen present in the agar itself varying across batches used in these experimental replicates. Although the agar concentration was fairly low (0.22%), differences in nitrogen content per batch may have masked the swimming effect of varying amino acid sources. To avoid this effect, agarose was tested as a nitrogen-free alternative to agar. However agarose has different gelling properties and produced inconsistent results. If there is an impact on swimming between glutamate and glutamine, it is a much smaller effect than that due to carbon sources.

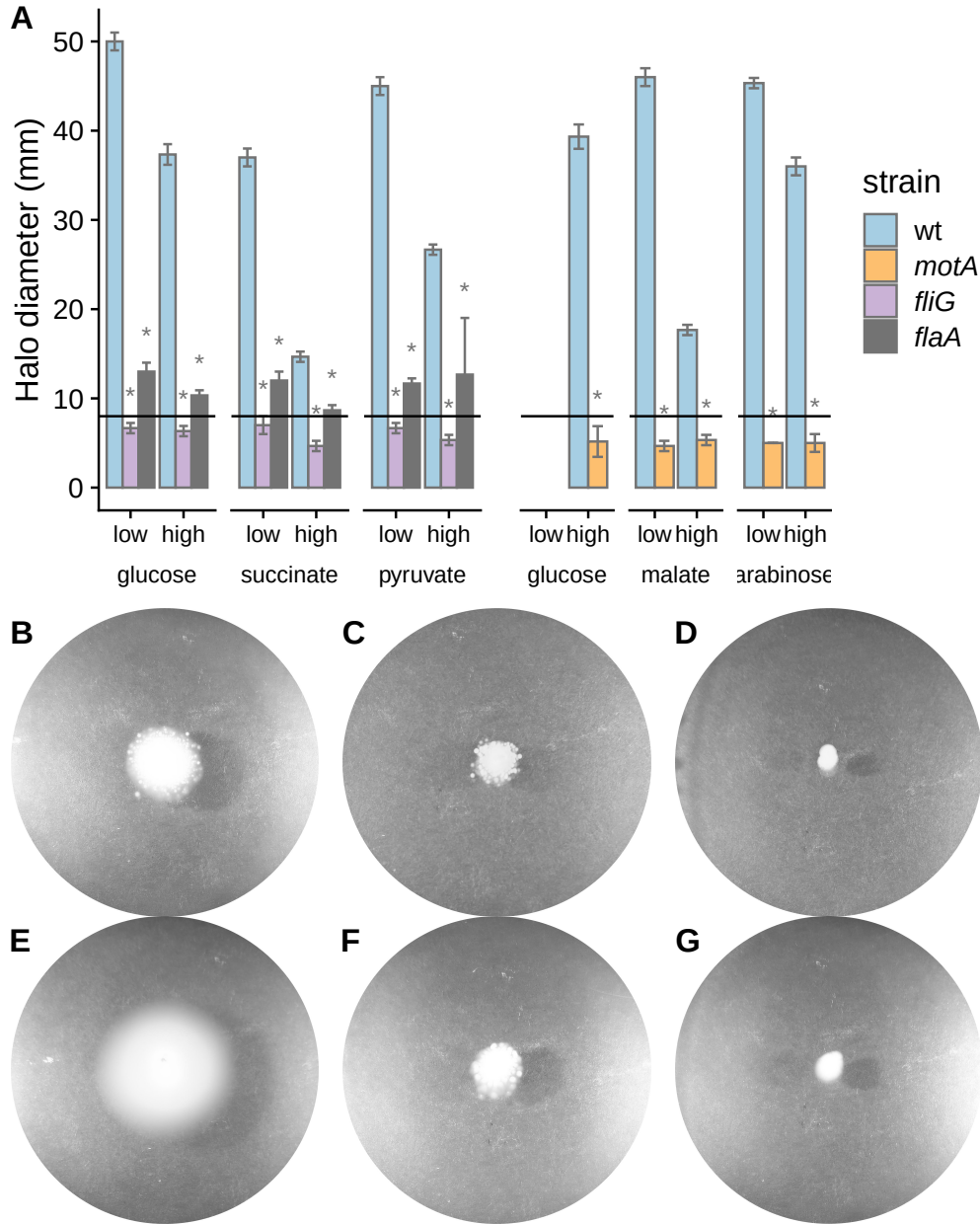


Figure 3.2: Motility mutant swimming in semi-solid plates. Swimming halo diameters of wt (Rlv3841) vs *motA* (OPS2107), *fliG* (OPS2163) and *flaA* (OPS2234). **A** Data was modelled with a Gaussian linear model with the factors strain, carbon source, carbon concentration and their interactions. All parameters were found to be significant, $P < 1e-10$, ANOVA. Dunnett's Post-hoc test found means of *motA*, *fliG* and *flaA* to be significantly different from wt under all media conditions with * $P < 0.0001$. Bars are mean $\pm$ SD, $N=3$, horizontal line represents estimated maximum non-motile diameter. **B** Representative photo of wt on high succinate. **C** Representative photo of *flaA* on high succinate. **D** Representative photo of *fliG* on high succinate. **E** Representative photo of wt on high pyruvate. **F** Representative photo of *flaA* on high pyruvate. **G** Representative photo of *fliG* on high pyruvate.

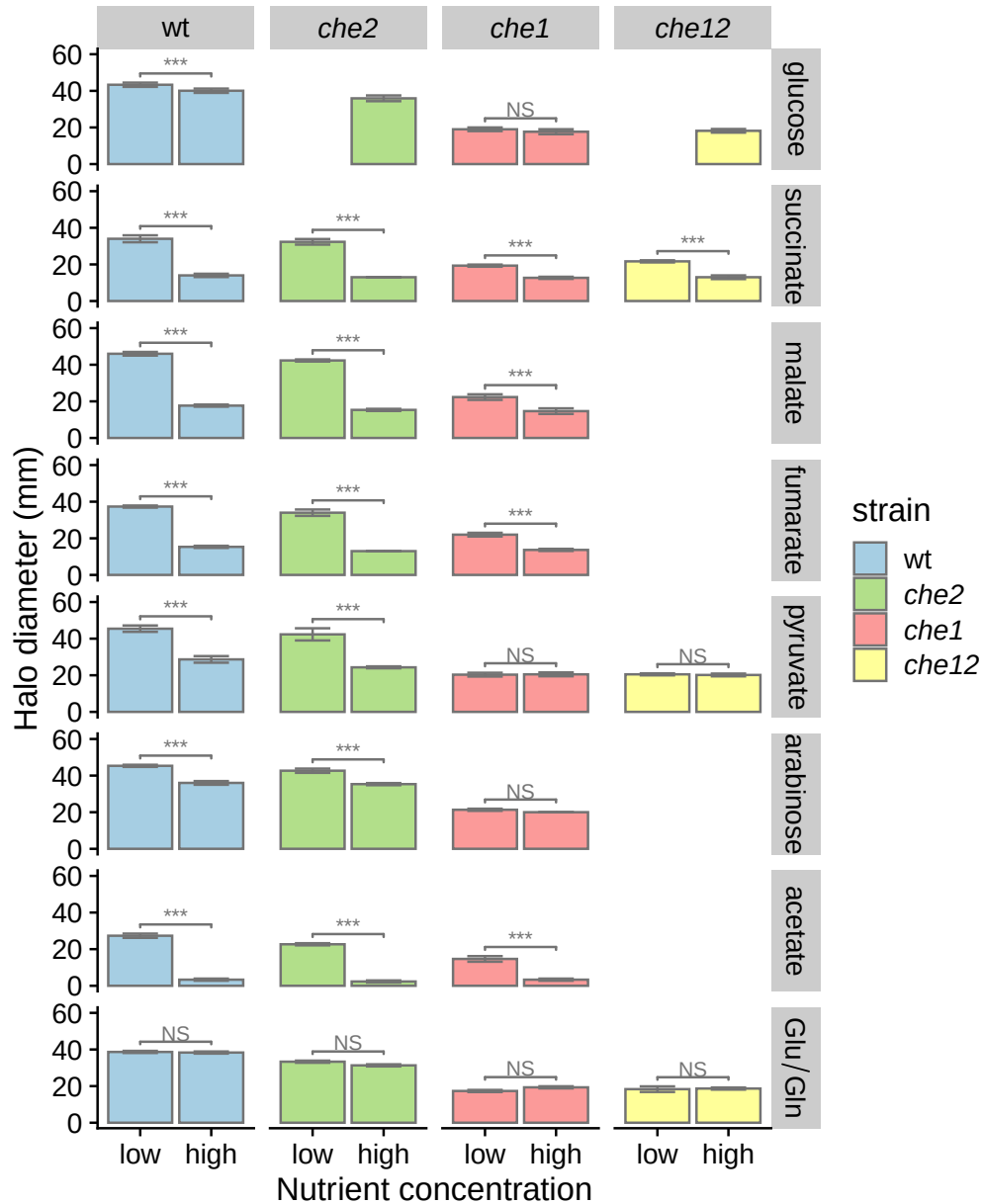


Figure 3.3: Chemotactic mutant swimming in semi-solid plates. Swimming halo diameters of wt (Rlv3841) vs *che1* (LM100), *che2* (LM400) and *che1,2* (LM300) on minimal media with high or low nutrient concentration as follows: glucose (10 mM vs 1 mM), succinate (20 mM vs 2 mM), malate (20 mM vs 2 mM), fumarate (20 mM vs 2 mM), pyruvate (30 mM vs 3 mM), arabinose (10 mM vs 1 mM), acetate (30 mM vs 5 mM), Glu/Gln (10 mM glutamine vs 10 mM glutamate). Data was modelled with a Gaussian linear model with the factors strain, nutrient, nutrient concentration, and their interactions. All factors were found to be significant, $P < 1e-29$, ANOVA. Dunnett's Post-hoc test found mean difference of nutrient concentration within strain and nutrient to be significantly different from 0. Significance levels are NS $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Bars are mean $\pm$ SD, N=3.

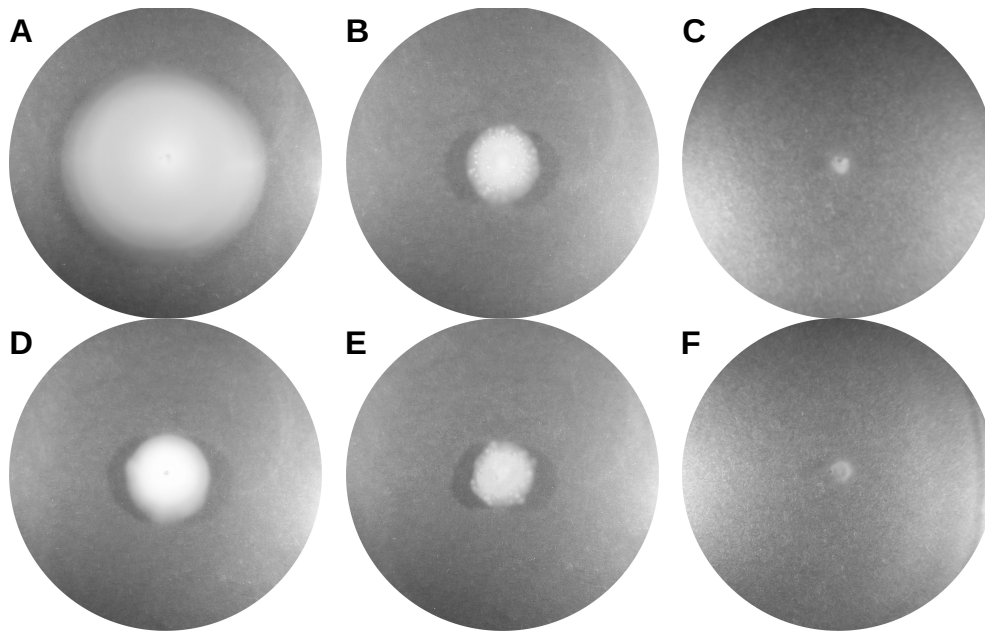


Figure 3.4: Photos of chemotactic mutant swimming in semi-solid plates. Swimming halos of wt (Rlv3841) vs *che1* (LM100), *che2* (LM400) and *che1,2* (LM300) from Figure 3.3. **A** Representative photo of wt on high glucose. **B** Representative photo of wt on high succinate. **C** Representative photo of wt on high acetate. **D** Representative photo of *che1* on high glucose. **E** Representative photo of *che1* on high succinate. **F** Representative photo of *che1* on high acetate.

Among the mutants tested, *che2* displayed reductions in swimming halo diameter under the same media conditions as wildtype. However, it also had significantly reduced diameter over wildtype on all low-concentration carbon sources, high-concentration pyruvate, glutamate and glutamine. Although significant, these reductions were minor in comparison to the *che1* mutant, reinforcing Che1 as the main chemotaxis system. The *che1* and *che1,2* mutants had comparable results where tested. This was a reduction in swimming diameter over wildtype in all conditions except high succinate, fumarate and acetate. In addition, there was no significant difference found between high and low concentrations of glucose, pyruvate and arabinose or between glutamate and glutamine. This agrees with above that the chemotaxis cluster encoded by *che1* is responsible for the majority of chemotactic response (i.e. for glucose, pyruvate and arabinose). Interestingly, deletion of *che1* did not remove the difference between high and low concentration of the TCA-cycle intermediates succinate, malate and fumarate or for acetate. Indeed, this reduction

brought each strain to the same diameter as wildtype (except for a minor, but significant, reduction of the *che1* mutant vs wildtype on high-concentration malate).

Swimming on high acetate matched that observed above on wildtype Rlv3841 and Sme2011, producing small, faint halos indicative of low growth rate (Figure 3.4C and F). This is thus independent of chemotaxis, as would be expected of a growth-rate dependent response. However, all halos of high succinate, malate and fumarate were again thick and clumped (Figure 3.4B and E). We theorised that this is a chemotaxis-independent effect. The shape and structure of the halos can be explained by a system that shuts down motility or causes non-flagellation under high TCA-cycle intermediate concentrations. The structure is then formed by high consumption of the nutrient source in the centre of the plate, leading to its exhaustion and causing motility to be reactivated or flagella resynthesised until swimming takes the bacteria back to a higher concentration. Growth from those positions reached by the temporarily swimming bacteria causes the clumping. It should be noted that high concentrations of these metabolites (20-30 mM) may be saturating responses and thus oversimplifying any relationships between the concentration and the halo diameter.

To confirm that this result was independent of growth rate, the swimming behaviour of individual bacteria were tracked via image analysis of microscopy videos. This growth-independence was considered very likely since the media conditions expected to have higher growth rate (high concentrations) had smaller swimming halos.

3.5 Chemotaxis-gutted mutant still responds to succinate

To study the effect of the TCA-cycle intermediates, the swimming behaviour of individual wildtype (Rlv3841) and the *che1* cluster mutant (LM100) bacteria were initially tested in minimal media (UMS with 10 mM NH_4Cl) supplemented with various succinate concentrations (20 mM as high and 2 mM as low). If a succinate concentration dependent effect was found on the swimming of individual bacteria,

there would be strong evidence that the effect is growth-independent. The *che1* cluster mutant was tested to confirm chemotaxis-independence of the succinate concentration effect. The assay was performed as detailed in Section 2.7.4. Tracking bacteria on surfaces has previously been found to cause some artefacts: the tendency to increased track curvature, hydrodynamic attraction reducing tumbling off the surface and increased effect of Brownian motion (Frymier *et al.* 1995; Berke *et al.* 2008; Li, Tam and Tang 2008). Despite this, tracking on the surface was chosen to prevent tumbles taking the cells out of the 2D plane of focus being recorded, which would otherwise lead to an abundance of short tracks with biased tumble rate. In addition, the surface was found to have an increased swimming population and acts as an approximation to root surface swimming.

Initial analysis of the motile fraction of cells revealed significant differences between low and high succinate conditions (Figure 3.5A). This was calculated as the overall motility of the video (recorded time of motile spots / total time of spots), to avoid bias from counting motile cells multiple times. Both wildtype and *che1* had very similar effects, having significantly higher motile fractions in low succinate. This indicates that there is a succinate dependent effect on motility that is independent of the presence or absence of Che system chemotaxis. Indeed the strain factor was not found to be significant (ANOVA). This data agrees with the hypothesis described above (Section 3.4), that high succinate concentration is triggering a low-motility state, through halting or deflagellation.

The overall tumble rate for each replicate was also calculated (Figure 3.5B). This data describes both a change by strain (reduced tumble rate for the *che1* mutant) and by media (reduced tumble rate in high succinate). The data for wildtype at low succinate had high variability, so all its comparisons were non-significant. To avoid this discrepancy, more replicates are needed. Otherwise, the data shows significant differences between the *che1* mutant at both media conditions and between wildtype and the *che1* mutant at high succinate. This indicates that the *che1* mutation does impact tumble-rate, reducing it as expected. However, there is a change in tumble-rate between media conditions that is not expected. Tumble rate is thought

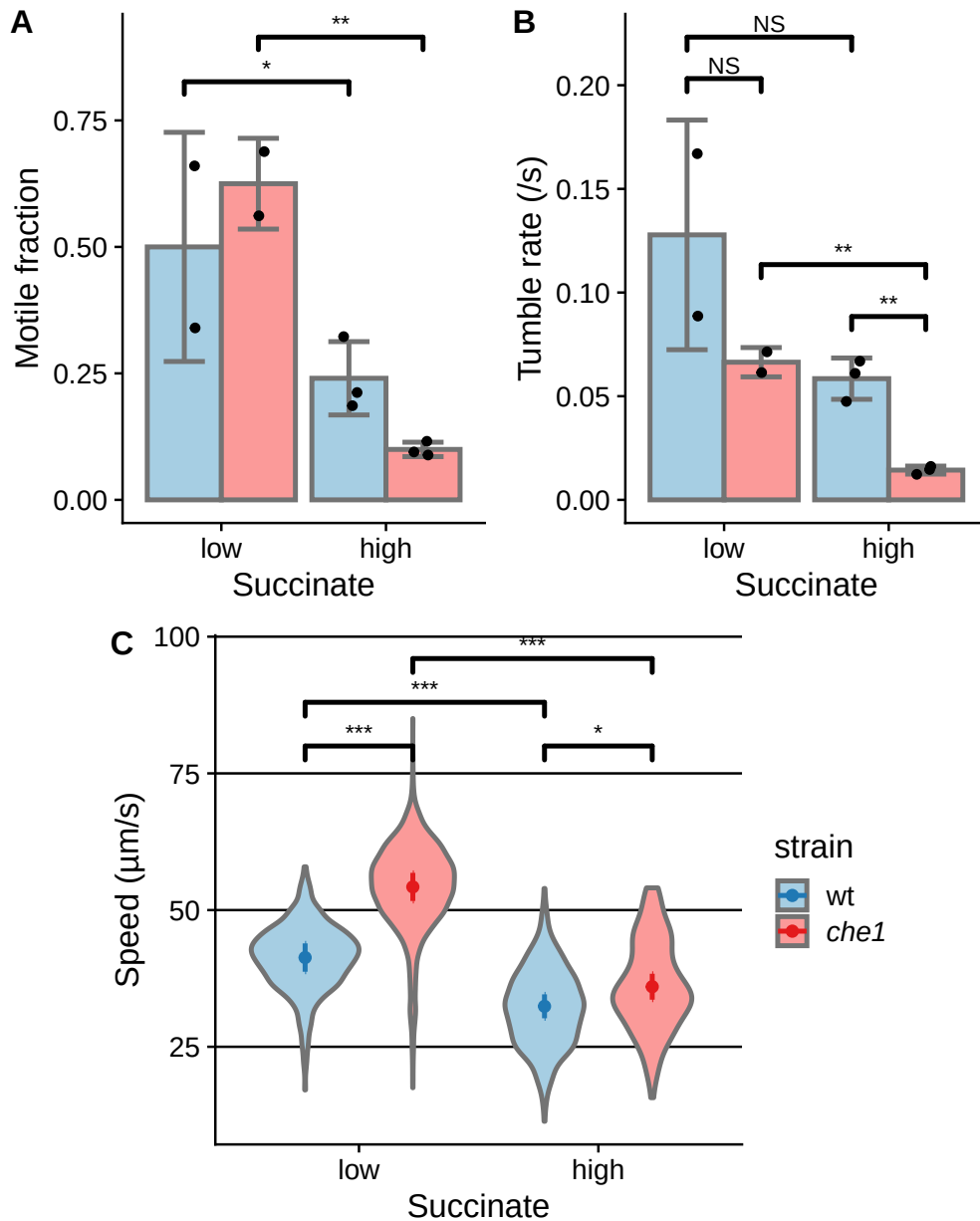


Figure 3.5: Wildtype and chemotaxis mutant swimming tracks. Bacterium tracking analysis were carried out in capillaries with liquid cultures grown in minimal media (UMS) with 2mM or 20mM succinate of wt (Rlv3841) and *che1* (LM100). The analysis from each replicate was done from 160-1,700 tracks per strain, N=2-3. **A** Fractions of motile tracks. Data was modelled with a Gaussian linear model with the factors strain and media and their interaction. Only media was significant, $P < 0.01$, ANOVA. Dunnett's Post-hoc test found differences between levels (brackets). **B** Average tumble rate of motile bacteria. Data was modelled as in A, using generalised least squares fit to model unequal variance. All factors significant, $P < 0.01$, ANOVA. Dunnett's Post-hoc test found differences between levels (brackets). **C** Average swimming speed of motile bacteria. Data was modelled as above, with random intercept by technical replicates. All factors were significant, $P < 0.05$, ANOVA. Dunnett's Post-hoc test found differences between levels (brackets). Bars are mean $\pm$ SD, density is across all individual tracks and points are mean $\pm$ 95% CI, N=2-3. Indicated significance levels are NS $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, insignificant differences were not indicated.

to mostly be a result of the chemotaxis system. So, we would expect the *che1* mutant's tumble rate to be very low and consistent across media conditions. This discrepancy may be due to the Che2 system that is still present in that mutant (investigated in Section 4). Alternatively, it could be due to changes in the proton-motive force driving motility due to succinate increasing the oxygen consumption through the electron-transport chain. In addition, it may be due to a small number of tracks that were found to have high tumble rates. These tracks tended to have unrealistic trajectories and were likely an artefact. They seemed to generally be formed by the tracking software combining motile and non-motile tracks and thus would have a higher expected presence in low succinate with a higher motile fraction. Therefore the data generated from these assays was obtained by tracking first the non-motile spots and removing them from later analysis. Although this greatly reduced the impact of these tracks and reduced the variability between replicates, some were still present, causing a baseline tumble rate.

Finally, the average swimming speeds were analysed (Figure 3.5C). Since this was a continuous measurement, the individual bacteria could be plotted. As for the tumble rate, the data describes both a change by strain (increased speed for the *che1* mutant) and by media (increased speed in low succinate). The difference between wildtype and the *che1* mutant may be explained by the reduction in tumble rate, since the bacteria slow down and stop with each tumble, and this timing was not removed from the speed calculations. However, the media effect cannot be explained in this way, since the bacteria are both swimming faster and tumbling more often in low succinate. This effect may be due to the same mechanism that causes the succinate effect on the motile fraction. The bacteria are both swimming less often and swimming at slower speed in high succinate. This effect also clearly occurs for the *che1* mutant bacteria.

The succinate-dependent control of motility appears to be independent of both growth-rate and the chemotaxis cluster 1. In Section 4, I discuss motility tracking for a wider range of mutants under more media conditions. To check whether this

effect was due to chemoreceptors signalling via a different pathway, each receptor was bioinformatically analysed as reported in the next section.

3.6 The chemoreceptors of *Rhizobium leguminosarum* biovar *viciae* 3841

The chemoreceptors, or methyl-accepting chemotaxis proteins (MCPs), of *Rhizobium leguminosarum* biovar *viciae* 3841 were bioinformatically analysed. This involved searching the genome for MCP signal domains, aligning all the candidate proteins and generating a phylogenetic tree as displayed in Figure 3.6. The MCP signal domain was chosen due to its high conservation across chemoreceptors, allowing easy identification. These MCP domains represent the protein segment responsible for CheA/CheW binding. The phylogenetic analysis compared the 27 candidate chemoreceptors from Rlv3841 to eight chemoreceptors from the well-studied *Sinorhizobium meliloti* RU11/001.

In addition to the MCP signal domain, most chemoreceptors contained one or more of the following domains: HAMP domains, acting as signal transmission from sensor domain to the signalling domain (Salah Ud-Din and Roujeinikova 2017); sensory domains, highly varied and including a number of different classes (e.g. Cache, PAS, protoglobin); transmembrane domains, typically present in a pair and surrounding the sensory domain, which causes most sensory domains to sense in the periplasm, with the remaining sensors being detached from the membrane and sensing in the cytoplasm.

Table 3.2: The chemoreceptors of Rlv3841. a: amino acids and small carbohydrates.

Gene number	Gene name	Sensory domain	Sensor type	Predicted sensing
RL0426	<i>mcpX</i>	dCache1	Periplasm	Quaternary ammonium
RL0429	<i>hemAT</i>	Protoglobin	Cytoplasm	Oxygen
RL0455	<i>mcpT</i>	Unknown	Periplasm	-
RL0564	<i>mcpE</i>	Unknown	Periplasm	-
RL0685	<i>icpA</i>	Protoglobin	Cytoplasm	Unknown
RL0757	<i>mcpZ</i>	Unknown	Periplasm	-
RL0758	<i>mcpI</i>	Cache3-Cache2	Periplasm	AA and small carbohydrates ^a
RL0949	<i>mcpJ</i>	sCache2	Periplasm	AA and small carbohydrates ^a

Gene number	Gene name	Sensory domain	Sensor type	Predicted sensing
RL0972	<i>mcpK</i>	CHASE3	Periplasm	Small peptides
RL1065	<i>mcpL</i>	4HB	Periplasm	Unknown
RL1318	<i>mcpM</i>	dCache1	Periplasm	AA and small carbohydrates ^a
RL1386		N/A	N/A	N/A
RL1447		N/A	N/A	N/A
RL2683	<i>mcpD</i>	Unknown	Periplasm	-
RL2931	<i>mcpN</i>	sCache2	Periplasm	Short-chain carboxylates
RL3985	<i>mcpO</i>	dCache1	Periplasm	AA and small carbohydrates ^a
RL4031	<i>mcrA</i>	4HB	Periplasm	Unknown
RL4032	<i>mcrB</i>	4HB	Periplasm	Unknown
RL4033	<i>mcrC</i>	4HB	Periplasm	Unknown
RL4277	<i>mcpP</i>	Unknown	Periplasm	-
RL4387	<i>mcpQ</i>	PAS	Cytoplasm	Oxygen, redox, light
pRL80031	<i>mcpS</i>	sCache2	Periplasm	AA and small carbohydrates ^a
pRL100403	<i>mcpG</i>	4HB	Periplasm	Unknown
pRL120056	<i>mcpR</i>	Unknown	Periplasm	-
pRL120068	<i>mcpY2</i>	PAS	Cytoplasm	Oxygen, redox, light
pRL120312	<i>mcpC</i>	4HB	Periplasm	Unknown
pRL120683	<i>mcpB</i>	HBM	Periplasm	Organic acids

The Rlv3841 genome consists of seven DNA molecules, a large chromosome and six different plasmids (Young *et al.* 2006). Chemoreceptor candidate genes were found distributed across the entire genome, with 21 in the main chromosome, one in pRL8, one in pRL10 and four in pRL12 (see Table 3.2). Interestingly, one chemoreceptor was located within the *che1* cluster in the chromosome (RL0685), and three were located within the *che2* cluster in the chromosome (RL4031-4033).

3.6.1 MCP heptad classes

In addition to the domain classification, chemoreceptors can be categorised by heptad counts. The heptad is a repeated 7-amino acid structure that is mirrored from the centre of the MCP signal domain upstream to the HAMP and transmembrane domains, and downstream to the C-terminus. All Rlv3841 chemoreceptors are class 36H (consist of 36 heptads) except McrA, McrB and McrC (within the *che2* cluster) which are class 34H, missing two heptads in the flexible bundle subdomain (Alexander and Zhulin 2007; Ortega and Zhulin 2018). This matches

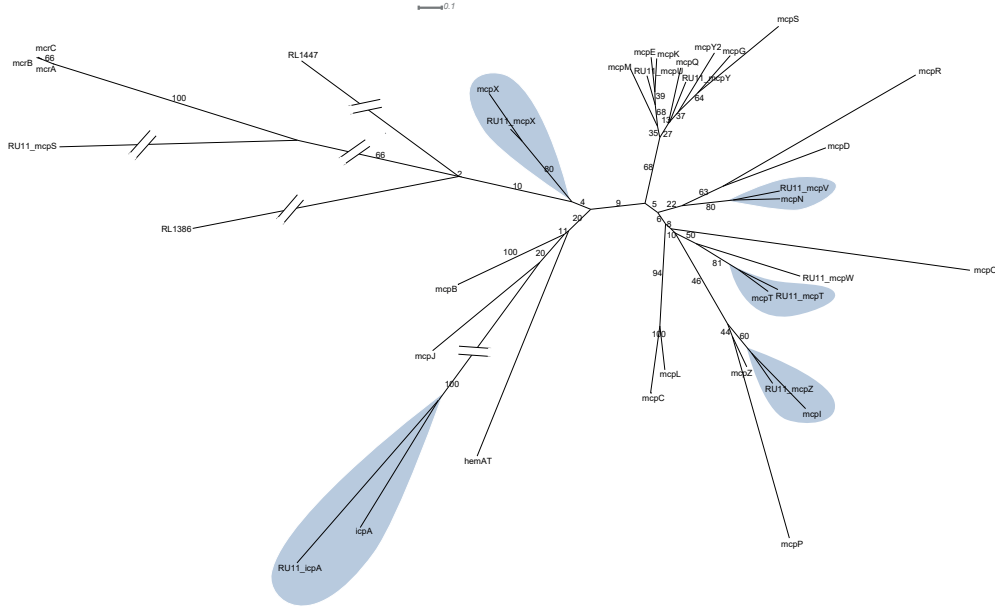


Figure 3.6: Maximum likelihood phylogenetic tree of chemoreceptor sequences from Rlv3841 and *Sinorhizobium meliloti* RU11/001 based on MCP signal domain region. Orthologous groups of sequences are highlighted. The tree was generated following a standard MCP analysis pipeline (Ortega and Zhulin 2018) using the following tools: MAFFT 7.407, RAxML 8.2.12, HMMER 3.2.1, Jalview 2.10.5 and Dendroscope 3.5.10 (Waterhouse *et al.* 2009; Eddy 2011; Huson and Scornavacca 2012; Katoh and Standley 2013; Stamatakis 2014).

the chemoreceptors in *S. meliloti* RU11, which were all class 36H, save McpS (within a secondary chemotaxis cluster) which is class 40H (Alexander and Zhulin 2007). These classes mainly seem to indicate evolutionary history, with no particular functional difference between the 34H and 36H classes.

3.6.2 Methylation

The final area for general MCP analysis is a description of methylation residues. These residues are methylated by CheR, demethylated by CheB and deamidated by CheD to finely tune the sensitivity of the receptors. The methylation zones of most chemoreceptors follow the consensus sequence of (3.1) below, where the highlighted residue is the methylation target (Salah Ud-Din and Roujeinikova 2017). This notation is matched in the putative methylation residues described below.

$$[ASTG][ASTG]xx[EQ][\mathbf{EQ}]xx[ASTG][ASTG] \quad (3.1)$$

N-terminal methylation area

The first zone is in the N-terminal side of the MCP signal domain from heptads N14 to N11. In the well studied *E. coli* Tsr chemoreceptors, the sequence across these heptads is (3.2)A. The highlighted residues are experimentally confirmed methylation sites (Rice and Dahlquist 1991). Most of the Rlv3841 chemoreceptors instead match the consensus motif of (3.2)B. Where residues are highlighted to match the *E. coli* methylation residue.

$$\begin{array}{l} \text{A)} T \ E \ \mathbf{Q} \ Q \ A \ A \ S \ L \ E \ \mathbf{E} \ T \ A \ A \ S \ M \ E \ \mathbf{Q} \ L \ T \ A \ T \ V \\ \text{B)} T \ E \ \mathbf{Q} \ Q \ A \ A \ S \ L \ E \ \mathbf{E} \ T \ A \ A \ A \ L \ E \ \mathbf{Q} \ x \ T \ T \ T \ x \end{array} \quad (3.2)$$

The close identity with the *E. coli* sequence, combined with the conservation of the EQ methylation sites and neighbouring EQ, indicates that these methylation sites are likely to be active in Rlv3841. There are, however, some interesting discrepancies. McrA, McrB and McrC are again divergent, matching instead the *E. coli* sequence with 100% identity. In addition, McpI, McpJ and McpP are missing the first EQ methylation site; McpO is missing the first and third; and IcpA is missing all three sites. Finally, RL1386 and RL1447 (putative MCPs) both have poor alignment in this area.

C-terminal methylation area

The second zone is in the C-terminal side of the MCP signal domain from heptads C13 to C14. The *E. coli* Tsr sequence in this region is (3.3)A, where again, the highlighted residue is an experimentally confirmed methylation site (Rice and Dahlquist 1991). This region also has high identity with the Rlv3841 consensus, (3.3)B. And the McrA, McrB and McrC protein sequences are (3.3)C. Where residues are highlighted to match the *E. coli* methylation residue.

$$\begin{array}{l} \text{A)} V \ T \ Q \ Q \ N \ A \ A \ L \ V \ E \ \mathbf{E} \ S \ A \ A \\ \text{B)} V \ T \ Q \ Q \ N \ A \ A \ M \ V \ E \ \mathbf{E} \ [ST] \ T \ A \\ \text{C)} V \ T \ Q \ Q \ N \ A \ G \ A \ S \ E \ \mathbf{Q} \ M \ S \ A \end{array} \quad (3.3)$$

Again, these sequences seem close enough to be active in Rlv3841. There are also some discrepancies with this region: IcpA and HemAT are missing the EQ methylation site and RL1386 and RL1447 again have poor alignment in this area.

This means that only IcpA is missing all of the E and Q methylation residues. And despite diverging from the main chemoreceptor consensus, McrA, McrB and McrC have all of the methylation residues.

CheR/CheB binding motif

In addition to the methylation motifs, CheR and CheB methylation and demethylation efficiency can be increased through binding on the flexible C-terminus of the chemoreceptor. The binding in *E. coli* and *S. meliloti* was found to match the (3.4)A and B motifs respectively. Or, in general (3.4)C (Alexander and Zhulin 2007; Meier, Muschler and Scharf 2007). In Rlv3841, the consensus is (3.4)D and is found in a number of membrane-bound (McpD, McpE, McpG, McpK, McpM, McpT and McpX) and cytoplasmic chemoreceptors (McpQ and McpY2).

$$\begin{array}{lllll}
 \text{A)} & N & W & E & T & F \\
 \text{B)} & x & W & E & x & F \\
 \text{C)} & x & [FW] & x & x & [FW] \\
 \text{D)} & x & W & [ET] & E & F
 \end{array} \tag{3.4}$$

3.7 Rlv3841 chemoreceptor classification

The Rlv3841 chemoreceptors were also individually analysed based on relationships determined in the maximum likelihood phylogenetic tree of chemoreceptor sequences (Figure 3.6) and putative functions determined by analysis of their sensor domains. Gene images were taken from CDvist analysis with grey boxes indicating transmembrane domains and red outlines indicating tentative annotations (>60%) (Adebali, Ortega and Zhulin 2015). The results are summarised in Table 3.2.

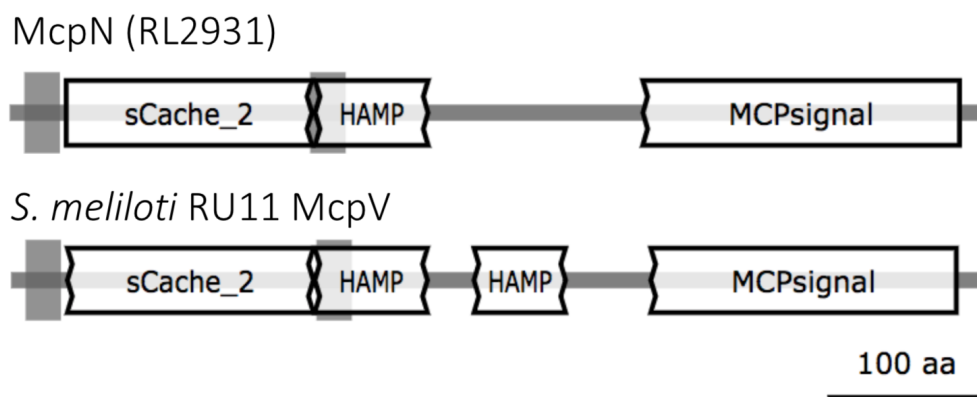


Figure 3.7: McpN domain structure.

3.7.1 Short-chain carboxylate sensing

The phylogenetic tree indicates that McpN (RL3931) is closely related to *S. meliloti* McpV (Figure 3.6), confirmed with a sensor domain alignment (67.5% amino acid identity). The annotated sCache2 domain is a single Cache domain based on the original ‘calcium channels and chemotaxis receptor’ based domain which has been found to bind amino acids and small carbohydrates (Anantharaman and Aravind 2000) (Figure 3.7). Searching the McpN sCache2 domain on the RCSB protein data bank showed 56% identity with generic Cache chemosensory domain binding carboxylic acids and 35-49% identity to various other chemoreceptors binding small carbohydrates (Berman 2000). Indeed, the closely related *S. meliloti* McpV was found to sense short-chain carboxylates (including acetate, propionate, pyruvate, glycolate, formate and acetoacetate) (Compton *et al.* 2018). In addition, *S. meliloti* McpV was linked to Che1 due to localisation to cell poles and was found in high abundance in the cell (Meier and Scharf 2009; Zatakia *et al.* 2017).

3.7.2 Oxygen, redox or light sensors

The phylogenetic tree indicates that McpQ (RL4387) and McpY2 (pRL120068) are clustered with *S. meliloti* McpY (Figure 3.6), confirmed by a sensor domain alignment (70.6% and 60.5% identity respectively; McpG and McpS were also clustered on the phylogenetic tree but had greatly divergent sensor domains). *S. meliloti* McpY has an unknown role, with very low abundance detected in the

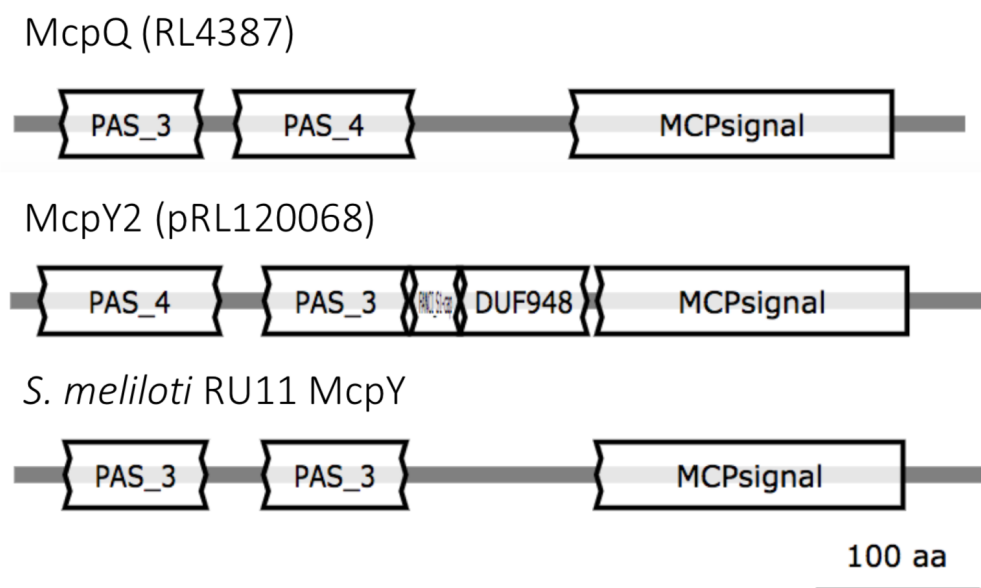


Figure 3.8: McpQ and McpY2 domain structure.

cell, although with 30-40% residual expression in alfalfa nodules. Thus, it was hypothesised to play a role *in planta* and to be linked to Che1 due to its localisation to cell poles (Meier and Scharf 2009; Zatakia *et al.* 2017). The proteins are all of topology type IVa, meaning that they are soluble receptors binding intracellular signals (Salah Ud-Din and Roujeinikova 2017). In particular, the PAS domains found in each protein are known to bind a number of cofactors: haem to detect oxygen, FAD to detect redox potential, or FMN to detect light (Möglich and Moffat 2007; Collins, Lacal and Ottemann 2014) (Figure 3.8). Searching the sensor domains on the RCSB protein data bank showed 25% and 32% identity with an FAD-binding redox sensor and, 28% and 32% identity with FMN-binding photosensor for McpQ and McpY2 respectively (Berman 2000). McpY2 is also found adjacent to a *flgLKE* operon on pRL12.

3.7.3 Other haem binding proteins

IcpA (RL0685) and HemAT (RL0429) are both soluble receptors of topology type IVa containing a protoglobin domain, known for haem binding (Figure 3.9). This domain was originally described in aerotactic *Bacillus subtilis* HemAT which binds haem in its sensor domain (residues 1-178) via residues Tyr-70 and His-123

(corresponding to Tyr-41 and His-94 in model) (Hou *et al.* 2001; Zhang, Olson and Phillips 2005). The haem has two distinct O₂ binding components with a biphasic pattern (1-2 μ M and 50-100 μ M components). As a result of having both high- and low-affinity components, HemAT can respond to oxygen concentration gradients under both hypoxic (0-10 μ M) and aerobic (50-250 μ M) conditions (Zhang, Olson and Phillips 2005). In addition to the binding residues, Tyr-70, Trp-132 and Tyr-133 were found to have a conformational change on ligand binding important for signal propagation (El-Mashtoly *et al.* 2008). HemAT (originally annotated McpH) has methylation residues, the haem binding residues (Tyr-46 and His-99) and signal propagation residues (Trp-108 and Tyr-109), it also has a 21.9% identity (45.3% positive identity) over the binding region. IcpA (originally annotated HemAT), on the other hand, is missing both haem binding residues (Glu-107 and Phe-54) and one of the signal propagation residues (Trp-116 matches but His-117), finally having less identity with *B. subtilis* HemAT (17.9% with 36.4% positive identity).

IcpA correlates better with *S. meliloti* IcpA. In addition, the phylogenetic tree indicates that *S. meliloti* IcpA is more closely related to IcpA (Figure 3.6), confirmed with a sensor domain alignment (58.9% identity). *S. meliloti* IcpA has an unknown role, with low abundance detected in the cell, but is likely linked to the Che1 cluster, being found in the *che1* operon in both species and, despite being soluble, due to its localisation to cell poles (Meier and Scharf 2009; Zatakia *et al.* 2017). A deletion mutant displayed a significant change in halo size on Bromfield agar, although the target of the sensor is currently unknown (Sourjik and Schmitt 1996; Meier, Muschler and Scharf 2007).

