Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Formalist features determining the tempo and mode of evolution in *Pseudomonas fluorescens* SBW25

A thesis submitted in partial fulfilment of the requirements for the degree

of

Ph.D.

In

Evolutionary Genetics

At Massey University, Auckland, New Zealand

Andrew David Farr

2015

Abstract

In order to explain the adaptive process, it is necessary to understand the generation of heritable phenotypic variation. For much of the history of evolutionary biology, the production of phenotypic variation was believed to be unbiased, and adaptation the primary outcome of selection acting on randomly generated variation (mutation). While true, 'internal' features of organisms may also play a role by increasing the rate of mutation at specific loci, or rendering certain genes better able to translate mutation into phenotypic variation. This thesis, using a bacterial model system, demonstrates how these internal features – localised mutation rates and genetic architectures – can influence the production of phenotypic variation.

Previous work involving the bacterium *Pseudomonas fluorescens* SBW25 has shown that mutations at three loci, *wsp, aws* and *mws*, can cause the adaptive wrinkly spreader (WS) phenotype. For each locus, the causal mutations are primarily in negative regulators of di-guanylate cyclase (DGC) activity, which readily convert mutation into the WS phenotype. Mutations causing WS at other loci were predicted to arise, but to do so with less frequent types of mutation. The data presented in this thesis confirms this prediction. My work began with the identification and characterisation of a single rare WS-causing mutation: an in-frame deletion that generates a translational fusion of genes *fadA* and *fwsR*. The fusion couples a DGC (encoded by *fwsR*) to a membrane-spanning domain (encoded by *fadA*) causing relocalisation of the DGC to the cell membrane and the WS phenotype. This is one of the few examples of adaptation caused by gene fusion and protein relocation in a real-time evolution experiment. I next took an experimental evolution approach to isolate further rare WS types and characterized these, revealing a range of rarely taken mutational pathways to WS. Lastly, I describe an example of extreme molecular parallelism, in which a cell chaining phenotype is caused – without exception – by a single nucleotide substitution within the gene *nlpD*, despite multiple mutational pathways to this phenotype. Characterisation of different *nlpD* mutants suggests this molecular parallelism is caused by a high local mutation rate, possibly related to the initiation of transcription within this gene.

Acknowledgements

Foremost, I would like to thank my supervisor, Professor Paul Rainey for discussion and ideas, and for providing the opportunity to pursue my own research directions in such a quality environment. I am very grateful for the years of support. I am thankful for the opportunity to discuss and present research at the Gordon Research Conference in Microbial Population Biology (New Hampshire, USA, 2011). Funding for these projects was obtained by Prof Rainey, and was partially supported by the Marsden Fund of New Zealand. Thanks to the Royal Society of New Zealand and Marsden Fund Council for making fundamental research such as this possible. I am also very grateful to the New Zealand Institute for Advanced Study for financial support during my studies.

Thanks to the post-docs who helped school me on the theory and study of microbial evolution. Particular acknowledgement is due to Dr Peter Lind who provided guidance on the *nlpD* project and assisted with both the flow cytometry fitness assays, and performed the *nlpD* mutant reversion-frequency assay. Thanks to Dr Heather Hendrickson, who put an enormous effort in setting up the fluorescent microscope, keeping it maintained, and offering guidance in operation. Thanks to Dr Mike McDonald who made the deletions of the initial PBR716 type, and who – as main author on a review we wrote – pointed the way for the useful histories of the field. Thanks to Dr Jenna Gallie for help with my first transposon assay, provision of primers, advice whenever it was needed, and the pUIC3 Δ*wss* construct. Thanks to Dr Xue-Xian Zhang for his expansive protocols, Dr Monica Gerth for sage cloning advice, and to Dr Christian Kost for his encouragement and advice and gift of PBR923 and primers.

Thanks also to all those I've worked with over the years who made and make the Rainey Lab such a rewarding experience!

For providing suggestions for the improvement of this thesis gratefully thank Professor Paul Rainey, Dr Peter Lind, Katrina Rainey and Simon Coverdale.

Finally, sincere gratitude to my friends for making life fun, and the deepest appreciation to my family for tireless personal support – Linda, Gavin and Vera, I'm forever in your debt.

Table of Contents

Abstract I

Acknowledgements II

Table of Contents III

Table of Abbreviations IX

Chapter 1: Introduction 1

1.1 A very brief history of the theory of variation 1

1.1.1 Variation within Darwin's theory of natural selection 2

1.1.2 The first formalist alternative to Darwinian variation: orthogenesis 3

1.1.3 The second evolutionary alternative to Darwinian variation: saltation 5

1.1.4 Rejection of orthogenesis and saltation in the modern synthesis 6

1.2 Genetic architecture and the production of variation 8

1.2.1 Contemporary theories of the importance of variation 8

1.2.2 Contemporary theories of the role of genetic architecture on evolution 9

1.2.3 Insight on the constraint of genetic architecture from evo-devo 10

1.2.3.1 Parallel mutations to *shavenbaby* in *Drosophila* 10

1.2.3.2 Parallel mutations to *FRIGIDA* in *Arabidopsis* 11

1.2.3.3 Parallel mutations to *R2R3MYB* across Astrids 12

1.2.4 Implications of evo-devo examples on rate of variation production 12

1.2.5 Localised mutation rates may alter the production of phenotypic variation 14

1.3 Microbes as model systems to explore the relative bias of variation caused by genetic architecture 14

1.3.1 Microbes and the ability to identify the locus of mutation 15

1.3.2 Microbes and the ability to compare architectures via redundancy 16

1.3.3 The *Pseudomonas fluorescens* SBW25 model system 17

1.3.3.1 Adaptive radiation of SBW25 in experimental conditions 17

1.3.3.2 The structural basis of the WS phenotype 18

1.3.3.3 Mutations causing WS involve changes that affect the regulation of c-di-GMP 19

1.3.3.4 The wild type regulation and effects of individual DGCs remain unclear 19

1.3.4 DGCs provide a model to assess the effect of architecture on variation 20

1.3.4.1 The architectures of known WS-causing DGCs 21

1.3.4.1.1 The *wsp* operon 21

1.3.4.1.2 The *aws* operon 23

1.3.4.1.3 The *mws* composite gene 25
1.3.4.2 Summary of *wsp*, *aws* and *mws* architecture and relative rate of evolution of the three pathways 27
1.3.5 Negative regulation in *wsp*, *aws* and *mws* may increase the production of variation 28
1.4 Research objectives 30
Chapter 2: Materials and methods 31
2.1 Materials 31
2.1.1 Media and growth conditions 31
2.1.2 Antibiotics, reagents and enzymes 31
2.1.3 DNA Electrophoresis 32
2.1.4 Competent cell production 32
2.1.5 Photography and microscopy materials 32
2.1.6 Strain list 33
2.1.7 Transposon and plasmids 36
2.1.8 Primers 37
2.2 Methods 41
2.2.1 Plasmid purifications for electroporation and ligation 41
2.2.2 Purification of PCR products 41
2.2.3 Triparental conjugation 41
2.2.4 Transposon mutagenesis 42
2.2.5 Polymerase chain reaction (PCR) 42
2.2.5.1 Strand overlap extension PCR 43
2.2.6 Sanger sequencing 44
2.2.7 Evolution of ISWS types 45
2.2.8 Allelic exchange 45
2.2.9 Production and transformation of competent cells 46
2.2.10 Fitness assays 47
2.2.10.1 Fitness assays via colony morphology distinction 47
2.2.10.2 Competition fitness assays using flow cytometry 48
2.2.11 Transcriptional assay 49
2.2.12 Quantitative RT-qPCR 49
2.2.13 Fluorescence microscopy (ACP and GFP) 50
2.2.14 Whole-genome sequencing 51
2.2.15 SREE protocol 51
2.2.16 Ordering mutations of the SREE 52
2.2.17 Measuring *rifR* mutation rates 52

2.2.18 Statistical analyses 53

Chapter 3: Characterisation of an adaptive pathway created by domain fusion and protein relocalisation 55

3.1 Introduction 55

3.2 Aims 56

3.3 Results 57

3.3.1 Identification of the genetic basis of SWS and measurement of the effects on fitness 57

3.3.1.1 The SWS mutational locus spans two genes 57

3.3.1.2 The *fadA-fwsR* fusion is sufficient to form WS 60

3.3.1.3 The *fadA-fwsR* fusion causes WS of similar fitness to *wspF* mutants 62

3.3.2 Multiple *fadA-fwsR* fusion mutations cause WS types 63

3.3.3 The *fadA-fwsR* fusion requires an active DGC to cause WS 68

3.3.4 Investigation of how *fadA-fwsR* causes WS 70

3.3.4.1 Testing the hypothesis that promoter capture may increase expression of the DGC coding frame 70

3.3.4.1.1 Sequence data suggests the *fadA-fwsR* deletion alters the operator of transcription of *fwsR* 70

3.3.4.1.2 Promoter capture is not sufficient to cause SWS 71

3.3.4.2 The SWS phenotype requires FadA and FwsR to be translationally fused 72

3.3.4.3 Testing the hypothesis that relocalisation of the FwsR DGC causes SWS 74

3.3.4.3.1 Membrane localisation domains necessary for FwsR to cause the SWS phenotype 74

3.3.4.3.2 Fluorescent tagging shows relocation of the DGC to the membrane associated with the *fadA-fwsR* fusion 75

3.4 Discussion 81

3.4.1 Summary of results: relocalisation of FwsR to the membrane determines the WS phenotype 81

3.4.2 The potential contribution to evolution across loci is influenced by genetic architecture 81

3.4.3 *fadA-fwsR* as a model for the evolution of new genes 83

3.4.3.1 The *fadA-fwsR* mutation is a model example of the creative power of modularity 83

3.4.3.2 *FadA-fwsR* provides direct experimental evidence that translational fusion mutations cause adaptive change via protein relocalisation 84

3.4.3.3 Chimeric relocalisation as a novel mechanism behind new genes in the 'neofunctionalisation model' 85

3.4.4 Implications for secondary messenger research and future directions 86

Chapter 4: The identification of multiple loci that cause WS..... 89
4.1 Introduction 89
4.2 Aims 92
4.3 Results 93
4.3.1 The SREE methodology and the subsequent identification of mutations 93
4.3.2 Identification of genotypes that were not identified via sequencing of the endpoints 94
4.3.3 Identity of mutations at GGDEF-containing loci and associated WS phenotypes 96
4.3.3.1. Mutations to DGC loci during the SREE are rare and diverse, compared to the original REE 96
4.3.3.2 Mutations identified at rare WS loci 97
4.3.3.2.1 PFLU0085 97
4.3.3.2.2 *fadA-fwsR* 99
4.3.3.2.3 PFLU0458 and PFLU4744 101
4.3.3.2.4 PFLU0621 103
4.3.3.2.5 PFLU0956 105
4.3.3.2.6 PFLU4781 to PFLU4782 107
4.3.3.2.7 PFLU4858 109
4.3.3.2.8 PFLU2764 111
4.3.3.2.9 PFLU5698 113
4.3.4 The *wss* operon is the most frequent target of mutation in the SREE 115
4.3.5 Parallel evolution at non-DGC regulatory loci of *mvaT* and *nlpD*..... 116
4.3.5.1 PFLU4939 117
4.3.5.2 PFLU1301 (*nlpD*) 119
4.4 Discussion 122
4.4.1 General summary of putative DGC-regulatory mutations 122
4.4.2 Mutational set at PFLU0085 suggests post-translational negative regulation encoded at this locus 123
4.4.3 The remaining putative c-di-GMP regulators require co-occurring mutations, fusions or promoter mutations to cause WS 124
4.4.4 The *wss* operon is the source of the greatest parallelism in the SREE 125
4.4.5 Parallel evolution during the SREE is associated with negative regulators 126
4.4.5.1 Parallelism at locus PFLU4939 is associated with loss-of-function mutations 127
4.4.5.2 Accounting for parallelism in the *nlpD* gene 129
Chapter 5: The causes of parallel molecular evolution of *nlpD* mutations 131

5.1 Introduction 131
5.1.1 The observed parallelism to *nlpD* and the associated phenotype change .132
5.1.2 The parallel evolution of C565T mutants may have several causes 133
5.2 Aim.....134
5.3 Results 135
5.3.1 A guide to the nomenclature of mutations presented in this chapter 135
5.3.2 Identification of the possible spectrum of *nlpD* mutants evolved from SBW25 Δ*wss* 135
5.3.3 The chaining phenotype caused by *nlpD* mutations does not require transcription of *rpoS* 137
5.3.4 Alteration of the *nlpD* sequence causes a lower frequency of molecular parallelism 139
5.3.5 Assessing whether the fitness of *nlpD* mutants explains parallel evolution142
5.3.6 The reduced parallelism of *nlpD* subtle mutants is associated with reduced *rpoS* transcription 144
5.3.7 Increased transcription from P_{rpoS} correlates with the frequency of mutations within *nlpD* 147
5.4. Discussion 151
5.4.1 Summary of the causes of parallel mutations to *nlpD* 151
5.4.2 A mutational mechanism may result in the parallelism of C565T.....152
5.4.3 Implications of the potential mutability of nested promoters at *nlpD* and other loci 154
5.4.4 What is the benefit of a nested *rpoS* promoter?.....156
5.4.4.1 The nesting of P_{rpoS} in *nlpD* may preserve the promoter.....156
5.4.4.2 The nesting of P_{rpoS} may cause a stochastic change in cell length 157
Chapter 6: Concluding remarks..... 159
6.1. Project background.....159
6.2 Summary of findings.....160
6.3 Implications of these findings for predicting evolution 164
Reference list..... 167
Appendix..... 187
Appendix for Chapter 3.....187
Section 7.3.1 Transcription assays suggest *fadA* and *fwsR* are independently transcribed.....190
Section 7.3.2 FadA-FwsR-GFP fusions have greater foci intensities than FadA-GFP fusions.....196

Appendix for Chapter 4 199
Appendix for Chapter 5 206
Section 7.5.1: SBW25 *nlpD* C565T is less fit than WS types 206
Section 7.5.2: The degree of change to NlpD is associated with cell-division deficiency 209

Table of Abbreviations

Abbreviation	Meaning
ACP	Acetylated cellulosic polymer
ALI	Air-liquid interface
BLAST	Basic local alignment search tool
bp	Base pair
c-di-GMP	bis-(3′-5′)-cyclic dimeric guanosine monophosphate
CCF	Cross correlation function
CFU	Colony forming units
DGC	Di-guanylate cyclase
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Dinucleotide triphosphate
EE	Experimental evolution
FAD	Fatty acid desaturase
FS	Fuzzy Spreader
g	Gravity or gram
GFP	Green fluorescent protein
Gm	Gentamicin
h	Hour
HT	Homopolymeric tract
IWS	Independent wrinkly spreader
ISWS	Independent slow wrinkly spreader
Km	Kanamycin
kb	kilobase
KB	King’s B medium
LB	Lysogeny broth
LSWS	Large spreading wrinkly spreader
M-W-W	Mann-Whitney-Wilcoxon
MCs	Microcosms
min	Minute
4MU	7-hydroxy-4-methylcoumarin
4MUG	4-methylumbelliferyl-b-d-galactoside
NF	Nitrofurantoin
ORF	Open reading frame
OD	Optical density
PCR	Polymerase chain reaction
PDE	Phosphodiesterase
RBS	Ribosome binding site
REE	Reverse evolution experiment
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT-qPCR	Reverse transcription quantitative real-time PCR

Abbreviation	Meaning
SE	Standard error
SM	Smooth morphotype
SOC	Super optimal broth with catabolic repressor
SREE	Slow reverse evolution experiment
SWS	Slow wrinkly spreader
TAM	Transcription associated mutation
Tc	Tetracycline
TFB	Transformation buffer
TIAM	Transcription initiation associated mutation
TMD	Transmembrane domain
WGS	Whole genome sequencing
WS	Wrinkly spreader
WSS	Wrinkly spreader structural locus
WT	Wild type
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Chapter 1:
Introduction

"… new and important modifications may not arise from reversion and analogous variation …"

Charles Darwin, *On the Origin of Species*, 1st edition, 1859. (p. 169)

"Such discontinuity is not in the environment; may it not, then be in the living thing itself?"

William Bateson, *Materials for the Study of Variation*, 1894. (p. 17)

1.1 A very brief history of the theory of variation

A full understanding of evolution requires an understanding of the production of heritable phenotypic variation. Phenotypic variation within a population is the ultimate source of evolutionary change. Variation is caused by mutation in the heritable code of organisms, and mutation is assumed to be random in respect to the phenotypes it can alter. Does this mean the production of phenotypic variation is random? This thesis will address this question, paying particular attention to whether the rate of production of variation for a particular phenotype can be biased by internal features of an organism.

Despite the fundamental nature of this question, rarely has the existence of biases been assessed as determining the production of heritable phenotypic variation. This paucity of enquiry is in part due to technical limitations, but is also due to the history of evolutionary thought. The question as to whether internal features of organisms may bias the production of variation has a controversial history. To begin this introduction, I will discuss the history of the theory of variation, paying particular attention to formalist theories of variation.

1.1.1 Variation within Darwin's theory of natural selection

In *On The Origin of Species* (referred to from here on in as *Origin*) Charles Darwin presented a theory of natural selection that was responsible for the diversity of life on earth (Darwin 1859). According to this theory, the evolution of new species is wrought by selection, which acts upon variation among a population. Darwin viewed this variation as small, continuous and isotropic (Gould 2002). Analogous variation, the observation that related subspecies produce variation in similar ways, while not entirely dismissed by Darwin, was relegated as only applicable to phenotypes of minimal importance. Despite his conviction of this position, Darwin admitted "… our ignorance of the laws of variation is profound …". This statement would prove to be prudent, given the rediscovery of the laws of heredity, the ultimate source of variation, lay 40 years in the future.

Theories on the nature of variation, in particular the potential for variation to be biased in both directions and non-gradual in manner, have been subject to fierce debate since the publication of *Origin* (Bowler 1992; Gould 2002). Considered criticism from prominent teleologist St George Mivart in *On the Genesis of Species* (Mivart 1871) led Darwin to formally defend his account of the gradual and isotropic nature of variation in the 6th edition of *Origin*. Mivart had criticised the ability of natural selection to account for the evolution of complex traits. Mivart further believed organisms could evolve rapidly, with extreme suddenness, resulting in the appearance of coordinated, complex phenotypes such as the eye.

To illustrate his theories on the fast and non-gradual production of variation, Mivart invoked the work of contemporary polymath Thomas Galton (Darwin's cousin). Galton had described the evolution of species as proceeding like the motion of a polyhedron rolling along a surface. Environmental forces would guide the direction of evolution (analogous to the forces that roll the polyhedron) and unknown internal features of the organism (analogous to the shape of the polyhedron) would determine the direction of evolution and the stability of an organism with a particular phenotype (Galton 1869). This analogy was used by Mivart to describe how internal directed factors may be responsible for the production of new variants. New variants would not arise gradually, but in dramatic motions, occasionally resulting in complex traits. Mivart could not provide a mechanism to account for the nature of the internal factors that would bias the production of variation. However, as Mivart was a devout teleologist, there could be no mistaking the ultimate source of the internal direction discussed in his criticism.

Darwin refuted Mivart's criticism in the 6th edition of *Origin*, providing a mechanism to account for the evolution of complex traits. Darwin accounted for such traits with the mechanism of 'preadaptation', in which the pleiotropic nature of organs (the ability of a given feature to perform multiple functions) could generate intermediate steps to complex traits via the generation of gradual and isotropic variants upon which selection could act. Despite taking Mivart's criticism seriously, Darwin maintained his doubt for variation to proceed by "an internal force or tendency", stating "there is no need, as it seems to me, to invoke any internal force beyond the tendency to ordinary variability …" (Darwin 1872).

1.1.2 The first formalist alternative to Darwinian variation: orthogenesis

Adherents to the theory of orthogenesis (coined by Haacke 1893) would provide a theoretical account of evolution that suggested variation was neither isotropic nor gradual, yet did not invoke teleological theistic mechanisms (Gould 2002). The proponents of orthogenesis were united by a theory that internal features of the organism directed the production of variation. This variation was either channelled (there was bias in the amount of variation in a phenotype), directional (evolution proceeded down a channel in a certain direction, sometimes leading to the extinction of species), or both. Though an explicit mechanism that could account for orthogenesis was lacking – a lack that would later be the source of criticism – prominent figures supported their theories with observational data.

Theodor Eimer was the most prominent advocate for a bias in the production of variation of evolving species, whilst totally refuting the teleological implications of an 'inner perfecting force' proposed by contemporaries such as Karl Naegeli (Bowler 1979). Eimer presented a theory of orthogenesis in which internal factors would provide channels that could direct or constrain variation in a limited number of ways, with external environmental factors determining which direction evolution would proceed. To Eimer, these environmental factors could include selection, which would cull unsuited changes. Eimer developed these theories based upon his personal observations of colouration in lizards, which presented evidence for allopatric speciation that was difficult to account for under Darwinian selection. Eimer also used observations on the seasonal dimorphism of butterflies (which would undergo phenotypic changes following changes in climate) (Eimer 1890) to support his theory.

Even Eimer's Darwinian contemporaries, such as August Weismann, admitted some of Eimer's examples suggested variation could be constrained to certain directions (Bowler 1979). However, Eimer would later become convinced that orthogenetic processes were universal in nature, a claim made without any knowledge of the nature of the processes that caused such biased channels, nor the evidence to support such a claim. Such non-pluralistic claims made by the most widely known proponent of orthogenesis, together with Eimer's invocation of Lamarckism as determining the direction of variation along a channel (Gould 2002), would damage the chances of the wider theory of orthogenesis being considered by later evolutionary theorists.

Charles Otis Whitman proposed a softer critique on the potential of internal factors as biasing the production of variation. Whitman was a gradualist in the Darwinian sense, but cited observational data suggesting directionality in the production of variation (Gould 2002). Whitman used Darwin's favoured case study of the evolution of breeds of pigeons in support of orthogenesis. By a comparative study of lineages of domestic pigeon and distantly related species of dove, Whitman contended the changing pattern of colouration of pigeon feathers proceeded in a unidirectional manner down a particular variation channel (Whitman and Riddle 1919). He did not believe selection could change the directionality of evolution. However, unlike Eimer, Whitman placed his theory of biased direction in evolution as complementary to the Darwinian theory of isotropic gradualism "Natural selection, orthogenesis … I believe each stands for truth, and that reconciliation is not distant" (Whitman and Riddle 1919). However, despite Whitman's inclusive account of bias in variation within a gradualism framework, he would receive little publicity, in no small part due to the death of Whitman nine years prior to his sole articulation of this theory in 1919 (Gould 2002). In the early 20th century, few proponents of orthogenesis remained to continue research into the shape and direction of variation.

The post-Darwinian era featured additional proponents of orthogenesis, however Eimer and Whitman represented the less hard-line of the theorists. Eimer's earlier writings and Whitman's work represent ideas that could be interpreted as complementary to natural selection. The central question raised from these works is: is there a bias in the production of variation upon which selection can act? Though an intriguing question, convincing evidence to support the directional or channelled nature of variation could not be produced without the confounding effect of selection as an additional mechanism required in the process.

1.1.3 The second evolutionary alternative to Darwinian variation: saltation

That variation may undergo saltational change received concentrated attention at the turn of the 20th century. In relation to variation, saltation refers to discontinuous variation, in which significant changes in phenotypes may arise de novo. The concept of saltation was not new to evolution, and the exclusion of saltation from *Origin* was quickly realised. After reading *Origin*, Darwin's great defender T. H. Huxley wrote to Darwin stating "you have loaded yourself with unnecessary difficulty in adopting *Natura non facit saltum*[1] so unreservedly" (Bowler 1992). The saltationists were formalists and, like the orthogenesists, identified internal features of organisms as creative in causing evolutionary change, and de-emphasised the functionalism of natural selection. To the formalists, natural selection would be relegated to a force that would remove strongly deleterious variants that had been created by these internal forces (Gould 2002). The saltationists supported these formalist theories with the developing theory of genetics – in fact many were crucial figures in the development of genetics.

William Bateson was one such saltationist and a pioneering geneticist[2]. His main contribution to the study of variation was *Materials for the Study of Variation* (1894) as reviewed in Bowler (1992) and Gould (2002). *Materials* consisted of a compendium of observations of variations from natural history, mainly consisting of instances of 'meristic' variation – variation in repeating, discontinuous structures of an organism. Examples of meristic qualities can be seen in flower petals and segments of antennae. Bateson argued that variation seen in such qualities between species could not have arisen via continuous variation, but by discrete discontinuous change (Bowler 1992). Furthermore, he rejected natural selection as a cause of meristic variation (Gould 2002). Perhaps because of this controversial interpretation, *Materials for the Study of Variation* received little attention at the time of publication, and would only receive more attention following Bateson's role in the burgeoning science of genetics – he was first to publish Mendel's work in English, and would coin the term 'genetics' in 1905.

[1] *Nature does not make leaps.*

[2] Despite such influence in the Mendelian school, Bateson believed variation to have a vibratory nature (based on the patterns of chemical and physical interactions), and not particulate (Gould 2002).

Hugo de Vries (one of the men responsible for the rediscovery of Mendel's work) would evoke saltation (which was termed mutationism) to explain observations of the production of variation in populations of *Oenothera lamarckiana*. In these large, experimental populations, de Vries noticed the propensity for lineages to undergo dramatic changes in morphology (this species was later identified as a highly mutable hybrid). De Vries explained the dramatic production of variants using a theory of mutationism (Gould 2002). De Vries proposed that mutations cause variants in any given aspect of phenotypes and in a non-gradual manner "… current belief assumes that species are slowly changed. In contradiction … the theory of mutation assumes that new species and varieties are produced from existing forms by sudden leaps" (De Vries 1909). The theory of mutationism would challenge Darwin's theory of natural selection as the mechanism for creating new species, by relegating natural selection to being merely a means of preserving mutated variants that happened to suit the local environment.