3.7.4 Quaternary ammonium compound sensing

The phylogenetic tree indicates that McpX (RL0426) is closely related to *S. meliloti* McpX (Figure 3.6), confirmed with a sensor domain alignment (67.4% identity) (Figure 3.10). As mentioned above, proteins with Cache domains are known to bind amino acids and small carbohydrates, indeed a search on the RCSB protein data bank showed 25% identity with small carbohydrate binding chemoreceptors

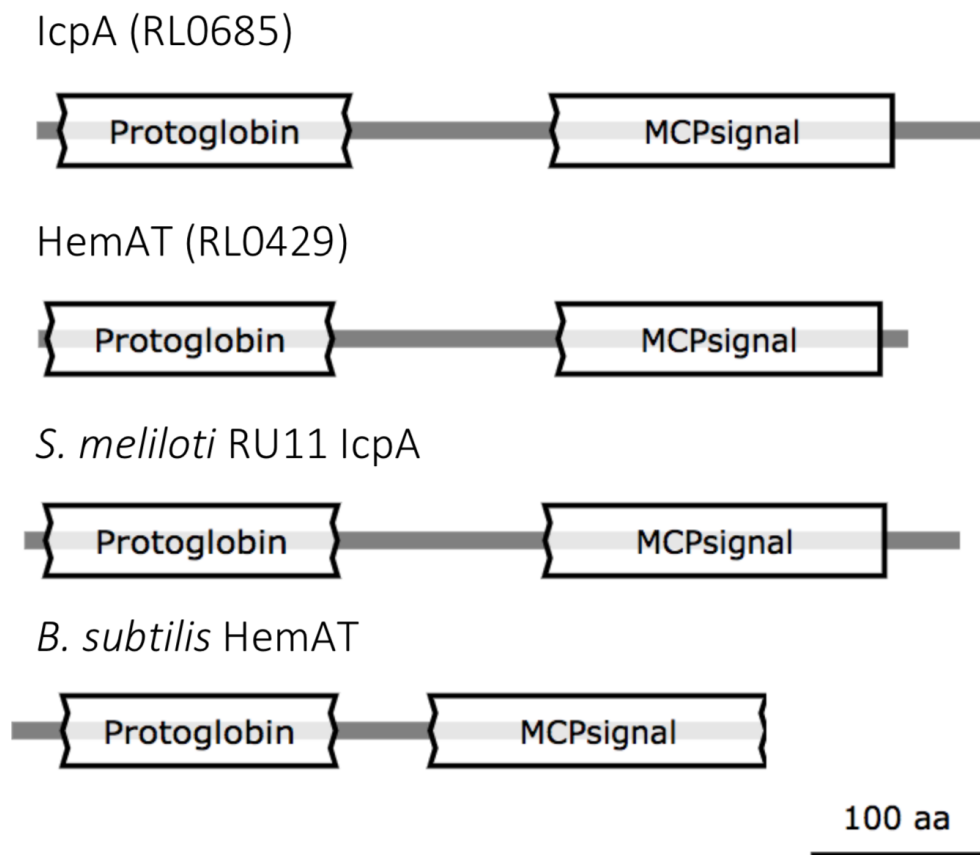


Figure 3.9: IcpA and HemAT domain structure.

(Berman 2000). Although found in low abundance in the cell, *S. meliloti* McpX is likely linked to the Che1 cluster, due to localisation to the cell poles and a deletion mutant having a significant change in halo size on Bromfield agar (Sourjik and Schmitt 1996; Meier, Muschler and Scharf 2007; Meier and Scharf 2009; Zatakia *et al.* 2017). It was also found to sense quaternary ammonium compounds (exuded by plant seeds and roots), including choline and betaines (Webb *et al.* 2017b). The binding pocket is composed of a number of aromatics (including Trp-109, Tyr-139, Phe-153 and Trp-161) and carboxylate binding residues (Ala-179, Gly-180, Asp-208), all of which are conserved in McpX (Shrestha *et al.* 2018).

3.7.5 Plant related sensors

The phylogenetic tree indicates that McpT is closely related to *S. meliloti* McpT (Figure 3.6), confirmed with a periplasmic region alignment (55.0% identity). *S.*

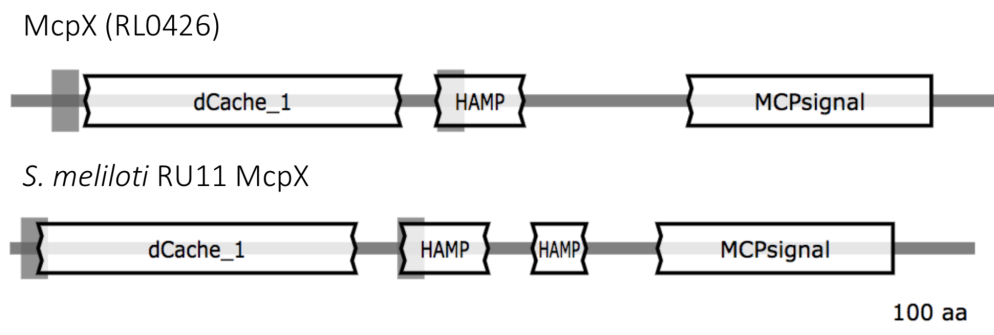


Figure 3.10: McpX domain structure.

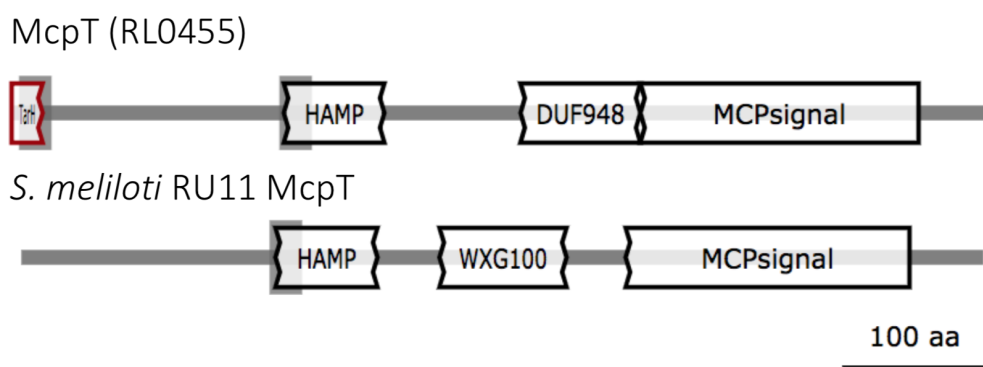


Figure 3.11: McpT domain structure.

meliloti McpT has an unknown role, being nearly undetectable in the cell, although having 30-40% residual expression in alfalfa nodules. Thus, it was hypothesised to play a role *in planta* and to be linked to Che1 due to its localisation to cell poles (Meier and Scharf 2009; Zatakia *et al.* 2017). In addition to a predicted TM at 180-200, *S. meliloti* McpT had non-significant hydrophobic residues at 1-20 (matching McpT architecture), potentially representing a modified transmembrane (TM) domain (Krogh *et al.* 2001) (Figure 3.11). The sensor domain for both proteins remains unidentified.

3.7.6 Putative organic acid sensing

McpB (pRL120683) is predicted to have an HBM domain, a 6 alpha helix bundle (Lacal *et al.* 2010a; Ortega and Krell 2014) (Figure 3.12). This domain was characterised in a *Pseudomonas putida* chemoreceptor, McpS. McpS was found to bind organic acids from the TCA cycle but has only 21.5% identity to the

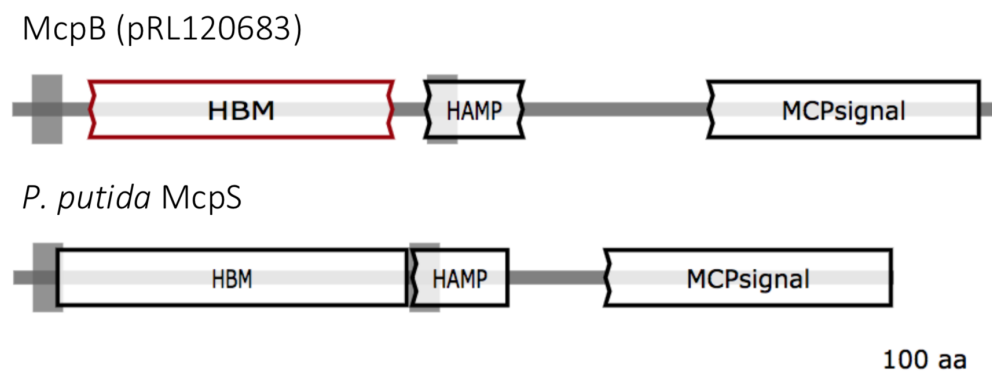


Figure 3.12: McpB domain structure.

McpB domain. In addition, none of the required residues for ligand binding were conserved in McpB. This diverged sensor domain was found to be important in *R. leguminosarum* biovar *viciae* VF39SM where the nearly identical (>98% identity) McpB mutant had greatly reduced nodule competitiveness on *Pisum sativum* (Yost, Rochepeau and Hynes 1998).

3.7.7 Four-helix bundle sensor domains

Chemoreceptors with the 4HB_MCP_1 domain include McpC (pRL120312), McpG (pRL100403) and McpL (RL1065) (Figure 3.13). 4HB_MCP_1 is a 4 alpha helix bundle sensor domain with no known targets. The only other information known about these is that the equivalent chemoreceptor to McpC in *R. leguminosarum* biovar *viciae* VF39SM had greatly reduced nodule competitiveness on *Pisum sativum* against the wildtype (>98% identity) (Yost, Rochepeau and Hynes 1998). The domain architecture seems standard in McpC, but McpG and McpL are missing the second TM domain. They both have nonsignificant hydrophobic residue clusters (at 200-220 and 180-200 respectively) that may indicate a modified TM domain (Krogh *et al.* 2001).

3.7.8 Dual Cache sensor domains

Chemoreceptors with the dCache1 domain include McpM (RL1318) and McpO (RL3985) (Figure 3.14). This dual Cache domain is based off the original ‘calcium

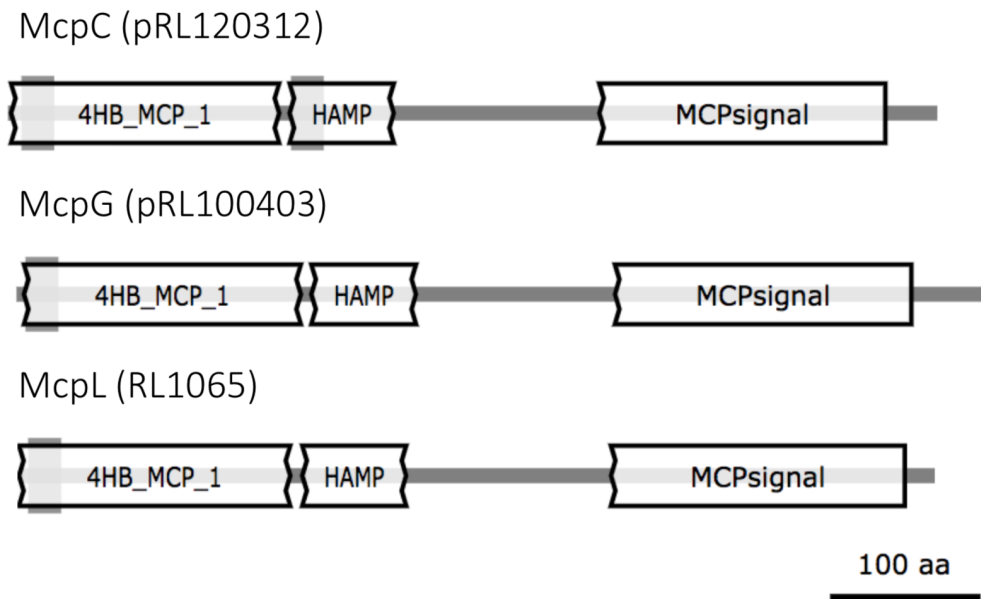


Figure 3.13: McpC, McpG and McpL domain structure.

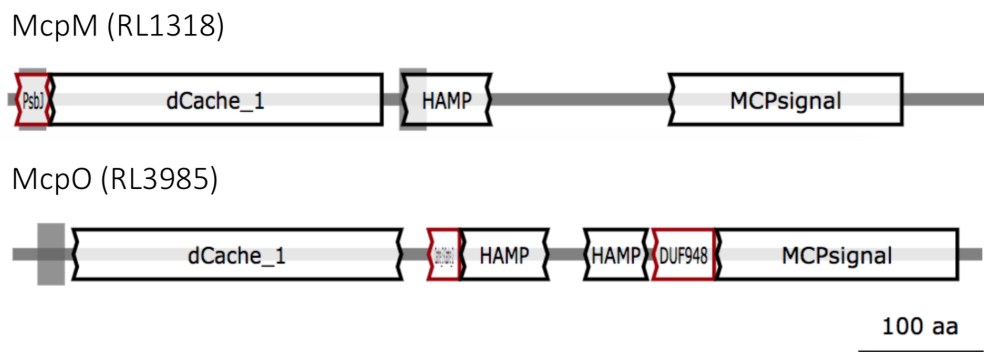


Figure 3.14: McpM and McpO domain structure.

channels and chemotaxis receptor' based domain that is known to bind amino acids and small carbohydrates (Anantharaman and Aravind 2000). Searching the RCSB protein data bank for both dCache1 domains showed 30% identity with receptors binding small carbohydrates for McpM (Berman 2000). The domain architecture is standard for McpM, but McpO is missing the second TM domain. It again has a nonsignificant hydrophobic residue cluster (at 360-390, before first HAMP domain), that may indicate a modified TM domain (Krogh *et al.* 2001).

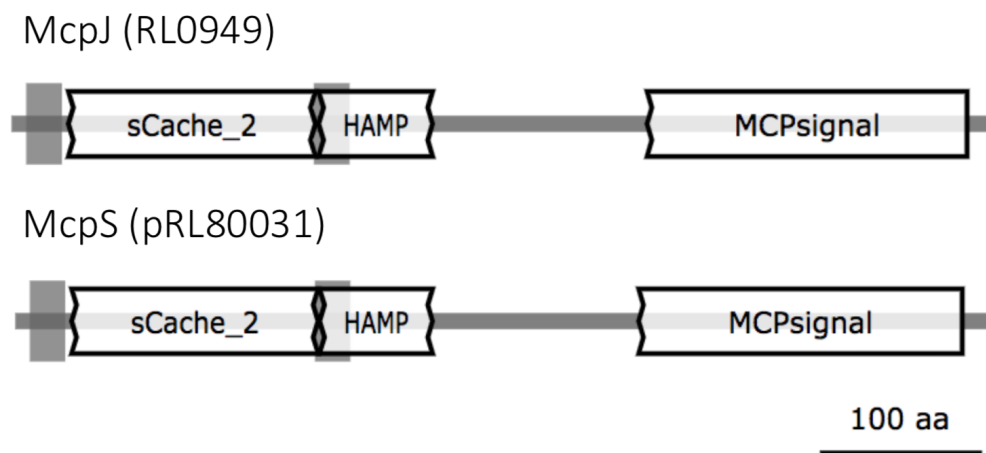


Figure 3.15: McpJ and McpS domain structure.

3.7.9 Single Cache sensor domains

Chemoreceptors with the sCache2 domain include McpJ (RL0949) and McpS (pRL80031) (Figure 3.15). This single Cache domain is also based on the original ‘calcium channels and chemotaxis receptor’ based domain that is known to bind amino acids and small carbohydrates (Anantharaman and Aravind 2000). Searching the RCSB protein data bank for both sCache2 domains showed 30-32% identity of McpS with a number of receptors binding small carbohydrates and 30% identity with generic Cache chemosensory domain binding carboxylic acids (Berman 2000). The domain architecture is standard for both proteins.

3.7.10 CHASE sensor domain

The sole chemoreceptor in Rlv3841 with the CHASE3 domain is McpK (RL0972) (Figure 3.16). This domain was characterised from ‘Cyclases/Histidine kinases Associated Sensory Extracellular’ and is speculated to bind small peptides (Mougel and Zhulin 2001). *Pseudomonas aeruginosa* PAO1 McpN has been found to bind nitrate and has 13.6 Z-score on a DALI search and 13% identity with *Rhodopseudomonas palustris* HK9’s CHASE3 domain (Martín-Mora *et al.* 2019). The CHASE3 domain of McpK has 15.9% identity with *P. aeruginosa* PAO1 McpN.

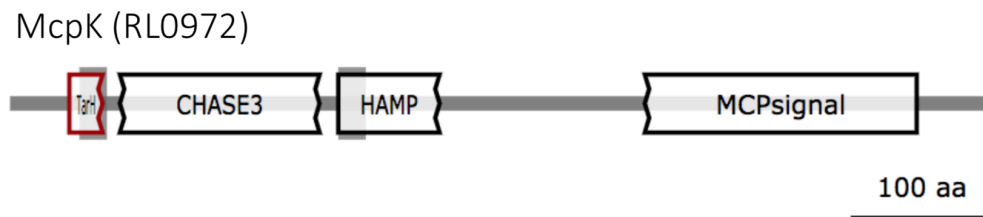


Figure 3.16: McpK domain structure.

3.7.11 Che2-linked relatives

A related group of chemoreceptors, McrA (RL4031), McrB (RL4032) and McrC (RL4033) are all found within the *che2* operon and have high identity (Figure 3.17). These proteins are distinct from the other *Rlv3841* chemoreceptors (Figure 3.6), being in the 34-heptad class (Alexander and Zhulin 2007). Although also found in its *che2* operon, *S. meliloti* McpS is phylogenetically distinct and in the 40-heptad class (Figure 3.6). It is, however, independent of other *S. meliloti* chemoreceptors, being evenly distributed throughout the cell, not maximally expressed during mid-exponential phase and appears inactive during normal chemotactic movement. It is hypothesised to be involved in plant-related chemosensing with an unknown sensor domain (Meier, Muschler and Scharf 2007; Meier and Scharf 2009). McrA, McrB and McrC, on the other hand, have the four-helix bundle sensor domain, described above to have no known targets. However, they are likely to be involved with the Che2 chemotaxis cluster, with the duplication allowing divergence of function.

3.7.12 >200 bp unknown sensor domains

Although the phylogenetic tree indicates that *S. meliloti* McpZ is more closely related to McpI (RL0758) (Figure 3.6), the periplasmic region of *S. meliloti* McpZ was better aligned with McpZ (RL0757; 36.7% identity), with McpI showing high differentiation with low identity (12.3%). Another chemoreceptor, McpP (RL4277) also had higher identity (24.1%). Indeed, the domain architecture of McpI is highly differentiated from *S. meliloti* McpZ, consisting of a short, merged Cache domain with three TM domains (Figure 3.18). *S. meliloti* McpZ, McpZ and McpP, on the other hand, have a long unannotated sensor domain. An additional complication,

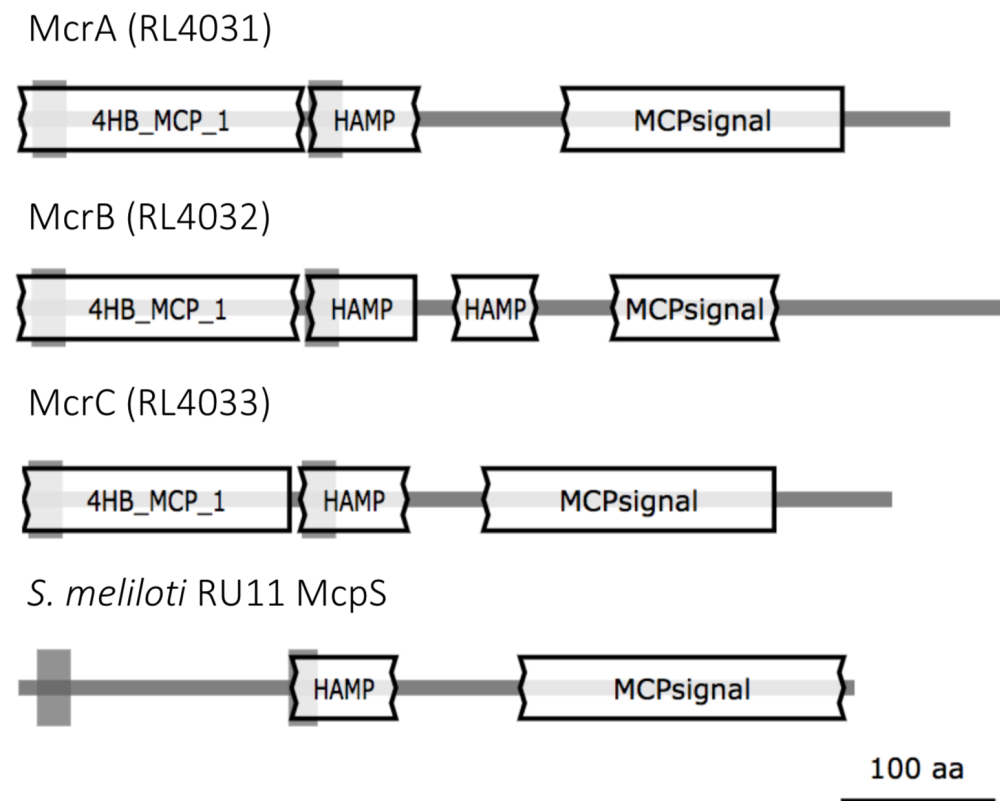


Figure 3.17: McrA, McrB and McrC domain structure.

S. meliloti McpZ and McpZ have only a single TM domain, indicating the sensor domain is located inside the cell (Krogh *et al.* 2001). These proteins fit topology type IIIc, mostly found in archaea (Salah Ud-Din and Roujeinikova 2017). Although *S. meliloti* McpZ was detected at a very low abundance in the cell and has an unknown sensor domain, it is likely linked to the Che1 cluster due to localisation to cell poles (Meier and Scharf 2009; Zatakia *et al.* 2017).

3.7.13 <200 bp unknown sensor domains

These chemoreceptors have no significant domain hits across their predicted sensor domains and include McpD (RL2683), McpE (RL0564) and McpR (pRL120056) (Figure 3.19). The proteins have poor alignment between their periplasmic domains, likely indicating different sensor domains. McpR is located near a *flgLKE* operon on pRL12. McpE appeared growth defective under 1% and 21% oxygen on both glucose and succinate in the INSeq data.

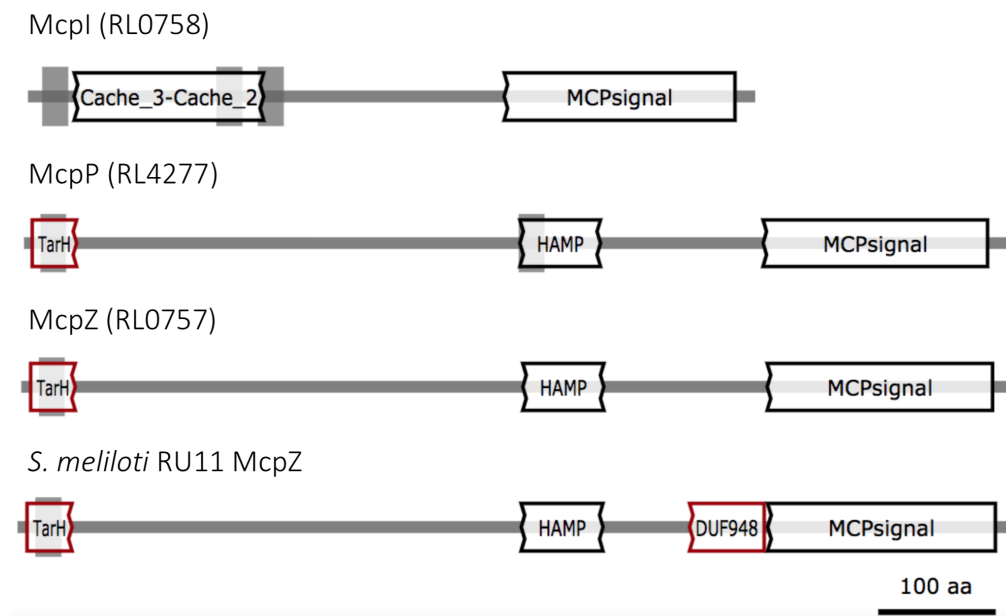


Figure 3.18: McpI, McpP and McpZ domain structure.

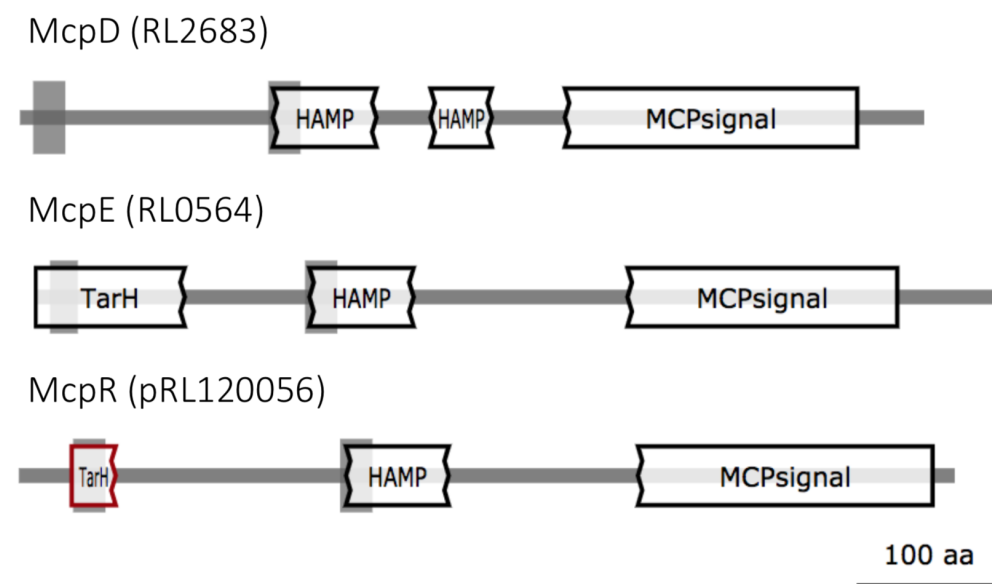


Figure 3.19: McpD, McpE and McpR domain structure.

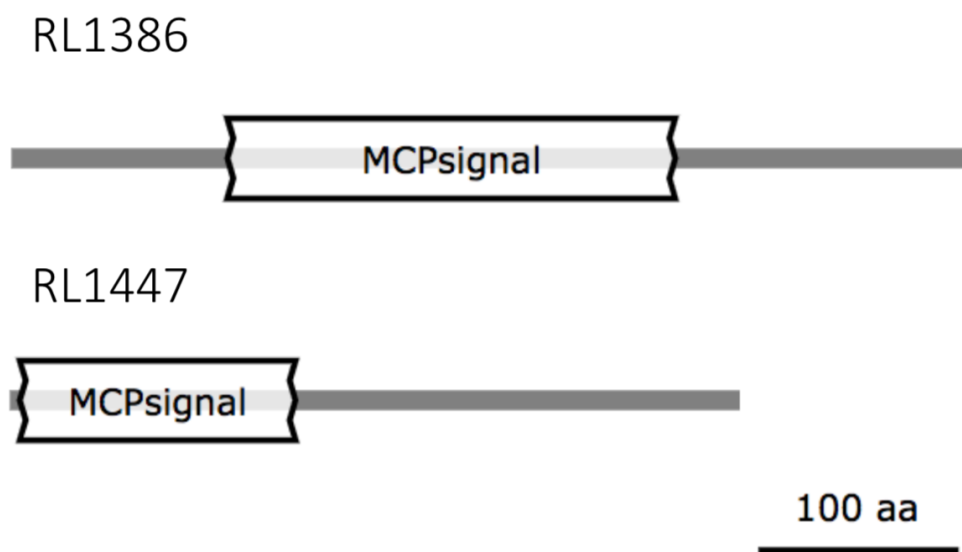


Figure 3.20: RL1386 and RL1447 domain structure.

3.7.14 No sensor domains

The final two proteins, RL1386 and RL1447, are both unannotated and do not match the typical chemoreceptor architecture (Figure 3.20). They mainly consist of the MCP signal domain with no sensor or transmembrane domains. They may be of topology type IVb, perhaps holding a regulatory role (Salah Ud-Din and Roujeinikova 2017).

3.8 Testing of internal chemoreceptors as candidates for starvation response

Since the starvation response was discovered for multiple compounds important for internal metabolism, global control is unlikely to be due to an externally sensing chemoreceptor. In addition, since *S. meliloti* Rm2011 did not display this starvation effect (Figure 3.1), this response is unlikely to be due to chemoreceptors that are closely related to *S. meliloti* chemoreceptors. Therefore, this only leaves HemAT, which is an internal receptor divergent from IcpA and *S. meliloti* IcpA. We also decided to study the other internal chemoreceptors (McpQ, McpY2 and IcpA) to find conditions in which they are active.

We obtained three strains with Tn5 mutations in these genes from Allan Downie's laboratory (John Innes Centre): A1296 (*mcpQ*), A1300 (*mcpY2*) and A1289 (*hemAT*). To make sure the Tn5 mutation was unique, they were transduced back into Rlv3841, producing the strains OPS1881, OPS1505 and OPS1500, respectively. The position of the Tn5 transposon in these strains was also confirmed by sequencing upstream and downstream genome fragments amplified using primers oxp2649+Tn5, oxp2660+Tn5 and oxp2646+Tn5, respectively (listed in Table 2.3).

To isolate a strain mutated in *icpA*, a double crossover mutation was isolated using a pK19mobSacB strategy, as described in Section 2.5.2, producing plasmid pOPS1411. After the first single crossover, the integration was mapped using primers oxp3917+oxp3918. A double crossover was isolated on 10% sucrose and the deletion of *icpA* was confirmed by PCR mapping using primers oxp4057+oxp4058 and sequencing the amplified fragments. This strain was named OPS2108 ($\Delta icpA$).

Finally, to isolate a double mutation of *hemAT* and *icpA*, the *hemAT* Tn5 mutation was transduced into the *icpA* deletion strain (OPS2108). The Tn5 transposon insertion was confirmed with PCR as above. This strain was named OPS2288 ($\Delta icpA$ and Tn5 *hemAT*).

These mutants were tested in swimming plates under a wide range of minimal media conditions. Firstly, on a variety of carbon sources together with the wildtype and the *che1* mutant as controls (Figure 3.21). These included glucose (10 mM vs 1 mM), succinate (20 mM vs 2 mM), malate (20 mM vs 2 mM), fumarate (20 mM vs 2 mM), pyruvate (30 mM vs 3 mM), arabinose (10 mM vs 1 mM), acetate (30 mM vs 5 mM), with NH_4Cl (10 mM) as nitrogen source in all conditions. The strain and media factors were all found to be significant, so all the strains were compared within each media condition at high and low concentrations. At low concentration, only the *icpA* mutant was significantly reduced on all the conditions tested except acetate. At high concentration, the *icpA* and *hemAT* mutants were reduced on glucose and the *icpA* mutant on arabinose. Across all media conditions, the *mcpQ* and *mcpY2* mutants were not significantly different from wildtype.

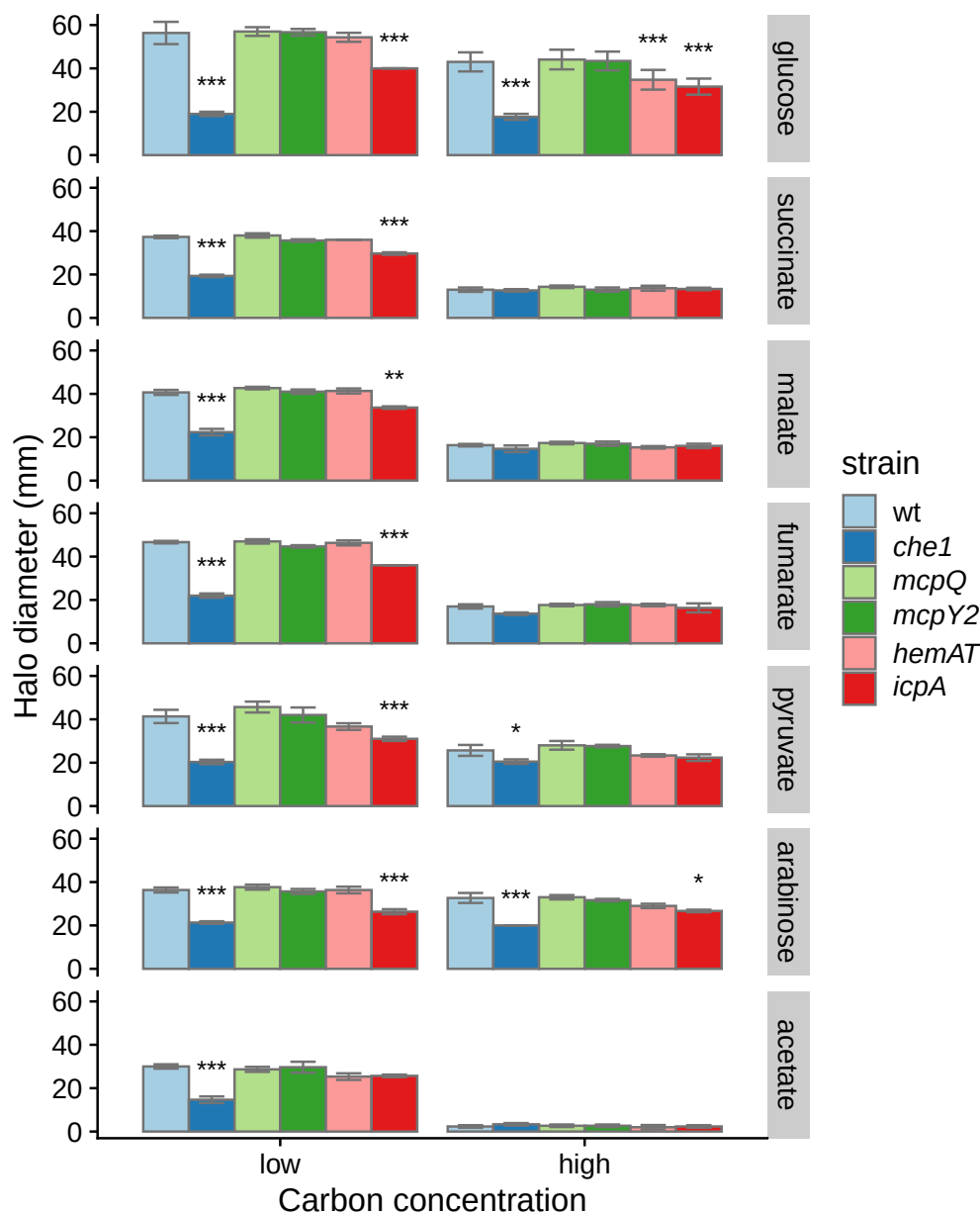


Figure 3.21: Swimming of internal chemoreceptor mutants in semi-solid agar plates with different carbon sources. Swimming halo diameters (mm) from plate assays done on minimal media with 10 mM NH_4Cl as nitrogen source and different carbon sources at high and low concentrations. The strains tested were wt (Rlv3841), *che1* (LM100), *mcpQ* (OPS1881), *mcpY2* (OPS1505), *hemAT* (OPS1500) and *icpA* (OPS2108). Data was modelled with a Gaussian linear model with the factors strain, carbon source and carbon concentration and their interactions. All parameters were found to be significant, $P < 0.01$, ANOVA. Dunnett's Post-hoc test found mean differences between strains with $P < 0.01$. Bars are mean $\pm$ SD, $N=3$. Indicated significance levels are * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, insignificant differences were not indicated.

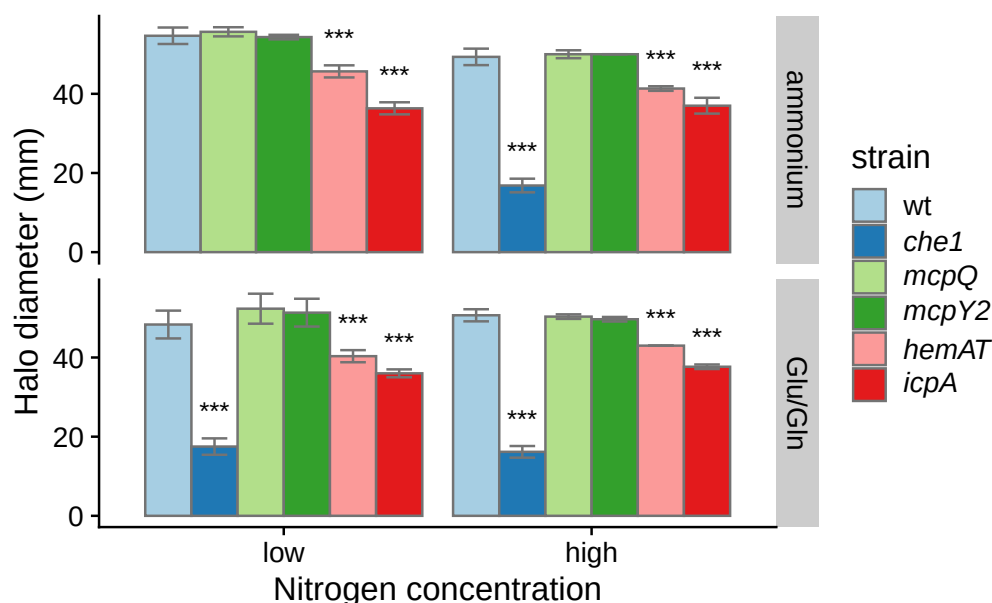


Figure 3.22: Swimming of internal chemoreceptor mutants in semi-solid agar plates with different nitrogen sources. Swimming halo diameters (mm) from plate assays done on minimal media with 10 mM glucose as carbon source and nitrogen sources at high and low concentrations. The strains tested were wt (Rlv3841), *che1* (LM100), *mcpQ* (OPS1881), *mcpY2* (OPS1505), *hemAT* (OPS1500) and *icpA* (OPS2108). Data was modelled with a Gaussian linear model with the factors strain, nitrogen source and nitrogen concentration and their interactions. All parameters except strain:nitrogen were significant, $P < 0.05$, ANOVA. Dunnett's Post-hoc test found mean differences between strains with $P < 0.05$. Bars are mean $\pm$ SD, $N=3$. Indicated significance levels are *** $P < 0.001$, insignificant differences were not indicated.

Next, the strains were tested under different nitrogen sources (10 mM vs 0.5 mM ammonium and 10 mM glutamate vs 10 mM glutamine) with glucose as a carbon source (Figure 3.22). Again, the strain and media factors were all found to be significant, so the strains were compared within each media condition at high and low concentrations. These nitrogen conditions had a similar pattern to the high glucose above, with the *icpA* and *hemAT* mutants having significantly reduced diameters in all conditions tested, but with no effect from the *mcpQ* and *mcpY2* mutations.

The strains were also tested in high glucose (10 mM), high pyruvate (30 mM) and low pyruvate (3 mM) minimal media supplemented with NH_4Cl (10 mM) under light and dark conditions (Figure 3.23). In this assay, there was no observed effect due to the light level. Only the strain and media factors were found to be significant, without interactions, indicating that the strain effect does not differ across media

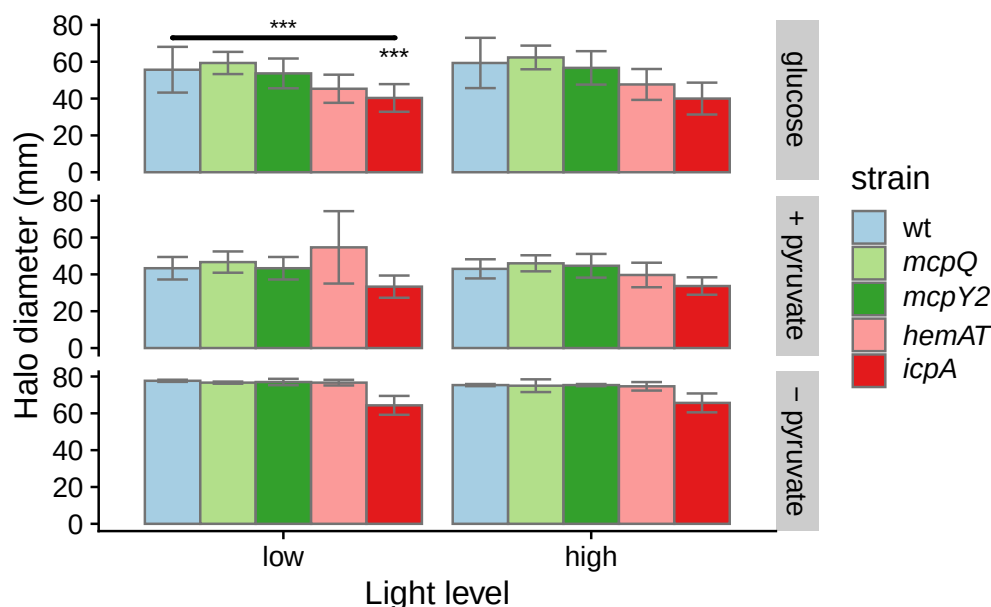


Figure 3.23: Swimming of internal chemoreceptor mutants in semi-solid agar plates with different light levels. Swimming halo diameters (mm) from swimming plate assays done on minimal media with 10 mM NH_4Cl as nitrogen source and high glucose (10 mM), high pyruvate (30 mM) and low pyruvate (3 mM) as carbon sources. The strains tested were wt (Rlv3841), *che1* (LM100), *mcpQ* (OPS1881), *mcpY2* (OPS1505), *hemAT* (OPS1500) and *icpA* (OPS2108). Data was modelled with a Gaussian linear model with the factors strain, media and light level and their interactions. Only strain and media terms were significant, $P < 1e-6$, ANOVA. Dunnett's Post-hoc test found mean differences between strains across all media conditions, $P < 1e-5$. Bars are mean $\pm$ SD, $N=3$. Indicated significance levels are *** $P < 0.001$, insignificant differences were not indicated.

conditions. This effect was found to be significant for the *icpA* mutant.

Finally, the strains were tested at 21% and 1% oxygen with different carbon sources (high glucose (10 mM), high succinate (20 mM), low succinate (2 mM), high arabinose (10 mM) and low arabinose (10 mM)) (Figure 3.24). The strain and media factors were all found to be significant, so the strains were compared within each media condition at high and low oxygen tension. The *icpA* mutant was found to be significantly reduced under all the conditions tested except for high succinate at 21% oxygen and high glucose at 1% oxygen. Comparing to previous results, the *icpA* mutant was expected to be significantly reduced at high glucose and not at high succinate at 1% oxygen. However, under both conditions, the *icpA* mutant displayed the highest variance, indicating that these results perhaps need more

replicates for a more robust response. Indeed, the *icpA* and *hemAT* double mutant matched the reduction under all conditions except that it was significantly reduced in high glucose and not at high succinate with 1% oxygen. The *hemAT* mutant displays some interesting differences between the 21% and 1% oxygen conditions. It was found to be significantly reduced in only high glucose and arabinose at 21% oxygen. However, at 1% oxygen, it was also found to be reduced in low succinate and arabinose. This effect of *hemAT* deletion appears to occur as an interaction of low carbon and low oxygen tension. Once again, the *mcpQ* and *mcpY2* mutants were not found to be significantly different to wildtype under any of the conditions tested.

Overall, none of the mutants displayed the effects expected for a sensor driving the TCA-cycle intermediate response, as they did not show an increased swimming diameter under high succinate, malate or fumarate. This confirms the hypothesis that the effect is chemotaxis and chemoreceptor independent and that another regulatory mechanism might be responsible.

The broad impact of the *icpA* deletion may indicate that this chemoreceptor is a general sensor of the metabolic state. Especially considering it was significantly reduced on nearly every media condition where the *che1* mutant was also reduced. Without the sensor encoded by *icpA*, the cell is unable to sense how unfavourable certain carbon sources are. This effect also appears robust to changing oxygen tension, indicating independence from oxygen sensing. This was expected, since the protein appears unable to bind haem (Section 3.7.3).

The action of the sensor encoded by *hemAT* appears to be more restricted. In general, it appears to be active at high carbon source concentration, with fairly low robustness. But at low carbon concentration, it only appears active when also at low oxygen tension. This indicates some level of oxygen sensing, as predicted in Section 3.7.3. But how can we explain the interaction effects? Well, at high carbon concentration, the cell is expected to have a high level of oxygen consumption. Whereas under low carbon concentration, the metabolism only has a significant impact on internal oxygen tension when already at low oxygen. An oxygen sensor would consider both conditions to be unfavourable, and so a

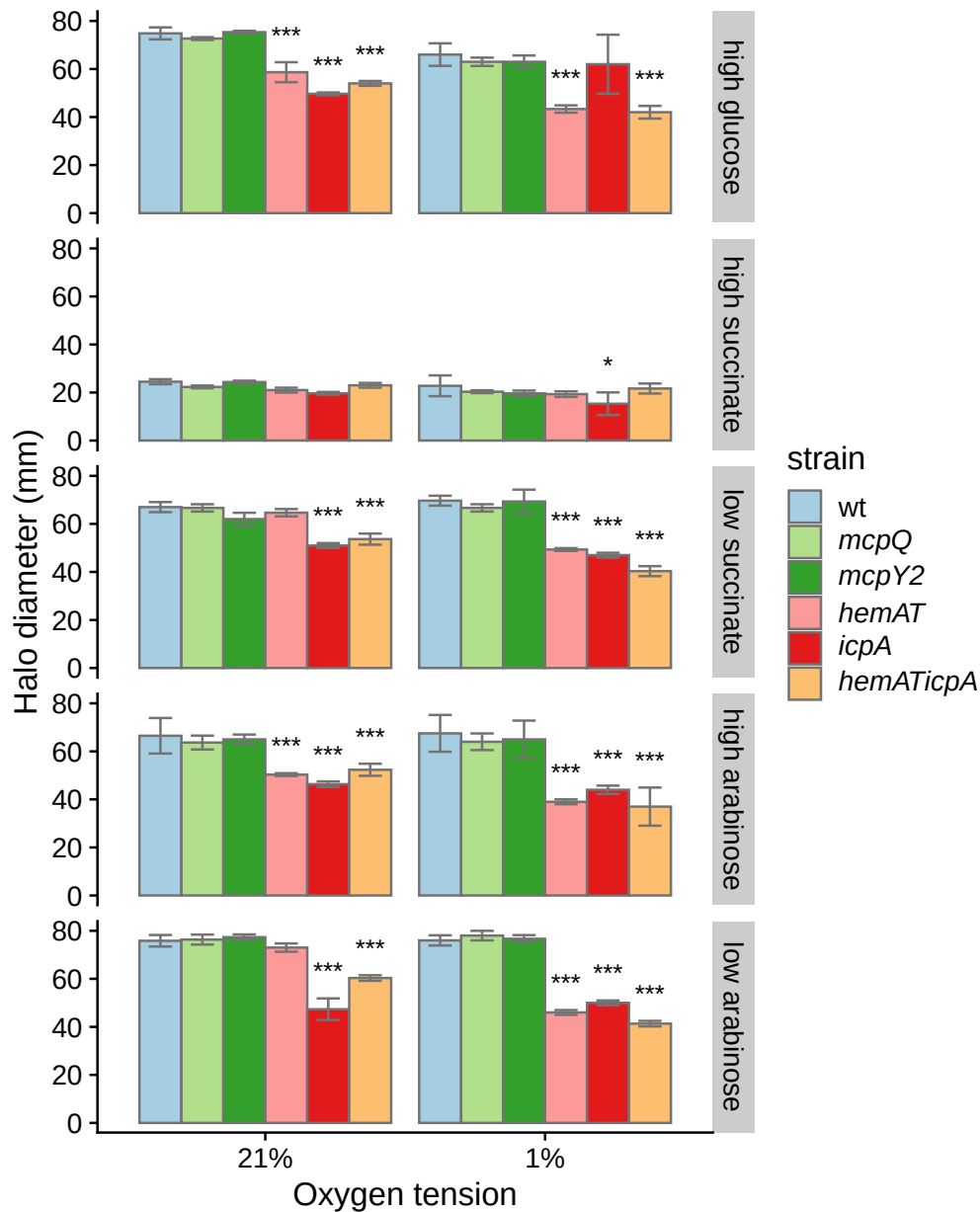


Figure 3.24: Swimming of internal chemoreceptor mutants in semi-solid agar plates with different oxygen environments. Swimming plate diameters of wt (Rlv3841), *che1* (LM100), *mcpQ* (OPS1881), *mcpY2* (OPS1505), *hemAT* (OPS1500), *icpA* (OPS2108) and *hemATicpA* (OPS2288). Data was modelled with a Gaussian linear model with the factors strain, media and oxygen tension and their interactions. All factors were found to be significant, $P < 0.01$, ANOVA. Dunnett's Post-hoc test found mean differences between strains with $P < 0.05$. Bars are mean $\pm$ SD, $N=3$. Indicated significance levels are * $P < 0.05$, *** $P < 0.001$, insignificant differences were not indicated.

mutant in such a sensor would have reduced swimming diameters under those conditions and therefore, it would also be reduced in high carbon and only in low carbon when grown at low oxygen.

It may also be expected that low nitrogen conditions would reduce oxygen consumption. However, under these conditions, the *hemAT* mutant was significantly reduced at 21% oxygen. This may be explained by the different relevance of the nitrogen effect vs the carbon effect. Whenever nitrogen effects on swimming were tested, they had a smaller effect than carbon. For example, wildtype swimming diameter increases from about 43 mm at 10 mM glucose by 30% to 56 mm at 1 mM glucose (Figure 3.21). In contrast, from about 49 mm at 10 mM ammonium by 12% to 55 mm at 0.5 mM ammonium (Figure 3.22). It may be that 0.5 mM ammonium is insufficiently nitrogen starved to see a great enough reduction in metabolism to abolish the *hemAT* response.

The *mcpQ* and *mcpY2* mutants were not significantly reduced from wildtype under any media conditions tested. This may indicate some degree of redundancy between the two receptors. Indeed, they do share sensor domains, as mentioned above (Section 3.7.2, Figure 3.8).

3.9 Discussion

In this chapter, we first described the swimming characteristics of two rhizobial strains: *Rhizobium leguminosarum* biovar *viciae* 3841 (Rlv3841) and *Sinorhizobium meliloti* Rm2011 (Sme2011). Both strains displayed a starvation response: increased halo diameter in low carbon concentrations and small, thick halos in high carbon concentrations (Figure 3.1). For Rlv3841, the most obvious effects were with acetate and succinate. High acetate produced tiny, faint halos which are likely due to very little growth. High succinate produced a different halo type, small and thick as above, but with a cloudy spots appearance. This appears to be a response to high succinate, that triggers a non-motile state. We were also able to further investigate the succinate effect, finding the same effect for other dicarboxylic acid TCA-cycle intermediates such as malate and fumarate (Figure

3.3). Interestingly, this response to succinate appears to be novel and characteristic of *R. leguminosarum*, as Sme2011 did not show the same response.

What, then, is the biological purpose of the TCA-cycle intermediate effect? The most obvious possibility is that it is simply an extension of the starvation effect. Increasing the concentration of a carbon source causes a reduction in swimming diameter. Starvation responses in bacteria tend to control the tumble frequency of flagella to increase the time spent in high energy environments. Perhaps high TCA-cycle intermediates are a more favourable environment than any other high carbon source, such as sugars (e.g. glucose and arabinose) or TCA-cycle precursors (e.g. pyruvate). This is supported by the fact that the low succinate media produces the same sized halos as the high glucose media. However, this does not explain the distinct shape of the halos produced in high TCA-cycle intermediate media, so an alternative mechanism might be interacting with flagella to completely stop swimming.

We have demonstrated that deletions of key flagellar genes (*motA*, *fliG* and *flaA*) and chemotaxis gene clusters (*che1* and *che2*) were able to disrupt the swimming ability of Rlv3841 as expected (Figures 3.2, 3.3). We also confirmed the major role of Che1 in chemotaxis, with only minor effects discovered for Che2 (Miller *et al.* 2007). Interestingly, although the flagellar gene mutants no longer displayed the TCA-cycle intermediate effect, the chemotaxis gene cluster mutants still did. Since the glucose, pyruvate and arabinose starvation effects are removed by *che1* deletion, it is unlikely that the TCA-cycle intermediate effect is produced by the *che1* cluster.

If not chemotaxis, another possible explanation could involve metabolism. If so, then it is possible that the swimming plates are confounded by the requirement for growth to be visible. In order to remove this possibility, we also tracked individual bacteria swimming in liquid cultures grown in both high and low succinate (Figure 3.5). Both wildtype and *che1* mutants displayed a significantly reduced motile cell fraction in high succinate media. This confirms that the high succinate media directly affects the motility of cells. In addition, the swimming speed and tumble rate of the bacteria were also reduced in high succinate. Whether these modifications to

swimming parameters are also due to the same mechanism is unclear. It is possible that the reductions in swimming speed shift the distribution of cells, such that there are an increased number of non-motile cells.

To further demonstrate the independence of the TCA-cycle intermediate effect from the effect of the chemotactic machinery, we investigated the chemosensory array of Rlv3841 (Figure 3.6). Rlv3841 has 25 chemoreceptors, with a wide variety of sensory domains and phylogenetic diversity. At this stage, we were interested in a potential link to metabolism, so the internally sensing chemoreceptors were chosen for further analysis. These included PAS-domain sensing McpQ and McpY2, and protoglobin-domain sensing IcpA and HemAT (Sections 3.7.2, 3.7.3). After generating the corresponding mutants, none of these chemoreceptors were found to alter the high succinate, malate or fumarate responses. McpQ and McpY2 are likely redundant, since they are highly similar and single mutants did not find any effects. IcpA was found to be important under a wide range of conditions, indicating that it may be a general sensor of metabolic state. HemAT appears to be an oxygen sensor, being important under conditions of high oxygen consumption or low oxygen tension. Thus, the response does not appear to be connected to any internally sensing chemoreceptor.

So what about a metabolic link? A literature search revealed one instance of metabolism being necessary for flagellar function. In gutted *Escherichia coli* cells, flagellar switching was only recovered with the addition of CheY and either fumarate, malate or succinate (Barak and Eisenbach 1992; Barak, Giebel and Eisenbach 1996). CheY alone causes a single CCW to CW rotation switch, with no further changes. The cause appears to be reducing the free energy of switching by fumarate reductase preferentially binding and stabilising CW rotation (Prasad, Caplan and Eisenbach 1998; Koganitsky *et al.* 2019). Although this appears superficially similar to the effect found in Rlv3841, there are several important differences. Firstly, although in *E. coli* CW rotation corresponds to a tumble, in Rlv3841, it is responsible for forward movement (Miller *et al.* 2007). Rlv3841 is also unable to change the direction of flagellar rotation. This means that any stabilisation of flagellar would likely

have minimal effect on the swimming behaviour of Rlv3841. Secondly, Rlv3841 does not have a fumarate reductase. Although it does have a related enzyme, succinate dehydrogenase, this enzyme was found to not have the same effect in *E. coli* (Koganitsky *et al.* 2019). Finally, fumarate reductase appears to only affect flagellar switching and is unlikely to trigger a non-motile state. However, there are also other descriptions of metabolism effects on swimming.

The interaction of metabolism with chemotaxis has only recently been accepted. The standard model has been metabolism-independent chemotaxis, due to initial descriptions of bacterial response to non-metabolizable compounds (Adler 1969). More recently, metabolism-dependent chemotaxis has been described (Alexandre and Zhulin 2001). This involves interactions between metabolic and chemotactic mechanisms. The authors of a modelling paper describe a third option: metabolism-based taxis, ostensibly invented to allow simulations of metabolism-dependent chemotaxis without modelling chemotaxis (Egbert, Barandiaran and di Paolo 2010). However, this mechanism appears similar to that proposed by the TCA-cycle intermediate effect: chemotaxis-independent but metabolism-dependent control of flagella. This mechanism is surprisingly powerful: it alone can produce movement up gradients of metabolizable compounds, down those of metabolism suppressors, switching between response modes based on metabolic state and finding spatial distributions that maximise combinations of metabolite gradients (Egbert, Barandiaran and di Paolo 2010). If this is the basis of the TCA-cycle-intermediate-response, the mechanism is a powerful addition to chemotaxis for more nuanced environment selection.

If, however, the response mechanism is not linked to chemotaxis, it is possible that its biological purpose is not for metabolically-favourable environment selection. Another explanation for the response comes from the environment within the host plant. During symbiosis, the bacteria are housed within nodule structures, in which the bacteria are fed malate and succinate (Prell and Poole 2006). One of the clear transitions that occurs to bacteria within nodules is the down-regulation of motility. Perhaps the signal for this transition comes from the TCA-cycle intermediates.

If this is the case, it is likely to be linked to a global metabolism control system. One candidate for such a system is the phospho-transfer system (PTS). This system has been shown to be linked to chemotaxis in *E. coli* and to transcriptionally inhibit motility in *Vibrio fischeri* (Lengeler *et al.* 1981; Visick *et al.* 2007). In Rlv3841, the PTS system has both carbon and nitrogen regulatory networks that regulate the bacterial response to environmental conditions. Analysis of links from this system to chemotaxis and testing it as a candidate mechanism behind the TCA-cycle intermediate response is discussed in the following chapter (Section 4).

Regulation of motility under starvation conditions

4.1 Introduction

Energy taxis encompasses all systems used to alter swimming direction in response to metabolically necessary compounds. Also called metabolism-dependent chemotaxis, it includes aerotaxis, phototaxis, redox taxis, taxis to alternative electron acceptors, and chemotaxis to carbon sources (Taylor, Zhulin and Johnson 1999). For example, in *Azospirillum brasilense*, cells are strongly attracted to organic acids and sugars (malate, succinate; fructose, galactose), but this attraction is abolished in the presence of high amounts of their carbon storage polymer (polyhydroxybutyrate) (Alexandre, Greer and Zhulin 2000). In addition, the strength of attraction was correlated with the efficiency of growth the bacteria have on that attractant, such that TCA cycle intermediates had a stronger attraction than sugars and amino acids (Alexandre, Greer and Zhulin 2000). However, among all these varieties of energy taxis, they all transmit motility signals rapidly through the chemotaxis system or generationally through transcriptional control of flagella synthesis (Taylor, Zhulin and Johnson 1999; Alexandre and Zhulin 2001; Visick *et al.* 2007; Schweinitzer and Josenhans 2010). Both of these systems are important during the swim plate assay timescales of 4-7 days.

Another option has been described theoretically: metabolism-based chemotaxis in which metabolic enzymes interact directly with the motility machinery (Egbert, Barandiaran and di Paolo 2010). This may describe the TCA cycle intermediate halting effect that we discovered (see Chapter 3): chemotaxis-independent, but metabolism-dependent modulation of motility. Although it is theoretically suggested that metabolism-based chemotaxis alone can produce a variety of complex behaviours (movement up gradients of metabolizable compounds, down those of metabolism suppressors, switching between response modes based on metabolic state and finding spatial distributions that maximise combinations of metabolite gradients), this would depend on the granularity of the response offered by the Rlv3841 TCA cycle intermediate halting effect. Thus far, the effect appears to be discrete: off and enabling swimming at 2 mM succinate, but on with disabled swimming at 20 mM succinate.

To further probe the responsiveness of the TCA-cycle intermediate halting effect and to find the system/s responsible, we performed the following experiments. Firstly, the swimming reaction to a gradient of concentrations of various carbon sources, including a TCA-cycle intermediate, were characterised. Secondly, the swimming behaviour of the phosphotransferase system (PTS; described in Section 1.8) mutants from both the ManX and PtsN branches were characterised to determine if they were linked to the TCA-cycle intermediate effect. Finally, 9 mutants that did not display the TCA-cycle intermediate effect were isolated from papillations formed from wildtype Rlv3841 inoculated in swimming plates. These mutants were sequenced, enabling the identification of two distinct classes of mutations that block the TCA-cycle intermediate effect. All these mutants were characterised for swimming behaviour under various conditions and complemented with native systems.

4.2 Characterisation of chemotaxis-independent but metabolism-based taxis

We began the investigation into the TCA-cycle intermediate swimming effect by characterising the behaviour of key mutants across varying concentrations of the relevant carbon sources in minimal media (Figure 4.1). The strains chosen were the wildtype strain, *Rhizobium leguminosarum* biovar *viciae* 3841 (Rlv3841); a motile chemotaxis system 1 mutant, *che1* (LM100); a motile chemotaxis system 2 mutant, *che2* (LM400); and finally, a non-motile flagellar mutant, *fliG* (OPS2163). The carbon sources chosen were glucose, which feeds into glycolysis; pyruvate, which is one reaction away from the TCA cycle; and succinate, a TCA cycle intermediate.

Firstly, Rlv3841 and the *che2* mutant tended to form larger halos than the *che1* mutant and larger still than the non-motile *fliG* mutant. This was expected due to the *che1* operon being responsible for the majority of the chemotactic response, with the *che2* operon having only a minor role under these conditions. The non-motile *fliG* mutant behaved as expected.

Rlv3841 displayed similar responses to those recorded in Chapter 3, glucose was found to have a 20% reduction and pyruvate a 30% reduction in swimming halo diameter with increasing solute concentration. For both of these carbon sources, the halos appeared with solid, circular morphology for all motile strains (Figure 4.1B, C, E and F). These reductions appear to be linked to the Che1 chemotaxis system, judging by the absence of a change in the *che1* mutant. Interestingly, the succinate results again show an entirely different trend. Rlv3841 was found to have a 70% reduction in swimming halo diameter with increasing concentration and at high succinate concentrations, the halos were small, with cloudy spots morphology (Figure 4.1D and G). We predicted that the cloudy spots morphology is caused by a halting of motility at high succinate. The morphology could be formed by a two-stage process, firstly, the bacteria consume the succinate after inoculation, eventually releasing motility control, enabling some cells to travel out to uncolonised areas. Once there, the bacteria again sense high succinate and halt, producing a new spot.

In addition, the intermediate concentrations show a transition from clearly distinct strain halos at low concentrations, to indistinguishable at high concentrations. Indeed, at 30 mM succinate, the wildtype behaves as if non-motile and non-chemotactic. A critical observation is that, while *che1* mutants have reduced movement relative to wildtype at low succinate, they still respond with reduced motility to high succinate. This indicates that there is a mechanism for motility control that is independent of *che1* and *che2*.

One possible mechanism is transcriptional control through ActSR. High concentrations of succinate would drive the TCA cycle, pushing high amounts of reductant through the ETC that would enable the activation of ActSR which could in turn inhibit motility through possible transcriptional inhibition of flagellar synthesis. Additionally, high flux through the ETC would produce a strong proton motive force which could have an impact on the operation of the flagellar motor.

Since the swimming plates display only overall colony reactions to media conditions, we also imaged the bacteria to track individual behaviour. The motile fraction was unchanged in high and low glucose (Figure 4.2A), reduced by 15% in high pyruvate (Figure 4.3A) and reduced by 30% in high succinate (Figure 4.4A). This correlates with the relative reductions in halo diameters on plates for each carbon source. This agrees with the hypothesis that high levels of TCA cycle intermediate halt the bacteria.

The swimming speeds were similarly affected, with no change between high and low concentrations of glucose and pyruvate (Figures 4.2B and 4.3B), but with an overall reduction in swimming speed at high succinate (Figure 4.4B). These speeds were calculated as overall mean swimming speed, including tumbles. Although the reduction in swimming speed is likely related to the TCA cycle intermediate response, since it still occurs in *che1* mutants, it cannot explain the entire effect. There is a clear separation between the slow swimmers and the halted bacteria.

The tumble rates also followed a similar pattern, with no difference between high and low glucose (Figure 4.2B), but significant differences were detected between high and low pyruvate and succinate (Figures 4.3B and 4.4B). The mutant strains,

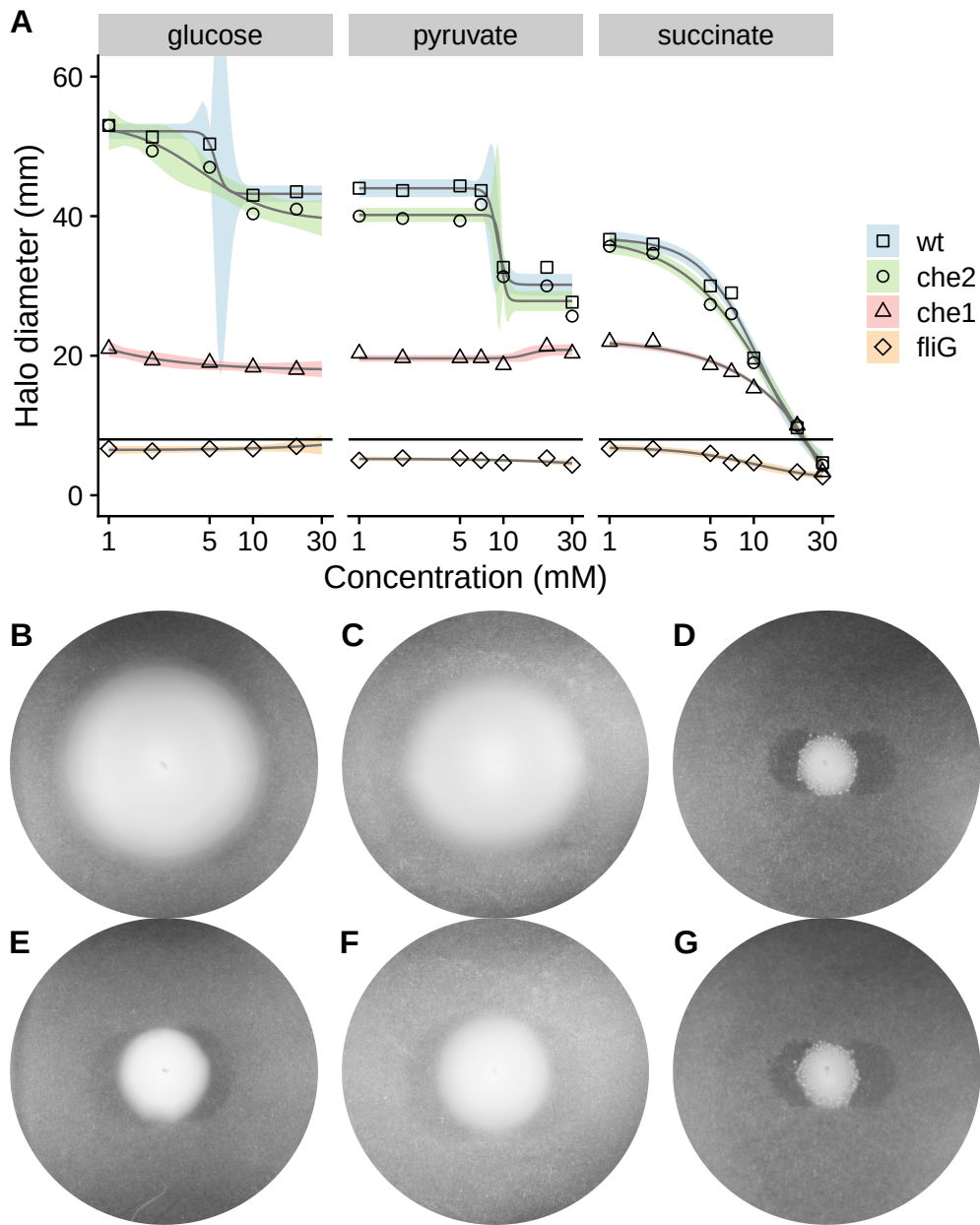


Figure 4.1: Motility mutant swimming in gradients of carbon sources. Swimming halo diameters of wt (Rlv3841) vs *che1* (LM100), *che2* (LM400) and *fliG* (OPS2163) on minimal media with varying concentrations of glucose, pyruvate and succinate. **A** Data was modelled with 4-parameter Log-logistic or linear models for each strain and carbon source where appropriate. Strains wt, *che2* and *che1* were motile except under high succinate concentrations, whereas *fliG* was non-motile under all conditions. Shapes are mean. Lines and ribbons are model predictions with 95% CI. N=3, horizontal line represents estimated maximum non-motile diameter. **B** Representative photo of wt on high glucose. **C** Representative photo of wt on high pyruvate. **D** Representative photo of wt on high succinate. **E** Representative photo of *che1* on high glucose. **F** Representative photo of *che1* on high pyruvate. **G** Representative photo of *che1* on high succinate.

however, behave differently. Deletion of the *che1* operon results in a significant reduction in tumbles and removes the difference between high and low succinate concentrations. This indicates that the tumble change in the wildtype is due to chemotaxis and thus cannot be responsible for the reduced swimming.

4.3 The TCA-cycle intermediate halting effect is specific to *R. leguminosarum*

A variety of soil bacteria were tested for their swimming response to high and low succinate. These included *Rhizobium leguminosarum* biovar *viciae* VF39 (RlvVF39), *Rhizobium leguminosarum* biovar *phaseoli* 4292 (Rlp4292), *Sinorhizobium meliloti* RCR2011, *Sinorhizobium meliloti* Rm2011 str3 (Sme2011), *Sinorhizobium meliloti* 1021 (Sme1021), *Azorhizobium caulinodans* ORS571, *Azospirillum brasilense* FP2, *Pseudomonas fluorescens* SBW25 and *Kosakonia radicincitans* DSM 16656 (Kos). Of these, the small, cloudy spots morphology was observed only in the various biovars of *Rhizobium leguminosarum* (Figure 4.5). The restriction to a single species indicates that it may be an adaptation to symbiosis with particular legumes or particular soils that the *Rhizobium leguminosarum* strains were isolated from.

4.4 Testing of PTS as a candidate for the halting response

As mentioned previously (Section 3), the phosphotransferase system (PTS) is a global metabolic control system that regulates bacterial responses to environmental conditions. The system has been shown to be linked to chemotaxis in *E. coli* and to transcriptionally inhibit motility in *Vibrio fischeri* (Lengeler *et al.* 1981; Visick *et al.* 2007). In Rlv3841, the PTS system has both carbon and nitrogen regulatory networks that regulate the bacterial response to environmental conditions. We thus chose to test the system as a candidate mechanism behind the TCA-cycle intermediate halting effect. Since the ManX branch of the PTS is typically involved with carbon metabolism, it was investigated first.

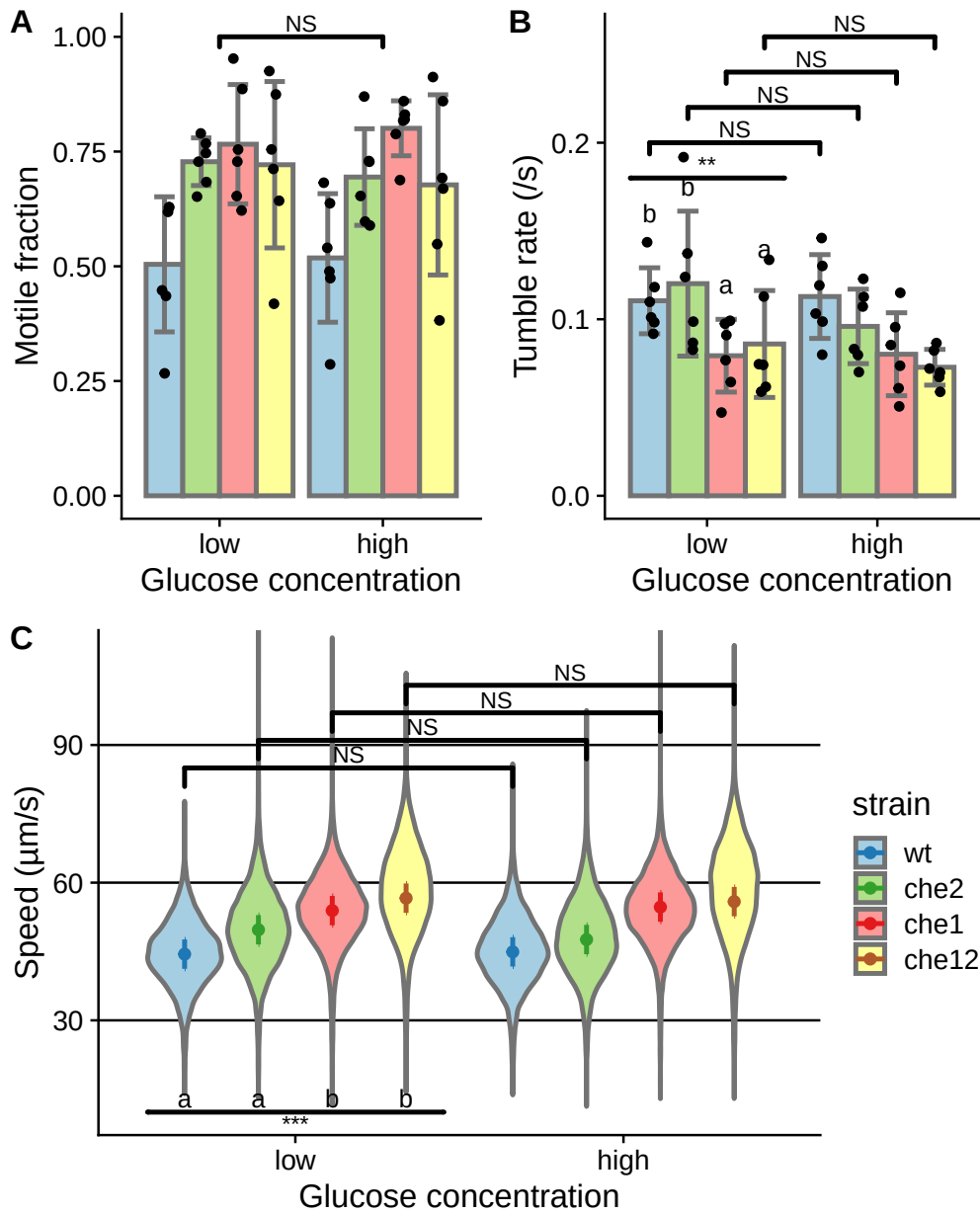


Figure 4.2: Glucose swimming tracks. Bacterium tracking analysis in minimal media (UMS) capillaries of wt (Rlv3841) vs *che1* (LM100), *che2* (LM400) and *che1,2* (LM300). *motA* (OPS2107) and *fliG* (OPS2163) were also tested but had no motile bacteria. Analysis from three independent experiments with 14,000-38,000 tracks per group. **A** Fractions of motile tracks. Data was modelled with a Gaussian linear model with parameters strain and media and their interaction. Strain was significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found no differences between media levels (brackets). **B** Average tumble rate of motile bacteria. Data was modelled as in A, using generalised least squares fit to model unequal variance. Strain was significant, $P < 0.01$, ANOVA. Tukey's Post-hoc test found differences between strains (letters) Dunnett's Post-hoc test found no differences between media levels (brackets). **C** Average swimming speed of motile bacteria. Data was modelled as in A, with random intercept by technical replicates. Strain was significant, $P < 0.001$, ANOVA. Tukey's Post-hoc test found differences between strains. Dunnett's Post-hoc test found no differences between media levels (brackets). Bars are mean $\pm$ SD, density is across all individual tracks and points are mean $\pm$ 95% CI, $N = 3$. Significance levels are NS $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

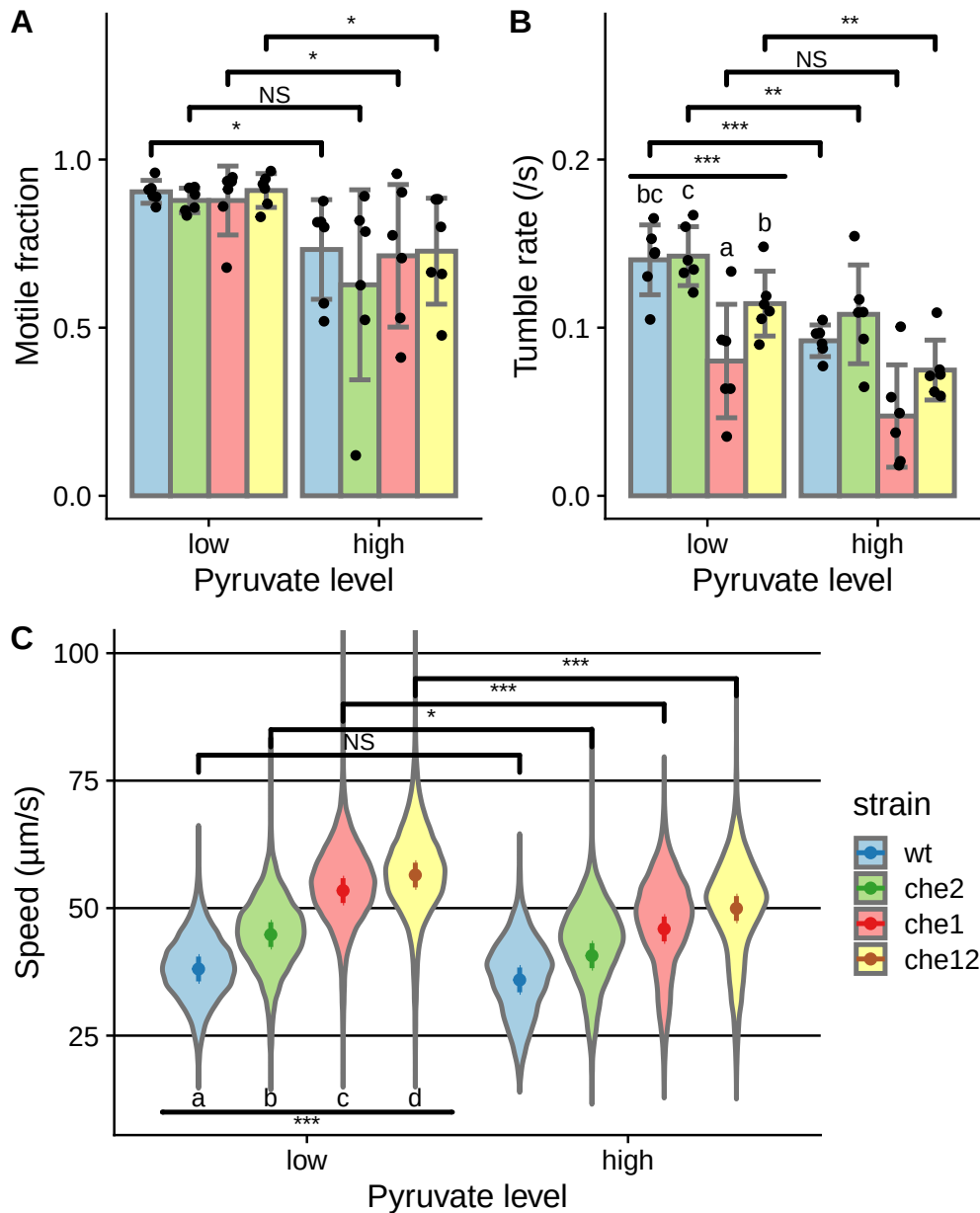


Figure 4.3: Pyruvate swimming tracks. Bacterium tracking analysis in minimal media (UMS) capillaries of wt (Rlv3841) vs *che1* (LM100), *che2* (LM400) and *che1,2* (LM300). *motA* (OPS2107) and *fliG* (OPS2163) were also tested but had no motile bacteria. Analysis from three independent experiments with 7,500-40,000 tracks per group. **A** Fractions of motile tracks. Data was modelled with a Gaussian linear model with parameters strain and media and their interaction. Media was significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found differences between media levels (brackets). **B** Average tumble rate of motile bacteria. Data was modelled as in A, using generalised least squares fit to model unequal variance. Strain and media were significant, $P < 0.0001$, ANOVA. Tukey's Post-hoc test found differences between strains (letters). Dunnett's Post-hoc test found differences between media levels (brackets). **C** Average swimming speed of motile bacteria. Data was modelled as in A, with random intercept by technical replicates. Strain and media were significant, $P < 0.001$, ANOVA. Tukey's Post-hoc test found differences between strains. Dunnett's Post-hoc test found differences between media levels (brackets). Bars are mean $\pm$ SD, density is across all individual tracks and points are mean $\pm$ 95% CI, $N=3$. Significance levels are NS $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

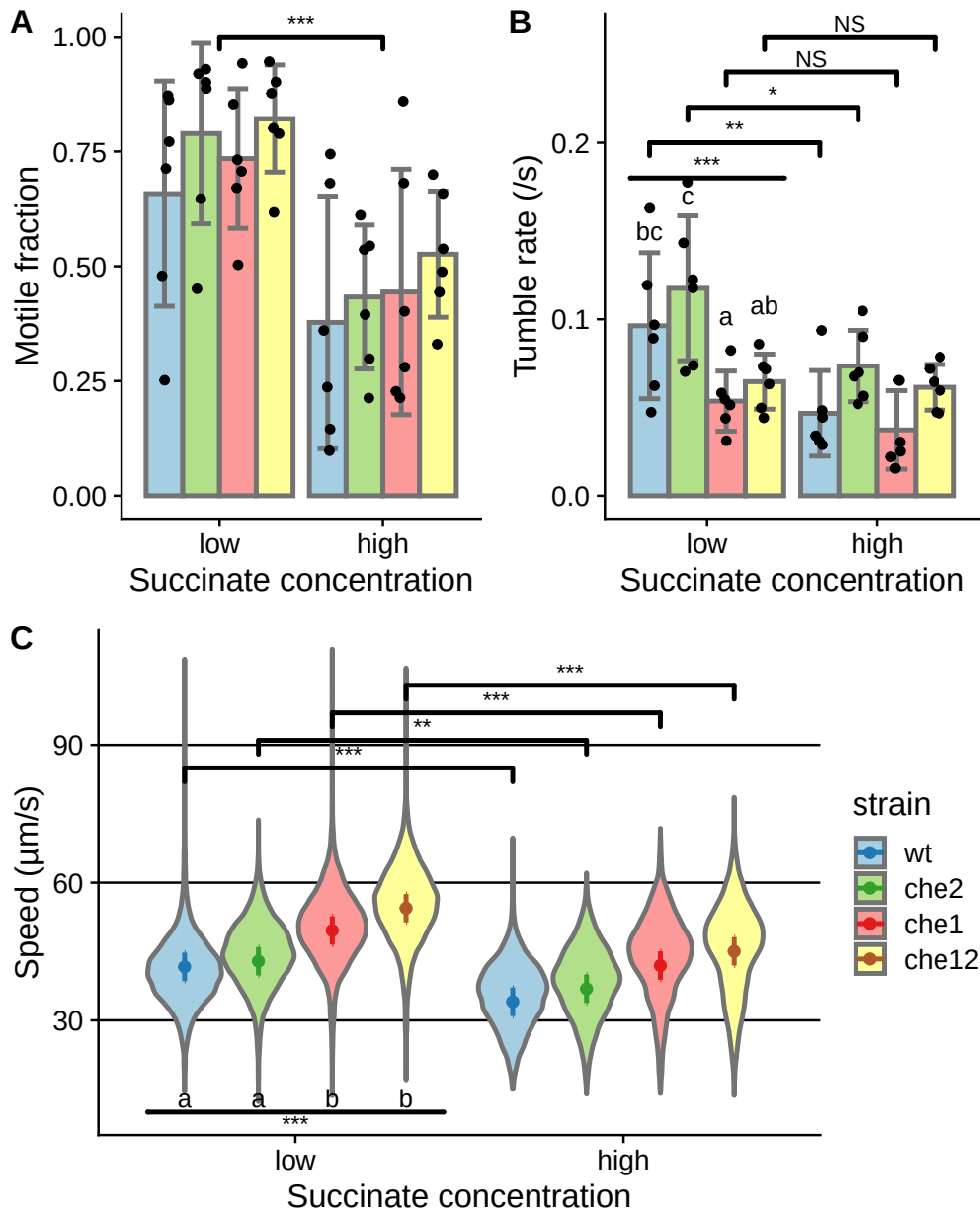


Figure 4.4: Succinate swimming tracks. Bacterium tracking analysis in minimal media (UMS) capillaries of wt (Rlv3841) vs *che1* (LM100), *che2* (LM400) and *che1,2* (LM300). *motA* (OPS2107) and *fliG* (OPS2163) were also tested but had no motile bacteria. Analysis from three independent experiments with 3,800-25,000 tracks per group. **A** Fractions of motile tracks. Data was modelled with a Gaussian linear model with parameters strain and media and their interaction. Media was significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found differences between media levels (brackets). **B** Average tumble rate of motile bacteria. Data was modelled as in A, using generalised least squares fit to model unequal variance. Strain and media were significant, $P < 0.001$, ANOVA. Tukey's Post-hoc test found differences between strains (letters). Dunnett's Post-hoc test found differences between media levels (brackets). **C** Average swimming speed of motile bacteria. Data was modelled as in A, with random intercept by technical replicates. Strain and media were significant, $P < 0.0001$, ANOVA. Tukey's Post-hoc test found differences between strains. Dunnett's Post-hoc test found differences between media levels (brackets). Bars are mean $\pm$ SD, density is across all individual tracks and points are mean $\pm$ 95% CI, $N=3$. Significance levels are NS $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

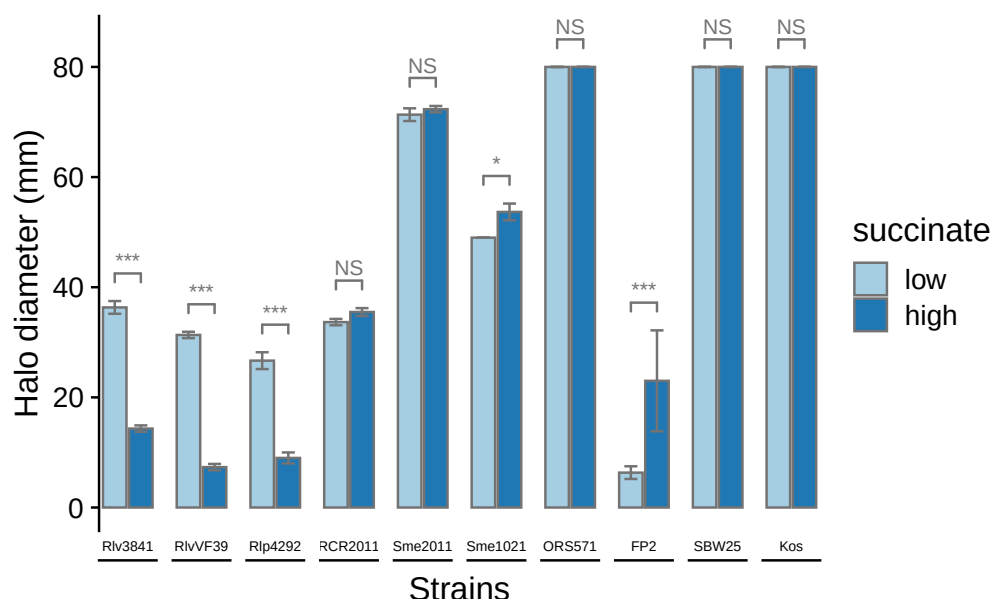


Figure 4.5: Broad-range bacterial succinate responses. Swimming halo diameters for *Rhizobium leguminosarum* biovar *viciae* 3841 (Rlv3841) vs *Rhizobium leguminosarum* biovar *viciae* VF39 (RlvVF39), *Rhizobium leguminosarum* biovar *phaseoli* 4292 (Rlp4292), *Sinorhizobium meliloti* RCR2011, *Sinorhizobium meliloti* Rm2011 str3 (Sme2011), *Sinorhizobium meliloti* 1021 (Sme1021), *Azorhizobium caulinodans* ORS571, *Azospirillum brasilense* FP2, *Pseudomonas fluorescens* SBW25 and *Kosakonia radicincitans* DSM 16656 (Kos). Data was modelled with a Gaussian linear model with the factors strain, media and their interaction. All factors were found to be significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found differences in media within each strain. Bars are mean $\pm$ SD, $N=3$. Significance levels are NS $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Interestingly, there was a clear link to halo size, with the *manX* mutant displaying significantly smaller halos than wildtype on all media tested. This was complemented by the non-phosphorylatable mutant (H9A) version of ManX (ManX*), but not by the phosphomimic version (H9D; ManX*P) (Figure 4.6). These results, however, may be compounded by the known growth defect of *manX*, which may be the cause of the smaller halos (S'anchez-Cañizares *et al.* 2020). Importantly, the *manX* mutant still displays the TCA-cycle intermediate halting response, formed the typical small, cloudy spots morphology halos on high succinate plates. The ManX branch can thus be ruled out as being the direct cause of the halting.

The other branch of the PTS, PtsN, is typically associated with nitrogen metabolism. To test the swimming response to various nitrogen sources, washed agar was used to remove excess nitrogen (Section 4.7.4). Similar results were found

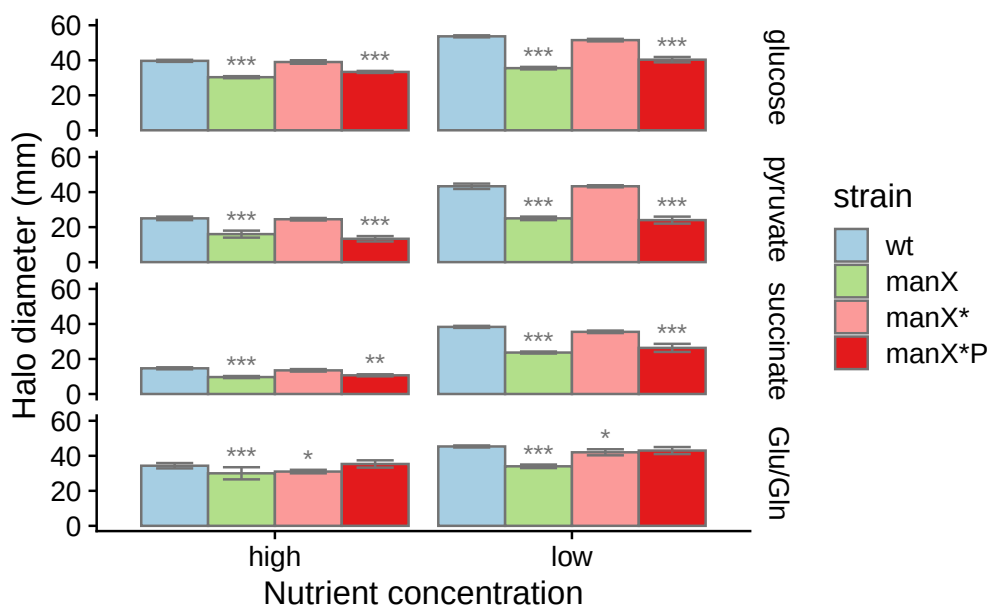


Figure 4.6: *manX* mutant plates. Swimming halo diameters for *Rhizobium leguminosarum* biovar *viciae* 3841 (Rlv3841) vs LMB692 (*manX*), OPS1012 (*manX**), OPS1099 (*manX*P*). Data was modelled with a Gaussian linear model with the parameters strain, nutrient, nutrient level and their interactions. All parameters were found to be significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found differences in strain compared to wildtype within each media condition. Bars are mean $\pm$ SD, $N=3$. Significance levels are NS $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

as above, with the *ptsN1* mutant displaying significantly smaller halos than wildtype, but, in reverse to *ManX*, being complemented by the phosphomimic (H66D; *PtsN*P*) and only partially by the non-phosphorylatable protein (H66A; *PtsN**). In addition, since *ptsN1* mutants do not have a growth defect, this is good evidence for a connection between the PTS system and motility (S'anchez-Cañizares *et al.* 2020). However, all of these mutants still perform the TCA-cycle intermediate halting response and thus the *PtsN* branch also can be ruled out of being the direct cause.