Richard Goldschmidt made further contributions to mutational theory and the production of variation. Goldschmidt would incorporate the study of ontogeny to provide an account for how mutations could produce radical new species. In *The Material Basis of Evolution* Goldschmidt claimed mutation could generate great changes in phenotypes by altering the timing of the developmental process in complex organisms (Goldschmidt 1940). His inclusion of development into evolutionary theory would foreshadow the contemporary field of evolutionary development, and was to influence the works of J. B. S. Haldane, one of the fathers of the modern synthesis. However, Goldschmidt's rejection of the particulate nature of genes, and continued rejection even following the discovery of DNA, would result in the exclusion of his earlier reasonable ideas in the modern synthesis (Gould 2002).

1.1.4 Rejection of orthogenesis and saltation in the modern synthesis

Mendelian inheritance would eventually become incorporated into evolutionary theory, resulting in the evolutionary synthesis. The synthesis was a consensus of evolutionary thinking and, as such, held the Darwinian theory of random isotropic variation at its core. The incorporation of Mendelian inheritance not only refined evolutionary theory but helped eliminate challenges to Darwinian processes such as 'blending inheritance'. The synthesis held core Darwinian principles on how populations evolve, and the idea of isotropic variation was central. Both orthogenesis and saltation were excluded from

the modern synthesis. This was not because of direct refutation against the two alternative evolutionary theories, but rather an understanding that the new evolutionary paradigm was sufficient to account for evolutionary processes (Gould 2002).

The first major work that unified Mendelian inheritance with selection was Fisher's *The Genetical Theory of Natural Selection* (1930). This theory depicted most mutations as causal of small phenotypic effects. In so doing, Fisher hypothesised a Mendelian basis for continuous variation. Mutations of large phenotypic effect were hypothesised to be mostly deleterious, and hence unlikely to be important for evolution generally (thus reducing the importance of saltation in creating new species) (Fisher 1930). Fisher also concluded that orthogenesis was unlikely. To reach this conclusion Fisher assumed orthogenesis would need to be caused by strong mutational pressure (mutations occurring at a high rate in a gene causing overall change in a population). Without evidence for a high mutation rate, such processes seemed unlikely.

Other founders of the modern synthesis did not so readily dismiss orthogenesis or saltation. Julian Huxley, who coined the term 'modern synthesis', accepted that species may be constrained in the production of variation along orthogenetic lines. Furthermore, he recognised there was no easy way of identifying such biases in variation – selection may often be mistaken for orthogenetic biases in the production of variation "we are not yet able to be sure … whether a limitation of variation as actually found in a group is due to a limitation in the supply of mutations or to selection, or to other causes" (Huxley 1942).

Despite the initial identification in the synthesis of saltation and orthogenesis as legitimate, albeit rare, evolutionary processes, these alternatives to Darwinian evolution were further relegated by later contributors of the modern synthesis. In what S. J. Gould described in *The Structure of Evolutionary Theory* as a 'hardening' of the synthesis, later writers would come to reject and ridicule such alternative explanations. This is partly due to proponents such as Eimer, Bateson and Goldschmidt, who had championed outdated ideas alongside the theories of orthogenesis and saltation. To confound matters, most of the proponents of these theories had died; by the coming of the synthesis few of these figures, other than Goldschmidt, would be living and able to try to defend these alternatives. Furthermore, the synthesis became rapidly convinced that alternative mechanisms to selection were not required to explain evolution. In this second phase, the mechanism of natural selection was considered all-powerful, to the

point that all features of organisms were believed to be wrought by natural selection (Gould and Lewontin 1979).

1.2 Genetic architecture and the production of variation

Modern molecular biology has enabled a clearer view of the genetic basis of phenotypes and an understanding of how mutation results in the production of variation. It is now clear that many phenotypes are encoded by sets of genes. In fact, many phenotypes are underpinned by complex genetic architectures. For future reference, I define genetic architecture as the connection and regulation of genes that determines a phenotype. Genome sequencing has brought the realisation that gene regulation and architecture are responsible for an inordinate amount of phenotypic change (King and Wilson 1975). Such studies have also demonstrated the need to understand molecular processes in order to understand evolutionary processes.

In 1985, stimulated by the increased knowledge of developmental constraints, and the recent interest in the sources of saltation and tempo of evolution (Gould and Eldredge 1977), an influential review brought to light the role of development in biasing the evolution of species (Maynard Smith et al. 1985). Maynard Smith and colleagues identified the importance of local constraints in biasing the course of evolution (local constraints being a subset of the universal constraints such as physicochemical laws, which confine the possible phenotypes available to selection). In their paper, local constraints were characterised as forces capable of biasing the production of variation. Amongst the mechanisms that can bias evolution, the authors questioned whether developmental constraints (genetic architectures) can constrict the production of variation and, moreover, whether development can act as a ‘directing force’ in the production of variation. The authors suggested ways of measuring the relative contribution of developmental constraints in selection. However, the tools available at that time for assessing influences such as developmental processes and selection were inadequate to directly disentangle the relative influence of biased variation and selection.

1.2.1 Contemporary theories of the importance of variation

We now have a better understanding of the importance of variation in guiding the course of evolutionary processes. The previous conviction of selection being the

mechanism shaping all evolutionary processes has been challenged by data that show a substantial impact of the production of variation in determining the course of evolution. Quantitative genetics has provided evidence for the concept of 'lines of least resistance' (Futuyma et al. 1993; Schluter 1996), which describes the evolution of species along the lines of greatest variance of multivariate traits. Multiple studies have presented quantitative data depicting this trend across multiple populations (Schluter 1996; Badyaev and Foresman 2000; Renaud et al. 2006; Boell 2013). However, it is extremely difficult to disentangle the effects of selection from such observational data, and the relative role of variation and selection requires further investigation in these specific populations (Schluter 1996).

1.2.2 Contemporary theories of the role of genetic architecture on evolution

That genetic architecture may be an important constraint of evolution has recently gained attention. There are now substantial theoretical studies on the effect of genetic architectures on the potential for phenotypes to vary (Hansen 2006; Wagner and Zhang 2011). A simple, yet important, underlying architectural feature is size. The number of genes determining a trait, as well as the size of the genes, creates a target that can increase the susceptibility of complex architectures to mutation (Houle 1998). This simple feature can bias the adaptive relevance of genes and genetic architectures given both mutation rate and the mutation type needed to produce adaptive outcomes.

The related architectural principles of epistasis (the mapping of several loci to a given trait) and pleiotropy (the effect of a single locus on several traits) have been identified as important constraints on the variability of traits, and hence on the capacity for organisms to evolve (Renaud et al. 2006; Pavlicev and Hansen 2011; Rajon and Plotkin 2013). Minimal pleiotropy of a phenotype (also known as a modular or autonomous phenotype) can allow increased variation of a trait by reducing the effect on other traits that may be deleterious when altered. Such architecture may be of varying benefit to an evolving organism (Orr 2000; Pavlicev and Hansen 2011; Wagner and Zhang 2011). In the special case of complex coordinated traits with low pleiotropy, such as a modular body part, mutations to the underlying genes may result in dramatic novelties (Kirschner and Gerhart 1998).

Similarly, the epistatic nature of a trait may determine the robustness of that trait in the event of mutation, which may be beneficial in certain environments (Rajon and Plotkin

2013). Epistasis of mutations underlying a given trait can be varied in nature (the effects can be additive, non-additive or opposing in fitness effect), and such interactions can produce cryptic variation that can be adaptive given further mutation or a change in environment (Hansen 2006).

Recent studies of genetic architecture (see Section 1.2.3) have demonstrated that architecture affects the evolutionary potential of organisms. These architectural studies are focused on the evolvability of organisms as a result of selection in genetic interactions. What has not been addressed is how architectures may bias variation directly (in the way Maynard Smith and colleagues imagined). Can genetic architectures bias the course of evolution by biasing the production of variation? What type of architectural features might result in a bias in the production of variation? Insight has been gained into these questions by the study of developmental systems and the architectures underpinning particular phenotypes.

1.2.3 Insight on the constraint of genetic architecture from evo-devo

The study of developmental systems and the architectures underpinning particular phenotypes has provided evidence that certain architectures may bias the production of variation. Developmental model systems have been utilised in the field of evolutionary developmental biology (evo-devo) to address how underlying developmental genetics can facilitate evolutionary change. An important observation from evo-devo studies is the repeated finding of mutations in certain loci of developmental cascades; mutations which result in complex phenotypic changes (Stern 2013). The confinement of such causal mutations is suggestive that certain genes – and possibly the regulatory position of these genes within the architecture underpinning the phenotypic trait – are more readily able to contribute to variation and hence evolution. Illustrative examples of the phenomenon of mutational parallelism in complex networks are found across multiple well-established model systems of development.

1.2.3.1 Parallel mutations to *shavenbaby* in *Drosophila*

The study of the developmental mechanisms underpinning epidermal cell fate in *Drosophila melanogaster* is one of the most well understood developmental pathways

in a complex species. One of the epidermal cell traits that has been studied is the formation of tricombes (epidermal hairs), and the variation of tricombes across *Drosophila* species (McGregor et al. 2007). Analysis of this variation between *D. melanogaster* and *D. sechellia* has revealed mutations at the regulatory region (specifically enhancer regions (McGregor et al. 2007)) of the transcription factor *shavenbaby* (*svb*) as being responsible for the loss of tricombes at particular dorsal regions of *D. sechellia* (Sucena and Stern 2000). This pattern of loss and the associated change in *svb* has been identified in the species group of *D. virilis* (Sucena et al. 2003), demonstrating the parallelism of this mutation in species with this particular epidermal developmental architecture.

Laboratory analysis has revealed mutations altering trichome patterning are not permitted by mutations elsewhere within the epidermal developmental pathway. Despite the large number of genes involved in regulation of trichome development, mutation of the downstream targets of *svb* are not capable of recapitulating the same phenotypic effects of mutations to *svb* (Chanut-Delalande et al. 2006). This clearly demonstrates the unique position of *svb* in this developmental 'module' of tricombe development. The pleiotropic effects of mutations on downstream effectors of *svb* are often invoked as the cause of this mutational constraint (Stern 2013). This model provides an example of how certain loci are relatively biased, due to their position in developmental processes, to convert mutation into phenotypic change.

1.2.3.2 Parallel mutations to *FRIGIDA* in *Arabidopsis*

The network of genes determining flowering time in *Arabidopsis thaliana* seems to display architectural constraints in the location of permissible mutations. The timing of flowering involves a complex interaction of physiological and developmental factors in response to environmental change and represents a dramatic developmental transition of plants from the vegetative to reproductive stage (Baurle and Dean 2006). This complex trait is underpinned by a complex network of over 100 genes (Srikanth and Schmid 2011). There is some variation in the timing of flowering, in which some accessions are rapid cycling and flower early, regardless of the climate. Most of this variation in flowering time is due to mutations to *FRIGIDA (FRI),* which by allowing early flowering, confers a selective benefit to the plant, presumably by allowing adaptation to local climates (Toomajian et al. 2006). Over 70% of the variation in flowering time in natural accessions of *A. thaliana* is associated with mutation to *FRI*

(Shindo et al. 2005). The prevalence of mutations in *FRI* is consistent with the role of FRI as a positive regulator of the early flowering repressor Flowering Locus C (FLC). Yet surprisingly, many genes in the flowering time network can mutate to form early flowering phenotypes. Of a set of over 40 experimentally-derived early flowering mutants (involving mutations to the network), only three involved mutations to *FRI* (Michaels and Amasino 1999). This implies a strong constraint on mutation to the rest of the flowering time network (Shindo et al. 2005). However, direct comparisons of phenotypes of early flowering mutants with mutations across the flowering network have not yet been reported.

1.2.3.3 Parallel mutations to *R2R3MYB* across Astrids

A similar mutational constraint has been identified in patterns of flower colour across Astrid plants. Across a diverse range of plant species, flower colour is determined by highly conserved flavonoid biosynthesis pathways (Winkel-Shirley 2001). Flux through the pathways determines the intensity of colour production, with loss-of-function mutations throughout the pathway resulting in colourless flowers (Streisfeld and Rausher 2011). Such mutations are not evolutionary dead ends, and may result in a transition of pollinators, such as bees to moths (Rausher 2008). The mutations resulting in deceasing flux may result via a spectrum of mutations, which may target enzymes of the pathway, cis-regulatory regions of these enzymes, and transcription factors (and their regulatory regions which regulate enzyme expression (Streisfeld and Rausher 2011)). However, across *Antirrhinum*, *Ipomoea* and *Petunia*, despite an even distribution of artificially induced mutations across this spectrum of targets, naturally arising variants of floral intensity were associated with mutations in transcription factors only, and in most cases the mutation was only identified in the transcription factor *R2R3MYB* (Streisfeld and Rausher 2011). Though no direct empirical tests have been made to test the relative fitness of mutants with mutation at different positions in the anthocyanin pathway, deleterious pleiotropy has been proposed as the most likely cause of this constraint (Wessinger and Rausher 2012).

1.2.4 Implications of evo-devo examples on rate of variation production

These examples provide evidence that genetic architecture is significant in the determination of the position of mutation underpinning phenotypic change. This bias in

mutational locus is due either to a functional constraint in the case of *Drosophila* (only mutations at a particular position can provide the phenotype), or constraint driven by selection as in the plant models (mutations are constrained due to selection against mutations elsewhere in networks). In all cases, it is suggested this bias is influenced by selection due to deleterious pleiotropy in most regions of the gene networks studied. It should be noted that selection as a cause of this constraint is not directly tested, with qualitative fitness measures inherently problematic in such developmental systems.

That mutations are biased to particular regions due to genetic architecture is fascinating, not just because it provides information on the mechanisms with which variation may be biased, but also because it provides clues that, given an understanding of architecture, the locus of evolution may be predictable (Stern and Orgogozo 2009; Stern 2013).

These examples also suggest (but do not fully articulate) a concept regarding the production of variation that is aligned with orthogenetic thinking (one wonders how the development of evolutionary theory might have progressed if only the proponents of orthogenesis had been provided with insight to the nature of genetic architecture). These studies hint that architecture may bias the production of variation relative to other architectures, which leads to the question: provided a given genetic architecture permits a mutable gene that confers an adaptive phenotypic change (such as *svb*, *FRI* or *R2R3MYB*), does this architecture confer a greater supply of variation, subject to selection, relative to different genetic architectures underpinning similar phenotypes? An answer to this question would not only exonerate the previously dismissed ideas of orthogenesis, but may also help make predictive claims about the course of evolution, given extensive knowledge of the genetic architectures underlying a beneficial phenotype of a given species.

To answer this question using biological models, the phenotypes underpinned by the various genetic architectures must be relatively comparable to each other. The regulatory interactions of the architectures must be identifiable, and the effects on fitness of mutations in these architectures must be quantifiable. Such a model system – capable of demonstrating a relative bias in the production of variation due to the inherent internal features of the organism – can be found within microbial model systems.

1.2.5 Localised mutation rates may alter the production of phenotypic variation

Differences in the mutation rate between regions of a genome provide a further mechanism that may bias the production of certain phenotypic variants. The process of mutation is considered to be spontaneous and rare (Wielgoss et al. 2011; Lee et al. 2012). However, recent evidence suggests mutation rates may vary across a given genome and between loci. Studies in *Escherichia coli* have revealed variation in the rate of mutation across the genome; slight increases in mutation rate have been detected on the leading strand of DNA replication (Gawel et al. 2014) and in regions adjacent to the origin of replication (Foster et al. 2013). At the level of individual genes, correlations have been observed between levels of transcription and the rate of mutation (Aguilera 2002; Hendriks et al. 2010; Kim and Jinks-Robertson 2012; Park et al. 2012). Specific secondary structures of DNA (such as stem-loop structures encoded by palindromic sequences) elevate the rate of particular point mutations (Viswanathan et al. 2000; Dutra and Lovett 2006).

High mutation rates that enable the variation of phenotypes may provide an adaptive benefit in stochastic environments. Across a wide range of taxa – including bacterial species – simple sequence repeats enable the rapid mutation of loci encoding epitopes recognisable by the adaptive immune system. This phenotypic variation enables pathogenic species such as *Bordetella pertussis* and *Haemophilus influenza* to rapidly evolve subpopulations that can evade detection by the adaptive immune system of a host (Moxon et al. 2006; Deitsch et al. 2009).

The studies above enable broad predictions of genomic regions, genes and particular sequences which may mutate at relatively high rates. Assuming the phenotypic consequences of such mutations are known, the identification of these mutable regions may allow predictions of the relative production of particular adaptive phenotypes that may be subject to selection.

1.3 Microbes as model systems to explore the relative bias of variation caused by genetic architecture

Experimental evolution (EE) with microbes has become a powerful tool to observe and test theories of evolutionary processes (Elena and Lenski 2003; Buckling et al. 2009;

Kawecki et al. 2012; Barrick and Lenski 2013), and offers great potential to explore the relative effects of architecture on variation. Microbial model systems have become widely used in EE studies for multiple reasons. These include experimental practicalities such as the rapidity of generation time (enabling many generations in a realistic experimental time), permissibility to cryogenic storage (enabling the comparison of ancestral and evolved types) and ability to generate large populations (allowing for selection to outweigh the effects of genetic drift).

1.3.1 Microbes and the ability to identify the locus of mutation

Two further advantages of the EE approach make it possible to ask fundamental questions about the evolutionary process: the ability to readily determine the mutational source of variation, and the ability to accurately assess the quantitative effects of that variation. Over the last decade the accuracy and decreasing cost of whole genome sequencing (WGS) technology have allowed the rapid detection of the mutational source of adaptive change of evolved populations. WGS, in combination with fitness assays that measure the change in fitness derived from a mutation, has allowed powerful assessments of evolutionary population processes.

The ability to identify the mutational source of adaptive traits also allows elucidation of the types and location of mutations. A high degree of mutational parallelism in adaptive traits has been determined in EE lineages across a diverse spectrum of experimental models (Bull et al. 1997; Wichman et al. 1999; Segre et al. 2006; Bantinaki et al. 2007; Barrick et al. 2009; Wang et al. 2010; Wong et al. 2012). In model organisms such as *Escherichia coli* where the function of genes is relatively well known, adaptation to nutrient-limited environments include parallelism to regulatory genes such as *fis* and *dusB* (which regulate superhelicity) (Crozat et al. 2010), *spoT* (which regulates dozens of genes to effect the stringent response) (Cooper et al. 2003), *nadR* (which acts to repress NAD synthesis and recycling) (Woods et al. 2006) and *rpoS* (which encodes a sigma factor which regulates genes activated or repressed in stationary phase) (Wang et al. 2010). Such genes can regulate a large number of genes; *rpoS* for example, in regulating the stress response in *E. coli*, regulates the expression – directly or indirectly – of up to 10% of the genome (Weber et al. 2005).

The finding of parallel mutations within regulatory genes of complex architectures suggests regulatory genes may commonly underpin adaptive traits. However, the effect

on fitness conferred by such mutations is the only explanation usually provided to account for the parallelism to such regulatory genes (Dettman et al. 2012; Barrick and Lenski 2013). This is surprising considering data that suggests the most frequent parallel mutations do not always confer the greatest fitness effect of the spectrum of adaptive mutants (Barrick et al. 2009; Conrad et al. 2009). An alternative explanation consistent with the above data is that the genetic architecture underpinning these traits elevates the production of variation. However, this explanation has not been invoked in relation to the above studies.

1.3.2 Microbes and the ability to compare architectures via redundancy

An aspect of microbial models that is suited to the study of variation is genetic redundancy. Microbial genetics has clearly demonstrated the great redundancy and resilience of microbes to gene deletion. This redundancy is demonstrated in the *E. coli* K-12 Keio collection of gene deletions, in which over 92% of genes were demonstrated as able to be individually deleted whilst permitting viability (Baba et al. 2006). Even when essential genes have been deleted, cellular viability may be maintained by overexpression of non-essential genes (Patrick et al. 2007; Bergmiller et al. 2012), further demonstrating the potential redundancy of many genes.

Moreover, experimental evolution (EE) has demonstrated the capacity for microbes to evolve the complex functionalities of deleted genes via mutations to previously unidentified genes and gene networks. Examples of this can be found in EE of both poxvirus (Elde et al. 2012) and phage (Heineman et al. 2009; Meyer et al. 2012). In these studies, the typical pathway encoding a complex trait has been deleted (or mutated in an environment with relaxed selection (Meyer et al. 2012)), and the functionality is returned via mutation to a previously unknown pathway. These studies demonstrate the multiple layers of functionality for a given trait, which are only uncovered following deletion of the main genetic pathway. Genetic redundancy will become important for the studies presented below, as it allows assessment of the relative production of variation generated by different genetic architectures in a microbial system.

Microbial systems hold great promise for direct investigations into whether genetic architecture can bias the production of variation. By using standard EE techniques, replicate lines of adaptive mutants arising in a given environment can be isolated and

the genetic basis of the adaptive change can be readily determined. The fitness of the adaptive trait can be measured via fitness assays, and mutants of equivalent fitness can be identified. Of the spectrum of mutants with equivalent fitness, architectures underpinning particular traits may present more or less variation subject to selection, resulting in more or less adaptive types. Very rare adaptive types may be identified via gene deletion of common pathways. The architectures underpinning bias in the production of adaptive types can then be assessed via molecular genetics. Such a strategy has been partially employed using a model system highly suited to such an approach – *Pseudomonas fluorescens* SBW25. This thesis presents data contributing to this general study.

1.3.3 The *Pseudomonas fluorescens* SBW25 model system

Pseudomonas fluorescens SBW25 was first isolated in 1989 from the phyllosphere (above-ground plant surface) of a sugar beet plant grown in Oxfordshire (Rainey and Bailey 1996). *Pseudomonas fluorescens* encompasses a genetically diverse range of saprophytic organisms capable of colonising a diverse range of terrestrial habitat niches, including soil, water and plant surfaces. The SBW25 isolate has been subject to extensive analysis and experimentation, and its genome is sequenced (Silby et al. 2009). This genomic analysis revealed an abundance of regulatory and chemotaxis genes, typical of soil-dwelling bacteria.

1.3.3.1 Adaptive radiation of SBW25 in experimental conditions

The potential of SBW25 as a model for EE was first described in Rainey and Travisano (1998). Here, the SBW25 ancestral isolate was inoculated in static microcosms containing King's medium B (KB) for 72 hours. Over this course of time, the ancestral type evolved niche specialist types, readily identifiable via distinct colony morphologies. The ancestral types would exhibit round, symmetrical colonies, termed 'smooth' (SM). Two main morphotypes were seen to evolve – wrinkly spreader (WS) types and fuzzy spreader (FS) types – which are both known to colonise the air-liquid interface (ALI) of statically incubated microcosms (see Figure 1.1) (Rainey and Travisano 1998; Ferguson et al. 2013). Despite slow doubling time of WS types, WS types readily invade SM types in statically incubated microcosms (Rainey and Travisano 1998). The WS types are capable of excluding the evolution of FS types in invasion from rare

assays (Rainey and Travisano 1998). Only recently have the ecology and underlying genetics of the FS types been described (Ferguson et al. 2013), with the majority of papers utilising SBW25 as a model for EE focusing on the evolution of WS types.

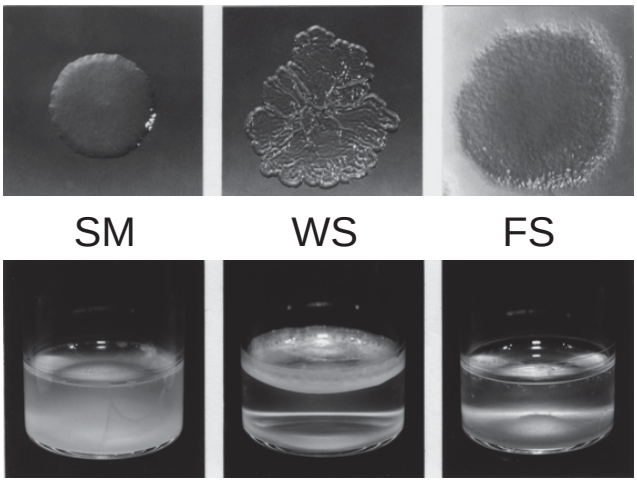

Figure 1.1: Morphologies and niches of smooth (SM), wrinkly spreader (WS) and fuzzy spreader (FS) types. The ancestral SM types diversify into WS and FS types (colony morphology above) after incubation in static microcosms for 72 hours, which allows colonisation of the air-liquid interface (ALI) (below). Although the FS type appears to colonise the bottom of the microcosms, it first grows at the ALI and then settles to the bottom of the microcosm. Figure adapted from Rainey and Travisano (1998).

The selective advantage of colonisation of the ALI has been identified. WS types can invade against the ancestral SM types via the evolution of a cooperative trait that enables a greater access to resources at the ALI (Rainey and Rainey 2003). A primary resource accessible at the ALI is oxygen – colonisation of the broth by ancestral SM types results in a steep oxygen gradient (Koza et al. 2011). The evolution of a biofilm is not unique to SBW25. Natural isolates of WS-like biofilm mutants – in both morphology, physiology and underlining genetic change – have been found in other *Pseudomonas* species (Kirisits et al. 2005) and *Enterobacteriaceae* (Romling 2005).

1.3.3.2 The structural basis of the WS phenotype

A combination of genetic and phenotypic characterisation has provided insight on the phenotypic changes that allow SBW25 WS types to colonise the ALI. Transposon

mutants of WS isolates – generated to identify the traits that allow ALI habitation – have demonstrated the importance of the *wss* operon for the expression of the WS phenotype (Spiers et al. 2002). The operon consists of 10 genes – *wssA* to *wssJ* – with partial homology in *wssB*, *wssC* and *wssE* genes in *E. coli* and *Acetobacter xylinum*, which are known to encode for cellulose synthase complexes (Spiers et al. 2002). The downstream genes – *wssF* to *wssJ* – are homologous to genes responsible for acetylation of alginate (a polysaccharide polymer) in *P. aeruginosa*. Biochemical analysis of mutants to *wssF* has demonstrated the role of these genes in acetylation of cellulose (Spiers et al. 2003). The *wss* operon is crucial to the phenotypic basis of WS types, with transposon mutation hits to the cellulose synthase complex returning the WS phenotype to the ancestral state. However, despite the central role in expression of the WS phenotype, mutations within the *wss* locus have not been identified as playing any role in generation of the WS phenotype.