4.5 Mutant isolation screening for TCA-intermediate swimming control

Since the PTS system does not regulate reduced swimming at high concentrations of TCA-cycle intermediate, an isolation screen was developed to identify the controlling system/s. It was noted previously that incubating swimming plates

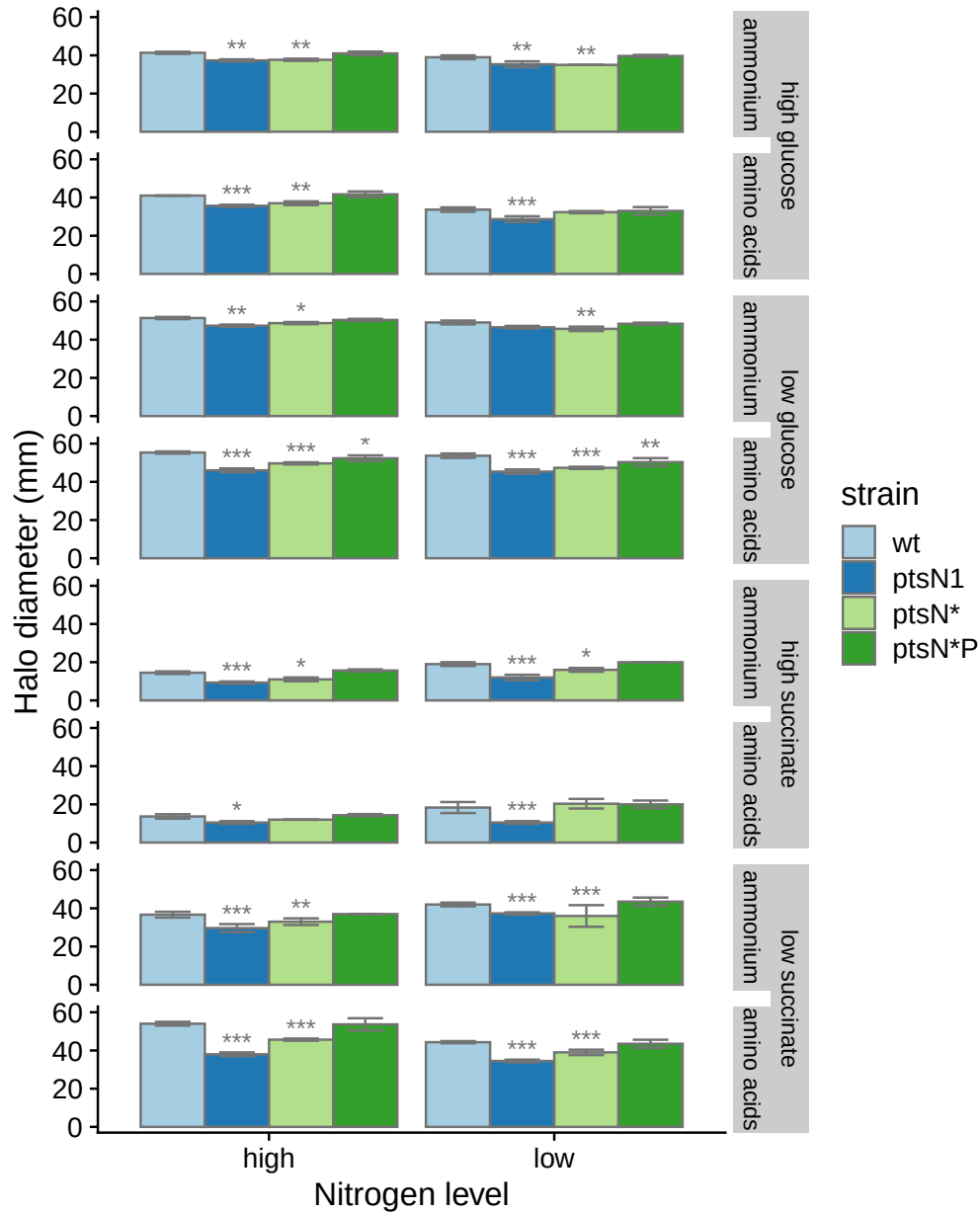


Figure 4.7: *ptsN* mutant plates. Swimming halo diameters for *Rhizobium leguminosarum* biovar *viciae* 3841 (Rlv3841) vs LMB271 (ptsN1), OPS1101 (ptsN*), OPS1103 (ptsN*P) in minimal media with 1.1% washed agar. Data was modelled with a Gaussian linear model with the parameters strain, nutrient, nutrient level and their interactions. All parameters were found to be significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found differences in strain compared to wildtype within each media condition. Bars are mean $\pm$ SD, $N=3$. Significance levels are NS $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

supplemented with high TCA-intermediate for periods longer than 4 days produced small papillations, protrusions out from the central halo, presumably caused by mutations in the control system (see Figure 4.14A top-right side). Fourteen such mutants were thus isolated to enable screening to find the genetic basis for the control system (see Section 4.7.1). Of these, 10 displayed wildtype-like swimming behaviour in glucose media but did not display the TCA-intermediate effect in high fumarate, malate or succinate media. These 10 were then sequenced, in addition to the wildtype and an isolate from the centre of one of the halos (see Section 4.7.2). The sequenced reads were aligned to the wildtype Rlv3841 sequence, allowing identification of mutations that were not present in the wildtype or the centre isolate (Table 4.1; see Section 4.7.3).

Table 4.1: SNP and INDEL mutations found in papillation isolates.

element	gene	mutant	position	wildtype	mutation
RLChr	inter-RL0072	OPS2237	93799	CG	CGTGGTT...
RLChr	inter-RL1270	OPS2238	1335005	A	C,<*>
RLChr	inter-RL3891	OPS2298	4120525	G	T,<*>
RLChr	inter-RL3891	OPS2290	4120539	G	C,<*>
RLChr	RL0542	OPS2300	586583	CG	CGCACGG...
RLChr	RL0543	OPS2300	587321	GCTTGCC...	GC
RLChr	RL0543	OPS2290	587340	TCC	TC
RLChr	RL0545	OPS2295	590435	G	T,<*>
RLChr	RL0545	OPS2297	590435	G	T,<*>
RLChr	RL0545	OPS2292	590463	C	A,<*>
RLChr	RL0675	OPS2236	725741	GAA	GA
RLChr	RL1061	OPS2295	1141455	G	A,<*>
RLChr	RL1094	OPS2238	1176749	A	T,<*>
RLChr	RL1471	OPS2237	1540201	CG	CGGCGAA...
RLChr	RL1584	OPS2237	1653614	C	A,<*>
RLChr	RL1824	OPS2292	1916585	G	A,<*>
RLChr	RL1824	OPS2294	1916585	G	A,<*>
RLChr	RL1824	OPS2295	1916585	G	A,<*>
RLChr	RL1825	OPS2237	1917685	C	T,<*>
RLChr	RL2750	OPS2298	2910792	A	AC
RLChr	RL3464	OPS2237	3635515	T	G,<*>
RLChr	RL3464	OPS2294	3635539	T	A,<*>
RLChr	RL3464	OPS2238	3635647	C	G,<*>
RLChr	RL3464	OPS2298	3635755	T	C,<*>
RLChr	RL3467	OPS2238	3638880	CT	C

Table 4.1: (continued)

element	gene	mutant	position	wildtype	mutation
RLChr	RL3496	OPS2290	3669336	T	A,<*>
RLChr	RL3496	OPS2300	3669359	G	T,<*>
RLChr	RL4742	OPS2294	5057141	AT	ATCAGCC...
pRL9	pRL90016	OPS2297	16484	G	A,<*>
pRL8	inter-pRL80142	OPS2292	147462	TC	TCGTGGA...
pRL8	pRL80001	OPS2294	4	GA	GACGTCTG...
pRL8	pRL80001	OPS2238	4	GA	GACGGAG...
pRL7	inter-pRL70047E	OPS2300	45485	TCCTC	TC
pRL7	inter-pRL70047E	OPS2237	45734	CA	CAAGGAG...
pRL7	inter-pRL70047E	OPS2295	45832	AA	AACAGGT...
pRL7	pRL70103	OPS2290	87844	A	T,<*>
pRL7	pRL70107	OPS2237	91998	A	T,<*>
pRL12	inter-pRL120460	OPS2290	495152	C	T,<*>
pRL12	pRL120344	OPS2298	370407	A	T,<*>
pRL12	pRL120564	OPS2290	616419	A	C,<*>
pRL12	pRL120564	OPS2297	616419	A	T,<*>
pRL11	pRL110270	OPS2238	293258	G	A,<*>
pRL11	pRL110270	OPS2297	293258	G	A,<*>
pRL11	pRL110276	OPS2237	300438	A	G,<*>
pRL11	pRL110276	OPS2297	300438	A	G,T,<*>
pRL11	pRL110583	OPS2290	628515	A	T,<*>
pRL10	pRL100001	OPS2290	6	GA	GATCATT...
pRL10	pRL100088	OPS2298	87468	GG	GGG
pRL10	pRL100152	OPS2290	151684	A	T,<*>

A total of 69 genes or intergenic regions were found to have at least one SNP or INDEL in at least one mutant. A number of these regions had multiple SNP/INDELs per mutant. The high variability of these genes was assumed to be due to sequencing errors or errors in contig formation due to sequence similarity. This was confirmed by alignment of individual reads to the contigs in question. These genes were thus eliminated from further analysis. This left 33 genes and intergenic regions that had SNP/INDELs confirmed by read analysis (listed in Table 4.1).

Within these genes, a clear pattern was observed: half of the mutants had mutations in one gene and the other half all had mutations in another, separate cluster (Table 4.2). The gene RL3464 had SNPs observed in the mutants OPS2237, OPS2238, OPS2294 and OPS2298. RL3464 encodes a putative SAM-dependent

methyltransferase of unknown function.

The second cluster of mutations were observed in the genes RL0540-7. The gene RL0542 had an INDEL observed in the mutant OPS2300 (22 bp inserted). The next gene RL0543 had INDELs observed in the mutants OPS2290 (1 bp deleted) and OPS2300 (52 bp deleted). The final gene RL0545 had SNPs observed in the mutants OPS2292, OPS2295 and OPS2297. These genes encode a system for inorganic phosphate uptake that is active under low phosphate conditions. The genes RL0542-5 produce an ABC transporter for inorganic phosphate uptake. Specifically, RL0542 is *pstS*, a periplasmic inorganic phosphate binding protein; RL0543 is *pstC*, an integral membrane protein forming the ABC transporter with RL0544, *pstA*; finally, RL0545 is *pstB*, the ABC subunit ATP binding protein (Yuan, Zaheer and Finan 2006).

Table 4.2: The mutations of RL0540-7 and RL3464 discovered in papillation isolates.

Strain	Gene	Gene (bp)	DNA mutation	Protein result
OPS2237	RL3464	320	T-G transversion	L107R
OPS2238	RL3464	452	C-G transversion	T151R
OPS2290	RL0543	102	1 bp deletion	frameshift (after 34aa)
OPS2292	RL0545	407	C-A transversion	S136* (truncating 136aa)
OPS2294	RL3464	344	T-A transversion	L115H
OPS2295	RL0545	379	G-T transversion	G127C
OPS2297	RL0545	379	G-T transversion	G127C
OPS2298	RL3464	560	T-C transition	L187P
OPS2300	RL0542	544	22 bp insertion	frameshift (after 181aa)
OPS2300	RL0543	83-134	52 bp deletion	frameshift (after 27aa)

To confirm the mutations, the RL0543 mutants were sequenced with *exp4636+exp4637*, the RL0545 mutants with *exp4638+exp4639* and the RL3464 mutants with *exp4640+exp4641* (see Section 2.3.5). The wildtype and the centre isolate were also sequenced with all primer sets to confirm absence of mutations. The individual sequencing confirmed all mutations (Figure 4.8). In-frame deletion mutants of each cluster were attempted to be isolated following the methods used previously, however, neither were successful from 40 colonies screened.

The phosphate uptake mutants were grown in minimal media under high and low phosphate conditions to compare their growth rate to wildtype. Each strain

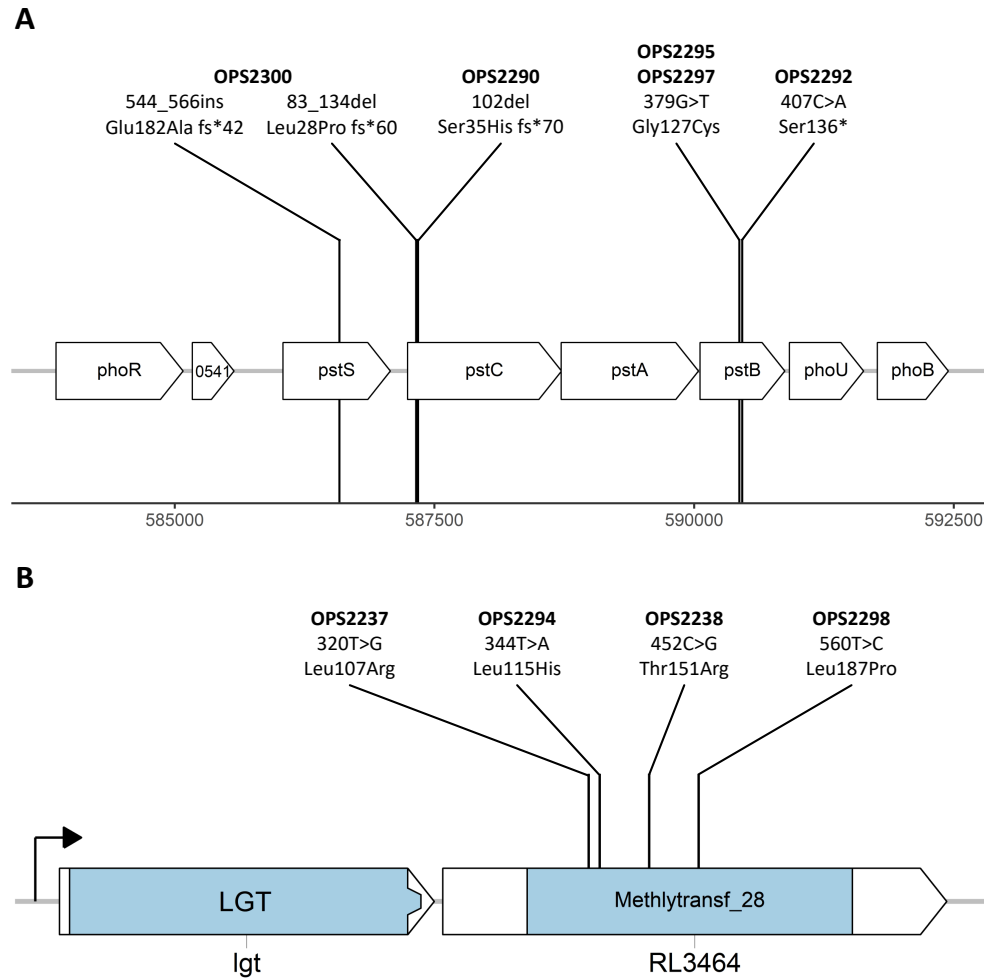


Figure 4.8: Papillation isolate mutations. Primary mutations found in papillation isolates, ordered by occurrence in the genome. **A** Single nucleotide polymorphisms and insertion/deletions found in RL0542-RL0545 (*pstSCAB*). **B** Single nucleotide polymorphisms found in RL3464 (putative SAM-dependent methyltransferase). *lgt* (RL3463) is predicted to encode a putative prolipoprotein diacylglycerol transferase. Both genes are predicted to be under the same promoter, shown by an arrow. Pfam annotation of domains are in blue.

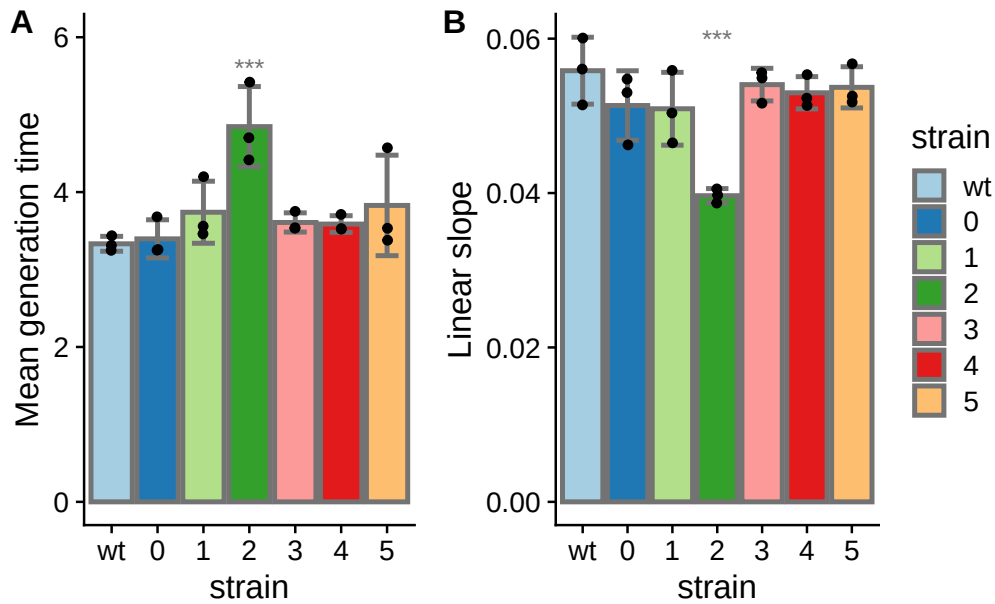


Figure 4.9: Papillation isolates growth rate. Growth rate in high and low phosphate for Rlv3841 (wt) vs OPS2301 (0), OPS2290 (1), OPS2292 (2), OPS2295 (3), OPS2297 (4) and OPS2300 (5). **A** High phosphate growth conditions. Data was modelled with individual Exponential linear models for each strain. Mean generation time was calculated from those models. **B** Low phosphate growth conditions. Data was modelling with individual Gaussian linear models for each strain. Dunnett's Post-hoc test found differences in strains compared to wildtype. Bars are mean $\pm$ SD, N=3. Significance levels are NS $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

grew exponentially in high phosphate ($500 \mu\text{M}$) media and so the mean generation time was calculated and compared. However, in low phosphate ($50 \mu\text{M}$), the strains grew linearly and so the slope of that growth was compared. None of the phosphate uptake mutants grew significantly differently to wildtype except for OPS2292, which was significantly slower growing under both conditions (Figure 4.9). OPS2292 has a large nonsense mutation (truncating half of the protein) in *pstB*, the ATP binding protein of the ABC transporter, so would be expected to have greatly reduced phosphate uptake. Interestingly, OPS2300, which has a 52 bp deletion with frame-shift, leaving only 27 amino acids intact of *pstC*, an integral part of the ABC transporter, had no growth defects, despite also likely having greatly reduced phosphate uptake. This may mean that the disrupted uptake of phosphate by the *pstSCAB* system is compensated by other phosphate uptake systems but that the major mutation in OPS2292 has an additional toxic effect.

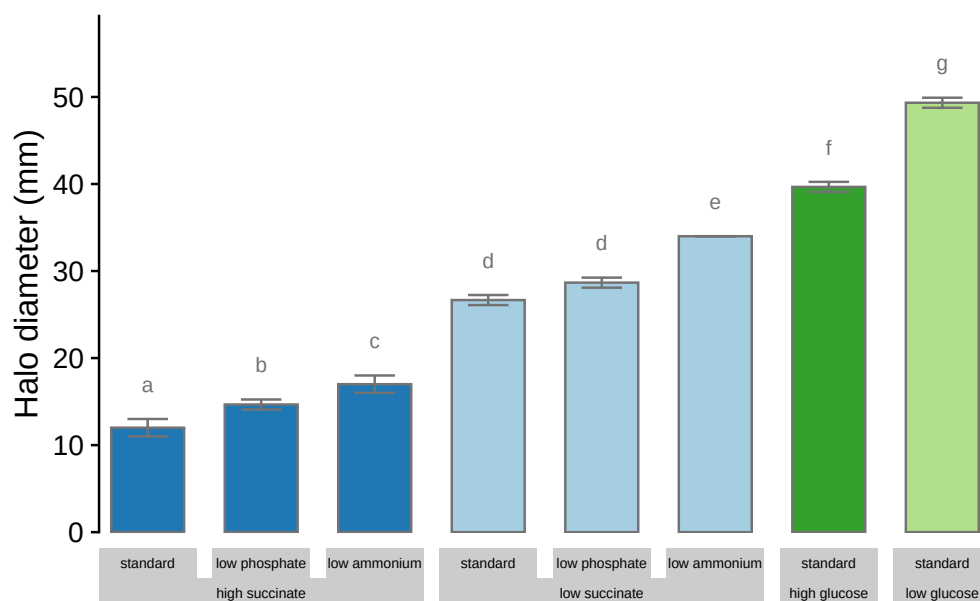


Figure 4.10: Rlv3841 reacts to nutrient concentrations. Swimming halo diameters of Rlv3841 in minimal media (UMS) with either 0.22% agar (with glucose) or 1.1% washed agar (with succinate). Data was modelled with a Gaussian linear model with the parameters carbon, misc and their interaction. All parameters were found to be significant, $P < 0.01$, ANOVA. Tukey's Post-hoc test was used to stratify data within strains into groups, where those groups not sharing a letter were significantly different at $\alpha = 0.05$. Bars are mean $\pm$ SD, $N = 3$.

4.5.1 Swimming responses of the two classes of mutation

The swimming behaviour of the papillation isolates was tested on varying concentrations of carbon sources (glucose and succinate) and other nutrients (ammonium and phosphate). Washed agar was again used to remove excess nutrients. The limiting phosphate conditions of the low ammonium media with washed agar were confirmed with the activation of alkaline phosphatase fused to GFP (using strain RU1235). Wildtype Rlv3841 again displayed increased swimming on low concentrations of carbon sources compared to high concentrations (Figure 4.10). In addition, at high succinate, there were significant increases in swimming with reduced concentrations of both phosphate (23% increase) and ammonium (42% increase). In contrast, at low succinate, there was only a significant increase with reduced concentrations of ammonium (27% increase).

The swimming behaviour of the papillation isolates were tested on the same

media conditions as above, varying concentrations of carbon sources (glucose and succinate) and other nutrients (ammonium and phosphate) (Figure 4.11). The diameters were then compared to the wildtype within each media condition, with significance differences noted by asterisks.

Firstly, with high concentrations of succinate as the carbon source, and with high concentrations of ammonium and phosphate, the strains behaved as expected, with the papillation isolates displaying increased halo diameters compared to the wildtype (60-70% for the phosphate uptake mutants and 90-130% for the methyltransferase mutants). In low ammonium and low phosphate media conditions, though the differences were reduced (30-90% increases over wildtype), the papillation isolates were all significantly larger than wildtype. This indicates that TCA-cycle intermediate halting response does not also require high nitrogen or high phosphate conditions. This is also further evidence that the phosphate uptake mutations are not altering the transporter functionality.

In low succinate media, the strains were again larger than wildtype with about 20% increases for phosphate uptake mutants and 20-50% increases for methyltransferase mutants. Moving to the low phosphate media conditions, the pattern was similar to on high succinate, with increases in the wildtype diameter causing the increase of the mutants to be lessened, although still larger than wildtype. The low ammonium media, however, displayed a different pattern, with the wildtype diameter increasing to the level of the phosphate uptake mutants. Since the low succinate media has only a mild halting effect, the increased swimming due to low ammonium was sufficient to overcome the difference.

Both the high and low glucose media displayed similar results. Although many of the mutants displayed significant differences to the wildtype, there were roughly equal numbers larger and smaller. The largest differences were an increase of 27% on high glucose and 14% on low glucose media by different methyltransferase mutants. Thus, in general, the papillation isolates only have increased swimming in the presence of TCA-cycle intermediates. This suggests that the effect of the mutations is limited to the halting effect and not flagella-based swimming in general.

One common thread throughout the media conditions is that the methyltransferase mutants tend to have a more dramatic increase in halo diameter than the phosphate uptake mutants. This may indicate that the methyltransferase is more directly linked to the TCA-cycle halting effect such that mutations more effectively prevent the halting of swimming.

The papillation mutants were also tracked as before, swimming in minimal media at high and low succinate concentrations (Figure 4.12). In general, it appears that the mutants had an increased motile fraction over the wildtype, as large as 20 percentage points on high succinate, although only three strains were significant due to high variance (Figure 4.12A). Similarly, in the tumble rate data there was high variance, but the general trend of an increase in tumble rate for the methyltransferase mutant (Figure 4.12B). Overall, the swimming speed was reduced in high succinate (Figure 4.12C). These results support the hypothesis that the mutations cause an increase in motile fraction by preventing the halting of swimming caused by sensing of high TCA-cycle intermediates.

Next, the isolates were complemented using a Tn7 integration system to integrate the entire operon, with native promoters, in which their respective mutation was found. In addition, a constitutively active promoter was fused to both the wildtype and a mutant (OPS2294) methyltransferase and integrated into each methyltransferase mutant and the wildtype. As a control, a blank Tn7 was also integrated into each mutant.

We first determined the swimming behaviour of these strains on high succinate (Figure 4.13). As before, the non-integrated mutants (plain) all had significantly larger halos than the wildtype. The negative controls (empty Tn7 integrated) behaved similarly to the non-integrated mutants, with large, standard halos. The native complementation recovered the wildtype halo diameters, being reduced in all of the mutants. Native complementation also recovered the cloudy spots morphology, although increased frequency of papillations on the complemented halos was noted (Figure 4.14F and G).

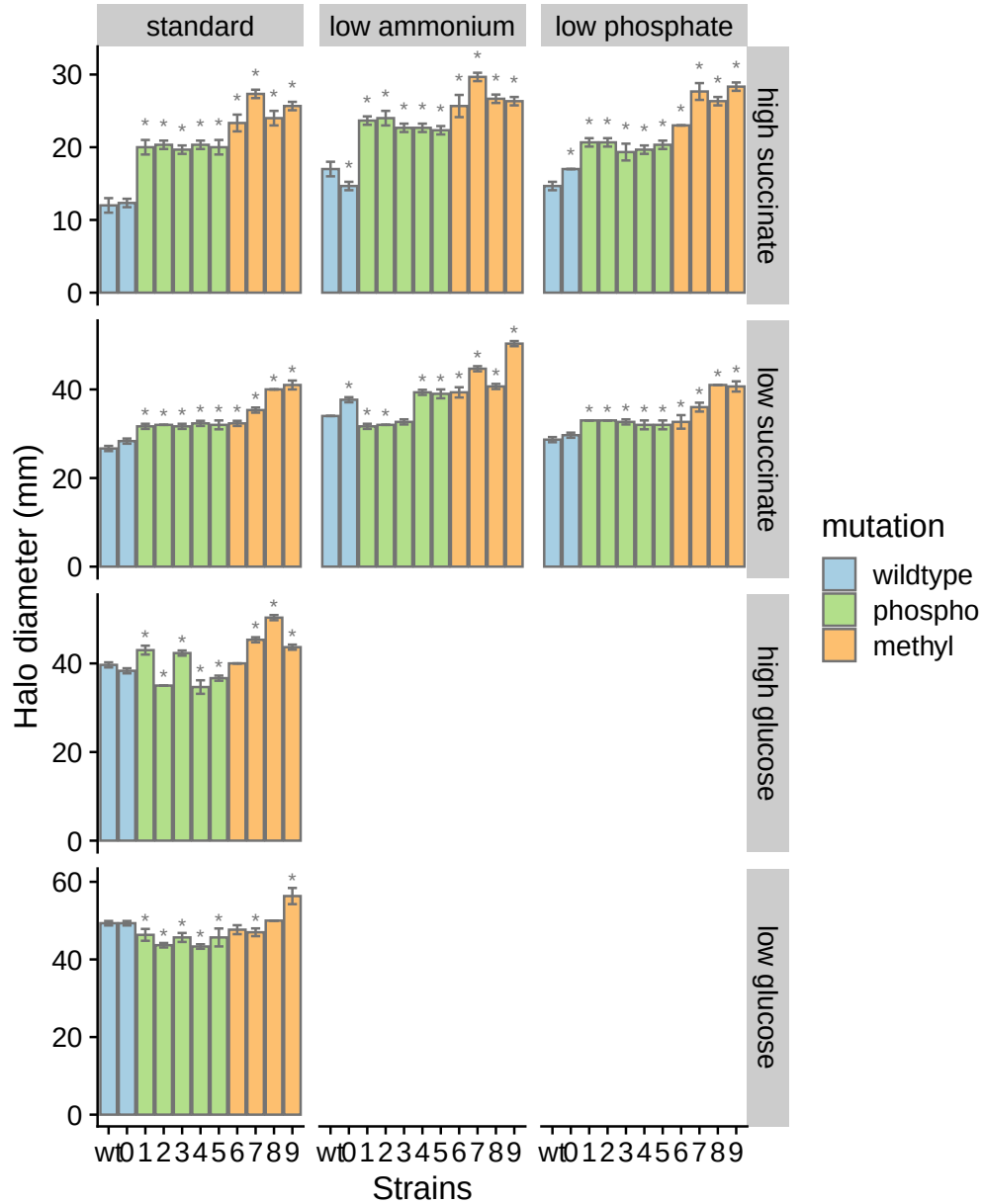


Figure 4.11: Papillation isolates on swimming plates. Swimming halo diameters of wt (Rlv3841) vs the phosphate uptake system mutants OPS2290 (1), OPS2292 (2), OPS2295 (3), OPS2297 (4) and OPS2300 (5) and the methyltransferase mutants OPS2237 (6), OPS2238 (7), OPS2294 (8) and OPS2298 (9) in minimal media (UMS) with either 0.22% agar (with glucose) or 1.1% washed agar (with succinate). Data was modelled with a Gaussian linear model with the parameters strain, carbon, misc and their interactions. All parameters were found to be significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found differences in strain compared to wildtype. Bars are mean $\pm$ SD, $N=3$. Significance levels are NS $P > 0.05$, * $P < 0.05$.

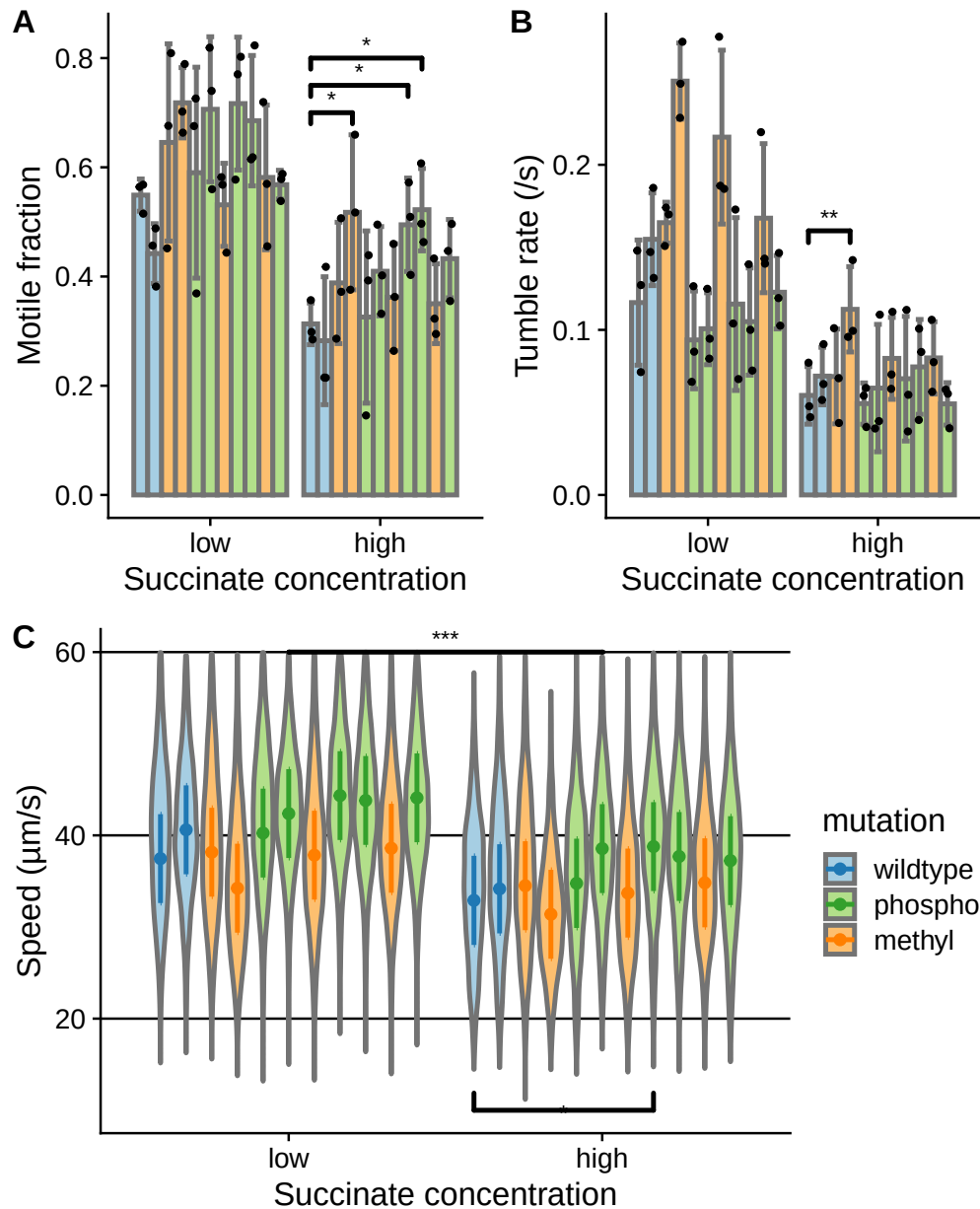


Figure 4.12: Papillation isolate swimming tracks. Bacterium tracking analysis in minimal media (UMS) capillaries of wt (Rlv3841) vs papillation isolates OPS2301, OPS2237, OPS2238, OPS2290, OPS2292, OPS2294, OPS2295, OPS2297, OPS2298 and OPS2300. Analysis from three independent experiments with 1,300-6,400 tracks per group. **A** Fractions of motile tracks. Data was modelled with a Gaussian linear model with parameters strain and media and their interaction. The parameters strain and media were significant, $P < 0.01$, ANOVA. Dunnett's Post-hoc test found differences between strains (brackets). **B** Average tumble rate of motile bacteria. Data was modelled as in A, using generalised least squares fit to model unequal variance. The parameters strain and media were significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found differences between strains (brackets). **C** Average swimming speed of motile bacteria. Data was modelled as above, with random intercept by technical replicates. The parameters strain and media were significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found differences between media levels (top bracket) and strains (bottom bracket). Bars are mean $\pm$ SD, density is across all individual tracks and points are mean $\pm$ 95% CI, $N=3$. Significance levels are NS $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

The constitutively expressed wildtype methyltransferase gene also reduced the halo diameter in cognate mutants to that of the wildtype. However, although the smaller halo size was restored by constitutive complementation, the morphology was not. The halos were two-layered, with a small, possibly non-motile centre, surrounded by an outer halo (Figure 4.14I). Although the two layers suggests suppressor mutations, the outer halo appears too uniform and thick to be a secondary mutation of the type that caused the papillations. Interestingly, the constitutive expression of the wildtype methyltransferase gene in the wildtype strain produced even smaller halos than the unmodified wildtype. These halos were the size and morphology of non-motile mutants (Figure 4.14H). Thus, the constitutive levels of methyltransferase produces the same behaviour at 20 mM as wildtype levels do at 30 mM succinate (see Figure 4.1).

Constitutive expression of the mutant methyltransferase gene L115H (mutation in strain OPS2294) did reduce the size of the halos in cognate mutants, but they were all still significantly larger than those in wildtype. Overexpressing the mutant methyltransferase gene is thus insufficient to completely overcome the mutations. This is consistent with the mutations in the methyltransferase impairing its function, although perhaps not completely abolishing its function.

Three of the complemented mutants (OPS2294, OPS2297 and OPS2298) were also tracked as before, swimming in minimal media at high and low succinate concentrations (Figure 4.15). The high variance in motile fraction again make it difficult to interpret, but the general pattern is an increase of motile fraction in the mutants that is unchanged in the negative integration and reduced to wildtype levels with native complementation (Figure 4.15A). For tumble rate, the two methyltransferase mutants (5 and 8) were significantly increased over wildtype in low succinate conditions that was reduced to wildtype levels with native complementation (Figure 4.15B). There were no differences between strains in the swimming speed, so it was not included in the figure for clarity.

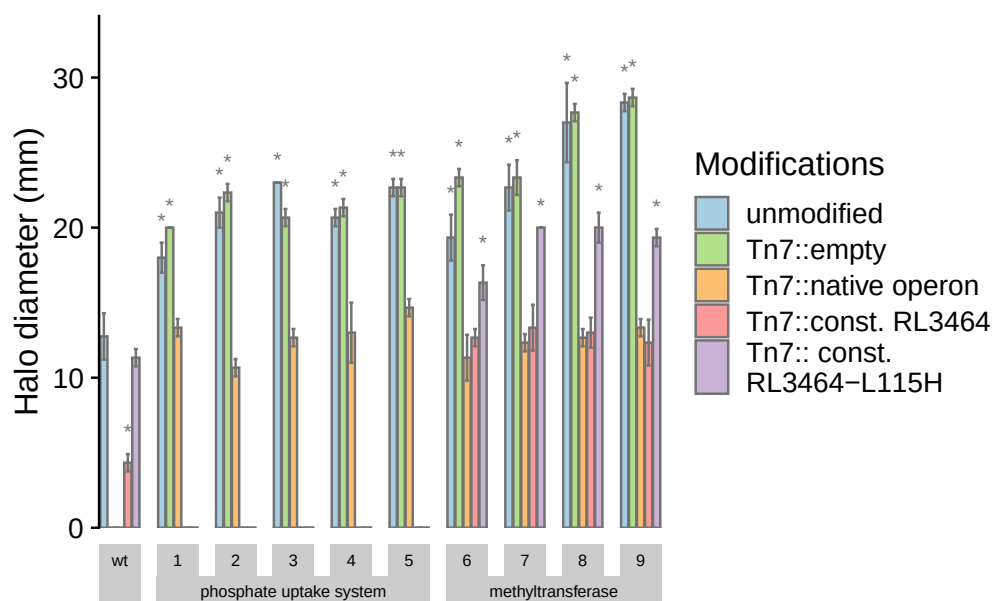


Figure 4.13: Papillation isolate complementation on swimming plates. Swimming halo diameters of wt (Rlv3841) vs the phosphate uptake system mutants OPS2290 (1), OPS2292 (2), OPS2295 (3), OPS2297 (4) and OPS2300 (5) and the methyltransferase mutants OPS2237 (6), OPS2238 (7), OPS2294 (8) and OPS2298 (9) on high succinate minimal media (UMS). Data was modelled with a Gaussian linear model with the parameters strain, media, Tn7 and their interactions. All parameters were found to be significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found differences in strain compared to wildtype. Bars are mean $\pm$ SD, $N=3$. Significance levels are NS $P > 0.05$, * $P < 0.05$.

4.6 Discussion

In this chapter, we first described the swimming behaviour of *Rhizobium leguminosarum* biovar *viciae* 3841 (Rlv3841) wildtype, chemotaxis and flagellar mutants in varying concentrations of carbon sources. As in the previous chapter, the TCA-cycle intermediate halting effect was found to independent of both chemotaxis system, occurring in the *che1* and *che2* mutants (Figure 4.1. All three strains displayed the reduction in swimming halo starting at 5 mM succinate until 30 mM where they were the same size as the non-motile strain. At intermediate concentrations, the halos displayed cloudy spots, indicating halting behaviour that is dynamically responsive to the current succinate concentration. We then proceeded to investigate the phospho-transferase system (PTS) as a candidate controlling system.

Although both the ManX and PtsN branches appeared to alter motility, mutants

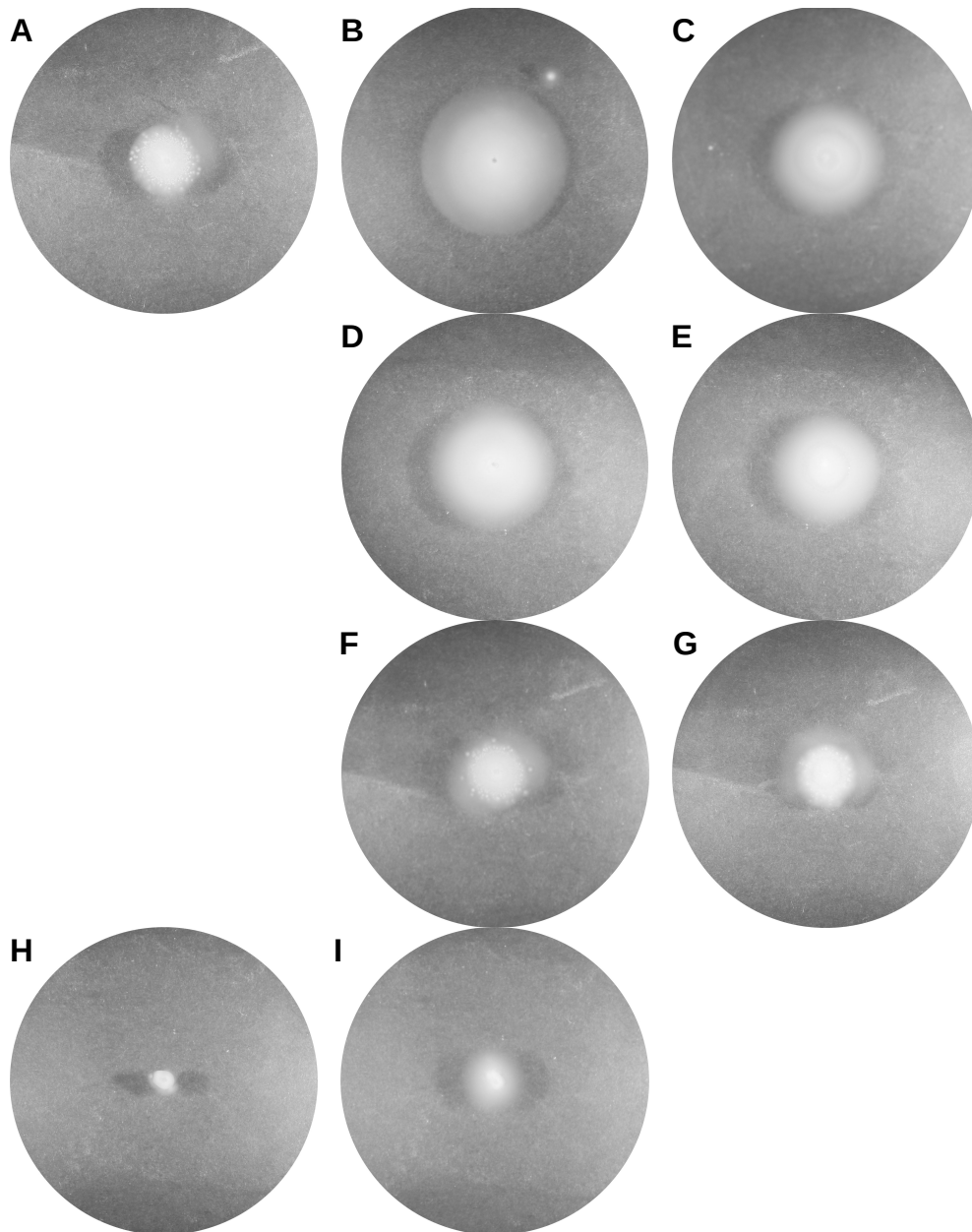


Figure 4.14: Photos of papillation isolate complementation. Swimming halos of wt (Rlv3841) vs OPS2237 (methyltransferase mutant) and OPS2297 (phosphate uptake mutant) on high succinate media from Figure 4.11. Representative photo of wt (**A**, **H**), the methyltransferase mutant OPS2237 (**B**, **D**, **F**, **I**) and the phosphate uptake system mutant OPS2297 (**C**, **E**, **G**). The strains are visualised without Tn7 (**A**, **B**, **C**), with empty integration (**D**, **E**), with native complementation (**F**, **G**) and with constitutively expressed methyltransferase gene (**H**, **I**).

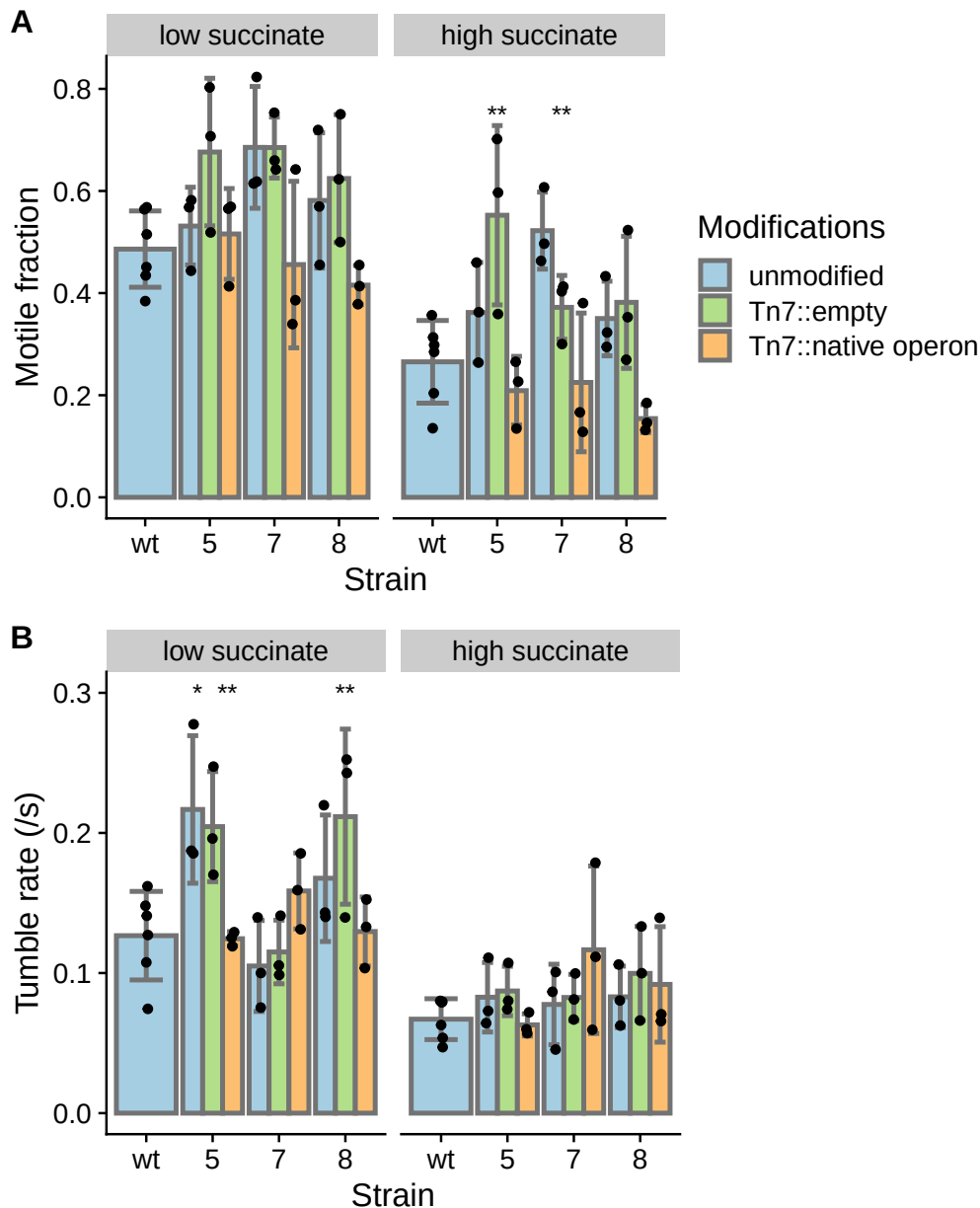


Figure 4.15: Papillation isolate complementation swimming tracks. Bacterium tracking analysis in minimal media (UMS) capillaries of wt (Rlv3841) vs papillation isolates OPS2294 (5), OPS2297 (7) and OPS2298 (8). Analysis from three independent experiments with 1,500-8,000 tracks per group. Wildtype overexpressing RL3464 had no motile tracks. There were no significant differences in swimming speed. **A** Fractions of motile tracks. Data was modelled with a Gaussian linear model with parameters strain, Tn7 and media and their interactions. The parameters strain, Tn7 and media were significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found significant differences to wildtype. **B** Average tumble rate of motile bacteria. Data was modelled as in A, using generalised least squares fit to model unequal variance. The parameters strain and media were significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found significant differences to wildtype. Bars are mean $\pm$ SD, $N=3$. Significance levels are NS $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

in both branches still displayed the halting effect (Figure 4.6 and 4.7), suggesting that PTS is not responsible for this effect. The *manX* mutant displayed reduced swimming behaviour in each media condition and was complemented by a non-phosphorylatable version of the protein (H9A; ManX*) but not by the phospho-mimic version (H9D; ManX*P). In contrast, although the *ptsN1* mutant also displayed reduced swimming behaviour in each media condition, it was complemented by the phospho-mimic version of the protein (H66D; PtsN*P), but not by the non-phosphorylatable version (H66A; PtsN*). Since both proteins tend to be phosphorylated under the same conditions, this indicates a branching control of motility. Both proteins tend to be phosphorylated under nitrogen limitation and carbon excess, and not phosphorylated under nitrogen excess and carbon limitation (S'anchez-Cañizares *et al.* 2020). So under nitrogen limitation, perhaps PtsN-P stimulates motility, whereas in carbon limitation, ManX stimulates motility. In summary, PTS effects appear to be separate from the high succinate restriction in motility.

We then isolated mutants that no longer displayed the TCA-cycle intermediate halting effect by recovering bacteria from papillations formed by the wildtype in swimming plates containing a high concentration of fumarate (20 mM). Sequencing 9 of these mutants along with the wildtype and a neutral isolate revealed two general types of mutation: mutants in a putative SAM-dependent methyltransferase (RL3464) and mutants in an ABC transporter for inorganic phosphate uptake (*pstSCAB*, RL0542-5). Both types of mutants were found to abolish the cloudy spots morphology in high TCA-cycle intermediate media and cause increased halo diameters (Figure 4.14B and C and Figure 4.11). In addition, the mutants tended to have an increased motile fraction at high succinate over the wildtype (Figure 4.12A). Finally, these effects were restored to wildtype by complementation with the native operons, again forming small, cloudy halos with papillations in high succinate media (Figure 4.13 and Figure 4.14F and G).

4.6.1 The phosphate transporter PstSCAB

Rhizobium leguminosarum encodes four phosphate transporters: *pit1* (RL0857), *pit2* (pRL100343), *phoCDET* (RL0170, RL0168-6) and *pstSCAB* (RL0542-5). In *E. coli*, the PstSCAB system has two orders of magnitude higher affinity for inorganic phosphate than Pit, but is induced only in phosphate limiting conditions, being inhibited by >1 mM phosphate (Rosenberg, Gerdes and Chegwidden 1977). In *Sinorhizobium meliloti* and *S. fredii*, the Pit transporter is also low affinity and is unnecessary for symbiosis, whereas the PhoCDET and PstSCAB transporters are high affinity and, although coredundant, are together necessary for effective symbiosis and fixation (Yuan, Zaheer and Finan 2006; Hu *et al.* 2018). In addition, the PstSCAB transporter is associated with the two-component system PhoRB in both *E. coli* and *S. meliloti*. Under limiting phosphate conditions, PstSCAB transports phosphate, the linker protein PhoU is dissociated and PhoRB is active (Makino *et al.* 1989; Hsieh and Wanner 2010; diCenzo *et al.* 2017) (Figure 4.16). Under excess phosphate conditions, PhoU binds both the transporter to reduce transport and PhoR to prevent phosphorylation of PhoB. When phosphorylated, PhoB transmits the phosphate limitation signal by activating and repressing a wide range of genes through binding of a Pho box motif (Krol and Becker 2004). Interestingly, many basal body and motor genes were downregulated in limiting phosphate but this appeared to be independent of PhoB (Krol and Becker 2004). In contrast, in *Vibrio cholerae*, PhoB regulates *acgAB* which encodes c-di-GMP metabolic enzymes which in turn positively regulate motility (Pratt, McDonough and Camilli 2009).

In *S. meliloti* Rm2011, the effect of PhoB is only active under limiting phosphate conditions. Interestingly, in the related *S. meliloti* Rm1021, PhoB is active even under excess phosphate conditions. This strain has a 1 bp insertion in the *pstC* gene, causing a frame-shift that presumably alters the conformation of the transporter such that PhoU cannot be activated as normal under excess phosphate conditions (Krol and Becker 2004; Yuan, Zaheer and Finan 2006). This matches two of the mutations observed in the papillation isolates, both having deletions in *pstC* causing frame-shift. Together with the SNP found in *pstB*, the mutants are unlikely to

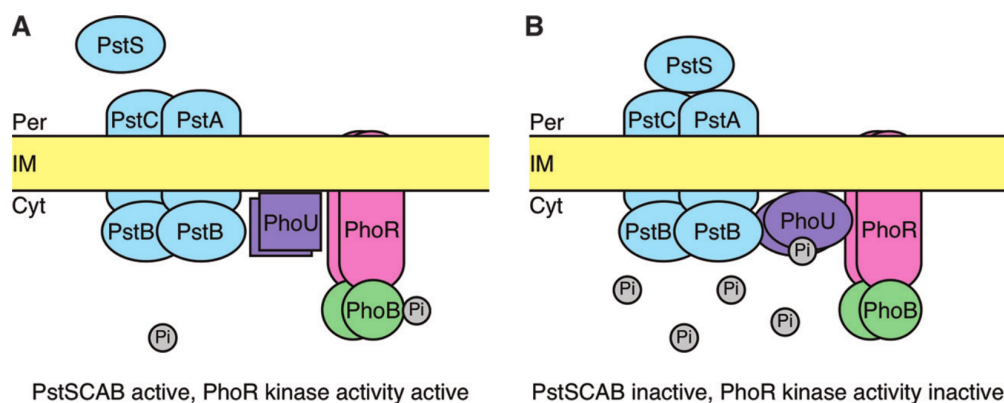


Figure 4.16: PhoU regulation of PstSCAB transporter and PhoRB two-component system. Model proposed by diCenzo *et al.* (2017) from *Sinorhizobium meliloti*. Dynamic equilibrium between state **A**, with phosphate limitation, low transport of PstSCAB, dissociation of PhoU and activation of PhoB by PhoR; and state **B**, with phosphate excess, which would increase transport by PstSCAB, except association of PhoU to PstB reduces PstSCAB transport and also association with PhoR prevents activation of PhoB. Reproduced from diCenzo *et al.* (2017).

effect global phosphate uptake (only one had a growth defect in low phosphate as shown in Figure 4.9), instead they may change the activity of PhoB via PhoU in a similar way to Rm1021 such that PhoB is constitutively phosphorylated and active. If so, it could be expected that the halting effect would also be avoided in the wildtype under phosphate limiting conditions, which would also cause increased activity of PhoB. This was, however, not the case, with only a minor, though significant, increase in wildtype halo diameter under phosphate limitation (Figure 4.11). The effect of these mutations is thus likely to be through an unknown independent regulatory role of some part of the system, perhaps the PhoU protein. Alternatively, a constitutively active PhoB-P may be a more potent activator of motility than when PhoB is phosphorylated transiently in low phosphate agar. Presumably as cells move through the agar to new reserves of phosphate, PhoB may be dephosphorylated and only phosphorylated when the phosphate is consumed.

To determine how many of these effects are active in Rlv3841, mutants in the relevant genes could be made. Firstly, a deletion of the entire *pstSCAB* operon or one of the integral genes, could be used to check co-redundancy with PhoCDET and determine if growth defects are expected with loss of PstSCAB function. Secondly,

a deletion of *phoU* would be expected to produce a strain with constitutively active PhoB and PstSCAB. If constitutive PhoB is indeed responsible for the loss of the halting effect in the mutants with *pstC* mutations, then this strain would again be expected to not halt under high TCA cycle intermediate concentration. Phosphorylation of PhoB is via an aspartate residue, causing a steric change in the effector domain, activating the protein (Bachhawat, Swapna and Stock 2005). Thus it would be difficult to produce a phosphomimic mutant of PhoB, since phosphomimic versions usually substitute an aspartate for a native residue. Finally, a deletion of *phoB* itself could influence the substrate specificity or concentration sensitivity of the halting effect, if PhoB is indeed a negative regulator of the effect.

4.6.2 Methyltransferase

In contrast to PstSCAB, nearly nothing is known about the protein encoded by RL3464 except its Pfam classification as a putative (S)-Adenosyl-L-methionine (SAM)-dependent methyltransferase and its shared operon with *lgt* (RL3463), a putative prolipoprotein diacylglycerol transferase. Mutations in the gene, which presumably alter the function of the translated protein, prevented the TCA-cycle intermediate halting effect. A particularly dramatic confirmation of this was constitutive expression of the gene in wildtype cells causing a non-motile phenotype.

Its classification as a methyltransferase suggests a possible connection with the methyl-accepting chemotaxis proteins (MCPs). MCPs are chemoreceptors that are methylated by CheR, ultimately to increase phosphorylation of CheY in response to higher levels of stimuli and thus increase tumbles through slowing flagella rotation (see Section 1.5). Together with demethylation by CheB, the two proteins control adaptation to stimulus concentration. Methylation of these receptors by the methyltransferase encoded by RL3464 could then also cause increased phosphorylation of CheY and slowed flagella rotation, perhaps to the point of halting through constitutive phosphorylation of CheY. However, the methyltransferase encoded by RL3464 had low identity across the methyltransferase domain (<40%) to both CheR1 and CheR2 and did not contain the response regulator domain

found in both CheRs. In addition, both the *che1* and *che2* operon deletions and the *che1,2* double deletion still displayed the halting effect (Figures 4.1 and 3.3). Thus, it is expected that the methyltransferase has a different regulatory role, somehow being linked to motility, perhaps post-translationally via the flagella motor or transcriptionally via VisNR.

To determine conclusively if the halting effect is through post-translational control of flagella or transcriptional down-regulation, the flagella could be stained and counted under high and low TCA-cycle intermediates to check for loss of flagella. In addition, a culture growing in low succinate conditions could be shocked with high succinate and their swimming tracked immediately afterwards to determine the kinetics of halting. If the change can occur within a few minutes, then it is highly unlikely to be transcriptional control, which would be expected to occur via dilution from cell division, thus taking several hours.

Bioinformatic analysis of relatives of the methyltransferase encoded by RL3464 allowed construction of a phylogenetic tree (Figure 4.17). The tree grouped methyltransferases from various *Rhizobium* strains together, each with >92% amino acid identity with the methyltransferase from Rlv3841. Another clade was formed of methyltransferases from various *Sinorhizobium* strains, each with about 70% identity with the methyltransferase from Rlv3841. The various other methyltransferases had 50% or less identity with the methyltransferase from Rlv3841. The lack of halting effect in Rm2011 is thus possibly due to differences in the methyltransferase protein function. Another explanation is that Rm2011 is missing or contains altered versions of downstream elements that are responsible for causing the halting. Expression of the Rlv3841 RL3464 gene in Rm2011 could reveal which is the case—if the strain displays halting, then it is likely to be changes in the methyltransferase that are important, if not, then there are likely to be other altered elements.

Future experiments could discover downstream elements or confirm connections between the methyltransferase and the candidate protein targets proposed above, revealing which of the scenarios is accurate. Firstly, RL3464 could be expressed in an *E. coli* or Rlv3841 vector to produce tagged, purified protein that could be

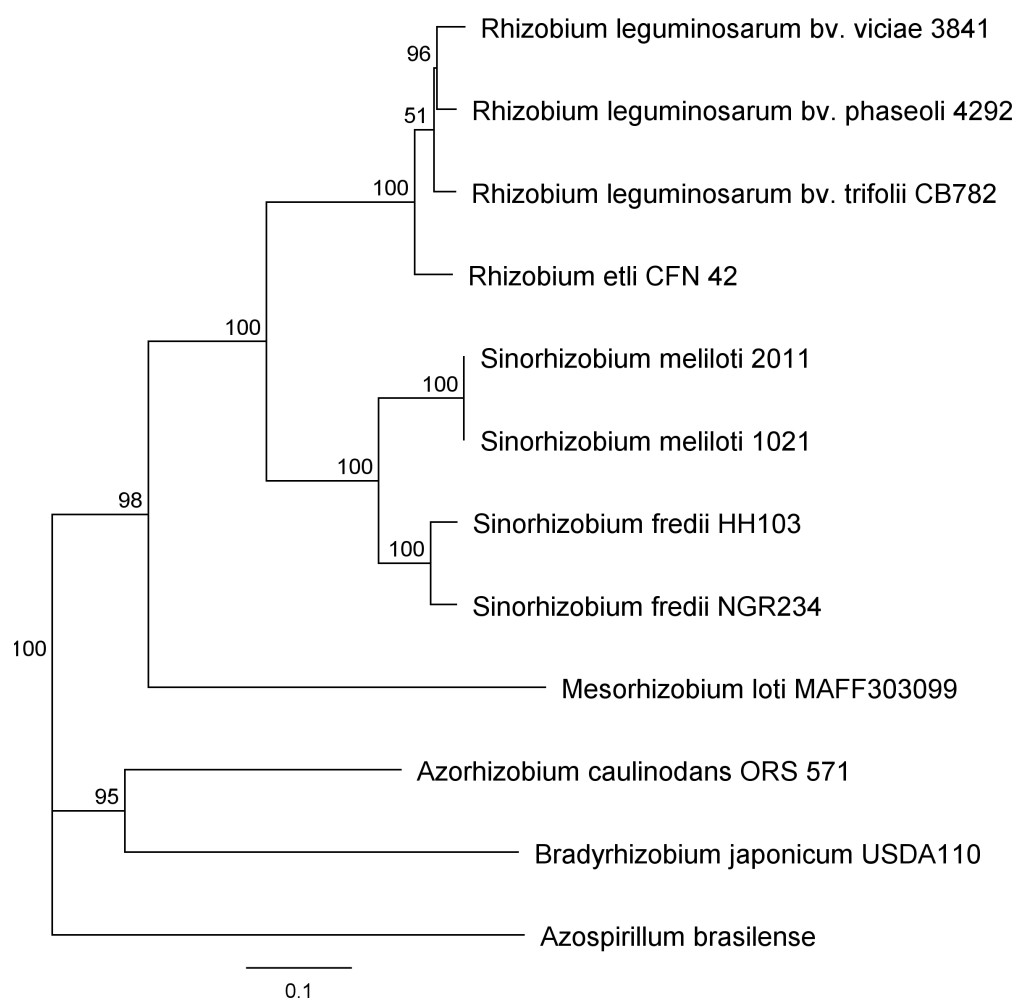


Figure 4.17: Neighbour-joining consensus tree of RL3464 methyltransferase related protein sequences. Sequences chosen by BLAST identity and aligned using MUSCLE. Tree generated using neighbour-joining algorithm with 100 bootstrap replicates.

used in cell lysate pull-downs as a binding screen. These pull-downs could also determine if any methylation is detected on any of the targets via mass spectrometry. Secondly, more focused two-hybrid assays could find if any of the proposed binding interactions occur. If any methylation targets are discovered, further experiments could determine the kinetics of the reaction under different conditions, to see if the methyltransferase somehow responds to TCA cycle intermediates directly (Dorgan *et al.* 2006; Ibáñez *et al.* 2010). In addition, cross-complementation with the native PstSCAB operon could reveal if the two systems are linked together or acting independently. If the expression of the methyltransferase gene is somehow influenced by PstB, then complementation of the methyltransferase in the phosphate

uptake mutants may recover halting by recovering a sufficient level of expression of the methyltransferase gene.

4.7 Methods

4.7.1 Random mutagenesis screening

Mixed colonies of Rlv3841 were inoculated into the centre of semi-solid agar UMS media plates supplemented with 20 mM fumarate and 10 mM ammonium (as per Section 2.7.3). The plates were incubated for 11 days at 28°C until papillations extended out of the inner halos were formed. Bacteria were isolated from one papillation per plate and double colony purified on TY agar plates. These isolates were then stocked at -80°C as OPS2236, OPS2237, OPS2238, OPS2290, OPS2291, OPS2292, OPS2293, OPS2294, OPS2295, OPS2296, OPS2297, OPS2298, OPS2299 and OPS2300. An additional isolate was taken from the centre of one of the halos and stocked as OPS2301.

4.7.2 Whole genome sequencing

Single colonies of Rlv3841, OPS2236, OPS2237, OPS2238, OPS2290, OPS2292, OPS2294, OPS2295, OPS2297, OPS2298, OPS2300 and OPS2301 were individually resuspended in 100 μ L sterile TY. These were streaked out on TY agar plates to form a lawn of bacteria on 1/3 of the plate and streaked further across the rest of the plates. The plates were incubated for three days at 28°C. After confirming the absence of colony impurities in the streaks, the bacterial cultures were removed with sterile loops into the barcoded bead tubes supplied by MicrobesNG. The tubes were sealed, mixed by inverting 10 times and sent at room temperature to MicrobesNG for sequencing.

4.7.3 SNP analysis

Contigs from the strains sequenced above (Section 4.7.2) were aligned against the genome of Rlv3841 (NCBI:txid216596) and mutations (both Single Nucleotide Polymorphisms (SNPs) and Insertion/Deletions (INDELs)) were detected using

a bash pipeline (github.com/AroneyS/3841_contig_SNP_identification). This pipeline uses Docker based BioContainers (da Veiga Leprevost *et al.* 2017) of the following tools: Bowtie2 v2.3.4 (Langmead and Salzberg 2012), Samtools v1.9 (Li *et al.* 2009), BWA v0.7.17 (Li and Durbin 2009), Picard v2.3.0 (Picard toolkit 2019) and BCFtools v1.9 (Li 2011).

4.7.4 Agar washing

Agar was washed to remove excess nutrients by mixing with 0.5 M NaCl to form a slurry. The slurry was then centrifuged at 4,500 rpm for 5 minutes, the supernatant was discarded, replaced with washing solution and shaken at 10,000 rpm for 1 minute. This was repeated 3 times with water and 2 times with distilled water. After the final centrifugation, the pellet was dried overnight at 50 °C. The absorbed water was estimated at 5x weight by weighing the agar before and after washing. Thus 1.1% of washed agar was equivalent to 0.22% agar.

Chemotaxis and motility in symbiosis

5.1 Introduction

As described in Chapter 1, insertion sequencing (INSeq) is a technique that indicates the relative importance of all genes in a genome by generating a library of transposon mutants which ideally saturates the transposon insertion sites in the genome. These mutants are each disrupted in a single gene or area in the genome, and so increases and reductions in their relative abundance indicate the importance of the disrupted region (Perry and Yost 2014; Perry, Akter and Yost 2016).

This technique has been used in our laboratory over the last few years to characterise the genes involved in growth under different carbon sources and oxygen tensions (Wheatley *et al.* 2016), and more recently, to identify the genes involved in the *Rhizobium*-legume symbiosis (Wheatley *et al.* 2020). This last INSeq experiment was conducted for rhizosphere, root colonised and nodule bacteria, and differentiated nodule bacteroids. Our reference strain *Rhizobium leguminosarum* biovar *viciae* 3841 was used as model rhizobia and the base for the transposon library, and the pea plant *Pisum sativum* was used as the model plant and symbiosis partner. For this project, the set of genes related to chemotaxis and motility was analysed across all the experimental conditions. As explained in this chapter, our analysis found that bacterial strains mutated in chemotaxis and motility genes

tended to be neutral for recovery in the rhizosphere and on the root, yet with a highly reduced presence in the nodule. This unexpected lack of importance in the rhizosphere and on the root is likely due to experimental conditions, as discussed below, since chemotaxis and motility has long been shown to be important in these environments (Scharf, Hynes and Alexandre 2016; Kawasaki *et al.* 2016). After these two stages, the role of chemotaxis and motility has not been well-characterised, therefore, in this work we carried out a broader analysis of these two systems across all the stages of the symbiotic process.

To examine all the stages more closely, we used the chemotaxis and motility mutants previously isolated (Chapter 3). These mutants were marked with different reporter genes that allowed us to design experiments able to monitor our cells thanks to assays based on fluorescence, luminescence and enzymatic activity. This enabled us to examine the effects of the mutants on root attachment, colonisation, initiation of symbiosis and nodule development.

5.2 Analysis of chemotaxis and motility genes in insertion sequencing data

Once the INSeq experiment was conducted in our laboratory and the sequence reads were obtained, we carried out the data processing and analysis of the chemotaxis and motility genes (methods described in Section 5.11.1). Of the 93 genes with putative roles in chemotaxis and motility (Chapter 1), 39 are required for nodule formation and function and for 11, their mutation is advantageous for nodule bacteria (Table A.1, Figure 5.1).

The majority of the genes required for the flagellar assembly and basal body apparatus are important in nodules and essential for the two types of rhizobial cells they harbour: undifferentiated nodule bacteria and bacteroids (Table A.1, Figure 5.1). The gene encoding the critical motor protein FliG (RL0700) is neutral across all stages of symbiosis. In contrast, the gene encoding an interacting integral motor protein MotA (RL0703) is required in nodule bacteria and bacteroids. Interestingly, these strains are expected to be flagellated differently, with mutants in the gene

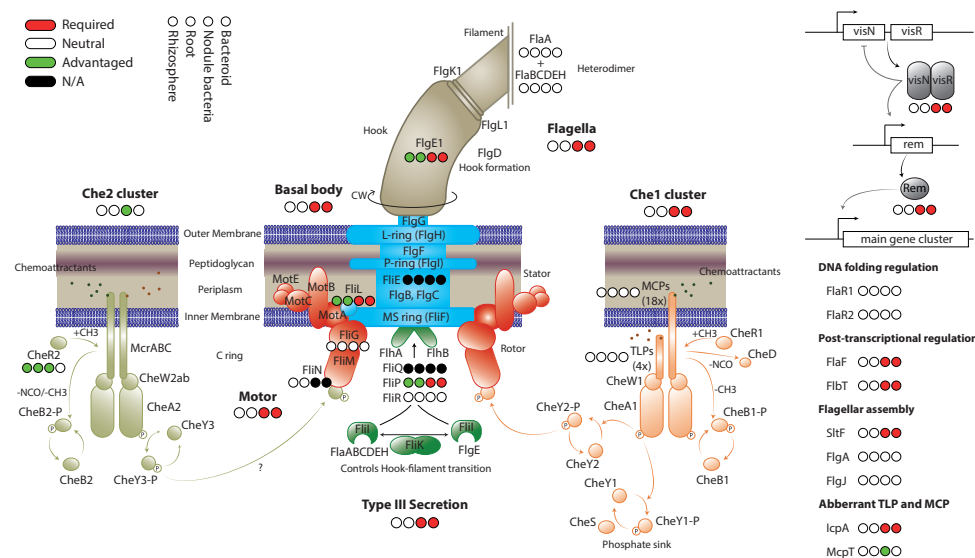


Figure 5.1: Rlv3841 chemotaxis and motility gene requirement at different symbiotic stages. INSeq analysis of the phenotypes of motility and chemotaxis function encoding genes across the rhizosphere, root colonisation, nodule bacteria and bacteroid libraries. Libraries are represented in the diagram as ordered, coloured circles. INSeq phenotypes are represented in colour as red for advantaged (AD), white for neutral (NE), yellow for defective (DE), green for essential (ES) and black for unclassified due to too few TA sites (NA).

encoding *FliG* likely producing non-flagellated strains whereas mutants in the gene encoding *MotA* likely form non-rotating flagella. In addition, the genes encoding the master flagella and chemotaxis regulator pair, *VisNR* (RL0696 – RL0697), are required in nodule bacteria and bacteroids, although when examined in detail, the C-terminal GerE-encoding domain of *visN* displays a dramatic increase in transposon insertions, suggesting its deletion does not impair *VisN* function and may even enhance it. A number of flagellar biosynthesis-encoding genes required in nodule bacteria are not essential in the rhizosphere and for root colonization, or their mutation is even advantageous. Non-motile bacteria may become stranded in the rhizosphere or on the root surface, unable to continue to symbiosis.

In the case of the chemotaxis systems, each of the individual genes in the *che1* cluster (RL0686-93) are required only for nodule bacteria and differentiated bacteroids (Table A.1, Figure 5.1). However, this cluster was not essential at the initial stages for optimal colonisation of the rhizosphere or pea plant roots. This suggests that, under these experimental conditions, the *Che1* system is required at

a later stage of symbiosis, either for chemotaxis specifically to root hairs or early progression into infection threads. An experimental investigation of the initiation of symbiosis at the root hairs and its progression was subsequently performed and described in Section 5.6.

Mutations in all the genes present in the *che2* cluster (RL4028-36) are advantageous for undifferentiated nodule bacteria. While the reason is not known, we speculate that the *che2* gene cluster is involved in the later stages of bacterial movement through infection threads. A mutation which delays progression from infection threads to differentiated bacteroids would result in their over-representation in the nodule bacteria library. Experimental investigations into the impact of chemotaxis mutants on nodule development and on the abundances of nodule bacteria and bacteroids in mature nodules are described in Sections 5.8 and 5.9 respectively.

Overall, this analysis suggests that under our experimental conditions, motility and chemotaxis are not essential for colonisation of the root surface, but may have a particularly important role in the nodule infection pathway. While emphasising a very important role for motility and chemotaxis towards root hairs or, possibly, progression down infection threads, it does not preclude a more general role in the environment. The INSeq library was inoculated from a frozen culture at stationary phase (likely non-motile) onto sterile vermiculite-grown plants. However, bacterial growth in soil takes place in competition with a complex, motile microbiome. Therefore, the soil environment is likely to impose different motility and chemotaxis selection upon the bacteria, with loss of chemotaxis and motility likely to be catastrophic for long-term survival in soil and root colonization (Turnbull *et al.* 2001; Zheng *et al.* 2015). Experimental investigations into the impact of chemotaxis and motility mutants on attachment and colonisation are described in Sections 5.4 and 5.5 respectively.

5.3 Generation of marked strains to monitor rhizobia across different symbiotic stages

Chemotaxis and motility mutants from Chapter 3 were marked with different bioreporter genes to enable various experimental techniques targeted at different stages of the symbiosis cycle (Figure 1.1). Luminescence was used to count cells on plant roots and to identify sites of high promoter activity, testing attraction and communication stages (Sections 5.4 and 5.6). Fluorescence was used to count cells and differentiate by strain and cell-type using flow cytometry, testing colonisation, infection thread and nodulation stages (Sections 5.5 and 5.9). Finally, enzyme activity staining was used in competition experiments to distinguish strains in nodules and individually to assay nodule maturation (Sections 5.7 and 5.8).

Constitutive luminescence was achieved through the pIJ11282 plasmid, which contains the *lux* gene cluster under a constitutive pNeo promoter (Frederix *et al.* 2014). This plasmid had been previously conjugated into wildtype Rlv3841 (named D5250). In this work, we additionally conjugated the plasmid into the *che1* mutant LM100 (named OPS2335), the *che2* mutant LM400 (OPS2337) and the *motA* mutant OPS2107 (OPS2338). To assess the importance of the motility and chemotaxis systems at the first stages in symbiosis, we also used the pLMB712 plasmid, which contains the *lux* cluster under the promoter of the Rlv3841 *nodABC* operon (pRL100185-7). This promoter is activated by flavonoids produced by pea plant roots at infection zones (Pini *et al.* 2017). Again, this plasmid had been previously conjugated into wildtype Rlv3841 (named strain LMB612) and we conjugated it into the *che1* mutant LM100 (named OPS2139) and the *che2* mutant LM400 (OPS2141).

Strains were fluorescently marked with Tn7 that integrates the transposon after the *glmS* gene (Choi and Schweizer 2006). Each strain was marked with sYFP (Kreiners *et al.* 2006), mCherry (Shaner *et al.* 2004) and sfGFP (P’edelacq *et al.* 2006) all under a constitutive J23104 promoter (through plasmids pOPS0860, pOPS0856 and pOPS0858 respectively). The following strains were generated in this work:

Strain name	Parent strain	Gene deletion	Fluorescent protein
OPS1869	Rlv3841	-	sYFP
OPS1897	Rlv3841	-	mCherry
OPS1953	Rlv3841	-	sfGFP
OPS1848	LM100	<i>che1</i>	sYFP
OPS1846	LM100	<i>che1</i>	mCherry
OPS1847	LM100	<i>che1</i>	sfGFP
OPS1860	LM400	<i>che2</i>	sYFP
OPS1858	LM400	<i>che2</i>	mCherry
OPS1859	LM400	<i>che2</i>	sfGFP
OPS1854	LM300	<i>che1,2</i>	sYFP
OPS1855	LM300	<i>che1,2</i>	mCherry
OPS2442	OPS2107	<i>motA</i>	sYFP
OPS2443	OPS2107	<i>motA</i>	mCherry
OPS2445	OPS2163	<i>fliG</i>	sYFP
OPS2446	OPS2163	<i>fliG</i>	mCherry
OPS2449	OPS2234	<i>flaA</i>	sYFP
OPS2450	OPS2234	<i>flaA</i>	mCherry

To carry out enzymatic activity staining, two marker genes were used: *gusA* (encoding a β -glucuronidase that turns magenta with Magenta-GlcA) and *celB* (thermostable β -galactosidase that turns blue with X-Gal) (See Section 2.8.6) (Sánchez-Cañizares and Palacios 2013). Genes were integrated with Tn7 as above, under a constitutive J23104 promoter. We isolated integrants of *gusA* using plasmid pOPS0857 into Rlv3841 wildtype (named OPS1896), the *che1* mutant LM100 (OPS1871), the *che2* mutant LM400 (OPS1873) and the *motA* mutant OPS2106 (OPS2267). Additionally, a *celB* integration mutant using the plasmid pOPS1376 was integrated on the Rlv3841 wildtype (OPS2085) for comparison.

5.4 Chemotaxis mutants but not motility mutants are impaired in initial attachment

The first stage of symbiosis tested was the early attachment phase. For this experiment, pea roots were incubated in a liquid suspension of constitutively luminescent bacteria (Section 5.11.2). After an hour of incubation, the roots were imaged to count the total luminescence due to bacteria attached to roots.

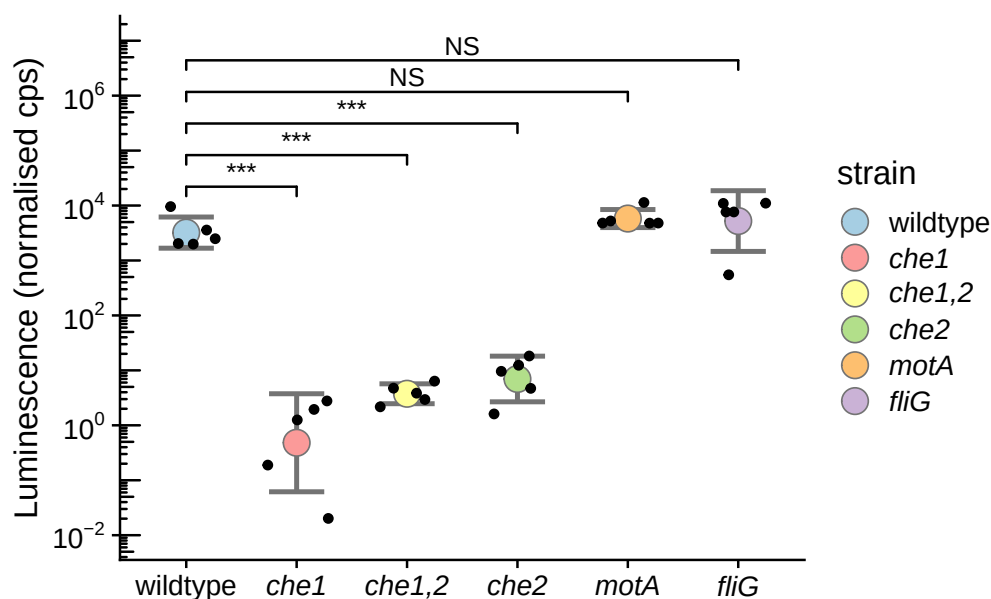


Figure 5.2: Attachment phenotypes for chemotaxis and motility mutants. Tested pNeo-lux marked strains, wildtype (D5250), *che2* (OPS2337), *che1* (OPS2335) and *motA* (OPS2338). Data magnitudes were modelled with Gaussian linear model with parameter strain, found to be significant, $P < 0.001$, ANOVA. Dunnett's Post-hoc test found significant differences between strain means with NS $P > 0.05$, *** $P < 0.001$. Bars are mean $\pm$ SD, with samples as points, $N=5$.

Afterwards, bacteria from each inoculant were grown on swimming plates to confirm their motility phenotypes. A strong effect in this experiment was expected to cascade to following stages. Although, since this stage is the only one experimentally tested with bacteria suspended in liquid, there may be some differences in the magnitude of the effect, compared to later experiments.

Analysis of the luminescence of the attached cell fraction revealed significant reductions by 3-4 orders of magnitude for the motile *che1*, *che1,2* and *che2* mutants compared to wildtype, whereas there were no significant differences found for the non-motile *motA* and *fliG* mutants (Figure 5.2). This indicates that the Che, but not motility, systems are essential for attachment of rhizobia to plant roots in liquid. The possible reasons for this apparent contradiction are discussed below (Section 5.10). To check the effects of these mutants on competitive, long-term attachment on roots, we next investigated root colonisation.

5.5 Motility and chemotaxis are important for root colonisation

To efficiently determine the competitive effect of chemotaxis and motility on root colonisation, we used flow cytometry (Section 5.11.7). We first tested the number of particles detected per mL of $OD_{600} = 1$ liquid cultures for both the wildtype and *che1* mutant. The results are shown in Figure 5.3A. On average, wildtype marked with sYFP and mCherry, and *che1* marked with mCherry appear to be about 5×10^8 . However, *che1* marked with sYFP was found to be significantly higher at about 1×10^9 . Due to these differences, we decided to perform the experiments with both combinations, to average over any individual fluorescence differences.

We then conducted competition experiments using the wildtype marked with sYFP against wildtype marked with mCherry. Pre-germinated pea seeds were grown in boiling tubes with sterile vermiculite and the plants were subsequently inoculated with an equal mixture of both strains and harvested at different times after inoculation (1, 3, 5 and 7 days) (Section 5.11.3). Watering around the roots was standardised to allow a fair comparison of the contributions of motility to colonization. The fluorescent bacteria present on the root at harvest time were then collected and counted using flow cytometry (Figure 5.3B). The results did not show any effect due to the different fluorescent markers used in the study, indicating that the strains behave equally. However, there was a clear increase in bacterial population up to 5 days post inoculation. Thus, this day was chosen for the comparison experiments as the point at which the population reaches maximal capacity but just before the population decline.

With the experimental conditions defined, we then competed the wildtype against itself and the *che1*, *che12*, *che2*, *motA*, *fliG* and *flaA* mutants as sYFP vs mCherry-tagged strains and vice versa (Figure 5.4). The wildtype data are mirrored for better comparison to the mirrored competitions against the test strains.

There were significant competitive disadvantages for all mutant strains relative to the wildtype. Notably, the *che1* mutant had the greatest reduction (by 76%), followed by *motA* and *fliG* (by 50% and 55% respectively). Although it is unclear if

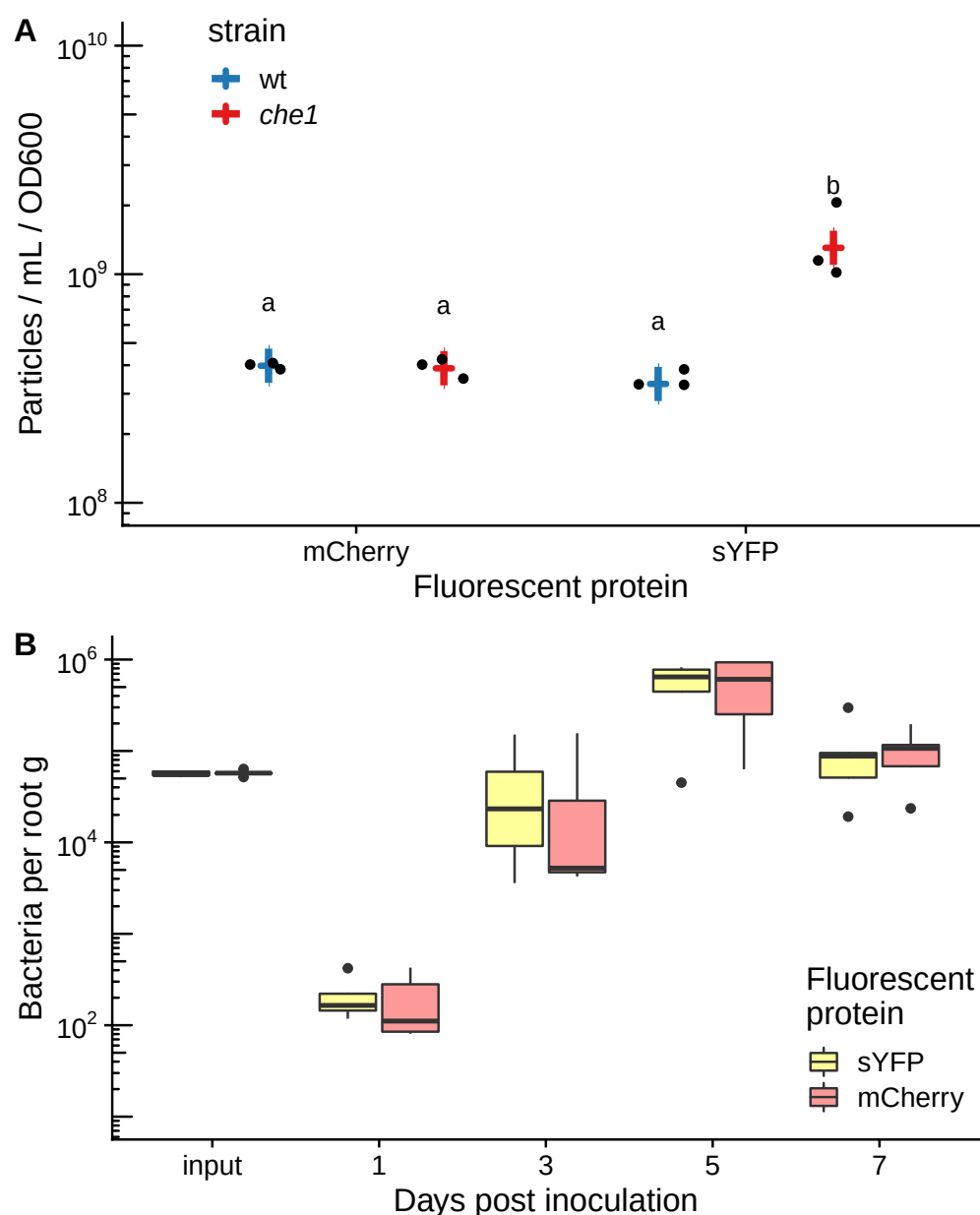


Figure 5.3: Cell counts from pea root colonisation experiments. **A** Flow cytometer fluorescent particle counting per OD₆₀₀ of the liquid cultures used for inoculation. sYFP and mCherry marked wildtype (OPS1869 and OPS1897, respectively) compared to *che1* (OPS1848 and OPS1846, respectively). Data was modelled with a Poisson linear model with the factors strain, fluorescence and their interaction with random intercepts per replicate and offset by volume. Only the interaction factor was found to be significant, $P < 1e-5$, ANOVA. Tukey's Post-hoc test was used to stratify data into groups, where those groups not sharing a letter were significantly different at $\alpha = 0.05$. Points are estimated mean $\pm$ 95% CI, $N = 3$. **B** Wildtype colonisation assays done with pea plants in sterile vermiculite harvested at different times. The wildtype strains used were marked with sYFP vs mCherry (OPS1869 and OPS1897, respectively). Boxes represent quartiles with a centre line for median, $N = 5$.

the differences between the mutants are significant, it is unlikely to be directly due to motility differences since then the *motA* strain would be most disadvantaged, combining the high energy cost of flagella synthesis without the advantages of swimming. This is especially important in this experiment as it is testing bacterial movement towards the plant roots. To test this directly, future experiments could compete *che1* against non-motile strains. Either way, it is interesting that both the non-chemotactic and non-motile results seemingly contradict the INSeq root colonisation results where, as seen previously, *che1* was shown to be a neutral mutation. The INSeq experimental conditions may have masked the colonisation and attachment effects because the INSeq library inoculated onto plants contained bacteria in a non-motile state. Alternatively, the above INSeq may not have been sensitive enough to detect these reductions, given that a later barcoded INSeq experiment with higher sensitivity did find that chemotaxis and motility are important for colonisation.