1.3.3.3 Mutations causing WS involve changes that affect the regulation of c-di-GMP

Mutations to genes encoding di-guanylate cyclases (DGCs) are solely responsible for the evolution of WS types. DGCs are responsible for synthesis of the secondary messenger bis-(3'-5-)-cyclic dimeric guanosine monophosphate (c-di-GMP). This messenger is near-ubiquitous in bacteria (Galperin 2005), and has an important role in the transition between motile to sessile states (Jenal and Malone 2006; Hengge 2009). The active signalling molecule c-di-GMP is synthesised from two GMP molecules by DGC enzymes containing GGDEF domains (the GGDEF motif within the DGC is highly conserved (Galperin et al. 2001; Malone et al. 2007)). Conversely, the c-di-GMP molecules can by hydrolysed and inactivated by phosphodiesterase (PDE) enzymes containing EAL domains. In many species, genes encoding both domains can be found.

1.3.3.4 The wild type regulation and effects of individual DGCs remain unclear

The regulation of c-di-GMP production is highly variable. Typically, the GGDEF domain features a sub-domain capable of sensing a signal, and which is then able to transduce that signal to activate the DGC or PDE (Schirmer and Jenal 2009). The signals and sensing domains are diverse, including GAF domains (which bind small molecular

weight effectors), PAS domains (which bind heme or flavin molecules capable of binding oxygen) and response regulator domains (which can transduce a signal following phosphorylation) (Ryan et al. 2006). The diversity of signals capable of being sensed by each domain (such as PAS and GAF domains, which are near-ubiquitous in bacterial species (Galperin 2004)) prevent prediction of the environmental or physiological signal causing the activity of a specific DGC/PDE.

The mechanism of signalling involving c-di-GMP is complex and poorly understood. The c-di-GMP messenger mediates the response to specific environmental or physiological changes, such as growth of surfaces or the cell cycle, by regulating specific responses, such as decreased flagella synthesis or EPS production (Jenal and Malone 2006). Such specific regulation is believed to result from the effects of c-di-GMP on the level of transcription, translation and post-translation of the target effectors (Hengge 2009). In the case of the WS phenotype, the production of acetylated cellulosic polymer (ACP) is regulated by allosteric regulation of the cellulose biosynthetic machinery (encoded by *wssB* to *wssE*). However, transcription of the *wss* operon is also increased by DGC activity (Spiers et al. 2003) as well as other regulators of transcription (Giddens et al. 2007).

1.3.4 DGCs provide a model to assess the effect of architecture on variation

The multitude of effectors of c-di-GMP signalling is matched by a multitude of DGC and PDE encoding genes. Most bacteria have multiple DGC and PDE domains, with *E. coli* encoding approximately 19, *Vibrio cholera* 41 and *P. fluorescens* 33 (Galperin 2004). Recent studies have demonstrated a high degree of specificity of DGC/PDE c-di-GMP-producing enzymes for target effectors (Massie et al. 2012). This specificity is believed to be mediated by temporal and spatial sequestration of c-di-GMP (Hengge 2009).

The number of genes encoding GGDEF domains in SBW25 (with at least 24 genes encoding non-degenerate motifs) suggests many separate genes may be capable of synthesising c-di-GMP to levels capable of causing WS types. This provides a set of different architectures that may each regulate ACP production and the WS phenotype. By studying the architecture of each GGDEF gene, and determining each gene's

contribution to adaptive WS variants, we may gain insight to how architecture may bias the production of variation.

1.3.4.1 The architectures of known WS-causing DGCs

Prior to the work presented in this thesis, the only genes identified as capable of causing the WS phenotype were three DGC-encoding loci – *wsp*, *aws* and *mws*. Each locus was identified via iterative rounds of selection for WS types, identification of the causal mutational locus, and deletion of that locus (Spiers et al. 2002; Goymer et al. 2006; Bantinaki et al. 2007; McDonald et al. 2009). It was not until the final paper in the series, McDonald et al. (2009), that two pathways aside from *wsp* were identified that could readily cause WS from the ancestral SM type. Presented below are the molecular details of each pathway, and details of how known mutations result in DGC activity and the WS phenotype.

1.3.4.1.1 The *wsp* operon

The *wsp* (wrinkly spreader phenotype) operon is the most frequent target of mutation underpinning WS evolved from SBW25. The *wsp* operon is encoded by an 8.4 kb region consisting of seven genes termed *wspABCDEFR* (see Figure 1.2). The final gene in the operon – *wspR* – encodes a DGC, with naturally arising mutations leading to WS, resulting in increased c-di-GMP levels, presumably through increased activation and activity of WspR (Malone et al. 2007). Like all putative DGCs encoded by SBW25, the precise role of c-di-GMP produced by WspR in WT SBW25 remains unknown.

Figure 1.2: The organisation of the *wsp* operon. The *wsp* operon is composed of seven genes with the GGDEF domain (represented in grey**)** encoded by *wspR***.**

The predicted proteins encoded by the *wsp* operon share high homology with the *che* chemosensory pathway in *E. coli* (Bantinaki et al. 2007). The *che* pathway is one of the

best-characterised sensory systems (Falke et al. 1997). In *E. coli* and *Salmonella* species the *che* pathway regulates chemotaxis. In these species *che* encodes a sensory receptor for small molecules. Increased sensing of these small molecules is transduced by the rest of the *che* pathway, resulting in alterations to flagella activity and the rate of 'tumbling', which in turn alters the direction of cells towards these small molecules. A notable difference of *wsp* from the *che* pathway is the output effector (*wspR* and *cheY* respectively), both of which are activated via phosphorylation (Goymer et al. 2006). The *che* pathway does regulate c-di-GMP, but results in the phosphorylation of *cheY*, which diffuses through the cytoplasm and acts directly on flagella proteins.

The well-developed model of regulation of *che* has allowed for a similar model to be developed for the regulation of *wsp* (see Figure 1.3) (Bantinaki et al. 2007). The model predicts – in the WT case – that an unknown signal from WspA (located at the inner membrane of the cell) is transduced via the structural domains of WspB and WspD to WspE (the kinase), which activates WspR via phosphorylation. The level of phosphorylation by WspE is regulated by the antagonistic enzymes WspC and WspF. WspC acts as a constitutive methyltransferase, and as such, WspC adds methyl groups to WspA, resulting in constitutive wspA activity (and DGC activity). WspF is a methylesterase of WspA, with demethylation of wspA by WspF resulting in lower WspA activity and hence less phosphorylation of wspR. WspF causes negative feedback, as WspF is activated via phosphorylation of WspE. WspF hence has a unique place within the Wsp post-translational complex, with deactivation of WspF via mutation resulting in constitutive phosphorylation of WspR and high DGC activity.

As a consequence of this post-translational architecture, in particular the negative feedback of WspF, the *wsp* operon is the target of most known mutations identified as causal of the WS phenotype. Of a set of 26 mutations causal of the WS phenotype, 17 mutations are identified as occurring within the *wsp* operon (Bantinaki et al. 2007). The negative regulation within the Wsp pathway clearly biases mutations to occur within *wspF*, with 14 mutations isolated to *wspF*. A further three independent mutations were identified in *wspE* (though it is not known how these cause WspE kinase activity) (McDonald et al. 2009). This parallelism of mutation to *wspF* is suggestive that the negative regulation of WspF allows the predominance of loss-of-function mutations, which is the predominant effect of mutation (Kimura 1968), to bias the production of WS types to the *wspF* locus.

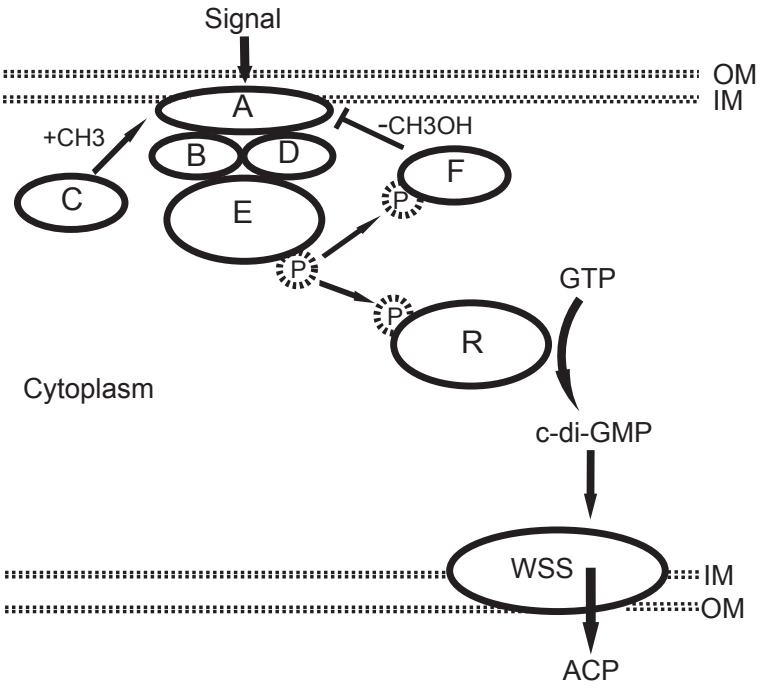

Figure 1.3: The proposed model for the WspA to WspR chemosensory pathway. In the ancestral case, the activation of WspA causes the kinase WspE to become active. WspE then phosphorylates WspF and WspR. Both WspF and WspR are then activated, causing WspF to negatively regulate WspA via demethylation, and WspR to produce c-di-GMP and activate ACP via allosteric activation of WSS. WspC acts to counter WspF by constitutive methylation of WspA, allowing WspA to become active. Loss-of-function of wspF results in a loss of repression of WspA and the WS phenotype. Figure adapted from Bantinaki et al. (2007). 'OM' and 'IM' represent the outer and inner membranes, and 'P' represents inorganic phosphate.

1.3.4.1.2 The *aws* operon

The *aws* (alternate wrinkly spreader) operon was the second operon identified as capable of mutating and causing WS types (McDonald et al. 2009). The operon consists of three genes termed *awsX*, *awsR* and *awsO,* encoded over ~2.3 kb (see Figure 1.4). The *aws* operon shares great homology in synteny, domain structure and predicted mode of action to the *yfiBNR* locus in *P. aeruginosa* PA01 (Malone et al. 2012). Mutations to this locus have been seen, in vitro, to underpin *P. aeruginosa* small-colony variants, which have been isolated from the lungs of cystic fibrosis patients. The gene *awsR* (homologous to *yfiN*) encodes two transmembrane helices, a HAMP domain (a common domain found in signal transduction proteins, which allow multimerisation of YfiN (Malone et al. 2010)), and a GGDEF domain (McDonald et al. 2009). The gene *awsO* encodes an outer membrane porin (OmpA), and *awsX* encodes

a protein of no known homology, except for a proteolitic cleavage site suggesting the encoded protein is periplasmic. Both *awsX* and *awsO* encode proteins predicted to regulate the DGC encoded by *awsR*.

Figure 1.4: The organisation of the *aws* operon. The *aws* operon is composed of three genes. The GGDEF domain (represented in grey) is encoded by *awsR*. Both *awsR* and *awsO* encode regions that allow binding to the inner and outer membranes respectively (represented by striped regions).

The model of WT regulation of AwsR DGC activity is presented in Figure 1.5. This model is based upon the putative function and mutations in the *aws* operon (McDonald et al. 2009), and homology of *awsXRO* to *yfiBNR*, the encoded protein-protein associations of which have been studied (Malone et al. 2010; Malone et al. 2012). These models describe *awsX* as encoding a post-translational negative regulator of AwsR activity. In the inactive state of AwsR, AwsX is predicted as being bound to the periplasmic region of AwsR. This either directly represses AwsR, prevents activation of AwsR by periplasmic signals, or prevents dimerisation of AwsR (DGCs are often active in dimers (Paul et al. 2007)). Upon activation of AwsO, situated in the outer membrane, by an unknown signal, AwsX translocates from AWSR to AWSO, thus relieving repression of AwsR. Based upon this model, it can be expected that mutations to the structure of either AwsX or AwsR that prevent association between these two proteins are capable of causing WS.

Of the 26 known mutations identified as causing WS types from SBW25, seven independent mutations were found in the *aws* operon (McDonald et al. 2009). As a demonstration of the above model, *awsX* was the target of two in-frame deletions and a nonsense mutation, and two mutations to *awsR* were both located at the predicted periplasmic region of *awsR*. These mutations, and the corresponding models of AwsR activation, demonstrate the constitutive negative regulation encoded at the *aws* locus.

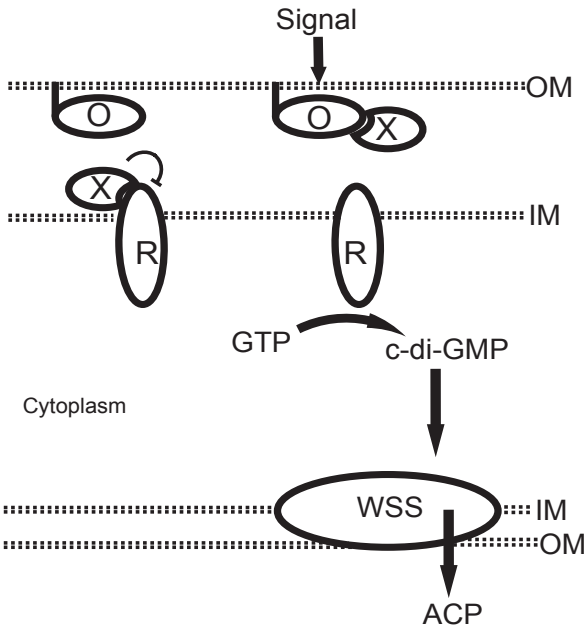

Figure 1.5: The predicted model for Aws pathway. In the inactive state, AwsR is repressed by binding of AwsX. An unknown signal activates AwsO, which causes sequestration on AwsX and a relief of the repressor. AwsR is then active – or is able to bind together in homodimers where it is active – resulting in the production of c-di-GMP, and activation of WSS enzymes. Mutations affecting the binding of AwsX to AwsR are capable of causing constitutive AwsR activity. This model is a synthesis of models from both McDonald et al. (2009) and Malone et al. (2012). 'OM' and 'IM' represent the outer and inner membranes.

1.3.4.1.3 The *mws* composite gene

The final gene known to cause WS, *mwsR,* is a single gene ~3.9 kb long (see Figure 1.6). This gene encodes two probable transmembrane helices, at least three distinct PAS domains, a GGDEF domain and an EAL domain. This gene thus represents a composite gene with two distinct functional domains regulating c-di-GMP levels.

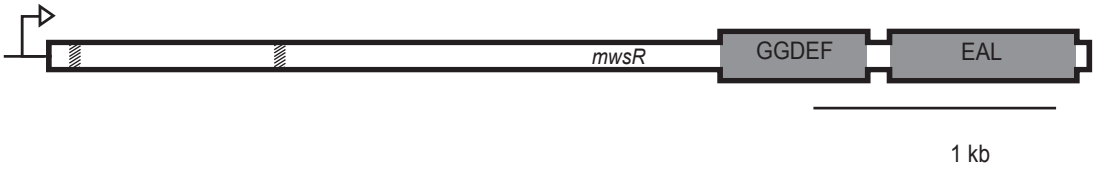

Figure 1.6: The organisation of the *mwsR* gene. The gene *mwsR* encodes both a GGDEF domain and a region of predicted negative regulation (the EAL domain). The GGDEF domain is transcribed prior to the EAL domain. Striped regions represent predicted transmembrane helices.

Little is known about this gene from direct molecular analysis. Several composite genes have been described; however, all appear to have lost function in either one of the two domains (Jenal and Malone 2006), with no composite genes previous recognised as encoding active GGDEF domains. Genetic constructs of the *mwsR* in SBW25 suggest both domains are functional. Deletion of the C-terminal region of *mwsR* encoding the EAL domain resulted in WS. Deletion of the rest of the gene reverted this WS to SM, indicating the WS phenotype was caused by the GGDEF domain of MwsR (McDonald et al. 2009).

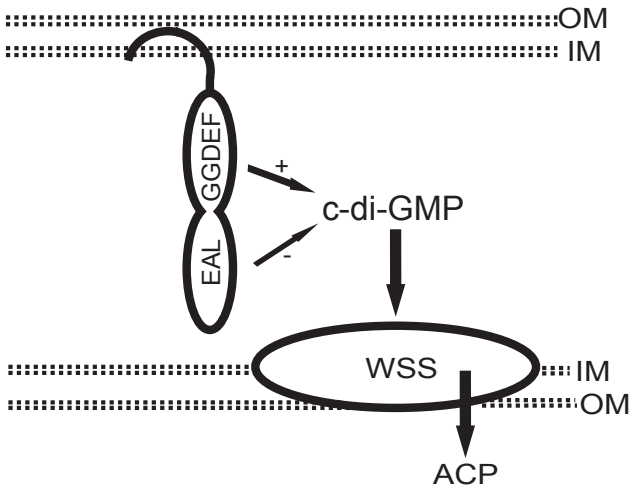

Figure 1.7: The predicted regulation of c-di-GMP by MwsR. The GGDEF and EAL domains are predicted to antagonistically regulate levels of c-di-GMP, which can then activate WSS enzymes. MwsR is predicted to be attached to the inner membrane at the N-terminal. 'OM' and 'IM' represent the outer and inner membranes.

The domains encoded by *mwsR* and genetic analysis of this gene (McDonald et al. 2009) allow for a model to be constructed that suggests MwsR is negatively regulated at the post-translational level (see Figure 1.7). This data suggests a model whereby the GGDEF domain encodes an active DGC, and the EAL domain encodes an active PDE, and both domains antagonistically determine c-di-GMP levels. Artificial mutations that deactivate the EAL domains, and hence reduce PDE activity, result in WS, presumably by increasing the net production of c-di-GMP. This model has been supported by two mutations causing WS (of the 26 identified mutations), which are located within or adjacent to the EAL domain. Taken together, this data indicates the negative regulation of the EAL domain within the architecture of MwsR.

1.3.4.2 Summary of *wsp*, *aws* and *mws* architecture and relative rate of evolution of the three pathways

Identification of the genetic routes underpinning WS types evolved from SBW25 has provided evidence that genetic architecture biases the relative production of variation from the different loci. The set of 26 mutants discussed above contains mutants of *wsp*, *aws* and *mws* of varying frequencies (with mutations to *wspF* accounting for half of the mutants), yet the fitness of mutants from all three loci are largely equivalent (Bantinaki et al. 2007; McDonald et al. 2009). This suggests that features of genetic architecture of each of the three pathways, relative to each other, may bias the production of variation.

An explanation for the relative frequency of mutation to the different pathways includes both target size and post-translational interactions of the negative regulators of each pathway. Deleterious mutation to the negative regulator of each pathway (such as *wspF*, *awsX* and the EAL domain of *mwsR*) has been shown as sufficient to cause WS. However, frequency of mutation appears to be biased towards mutation to *wspF.* An explanation for this bias is the target size of *wspF*. Of the three pathways, *wspF* encodes the largest gene encoding a negative regulator (1011 bp), compared to *awsX* (573 bp) (*awsX*) and the EAL domain of *mwsR* (726 bp). Differences in the target size of genes, assuming equality of other factors such as mutation rate and selection coefficient, has been shown to bias the comparative genetic variance of genes (Houle 1998; Chevin et al. 2010).

However, target size cannot entirely account for the bias of mutants to *wspF*. Further differences between the observed frequency of mutations at *wspF* to the other negative regulators have been explained by additional constraints (McDonald et al. 2009). These include an exclusion of frameshift mutations from *awsX* due to deleterious effects on the translation of downstream *awsR*. Similarly, in *mwsR*, frameshift mutations in the EAL domain may interfere with the activity of the encoded DGC. However, these biases may also be explained by small sample size.

1.3.5 Negative regulation in *wsp*, *aws* and *mws* may increase the production of variation

A further prediction made by McDonald et al. (2009) is that genetic architecture biases the production of variation between the common three routes and the rarer routes. The common feature of *wsp, aws* and *mws* is the inferred post-translational negative regulation encoded by these genes. This common architectural motif is believed to increase the production of variation at these three loci relative to any other loci that encode for DGCs. Negative regulatory architecture may increase variation because of the prevalence of loss-of-function mutations that can uniquely generate WS at these loci. This idea is supported by the number of loci potentially able to mutate to cause WS. There are approximately 29 DGC encoded in the *P. fluorescens* SBW25 genome, and yet mutations were found in only three such pathways (see Figure 1.8).

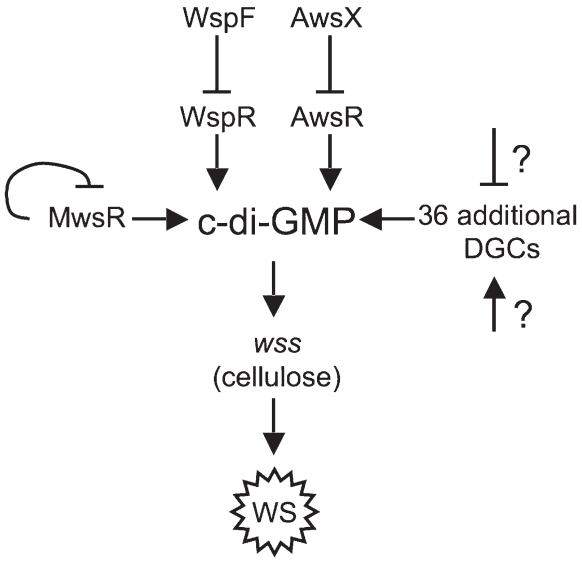

Figure 1.8: Multiple genetic pathways determine c-di-GMP levels and further paths may be identified. Genes *wspF*, *awsX* and the EAL domain of *mwsR* encode negative regulators of DGCs. However, additional pathways may also mutate to cause WS types, the regulation of which is unknown. Figure used with permission from McDonald et al. (2009).

Furthermore, McDonald and colleagues have demonstrated the genetic potential for SBW25 to evolve WS via rare pathways outside of *wsp, aws* and *mws*. By deleting the three common genes, the resulting SM type was inoculated in static microcosms and tested for the ability to evolve WS types via an alternative pathway. After 5 days – a

longer period of time than usually required – WS type evolved, termed SWS (slow wrinkly spreader) (McDonald et al. 2009). The evolution of SWS clearly demonstrates the potential for additional pathways, most likely encoding DGCs, to mutate to produce WS.

However, despite the discovery of SWS, it remains unclear whether the production of variation biases the relative production of common routes compared to rare routes. An attempt to determine the mutational basis of the WS phenotype in SWS was unsuccessful. Without the information on the mutation causing SWS, the mutation cannot be reconstructed in a directly comparable background – and thus the effect on fitness of the mutation cannot be assayed. Hence, the relative role of selection in biasing the production of variation remains unclear. Furthermore, the unknown architecture of the rare route underpinning SWS remains unknown. The architecture of this rare route requires investigation to directly demonstrate how genetic architecture biases the production of variation from common to rare routes.

The mutational basis of SWS, the effect on fitness of this mutation, and the architecture of this rare route are investigated in this thesis. The findings presented allow direct claims about the relative production of variation from common to rare routes to be made. Furthermore, the mutations and architectures of several additional routes to WS are identified. This study has serendipitously led to several remarkable findings concerning the evolution of new genes and causes of molecular parallelism in SBW25.

1.4 Research objectives

-Determine the locus responsible for the SWS type, reconstruct the causal mutation and determine the fitness of this mutant as compared to common WS types. Assess whether the fitness of SWS prevents the adaptive evolution of this mutant. Identify how the architecture at the locus underpinning SWS reduces the rate of production of WS (chapter 3).

-Utilise experimental evolution to identify multiple additional pathways to WS outside of *wsp*, *aws*, *mws* and the pathway underpinning SWS. Describe the architectures of these additional loci to explain why these routes rarely cause WS (chapter 4).

-Describe the architectures of the most common parallel mutants identified outside of DGC-encoding genes (from chapter 4). Investigate whether genetic architecture biases the production of these adaptive types (chapter 5).

Chapter 2:
Materials and methods

2.1 Materials

2.1.1 Media and growth conditions

Unless otherwise specified, all *Pseudomonas fluorescens* strains were cultured in King's B (KB) media (10 g glycerol, 20 g Proteose Peptone No.3 (for liquid media, BDH) or tryptone (for solid media, BDH), 1.5 g $MgSO_4.7H_2O$, 1.5 g $K_2PO_4.3H_2O$ L^{-1}. All *Escherichia coli* strains were cultured in Lysogeny Broth (LB) media (10 g tryptone, 5 g yeast extract, 10 g NaCl L^{-1}. To prepare plates of solid media, bacteriological agar was added to the media at 15 g L^{-1}. Where specified, *P. fluorescens* strains intended for fluorescent microscopy were cultured in M9 media (12.8 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 1 g NH_4Cl, 0.1 mM $CaCl_2.6H_2O$, 2 mM $MgSO_4.7H_2O$, 100 mL 20% ($^{w}/_{v}$) glucose L^{-1}). All bacterial overnight cultures were grown shaking at 160 rpm for approximately 16 h in glass microcosms containing 6 mL of media. *P. fluorescens* SBW25 strains were grown at 28°C, *E. coli* strains at 37°C. All bacteria were stored indefinitely at -80°C in 45% glycerol saline (70% ($^{V}/_{V}$) glycerol, 8.5 g NaCl L^{-1}).

2.1.2 Antibiotics, reagents and enzymes

Antibiotics were purchased from Melford Laboratories and used at the following concentrations: gentamycin (Gm) 25 μg mL^{-1}, kanamycin (Km) 100 μg mL^{-1}, tetracycline (Tc) 15 μg mL^{-1} (where the stock solution, 15 mg mL^{-1}, was made up in 1:1 ethanol:water solution), nitrofurantoin (NF) 100 μg mL^{-1} dissolved in dimethyl sulfoxide (DMSO) and cycloserine (800 μg mL^{-1}). X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) was used at a concentration of 60 mg L^{-1} in agar plates as an indicator of β-galactosidase activity. All restriction enzymes were purchased from Thermo Scientific and used with supplied reagents as per manufacturer's instructions.

2.1.3 DNA Electrophoresis

Electrophoresis gels were made with 1% ($^{w}/_{v}$) ultrapure agarose (Invitrogen), and TBE buffer (90 mM Tris-HCL pH 8.0, 0.55% ($^{w}/_{v}$) boric acid, 2 mM EDTA). Visualisation of DNA was enabled via addition of 1x SYBR® Safe DNA gel stain. Gels were run in TBE buffer. DNA was visualised using a UV transilluminator or a Safe Imager™ 2.0 Blue-Light Transilluminator (Invitrogen).

2.1.4 Competent cell production

The production of chemically competent *E. coli* cells for transformation required filter sterilised TFBI (100 mM RbCl, 50 mM $MnCl_2.4H_2O$, 30mM potassium acetate, 15 mM $CaCl_2.2H_2O$ L^{-1}, adjusted to pH 5.8) and TFBII (10 mM MOPS, 75 mM $CaCl_2.2H_2O$, 10 mM $RbCl_2$, 15 % ($^{v}/_{v}$) glycerol L^{-1}, adjusted to pH 7.0) buffer. The production of *P. fluorescens* electrocompetent cells required the preparation of filter sterilised 300 mM sucrose.

2.1.5 Photography and microscopy materials

Photography of microcosms was performed using a Canon EOS 400D camera. Colony-level photography was performed using a Canon Powershot A640 camera in conjunction with a Zeiss Axiostar Plus light microscope. In order to detect the production of acetylated cellulosic polymer (ACP), Calcofluor (Sigma, also called Fluorescent Brightener 28) was added to agar plates at a concentration of 35 μg mL^{-1}. Cell-level microscopy was performed using an Olympus BX61 upright fluorescence microscope and an F-View II monochrome camera as per Section 2.2.13.

2.1.6 Strain list

Strain	Genotype description	Reference
Pseudomonas fluorescens		
SBW25 (PBR340)	Ancestor of all evolved and genetically manipulated *P. fluorescens* strains in this study. A WT strain isolate with a SM morphology.	Rainey and Bailey 1996
SM Δ*wsp* Δ*aws* Δ*mws* (PBR716)	SBW25 ancestor with ORF of *wsp aws mws* deleted, used to evolve rare WS types	McDonald et al. 2009
LSWS	Ancestor WS evolved from SBW25 resulting from a mutation in *wspF* (A901C)	Spiers et al. 2002
LSWS *lacZ*	Ancestor WS evolved from SBW25 resulting from a mutation in *wspF* (A901C), containing a *lacZ* gene	Zhang and Rainey 2007a
IWSG	WS type evolved from SBW25 resulting in a mutation to *wspF* (C556T), measured by Bantinaki and colleagues as exhibiting the lowest fitness of the independent WS collection	Bantinaki et al. 2007
Original SWS	A WS type evolved from SM Δ*wsp* Δ*aws* Δ*mws,* selected in static microcosms over 5 days	McDonald et al. 2009
SWSTN(1-36)	The collection of 36 independent SM mutants resulting from transposon mutagenesis with IS-Ω-Km/hah of 'original SWS' (33 mutant genotypes are identified)	This study
PBR716 *fadA-fwsR*	SM Δ *wsp* Δ*aws* Δ*mws* genotype in which genes PFLU0184-PFLU0183 have been exchanged with the *fadA-fwsR* deletion mutation from 'original SWS' (using primer pairs BglII-PF184f and BglII-PF183R2)	This study
SBW25 *fadA-fwsR*	Genotype SBW25 genotype in which genes PFLU0184-PFLU0183 have been exchanged with the *fadA-fwsR* deletion mutation from 'original SWS' (using primer pairs BglII-PF184f and BglII-PF183R2)	This study
SBW25 *fadA-fwsR-lacZ*	Genotype SBW25 *fadA-fwsR* that contains a chromosome integrated mini-Tn7T-*lacZ* gene used for fitness assays	This study
ISWS 1-92	A collection of 92 independent WS types evolved from SM Δ *wsp* Δ*aws* Δ*mws*	This study
PBR716 ISWS32-*fadA-fwsR*	SM Δ *wsp* Δ*aws* Δ*mws* genotype in which genes PFLU0184-PFLU0183 have been exchanged with the *fadA-fwsR* fusion from 'ISWS32" (using primer pairs PF184f and PF183PROD-4R)	This study
SBW25 IWS32-*fadA-fwsR*	SBW25 genotype in which genes PFLU0184-PFLU0183 have been exchanged with the *fadA-fwsR* fusion from 'ISWS32" (using primer pairs PF184f and PF183PROD-4R)	This study
PBR716 *fadA-fwsR* EE255-256AA	SM Δ*wsp* Δ*aws* Δ*mws* genotype featuring the reconstructed *fadA-fwsR* chimera, in which the GGEEF residue motif is replaced with GGAAF (using primer pairs DGCF1 and GGAAF-R, GGAAF-F and DGCR4)	This study

Strain	Genotype description	Reference
SBW25 *fadA-fwsR* EE255-256AA	SBW25 genotype featuring the reconstructed *fadA-fwsR*, in which the GGEEF residue motif is replaced with GGAAF (using primer pairs DGCF1 and GGAAF-R, GGAAF-F and DGCR4)	This study
PBR716 *fadA-fwsR* F257L	SM Δ*wsp* Δ*aws* Δ*mws* genotype featuring the reconstructed *fadA-fwsR* chimera, in which the GGEEF residue motif is replaced with GGEEL (using primer pairs DGCF1 and DGCR2, DGCF3 and DGCR4)	This study
SBW25 *fadA-fwsR* F257L	SBW25 genotype featuring the reconstructed *fadA-fwsR* chimera, in which the GGEEF residue motif is replaced with GGEEL (using primer pairs DGCF1 and DGCR2, DGCF3 and DGCR4)	This study
PBR716 *PfadA-fwsR*	SM Δ*wsp* Δ*aws* Δ*mws* featuring the transcriptional fusion of the promoter of *fad* to the ORF of *fwsR*, produced via allelic exchange (using primer pairs PF184ORFD1F and PF184ORFD2R, PF184ORFD3F and PF184ORFD4R)	This study
SBW25 *PfadA-fwsR*	SBW25 featuring the transcriptional fusion of the promoter of *fadA* to the ORF of *fwsR*, produced via allelic exchange (using primer pairs PF184ORFD1F and PF184ORFD2R, PF184ORFD3F and PF184ORFD4R)	This study
PBR716 *fadA*-3X-*fwsR*	SM Δ*wsp* Δ*aws* Δ*mws* featuring a transcriptional fusion of *fadA* and *fwsR*, separated by three stop codons in each reading frame (using primer pairs PF184f and NEW183PROD2R, NEW183PROD3F and PF183PROD-4R)	This study
SBW25 *fad*A-3X-*fwsR*	SBW25 featuring the a transcriptional fusion of WT *fadA* and *fwsR*, separated by three stop codons in each reading frame (using primer pairs PF184f and NEW183PROD2R, NEW183PROD3F and PF183PROD-4R)	This study
PBR716 *fad*A-1X-*fwsR*	SM Δ*wsp* Δ*aws* Δ*mws* featuring a transcriptional fusion of *fadA* and *fwsR*, separated by one stop codon in each reading frame (using primer pairs PF184f and PF183PROD-2R, PF183PROD-3F and PF183PROD-4R)	This study
SBW25 *fad*A-1X-*fwsR*	SBW25 featuring a transcriptional fusion of WT *fadA* and *fwsR*, separated by one stop codon in each reading frame (using primer pairs PF184f and PF183PROD-2R, PF183PROD-3F and PF183PROD-4R)	This study
PBR716 2w2	SM Δ*wsp* Δ*aws* Δ*mws* in which genes PFLU0184-PFLU0183 have been exchanged with the *fadA-fwsR* fusion from 'SREE strain 2w2' (using primer pairs P184BGLF and PF183PROD-2R)	This study
SBW25 2w2	SBW25 in which genes PFLU0184-PFLU0183 have been exchanged with the *fadA-fwsR* deletion mutation from 'SREE strain 2w2' (using primer pairs P184BGLF and PF183PROD-2R)	This study
SMΔ*mws* (PBR711)	2w2' SBW25 featuring a knockout of *mwsR* produced via allelic exchange	McDonald et al. 2009
SM Δ*mws* P_{fad}-*mwsr1218-fwsR*	SM Δ*mws* featuring an allelic replacement of *fadA* with the initial translated 1218 bp of *mwsR* creating a translation fusion to WT *fwsR* (using primer pairs TMD_F1 and TMD_MWB_R2, TMD_MWB_F3 and TMD_MWA_R4, TMD_MWB_F5 and TMD_R6)	This study

Strain	Genotype description	Reference
SBW25 mini-Tn7T-LAC-GFP(-)	SBW25 featuring a chromosome integrated mini-Tn7T-LAC plasmid. The negative control for GFP fluorescence	This study
SBW25 mini-Tn7T-LAC-GFP(+)	SBW25 featuring a chromosome integrated mini-Tn7T-LAC plasmid expressing *gfp.* The positive control for GFP fluorescence (using primer pairs GFPmut3_Tn7LAC and GFP_KpnI_R)	This study
SBW25 mini-Tn7T-LAC-*fadA-gfp*	SBW25 featuring a chromosome integrated mini-Tn7T-LAC plasmid expressing a transcriptional fusion of *fadA* to *gfp* (using primer pairs 184_FP_SPEIRBSF and 184_FP_SBFI-R, GFP_XhoI_F and GFP_KpnI_R)	This study
SBW25 mini-Tn7T-LAC-*fwsR-gfp*	SBW25 featuring a chromosome integrated mini-Tn7T-LAC plasmid expressing a transcriptional fusion of WT *fwsR* to *gfp* (using primer pairs 183_FP_SPEIRBSF and 183_FP_SBFI_R, GFP_XhoI_F and GFP_KpnI_R)	This study
SBW25 mini-Tn7T-LAC-*fwsR-gfp*	SBW25 featuring a chromosome integrated mini-Tn7T-LAC plasmid expressing a transcriptional fusion of *fadA-fwsR* to *gfp* (using primer pairs 184_FP_SPEIRBSF and 183_FP_SBFI_R, GFP_XhoI_F and GFP_KpnI_R)	This study
SBW25 mini-Tn7T-LAC-*fadA::3X-fwsR-gfp*	SBW25 featuring a chromosome integrated mini-Tn7T-LAC plasmid expressing a transcriptional fusion of *fadA-3X-fwsR* (see above) to *gfp* (using primer pairs 184_FP_SPEIRBSF and 183_FP_SBFI_R, GFP_XhoI_F and GFP_KpnI_R)	This study
SBW25 mini-Tn7T-LAC- fws_{2w2}-*gfp*	SBW25 featuring a chromosome integrated mini-Tn7T-LAC plasmid expressing a transcriptional fusion of *fadA-fwsR* (isolated from strain '2w2') to *gfp* (using primer pairs 184_FP_SPEIRBSF and 183_FP_SBFI_R, GFP_XhoI_F and GFP_KpnI_R)	This study
SREE mutants L1-M1 – L8-M8	64 mutants evolved from an SM Δ*wsp* Δ*aws* Δ*mws* ancestor across 8 lines, each isolate adapted to a given alternating environment	This study
SBW25 *nlpD* (C565T)	SBW25 featuring a nonsense mutation to *nlpD* (C565T)	Dr C Kost
SBW25 Δ*wss*	SBW25 featuring a knockout of the *wss* operon (*wss a-j*), using a pUIC3 plasmid gifted by Dr J. Gallie	This study
SBW25 Δ*wss nlpD* (C565T)	SBW25 Δ*wss* featuring the allelic exchange of WT *nlpD* with a non-synonymous C565T mutation (using primer pairs PF1302_bglII_F and PF1302_bglII_R)	This study
SBW25 pUIC3-*nlpD*(WT)	SBW25 featuring pUIC3 recombined between *nlpD* and *rpoS,* resulting in two WT *nlpD* regions (using primer pairs PF1302_bglII_F and PF1302_bglII_R)	This study
SBW25 pUIC3-3'*nlpD* (C565T)	SBW25 featuring pUIC3 recombined between *nlpD* and *rpoS,* resulting in WT *nlpD* 5' of pUIC3, and mutant *nlpD* (C565T) 3' of pUIC3 (using primer pairs PF1302_bglII_F and PF1302_bglII_R)	This study
SBW25 pUIC3-5'*nlpD* (C565T)	SBW25 featuring pUIC3 recombined between *nlpD* and *rpoS,* resulting in mutant *nlpD* (C565T) 5' of pUIC3, and WT *nlpD* 3' of pUIC3 (using primer pairs PF1302_bglII_F and PF1302_bglII_R)	This study

Strain	Genotype description	Reference
SBW25 *nlpD* (C565T)-pUIC3 (C565T)	SBW25 featuring pUIC3 recombined between *nlpD* and *rpoS,* resulting in mutant *nlpD* (C565T) 5' of pUIC3, and mutant *nlpD* (C565T) 3' of pUIC3 (using primer pairs PF1302_bglII_F and PF1302_bglII_R)	This study
SBW25 *nlpD* (C565T A566G)	SBW25 featuring a knockout of the *wss* operon (*wssa-j*), and the allelic exchange of WT *nlpD* with a non-synonymous C565T and A566G mutation (using primer pairs PF1302_bglII_F and PF1302_bglII_R)	This study
SBW25 *nlpD* C565T A566G (1 -38)	33 *nlpD* chaining mutants evolved from SBW25 featuring a knockout of the *wss* operon (*wssa-j*), and the allelic exchange of WT *nlpD* with a non-synonymous A566G mutation (using primer pairs PF1302_bglII_F and PF1302_bglII_R)	This study
Escherichia coli		
TOP10	F' , mcrA , Δ (mrr-*hsd* RMS-*mcr*BC), Φ 80*lac*ZΔ M15, Δ*lac* X74, *deoR* , *recA1*, *ara*D139, Δ(*ara-leu*) 7697, *gal*U, *gal*K, *rps*L, Str^R , *endA*1, *nup*G	Invitrogen
DH5α (λpir)	*sup*E, Δ*lac*U169 (φ 80 *lac* ZΔM15) , *hsd*R , *rec*A , *end*A , *gyrA*, *thi*, *relA* (oriR6K replication)	Invitrogen

Table 2.1: Strains used, evolved or constructed. WT=wild type, SM=smooth, ORF=open reading frame, IWSG=independent wrinkly spreader G, SWS=slow wrinkly spreader, SREE=slow reverse evolution experiment, GFP=green fluorescent protein.

2.1.7 Transposon and plasmids

Name	Characteristics	Reference
Transposon		
IS-Ω-Km/hah	ColE1, *nptII* promoter, IE, OE, *LoxP, km*^{R}	Giddens et al. 2007
Plasmids		
pUIC3	Used for allelic exchange; TcR , mob , oriR6K , bla , Δ promoter-lacZY	Rainey 1999
pCR8/GM/TOPO	High copy number cloning vector; SpeR , pUC ori	Invitrogen
pRK2013	Helper plasmid for tri-parental mating; KmR , incP4, tra, mob.	Figurski and Helinski 1979
pUC18-mini-Tn7T-LAC	Chromosome integration vector with inducible P_{tac} promoter; Amp^R, Gm^R, *Ptac* expression cassette, mini-Tn7T vector	Choi and Schweizer 2006
pUX-BF13	Helper plasmid to allow Tmn7T recombination at the attB site; *mob*, Amp^R	Bao et al. 1991

Table 2.2: Name and characteristics of plasmids and transposon used.

2.1.8 Primers

All primers were manufactured by Integrated DNA Technologies. Primers were resuspended in deionised water to a stock solution of 100 μM, and diluted to a working concentration of 10 μM. Primers were stored at -20°C prior to use and thawed on ice.

Name	Sequence (5' to 3')	Target region
Primers used for transposon mutagenesis		
Tn*pho*A-II	GTGCAGTAATATCGCCCTGAGCA	IS-Ω-Km/hah
CEKG 2A	GGCCACGCGTCGACTAGTACNNNNNNNNNNAGAG	Non-specific
CEKG 2B	GGCCACGCGTCGACTAGTACNNNNNNNNNNACGCC	Non-specific
CEKG 2C	GGCCACGCGTCGACTAGTACNNNNNNNNNNGATAT	Non-specific
Hah-1	ATCCCCCTGGATGGAAAACGG	IS-Ω-Km/hah
CEKG 4	GGCCACGCGTCGACTAGTAC	5' end of CEKG 2A, B, & C
Primers used for generic plasmid cloning		
MLG021-PF	CAGGGTTATTGTCTCATGAGC	pUIC3
MLG022-PR	TGGGATTAACTGCGCGTCGCC	pUIC3
MLG056-PF	GTAAAACGACGGCCAG	pCR8
MLG057-PR	CAGGAAACAGCTATGAC	pCR8
SBW25-*glms*	CACCAAAGCTTTCACCACCCAA	*glms*
Tn7R109	CAGCATAACTGGACTGATTTCAG	Mini-tn7
Primers used in the reconstruction and genetic analysis of the fadA-fwsR mutation		
PF184f	CCCACAAAGGCTGCGGATTATAG	5' of PFLU184
Pf184r	CGGTGGAATTGGTGACGAACG	PFLU184
PF183F	CAGATCATGGCTCAGTACCGC	PFLU184
PF183r	CTGTGTTCGGCGTAGTCGAG	PFLU183
PF183F2	CTTACGAGCGGGTGTTCGG	PFLU183
PF183R2	GATCAGTCGACCACTAGCCG	3' of PFLU183
BglII PF184f	AGATCTCCCACAAAGGCTGCGGATTATAG	5' of PFLU184
BglII PF183R2	AGATCTGATCAGTCGACCACTAGCCG	3' of PFLU183
PF184x1	CTGAGTGCCACCCACACATAG	PFLU184
PF183R2x1	GGCGATTTATTCGTTGGAGGGTG	3' of PFLU183
PF184ORFD1F	AGATCTGGTCCAGATCGAGCAACAGC	PFLU185
PF184ORFD2R	GTGAACCATCGGCATAGGCTGTATGGCCCTCAG	5' of PFLU184
PF184ORFD3F	ATACAGCCTATGCCGATGGTTCACGAAAAGTCCTC	PFLU183
PF184ORFD4R	AGATCTGATCAGTCGACCACTAGCCG	3' of PFLU183
PF184ORFxF	CAGTGCCTTGAGCCGTTCAC	PFLU185
PF183PROD-2R	CAACTCTCTCTGTCATATGGATGCAGGGCGC	PFLU184
PF183PROD-3F	GCATCCATATGACAGAGAGAGTTGTGCCGATGG	PFLU183

Name	Sequence (5' to 3')	Target region
NEW183PROD2R	TGGTCACTCACTCATATGGATGCAGGGCGC	PFLU184
NEW183PROD3F	ATGAGTGAGTGACCAGAGAGAGTTGTGCCGATGG	5' of PFLU183
PF183PROD-4R	ACTAGTGCTGATGGTCACGCTGAAGG	PFLU183
DGCF1	AGATCTAAGTCGAGCGATTGAGCGAGC	PFLU183
DGCR2	GACAGCGGCGAGCTCTTCGCCAC	PFLU183
DGCF3	GTGGCGAAGAGCTCGCCGCTGTC	PFLU183
DGCR4	ACTAGTGCACAGCATGAACCTGATCCTCG	PFLU182
GGAAF-F	CATTGGTGGCGCAGCGTTCGCCG	PFLU183
GGAAF-R	CGGCGAACGCTGCGCCACCAATG	PFLU183
P184BGLF	AGATCTAATCGAACACGGCAGCGGC	PFLU185
P184BGLR	AGATCTGTGCAGATAGACCGTGACCGC	PFLU184
END183F	AGATCTCAACTTGAGCTGGAACTGCACCG	PFLU183
END183BglIIR	AGATCTCTAGCCGGCGATGACGCG	3' end of PFLU183
TMD_F1	AGATCTCACGGCTGTAACCGGTGATTTCC	PFLU185
TMD_MWB_R2	CGTGACTTTGGACATAGGCTGTATGGCCCTCAG	5' end of PFLU184
TMD_MWB_F3	ATACAGCCTATGTCCAAAGTCACGCCGCCAACTC	5' end of *mwsR*
TMD_MWA_R4	GTGAACCATCGGCTTGAGCTGGATTTCCAGCTCCAAC	mwsR
TMD_MWB_F5	ATCCAGCTCAAGCCGATGGTTCACGAAAAGTCCTC	5' end of PFLU183
TMD_R6	ACTAGTGTTGCTGCTCTGGGTCAGCAC	PFLU183
GFPmut3_Tn7LAC	GACTCGAGGAGGAGAATTAAGCATGCGTAAAGGAGAAG AACTTTTC	5' end of *gfp*
GFP_KpnI_R	GGTACCTTATTTGTATAGTTCATCCATGCC	3' end of *gfp*
GFP_XhoI_F	CTCGAGATGCGTAAAGGAGAAGAACTTTTC	5' end of *gfp*
184_FP_SPEIRBSF	ACTAGTGAGGAATACAGCCTATGTGGTACAACG	5' end of PFLU184
184_FP_SBFI-R	CCTGCAGGGATATGGATGCAGGGCGCAGC	3' end of PFLU184
183_FP_SPEIRBSF	ACTAGTGAGGAGTTGTGCCGATGGTTCACG	5' end of PFLU183
183_FP_SBFI_R	CCTGCAGGGAGCCGGCGATGACGCGGTTTTTG	3' end of PFLU183
Primers used in the confirmation and temporal ordering of mutations in the SREE		
PFLU0085_1010_F	CCTACAGCTCCGTAGCCAAG	PFLU0085
PFLU0085_D_R	CTCAAGGTTGGTGGAGGTGT	PFLU0085
PFLU0085_1730_R	ATCCAAGTCCAGCATCGCC	PFLU0085
PFLU0301_632F	AGGTCGGCGTGAACTACATC	PFLU0301
PFLU0301_1629R	CAACTGGTGGATACCGAAGC	PFLU0301
wss21Pf	TGCTGTACGCCAACGTGCTGC	wssE (PFLU0304)
PFLU0304_3060	TGGGCTTCCTCTACAACACC	PFLU0304
PFLU0304_DR	CCTTGCTTGGTGAGGTTGTC	PFLU0304
3-13f	CACCGACAACCTGACCAAC	PFLU0458
3-13r	CTTTTCCGGATGGTTCCAGC	PFLU0458
PFLU0476_F	ATGCTGAACCGCTTCCAGGC	PFLU0476
PFLU0476_R	CGATACACCAACTGCACGGC	PFLU0476
PFLU0477_F	AGCGTTACCTCAACCTCACCG	PFLU0477
PFLU0477_R	CACTCAAGCCCCTGCAACG	PFLU0477
PFLU0478_F	TTACCTGATCGGCCTCAACTGC	PFLU0478

Name	Sequence (5' to 3')	Target region
PFLU0478_R	CGATATTCGGGTACGTCTGGCC	PFLU0478
PF0621_400_F	CGATCTACGTCATGCTGGCC	PFLU0621
PF0621_850_R	ATCGGTTCGATCAGCGTGC	PFLU0621
PFLU0956_U_F	ATGAGTACCGGGTGATCGTC	PFLU0956
PFLU0956_870_F	GTGTTCGCTCCCTGGAAACGC	PFLU0956
PFLU0956_250_R	TTGGTGGTTGAGCGGGTAGC	PFLU0956
PFLU0956_D_R	GAGATAAAAGTCGCCAGGCA	PFLU0956
1s4Deletion_F	TGGCACTCATCGCGTGTATGC	PFLU1732
1s4Deletion_R	AGTTTCCGGTGGGGTATGACC	PFLU1756
PF1302_bglII_F	AGAGATCTTACCGGATCTTCGCCCTCGTC	PFLU1301
PF1302_bglII_R	AGAGATCTCAGCAGGTCCCTCATTCGAC	PFLU1302
PFLU1467_U_F	GTGTGTGCCAAATCAGTGCG	PFLU1467
PFLU1467_400_R	TGAACGGTGCCTTCGATCTCG	PFLU1467
PFLU1468_F	CATGAGTCGTGATGCCCTGC	PFLU1468
PFLU1468_R	GCAAAAGCGTAAGGAGCGG	PFLU1468
PFLU1739_2300_F	CCATCGAAGTGGCGTACAACTC	PFLU1739
PFLU1739_2900_R	ACCAGCACATTGAGTACGCGC	PFLU1739
PFLU2441_150_F	CGGGTGTTTCATGCAAGCGG	PFLU2441
PFLU2441_600_R	GTCGAGGTCTTCCACACTGTCC	PFLU2441
PFLU2485_U_F	GCAAGCCAAATCCCCACG	PFLU2485
PFLU2753_F	CGAGAAGTCCCATGCTCATCGG	PFLU2753
PFLU2753_650_R	CCAGCACCAGATTATCCAGGC	PFLU2753
PFLU2764_100U_F	GTCACTGGCGTTGAATCTGC	PFLU2764
PFLU2764_900_R	GATCACCCCACCAATCTGGC	PFLU2764
PF3016U_F	GCGCTTACTCAAGGAAGAGGCC	PFLU3016
PF3016D_R	GCCATACAGGTAGGCTTCCTTGGC	PFLU3016
PF3018_U_F	GCCAAGCCGATTGATCTGGATCG	PFLU3018
PF3018_600_R	TGAGAATACGCTATCGGTCGCC	PFLU3018
PFLU3409_U_F	GACGATGTTGTTCTGCGACAGGC	PFLU3409
PFLU3409_R	GGCCTTGTTTGAGCGCACC	PFLU3409
PFLU3677_F	CAGCGATCCCGTTCAGTGC	PFLU3677
PFLU3677_R	GTCAATTCCACCTCCCGCACC	PFLU3677
PFLU4287_180_F	AGATGGTCACCGCGTATGCC	PFLU4287
PFLU4287_900_R	AGGTACACCACTTCAGGCACG	PFLU4287
PFLU4414_560_F	CGACCAGCGAAGACATTACCG	PFLU4414
PFLU4414_1330_R	TTCACCCACCAGTTCCAGG	PFLU4414
PFU4417_400_F	TGAACGATACCTTGGGCAGC	PFLU4417
PFLU4417_300_DR	CCAGAAAGTCGATGCTCCCG	PFLU4417
PFLU4744_UF	CTGTGGTGGCCTGGACTTAC	PFLU4744
PFLU4744_DR	AGCTTGCTCGCTAAAGAAGC	PFLU4744
PFLU4782_250_F	GCGACGCTGCAATTTACACC	PFLU4782
PFLU4782_443R	CAGTGCGTCATACAGCAGGT	PFLU4782
PFLU4782_900_R	CTGTGGGTTGCGGTAATTGC	PFLU4782
PFLU4858_350F	GGAACGCACCTTTACCACGC	PFU4858
PFLU4858_530F	GTCTATCACTGGCTGCTGACC	PFLU4858