As for the other strains, both the *che12* and *flaA* mutants showed a high variance in the growth of the plants (Figure 5.4). Even so, they both had significant reductions of 17% and 31% respectively. These seems to be driven by a subset of the samples, in which more extreme reductions were found. These samples may be a better representation of the underlying relationship, and indeed for *flaA* they had larger total cell counts. The final strain, the *che2* mutant, had a small reduction of 27%, although the starting ratios were not as balanced as for the other strains. This imbalance may have resulted in the reduction for this case.

Overall, both the motility and chemotaxis systems have shown to be important for competitive colonisation of pea roots. Then, once the bacteria have arrived at and colonised the root, they still have to reach the root hair tips to initiate symbiosis. In the next section, we describe testing the initiation stage.

5.6 Chemotaxis is involved in symbiosis initiation

The effect of chemotaxis on symbiosis initiation was tested with a time course experiment for nodulation dynamics conducted with the wildtype and the *che1* and

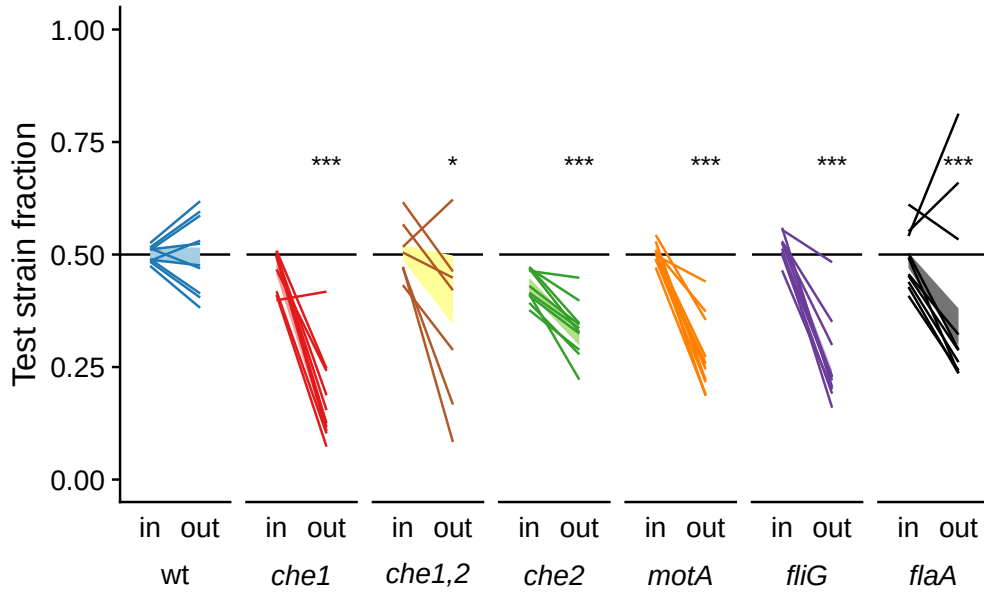


Figure 5.4: Colonisation assays in competition against the wildtype. Both combinations of sYFP and mCherry marked strains were used for the wildtype (OPS1869 and OPS1897) against the following chemotaxis and motility mutants: *che1* (OPS1848 and OPS1846), *che12* (OPS1854 and OPS1855), *che2* (OPS1860 and OPS1858), *motA* (OPS2442 and OPS2443), *fliG* (OPS2445 and OPS2446) and *flaA* (OPS2449 and OPS2450). The y-axis represents the fraction of each of the mutants (test strains) that were counted in each sample. Wildtype data are mirrored for better comparison. Data modelled with binomial linear model with the following factors: input (in)/output (out) and strain and their interaction with random intercepts per plant. All factors were found to be significant, $P < 1e-10$, ANOVA. Dunnett's Post-hoc test found mean differences between input/output with NS $P > 0.05$, * $P < 0.05$, *** $P < 0.001$. Lines represent the mean ratio for each sample at inoculation and after harvesting, and ribbons represent 95% CI of the overall mean, $N=10$.

che2 mutants harbouring the pLMB712 plasmid (pNodA-lux promoter fusion) with the plant *Vicia villosa* (hairy vetch). The plants were grown as described in Sections 2.8.2 and 2.8.4. During this experiment, nodule initiation events (visualised as spikes of luminescence) were recorded together with the eventual physical appearance of nodules on the roots. These spikes of luminescence were caused by the activation of the Rlv3841 *nodABC* promoter by the flavonoids produced by the plant at the onset of the symbiosis to communicate with their symbiotic partner. The timing of the appearance of each nodule, together with the corresponding spike of luminescence were analysed both separately and paired together.

Figure 5.5A displays the distributions of the initiation events that were found

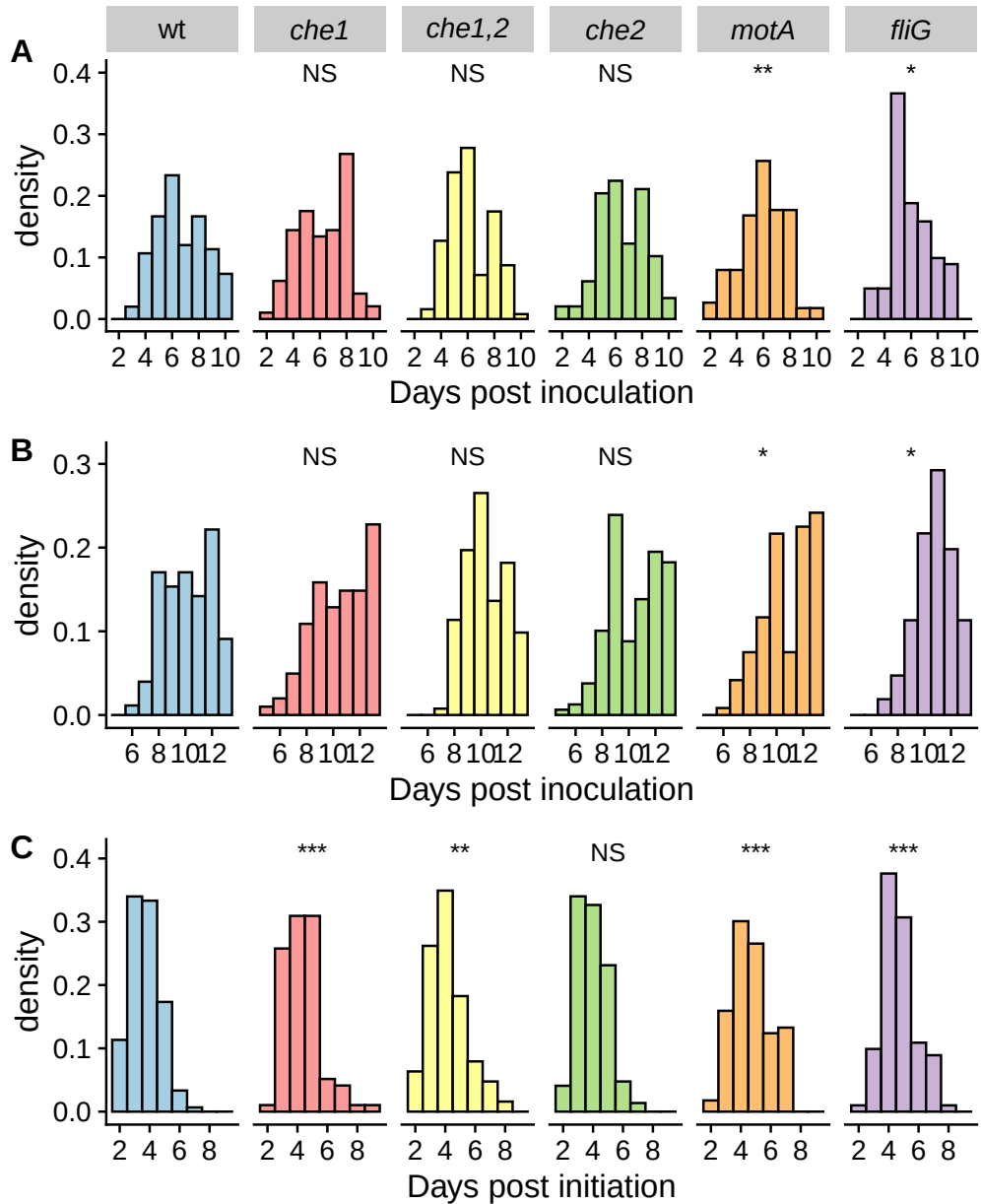


Figure 5.5: Nodulation dynamics assay. Tested wildtype (LMB612), *che1* (OPS2139), *che1,2* (OPS2140), *che2* (OPS2141), *motA* (OPS2528) and *fliG* (OPS2526) harbouring a pNodA-lux promoter fusion. **A** Lux point start times for those spikes of luminescence that resulted in nodule formation. **B** Nodule initial appearance times. **C** Times between the luminescence spike start and the corresponding nodules appearance. Distributions of each type modelled with a Gaussian distribution with strain parameter, significant in all types, $P < 0.05$, ANOVA. Dunnett's Post-hoc test found significant differences in wildtype-strain comparisons with NS $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Histograms represent distributions of data from 5-8 independent growth plates each containing 5 plants, per strain.

to later transition into nodules. Significant reductions were found for the *motA* and *fliG* mutants compared to wildtype, being on average half a day earlier, whereas there were no significant differences found for the *che1*, *che1,2* and *che2* mutants. These differences may be due to earlier detection of the flavonoid signal by the *motA* and *fliG* mutants, or more likely, are simply due to plant variation. The nodule distributions showed the reverse pattern, with significant increases found for the *motA* and *fliG* mutants, again by half a day, with no significant differences for the other mutants 5.5B).

To check the early development timings of the strains, the time between initiation and nodule appearance for each nodule was calculated. The nodule development distributions (Figure 5.5C) of the *che1*, *che1,2*, *motA* and *fliG* mutants were significantly increased over the wildtype, on average half to a full day later. In contrast, the *che2* mutant was not significantly different to wildtype. This indicates that removing motility or the main chemotaxis system delays nodule initiation. Since the bacteria are already colonised on the roots when the initial signal is detected, it is likely that the delay occurs due to movement required to the root hair tips or down the infection thread.

These results show that there is still some unresolved questions on the effect of chemotaxis on nodulation. We thus progressed to measuring nodulation competition directly against the wildtype to determine the ultimate effects of the chemotaxis mutations.

5.7 Motility and chemotaxis are important for nodulation

The effects of chemotaxis and motility on nodulation competition was measured by competing wildtype and the *che1*, *che2* and *motA* mutants marked with *gusA* against *celB* marked wildtype (Section 5.11.4). The pre-germinated peas were planted in sterile silver-sand-vermiculite pots and inoculated with this mixture, then harvested after three weeks and stained to enable nodule counting. The strain ratio at the nodule stage was then compared to the strain ratio in the inoculant.

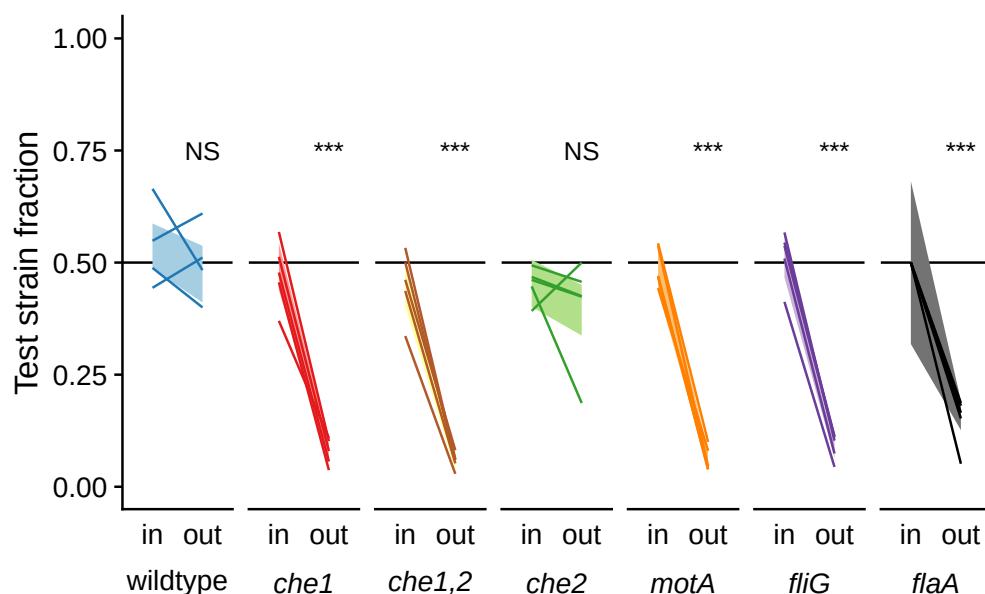


Figure 5.6: Competition assays on nodulation for chemotaxis and motility mutants. *gusA* marked strains of wildtype (OPS1896), *che1* (OPS1871), *che1,2* (OPS1872), *che2* (OPS1873), *motA* (OPS2267), *fliG* (OPS2447) and *flaA* (OPS2451) mutants were tested against *celB* marked wildtype (OPS2085). Data modelled with binomial linear model with the factors input (in)/output (out), strain and their interaction with random intercepts per plant. Input/output and the interaction factors were found to be significant, $P < 1e-10$, ANOVA. Dunnett's Post-hoc test found mean differences between input/output with NS $P > 0.05$, *** $P < 0.001$. Lines represent the mean ratio for each sample at inoculation and after harvesting, and ribbons represent 95% CI of the overall mean, $N=5$.

All of the *che1*, *che1,2*, *motA*, *fliG* and *flaA* mutants showed a significant reduction in competition compared to wildtype, by 85%, 89%, 88%, 84% and 69% respectively (Figure 5.6). Loss of motility and control of motility is clearly important across the whole symbiosis cycle. The colonisation data makes clear that movement along the roots, at least, is an important stage of symbiosis, but there may be another later stage requiring chemotaxis and motility: movement to root hair tips. In order to initiate symbiosis, bacteria have to be at the tip of root hairs (Oldroyd *et al.* 2011). Although this experiment competes strains across both stages, they could theoretically be separated by masking the initial colonisation impact, perhaps by inoculating the strains in a non-motile state. This could be used to confirm the INSeq results by showing a minimal effect at initial colonisation, but a large effect on nodulation.

The *che2* cluster mutant did not have a significant reduction in nodule counts (Figure 5.6). Although *che2* displayed reduced early attachment and colonisation efficiency (Sections 5.4 and 5.5), this had no significant impact on overall nodulation. Instead, it may be that the *che2* cluster has a role in nodule maturation, with mutants having slower nodule development. This could then cause the increased count of nodule bacteria found by the INSeq screen (Figure 5.1). To investigate this, an experiment that surveyed the sizes of nodules formed by wildtype and chemotaxis mutants was performed in the following section.

5.8 Chemotaxis is important for nodule development

The effect of the chemotaxis clusters on nodule development was determined by measuring the distribution of nodule areas over time (Section 5.11.5). The plants were inoculated individually with wildtype and chemotaxis mutants *che1* and *che2*, each marked with constitutively expressed *celB* to enable staining for nodule visualisation, and were grown in sterile vermiculite pots and harvested after 2, 3 and 4 weeks. After weighing, staining and photographing the roots, the images were analysed using a pipeline to produce the area of each nodule for each plant.

The plants displayed no significant differences between the inoculated strains in weeks 2 and 3, but in week 4, roots of plants harbouring the *che2* mutant were significantly smaller (29% reduction) than wildtype (Figure 5.7A). Indeed, one of the *che2* assigned plants even had yellow leaves, the same shade as the water controls, despite displaying nodules. Earlier, in week 2, both the *che1* and *che2* mutants produced significantly less nodules, by about 50%, than the wildtype, although this was recovered in weeks 3 and 4 (Figure 5.7B). In week 3, the *che2* mutant nodules were significantly smaller (19% reduction in average area), although they had recovered by week 4 (with a non-significant reduction by 7%). In contrast, the *che1* mutant nodules were consistent with the wildtype until week 4, where they were significantly smaller (15% reduction in average area).

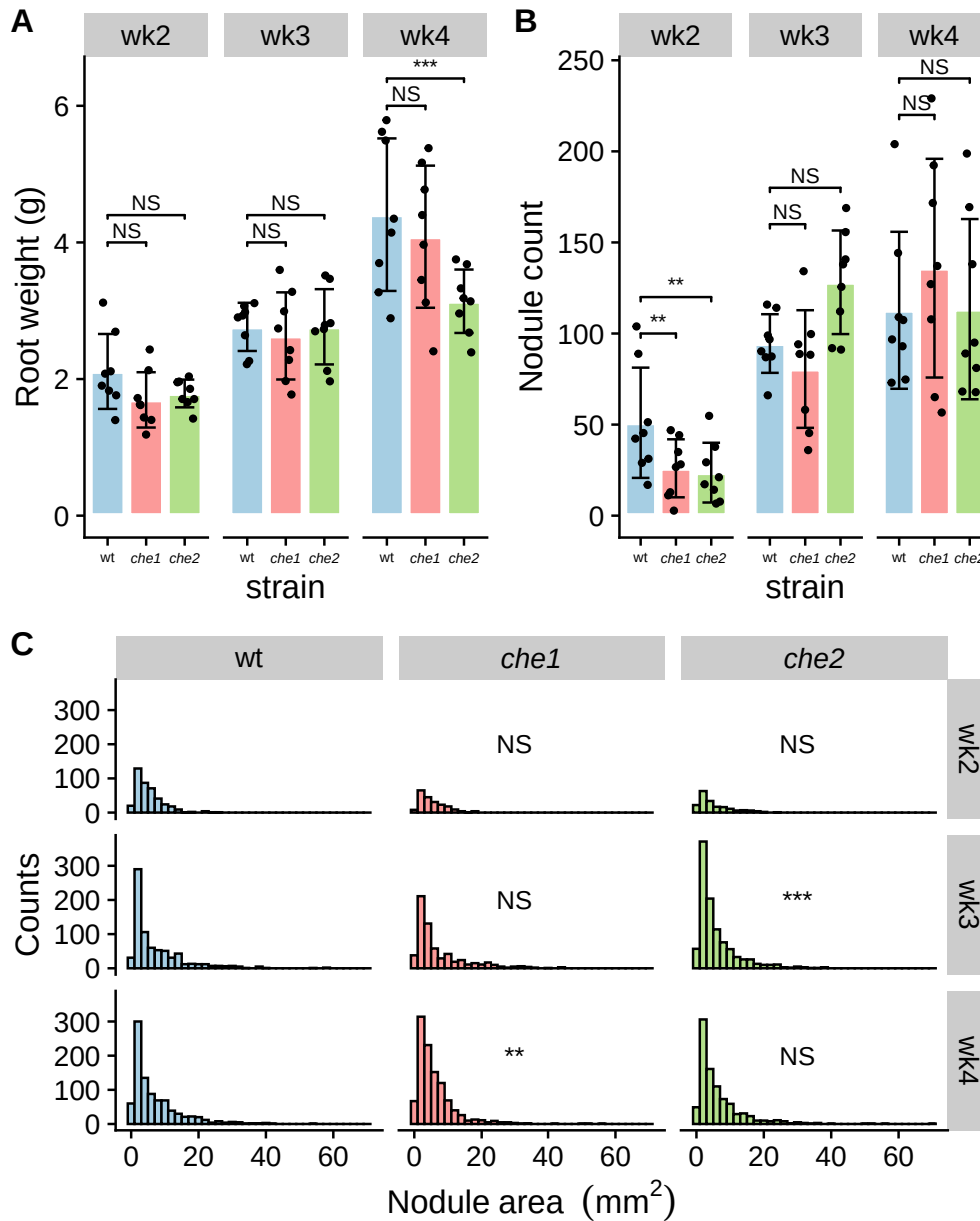


Figure 5.7: Nodulation maturity. *celB* marked strains of wildtype (OPS1896), *che1* (OPS1871) and *che2* (OPS1873) were inoculated on *Pisum sativum* and harvested after 2, 3 and 4 weeks. **A** Root weights modelled with Gaussian linear model, as above. Dunnett's Post-hoc test found mean differences. **B** Nodule counts modelled with Poisson linear model with the factors strain and timepoint, all significant, $P < 0.05$, ANOVA. Dunnett's Post-hoc test found mean differences. **C** Nodule areas modelled with Gamma linear model, as above. Dunnett's Post-hoc test found mean differences. NS $P > 0.05$, ** $P < 0.01$, *** $P < 0.001$. Bars are mean $\pm$ SD, histograms represent distributions of data, $N=8$.

For the *che1* mutant, the initial drop of nodules is consistent with the nodulation results (Section 5.7), although the mutant does recover nodule count as a single strain. Interestingly, the nodule area distribution only lags behind the wildtype after 4 weeks. The *che2* mutant also had an initial drop of nodules compared to the wildtype, although this recovered in later weeks. The nodule area distribution lagging behind the wildtype by the 3rd week combined with having no competitive disadvantage noted with nodulation, indicates that *che2* has slower nodule development than the wildtype (Section 5.7). This is also supported by the defect in plant health by the 4th week, likely due to reduced nitrogen fixation by less mature nodules.

If the *che2* cluster is indeed important for nodule maturation, it is expected that the *che2* cluster mutant would have some impact on bacteroid differentiation. To investigate this, an experiment that measured the amounts of bacteria and bacteroids in mature nodules formed by each of the strains was performed in the following section.

5.9 The *che2* cluster is important for nodule maturation

The effect of the chemotaxis clusters on the development of mature nodules was determined by counting their populations of bacteria and bacteroids (Section 5.11.6). The plants were inoculated individually in sterile vermiculite pots and harvested after 3 weeks. The strains used were the wildtype and chemotaxis mutants *che1* and *che2*, each marked with constitutively expressed sfGFP to enable counting in a flow cytometer. Bacteria were differentiated from bacteroids on the basis of size and granularity measured as forward and side scatter, respectively.

There was no significant difference between strains for counts of undifferentiated bacteria per gram of nodule (Figure 5.8A). There was, however, a significant reduction in the *che2* mutant bacteroid counts over wildtype. This also resulted in a higher proportion of bacteria relative to bacteroids in *che2* mutant nodules (Figure 5.8B). This suggests that the *che2* cluster has a role in nodule maturation, such that mutation causes a reduction in differentiated bacteroids by about 30%.

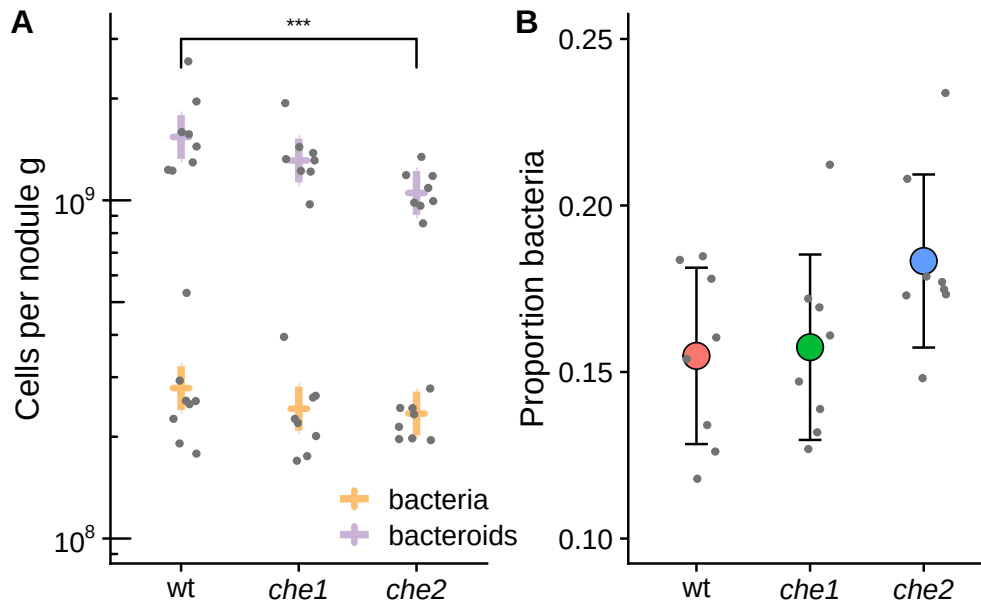


Figure 5.8: Undifferentiated bacteria and bacteroid counts per nodule. The strains tested were sfGFP marked wildtype (OPS1953), *che1* (OPS1847) and *che2* (OPS1859). **A** Cells per nodule gram obtained from 10 nodules. Data modelled with Poisson linear model with the factors strain and cell type (bacteria or bacteroids) with random intercepts per plant and offset by volume and nodule mass. The single terms were all found to be significant, $P < 1e-10$, ANOVA. Dunnett's Post-hoc test found mean differences between strains with NS $P > 0.05$, *** $P < 0.001$. Lines are model prediction $\pm$ 95% CI with samples as points, $N=8$. **B** Proportion of bacteria in the above nodules. Circles are mean $\pm$ SD with samples as points, $N=8$.

Indeed, this effect was likely even minimised by the selection of only mature nodules for the experiment. A clearer understanding of the maturation of *che2* nodules can only come from the distribution of bacteria and bacteroid counts across nodule development. Mature nodules were initially chosen to prevent bias between strains, indeed, each group of 10 nodules weighed approximately the same (0.05g). This may, however, have masked the main effect of the *che2* cluster, by removing any impact of a change in the rate of nodule maturation and therefore in the proportion of mature nodules. Future experiments may investigate this in more detail by harvesting all nodules and forming a more complete dataset or by harvesting nodules at different time points.

5.10 Discussion

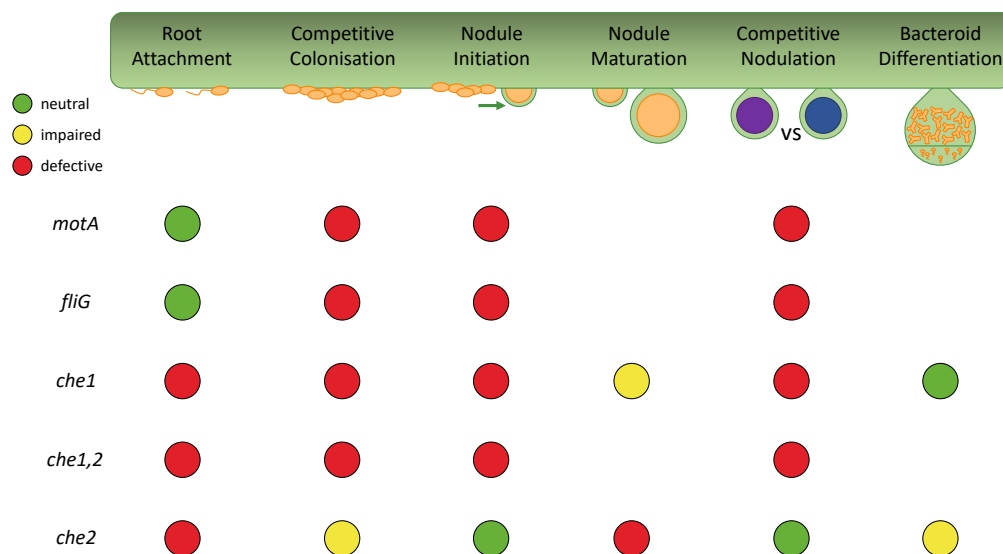


Figure 5.9: Overview of experiments.

5.10.1 The *che1* cluster, in coordination with flagella, appears to have a role in attraction to both the root surface and to root hair tips

We first described the effects of transposon insertions in flagellar and chemotactic genes based on an INSeq experiment. The majority of flagellar and *che1* cluster genes were found to be neutral in the rhizosphere and in root colonisation, but essential for nodule bacteria and bacteroids (Figure 5.1). This is likely due to the library being grown on TY agar without selecting for motility. During the early stages of colonisation, all strains were likely to be non-motile. So the early effects of chemotaxis and motility would be masked, allowing later effects to be revealed. This means that, under the experimental conditions of the INSeq experiment, chemotaxis-based flagella movement is important for some stage after root colonisation.

To investigate this hypothesis, we first tested the attachment of mutant strains to pea plant roots. We found a significant reduction in attachment for the motile *che1*, *che1,2* and *che2* mutants but not for the non-motile *fliG* and *motA* mutants (Figure 5.2). This suggests that motility is not important under these conditions but that disruption to motility via the chemotaxis mutants is highly detrimental. These results are clearly paradoxical and led us to check the phenotypes in 4 independent

experiments. Furthermore, to ensure that the *fliG* and *motA* mutants were behaving as expected, the cultures used for attachment assays were also stabbed into swimming plates and in all cases the strains were confirmed to be non-motile as expected. Work in Chapter 4 suggests that the *che1* mutant is hyper-motile, having a reduced tumble rate and thus increased smooth swimming and that the *che2* mutant has an increased tumble rate. Since these modifications to tumble rate are artificial—due to disruption of the system rather than in response to a signal—the response to signal gradients were impaired (see Chapter 3). The chemotaxis mutants may be moving too fast or too erratically to effectively transition to a sessile state upon reaching the boundary layer at the root surface and thus display reduced attachment. However, an alternative explanation is that the Che proteins control a cellular factor required for attachment that is independent of motility. Furthermore, while the *che1* and *che2* strains do have the expected swimming phenotypes, they have large deletions with antibiotic cassettes inserted in the chromosome. These strains were obtained from the Hynes laboratory and in future work it would be sensible to make new in-frame deletions in both clusters as well as in individual *che* genes.

Interestingly, Secchi *et al.* (2020) recently found that although motility is important for attachment to surfaces from a slowly flowing fluid, increasing the fluid flow reduces the effect of motility with the strains becoming indistinguishable at speeds over 1000 $\mu\text{m/s}$. In addition, they report that the motility advantage was due to directed movement towards the surface and that cells with disrupted tumbling, both increased and decreased, had altered attachment. This supports the idea that uncontrolled swimming is worse for attachment than non-motility. It might be possible to test this by treating cells with a proton ionophore to dissipate the transmembrane pH gradient and thus externally stopping motility (Manson *et al.* 1977). If the wildtype and mutants attach similarly after this treatment, then the hyper-motility of the chemotaxis mutants is likely the cause of the reduced attachment. It does of course raise a separate and interesting question of whether an energized membrane is needed by cells for attachment. An alternative way to reduce motility would be to increase the viscosity of the medium by adding a reagent such

as Ficoll to increase the strength of the root surface boundary effect, potentially slowing the hyper-motile bacteria sufficiently to allow effective attachment.

Interestingly, previous experiments that inoculated roots with bacteria taken directly from agar slope cultures (data not shown) produced an order of magnitude less attachment than the data described above, in which inoculation was from overnight liquid cultures. It was initially assumed that the differences were due to most of the cells grown on agar slopes forming biofilms and thus becoming largely non-motile. However, the genetically non-motile mutants, *motA* and *fliG*, grown in liquid culture displayed the same attachment as wildtype. This is especially interesting because it is expected that *fliG* is non-flagellated, whereas *motA* is expected to produce non-rotating flagella. This suggests another, alternative explanation of the results, that the discrepancy is due to exopolysaccharide (EPS) regulation; EPS is required for rhizobia to form strong attachments to roots and is necessary for biofilm formation (Janczarek 2011; Wheatley and Poole 2018). Indeed, EPS and motility are commonly antagonistically regulated by levels of 3',5'-cyclic di GMP (c-di-GMP): with low levels of c-di-GMP corresponding to motility phenotypes and high levels to sessility. For instance, *Sinorhizobium meliloti* Rm2011 has c-di-GMP receptors that negatively regulate motility and promote EPS production (Schaper *et al.* 2016, 2017). In addition, *Rhizobium leguminosarum* biovar *viciae* UPM791 displayed increased biofilm formation and root attachment upon overexpression of a gene encoding a c-di-GMP-producing diguanylate cyclase (*pleD*) (Perez-Mendoza *et al.* 2014).

Some bacteria have also been found to have regulatory links back from flagella to c-di-GMP levels. One of the flagellar stators of *Pseudomonas aeruginosa* interacts with the diguanylate cyclase SadC, stimulating activity and repressing motility (Baker *et al.* 2019). The flagella of *Caulobacter crescentus* acts as a sensor, stimulating the diguanylate cyclase DgcB upon surface-induced motor interference (Hug *et al.* 2017). In addition, disruption of its flagellar assembly also stimulates c-di-GMP levels via PleD for early-stage disruption and through MotAB binding of DgcB for late-stage disruption (pre-print Hershey, Fiebig and Crosson (2020)).

Perhaps the *fliG* and *motA* deletions induce a similar kind of system, increasing c-di-GMP levels which induce EPS production and increase attachment. This would have to be linked to both the non-rotation of flagella (in *motA*) and the absence of flagella (presumably in *fliG*).

To find the effects of these mutants on competitive, long-term attachment on pea roots, we next investigated root colonisation. Mutants on *che1*, *motA* and *fliG* all displayed reduced colonisation in competition against wildtype, although to differing degrees (76%, 50% and 55% reductions respectively) (Figure 5.4). So whatever the cause of the attachment discrepancies, they are masked during colonisation, perhaps due to the longer timescale or competition with wildtype.

However, these results seem to contradict the INSeq results and suggest that the *che1* cluster and flagellar motility are indeed important for competitive root colonisation. This discrepancy may simply be due to differing experimental conditions. Although the colonisation experiment used the same inoculation density (10^5 CFU per plant), it was carried out with a one-to-one strain competition compare to the complex mixture of mutants used in the INSeq experiment. Most importantly the cultures were carefully grown to be at maximum motility before incubation of roots in a suspension of bacteria.

Competition experiments were also performed at the level of nodulation. This experiment again showed a significant reduction in nodules formed by *che1*, *che1,2*, *motA* and *fliG* mutants, with a higher reduction than for colonisation (85%, 89%, 88% and 84% reductions respectively) (Figure 5.6). The further reduction in competitiveness may be due to another role of the *che1* cluster and flagellar-based motility in the intervening stages between colonisation and formation of mature nodules. It is critical that for the nodulation results, INSeq and the work described here are in agreement.

To probe the intervening stages, our next experiment was designed to allow investigation into the initiation and development of nodules. The first stage was visualised using a plasmid containing the genes for light emitting luminescence under a *nodABC* operon promoter. This operon is activated by plant-produced flavonoids,

and so activation of the operon is linked to the earliest stages of infection where plants boost the synthesis of flavonoids in response to bacterial lipo-chitooligosaccharide (LCO) signals (Pini *et al.* 2017). The time when a nodule first appeared was compared to the time at which the luminescence signal was first observed in that spot to obtain initiation timings. The nodule development distributions of the *che1*, *che1,2*, *motA* and *fliG* mutants were shifted significantly longer than the wildtype (by 0.7, 0.5, 1 and 1 day respectively; Figure 5.5C). In addition, we found that the *che1* mutant inoculated individually formed fewer nodules than the wildtype at the 2nd week and that the nodules formed were 15% smaller on average at the 4th week (Figure 5.7B and C).

This indicates that removing motility or the main chemotaxis system delays nodule initiation. The extent of the luminescence signal along large portions of the roots indicates that it comes from the colonised root near root hairs. As such, chemotaxis and motility are likely important for movement from the root surface to the root hair tips, and thus causing the delay in nodule formation after activation of the lux signal at the root surface. Another possibility is that chemotaxis is important for movement down the root hairs inside the infection threads, and thus causing a delay after entering the root hairs. To distinguish between the two, experiments were attempted to count the root hairs produced by the *che1* mutant under competition. Unfortunately, finding infection threads or colonised root hair tips proved difficult, to the point that we were unable to consistently score their occurrence per root. A potential reason for this is that the number of infection threads on a root at a particular time is likely restricted. Instead, future experiments could take a different approach by comparing the exudates of root hairs versus the root proper. These could examine their chemoattractive effects and metabolic profiles to determine if it is plausible that there is chemotactic movement to root hair tips.

As for movement down infection threads, this was indirectly tested in the nodule development experiment. These experiments found no difference in bacteria or bacteroid counts in mature nodules between the *che1* mutant and wildtype (Figure 5.8A). Since infection threads are continuously formed as the nodule grows

and matures, it is expected that any disruption of infection thread movement would affect the bacteria and bacteroid counts (Vasse *et al.* 1990; Monahan-Giovanelli, Pinedo and Gage 2006). It is thus unlikely that the *che1* cluster is important for this stage of symbiosis.

Either way, the role appears to be after biofilm formation on the root. The colonisation effect was apparently masked in the INSeq experimental conditions, probably as a result of the whole library being non-motile. However, this enabled the role of chemotaxis independent of colonisation to be seen in nodulation competitiveness.

5.10.2 The *che2* cluster has a role in nodule maturation

We also investigated the *che2* cluster using the same experiments mentioned above. However, the results of the *che2* mutant were distinct from the *che1* mutant, resulting in different conclusions. Firstly, in the INSeq experiment, insertions in the *che2* cluster genes were found to be neutral in the rhizosphere, on plant root and for bacteroids but advantaged for nodule bacteria (Figure 5.1). Whatever role the *che2* cluster performs, mutating it produces nodules that overall have an increased number of bacteria.

A broad view of symbiosis was examined in light of the *che2* mutation, specifically across attachment, colonisation and nodulation competition. These experiments found disparate results, with a 3 order of magnitude difference on attachment, a 27% reduction in competitive colonisation and no difference in competitive nodulation between the *che2* mutant and wildtype (Figures 5.2, 5.4 and 5.6). It appears that the disrupted motility of the *che2* mutant may prevent an effective transition to a sessile state at the root surface, as discussed above. Perhaps then this effect also causes the reduction in colonisation, although ameliorated over the longer timescale. Alternatively, the *che2* deletion is partially complemented *in trans* by the presence of the wildtype in the competitive colonisation and nodulation experiments.

Any role of the *che2* cluster in early attachment is unlikely to cause the observed INSeq phenotype that deletion of *che2* genes produced an increase in recovered nodule bacteria. Further experiments thus examined later stages of symbiosis,

beginning with the transition from root colonisation to early nodule development. This experiment involved scoring the timing of initialisation (through *nodABC* lux) and formation of nodules (when visible). The *che2* mutant was not significantly different to the wildtype in either initiation timing or duration (Figure 5.5). The next experiment observed nodule development by measuring the distribution of nodule areas at different timepoints. Interestingly, the *che2* mutant displayed significant differences to the wildtype at each timepoint, forming 50% less nodules in week 2; having a smaller distribution of nodules in week 3 (19% reduction in average area); and, in week 4, the roots of plants harbouring *che2* mutant were 29% smaller (Figure 5.7). These changes combined indicate that the *che2* mutant is defective for nodule maturation, forming smaller nodules at a slower rate and thus impacting the plant health. To discover why, the next experiment analysed the composition of mature nodules.

To compare the composition of wildtype nodules to those of the *che2* mutant, bacteria and bacteroids were extracted from mature nodules and counted. We found that although there was no significant difference between counts of bacteria in wildtype and *che2* mutant nodules, there was a 30% reduction of bacteroids (Figure 5.8A). This result seems to contradict the INSeq result of an increase of nodule bacteria mutated in the *che2* cluster. The difference in this experiment, is that it only examined mature nodules. All the nodules on each plant's roots were harvested for the INSeq experiment. Since the *che2* mutant forms nodules that mature at a slower rate than the wildtype, mutants with insertions inside the cluster are expected to be at an earlier stage in development when there would be more bacteria that have yet to differentiate.

The role of the Che2 cluster in nodule maturation may be guiding movement down infection threads. Firstly, although sliding motility via growth is the dominant way to move down the infection threads, there is some evidence for movement independent from growth (Gage and Margolin 2000; Gage 2002; Fournier *et al.* 2008). Interestingly, due to the oxygen gradient within nodules, the HemAT sensor described in Section 3.7.3, would favour swimming in the opposite direction. If this

is the case, perhaps the action of the Che2 system is to override this and guide the sensory system down the infection threads. Indeed, the methylation protein CheB2 of the *che2* cluster has been shown to alter the methylation of chemoreceptors linked to the *che1* cluster (Miller *et al.* 2007). Future experiments could examine the effect of the double *che1* and *che2* mutant on nodule composition to determine if the link between the two systems is important.

However, it is more likely that the role of the Che2 cluster in nodule development is even independent of flagella. The whole flagellar machinery is down-regulated in infection threads and subsequently in nodules (Karunakaran *et al.* 2009). Future experiments could investigate this by testing the nodule composition of the *motA* and *fliG* mutants, with paralysed and absent flagella, respectively. If so, then the Che2 cluster must act by some other mechanism, perhaps driving motility through pili.

5.11 Methods

5.11.1 INSeq assay

As described by Wheatley *et al.* (2020), the INSeq experiment was performed as follows. Rlv3841 mariner insertion mutants (input1) were inoculated onto 7-day-old pea seedlings and subsequently recovered at 5 days post inoculation (dpi) to construct two libraries for INSeq analysis; 1) the rhizosphere and 2) the roots. Rlv3841 mariner insertion mutants (input2) were inoculated onto 7-day-old pea seedlings and subsequently recovered at 28 days post inoculation (dpi) from the nodules. A total of 150,000 nodules were picked from approximately 1,500 plants and a further two libraries for INSeq analysis were constructed; 3) undifferentiated nodule bacteria, generated by growing bacteria from crushed nodules (which will represent nodule bacteria that are not terminally differentiated along with those recently released from infection threads) for 12 h in rich medium before extracting DNA, and 4) bacteroids, generated from DNA extraction from the nodules. A total of six INSeq libraries (input 1, input 2, the rhizosphere, roots, nodule bacteria and bacteroids) were sequenced and analysed by a Hidden Markov Model (HMM) (DeJesus and Ioerger 2013). Each of the 7,357 Rlv3841 genes which contain TA

insertion sites (99.9% of the genome), was classified as essential (ES), defective (DE), neutral (NE) or advantaged (AD). Genes classified as NE in the respective input library (6,429 for input1 and 6,656 for input2) were considered for each test condition, with genes classified as either ES or DE (ES/DE) regarded as required at that stage of symbiosis.

5.11.2 Attachment assay

This assay was used to determine the ability of a strain to attach to plant roots. Plasmid pIJ11282, containing lux under a constitutive pNeo promoter, was conjugated into bacteria. Marked bacteria were then grown on TY slopes for 3 days before being resuspended into a TY liquid culture overnight without antibiotics. Peas were germinated as per Section 2.8.1 and planted in boiling tubes as per Section 2.8.3. The bacteria were resuspended in 4 mL of 15 mM MES/HEPES buffer (pH 7) and diluted to an optical density of $OD_{600} = 0.1$ in a volume of 100 mL. These mixtures were measured in triplicate on a PlateReader to determine luminescence and optical density for normalisation. Next the peas were removed from the tubes, dipped in sterile milliq H₂O and taped to the side of boiling tubes. The peas were then incubated in the bacterial suspensions for 1 hour with shaking at 150 rpm. The plants were then swirled gently in sterile milliq H₂O and blot dry on tissue paper. The roots were then removed and placed on square petri dishes for imaging using a NightOWL II LB 983 system (Berthold). The NightOWL system was run with the following conditions: luminescence, exposure time: 20s, gain: high, readout: slow, binning factors: 2x2 binning, cosmic suppression: enabled, subtract background image: disabled, background correction: enabled, Flatfield correction: disabled, image accumulation: disabled, emission filter wheel A, filter photo and wavelength 560nm.