Name	Sequence (5' to 3')	Target region
PFLU4858_1070_R	TTCCAGGTCGTCGAGGATGC	PFLU4858
PFLU4939_340U_F	CAGAACTTCGCAAAGGCCGC	PFLU4939
PF4939_40U_F	GCCACTATCCCTGTCAATCAGCG	PFLU4939
PFLU4939_150_R	AGCAGCGCGATGATGTCACG	PFLU4939
PF4939_D_R	GCGGCTCACAGAATAGACAGGC	PFLU4939
PF5034_660_F	CGAGCACTGCCGTCACAAGATC	PFLU5034
PF5034_1210_R	GTGGTAACCGCGCACTTCATC	PFLU5034
PFLU5698_400_UF	TCCTTGAAGCCGCTCATCAGC	PFLU5698
PFLU5698_270_R	GGCACTGAGGAAGATGATCGG	PFLU5698
PFLU5698_950_F	TCTACCTGGACCTTGACGGC	PFLU5698
PFLU5698_1570_R	CCTGATCGAAGGTGGCTTGC	PFLU5698
PF5748_3300_F	GATCGTCGCTACAGCTACTCTCC	PFLU5748
PF5748_4000_R	TGACCAAGTCGTCGAGCAGC	PFLU5748
PFLU6004_100_F	TACGGAAGTCGCGGATCTACC	PFLU6004
PFLU6004_750_R	CACGATGGCATCTTCCACC	PFLU6004
Primers used in the genetic analysis of mutations in nlpD		
PF1302_bglII_F	AGAGATCTTACCGGATCTTCGCCCTCGTC	PFLU1301
PF1302_bglII_R	AGAGATCTCAGCAGGTCCCTCATTCGAC	3' of PFLU1302
nlpDmutF	ACTACAAGGCGCTTGCAGCTC	PFLU1301
PF1301_BGLII_F	AGAGATCTGGAGCGCATCAAGGTGTTGC	5' PFLU1301
PF1301_BGLII_R	AGAGATCTCACGGAACTGAAGACGAGGC	PFLU1301
PF1301_80F	TGTTGGTAGGGTGCTCAAGC	PFLU1301
PF1301_180R	ACATACTGCCCGGTGGTTAC	PFLU1301
PF1301_670F	TACGTCAGTGCTTACGGTCAC	PFLU1301
PF1301_7600R	TGTACCCGTTGACCCCATTTC	PFLU1301
PF1302_110F	ACCTTCAGTTCGCGCCAAATC	PFLU1302
PF1302_210R	ATTGAGGTACAGCTGGGTCG	PFLU1302

Table 2.3: Primer names, sequences and intended targets of amplification. SREE=slow reverse evolution experiment.

2.2 Methods

2.2.1 Plasmid purifications for electroporation and ligation

Plasmid DNA was extracted from overnight cultures using an E.Z.N.A.® Plasmid Mini Kit 1 (Omega Bio-Tek), and stored in TE elution buffer at -20°C. PCR products used in cloning were isolated via gel extraction using a QIAquick® Gel Extraction Kit (Qiagen). Purified PCR products were cloned into pCR8/GW/TOPO (Invitrogen) using the TOPO® TA cloning kit (Invitrogen). Ligation of restriction enzyme digested fragments was performed in a reaction involving Quick ligase (NEB) as per the manufacturer's instructions.

2.2.2 Purification of PCR products

All PCR products intended for sequencing were purified with an enzymatic reaction. To do this, 50 μL PCR products were enzymatically treated by addition of 4 units of Exonuclease I (NEB) and 4 units of Calf Intestinal Phosphatase (NEB) and incubated at 37°C for 30 min followed by 85°C for 15 min. Products were then stored at 4°C prior to sequencing.

2.2.3 Triparental conjugation

Cultures of the *P. fluorescens* recipient, the *E. coli* donor strain and the helper strain (*E. coli* pRK2013) were grown overnight in shaken (160 rpm 18 h) LB (with appropriate antibiotics). Aliquots of the recipient strain were heat shocked, with 1 mL of culture incubated at 42°C for 20 min. Aliquots of 300 μL of donor and recipient culture was pelleted (13,000 x g, 2 min), and resuspended together in 250 μL LB. The heat-shocked recipient culture was then pelleted (13,000 x g, 1 min), and resuspended in the donor/helper mix. The mixed culture was then pelleted (13,000 x g, 1 min), and resuspended in 50 μL LB. The resulting concentrated suspension of cells was spread onto an LBA plate (in a spot 20 mm in diameter), and incubated at 28°C for 18 h. The cells were resuspended in 3 mL of Ringer's solution and a dilution series ranging from

10^0 to 10^{-2} was plated on KBA plates containing kanamycin or tetracycline (to select for the transposon insertion or pUIC3 recombination into the recipient genome), and nitrofurantoin (to select against the *E. coli* donor).

2.2.4 Transposon mutagenesis

The IS-Ω-km/hah transposon (located on plasmid pSCR001) (Giddens et al. 2007) was used to randomly mutagenise wrinkly spreader mutants and identify genes associated with the WS phenotype. Triparental conjugation was used to generate mutants of the WS, using the *P. fluorescens* recipient, the donor strain (*E. coli* S17-1 λ_{pir} pSCR001) and the helper strain (*E. coli* pRK2013). Of the resulting mutagenised colonies from each conjugation, 2000-3000 were screened visually for qualitative morphological similarity to the SM ancestor. Mutants with SM-like morphology were then streaked on KB plates and used to make overnight cultures for preservation. The site of the transposon insertion was then precisely determined by arbitrary-primed polymerase chain reaction (PCR) (Manoil 2000), as described in Section 2.2.5, and the insertion sites were mapped in Artemis (Rutherford et al. 2000).

2.2.5 Polymerase chain reaction (PCR)

All standard reactions for electrophoresis contained 1 unit of Taq polymerase, 1 μL of 10 μM primer, 2.5 μL 10× Thempol® Buffer (NEB), 0.5 μL dimethyl sulfoxide (DMSO), 5 nM dNTP mix, and were then adjusted to 25 μL with RNase and DNAse free water. Cells taken directly from colonies were used as a template for the reaction. Cycling was as follows: denaturation was performed for 4 min at 95°C, followed by 35 cycles of amplification, involving denaturation at 95°C for 15 s, annealing at approximately 56°C to 60°C (depending on primer sequence) and elongation at 72°C for 60 s per kilobase (kb) of target DNA amplified.

To amplify the chromosome-transposon junction of IS-Ω-Km/hah insertions, arbitrary-primed PCR (AP-PCR (Manoil 2000)) was used to generate amplicons of the junction to allow sequencing and identity of the insertion site. This involved a round of PCR involving the same concentration of reagents as in standard PCR; however, to ensure consistent levels of template, 3 μL of template cells was added to the reaction from a 1

mL overnight culture, pelleted (13,000 x g for 1 min) and resuspended in 400 μL deionised water. Cycling conditions were as follows: an initial denaturation was performed at 95°C for 10 min, followed by five cycles of amplification involving denaturation at 95°C for 30 s, annealing at 42°C for 30 s, extension at 72°C for 3 min. With each of these five amplifications, the annealing temperature was decreased by 1°C per cycle. This was followed by 25 more cycles with an increased annealing temperature of 65°C. A second round of PCR (25 μL) was performed using nested primers CEKG-4 and Hah-1 (see Table 2.3). Reagents for this PCR round were at the standard concentration, using 2 μL of the PCR product from the first round diluted 10 times using deionised water and 2 μL of each 10 μM primer stock. The initial denaturation was performed at 95°C for 10 min, with 30 cycles of amplification, each involving 95°C for 30 s, 65°C for 30 s and 72°C for 3 min. The sample was then stored at 4°C prior to purification.

2.2.5.1 Strand overlap extension PCR

Strand overlap extension (SOE-PCR) (Ho et al. 1989) was employed to generate all site-directed genomic mutations and deletions. This technique involved the production of amplicons of either side of the target locus (for mutation or deletion). In order to minimise mutations in these products, Phusion® High-Fidelity DNA polymerase (New England Biolabs (NEB)) was used in PCR reactions involving 10 μL of 5×reaction buffer, 1 μL 10 mM dNTP, 5 μL primers 10 mM, 0.5 μL Phusion® High-Fidelity DNA polymerase, 2 μL DMSO and 2 μL of a colony, grown for 48 h and mixed in 100 μL deionised water. The reaction was conducted in a 50 μL volume with deionised water. In the case of deletions, the primers used to amplify these regions contained regions (minimum 15bp) at the 5' region of the primers that were complementary to the 5' region of primers used in PCR of the other amplicon. In the case of site-directed mutagenesis, a longer complementary region was used that contained the required mutation sequence. Cycling conditions were standard for Phusion® High-Fidelity DNA polymerase – an initial denaturation was performed at 98°C for 30 s, with 35 cycles of amplification, each involving 98°C for 10 s, 65°C for 10 s and 72°C for 30 s per kb of template DNA. This was followed by a final elongation step of 72°C for 5 min. The resulting PCR product was then verified for size using gel electrophoresis.

A second round of PCR was then performed. This involved two reaction stages. The first stage used 20 ng μL^{-1} of the two PCR products from the first round as a template. The concentrations of reagents were maintained in this reaction, except for the template, and initially both primers were omitted from the reaction, which was adjusted to 40 μL with deionised water. The reaction was then subjected to 10 cycles to ensure annealing of the complementary regions and subsequent elongation of the fragments. This was performed at the temperatures standard for High-Fidelity polymerase, except the annealing temperature was 50°C. Primers were then added to the reaction (5 μL primers 10 mM), and the reaction was run as per standard conditions for 25 cycles. The resulting PCR product was verified for size using gel electrophoresis. Where the SOE-PCR products were to be used for cloning into pCR8/GW/TOPO, the products were gel extracted, and an extra round of elongation were performed, to ensure the PCR product contained an adenine 3' overhang, which is required for TOPO® cloning. To perform this elongation, 17.4 μL of PCR product (eluted in TE buffer), 0.4 μL of dATP (10 mM), 0.2 μL of standard Taq polymerase and 2.0 μL 10x reaction buffer were mixed and the reaction incubated at 72°C for 15 minutes. The resulting PCR product was then stored at 4°C prior to further cloning.

Agarose gels were made with TBE buffer (Life Technologies), and run at 90 V until separation of 1 kb ladder DNA bands. DNA bands were either viewed using a UV transilluminator and gel documentation system, or individual DNA bands were excised from the gel under a Safe-imager™ 2.0 Blue-Light Transilluminator (Life Technologies) using a sterile razor blade, and the DNA of interest extracted from the agarose gel using a QIAquick Gel Extraction Kit (Qiagen).

2.2.6 Sanger sequencing

PCR products were purified as per Section 2.2.2. Purified products were Sanger sequenced by Macrogen Inc. (Seoul) and the resulting chromatogram and sequence aligned to the WT sequence region from the *P. fluorescens* SBW25 genome using Geneious version 6.1.7. Protein domains were identified using the pfam database (Finn et al. 2010).

2.2.7 Evolution of ISWS types

Experimental evolution approaches adapted from McDonald et al. (2009) were used to generate the large library of independent slow wrinkly spreader (ISWS) types. The ancestor strain PBR716 SM (Δ*wsp* Δ*aws* Δ*mws*) was streaked to individual colonies on KB agar plates. Individual colonies were then used to innoculate 200 KB microcosms, which were incubated statically for 3 days. At the end of this period, the microcosms were vortexed and diluted 10^{-3} in fresh KB MCs. The MCs were then statically incubated for another 3 days at 28°C, the cultures spread onto KB agar plates, and colonies screened for the appearance of morphologically WS types (screening a minimum of 500 and a maximum of 2000 colonies per plated microcosm). WS types were then streaked on KB plates from which overnight cultures (18 h 180 rpm) were made and frozen (-80°C).

2.2.8 Allelic exchange

Two-step allelic exchange (Kitten et al. 1998) involving the pUIC3 plasmid system was used to manipulate the genome and introduce mutations. In order to introduce a mutation, the mutation of interest was amplified using Phusion® High-Fidelity DNA polymerase (or generated with SOE-PCR if the mutation was not pre-existing) and primers that contained restriction enzyme sites to facilitate further cloning. The product was then cloned into a pCR8/GW/TOPO vector using the TOPO® TA cloning kit (as per kit instructions), and transformed into One Shot® TOP10 chemically competent *E. coli* cells. Clones were then screened by PCR for successful transformants, and Sanger sequencing was used to check for any introduced mutations in the cloned fragment.

The cloned fragment was then excised by restriction enzyme digestion, ligated into double-digested pUIC3 vector, and 10 ng of the ligated vector used to transform *E. coli* DH5α-*λpir* chemically competent cells. Transformation of pUIC3 with the intended fragment was confirmed by standard PCR. Triparental conjugation (Section 2.2.3) was used to integrate the fragment at the locus of the wild type (WT) allele, resulting in two copies of the locus (one WT and the other mutant), and the pUIC3 vector. Transconjugants were then grown in overnight culture, and 10 μL of culture used to inoculate 200 mL LB in a 1 L flask. This was incubated for 18 h (28°C 160 rpm) in the

absence of antibiotics to select for loss of vector via homologous recombination. A 400 mL aliquot of culture was then transferred to a 100 mL flask, and incubated for 30 min (28°C 160 rpm). Tetracycline was added to the culture, which was incubated for a further 2 h at 28°C (160 rpm) to allow the growth of transconjugants that had not lost the pUIC3 vector (and accompanying resistance) during the overnight culture. Cycloserine (1 mg mL^{-1}) was added to the culture, which was incubated for a further 5 h at 28°C (160 rpm). The addition of cycloserine was to kill the growing cells that were tetracycline resistant and had not undergone homologous recombination during the previous overnight culture. Rare types that had undergone homologous recombination and had lost the pUIC3 vector backbone were thus selected, leaving one copy of the target locus (either WT or mutant). A 1 mL aliquot of culture containing tetracycline and cycloserine was then centrifuged and the cell pellet washed in 1 mL LB, diluted and spread on KB agar plates containing X-gal (60 μg mL^{-1}). Plates were incubated for 2 days at 28°C, and white colonies (types that had undergone a second round of homologous recombination and had thus lost the pUIC3 vector containing *lacZ*) were picked and streaked out on KB agar pates, from which cultures were made and frozen at -80°C. The incorporation of the mutant allele (as opposed to the WT allele) was not confirmed by phenotypic change, but was then confirmed by PCR and Sanger sequencing.

2.2.9 Production and transformation of competent cells

Electrocompetent *P. fluorescens* cells were made from overnight cultures grown in LB incubated shaken at 28°C (180 rpm) using a method adapted from Choi et al. (2006). 1 mL of culture was aliquoted into a pre-chilled microcentrifuge tube and pelleted (16 000 x g, 1 min). The pellet was then resuspended in 1 mL 300 mM sucrose and centrifuged (16 000 x g, 1 min). Resuspension and centrifugation was repeated, and the final pellet resuspended in 50 μL 300 mM sucrose and stored on ice until transformation.

Electrocompetent *Pseudomonas* cells were solely transformed with plasmids that had been extracted from *E. coli*. Prior to transformation, aliquots of electrocompetent *Pseudomonas* cells were thawed on ice for 5 min prior to the addition of 250 ng plasmid (equimolar helper plasmid was applied at the same time in the case of transformation with mini-Tn7 plasmids). Reaction mixes were incubated on ice for a further 10 min, and then transferred to a 1 mm electroporation cuvette (Bio-Rad) and

electroporated at 1.8 kV. Cells were then revived with 500 μL pre-incubated LB media and incubated at 28°C for 1 h, and then plated on selective media.

Escherichia coli DH5α-*λpir* was used for the transformation of all freshly ligated plasmids other than pCR8/GW/TOPO. Overnight cultures of DH5α-*λpir* were prepared in LB incubated shaken at 37°C (180 rpm), and 4 mL of culture was used to inoculate 100 mL LB, which was incubated shaken at 37°C until mid-log phase (approximately OD_{600} 0.5). The culture was cooled on ice, and two 50 mL aliquots were pelleted in Falcon tubes (4000 × g, 10 min, 4°C). The supernatant was removed and pellets of cells were resuspended using an inoculation loop in 50 mL of 4°C 10% glycerol solution ($^{V}/_{V}$), and the centrifugation was repeated. This cycle of resuspension and centrifugation was repeated. The pellet in each tube was resuspended in 1 mL of cold 10% glycerol solution ($^{V}/_{V}$). Cells were frozen as 50 μL aliquots and stored at -80°C.

Electrocompetent cells were primarily used as hosts for the transformation of freshly ligated plasmids. Prior to transformation, aliquots of electrocompetent cells were thawed on ice for 5 min prior to addition of 5 μL ligation reaction mixture. Cells were then incubated on ice for 20 min prior to electroporation. Cells were then immediately revived by addition of 500 μL LB media, and incubated at 37°C for 1 h and plated on appropriate selective media.

2.2.10 Fitness assays

2.2.10.1 Fitness assays via colony morphology distinction

Fitness assays were performed to test the relative fitness of strains either morphologically distinct to each other or marked with a neutral *lacZ* marker. Strains were streaked on KB agar and incubated at 28°C for approximately 2 days. Individual colonies were used to inoculate individual KB microcosm replicates, which were then incubated overnight for approximately 16 h (28°C, 180 rpm). Cultures were then vortexed for 30 s, and 5×10^6 cells of both tested and reference strains were used to inoculate microcosms. Sampling of cultures was performed at 0 h, and the microcosms incubated statically at 28°C (incubation times are specified in results). Samples were then taken of the incubated culture plated on KB agar and incubated at 28°C for 2 days. Colony forming units (CFU) counts of competing types were used to determine

relative fitness. The natural logarithm of CFU counts at the final time point, divided by the CFU counts from the initial time point, were used to calculate a Malthusian parameter of each competitor. The ratio of the Malthusian parameters of each competitor served as the parameter for relative fitness (Lenski et al. 1991).

2.2.10.2 Competition fitness assays using flow cytometry

The competitive fitness of strains measured via flow cytometry was measured relative to a strain (as indicated in the text) marked with mini-tn7-GFP. To measure relative fitness using this method, biological replicates of all strains were grown for 16 h at 28°C. Cultures were then vortexed and test and marked strains were combined into single microcosms of 6 mL of KB media, resulting in a 10^{-6} dilution. The mixed strains were then grown shaken for 4 h to allow for a similar physiological state of the tested and marked strains. The initial ratio of the test and marked strain was then measured via flow cytometry. The ratio of the mixed strains was measured by counting a total of 100,000 cells using a flow cytometer (BD FACSCanto) that detected fluorescence using a 488 nm laser (530/30 bandwidth filter). Gating of fluorescence values was used to assign event counts to the marked or test genotype. Dilutions of each mixed culture were also plated on plates of KBA to determine the CFU of the initial mixed culture.

The mixed cultures were then diluted by 10^{-3} in KB media and incubated statically for 48 h at 28°C. At the end of this competition time, the final ratio of fluorescent and non-fluorescent cells and the final CFU were measured as per the initial counts. Control strains were included with and without the GFP marking, to assess the fitness cost of GFP expression during competition. The selection coefficient of each strain was calculated from the ratio of competing cell types and number of generations. The selection coefficient for each test population was calculated using the model $s = [\ln(R(t)/R(0))]/[t]$, where R is the ratio of measured to marked cell counts, and t is the number of generations (Dykhuizen 1990). Generations were calculated using the model ln(final CFU/initial CFU)/ln(2). The fitness cost of GFP expression was used to correct the selection coefficient for each test strain.

2.2.11 Transcriptional assay

Relative levels of transcription were determined from promoterless '*lacZ* fusions using transconjugants of tested strains and the pUIC3 vector as previously described (Zhang and Rainey 2007b). DNA regions tested for transcription were cloned into pUIC3 using PCR conditions and cloning methods described in allelic exchange (Section 2.2.8). However, the cloning protocol was halted prior to the second step of allelic exchange, resulting in a fusion in which the region of interest was transcriptionally fused to '*lacZ* of pUIC3, separated by three stop codons in alternative frames. Transconjugants were then grown in tested media conditions as indicated, and the expression of transconjugants measured by β-galactosidase assay. Here, cells were lysed and the rate of enzymatic conversion of 4-methylumbelliferyl-ß-d-galactoside (4MUG) to fluorescent 7-hydroxy-4-methylcoumarin (4MU) was measured *in vitro* at 460 nm after excitation at 365 nm using a Synergy 2 plate reader (BioTek Instruments). Levels of fluorescence were compared to a standard curve of 4MU. The density of cells grown in the tested conditions was measured at 600 nm, and used to express enzymatic activity as (nM 4MU OD_{600}^{-1} min^{-1}).

2.2.12 Reverse transcription quantitative real-time PCR

The relative level of transcription across the *nlpD* and *rpoS* loci required accurate measurement without disturbance of the gene structure. To assess this relative transcription, reverse transcription quantitative real-time PCR (RT-qPCR) was employed to assess the relative production of RNA from these genes. To isolate RNA from this region, the total RNA was isolated from *nlpD* mutants using the SV Total RNA Isolation System (Promega). RNA was isolated from cells subcultured and harvested at OD_{600} = 2.5–3.0 (to ensure the cells were in stationary phase). 1 mL of culture was centrifuged (13,000 x g, 1 min) and cells were resuspended in TE buffer (10 mM Tris, 1 mM EDTA pH 8.0) with 0.4 mg/mL lysozyme to lyse the cells. The RNA was isolated using the supplied kit protocol. The isolated RNA was reverse transcribed into cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and diluted 40 times before use in quantitative RT-qPCR (DyNAmo Colorflash SYBR Green qPCR Kit (Thermo Scientific), PikoReal 96 Real-Time PCR System (Thermo Scientific)). The relative levels of mRNA between mutant loci were assessed using the ΔΔCq method using *recA* transcript as an internal control (Livak and Schmittgen 2001).

Four technical replicates and two biological replicates were used over two separate experiments.

2.2.13 Fluorescence microscopy (ACP and GFP)

Acetylated cellulose polymer (ACP) production was detected by *in vivo* staining of cell colonies with calcofluor, and fluorescent ACP was visualised by fluorescence microscopy. Colonies of *P. fluorescens* were grown for 48 h on KBA plates containing calcofluor (35 μg mL^{-1}), and several colonies were scraped from the plates and resuspended in 100 μL distilled water. 10 μL of mixed cells were pipetted onto microscope slides, covered with coverslips, and visualised using 4',6-diamidino-2-phenylindole (DAPI) excitation and emission at 60× magnification.

The localisation of fluorescently tagged proteins was visualised *in vivo* via fluorescence microscopy. The open reading frames of genes to be tagged were cloned using standard procedures and ligated into the multiple cloning site (MCS) of mini-Tn7-LAC (Choi and Schweizer 2006). This plasmid was modified to contain both a *Pseudomonas*-specific ribosome binding site (RBS) and a GFP*3 gene 3' to the ligated gene. Plasmids were then transformed into electrocompetent *P. fluorescens* cells and confirmed for integration into the neutral *attB* site by PCR and electrophoresis (as described by Choi and Schweizer (2006)). Strains were prepared for microscopy by inoculation from colonies in 6 mL sterile M9 media supplemented with 2% glycerol (with appropriate antibiotics), and grown overnight (160 rpm 18 h). An aliquot of 100 μL of overnight culture was then subcultured into 6 mL of M9 media supplemented with 2% glycerol (without addition of antibiotics) and the culture incubated shaken (160 rpm) for approximately 1 h, after which 6 μl 1 M ITPG was added to induce expression of the cloned gene. The culture was returned to the incubator for approximately 2 h until OD_{600} 0.1-0.3 was reached. To increase the concentration of cells to an appropriate density for microscopy, 1 mL of induced culture was centrifuged (13,000 x g, 1 min) and the pellet resuspended in approximately 100 μL M9 media. Agarose pads were prepared, consisting of 250 μL M9 media set with 1% agarose (Invitrogen), and were compressed with an additional slide. 3 μL of resuspended culture was pipetted to the pads, which were left to dry at room temperature for approximately 10 min, and a glass coverslip was placed on the pad. Cells were visualised using an Olympus BX61 upright fluorescence microscope at 100× magnification, and photographic images recorded on

an F-View II monochrome camera. Fluorescent images were captured using a constant exposure time of 3500 ms across samples. Three biological replicates were used per strain, and all image processing was performed using FIJI (Schindelin et al. 2012).

2.2.14 Whole-genome sequencing

Genomic DNA of isolates was extracted from overnight cultures inoculated from samples stored at -80°C. Genomic DNA was prepared for sequencing using a Wizard® Genomic DNA Purification Kit (Promega). Aliquots of genomic DNA were retained for further Sanger sequencing. Genomic DNA was resuspended in TE buffer (pH 8.0), with a minimum concentration of 100 ng μL^{-1}. The quality and quantity of extracted genomic DNA was confirmed by both spectrophotometry (a 260/280 ratio of approximately 1.8-2.0) and agarose gel electrophoresis (0.8% agarose). Sequencing of the genomic DNA was performed using an Illumina Hi-Seq 2000 platform (Australian Genome Research Facility, Melbourne) based on the preparation of 50 bp single end reads with a depth of approximately 50 reads. Sequence reads were aligned to the *P. fluorescens* SBW25 genome (Silby et al. 2009) using Geneious version 6.1.7 (Biomatters). Point mutations and genomic re-arrangements were identified by deviations from the aligned reference – a minimum of 60% deviation from a minimum of 10 overlapping reads was required for the designation of a putative SNP.

2.2.15 SREE protocol

The methodology of the slow reverse evolution experiment (SREE) was replicated from the original reverse evolution experiment (REE) but with minor modifications to the initial genotype, the number of lines and number of reversals (Beaumont et al. 2009). Approximately 6 μL of frozen -80°C glycerol stock of the founding genotype (PBR716) was used to inoculate 25 mL microcosms containing 6 mL KB medium. Eight lines were inoculated in total, along with eight control lines that were grown in shaken environments for the same period of time as each matched experimental line. The experimental lines were then incubated static for 72 h, whereupon 6 μL was transferred to fresh microcosms. An aliquot of microcosm was stored in glycerol at -80°C to allow revival of the experiment in the event of contamination. All microcosms were diluted (between 10^{-5} and 10^{-6}) and plated on KBA between transfers. Plates were incubated

for 48 h at 28°C, and visually screened for morphologically distinct types compared to the direct ancestor (minimum of 500 colonies were visually inspected). Distinct colonies were then streaked to single colonies on both a 'bottlenecking' KBA plate and a 'comparison' plate with an ancestral colony streaked to compare the morphological difference. Colonies that were confirmed as morphologically distinct were isolated from the control plate, inoculated in 6 mL KB media and grown overnight shaken at 28°C. The 'bottlenecking' plate served to bottleneck the population to a single CFU, and the 'comparison' served to ensure heritability of the visual distinction. The overnight culture was then used to make a glycerol stock and stored at -80°C, marking a single round. This round was then repeated seven more times, with shaken environments used alternatively, whilst maintaining all other parameters. A genotype that had passaged over both static and shaken environments was considered to have undergone a single bout of evolution. The ancestral morphological type was maintained in the control lines via random isolation of the ancestral SM types at the same time as the nascent evolved types.

2.2.16 Ordering mutations of the SREE

Mutations that arose at each isolated mutant of the SREE were identified using a combination of WGS and Sanger sequencing. To do this, whole genome sequencing of the final sequential genotype of each line was performed. PCR of identified mutations was performed using frozen glycerol stock of the isolated mutants and Sanger sequencing of the resulting PCR products was performed. To confirm mutations, the immediate ancestor and suspected mutant encoding the mutation were used as a template for PCR. Successive rounds of evolution at which mutations were not detected were screened for direct reversals of mutation.

2.2.17 Measuring *rifR* mutation rates

Genome-wide mutation rates were estimated by determining the frequency with which genotypes cause rifampicin-resistant (*rifR*) mutants. To do this, entire overnight cultures (200 µL) of each genotype were plated on KBA containing rifampicin, and subsequent resistant colonies were counted. Both the number of resistant mutants and the total population number were used to estimate the global mutation rate of

genotypes (using the Ma-Sandri-Sarkar Maximum Likelihood Estimator (MSS-MLE) method, using the web tool 'FALCOR', a fluctuation analysis calculator (Hall et al. 2009)).

2.2.18 Statistical analyses

All statistical analyses were performed with the R 2013 version 3.0.1 (the R Foundation for Statistical Computing). Normal distributions were assessed using the Shapiro-Wilk normality test, and comparability of variance assessed using the *F*-test. The significance of difference between means was detected using two-sample Student's *t*-test. Where normality standards could not be met of compared data sets, the Welch two-sample *t*-test was employed.

Chapter 3:
Characterisation of an adaptive pathway created by domain fusion and protein relocalisation

3.1 Introduction

This chapter addresses how the genetic architecture of a gene – the function and regulatory connectivities made by the gene of interest – determines the frequency at which evolution uses loci to generate (upon mutation) adaptive WS types. This work builds upon previous studies suggesting types of regulation may bias the frequency at which mutations are translated into adaptive phenotypic variation (McDonald et al. 2009).

In populations of ancestral SBW25, McDonald and colleagues identified three loci at which mutations causing WS arose (McDonald et al. 2009). From an initial study of 26 independent WS types, mutations generating WS were found exclusively in *wsp*, *aws* or *mws*, despite the existence of approximately 30 additional di-guanylate cyclase (DGC) encoding genes that could, upon mutation, provide additional opportunity to achieve the WS state. McDonald et al. (2009) argued that the apparent bias toward *wsp*, *aws* and *mws* could be understood in terms of the size of the mutational 'target size'. Each locus encodes a DGC subject to post-translational negative regulation. In each instance, activation of the DGC can be achieved via a loss-of-function mutation in the negative regulator. Given that most mutations are deleterious, negative regulation of *wsp*, *aws* and *mws* provides selection with a large mutational target.

To support the claim that genetic architecture (here, the mode of regulation) may bias the observed spectrum of WS types, McDonald and colleagues showed that WS types could arise via mutation in pathways other than *wsp*, *aws* or *mws*. To demonstrate this, the authors deleted the *wsp*, *aws* and *mws* loci from the ancestral SM genotype (PBR716). PBR716 was placed in a structured microcosm and propagated for 5 days until WS were observed. One of these WS types, termed SWS (slow wrinkly spreader), was then subject to further analysis (McDonald et al. 2009). Importantly, McDonald et al. showed that the longer time taken (5 days compared with 2-3 days in ancestral

SBW25) for WS types derived from PBR716 to arise was not a consequence of reduced fitness of SWS compared to WS types (that arose as a consequence of a mutation in either *wsp*, *aws* or *mws*). This led the authors to conclude that the mutation or mutations causing SWS was unlikely to reside in a negative regulator (a large mutational target) and more likely to reside in a locus affording a reduced mutational target. If true, this would show that the pathways that evolution takes are strongly influenced by genetic architecture.

McDonald and colleagues did not identify the mutation causing SWS, despite a mutagenesis screen intended to achieve this. Identifying the causal mutation is important in order to determine whether SWS is caused by a mutation that affects a small mutational target (relative to targets in *wsp, aws* or *mws*). Analysis of the mutational target has the potential to explain why WS types arising by pathways other than *wsp*, *aws* and *mws* do so at a reduced rate.

3.2 Aims

-To identify the mutational basis of the SWS phenotype, reconstruct this mutation in ancestral SBW25, and determine the fitness of the resulting WS relative to commonly evolved WS (to confirm the finding of McDonald et al. (2009))

-Determine whether features of regulation and functional interaction encoded by this locus – the genetic architecture – limit the translation of mutations into adaptive WS phenotypes.

3.3 Results

3.3.1 Identification of the genetic basis of SWS and measurement of the effects on fitness

In order to determine the role of selection and genetic architecture in determining SWS types, the mutation or mutations underpinning SWS require identification. This section describes the process of that identification, the causal nature of the identified mutation, and the effect of these mutations on fitness relative to other known WS types.

3.3.1.1 The SWS mutational locus spans two genes

In order to determine the mutational basis of the SWS type, the strategy employed by McDonald et al. (2009) was utilised. Here, transposon mutagenesis was performed on the 'SWS' type (PBR715) using the IS-Ω-Km/hah transposon (Bailey and Manoil 2002; Giddens et al. 2007)). A total of 54,000 mutant types were screened for loss of wrinkly spreader (WS) colony morphology and a return to the ancestral smooth (SM) morphology. A total of 36 mutants were identified as morphologically indistinguishable from that of the ancestral SM morphology. Of this collection, the position of the transposon was determined in 31 instances[3]. The coordinates of the transposon insertion in each case are detailed in Appendix Table 7.3.1. Of 31 transposon mutants, 11 insertions were mapped to PFLU0183 and 17 were mapped to the neighbouring gene PFLU0184. Insertions were also found in PFLU1667 (a predicted aminotransferase), PFLU3346 (encoding no known functionality), and *wssE* (which is an essential gene for cellulose production and the WS phenotype) (Spiers et al. 2002).

Sanger sequencing of PFLU0184 and PFLU0183 of SWS revealed a 467 bp deletion spanning the two predicted genes (see Figure 3.1), removing 200 bp from the 3' region of PFLU0184, the intergenic region (182 bp), and 85 bp from the 5' region of PFLU0183. This deletion generated an in-frame fusion resulting in a single predicted 1908 bp open reading frame.

[3] PCR using random primers generates a low quantity product which provides an unreliable template for Sanger sequencing

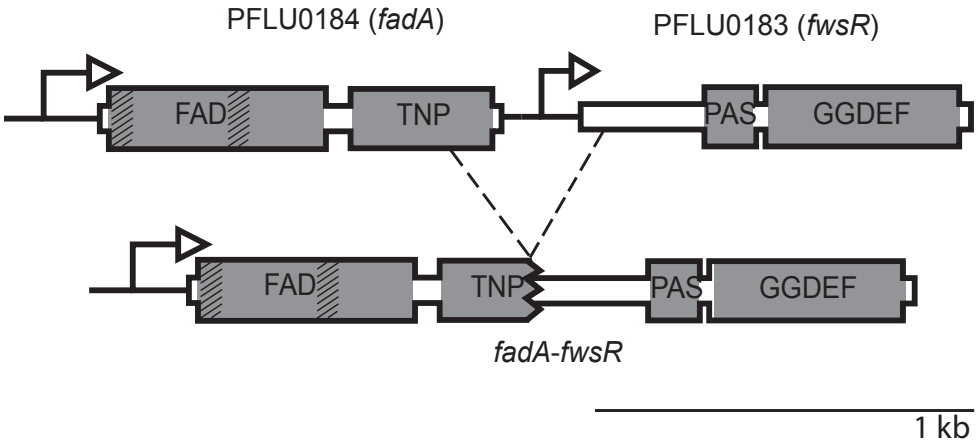

Figure 3.1: Illustration of the arrangement of PFLU0184 and PFLU0183 in the ancestral SM genotype and in SWS. Grey areas represent predicted domains and diagonal lines within grey areas represent predicted transmembrane domains. Dashed lines represent the region of the deletion that results in the fusion.

The predicted protein encoded by PFLU0183 suggests this deletion mutation may cause the SWS phenotype. PFLU0183 is predicted to encode a di-guanylate cyclase (DGC) and is henceforth referred to as '*fwsR*'[4]. The gene *fwsR* forms a predicted protein (FwsR) of 335 residues in length with a predicted PAS fold (of the PAS_3 family) at residues 68 to 157 (Pfam *E*-value 2.4 × 10^{-18}), and a predicted GGDEF domain at residues 174-331 (Pfam *E*-value 1.1 × 10^{-41}). The neighbouring gene PFLU0184 encodes a predicted protein 394 residues in length, including a predicted fatty acid desaturase at residues 11 to 246 (Pfam *E*-value 1.4 × 10^{-14}) and a predicted transposase element (DDE_Tnp_ISL3 family) at residues 261 to 389 (Pfam *E*-value 2.3 × 10^{-9}). PFLU0184 is hereafter referred to as '*fadA'*. The putative FadA protein contains two predicted transmembrane domains (TMDs) between residues 10-32 and 135-157 as predicted by the 'TMHMM server v2.0' that uses hidden Markov models (Krogh et al. 2001). The predicted gene resulting from the deletion mutation is hereafter termed '*fadA-fwsR'*.

Little is known about the functions of proteins encoded by either *fadA* or *fwsR* in *Pseudomonas fluorescens* SBW25. FadA shares 83% amino acid homology with DesA (PA0286) in *P. aeruginosa*, which functions as a fatty acid desaturase (Zhu et al.

[4] The abbreviation *fws* stands for *F*arr's *w*rinkly *s*preader. A locus encoding a DGC (PFLU1349) was identified in McDonald et al. (2009) during a transposon mutagenesis screen and labelled *sws*; however, no mutation was identified at this locus. To avoid confusion, the term '*sws'* has been retained for PFLU1349, and *fwsR* has been adopted for PFLU0183 ORF.

2006). Fatty acid desaturases modify phospholipids in the cell membranes in order to modify membrane fluidity in response to environmental change (Zhang and Rock 2008). In the case of *P. aeruginosa*, transcription of *desA* is promoted during anaerobic conditions and the encoded DesA introduces a double bond at the 9-position of fatty acid acyl chains to increase fluidity (Zhu et al. 2006).

The gene *fwsR* also has many orthologs across *Pseudomonas* spp.; however, none of these orthologs have been studied. The presence of a GGDEF domain (in this case featuring a GGEEF motif) is indicative of DGC function. The orthologs of *fwsR* show some synteny in relation to orthologs of *fadA* across *Pseudomonas* spp. (see Figure 3.2). Species such as *P. putida* and *P. stutzeri* share the same synteny as *P. fluorescens*. However, in *P. aeruginosa* the locus of the GGDEF domain-encoding gene (PA0260) is separated from *desA* by three open reading frames (ORFs). It is unknown whether the proximal relationship of *fadA* and *fwsR* orthologs across many *Pseudomonas* spp. represents a functional or regulatory relationship between the genes or encoded proteins.

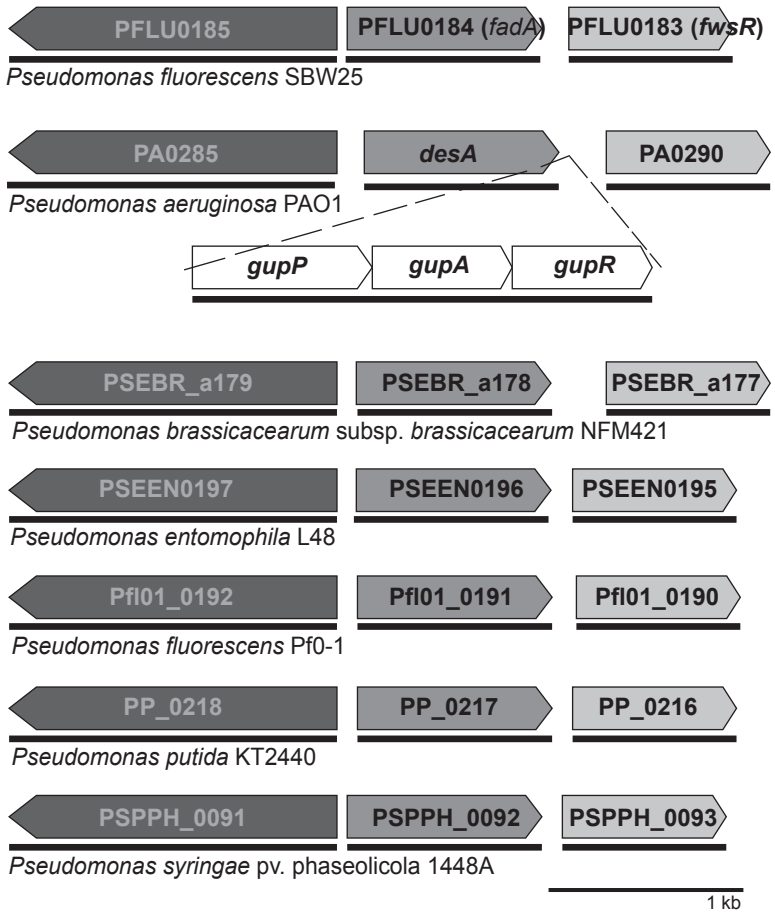

Figure 3.2: Organisation of homologs of *fadA* and *fwsR* across *Pseudomonas* spp. The arrangement of neighbouring genes of the *fwsR* locus is conserved across many species of the genus *Pseudomonas*. Notable exceptions are different strains of *P. aeruginosa*, in which the paralogs *fadA* and *fwsR* are separated by a single predicted operon of genes *gupPAR*. There is some variation in the intergenic space between predicted genes. There is no evidence of recombination between *fadA* and *fwsR* in other strains. Genes sharing the same shade represent likely paralogs as determined by the *Ortholuge* web program (Whiteside et al. 2013), with thin black bars (underneath depicted ORFs) representing operons as predicted by the *DOOR* web program (Mao et al. 2014).

3.3.1.2 The *fadA-fwsR* fusion is sufficient to form WS

Evidence the *fadA-fwsR* fusion mutation is sufficient to cause the SWS phenotype was obtained by replacement of ancestral *fadA* and *fwsR* with the *fadA-fwsR* fusion in both PBR716 (SM Δ*wsp* Δ*aws* Δ*mws*) and ancestral SBW25.

Three independent allelic replacements were performed in each genetic background. In each instance, introduction of the fusion resulted in morphology, ability to colonise the air-liquid interface (ALI), and production of acetylated cellulosic polymer (ACP) characteristic of the SWS phenotype (see Figure 3.3). The WS phenotypes caused by introduction of *fadA-fwsR* in both PBR716 and SBW25 demonstrate that this SWS phenotype is not dependent on the absence of *wsp*, *aws* and *mws*. Furthermore, the reconstruction demonstrates additional mutations causing epistatic interactions are not required for the *fadA-fwsR* fusion to cause the SWS phenotype.

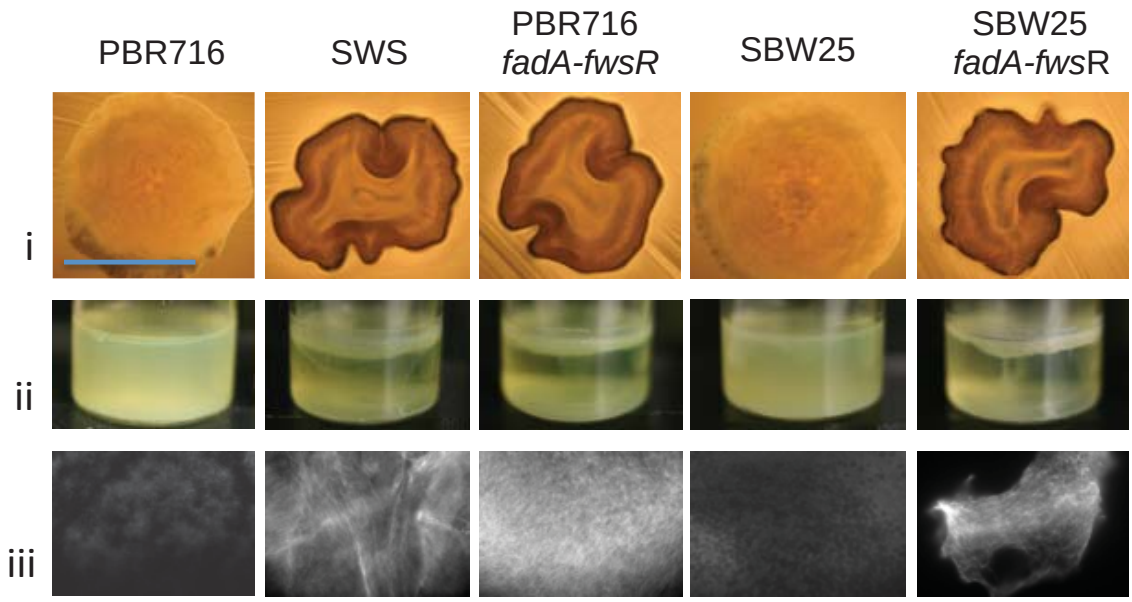

Figure 3.3: Reconstruction of the *fadA-fwsR* mutation in SM backgrounds causes WS. Genotypes featuring *fadA-fwsR* (either naturally evolved or reconstructed in SM backgrounds) are characteristically WS in terms of: i) morphology (visualised by light microscopy (10x) of ~32 h colonies grown on King's B agar (KBA), blue scale bar is ~2 mm); ii) formation of mats at the air-liquid interface of microcosms (incubated statically at 28^{o}C for 3 days); iii) calcofluor binding as indicative of ACP biosynthesis (visualised by fluorescent microscopy, 60x objective). SWS is the original WS genotype derived from PBR716 SM (Δ*wsp* Δ*aws* Δ*mws*) carrying the causal *fadA-fwsR* fusion mutation, PBR716 *fadA-fwsR* has the fusion mutation reconstructed in the PBR716 background. SBW25 is ancestral SM *P. fluorescens* SBW25, SBW25 *fadA-fwsR* has the causal SWS mutation reconstructed into the ancestral SM background.

The *fadA-fwsR* fusion reconstructed in ancestral SM has particular significance for the remainder of the study (this genotype is hereafter referred to as PBR955). Its significance stems from the fact that it is directly comparable to other WS genotypes, which also differ from ancestral genotypes at only the WS-generating locus of interest.

3.3.1.3 The *fadA-fwsR* fusion causes WS of similar fitness to *wspF* mutants

The delayed appearance of SWS relative to readily evolved WS types and the infrequent observation of WS caused by mutations outside of *wsp*, *aws* and *mws*, have two possible, but not mutually exclusive, explanations. WS generated by mutation in *wsp*, *aws* and *mws* may have high fitness relative to SWS. Alternatively, the regulation of *wsp*, *aws* and *mws* may differ significantly from that of *fadA* and *fwsR* – the former thus offering a much larger mutational target than the latter.

McDonald and colleagues had indirectly performed a comparison of the fitness of SWS with WS evolved from SBW25. However, without identification and reconstruction of the causal mutation in the ancestral SM genotype, the effect of additional mutations on the fitness to SWS could not be excluded.

SBW25 *fadA-fwsR* (PBR955) was used to provide a direct measure of fitness of this rarely observed WS. PBR955 was competed with two WS types that had evolved from SBW25 via mutations in *wspF*. These two types were LSWS and $WS_{G,}$[5] the fitness of each are found at the higher and lower ends of the spectrum of WS fitness respectively (Bantinaki et al. 2007). Competitions were performed over 3 days in static microcosms (see Appendix Tables 7.3.2 and 7.3.3 for raw data and details of statistics).

The fitness of SBW25 *fadA-fwsR* is within the range of fitness displayed by other WS that have evolved via mutations to common pathways (see Table 3.1). In a direct comparison of PBR955 (SBW25 *fadA-fwsR*) with LSWS, PBR955 was significantly less fit (mean relative fitness = 0.966, *t*-test *p-value* = 0.046). In a direct comparison with WS_G, PBR955 was significantly more fit (mean relative fitness = 1.046, *t*-test *p-value* = 0.044). A control test was also included of PBR955 tested against PBR955 expressing *lacZ*, to test for the neutral effect of the chromosomally integrated *lacZ* marker. There was no significant effect of the *lacZ* marker (mean relative fitness of the non-marked strain = 1.019, *t*-test *p*-value = 0.166).

[5] LSWS stands for 'large spreading wrinkly spreader' and has an A901C transition in *wspF*, resulting in an S301R amino acid substitution. WS_G is one of a set of independent WS types (A to Z) evolved from SBW25, and features a C556T transversion in *wspF* that results in an H186Y amino acid substitution (Bantinaki et al. 2007).

The relative fitness of PBR955 suggests that the rare appearance of *fadA-fwsR* mutants derived from SBW25 is not a consequence of reduced fitness relative to commonly generated WS types.

Competition	Competitor 1	Competitor 2	Mean RF ± 95% CI[a]	d.f.[b]	*t*-test p-value[c]
A	PBR955-*lacZ*	LSWS	0.966 ± 0.033	10	0.046
B	PBR955	PBR955-*lacZ*	1.019 ± 0.028	10	0.166
C	PBR955-l*acZ*	WS_G	1.046 ± 0.045	10	0.044

Table 3.1: Relative fitness of PBR955 directly compared to *wspF* mutants. Direct competitions of PBR955 with LSWS and WS^G were made to check the relative fitness of these strains. A competition with PBR955-*lacZ* was made to check the neutrality of the *lacZ* marker. [a]Mean relative frequency (RF) ± 95% confidence interval (CI). [b]Degrees of freedom (d.f = number of replicates – 1). [c]Student's *t*-test *p*-value for one-sample, two-tailed, *t*-test that the relative frequency differs from 1.

3.3.2 Multiple *fadA-fwsR* fusion mutations cause WS types

To understand how the *fadA-fwsR* fusion mutation causes the SWS phenotype, a diversity of *fadA-fwsR* mutants were obtained. PBR716 (SM Δ*wsp* Δ*aws* Δ*mws*) was inoculated in 200 independent, statically incubated microcosms (see methods and materials Section 2.2.7 for details) and propagated for 6 days. Of this set of microcosms, 91 yielded WS colonies in a visual screen for WS types across a minimum of 500 plated colonies (a maximum of 2000 colonies were visually screened). These types are referred to as 'independent slow wrinkly spreaders' (ISWS). PCR was performed using primers sufficient to detect size polymorphisms within or between *fadA* and *fwsR*, as well as in the predicted promoter region of *fadA*.

From the ISWS collection, eight WS types were identified that contained discernable size polymorphisms at the *fadA-fwsR* locus consistent with a range of deletions having occurred between *fadA* and *fwsR*. Sequence analysis of the fusion junctions showed all deletions resulted in fusions between *fadA* and *fwsR*. These fusions consisted of four unique deletion mutations that ranged in size from 161 bp to 872 bp (see Figure 3.4 and Table 3.2 for coordinates). Four of the independent fusion mutations were

identical, possibly generated by a region of homology at the junction of the deletion (see Figure 3.5). In all instances, *fwsR* remained in-frame, the fusions did not remove the conserved domains of *fwsR*, and the resulting phenotype was characteristic of SWS (see Figure 3.6). Analysis of the remaining 83 ISWS showed no additional mutations at this locus that might have been missed by the initial PCR screen (P.A. Lind, personal communication).