5.11.3 Colonisation assay

This assay was used to determine the relative colonisation competitive fitness of a strain. The bacteria were marked with both mCherry and sYFP fluorescent

proteins. Peas were germinated as per Section 2.8.1 and planted in boiling tubes as per Section 2.8.3. The next day, marked bacteria were grown on TY slopes for 3 days. After a further 4 days of plant and bacterial growth, the plants were inoculated with a 10^5 1:1 mix of the bacterial strains in 1 mL of 0.9% NaCl. Both combinations of mCherry and sYFP marking of the strains were tested. The 10^6 serial dilution of the inoculant was measured on a flow cytometer to determine the initial ratio of strains (Section 5.11.7). The plants were then harvested at either 1, 3, 5 or 7 days. The roots were separated, weighed and mixed with 25 mL of 0.9% NaCl supplemented with 0.02% Silwet-L77. The roots were then vortexed at maximum speed (2500 rpm) for 1 minute each. The liquid was then passed through a 40 μ m filter to remove excess plant debris. This filtrate was then measured on a flow cytometer to determine the ratio of strains on the root surface (Section 5.11.7).

5.11.4 Nodule competition assay

This assay was used to determine overall competitive fitness of a strain against the wild type. The bacteria were marked with the genes *gusA* or *celB* for β -glucuronidase or thermostable β -galactosidase activity, respectively (Sánchez-Cañizares and Palacios 2013). Tagged bacteria were then grown on TY slopes for 3 days. Peas were germinated as per Section 2.8.1, and two peas per 500 mL pot were planted within a mix of 50/50 fine vermiculite and silver sand as per Section 2.8.3. A 10^5 1:1 mix of the bacterial strains for comparison was added to each pot. The bacterial mixes were also plated in triplicate on TY plates supplemented with X-glcA (5-bromo-4-chloro-3-indolyl- β -D-glucuronic acid CHA salt) to confirm the mixture ratios. Negative controls were also made by adding water instead of the bacterial mixtures. The plants were harvested after 21 days and stained as per Section 2.8.6. The nodules were then counted by colour to determine the output ratio of strains.

5.11.5 Nodule development assay

The bacteria and plants were grown as per Section 5.11.4, with single inoculum of *celB* marked strains. The plants were harvested after 14, 21 and 28 days and

stained as per Section 2.8.6. The roots were then photographed and analysed using an image analysis pipeline (github.com/AroneyS/Nodule-segmentation).

5.11.6 Nodule harvest and bacteroid isolation

This assay was used to determine the proportion of bacteria and bacteroids in the nodules formed by various strains. The bacteria were first marked with sfGFP fluorescent protein. The peas were germinated as per Section 2.8.1, and two peas per 500 mL pot were planted within fine vermiculite as per Section 2.8.3. A 1 mL 10^5 inoculum of bacteria was added to each pot. The plants were harvested after 21 days and 10 mature nodules per plant were removed. The nodules were weighed, then crushed into 1 mL of bacteroid isolation buffer. This buffer contained 1M K_2HPO_4 , KH_2PO_4 , 300 mM sucrose and 2 mM $MgCl_2$, and was filter sterilised with 0.22 μm filters. The mixture was then centrifuged at 1000 rpm for 10 minutes to pellet plant debris. The liquid was reserved, being centrifuged again at 6000 rpm for 5 minutes, with the pellet resuspended in 1 mL 0.9% NaCl. The wash was repeated once. The liquid was then diluted 1:10 in flow-cytometer grade PBS for flow cytometer analysis (Section 5.11.7).

5.11.7 Flow cytometer measurement

Samples were measured in a flow cytometer as follows. 100 μL of each sample were placed in triplicate in 96-well plates, alternating rows with washes of 0.9% NaCl. The flow cytometer was set to measure the forward and side scatter of each particle alongside its fluorescence. Once the number of reasonable particles for a sample reached 5000, the machine was washed with 30 μL of 0.9% NaCl. sYFP, mCherry and sfGFP labelled cells were distinguished by setting a threshold of fluorescence (4000 for each, Figure 5.10A-B). Further gates were set to differentiate bacteria and bacteroids based on forward and size scatter. The bacteroids were larger (forward scatter) and more granular (side scatter) than the bacteria (Figure 5.10C-D). These raw counts were then analysed.

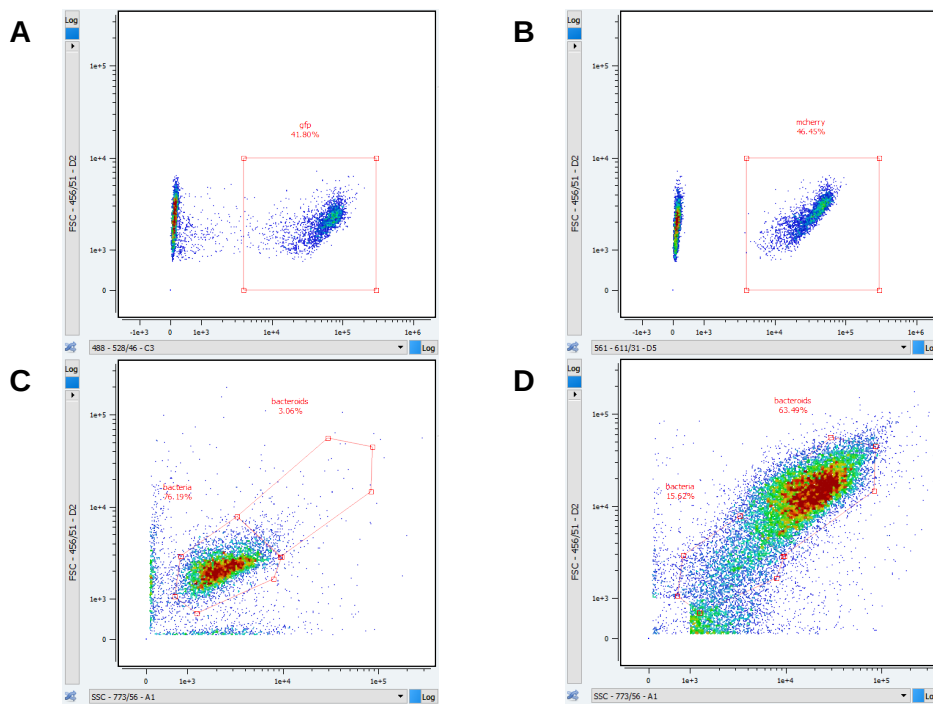


Figure 5.10: Flow cytometer examples. **A** Example sfGFP and sYFP gate using forward and 488-528/46 laser scatter for a bacteria sample. **B** Example mCherry gate using forward and 561-611/31 laser scatter for a bacteria sample. **C** Example bacteria and bacteroid gates using forward and side scatter for an inoculant sample containing bacteria. **D** Example bacteria and bacteroid gates for an extracted nodule sample containing both bacteria and bacteroids.

Conclusions and future perspectives

6.1 Impact of the 2020 Pandemic

The thesis work was partially conducted during extensive periods of COVID-19 pandemic lockdown from March 2020-21. This meant that several experiments were unable to be completed, including flagella presence scoring. In addition, the bioinformatic analysis was extended during home-work.

6.2 General overview

Fertilisers are used to provide crop plants with essential nutrients such as nitrogen. However, fertiliser generation is energetically expensive and must be provided in excess, with most fertiliser nitrogen leaching into surrounding water-systems (Cheng 2008; Smil 2011; Oldroyd and Dixon 2014). The symbiosis of legumes with rhizobia exhibits none of these problems since the bacteria provide fixed nitrogen directly to the plant. However, while efforts have been made to transfer this symbiosis to non-legume crops, the system remains unique to legumes (Geddes *et al.* 2015). Further understanding of the initiation of symbiosis may enable improvements to the farming of crop legumes and enable the same processes to be mimicked in engineered systems of other crops. Although motility and chemotaxis are known to be crucial for competitive plant root colonisation (the first stage of symbiosis), little

is known about their importance after this stage (Rodriguez-Navarro, Dardanelli and Ruiz-Sainz 2007; Scharf, Hynes and Alexandre 2016; Wheatley and Poole 2018). In this thesis, I aimed to describe the importance of chemotaxis and motility in sequential stages of symbiosis up to the generation of root nodules filled with nitrogen-fixing rhizobia. To this end, the chemotaxis and motility systems of the model rhizobia *Rhizobium leguminosarum* biovar *viciae* 3841 (Rlv3841) were characterised, in both free-living and plant-associated states.

Similar to standard models such as *E. coli*, Rlv3841 motility is driven by flagella rotation modulated by a chemotaxis system that responds to signals sensed by chemoreceptors. In this work we found that Rlv3841 has a diverse set of chemoreceptors that enable the bacteria to sense and respond to a variety of chemical gradients. It was found that the chemoreceptors operate both as external sensors and internal sensors of the cell's metabolic state (Chapter 3). We also found a novel system that appears to halt the bacteria in response to high concentrations of TCA cycle intermediates (Chapter 4). Critical components of this system were identified, but the link to the flagella is currently unknown. Finally, we characterised the importance of the chemotaxis and flagella systems of Rlv3841 at sequential stages of the symbiosis (Chapter 5). We found that there were several critical stages in which chemotaxis and motility are essential and that these occur after the plant root has been colonised by the rhizobia.

6.3 The multiple sensory systems of Rlv3841

The chemotaxis and motility systems of Rlv3841 were first characterised, including its 25 chemoreceptors (Chapter 3). Various mutant strains were examined by measuring the swimming halos produced in UMS semi-solid agar supplemented with various carbon and nitrogen sources. As expected, mutants of critical flagella genes (*motA* and *fliG*) were found to be non-motile, producing solid halos of limited size. Another mutant of the gene encoding the main flagellar filament protein (*flaA*) was expected to be unable to form stable flagellar filaments (Scharf *et al.* 2001; Tambalo *et al.* 2010a). Although the *flaA* mutant displayed greatly reduced

motility, it was still motile—possibly because it was partially rescued by other copies of the gene (*flaB*, *flaC*, *flaD*, *flaE* or *flaH*).

Wildtype Rlv3841 was tested on a variety of carbon sources and was found to display increase swimming in response to reduced levels of a number of metabolites including glucose, arabinose, pyruvate, acetate, succinate, malate and fumarate. In addition to the flagella mutants referred to above, mutants in both chemotaxis operons (*che1* and *che2*) were obtained. The *che1* mutant no longer displayed swimming responses to varying levels of glucose, pyruvate or arabinose, whereas the *che2* mutant displayed the same responses as wildtype. As expected, the Che1 cluster appears to be the dominant chemotaxis system for free-living control of swimming (Miller *et al.* 2007). Interestingly, the acetate and organic acid effects were not removed with chemotaxis operon deletion.

High acetate concentration produced faint halos that were smaller than the non-motile mutants. This was assumed to be a toxicity effect. However, the halos formed on high concentrations of organic acids were larger and cloudy, and appeared to be formed by halting at high concentrations followed by swimming after consumption brought the concentration down. The investigation undertaken to find the mechanism behind the TCA cycle intermediate halting effect is described below (Section 6.4).

The various chemoreceptors encoded by Rlv3841 were also characterised. These receptors were found to have a wide variety of sensory domains and sensor types, including periplasmic sensors predicted to sense amino acids, peptides, carbohydrates, carboxylates and organic acids, and cytoplasmic sensors predicted to sense oxygen, redox levels or light. Characterisation of the domains of the chemoreceptors revealed two distinct groups: those encoding proteins with longer signalling domains that were associated with the *che2* cluster (*mcrA*, *mcrB* and *mcrC*) and those with smaller signalling domains which included all the other chemoreceptors. It may be that the association of most chemoreceptors with the Che1 cluster, with only those three associated with Che2, is what drives the dominance of the Che1

cluster. Additionally, some of the receptors were found not to have transmembrane domains, indicating that they are internal sensors.

Two such cytoplasmic sensors were IcpA and HemAT, both having a Protoglobin sensor domain predicted to sense oxygen via a haem group, although only HemAT has the conserved residues important for haem binding in *Bacillus subtilis* (Hou *et al.* 2001; Zhang, Olson and Phillips 2005). Analysis of mutants of both sensors found that the receptor encoded by *icpA* senses the metabolic state of the cell, with mutants being unable to stimulate motility to the extent of the wildtype under nearly all carbon sources tested but independent of oxygen tension. In contrast, the HemAT receptor appears to sense oxygen concentrations, presumably by binding haem. Mutants of *hemAT* had reduced motility in conditions with high oxygen consumption or low oxygen tension.

6.4 TCA cycle intermediate halting effect

In Chapters 3 and 4, the discovery and investigation of a TCA cycle intermediate halting effect is described. High concentrations of succinate, malate or fumarate produced cloudy, spotted halos on swimming plates and even higher concentrations of succinate produced non-motile halos. It is expected that this morphology is caused by a halting effect from high concentrations of TCA cycle intermediates. The bacteria are initially halted but begin swimming once sufficient organic acid has been consumed; the bacteria then reach an area with high substrate concentration and halt again, forming another spot. As this process repeats, many spots are formed around the edges of the halo to produce the cloudy spots effect. Tracking individual bacteria in varying concentrations of succinate showed a reduced motile fraction in high succinate confirming this explanation.

The halting effect was next tested across a range of soil bacteria including representatives from *Rhizobium*, *Sinorhizobium*, *Azorhizobium*, *Azospirillum*, *Pseudomonas* and *Kosakonia*. However, responses to a high concentration of succinate were only observed in the *Rhizobium leguminosarum* strains.

The TCA cycle intermediate halting effect was found to be independent of both chemotaxis clusters and internal chemoreceptors, still being observed in *che* mutants. In addition, both the carbon and nitrogen branches of the phosphotransferase system (PTS) were found to affect motility but are not responsible for the halting effect. Deletions of the gene encoding the key carbon branch protein, *manX*, and the gene encoding the key nitrogen branch protein, *pstN1*, were found to reduce motility. These reductions were recovered by the non-phosphorylatable mutant of ManX (H9A) and by the phosphomimic mutant of PtsN1 (H66D) respectively but not by the reverse mimics. Both proteins tend to be phosphorylated under nitrogen limitation and carbon excess, and not phosphorylated under nitrogen excess and carbon limitation (S'anchez-Cañizares *et al.* 2020). It is possible that under nitrogen limitation, PtsN-P stimulates motility, whereas in carbon limitation, ManX stimulates motility.

In order to identify systems that were connected to the halting effect, a screen was performed by growing wildtype in the presence of high TCA cycle intermediates and isolating mutants from the papillations that formed. From this we isolated and then sequenced 9 independent strains that no longer responded to high concentrations of succinate, malate or fumarate. These strains were discovered to have mutations in either an ABC transporter for inorganic phosphate uptake (*pstSCAB*, RL0542-5) or a putative SAM-dependent methyltransferase (RL3464). Both types of mutants were found to abolish the cloudy spots morphology in high organic acid media and to have an increased motile fraction at high succinate over the wildtype. These effects were restored to wildtype by complementation with the native operons. In addition, constitutive expression of the gene encoding the putative methyltransferase in the wildtype strain caused a non-motile phenotype.

The protein encoded by RL3464 has a Pfam classification as a putative (S)-Adenosyl-L-methionine (SAM)-dependent methyltransferase. Bioinformatic analysis grouped methyltransferases from various *Rhizobium* strains together, each with >92% amino acid identity with the methyltransferase from Rlv3841. Another clade was formed of methyltransferases from various *Sinorhizobium* strains, each

with about 70% identity with the methyltransferase from Rlv3841. The lack of a halting effect in *Sinorhizobium* strains could be due to these differences or to absence of downstream effectors

The other primary mutations were observed in the *pstSCAB* genes (RL0542-5). These genes encode PstSCAB, a high affinity inorganic phosphate transporter that appears coredundant with PhoCDET, with either function sufficient for effective symbiosis and fixation (Yuan, Zaheer and Finan 2006; Hu *et al.* 2018). PstSCAB is additionally associated with a PhoRB two-component system (encoded by RL0540 and RL0547) via a linker protein PhoU (encoded by RL0546). PhoU negatively regulates both PhoRB and PstSCAB when under excess phosphate conditions and dissociates under limiting phosphate conditions (Makino *et al.* 1989; Hsieh and Wanner 2010; diCenzo *et al.* 2017). A frame-shift mutation in *pstC* was observed in *S. meliloti* Rm1021, causing constitutive activation of the PhoB response regulator (Krol and Becker 2004; Yuan, Zaheer and Finan 2006). This mutation was matched in strains OPS2290 and OPS2300, both having frame-shift mutations in *pstC*. Although this suggests that constitutive activation of PhoB, likely caused by these mutations, could be responsible for the mutant phenotype, the halting effect still occurs under phosphate limiting conditions. These mutations are thus likely to have this effect through the independent regulatory role of some unidentified part of the system, possibly the PhoU protein.

Another possible pathway is through regulation by a separate system that senses high levels of TCA-cycle intermediates. For instance, ActSR is activated by a reduced electron transport chain (ETC) which would be produced by high concentrations of succinate driving the TCA cycle and pushing high amounts of reductant through the ETC (Tiwari *et al.* 1996; Priefer *et al.* 2001; Elsen *et al.* 2004; Fenner *et al.* 2004). ActSR could thus be the sensory trigger for the TCA-cycle intermediate halting effect if it transcriptionally regulates *pstSCAB* or RL3464, which then presumably interact with motility in some way. Alternatively, PstSCAB and the SAM-dependent methyltransferase could be responding to the proton motive force, which would also be increased with high TCA cycle activity.

Since PstSCAB transports via ATPase activity and the increased proton motive force would drive increased ATP synthase activity, the activity of the PstSCAB transporter could also be increased under high organic acids. This may then cause a downstream response leading to non-motility.

6.5 Chemotaxis and motility in symbiosis

An insertion sequencing (INSeq) experiment was conducted for rhizosphere, root colonised and nodule bacteria, and differentiated nodule bacteroids to cover various stages of the *Rhizobium*-legume symbiosis. In general, the experiment found that genes required for flagella assembly and genes in the *che1* cluster were essential for both nodule bacteria and bacteroids but not for rhizosphere growth and root colonisation. Strains with mutations in genes in the *che2* cluster were advantaged in nodule bacteria but neutral in all other categories.

To confirm these results and examine the stages more closely, the chemotaxis and motility mutants described above were tested against wildtype at sequential stages of the symbiosis (Chapter 5; Figure 6.1). Firstly, attachment to roots was assayed by incubating pea roots in a liquid suspension of constitutively luminescent bacteria for 1 hour and imaging to count the total luminescence from attached bacteria. Next, root colonisation was assayed by inoculating pea plants grown in vermiculite with a competitive mixture of fluorescent bacteria and, after 5 days, counting colonised bacteria through flow cytometry. Then, symbiosis initiation was assayed by inoculating vetch plants grown on FP media plates with flavonoid-responsive luminescent bacteria and, over a period of 13 days, measuring the time between nodule initiation (visualised as spikes of luminescence) and the eventual physical appearance of nodules on the root. Next, nodule maturation was assayed by inoculating pea plants grown in vermiculite with enzymatically marked bacteria and, after 2, 3 and 4 weeks, measuring the area of each nodule through an image analysis pipeline. Nodulation competition was then assayed by inoculating pea plants grown in vermiculite with a competitive mixture of enzymatically marked bacteria and, after 3 weeks, counting stained nodules of each strain. Finally, bacteroid

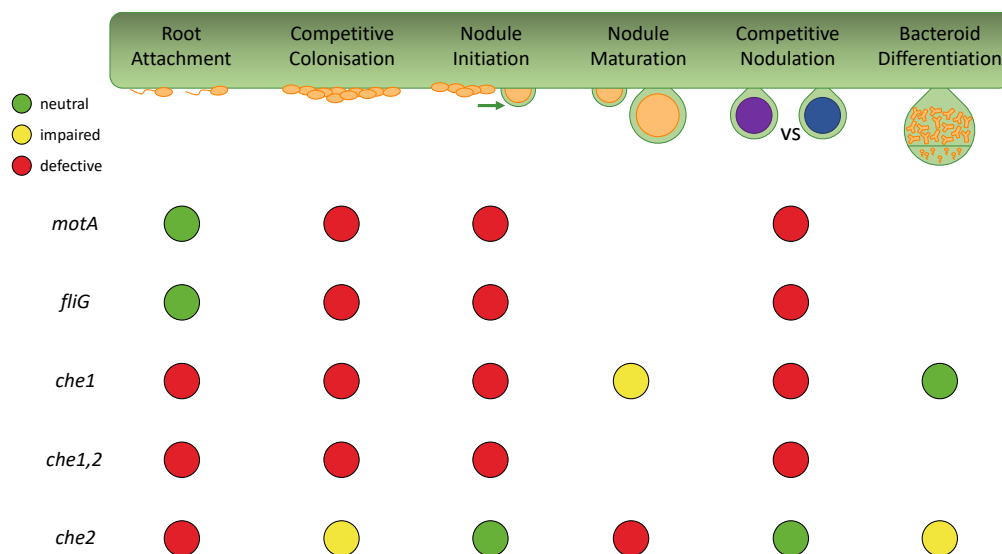


Figure 6.1: Overview of chemotaxis and motility in symbiosis.

differentiation was assayed by inoculating pea plants grown in vermiculite with fluorescently marked bacteria and, after 3 weeks, counting bacteria and bacteroids in mature nodules through flow cytometry.

6.5.1 The *Che1* cluster is important in early stages of symbiosis

Firstly, we found that the chemotaxis mutants in *che1*, *che1,2* and *che2* were defective for early attachment but the flagella mutants of *fliG* and *motA* were not significantly different to wildtype. Since the results appear paradoxical—motility being unimportant but disruption to motility via chemotaxis being highly detrimental—we independently verified the results in multiple ways. One possible explanation is that the chemotaxis mutants were moving too fast or too erratically to effectively transition to a sessile state upon reaching the boundary layer at the root surface and thus display reduced attachment. Indeed, Secchi *et al.* (2020) found that motile and non-motile strains were indistinguishable for attachment when under high fluid flow and that disruptions to tumbling, such as via chemotaxis mutants, altered attachment. Alternatively, the flagella mutants may have altered exopolysaccharide (EPS) regulation, noting that EPS is required for strong attachment and is commonly

antagonistically regulated with flagella (Janczarek 2011; Schaper *et al.* 2016, 2017; Wheatley and Poole 2018).

Results for later stages were more straightforward, with *che1*, *che1,2*, *fliG* and *motA* mutants all having defects in competitive colonisation and nodulation, and in nodule initiation. These results suggest that Che1 and flagella have an early role in initial colonisation and a later role between colonisation and nodule initiation. Swarming motility along the roots is expected to be important for competitive root colonisation, however, a later role for motility has not yet been described. There are two later stages that may require motility: movement from the root surface to the root hair tips and movement down infection threads. Infection threads are continuously formed as the nodule grows and matures, so it is expected that any disruption of infection thread movement would affect the bacteria and bacteroid counts (Vasse *et al.* 1990; Monahan-Giovanelli, Pinedo and Gage 2006). However, since the *che1* mutant had no significant difference in nodule composition, it is unlikely that Che1 is important for this stage. Thus it is likely that chemotaxis and motility are important for efficient movement to the root hair tips in order to competitively initiate symbiosis.

6.5.2 The Che2 cluster is important for nodule development

In addition to the defects on early attachment, the *che2* mutant was also found to be defective in later stages. Despite having wildtype timings for initiation of nodules, the *che2* mutant displayed delays in nodule maturation. After 2 weeks, the *che2* mutant had significantly less nodules than wildtype; after 3 weeks, the nodules were significantly smaller; and after 4 weeks, the roots of plants harbouring the mutant were significantly smaller. Together, these results indicate that the Che2 system has a role in nodule development such that the mutant forms smaller nodules at a slower rate to the extent that it impacts plant health.

We also investigated the composition of nodules since the INSeq screen recovered an increased number of bacteria with *che2* insertions from nodules. We found that although there was no significant change in bacterial counts, there were 30%

less bacteroids counted in mature nodules formed by the *che2* mutant than by the wildtype. These results again suggest a defect in nodule maturation for *che2* mutants and thus a role for the Che2 cluster in nodule development. If this role of the Che2 cluster is linked to flagella-based movement, it may be to guide motility down infection threads at nodule initiation and during nodule maturation. Although sliding motility via growth is the dominant way to move down the infection threads, there is some evidence for movement independent from growth (Gage and Margolin 2000; Gage 2002; Fournier *et al.* 2008). However, the entire flagella machinery is down-regulated in infection threads and subsequently in nodules (Karunakaran *et al.* 2009). Since there is no genetic evidence of a secondary flagellar system, this motility may instead be driven by pili. In that case, the Che2 cluster may instead be linked to pili and control twitching motility down infection threads.

6.6 Future directions

This work has provided a theoretical overview of all the chemoreceptors encoded by Rlv3841 by bioinformatic analysis. Future work could physically characterise these chemoreceptors, especially focussing on those with unknown sensor domains. Although deletions and multi-deletions would be powerful ways to discover the signals connected to each receptor, the abundance of genes and likely redundancy between them would make this difficult. Instead, the genes could be individually transferred to a chemoreceptor-gutted model bacteria to test for positive reactions to a library of signal compounds. Once candidate compounds have been discovered, protein tests by differential scanning fluorimetry and isothermal titration calorimetry would allow the binding kinetics to be characterised (Niesen, Berglund and Vedadi 2007; Webb *et al.* 2014).

An obvious extension to the investigation into the TCA-cycle intermediate halting effect is to determine the TCA-cycle-intermediate sensor and whether it is directly or indirectly sensing (See Table 6.1). The first candidate is the PstSCAB system which could be modulating PhoU or PhoB activity in response to organic acids, potentially indirectly via effects on the proton motive force and ATP. However,

since low phosphate conditions (expected to activate PhoB) did not trigger a halting response, it is unlikely that sensing is due to ATP-powered phosphate transport. This also challenges a role for PstSCAB in TCA-cycle-intermediate-halting-effect signal transduction which must then act via some phosphate-transport-independent activity of the system. However, further testing of the activity of PhoU and PhoB in relation to organic acid concentrations could reveal a signal transduction role. On the other hand, the methyltransferase encoded by RL3464 is unlikely to be the sensor since it has no annotated sensor domain and consists nearly entirely of the Pfam Methyltransf_28 domain. Alternatively, the sensor could be the electron-transport-chain-associated ActSR two-component system, which could sense effects of organic acids on the ETC. Although none of the papillation mutants had mutations in *actSR*, the system may be too critical to harbour mutations, especially in competition against wildtype in the semi-solid plate environment. Future work could attempt to generate *actSR* deletion mutants and then assess their swimming response to TCA-cycle intermediates and their expression of *pstSCAB* and RL3464. These experiments could determine which, if any, of these systems is sensing TCA-cycle intermediate concentrations.

Another unknown is if the TCA-cycle intermediate halting effect acts through post-translational control of flagella or instead by transcriptional down-regulation (See Table 6.1). Although the timescale of the plate experiments favours transcriptional regulation, tracked bacteria responded within 3 hours. To more precisely determine the kinetics of halting, a culture growing in low succinate conditions could be shocked with high succinate then have their swimming tracked immediately afterwards. If the change can occur within a few minutes, then it is highly unlikely to be transcriptional control, which would be expected to occur via dilution from cell division, thus taking several hours. In addition, the relative proportion of flagellated cells under high and low succinate concentrations could be determined through staining and microscopy (Kodaka *et al.* 1982b; Heimbrook, Wang and Campbell 1989b).

If the effect is through post-translational control of flagella, there are two candidates: the putative methyltransferase encoded by RL3464 and the CheY proteins (See Table 6.1). Since acetylation of CheY causes increased motor binding that encourages tumbling, excessive amounts of acetylated CheY could lead to a locked motor. However, the TCA-cycle intermediate halting effect still occurred in *che1* and *che2* and the double *che1,2* mutant. Alternatively, the putative methyltransferase could methylate some component of the flagellar motor or some kind of flagellar brake protein to cause a halt. Indeed, constitutive expression of RL3464 in wildtype Rlv3841 produced a non-motile strain. Mutation of the RL3464 gene could reveal these downstream effects. In addition, RL3464 could be expressed in a protein vector to produce tagged, purified protein for cell lysate pull-downs as a binding screen. It is likely that downstream elements of the putative methyltransferase are causing the halting effect. It has proved particularly effective in the Poole lab to Flag tag proteins and express them in Rlv3841. They are then gently lysed under native conditions, proteins bound to anti-Flag resin in a chromatography column, washed, then eluted with Flag peptide before mass spectrometry analysis. This has revealed interacting partners in native Rlv3841 for FixAB and PtsN proteins and could be a powerful approach to study RL3464.

If, in contrast, the effect is found to be via transcriptional regulation, it is more likely to be controlled via PstSCAB or ActSR (See Table 6.1). Either of these systems could directly affect flagellar synthesis through controlling transcription of VisNR or by increasing c-di-GMP levels to indirectly downregulate motility. However, it is unlikely to be mediated solely by ActSR since this would bypass both the PstSCAB system and the putative methyltransferase and thus mutations in these systems would not be expected to prevent the halting effect. In addition, since low phosphate conditions did not trigger the halting effect, it is unlikely that simple action of PhoB is transcriptionally regulating flagella. Importantly, further investigations are necessary to confirm that the system behaves in Rlv3841 as described in other species; perhaps by mutating individual genes or the entire operon. The current data,

however, supports post-translational regulation by the putative methyltransferase more strongly than transcriptional regulation by either PstSCAB or ActSR.

Finally, there is the question as to the relevance of the system to the free-living and symbiotic lifestyles of Rlv3841. It may be that during free-living conditions the chemotaxis systems are dominant in their control of motility since high levels of TCA-cycle intermediates in the soil are likely rare. In the rhizosphere or at plant root surfaces, there likely would be higher concentrations of TCA-cycle intermediates due to root exudation. Perhaps the TCA-cycle intermediate halting effect is active at the plant to assist in biofilm formation or symbiosis initiation. Alternatively, if concentrations only reach sufficient levels inside the plant, the halting effect may support the transition to the mostly sessile symbiotic life within the nodule. Since Rlv3841 nodulates a wide variety of plants, it would need to respond to widely relevant signals for differentiation, one of which may be the high levels of organic acids that legumes feed their symbionts.

Table 6.1: The possible mechanisms controlling the TCA-cycle intermediate halting effect.

Statement	Evidence
Sensing via PstSCAB	<i>For</i>
	• Mutations in transporter critical
	<i>Against</i>
	• Sensing via ATP levels unlikely due to low phosphate not triggering non-motility
	<i>Future expected results</i>
	• Change in PhoU response under organic acids • Change in PhoB activity under organic acids

Statement	Evidence
Sensing via RL3464 protein product	<i>For</i> • Mutations in methylation domain critical <i>Against</i>
Sensing via ActSR	<i>For</i> • No sensor domain • Mechanism: Responds to organic acid perturbation of electron transport chain <i>Against</i> • No random mutants found <i>Future expected results</i>
Motility control post-translational	<i>For</i> • <i>actSR</i> mutants with altered <i>pstSCAB</i> and RL3464 expression • Mechanism: Methylation of flagellar proteins • <3 hour response time in liquid <i>Future expected results</i> • Flagellated but non-motile cells under organic acids
Motility control transcriptional	<i>For</i> • Fast halting kinetics • Mechanism: Response regulator control or methylation of DNA • 4-7 day timeline in semi-solid media <i>Future expected results</i> • Non-flagellated cells under organic acids • Slow halting kinetics

Statement	Evidence
Control via RL3464 protein product	<i>For</i> • Constitutively expressed $\rightarrow$ non-motile • Mutations in methylation domain critical <i>Future expected results</i> • Flagellation with non-motility under constitutive expression • Complementation of <i>pstSCAB</i> mutants with RL3464 • Pull-down of flagella components by RL3464 protein product
Control via CheY	<i>For</i> • Mechanism: CheY binding to motor (likely acetylated or phosphorylated) <i>Against</i> • <i>che1</i> and <i>che2</i> mutants still display the effect
Control via PstSCAB	<i>For</i> • Mechanism: PhoB response regulator transcriptional control <i>Against</i> • Low phosphate did not trigger non-motility <i>Future expected results</i> • PhoB binding to flagellar genes

Statement	Evidence
Control via ActSR	<i>Against</i> • No known links to motility • <i>pstSCAB</i> and RL3464 mutations unexplained <i>Future expected results</i> • <i>actSR</i> mutants with altered flagellar synthesis

In addition to the halting effect, this work revealed a potential role for Che1 and flagella in the movement from the root surface to root hair tips in order to initiate symbiosis. This could be further investigated by comparing the exudates of root hairs versus the root proper. Differences in metabolic profiles may produce different chemoattraction effects and thus explain the chemotactic movement to root hair tips.

Finally, the role of Che2 could be further investigated, to determine if infection thread movement or bacteroid differentiation are more likely (See Table 6.2). Impacts on bacteroid differentiation were found with *che2* INSeq mutants being advantaged for nodule bacteria and *che2* deletion mutants forming 30% less bacteroids per nodule mass and having delayed nodule maturation. Perhaps Che2 has a role in sensing a plant signal to prepare the bacteria for differentiation or works as a redundant trigger for differentiation. However, there are no known links from chemotaxis to bacteroid differentiation, so more work is needed for a conclusion to be reached. Alternatively, Che2 could have a role in movement down infection threads. In addition to control of active motility being a known role for chemotaxis, since infection threads occur throughout maturation, disruption of this movement could cause the bacteroid effects noted above. Although the *che2* mutant was not found to be delayed in symbiosis initiation on Vetch, the delays in nodule maturation on Pea indicate that the mutation has a systemic impact across the symbiosis. Perhaps the role of *che2* is emphasised in later infection threads through maturation rather than the initiation infection threads where free-living systems may still be active.

Another unknown is if the Che2 system acts via flagella, pili or another, unknown system (See Table 6.2). At first glance flagella appears the most likely since the Che2 system has a known connection to swimming, with *che2* deletion mutants displaying reduced halo diameters and increased tumble rate, and CheB2 being known to methylate Che1-associated chemoreceptors (Miller *et al.* 2007). However, the whole flagellar machinery is also known to be downregulated in infection threads and subsequently in the nodule (Karunakaran *et al.* 2009). Future experiments could determine if the flagella are still present during infection through antibody-targetted staining of flagella in nodule sections. In addition, comparing the nodule compositions and maturation distributions of the *motA* and *fliG* mutants to the *che2* mutant could reveal whether the *che2* mutants effects are through flagella.

On the other hand, pili also appear to be a good candidate for the Che2 effector. Pili could be used to perform twitching motility down the surface of the infection threads. Chemotaxis systems have been found to regulate type IV pili and twitching motility in bacteria such as *Synechocystis*, *Myxococcus xanthus* and *Pseudomonas aeruginosa* (Bhaya, Takahashi and Grossman 2001; Shi and Sun 2002; Alexandre 2015; Oliveira, Foster and Durham 2016). Indeed, although the *tad* pilus of Rlv3841 was found to be downregulated in bacteroids, the genes were upregulated in the *bacA* mutant, which stalls symbiosis at the infection thread stage (Karunakaran *et al.* 2010). Thus it is possible that the *tad* pilus is upregulated in infection threads to enable twitching motility. In contrast, however, the entire pilus operon was found to be neutral across all symbiotic stages in the INSeq experiments. To confirm these results, future experiments could examine the behaviour of pilus mutants compared to the *che2* mutant to determine if they are indeed the Che2 effector. If pili are also not involved, then the search for the unknown downstream effectors could begin by screening for phosphorylation targets of CheY3, the most likely candidate for downstream signalling. Importantly, the INSeq only found various transporters, unknown genes and *mcpT* to have the same pattern as *che2*: advantaged for nodule bacteria and neutral for all other symbiotic stages. This suggests both that Che2 is unlikely to have a role in differentiation via a known regulatory system but also that

perhaps it is sensing via McpT. Interestingly, *S. meliloti* McpT was hypothesised to have a role *in planta* but has an uncharacterised sensor domain (Meier and Scharf 2009; Zatakia *et al.* 2017). Thus the current data supports the role of Che2 being navigation of infection threads during nodule maturation through pili control, perhaps by reacting to a signal sensed by McpT.

Table 6.2: The possible roles and mechanisms of the Che2 system.

Statement	Evidence
Che2 role is infection thread movement	<i>For</i> • Active motility noted in infection threads • Known role for chemotaxis • Bacteria move down infection threads throughout maturation <i>Against</i>
Che2 role is bacteroid differentiation	<i>For</i> • <i>che2</i> mutant unaffected in initiation timing or duration • 30% less bacteroids per nodule mass • <i>che2</i> mutants advantaged for nodule bacteria • Delay in nodule maturation <i>Against</i> • Unknown mode of action

Statement	Evidence
Che2 controls flagella	<i>For</i> • Swimming motility possible through thread • CheB2 methylates Che1-associated chemoreceptors • <i>che2</i> deletion generally increases tumble rate <i>Against</i> • Flagella down-regulated in infection threads and nodules • <i>che2</i> deletion only has small impact on swimming halos <i>Future expected results</i> • Flagellation of infection thread bacteria confirmed by antibody staining of sections • Flagella mutants behave as <i>che2</i> mutant in nodules
Che2 controls pili	<i>For</i> • Twitching motility possible along surface • Pili genes upregulated in <i>bacA</i> mutant on pea <i>Against</i> • INSeq mutants in pili genes were neutral across symbiosis • Pili genes downregulated in bacteroid <i>Future expected results</i> • Pili mutants behave as <i>che2</i> mutant in nodules • CheY3 interacts with pili components

Statement	Evidence
Che2 controls other unknown system	<i>Against</i> • INSeq mutants matching <i>che2</i> pattern only includes transporters and <i>mcpT</i> <i>Future expected results</i> • CheY3 binds other system proteins

In conclusion, this work has shown that chemotaxis and motility are indeed essential for effective symbiosis of rhizobia with legumes, having multiple roles across the entire cycle. This conclusion is also supported by the intricate and complex mechanisms controlling motility in Rlv3841, which display the importance of effective use of motility in the highly competitive soil environment.

A

INSeq results

Table A.1: Rlv3841 chemotaxis and motility gene requirement at different symbiotic stages. INSeq analysis of the phenotypes of motility and chemotaxis function encoding genes across the rhizosphere, root colonisation, nodule bacteria and bacteroid libraries. INSeq phenotypes are as follows: advantaged (AD), neutral (NE), defective (DE), essential (ES) and unclassified due to too few TA sites (N/A).

Name	Gene	Clusters	Input1	Rhizosphere	Root	Input2	Nodule bacteria	Bacteroids
mcpX	RL0426		NE	NE	NE	NE	NE	NE
hemAT	RL0429		NE	NE	NE	NE	NE	NE
mcpT ^c	RL0455		NE	NE	NE	NE	AD	NE
mcpE	RL0564		NE	NE	NE	NE	NE	NE
icpA ^b	RL0685	che1	NE	NE	NE	NE	ES	ES
cheS ^b	RL0686	che1	NE	NE	NE	NE	ES	ES
cheY1 ^b	RL0687	che1	NE	NE	NE	NE	ES	ES
cheA1 ^b	RL0688	che1	NE	NE	NE	NE	ES	ES
cheW1 ^b	RL0689	che1	NE	NE	NE	NE	ES	ES
cheR1 ^b	RL0690	che1	NE	NE	NE	NE	ES	ES
cheB1 ^b	RL0691	che1	NE	NE	NE	NE	ES	ES
cheY2 ^b	RL0692	che1	NE	NE	NE	NE	ES	ES
cheD ^b	RL0693	che1	NE	NE	NE	NE	ES	ES
	RL0694		NE	NE	NE	NE	NE	NE
fliF ^{b*}	RL0695	flg1	NE	NE	NE	NE	DE	ES/DE
visN ^{b*}	RL0696	flg1	AD	NE	NE	NE	ES/DE	ES/DE
visR ^{b*}	RL0697	flg1	NE	NE	NE	NE	ES/DE	ES/DE
	RL0698	flg1	NE	NE	NE	NE	NE	NE
flhB ^b	RL0699	flg1	NE	NE	NE	NE	ES	ES
fliG	RL0700	flg1	NE	NE	NE	NE	NE	NE
fliN [*]	RL0701	flg1	NE	NE	NE	N/A	N/A	N/A
fliM ^{b*}	RL0702	flg1	NE	NE	NE	NE	ES/DE	ES/DE
motA ^{b*}	RL0703	flg1	NE	NE	NE	NE	ES/DE	ES

Table A.1: (continued)

Name	Gene	Clusters	Input1	Rhizosphere	Root	Input2	Nodule bacteria	Bacteroids
flgF ^{b*}	RL0704	flg1	NE	NE	NE	NE	ES/DE	ES
fliI ^{b*}	RL0705	flg1	NE	NE	NE	NE	ES/DE	ES/DE
	RL0706	flg1	NE	NE	NE	NE	NE	NE
flgB ^{b*}	RL0707	flg1	NE	NE	NE	NE	ES/DE	ES/DE
flgC ^{b*}	RL0708	flg1	NE	NE	NE	NE	ES/DE	ES/DE
fliE [*]	RL0709	flg1	N/A	N/A	N/A	N/A	N/A	N/A
flgG ^{b*}	RL0710	flg1	NE	NE	NE	NE	ES/DE	ES/DE
flgA	RL0711	flg1	NE	NE	NE	NE	NE	NE
flgI ^b	RL0712	flg1	NE	NE	NE	NE	ES	ES
motE ^b	RL0713	flg1	NE	NE	NE	NE	DE	DE
flgH ^b	RL0714	flg1	NE	NE	NE	NE	DE	DE
fliL ^{ab}	RL0715	flg1	NE	AD	AD	NE	DE	DE
fliP ^{ab*}	RL0716	flg1	NE	AD	AD	NE	DE	ES/DE
	RL0717	flg1	NE	AD	AD	NE	NE	NE
flaA	RL0718	flg1	NE	NE	NE	NE	NE	NE
flaB	RL0719	flg1	NE	NE	NE	NE	NE	NE
flaC	RL0720	flg1	NE	NE	NE	NE	NE	NE
flaD	RL0721	flg1	NE	NE	NE	NE	NE	NE
	RL0722	flg1	NE	NE	NE	NE	NE	NE
motB ^{b*}	RL0723	flg1	NE	NE	NE	NE	DE	ES/DE
motC ^{b*}	RL0724	flg1	NE	NE	NE	NE	DE	ES/DE
fliK ^{b*}	RL0725	flg1	NE	NE	NE	NE	ES/DE	ES/DE
sltF ^b	RL0726	flg1	NE	NE	NE	NE	DE	DE
rem ^b	RL0727	flg1	NE	NE	NE	NE	DE	DE
flgE1 ^{ab*}	RL0728	flg1	NE	AD	AD	NE	DE	ES/DE

Table A.1: *(continued)*

Name	Gene	Clusters	Input1	Rhizosphere	Root	Input2	Nodule bacteria	Bacteroids
flgK1 ^{b*}	RL0729	flg1	NE	NE	NE	NE	ES/DE	ES/DE
flgL1 ^{b*}	RL0730	flg1	NE	NE	NE	NE	ES/DE	DE
flaF ^b	RL0731	flg1	NE	NE	NE	NE	DE	DE
flbT ^b	RL0732	flg1	NE	NE	NE	NE	DE	DE
flgD ^{b*}	RL0733	flg1	NE	NE	NE	NE	DE	ES/DE
fliQ	RL0734	flg1	N/A	N/A	N/A	N/A	N/A	N/A
flhA ^{b*}	RL0735	flg1	NE	NE	NE	NE	DE	ES/DE
fliR	RL0736	flg1	NE	NE	NE	NE	NE	NE
mcpZ	RL0757		NE	NE	NE	NE	NE	NE
mcpI	RL0758		NE	NE	NE	NE	NE	NE
mcpJ	RL0949		NE	NE	NE	NE	NE	NE
mcpK	RL0972		NE	NE	NE	NE	NE	NE
mcpL	RL1065		NE	NE	NE	NE	NE	NE
mcpM	RL1318		NE	NE	NE	NE	NE	NE
MCP putative	RL1386		NE	NE	NE	NE	NE	NE
MCP putative	RL1447		NE	NE	NE	NE	NE	NE
flaR1	RL2250		NE	NE	NE	NE	NE	NE
flaR2	RL2251		NE	NE	NE	NE	NE	NE
flgC hypothetical ^{b*}	RL2575		NE	NE	NE	NE	ES/DE	ES/DE
mcpD	RL2683		NE	NE	NE	NE	NE	NE
mcpN	RL2931		NE	NE	NE	NE	NE	NE
flaH ^c	RL3268		NE	NE	NE	NE	AD	NE
cheW3	RL3289		NE	NE	NE	NE	NE	NE
flgJ	RL3927		AD	NE	NE	NE	NE	NE
mcpO	RL3985		NE	NE	NE	NE	NE	NE

Table A.1: (*continued*)

Name	Gene	Clusters	Input1	Rhizosphere	Root	Input2	Nodule bacteria	Bacteroids
motB putative	RL4023		NE	NE	NE	NE	NE	NE
cheB2 ^c	RL4028	che2	NE	NE	NE	NE	AD	NE
cheR2 ^{ac}	RL4029	che2	NE	AD	AD	NE	AD	NE
cheW2b ^c	RL4030	che2	NE	NE	NE	NE	AD	NE
mcrA ^c	RL4031	che2	NE	NE	NE	NE	AD	NE
mcrB ^c	RL4032	che2	NE	NE	NE	NE	AD	NE
mcrC ^c	RL4033	che2	NE	NE	NE	NE	AD	NE
cheW2a ^c	RL4034	che2	NE	NE	NE	NE	AD	NE
cheA2 ^c	RL4035	che2	NE	NE	NE	NE	AD	NE
cheY3 ^c	RL4036	che2	NE	NE	NE	NE	AD	NE
mcpP	RL4277		NE	NE	NE	NE	NE	NE
cheW4 ^{*b}	RL4386		NE	NE	NE	NE	DE	ES/DE
mcpQ	RL4387		NE	NE	NE	NE	NE	NE
fliY	RL4634		NE	NE	NE	NE	NE	NE
flaG	RL4729		NE	NE	NE	NE	NE	NE
mcpS	pRL80031		NE	NE	NE	NE	NE	NE
mcpG	pRL100403		NE	NE	NE	NE	NE	NE
flaE	pRL110518		NE	NE	NE	NE	NE	NE
mcpR	pRL120056		NE	NE	NE	NE	NE	NE
flgL2	pRL120062	flg2	NE	NE	NE	NE	NE	NE
flgK2	pRL120063	flg2	NE	NE	NE	NE	NE	NE
flgE2	pRL120064	flg2	NE	NE	NE	NE	NE	NE
mcpY2	pRL120068		NE	NE	NE	NE	NE	NE
mcpC	pRL120312		NE	NE	NE	NE	NE	NE
mcpB	pRL120683		NE	NE	NE	NE	NE	NE

Table A.1: (*continued*)

Name	Gene	Clusters	Input1	Rhizosphere	Root	Input2	Nodule bacteria	Bacteroids
------	------	----------	--------	-------------	------	--------	-----------------	------------

^a Advantaged rhizosphere and root-colonised

^b Nodule-specific

^c Advantaged nodule-bacteria

* manually curated

References

- 10 Adebali O, Ortega DR, Zhulin IB. CDvist: A webserver for identification and visualization of conserved domains in protein sequences. *Bioinformatics* 2015;**31**:1475–7.
- Adler J. Chemoreceptors in bacteria. *Science* 1969;**166**:1588–97.
- Adler J, Epstein W. Phosphotransferase-system enzymes as chemoreceptors for certain sugars in *Escherichia coli* chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America* 1974;**71**:2895–9.
- Afanzar O, Di Paolo D, Eisenstein M *et al.* The switching mechanism of the bacterial rotary motor combines tight regulation with inherent flexibility. *The EMBO Journal* 2021;**40**:e104683.
- Ahlmann-Eltze C. *Ggsignif: Significance Brackets for 'Ggplot2'*, 2019.
- Alexander RP, Zhulin IB. Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. *Proceedings of the National Academy of Sciences* 2007;**104**:2885–90.
- Alexandre G. Chemotaxis Control of Transient Cell Aggregation. *J Bacteriol* 2015;**197**:3230–7.
- Alexandre G, Greer SE, Zhulin IB. Energy taxis is the dominant behavior in *Azospirillum brasilense*. *Journal of Bacteriology* 2000;**182**:6042–8.
- Alexandre G, Zhulin IB. More Than One Way To Sense Chemicals. *Journal of Bacteriology* 2001;**183**:4681–6.