Details of an additional *fadA-fwsR* fusion are included in the figures below. This additional fusion was obtained from a separate evolution experiment initiated with populations of PBR716 (SM Δ*wsp* Δ*aws* Δ*mws*) and was of special interest given the size of the deletion in *fadA*. This mutant, termed 'L2-M5', contains a *fadA-fwsR* deletion that removed 65% of the predicted FadA domain, and fused the remaining *fadA* region to *fwsR* (see Figure 3.4). To confirm this mutation was causal of the WS phenotype, without the effects of other mutations in the 'L2-M5' genotype, the *fadA* and *fwsR* allele was replaced by the 'L2-M5' *fadA-fwsR* fusion mutation in SBW25. The resulting SBW25 'L2-M5' *fadA-fwsR* was phenotypically WS (Figure 3.6). The removal of most of the predicted Fad domain strongly suggests fatty-acid desaturase activity is not required for the *fadA-fwsR* fusion to cause SWS. This mutation also suggests the structure of FadA is not required to cause the SWS phenotype, unless that structure is proximal to the *N*-terminus.

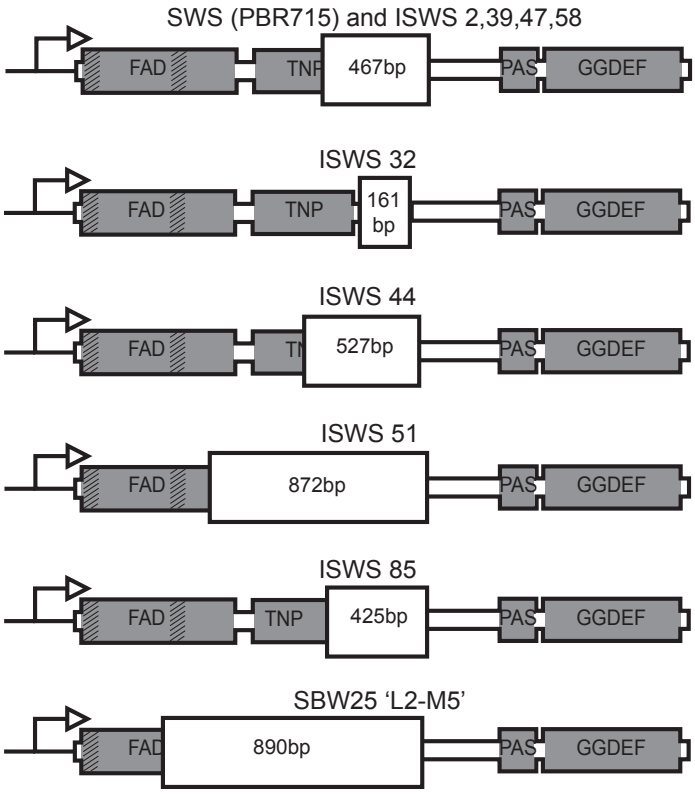

Figure 3.4: Illustration of the eight independent *fadA-fwsR* fusions at *fadA* and *fwsR*. The mutations in ISWS 2, 39, 47 and 58 share a common deletion junction with the *fadA-fwsR* fusion identified in the SWS genotype (PBR715). The L2-M5 *fadA-fwsR* deletion removes most of the FadA domain, including the second TMD. All deletions result in the ancestral predicted reading frame for the GGDEF domain.

Strain	Genome coordinate of mutation	Deletion size (bp)	Position PFLU184	Position PFLU183
SWS	208857 - 209323	467	986	85
ISWS 2	208857 - 209323	467	986	85
ISWS 32	208967 - 209127	161	1182	-26
ISWS 39	208857 - 209323	467	986	85
ISWS 44	208914 - 209440	527	869	28
ISWS 47	208857 - 209323	467	986	85
ISWS 51	208914 - 209440	872	574	78
ISWS 58	208857 - 209323	467	986	85
ISWS 85	208852 - 209276	425	1033	90
L2-M5	208860 - 209968	1108	341	82

Table 3.2: Details of the ISWS *fadA-fwsR* fusion mutations. Positions in genes are listed relative to the first bp of the predicted start codon. Genome coordinates are in reference to Genbank accession 'NC_012660'. SWS is the original SWS genotype (PBR716). L2-M5 is the *fadA-fwsR* mutation derived in a separate experiment (see Chapter 4).

Figure 3.5: Evidence of homology at the junction of the deletions resulting in several *fadA-fwsR* translational fusions. Presented are sequences found at the junction of the deletion for each *fadA-fwsR* fusion identified in the ISWS library. The top line for each type of fusion mutation is taken from the ancestral junction of *fadA*; the second line is taken from the ancestral junction of *fwsR*. The sequence that is removed as a result of the deletion mutation is coloured red. Vertical bars indicate homologous sequences.

One ISWS mutant in particular indicates the *fadA-fwsR* locus differs from the *wsp*, *aws* and *mws* loci in the type of mutation required to cause WS. ISWS 32 contains a small deletion that removes little of the open reading frame of either *fadA* or *fwsR*. The ISWS 32 deletion truncates 4 bp at the 3' end of *fadA* and removes the intergenic region of the two genes, leaving 25 bp, 5' to the start codon of *fwsR*. The ISWS 32 mutation involves minimal alteration to the coding frame of either gene, other than one non-synonymous mutation (I394M) in the *C*-terminal residue of ancestral FadA. This mutation was reconstructed in SBW25 and the resulting genotype was phenotypically WS, confirming this mutation causes WS without the possible effects of additional mutations (see Figure 3.6).

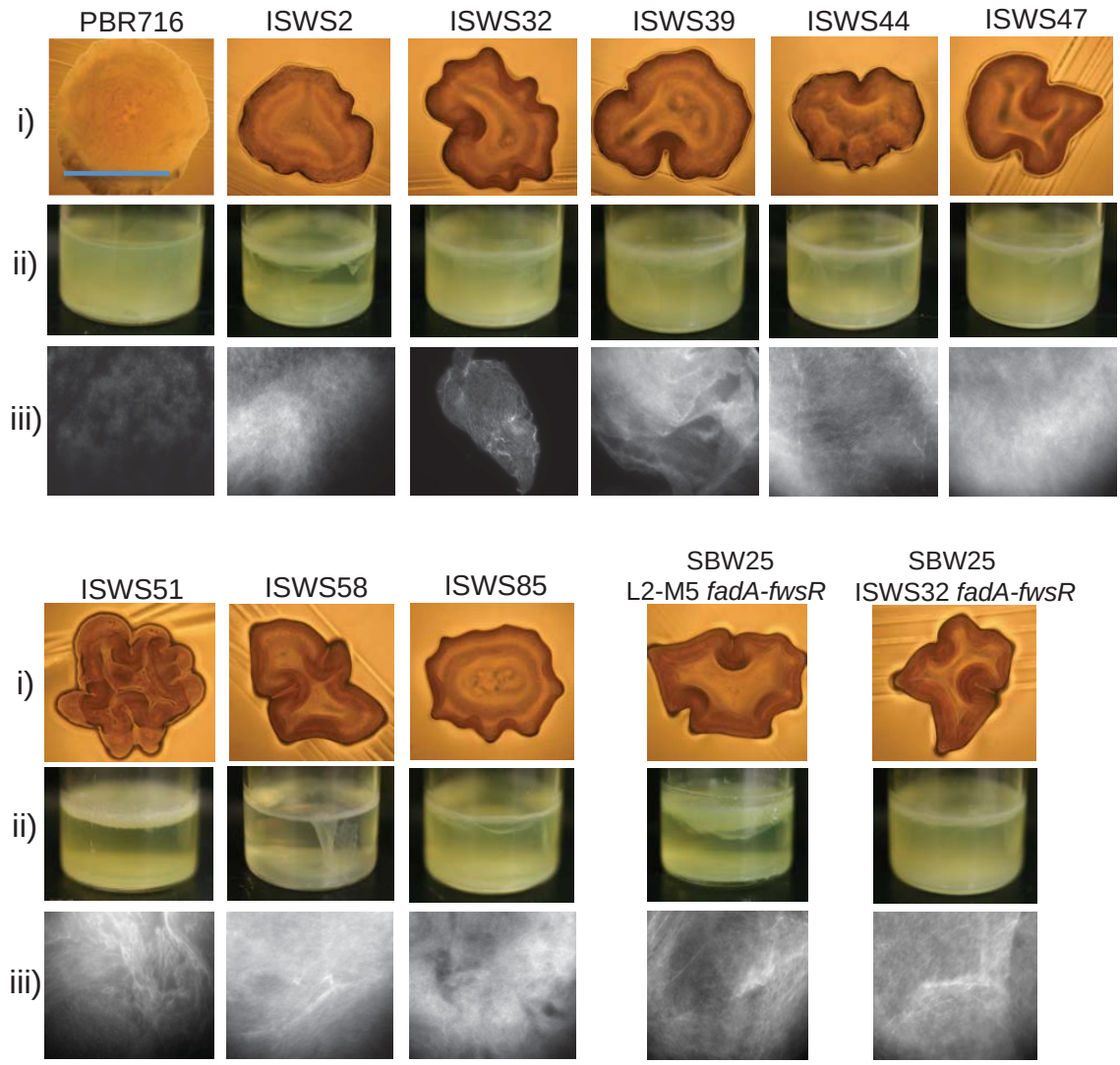

Figure 3.6: The spectrum *fadA-fwsR* fusions in the ISWS library express the WS phenotype. The phenotypes of the *fadA-fwsR* fusions identified in the 'independent slow wrinkly spreader' library are characteristic WS in: i) wrinkled morphology; ii) colonisation of the air-liquid interface; and iii) ACP production (for details of each image, see Figure 3.3. The blue bar represents ~2 mm).

The exclusive identification of fusions at this locus suggests that deletions resulting in the formation of a chimeric gene are required to generate the SWS phenotype. Furthermore, the exclusive finding of in-frame translational fusions that do not truncate the PAS and GGDEF domain, suggests DGC activity, in conjunction with a *fadA-fwsR* fusion mutation, is necessary to cause these nine WS phenotypes.

The nature of the resulting fusions allows predictions to be made concerning the molecular causes of the SWS phenotype. The predicted FadA-FwsR proteins all

feature two *N*-terminal TMDs and an intact PAS and GGDEF domain. This leads to the prediction that the *fadA-fwsR* fusion results in the coupling of the DGC to the inner cell membrane, resulting in the SWS phenotype. The putative DGC encoded by *fwsR* has no predicted TMD motif, and is predicted by the pSORTb v3.0 web program to locate to the cytoplasm (Yu et al. 2010). This location contrasts with the predicted or experimentally confirmed inner membrane locations for homologs of WspR, AwsR or MwsR – a transient membrane location in the case of WspR in *P. aeruginosa* (Guvener and Harwood 2007). By fusing FwsR to the membrane-localised FadA, it is likely that the DGC is relocated to the membrane, where it is either readily activated or where the produced c-di-GMP is more effective in allosteric activation of the WSS complex and other effectors of c-di-GMP signalling. This hypothesis for the cause of SWS was subject to further testing.

The only reasonable competing hypothesis is that the *fadA-fwsR* fusion, by transfer of the promoter for *fwsR* to the promoter of *fadA*, may result in altered transcription of *fwsR*, resulting in the greater production of c-di-GMP and the SWS phenotype. The altered transcription of *fwsR* may elevate the cellular levels of the DGC, resulting in higher levels of c-di-GMP. This change in transcription may be sufficient to cause the SWS phenotype. In studies on *P. fluorescens*, artificial over-expression of *wspR* is capable of inducing the WS phenotype (Goymer et al. 2006). Transcriptional up-regulation has been associated with the evolution of WS at other DGC-encoding genes (P.A. Lind, personal communication).

3.3.3 The *fadA-fwsR* fusion requires an active DGC to cause WS

The two competing hypotheses as to how the *fadA-fwsR* fusion causes WS (relocalisation of FwsR and promoter capture) require activity of the predicted FadA-FwsR DGC. To test the connection between the DGC activity and the SWS phenotype, non-synonymous changes were made to the GGDEF domain of the *fadA-fwsR* fusion.

The central role of the GGDEF (often, in *Pseudomonas*, the conserved domain is defined by GGEEF) motif of the DGC in synthesis of c-di-GMP is well understood (Chan et al. 2004; Paul et al. 2004; Malone et al. 2007; Wassmann et al. 2007; De et al. 2009). This information was used to make targeted mutations in the conserved residues of the active site. Two alanine residues were substituted for the two glutamine residues at positions 255 and 256 of FwsR, generating *FadA-FwsR* AA255-256EE (the

GGEEF motif becoming GGAAF). When introduced into PBR955 (SBW25 *fadA-fwsR*), the site-directed mutant no longer expressed a wrinkled morphology (see Figure 3.7). In addition, the mutant was no longer able to colonise the air-liquid interface of static microcosms and showed no evidence of ACP production. Together, these findings confirm the central role of the DGC of FadA-FwsR in expression of the WS phenotype, suggesting the DGC must be intact in order to generate the SWS phenotype, following a *fadA-fwsR* fusion.

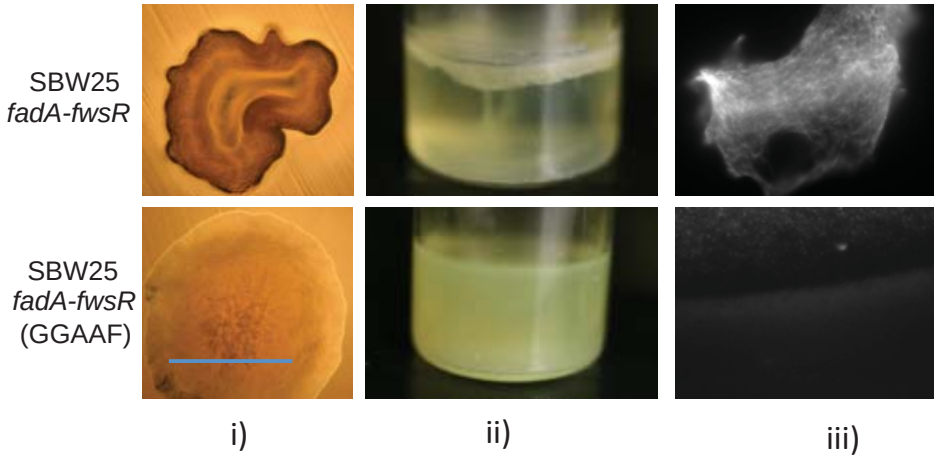

Figure 3.7: DGC activity encoded by *fadA-fwsR* is necessary to cause WS. In the context of *fadA-fwsR* reconstructed in SBW25, substitution of a GGAAF motif for the ancestral GGEEF motif within the predicted active site of the DGC domain results in SM phenotypes in: i) morphology; ii) colonised niche; and iii) ACP production. For details of each image, see Figure 3.3.

Interestingly, in an initial experiment intended to inactivate the DGC active site of FadA-FwsR, in which a phenylalanine residue was substituted for a leucine in the GGEEF motif (F257L, resulting in GGEEL), the resulting mutant manifested as a partial WS phenotype (see appendix Figure 7.3.1). This is in contrast to the observation of Malone et al. (2007) who made an identical change in the active site of WspR that obliterated the WS phenotype. The difference presumably reflects differences in the DGC active site of FwsR compared to WspR.

3.3.4 Investigation of how *fadA-fwsR* causes WS

3.3.4.1 Testing the hypothesis that promoter capture may increase expression of the DGC coding frame

Determining how the fusion mutation results in SWS requires direct assessment of whether promoter capture by *fwsR* (of the *fadA* promoter) is sufficient to cause the WS phenotype. Evidence that change in transcription of *fwsR* is not sufficient to cause this phenotype is presented below.

3.3.4.1.1 Sequence data suggests the *fadA-fwsR* deletion alters the operator of transcription of *fwsR*

Sequence data suggests ancestral *fadA* and *fwsR* are independently regulated, indicating the fusion mutation may change the operator region of *fwsR.* The WebGeSTer transcription terminator database identified the presence of a stem-loop structure indicative of a transcriptional terminator sequence (dG -22.31 kcal/mol) in the immediate 5' region of *fadA* (Mitra et al. 2011). This terminator appears to exhibit features, such as a poly-U tail, characteristic of a rho-independent transcription terminator (see Figure 3.8).

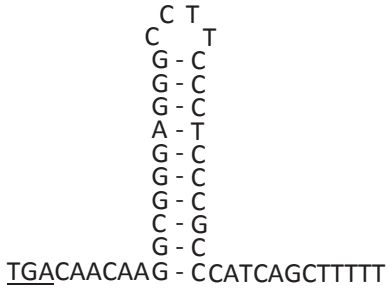

Figure 3.8: Sequence of the predicted transcription terminator of *fadA* (PFLU0184). This secondary structure is predicted by the WebGeSTer transcription terminator database (Mitra et al. 2011) to form a rho-independent transcription terminator directly downstream of the stop-codon of *fadA*. The stem-loop structure can cause a slower rate of transcription, and the poly-U tail causes a predicted dissociation of the RNA polymerase from the DNA. The underlined 'TGA' sequence indicates the stop codon of *fadA*.

The presence of such an element raises the possibility that SWS may be caused by enhanced transcription of *fwsR* resulting from the deletion of the putative intergenic transcriptional terminator and the capture of the promoter for *fadA* (P_{fad}). The presence of this transcription terminator was confirmed by measuring the transcription of *fwsR* (both with and without this intergenic region) via insertion of promoterless '*lacZ* into the 3' region of *fadA* and the 5' region of *fwsR* (see appendix Section 7.3.1).

3.3.4.1.2 Promoter capture is not sufficient to cause SWS

To test whether the capture of P_{fadA} by *fwsR* was sufficient to cause SWS, a construct was made in which P_{fadA} was adjoined to the open reading frame of *fwsR* (resulting in P_{fadA}-*fwsR,* Figure 3.9A). This construct was exchanged for the *fadA* and *fwsR* locus in ancestral SBW25. Derived mutants were then screened via PCR and Sanger sequencing to confirm cloning errors did not alter the sequence of P_{fadA} or *fwsR*. The artificial construct did not produce a WS phenotype, but rather, the SM phenotype in morphology, ecological habitation of the broth and undetectable ACP production (Figure 3.9B). This suggests the possible changes in transcription of *fwsR* caused by the *fadA*-*fwsR* fusion are not sufficient to cause SWS. This suggests that FadA plays a functionally necessary role in expression of the SWS phenotype.

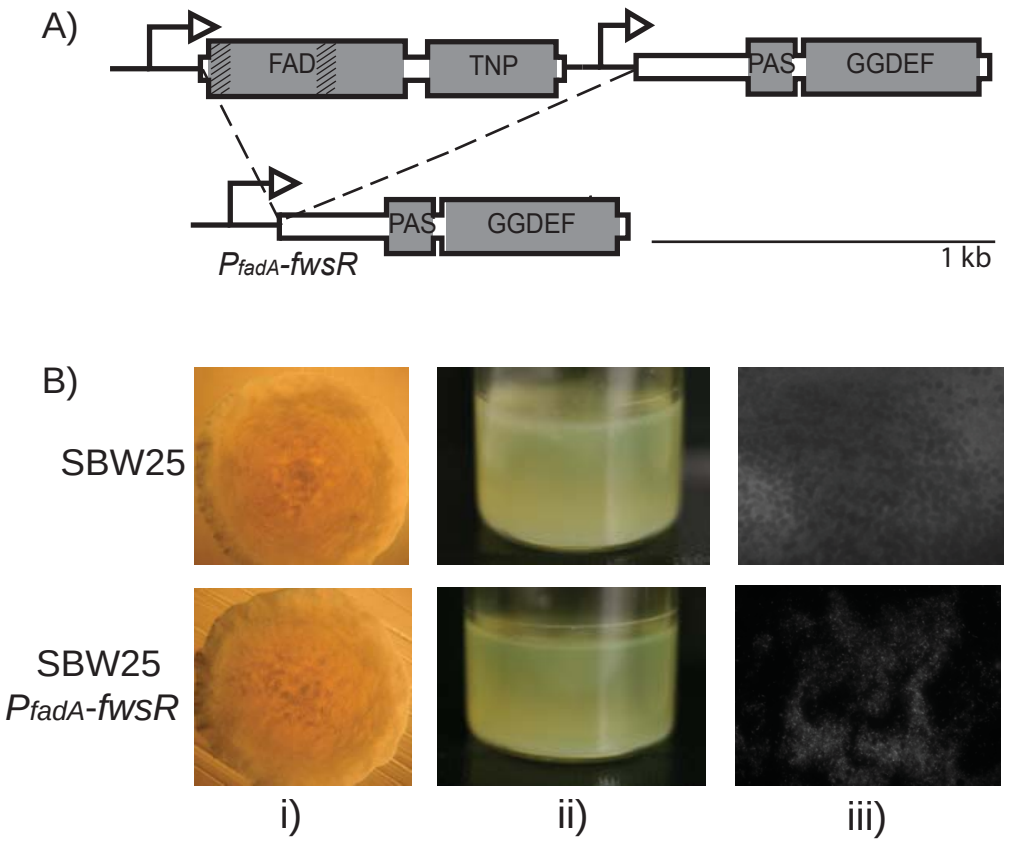

Figure 3.9: A) Illustration of the construct P_{fadA}-*fwsR*. In this construct the promoter for *fadA* is fused to *fwsR*. B) The potential change in transcription caused by promoter capture does not cause WS. The phenotype of the P_{fadA}-*fwsR* construct reconstructed in ancestral SBW25 is phenotypically SM in terms of: i) morphology: ii) broth colonisation; and iii) undetected ACP production. For details of each image, see Figure 3.3.

3.3.4.2 The SWS phenotype requires FadA and FwsR to be translationally fused

Having rejected the hypothesis that promoter capture underpins expressions of SWS I next sought to understand how the fusion might generate the WS morphology. The presence of in-frame deletions suggests the fusion couples the two genes into a single protein and that this translational fusion is the cause of SWS. This prediction was initially tested by the generation of a *fadA-fwsR* fusion in which translational coupling was eliminated. To generate such a fusion the region between the stop codon of *fadA* and the ATG start codon of *fwsR* was deleted and replaced by stop codons (one in each reading frame). The resulting genotype was designated *fadA-3X-fwsR* (see Figure 3.10A). The *in situ* RBS of *fwsR* was left intact to ensure translation initiation of an independent protein product for the *fwsR* region of *fadA-3X-fwsR*. Introduction of

the allele into ancestral SM by allelic exchange did not produce a WS phenotype, either in morphology, ecological niche or ability to produce detectable ACP (Figure 3.10B). This shows that *fadA* must be translationally fused to *fwsR* in order to cause WS.

Interestingly, in preliminary work, a construct was also made containing a single stop codon between *fadA* and *fwsR* (but was otherwise identical to *fadA-3X-fwsR*). This single stop codon did not abolish the WS phenotype to the same extent as the *fadA-3X-fwsR* construct (see Appendix Figure 7.3.4). This suggests complete translation uncoupling of FadA and FwsR is required to prevent the SWS phenotype.

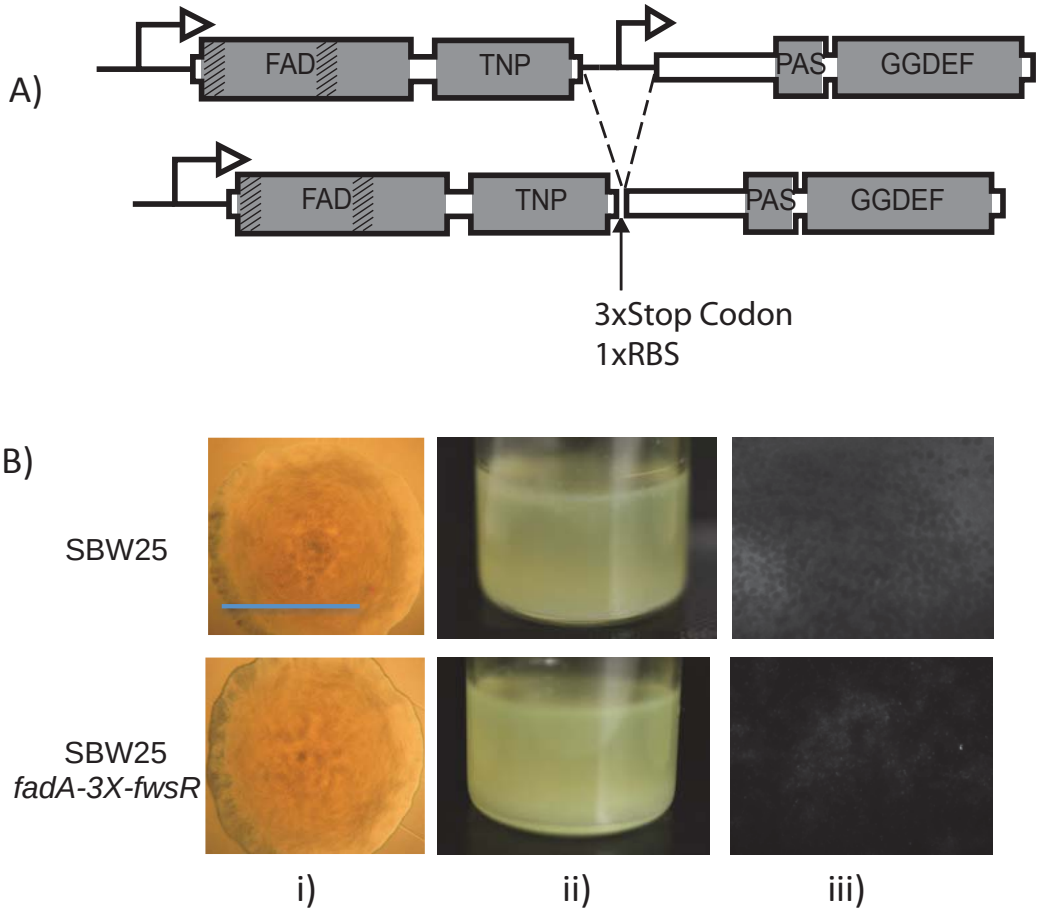

Figure 3.10: A) Illustration of the construct *fadA-3X-fwsR*. The ORF of *fadA* is fused to the ORF of *fwsR*, whilst being separated by three stop codons and the ancestral ribosome binding site (RBS) of *fwsR*. B) Transcriptional fusion and translational uncoupling of *fadA* and *fwsR* do not cause SWS. The phenotype of *fadA-3X-fwsR* reconstructed in SBW25 as compared to the direct ancestor (SBW25). Both constructs share a characteristic SM phenotype similar to the ancestor in: i) morphology; ii) broth colonisation; and iii) undetectable ACP production. For details of each image, see Figure 3.3.