- Allard-massicotte R, Tessier L, Lécuyer F *et al.* *Bacillus Subtilis* Early Colonization of *Arabidopsis Thaliana* Roots. 2016;**7**:1–0.
- Anantharaman V, Aravind L. Cache - A signaling domain common to animal Ca²⁺-channel subunits and a class of prokaryotic chemotaxis receptors. *Trends in Biochemical Sciences* 2000;**25**:535–7.
- Armitage JP, Schmitt R. Bacterial chemotaxis: *Rhodobacter sphaeroide* and *Sinorhizobium meliloti* - variations on a theme? *Microbiology* 1997;**143**:3671–82.
- Attmannspacher U, Scharf B, Schmitt R. Control of speed modulation (chemokinesis) in the unidirectional rotary motor of *Sinorhizobium meliloti*. *Molecular Microbiology* 2005;**56**:708–18.
- Auguie B. *gridExtra: Miscellaneous Functions for "Grid" Graphics.*, 2017.
- Bachhawat P, Swapna G, Stock AM. Mechanism of Activation for Transcription Factor PhoB Suggested by Different Modes of Dimerization in the Inactive and Active States. *Structure* 2005;**13**:1353–63.
- Baker AE, Webster SS, Diepold A *et al.* Flagellar Stators Stimulate c-di-GMP Production by *Pseudomonas aeruginosa*. *Journal of Bacteriology* 2019;**201**, DOI: 10.1128/JB.00741-18.
- Barak R, Eisenbach M. Fumarate or a fumarate metabolite restores switching ability to rotating flagella of bacterial envelopes. *Journal of Bacteriology* 1992;**174**:643–5.
- Barak R, Giebel I, Eisenbach M. The specificity of fumarate as a switching factor of the bacterial flagellar motor. *Molecular Microbiology* 1996;**19**:139–44.
- Baron S, Afanзар O, Eisenbach M. Methylation-independent adaptation in chemotaxis of *Escherichia coli* involves acetylation-dependent speed adaptation. *FEBS Lett* 2017;**591**:331–7.

- Baron S, Eisenbach M. CheY acetylation is required for ordinary adaptation time in *Escherichia coli* chemotaxis. *FEBS Letters* 2017;**591**:1958–65.
- Bates D, Mächler M, Bolker B *et al.* Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 2015;**67**:1–48.
- Berg HC. The Rotary Motor of Bacterial Flagella. *Annual Review of Biochemistry* 2003;**72**:19–54.
- Berke AP, Turner L, Berg HC *et al.* Hydrodynamic Attraction of Swimming Microorganisms by Surfaces. *Physical Review Letters* 2008;**101**:038102.
- Berman HM. The Protein Data Bank. *Nucleic Acids Research* 2000;**28**:235–42.
- Bernabéu-Roda L, Calatrava-Morales N, Cuéllar V *et al.* Characterization of surface motility in *Sinorhizobium meliloti*: Regulation and role in symbiosis. *Symbiosis* 2015;**67**:79–90.
- Berne C, Ducret A, Hardy GG *et al.* Adhesins involved in attachment to abiotic surfaces by Gram-negative bacteria. *Microbiology spectrum* 2015;**3**, DOI: 10.1128/microbiolspec.MB-0018-2015.
- Bhaya D, Takahashi A, Grossman AR. Light regulation of type IV pilus-dependent motility by chemosensor-like elements in *Synechocystis* PCC6803. *Proc Natl Acad Sci U S A* 2001;**98**:7540–5.
- Bibikov SI, Biran R, Rudd KE *et al.* A signal transducer for aerotaxis in *Escherichia coli*. *Journal of Bacteriology* 1997;**179**:4075–9.
- Brencic A, Winans SC. Detection of and Response to Signals Involved in Host-Microbe Interactions by Plant-Associated Bacteria. *Microbiology and Molecular Biology Reviews* 2005;**69**:155–94.
- Buchanan-Wollaston V. Generalized Transduction in *Rhizobium leguminosarum*. *Microbiology*, 1979;**112**:135–42.

- Catlow HY, Glenn AR, Dilworth MJ. Does rhizobial motility affect its ability to colonize along the legume root? *Soil Biology and Biochemistry* 1990;**22**:573–5.
- Cheng Q. Perspectives in Biological Nitrogen Fixation Research. *Journal of Integrative Plant Biology* 2008;**50**:786–98.
- Choi K-H, Schweizer HP. Mini-Tn7 insertion in bacteria with single attTn7 sites: Example *Pseudomonas aeruginosa*. *Nature protocols* 2006;**1**:153–61.
- Clock SA, Planet PJ, Perez BA *et al.* Outer Membrane Components of the Tad (Tight Adherence) Secretion of *Aggregatibacter actinomycetemcomitans*. *Journal of Bacteriology* 2008;**190**:980–90.
- Colin R, Sourjik V. Emergent properties of bacterial chemotaxis pathway. *Current Opinion in Microbiology* 2017;**39**:24–33.
- Collins KD, Lacal J, Ottemann KM. Internal Sense of Direction: Sensing and Signaling from Cytoplasmic Chemoreceptors. *Microbiology and Molecular Biology Reviews* 2014;**78**:672–84.
- Compton KK, Hildreth SB, Helm RF *et al.* *Sinorhizobium meliloti* Chemoreceptor McpV Senses Short-Chain Carboxylates via Direct Binding. Stock AM (ed.). *Journal of Bacteriology* 2018;**200**:1–6.
- da Veiga Leprevost F, Grüning BA, Alves Afrits S *et al.* BioContainers: An open-source and community-driven framework for software standardization. *Bioinformatics* 2017;**33**:2580–2.
- de Mendiburu F. *Agricolae: Statistical Procedures for Agricultural Research.*, 2020.
- Degli Esposti M, Martinez Romero E. A survey of the energy metabolism of nodulating symbionts reveals a new form of respiratory complex I. *FEMS Microbiology Ecology* 2016;**92**, DOI: 10.1093/femsec/fiw084.

- DeJesus MA, Ioerger TR. A Hidden Markov Model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data. *BMC bioinformatics* 2013;**14**:303.
- Del Campo AM, Ballado T, De La Mora J *et al.* Chemotactic control of the two flagellar systems of *Rhodobacter sphaeroides* is mediated by different sets of CheY and FliM proteins. *Journal of Bacteriology* 2007;**189**:8397–401.
- Dent D, Cocking E. Establishing symbiotic nitrogen fixation in cereals and other non-legume crops: The Greener Nitrogen Revolution. *Agriculture & Food Security* 2017;**6**:7.
- diCenzo GC, Sharthiya H, Nanda A *et al.* PhoU Allows Rapid Adaptation to High Phosphate Concentrations by Modulating PstSCAB Transport Rate in *Sinorhizobium meliloti*. *Journal of Bacteriology* 2017;**199**, DOI: 10.1128/JB.00143-17.
- Dorgan KM, Wooderchak WL, Wynn DP *et al.* An enzyme-coupled continuous spectrophotometric assay for S-adenosylmethionine-dependent methyltransferases. *Analytical Biochemistry* 2006;**350**:249–55.
- Downie JA. The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS microbiology reviews* 2010;**34**:150–70.
- Eddy SR. Accelerated profile HMM searches. *PLoS Computational Biology* 2011;**7**, DOI: 10.1371/journal.pcbi.1002195.
- Edwards JC, Johnson MS, Taylor BL. Differentiation between electron transport sensing and proton motive force sensing by the Aer and Tsr receptors for aerotaxis. *Molecular Microbiology* 2006;**62**:823–37.
- Egbert MD, Barandiaran XE, di Paolo EA. A minimal model of metabolism-based chemotaxis. *PLoS Computational Biology* 2010;**6**, DOI: 10.1371/journal.pcbi.1001004.

- Eggenhofer E, Haslbeck M, Scharf B. MotE serves as a new chaperone specific for the periplasmic motility protein, MotC, in *Sinorhizobium meliloti*. *Molecular Microbiology* 2004;**52**:701–12.
- El-Mashtoly SF, Gu Y, Yoshimura H *et al.* Protein conformation changes of HemAT-Bs upon ligand binding probed by ultraviolet resonance Raman spectroscopy. *Journal of Biological Chemistry* 2008;**283**:6942–9.
- Elsen S, Swem LR, Swem DL *et al.* RegB/RegA, a Highly Conserved Redox-Responding Global Two-Component Regulatory System. *Microbiol Mol Biol Rev* 2004;**68**:263–79.
- Fahraeus G. The Infection of Clover Root Hairs by Nodule Bacteria Studied by a Simple Glass Slide Technique. *Microbiology*, 1957;**16**:374–81.
- Fenner BJ, Tiwari RP, Reeve WG *et al.* *Sinorhizobium medicae* genes whose regulation involves the ActS and/or ActR signal transduction proteins. *FEMS Microbiology Letters* 2004;**236**:21–31.
- Fisher RF, Long SR. Rhizobiumplant signal exchange. *Nature* 1992;**357**:655–60.
- Fournier J, Timmers ACJ, Sieberer BJ *et al.* Mechanism of Infection Thread Elongation in Root Hairs of *Medicago truncatula* and Dynamic Interplay with Associated Rhizobial Colonization. *PLANT PHYSIOLOGY* 2008;**148**:1985–95.
- Fraiberg M, Afanзар O, Cassidy CK *et al.* CheY's acetylation sites responsible for generating clockwise flagellar rotation in *Escherichia coli*. *Mol Microbiol* 2015;**95**:231–44.
- Frederix M, Edwards A, Swiderska A *et al.* Mutation of praR in *Rhizobium leguminosarum* enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins. *Molecular Microbiology* 2014;**93**:464–78.
- Frymier PD, Ford RM, Berg HC *et al.* Three-dimensional tracking of motile bacteria near a solid planar surface. *PNAS* 1995;**92**:6195–9.

- Fu R, Wall JD, Voordouw G. DcrA, a c-type heme-containing methyl-accepting protein from *Desulfovibrio vulgaris* Hildenborough, senses the oxygen concentration or redox potential of the environment. *Journal of Bacteriology* 1994;**176**:344–50.
- Gage DJ. Analysis of infection thread development using Gfp- and DsRed-expressing *Sinorhizobium meliloti*. *Journal of Bacteriology* 2002;**184**:7042–6.
- Gage DJ. Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes. *Microbiology and Molecular Biology Reviews* 2004;**68**:280–300.
- Gage DJ, Margolin W. Hanging by a thread: Invasion of legume plants by rhizobia. *Current Opinion in Microbiology* 2000;**3**:613–7.
- Gao S, Wu H, Yu X *et al.* Swarming motility plays the major role in migration during tomato root colonization by *Bacillus subtilis* SWR01. *Biological Control* 2016;**98**:11–7.
- Gaworzewska ET, Carlile MJ. Positive Chemotaxis of *Rhizobium leguminosarum* and other Bacteria towards Root Exudates from Legumes and other Plants. *Microbiology* 1982;**128**:1179–88.
- Geddes BA, Ryu MH, Mus F *et al.* Use of plant colonizing bacteria as chassis for transfer of N₂-fixation to cereals. *Current Opinion in Biotechnology* 2015;**32**:216–22.
- Greer-Phillips SE, Alexandre G, Taylor BL *et al.* Aer and Tsr guide *Escherichia coli* in spatial gradients of oxidizable substrates. *Microbiology*, 2003;**149**:2661–7.
- Grün B, Leisch F. Fitting finite mixtures of generalized linear regressions in R. *Computational Statistics & Data Analysis* 2007;**51**:5247–52.
- Grün B, Leisch F. FlexMix version 2: Finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software*

2008;**28**:1–35.

Haag AF, Arnold MFF, Myka KK *et al.* Molecular insights into bacteroid development during *Rhizobium*-legume symbiosis. *FEMS Microbiol Rev* 2013;**37**, DOI: 10.1111/1574-6976.12003.

Haaker H, Szafran M, Wassink H *et al.* Respiratory control determines respiration and nitrogenase activity of *Rhizobium leguminosarum* bacteroids. *J Bacteriol* 1996;**178**:4555–62.

Heimbrook ME, Wang WL, Campbell G. Staining bacterial flagella easily. *Journal of Clinical Microbiology* 1989b;**27**:2612–5.

Heimbrook ME, Wang WL, Campbell G. Staining bacterial flagella easily. *Journal of Clinical Microbiology* 1989a;**27**:2612–5.

Hershey DM, Fiebig A, Crosson S. Flagellar perturbations activate adhesion through two distinct pathways in *Caulobacter crescentus*. *bioRxiv* 2020:2020.07.21.215269.

Hou S, Belisle C, Lam S *et al.* A globin-coupled oxygen sensor from the facultatively alkaliphilic *Bacillus halodurans* C-125. *Extremophiles* 2001;**5**:351–4.

Hsieh Y-J, Wanner BL. Global Regulation by the Seven-component Pi Signaling System. *Curr Opin Microbiol* 2010;**13**:198–203.

Hu Y, Jiao J, Liu LX *et al.* Evidence for Phosphate Starvation of Rhizobia without Terminal Differentiation in Legume Nodules. *MPMI* 2018;**31**:1060–8.

Hug I, Deshpande S, Sprecher KS *et al.* Second messenger-mediated tactile response by a bacterial rotary motor. *Science* 2017;**358**:531–4.

Huson DH, Scornavacca C. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic Biology* 2012;**61**:1061–7.

Ibáñez G, McBean JL, Astudillo YM *et al.* An enzyme-coupled ultrasensitive luminescence assay for protein methyltransferases. *Analytical Biochemistry* 2010;**401**:203–10.

- Imam S, Chen Z, Roos DS *et al.* Identification of Surprisingly Diverse Type IV Pili, across a Broad Range of Gram-Positive Bacteria. *PLoS One* 2011;**6**, DOI: 10.1371/journal.pone.0028919.
- Jackson LE, Schimel JP, Firestone MK. Short-term partitioning of ammonium and nitrate between plants and microbes in an annual grassland. *Soil Biology and Biochemistry* 1989;**21**:409–15.
- Jahn S, Haverkorn van Rijsewijk BR, Sauer U *et al.* A role for EIIA Ntr in controlling fluxes in the central metabolism of *E. Coli* K12. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 2013;**1833**:2879–89.
- Janczarek M. Environmental Signals and Regulatory Pathways That Influence Exopolysaccharide Production in Rhizobia. *International Journal of Molecular Sciences* 2011;**12**:7898–933.
- Johnston AWB, Beringer JE. Identification of the *Rhizobium* Strains in Pea Root Nodules Using Genetic Markers. *Journal of General Microbiology* 1975;**87**:343–50.
- Jones CW, Armitage JP. Positioning of bacterial chemoreceptors. *Trends in Microbiology* 2015, DOI: 10.1016/j.tim.2015.03.004.
- Jones KM, Kobayashi H, Davies BW *et al.* How rhizobial symbionts invade plants: The *Sinorhizobium* - *Medicago* model. *Nature Reviews Microbiology* 2007;**5**:619–33.
- Karunakaran R, Haag AF, East AK *et al.* BacA Is Essential for Bacteroid Development in Nodules of Galeoid, but not Phaseoloid, Legumes. *Journal of Bacteriology* 2010;**192**:2920–8.
- Karunakaran R, Ramachandran VK, Seaman JC *et al.* Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. *Journal of Bacteriology* 2009;**191**:4002–14.

- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 2013;**30**:772–80.
- Kawasaki A, Donn S, Ryan PR *et al.* Microbiome and Exudates of the Root and Rhizosphere of *Brachypodium distachyon*, a Model for Wheat. *PLoS ONE* 2016;**11**, DOI: 10.1371/journal.pone.0164533.
- Kim HJ, Lee CR, Kim M *et al.* Dephosphorylated NPr of the nitrogen PTS regulates lipid A biosynthesis by direct interaction with LpxD. *Biochemical and Biophysical Research Communications* 2011;**409**:556–61.
- Kodaka H, Armfield AY, Lombard GL *et al.* Practical procedure for demonstrating bacterial flagella. *J Clin Microbiol* 1982b;**16**:948–52.
- Kodaka H, Armfield AY, Lombard GL *et al.* Practical procedure for demonstrating bacterial flagella. *J Clin Microbiol* 1982a;**16**:948–52.
- Koganitsky A, Tworowski D, Dadosh T *et al.* A Mechanism of Modulating the Direction of Flagellar Rotation in Bacteria by Fumarate and Fumarate Reductase. *Journal of Molecular Biology* 2019;**431**:3662–76.
- Kojima S, Blair DF. Conformational Change in the Stator of the Bacterial Flagellar Motor. *Biochemistry* 2001;**40**:13041–50.
- Krehenbrink M, Downie JA. Identification of protein secretion systems and novel secreted proteins in *Rhizobium leguminosarum* bv. *Viciae*. *BMC Genomics* 2008;**9**:55.
- Kremers G-J, Goedhart J, van Munster EB *et al.* Cyan and Yellow Super Fluorescent Proteins with Improved Brightness, Protein Folding, and FRET Förster Radius,. *Biochemistry* 2006;**45**:6570–80.
- Krogh A, Larsson B, Von Heijne G *et al.* Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology* 2001;**305**:567–80.

- Krol E, Becker A. Global transcriptional analysis of the phosphate starvation response in *Sinorhizobium meliloti* strains 1021 and 2011. *Mol Genet Genomics* 2004;**272**:1–7.
- Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 2017;**82**:1–26.
- Lacal J, Alfonso C, Liu X *et al.* Identification of a chemoreceptor for tricarboxylic acid cycle intermediates: Differential chemotactic response towards receptor ligands. *Journal of Biological Chemistry* 2010a;**285**:23126–36.
- Lacal J, García-Fontana C, Muñoz-Martínez F *et al.* Sensing of environmental signals: Classification of chemoreceptors according to the size of their ligand binding regions. *Environmental Microbiology* 2010b;**12**:2873–84.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012;**9**:357–9.
- Lauffenburger D, Aris R, Keller K. Effects of cell motility and chemotaxis on microbial population growth. *Biophysical Journal* 1982;**40**:209–19.
- Leisch F. FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software* 2004;**11**:1–8.
- Lengeler J, Auburger AM, Mayer R *et al.* The phosphoenolpyruvate-dependent carbohydrate: Phosphotransferase system enzymes II as chemoreceptors in chemotaxis of *Escherichia coli* K12. *MGG Molecular & General Genetics* 1981;**183**:163–70.
- Lenth R. *Emmeans: Estimated Marginal Means, Aka Least-Squares Means.*, 2020.
- Li G, Tam L-K, Tang JX. Amplified effect of Brownian motion in bacterial near-surface swimming. *PNAS* 2008;**105**:18355–9.
- Li H. Improving SNP discovery by base alignment quality. *Bioinformatics (Oxford, England)* 2011;**27**:1157–8.

- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 2009;**25**:1754–60.
- Li H, Handsaker B, Wysoker A *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 2009;**25**:2078–9.
- Lux R, Jahreis K, Bettenbrock K *et al.* Coupling the phosphotransferase system and the methyl-accepting chemotaxis protein-dependent chemotaxis signaling pathways of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 1995;**92**:11583–7.
- Lüdecke D. Ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software* 2018;**3**:772.
- Lüdecke D. *sjPlot: Data Visualization for Statistics in Social Science.*, 2020.
- Macnab RM, Aizawa SI. Bacterial Motility and the Bacterial Flagellar Motor. *Annual Review of Biophysics and Bioengineering* 1984;**13**:51–83.
- Makino K, Shinagawa H, Amemura M *et al.* Signal transduction in the phosphate regulon of *Escherichia coli* involves phosphotransfer between PhoR and PhoB proteins. *Journal of Molecular Biology* 1989;**210**:551–9.
- Malpica R, Franco B, Rodriguez C *et al.* Identification of a quinone-sensitive redox switch in the ArcB sensor kinase. *PNAS* 2004;**101**:13318–23.
- Manson MD, Tedesco P, Berg HC *et al.* A protonmotive force drives bacterial flagella. *PNAS* 1977;**74**:3060–4.
- Martín-Mora D, Ortega Á, Matilla MA *et al.* The Molecular Mechanism of Nitrate Chemotaxis via Direct Ligand Binding to the PilJ Domain of McpN. Laub MT (ed.). *mBio* 2019;**10**:1–7.
- Mattick JS. Type IV Pili and Twitching Motility. *Annu Rev Microbiol* 2002;**56**:289–314.

- McCarter LL. Regulation of flagella. *Current Opinion in Microbiology* 2006;**9**:180–6.
- Meier VM, Muschler P, Scharf BE. Functional analysis of nine putative chemoreceptor proteins in *Sinorhizobium meliloti*. *Journal of Bacteriology* 2007;**189**:1816–26.
- Meier VM, Scharf BE. Cellular localization of predicted transmembrane and soluble chemoreceptors in *Sinorhizobium meliloti*. *Journal of Bacteriology* 2009;**191**:5724–33.
- Mergaert P, Uchiumi T, Alunni B *et al.* Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proceedings of the National Academy of Sciences* 2006;**103**:5230–5.
- Miller LD. Characterization of the chemotaxis system of the endosymbiotic bacterium *Rhizobium Leguminosarum* bv. *Viciae* strain 3841. 2007.
- Miller LD, Yost CK, Hynes MF *et al.* The major chemotaxis gene cluster of *Rhizobium leguminosarum* bv. *Viciae* is essential for competitive nodulation. *Molecular Microbiology* 2007;**63**:348–62.
- Mohari B, Licata NA, Kysela DT *et al.* Novel Pseudotaxis Mechanisms Improve Migration of Straight-Swimming Bacterial Mutants Through a Porous Environment. *mBio* 2015;**6**:1–2.
- Monahan-Giovanelli H, Pinedo CA, Gage DJ. Architecture of Infection Thread Networks in Developing Root Nodules Induced by the Symbiotic Bacterium *Sinorhizobium meliloti* on *Medicago truncatula*. *Plant Physiology* 2006;**140**:661–70.
- Mougel C, Zhulin IB. CHASE: An extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. *Trends in Biochemical Sciences* 2001;**26**:582–4.

- Möglich A, Moffat K. Domain of the Photosensor YtvA. *Journal of Molecular Biology* 2007;**373**:112–26.
- Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* 2007;**2**:2212–21.
- Nitzschke A, Bettenbrock K. All three quinone species play distinct roles in ensuring optimal growth under aerobic and fermentative conditions in *E. Coli* K12. *PLOS ONE* 2018;**13**:e0194699.
- Nochino N, Toya Y, Shimizu H. Transcription Factor ArcA is a Flux Sensor for the Oxygen Consumption Rate in *Escherichia coli*. *Biotechnology Journal* 2020;**15**:1900353.
- Oldroyd GED. Speak, friend, and enter: Signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews Microbiology* 2013;**11**, DOI: 10.1038/nrmicro2990.
- Oldroyd GED, Dixon R. Biotechnological solutions to the nitrogen problem. *Current Opinion in Biotechnology* 2014;**26**:19–24.
- Oldroyd GED, Downie JA. Coordinating Nodule Morphogenesis with Rhizobial Infection in Legumes. *Annual Review of Plant Biology* 2008;**59**:519–46.
- Oldroyd GED, Murray JD, Poole PS *et al.* The Rules of Engagement in the Legume-Rhizobial Symbiosis. *Annual Review of Genetics* 2011;**45**:119–44.
- Oliveira NM, Foster KR, Durham WM. Single-cell twitching chemotaxis in developing biofilms. *Proceedings of the National Academy of Sciences* 2016;**113**:6532–7.
- Ortega Á, Krell T. The HBM domain: Introducing bimodularity to bacterial sensing. *Protein Science* 2014;**23**:332–6.

- Ortega DR, Zhulin IB. Phylogenetic and Protein Sequence Analysis of Bacterial Chemoreceptors. In: Manson MD (ed.). *Bacterial Chemosensing: Methods and Protocols*. New York, NY: Springer New York, 2018, 373–85.
- P'edelacq J-D, Cabantous S, Tran T *et al.* Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* 2006;**24**:79–88.
- Park S, Park Y-H, Lee C-R *et al.* Glucose induces delocalization of a flagellar biosynthesis protein from the flagellated pole. *Molecular Microbiology* 2016;**101**:795–808.
- Perez-Mendoza D, Aragon IM, Prada-Ramirez HA *et al.* Responses to Elevated c-di-GMP Levels in Mutualistic and Pathogenic Plant-Interacting Bacteria. *PLOS ONE* 2014;**9**:e91645.
- Perry BJ, Akter MS, Yost CK. The Use of Transposon Insertion Sequencing to Interrogate the Core Functional Genome of the Legume Symbiont *Rhizobium leguminosarum*. *Frontiers in Microbiology* 2016;**7**, DOI: 10.3389/fmicb.2016.01873.
- Perry BJ, Yost CK. Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the Rhizobiaceae. *BMC Microbiology* 2014;**14**:298.
- Pflüger-Grau K, Görke B. Regulatory roles of the bacterial nitrogen-related phosphotransferase system. *Trends in Microbiology* 2010;**18**:205–14.
- Picard toolkit. *Broad Institute, GitHub repository* 2019.
- Pini F, East AK, Appia-Ayme C *et al.* Bacterial Biosensors for in Vivo Spatiotemporal Mapping of Root Secretion. *Plant Physiology* 2017;**174**:1289–306.
- Poole P, Ramachandran V, Terpolilli J. Rhizobia: From saprophytes to endosymbionts. *Nature Publishing Group* 2018, DOI: 10.1038/nrmicro.2017.171.
- Poole PS, Schofield NA, Reid CJ *et al.* Identification of chromosomal genes located downstream of *dctD* that affect the requirement for calcium and

- the lipopolysaccharide layer of *Rhizobium leguminosarum*. *Microbiology* 1994;**140**:2797–809.
- Porter SL, Wadhams GH, Armitage JP. Signal processing in complex chemotaxis pathways. *Nature Reviews Microbiology* 2011;**9**:153–65.
- Postma PW, Lengeler JW, Jacobson GR. Phosphoenolpyruvate: Carbohydrate phosphotransferase systems of bacteria. *Microbiological reviews* 1993;**57**:543–94.
- Prasad K, Caplan SR, Eisenbach M. Fumarate modulates bacterial flagellar rotation by lowering the free energy difference between the clockwise and counterclockwise states of the motor. *Journal of Molecular Biology* 1998;**280**:821–8.
- Pratt JT, McDonough E, Camilli A. PhoB Regulates Motility, Biofilms, and Cyclic di-GMP in *Vibrio cholerae*. *Journal of Bacteriology* 2009;**191**:6632–42.
- Prell J, Mulley G, Haufe F *et al.* The PTSNtr system globally regulates ATP-dependent transporters in *Rhizobium leguminosarum*. *Molecular Microbiology* 2012;**84**:117–29.
- Prell J, Poole P. Metabolic changes of rhizobia in legume nodules. *Trends in Microbiology* 2006;**14**:161–8.
- Prell J, White JP, Bourdes A *et al.* Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. *Proceedings of the National Academy of Sciences* 2009;**106**:12477–82.
- Priefer UB, Aurag J, Boesten B *et al.* Characterisation of *Phaseolus* symbionts isolated from Mediterranean soils and analysis of genetic factors related to pH tolerance. *Journal of Biotechnology* 2001;**91**:223–36.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019.

- Rabus R, Reizer J, Paulsen I *et al.* Enzyme I Ntr from *Escherichia coli*. *The Journal of biological chemistry* 1999;**274**:26185–91.
- Raina J-B, Fernandez V, Lambert B *et al.* The role of microbial motility and chemotaxis in symbiosis. *Nature Reviews Microbiology* 2019, DOI: 10.1038/s41579-019-0182-9.
- Rebbapragada A, Johnson MS, Harding GP *et al.* The Aer protein and the serine chemoreceptor Tsr independently sense intracellular energy levels and transduce oxygen, redox, and energy signals for *Escherichia coli* behavior. *PNAS* 1997;**94**:10541–6.
- Repik A, Rebbapragada A, Johnson MS *et al.* PAS domain residues involved in signal transduction by the Aer redox sensor of *Escherichia coli*. *Molecular Microbiology* 2000;**36**:806–16.
- Rice MS, Dahlquist FW. Sites of deamidation and methylation in Tsr, a bacterial chemotaxis sensory transducer. *The Journal of biological chemistry* 1991;**266**:9746–53.
- Richardson AE, Barea JM, McNeill AM *et al.* Acquisition of phosphorus and nitrogen in the rhizosphere and plant growth promotion by microorganisms. *Plant and Soil* 2009;**321**:305–39.
- Rodriguez-Navarro DN, Dardanelli MS, Ruiz-Sainz JE. Attachment of bacteria to the roots of higher plants. *FEMS Microbiology Letters* 2007;**272**:127–36.
- Rosenberg H, Gerdes RG, Chegwiddden K. Two systems for the uptake of phosphate in *Escherichia coli*. *J Bacteriol* 1977;**131**:505–11.
- Rotter C, Mühlbacher S, Salamon D *et al.* Rem, a new transcriptional activator of motility and chemotaxis in *Sinorhizobium meliloti*. *Journal of Bacteriology* 2006;**188**:6932–42.
- Rowell EH, Smith JM, Wolfe A *et al.* CheA, CheW, and CheY are required for chemotaxis to oxygen and sugars of the phosphotransferase system in

- Escherichia coli*. *Journal of Bacteriology* 1995;**177**:6011–4.
- S'anchez-Cañizares C, Prell J, Pini F *et al*. Global control of bacterial nitrogen and carbon metabolism by a PTSNtr-regulated switch. *Proceedings of the National Academy of Sciences* 2020;**117**:10234–45.
- Salah Ud-Din AIM, Roujeinikova A. Methyl-accepting chemotaxis proteins: A core sensing element in prokaryotes and archaea. *Cellular and Molecular Life Sciences* 2017;**74**:3293–303.
- Santi C, Bogusz D, Franche C. Biological nitrogen fixation in non-legume plants. *Annals of Botany* 2013;**111**, DOI: 10.1093/aob/mct048.
- Sánchez-Cañizares C, Palacios J. Construction of a marker system for the evaluation of competitiveness for legume nodulation in *Rhizobium* strains. *Journal of Microbiological Methods* 2013;**92**:246–9.
- Schaper S, Krol E, Skotnicka D *et al*. Cyclic Di-GMP Regulates Multiple Cellular Functions in the Symbiotic Alphaproteobacterium *Sinorhizobium meliloti*. *Journal of Bacteriology* 2016;**198**:521–35.
- Schaper S, Steinchen W, Krol E *et al*. AraC-like transcriptional activator CuxR binds c-di-GMP by a PilZ-like mechanism to regulate extracellular polysaccharide production. *PNAS* 2017;**114**:E4822–31.
- Scharf BE, Hynes MF, Alexandre GM. Chemotaxis signaling systems in model beneficial plantbacteria associations. *Plant Molecular Biology* 2016;**90**:549–59.
- Scharf B, Schuster-Wolff-Bühning H, Rachel R *et al*. Mutational analysis of the *Rhizobium lupini* H13-3 and *Sinorhizobium meliloti* flagellin genes: Importance of flagellin A for flagellar filament structure and transcriptional regulation. *Journal of Bacteriology* 2001;**183**:5334–42.
- Schlesinger WH. On the fate of anthropogenic nitrogen. *Proceedings of the National Academy of Sciences* 2009;**106**:203–8.

- Schweinitzer T, Josenhans C. Bacterial energy taxis: A global strategy? *Arch Microbiol* 2010;**192**:507–20.
- Secchi E, Vitale A, Miño GL *et al.* The effect of flow on swimming bacteria controls the initial colonization of curved surfaces. *Nature Communications* 2020;**11**:2851.
- Shaner NC, Campbell RE, Steinbach PA *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. Red fluorescent protein. *Nature Biotechnology* 2004;**22**:1567–72.
- Shi W, Sun H. Type IV Pilus-Dependent Motility and Its Possible Role in Bacterial Pathogenesis. *Infect Immun* 2002;**70**:1–4.
- Shrestha M, Compton KK, Mancl JM *et al.* Structure of the sensory domain of McpX from *Sinorhizobium meliloti* , the first known bacterial chemotactic sensor for quaternary ammonium compounds. *Biochemical Journal* 2018;**475**:3949–62.
- Simons M, van der Bij AJ, Brand I *et al.* Gnotobiotic system for studying rhizosphere colonization by plant growth-promoting *Pseudomonas* bacteria. *Molecular plant-microbe interactions : MPMI (USA)* 1996.
- Smil V. Harvesting the Biosphere: The Human Impact. *Population and Development Review* 2011;**37**:613–36.
- Sourjik V, Muschler P, Scharf B *et al.* VisN and VisR are global regulators of chemotaxis, flagellar, and motility genes in *Sinorhizobium* (*Rhizobium*) *meliloti*. *Journal of Bacteriology* 2000;**182**:782–8.
- Sourjik V, Schmitt R. Different roles of CheY1 and CheY2 in the chemotaxis of *Rhizobium meliloti*. *Molecular microbiology* 1996;**22**:427–36.
- Sourjik V, Schmitt R. Phosphotransfer between CheA, CheY1, and CheY2 in the chemotaxis signal transduction chain of *Rhizobium meliloti*. *Biochemistry* 1998;**37**:2327–35.

- Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**:1312–3.
- Stock D, Namba K, Lee LK. Nanorotors and self-assembling macromolecular machines: The torque ring of the bacterial flagellar motor. *Current Opinion in Biotechnology* 2012;**23**:545–54.
- Tambalo DD, Bustard DE, Del Bel KL *et al.* Characterization and functional analysis of seven flagellin genes in *Rhizobium leguminosarum* bv. *Viciae*. Characterization of *R. Leguminosarum* flagellins. *BMC Microbiology* 2010a;**10**:1–6.
- Tambalo DD, Del Bel KL, Bustard DE *et al.* Regulation of flagellar, motility and chemotaxis genes in *Rhizobium leguminosarum* by the VisN/R-Rem cascade. *Microbiology* 2010b;**156**:1673–85.
- Taylor BL, Zhulin IB, Johnson MS. Aerotaxis and Other Energy-Sensing Behavior in Bacteria. *Annual Review of Microbiology* 1999;**53**:103–28.
- Terpolilli JJ, Hood GA, Poole PS. What Determines the Efficiency of N₂-Fixing *Rhizobium*-Legume Symbioses? Poole RK (ed.). *Advances in Microbial Physiology* 2012;**60**:325–89.
- Theophrastus, Hort A. *Enquiry into plants*. Cambridge, MA: Cambridge, MA: Harvard University Press., 2015.
- Tiwari RP, Reeve WG, Dilworth MJ *et al.* Acid tolerance in *Rhizobium meliloti* strain WSM419 involves a two-component sensor-regulator system. *Microbiology (Reading)* 1996;**142** (Pt 7):1693–704.
- Tomich M, Planet PJ, Figurski DH. The tad locus: Postcards from the widespread colonization island. *Nature Reviews Microbiology* 2007;**5**:363–75.
- Turnbull GA, Morgan JAW, Whipps JM *et al.* The role of bacterial motility in the survival and spread of *Pseudomonas fluorescens* in soil and in the attachment and colonisation of wheat roots. *FEMS Microbiology Ecology* 2001;**36**:21–31.

- Udvardi M, Poole PS. Transport and metabolism in legume-rhizobia symbioses. *Annual review of plant biology* 2013;**64**:781–805.
- Untiet V, Karunakaran R, Krämer M *et al.* ABC Transport Is Inactivated by the PTSNtr under Potassium Limitation in *Rhizobium leguminosarum* 3841. *PLoS ONE* 2013;**8**:2–10.
- Vasse J, de Billy F, Camut S *et al.* Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. *Journal of Bacteriology* 1990;**172**:4295–306.
- Visick KL, O’Shea TM, Klein AH *et al.* The sugar phosphotransferase system of *Vibrio fischeri* inhibits both motility and bioluminescence. *Journal of Bacteriology* 2007;**189**:2571–4.
- Wang W, Jiang Z, Westermann M *et al.* Three mutations in *Escherichia coli* that generate transformable functional flagella. *Journal of Bacteriology* 2012;**194**:5856–63.
- Waterhouse AM, Procter JB, Martin DMA *et al.* Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;**25**:1189–91.
- Webb BA, Compton KK, del Campo JSM *et al.* *Sinorhizobium Meliloti* Chemotaxis to Multiple Amino Acids Is Mediated by the Chemoreceptor McpU. *Molecular Plant-Microbe Interactions* 2017a;**30**:MPMI-04-17-0096.
- Webb BA, Helm RF, Scharf BE. Contribution of Individual Chemoreceptors to *Sinorhizobium meliloti* Chemotaxis Towards Amino Acids of Host and Nonhost Seed Exudates. *Mpmi* 2016;**29**:231–9.
- Webb BA, Hildreth S, Helm RF *et al.* *Sinorhizobium meliloti* chemoreceptor McpU mediates chemotaxis toward host plant exudates through direct proline sensing. *Applied and Environmental Microbiology* 2014;**80**:3404–15.

- Webb BA, Karl Compton K, Castañeda Saldaña R *et al.* *Sinorhizobium meliloti* chemotaxis to quaternary ammonium compounds is mediated by the chemoreceptor McpX. *Molecular Microbiology* 2017b;**103**:333–46.
- Wei X, Bauer WD. Starvation-induced changes in motility, chemotaxis, and flagellation of *Rhizobium meliloti*. *Applied and Environmental Microbiology* 1998;**64**:1708–14.
- Wheatley RM, Ford BL, Li L *et al.* Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. *Proceedings of the National Academy of Sciences* 2020, DOI: 10.1073/pnas.2009094117.
- Wheatley RM, Poole PS. Mechanisms of bacterial attachment to roots. *FEMS Microbiology Reviews* 2018;**42**:448–61.
- Wheatley RM, Ramachandran VK, Geddes BA *et al.* Role of O₂ in the growth of *Rhizobium leguminosarum* bv. *Viciae* 3841 on glucose and succinate. *Journal of Bacteriology* 2016;**199**:e00572–16.
- Wickham H, Averick M, Bryan J *et al.* Welcome to the tidyverse. *Journal of Open Source Software* 2019;**4**:1686.
- Wickham H, Seidel D. *Scales: Scale Functions for Visualization.*, 2020.
- Wilke CO. *Cowplot: Streamlined Plot Theme and Plot Annotations for 'Ggplot2'*, 2019.
- Willett JW, Crosson S. Atypical modes of bacterial histidine kinase signaling. *Molecular Microbiology* 2017;**103**:197–202.
- Wu J, Li J, Li G *et al.* The receptor binding site for the methyltransferase of bacterial chemotaxis is distinct from the sites of methylation. *Biochemistry* 1996;**35**:4984–93.
- Yost CK. Characterization of *Rhizobium leguminosarum* genes homologous to chemotaxis chemoreceptors. 1998.

- Yost CK, Del Bel KL, Quandt J *et al.* *Rhizobium leguminosarum* methyl-accepting chemotaxis protein genes are down-regulated in the pea nodule. *Archives of Microbiology* 2004;**182**:505–13.
- Yost CK, Rochepeau P, Hynes MF. *Rhizobium leguminosarum* contains a group of genes that appear to code for methyl-accepting chemotaxis proteins. *Microbiology* 1998;**144**:1945–56.
- Young JPW, Crossman LC, Johnston AWB *et al.* The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology* 2006;**7**, DOI: 10.1186/gb-2006-7-4-r34.
- Yuan Z-C, Zaheer R, Finan TM. Regulation and properties of PstSCAB, a high-affinity, high-velocity phosphate transport system of *Sinorhizobium meliloti*. *J Bacteriol* 2006;**188**:1089–102.
- Zatakia HM, Arapov TD, Meier VM *et al.* Cellular Stoichiometry of Methyl-Accepting Chemotaxis Proteins in *Sinorhizobium meliloti*. Zhulin IB (ed.). *Journal of Bacteriology* 2017;**200**:e00614–17.
- Zhang W, Olson JS, Phillips GN. Biophysical and kinetic characterization of HemAT, an aerotaxis receptor from *Bacillus subtilis*. *Biophysical Journal* 2005;**88**:2801–14.
- Zheng H, Mao Y, Teng J *et al.* Flagellar-dependent motility in *Mesorhizobium tianshanense* is involved in the early stage of plant host interaction: Study of an flgE mutant. *Current Microbiology* 2015;**70**:219–27.
- Zipfel C, Oldroyd GED. Plant signalling in symbiosis and immunity. *Nature* 2017, DOI: 10.1038/nature22009.