3.3.4.3 Testing the hypothesis that relocalisation of the FwsR DGC causes SWS

3.3.4.3.1 Membrane localisation domains necessary for FwsR to cause the SWS phenotype

If relocalisation of FwsR is necessary for the *fadA-fwsR* mutation to cause the SWS phenotype, then fusion of alternative TMDs to FwsR should cause WS. To test this hypothesis, *fadA* was substituted for the TMD-encoding region of another gene – *mwsR* (PFLU5329) – in the context of *fadA-fwsR*. This region of MwsR has no observable amino acid homology to FadA. This region was made to replace the entire coding region of *fadA* and the intergenic region 5' of *fwsR* (see Figure 3.11A), resulting in a translational fusion of the *N*-terminal region of MwsR (residues 2 to 407) and FwsR (fused at residue 2). Ancestral *fadA* and *fwsR* was replaced with the P_{fadA}-*mwsR1218-fwsR* construct in PBR711 (SM Δ*mws*), a genotype used to prevent recombination of the construct in *mwsR*.

The resulting mutant was phenotypically WS (Figure 3.11B). This construct suggests that TMDs – encoded by *fadA* or potentially other genes – are the post-translational element necessary to cause SWS. Furthermore, this mutant confirms that the regulation of FwsR by the enzymatic activity of FadA does not cause SWS.

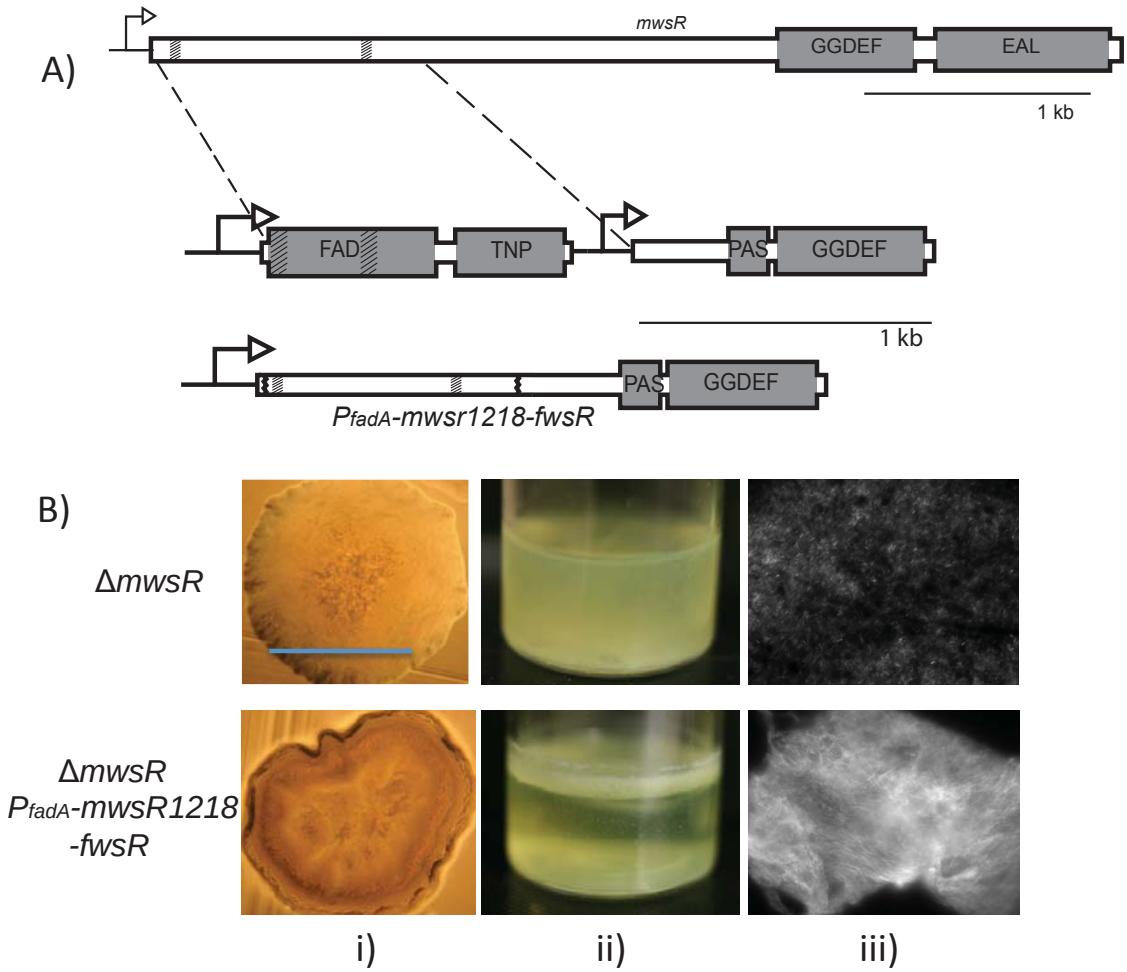

Figure 3.11: A) Illustration of the *P*$_{fadA}$*-mwsR1218-fwsR* construct. To demonstrate the lack of FadA enzymatic activity in generating the WS phenotype of *fadA-fwsR*, a version of *fadA-fwsR* was generated using the first 1218 bp of *mwsR*. The first 1218 bp of *mwsR* (which encodes for two TMD regions) was used to replace the *fadA* element of *fadA-fwsR.* Vertical grey bands within the ORFs represent TMDs. **B) The phenotype of *P*$_{fadA}$*-mwsR1218-fwsR* is phenotypically WS.** The *P*$_{fadA}$*-mwsR1218-fwsR* construct reconstructed in an SM ancestral background is characteristically WS in: i) morphology; ii) niche habitation; and iii) ACP production. For details of each image, see Figure 3.3.

3.3.4.3.2 Fluorescent tagging shows relocation of the DGC to the membrane associated with the *fadA-fwsR* fusion

To test the hypothesis that *fadA-fwsR* results in localisation of FwsR to the membrane, the cellular location of proteins encoded by *fadA*, *fwsR* and the *fadA-fwsR* fusion were visualised by creating a green fluorescent protein (GFP) translational fusion to the *C*-terminal region encoded by each gene. All fusions were cloned into the multiple cloning

site (MCS) of mini-Tn7-*lac* cassettes and were modified to contain a *Pseudomonas*-specific RBS. The cassette contained a *tac* promoter (repressed by LacI), which allowed the induced expression of the tagged genes via addition of IPTG to the growth media. Attempts to visualise GFP using native promoters of *P. fluorescens* cultured in M9 media proved unsuccessful (data not shown). The cassettes were then recombined into the neutral *attB* site of SBW25 and the integration was confirmed via PCR with site-specific primers. To reduce background fluorescence, cells were grown in M9 media substituted with 2% glycerol as a carbon source (for details of the growth, induction and microscopy, see methods and materials Section 2.2.13). Two controls were used during microscopy: a negative control of SBW25 without a *gfp* insert in the mini-Tn7-*lac* cassette; and a positive control of ancestral SBW25 containing *gfp* expressed in the integrated mini-Tn7-lac cassette. A cassette expressing *fadA-3X-fwsR-gfp* was also made to allow confirmation that the *fadA-3X-fwsR* construct allows the independent translation of FwsR despite the removed intergenic region (see Section 3.3.4.2).

Fluorescence microscopy demonstrates the fusion mutation alters the location of the DGC from the cytosol to the membrane location of FadA (Figure 3.12). The location of the fluorescent signal of cells expressing *fwsR-gfp* is visually unspecific and appears to be diffusely located within the cytoplasm (see Figure 3.12(iii)). Similarly, the *fadA-3X-fwsR-gfp* construct resulted in a dispersed fluorescent signal throughout the cytoplasm (see Figure 3.12(vi)). This shows that the break in translation conferred by the introduced stop codons returns the location of the DGC to a native cytoplasmic location.

The fluorescent signal from the different genotypes was measured to show that non-specific fluorescent signals are above the negative control (see appendix Figure 7.3.5 and Tables 7.3.6 and 7.3.7 for raw data and statistics of levels of fluorescence). These measures indicate FwsR-GFP is expressed at levels significantly above the negative control (*Welch t*-test, *p*-value = 5.86×10^{-11}), suggesting the diffuse expression of fluorescence results from the induced expression of FwsR-GFP.

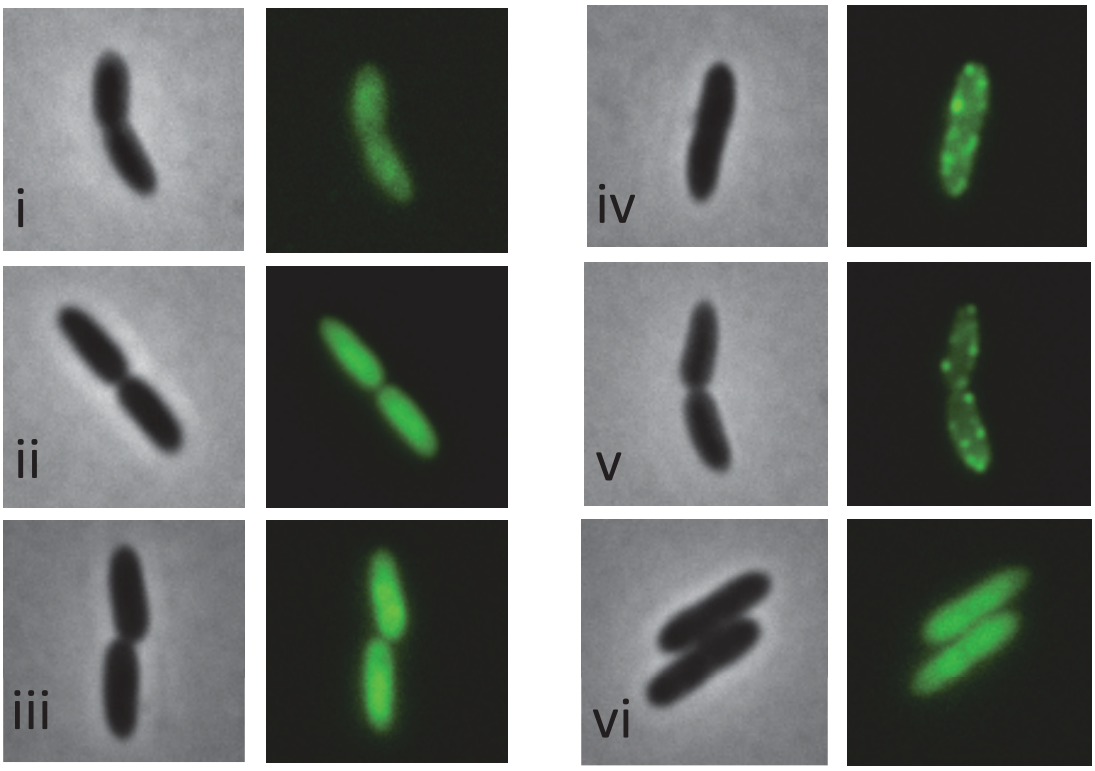

Figure 3.12: Fluorescence microscopy depicting the distribution of GFP tagged proteins encoded from the *fwsR* locus. Representative phase contrast (left) and GFP fluorescent images (right) of the subcellular distribution of fluorescence of: i) GFP(-) control; ii) GFP(+) control; iii) *fwsR-gfp*; iv) *fadA-gfp*; v) *fadA-fwsR-gfp*; and vi) *fadA-3X-fwsR-gfp*. Genotypes were grown in M9 overnight and subcultured in M9 with 1 mM IPTG for approximately 2 h.

Cells expressing *fadA-gfp* have clear foci localised to the edge of the cell, and, by inference, near the membrane (see Figure 3.12(iv)). The foci appear to be distributed predominately to the lateral edge of cells in visual reference to the phase contrast images, similar to observations of the localisation of WspA in *P. aeruginosa* (O'Connor et al. 2012). The protein expressed by *fadA-fwsR-gfp* has a localisation visually similar to that observed in cells expressing *fadA-gfp* (see Figure 3.12(v)). The *fadA-fwsR-gfp* fusion also results in fluorescent foci located to the periphery of cells. Together, the location of fluorescent foci seen in cells with induced *fadA-gfp* and *fadA-fwsR-gfp* demonstrates the *fadA-fwsR* fusion event relocated FwsR from the native cytosol to the membrane location of FadA.

An interesting observation (which is beyond the scope of this chapter) is that the intensities of foci of cells expressing *fadA-fwsR-gfp* were significantly brighter than cells expressing *fadA-gfp* (Mann-Whitney-Wilcoxon (M-W-W) test, *p*-value = 0.015). This

possibly represents the propensity of the DGC encoded by *fwsR* to cluster when active (see appendix Section 7.3.2 for further explanation and data of this observation). Such clustering has been previously observed of DGCs whilst in the active state (Huangyutitham et al. 2013).

In order to quantitatively validate the observation that cells expressing *fadA-gfp* and *fadA-fwsR-gfp* have foci located at the cell membrane, correlation analysis was performed of the fluorescent signal relative to the edges of the cells. Analysis of an image from both phase and fluorescent images from three biological replicates was undertaken for each *gfp*-fusion genotype. Approximately 100 cells were analysed per image (representing one biological replicate). Phase contrast images were used to define the edges of cells using the 'find edges' process tool available in the image-processing software FIJI (Schindelin et al. 2012). This process gives the pixels at the identified edges an increased grey value. These marked edges serve as a reference to which the corresponding fluorescent image can be tested for colocalisation using the JACoP plugin (Bolte and Cordelieres 2006). Van Steensel's approach was used to determine the colocalisation of the fluorescent and edge-marked images (van Steensel et al. 1996). This approach tests for the change in cross correlation function (CCF – a measure of the degree of colocalisation as measured by relating intensity of pixels at each position in two images), as reference and fluorescent images are moved in the x-plane relative to each another.

Van Steensel's approach confirms the observed peripheral localisation of both FadA and FwsR tagged proteins (see Figure 3.13). In the images where the fluorescent signal of the cells were visually assessed to be cytosolic, such as the negative control, the GFP positive control, *fwsR-gfp* and *fadA-3X-fwsR-gfp*, the van Steensel's distribution graphs present a maximum CCF at a pixel shift (δx) of approximately 6 to 9 and -6 to -9. These graphs also present a trough at δx of 0, providing evidence that the fluorescent signal seen in the image is not maximally colocalised with the edge of the cell. This trough further suggests that the strongest fluorescent signal is cytoplasmic. In contrast, the van Steensel's graphs for both the *fadA-gfp* and the *fadA-fwsR-gfp* genotypes depict a peak CCF at δx of 0. This demonstrates the strongest fluorescent signal for induced *fadA-gfp* and *fadA-fwsR-gfp* genotypes is correlated with the edge of the cells, and that this correlation decreases as the images are moved relative to each other. This provides evidence that the *fadA-fwsR* mutation relocates the *fwsR* encoded DGC to the periphery of the cell, most likely by integration in the inner cell membrane.

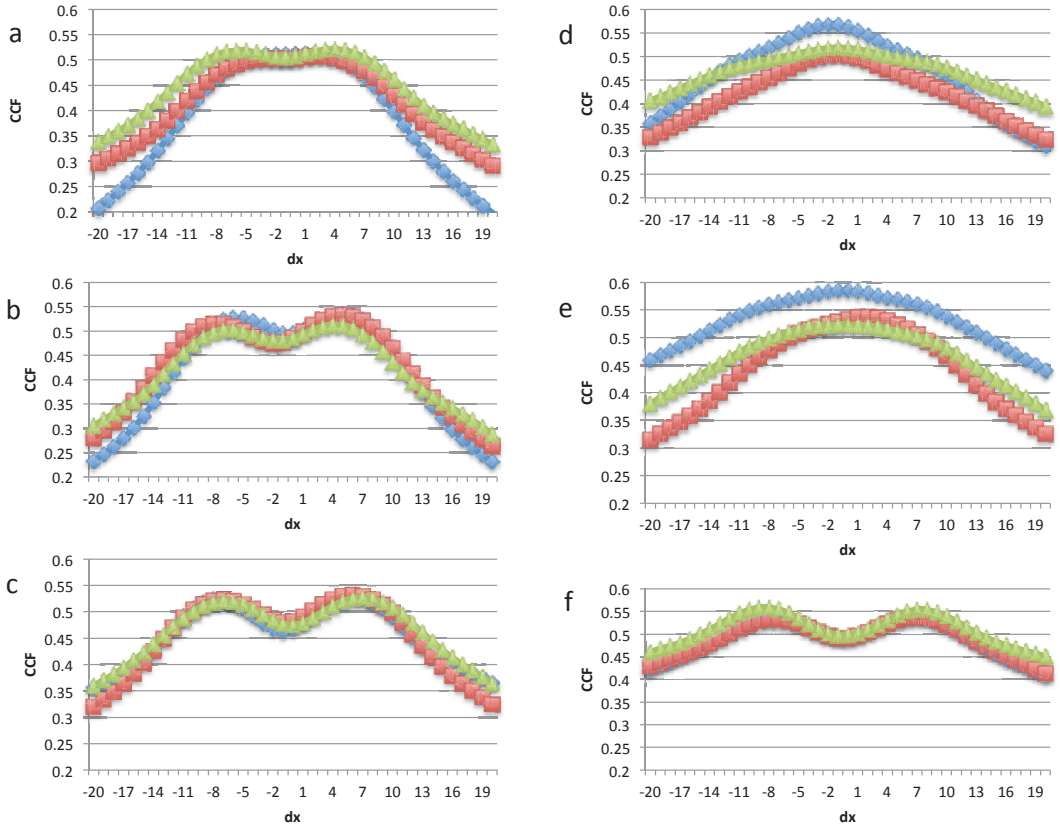

Figure 3.13: Van Steensel's distribution of colocalisation of fluorescent signal and the cell edge. The CCF represents the degree of colocalisation of the reference image (the marked edges of cells) and the fluorescent distribution of images. The van Steensel's distributions present the change in CCF as the images are moved horizontally (dx) relative to each other. Graphs represent van Steensel's distributions for: a) *gfp*(-) control; b) *gfp*(+) control; c) *fwsR-gfp*; d) *fadA-gfp*; e) *fadA-fwsR-gfp*; and f) *fadA-3X-fwsR-gfp*. Graphs a, b, c and f depict bimodal distributions indicative of a fluorescent distribution in between the reference markings of the cell edge. Distributions d and e are unimodal and the distribution peaks at approximately dx of 0, indicative of direct correlation of the fluorescent signal with the cell edge. Three biological replicates were used per genotype (one van Steensel's distribution is presented per replicate). Images were processed using FIJI software (Schindelin et al. 2012) and van Steensel's distributions calculated with the JACoP plugin (Bolte and Cordelieres 2006).

A further construct was made to confirm the importance of post-translational fusion of the DGC to the TMD-containing FadA. The 'L2-M5' mutation was predicted to provide the resulting *fadA-fwsR* protein with only a single TMD. The *fadA-fwsR* 'L2-M5' ORF was fused to *gfp* in the identical expression cassette used above. This fusion mutation was not regularly located to the periphery of the cell – most cells expressing 'L2-M5' *fadA-fwsR-gfp* translational fusions presented bright cytoplasmic fluorescent signals

(see Figure 3.14). However, a number of cells expressing these fusions presented peripherally located foci, indicating the remaining TMD is capable of causing localisation to the inner membrane, and suggesting the additional TMD may increase the stability of the DGC at the membrane location.

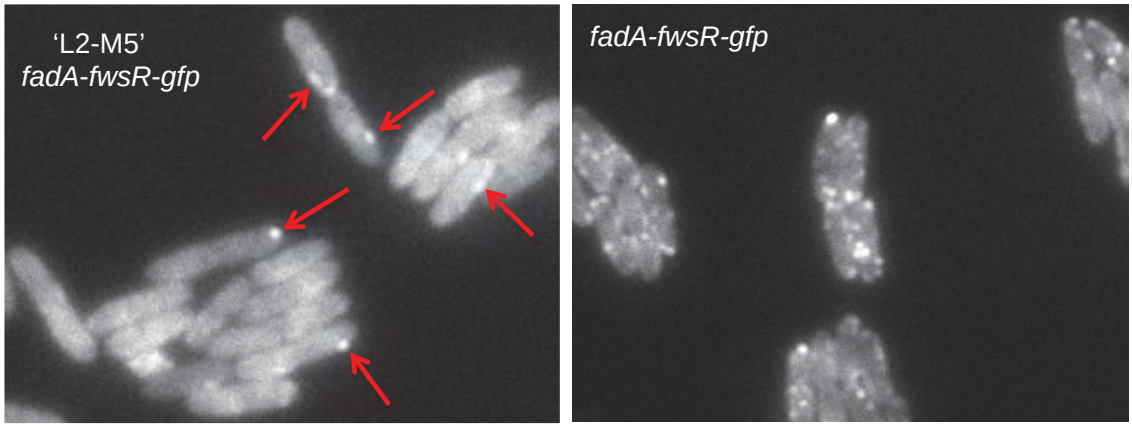

Figure 3.14: SBW25 expressing 'L2-M5' *fadA-fwsR-gfp* expresses lower frequency of foci. The 'L2-M5' mutation is a *fadA-fwsR* deletion mutation that retains only one TMD. The translational fusion bought about by the 'L2-M5' *fadA-fwsR* fusion generates a reduced number of foci, indicative of reduced stability of FadA-FwsR in the membrane. Image depicts fluorescence of the cells, which have been grown in M9 overnight and subcultured with 1 mM IPTG for approximately 2 h.

3.4 Discussion

3.4.1 Summary of results: relocalisation of FwsR to the membrane determines the WS phenotype

This chapter has provided an account of how the *fadA-fwsR* fusion mutation causes WS, and, concomitantly, explains why the ancestral *fadA* and *fwsR* locus is rarely observed to mutate and cause WS relative to *wsp, aws* and *mws*. The *fadA-fwsR* fusion mutation involves a translational fusion that combines open reading frames of both parent genes. Due to the genetically manipulable nature of bacterial model systems, the elements required from each gene in the fusion that cause the SWS phenotype have been identified. Changes in the regulation of transcription of *fwsR* resulting from promoter capture are not sufficient to cause WS – the relocalisation of FwsR to the cellular location of FadA is required. Translational tagging with fluorescent proteins provided evidence that FwsR is cytosolic. The fusion of FadA to FwsR generates a chimeric protein in which the membrane-spanning domains of FadA cause FwsR to be relocalised, most likely to the inner membrane. It appears that the native location of the DGC prevents the regular contribution of this gene to the spectrum of WS – relocalisation of FwsR requires specific fusion mutations between *fadA* and *fwsR*. In summary, the rare contribution of the *fwsR* locus in causing WS is due to a cellular constraint, in which the native location of the DGC prevents this locus causing WS – a constraint that is relieved by the *fadA-fwsR* fusion.

3.4.2 The potential contribution to evolution across loci is influenced by genetic architecture

This study sought to explain the delayed evolution of SWS compared to WS caused by mutations in *wsp*, *aws* or *mws*. The causal mutation of SWS was identified as the *fadA-fwsR* fusion, which, reconstructed in SBW25, caused a WS of similar fitness to LSWS and WS_G (see Section 3.3.1.3). This measure of fitness supported the claim of McDonald et al. (2009) that the relative fitness of SWS was not a likely cause of the delayed observation of this genotype relative to common WS.

The delayed evolution of SWS can be explained by a requirement at the causal locus for deletions that fuse the *C*-terminal region of FadA to the *N*-terminal region of FwsR, whilst leaving *fwsR* in-frame. This requirement for translational fusions was demonstrated by the loss of the WS phenotype upon introduction of stop codons that separated FadA and FwsR (see Section 3.3.4.2). The SWS phenotype also requires an active DGC encoded by *fwsR*, as indicated by the loss of WS characteristics upon degeneration of the GGEEF motif (see Section 3.3.3), which suggests the fusion must keep *fwsR* in-frame.

This requirement for *fadA-fwsR* deletions to cause translational fusions whilst leaving *fwsR* in-frame prevents two-thirds of *fadA-fwsR* deletions from causing WS. Furthermore, this deletion must be a minimum of approximately 150 bp (to ensure fusion of FadA and FwsR), and is likely to be restricted to deleting the first 300 bp of *fwsR* (to prevent mutation to the encoded domains). There are further possible constraints on the position and size of adaptive *fadA-fwsR* deletions resulting from the structure of the FadA-FwsR protein. The *fadA-fwsR* fusion is associated with the relocalisation of FwsR to the membrane location of FadA (see Section 3.3.4.3.2). In order to cause WS, these deletions may require sufficient space between the membrane-spanning domains of FadA and the tertiary structure of the *C*-terminal DGC, which may limit the maximum size of adaptive *fadA-fwsR* deletions.

This spectrum of permissible *fadA-fwsR* mutations that may cause WS is thus constricted, explaining the rarity of WS caused by mutations at *fwsR* compared to *wsp*, *aws* and *mws*. In comparison, numerous loss-of-function mutations – including deletions, indels and substitutions – in the negative-regulators encoded by *wsp*, *aws* and *mws* are sufficient to constitutively activate the associated DGC (McDonald et al. 2009). The small target size at *fadA* and *fwsR* that is capable of converting mutation into the WS phenotype provides an explanation for the low frequency at which this locus causes WS compared to the common routes (*wsp*, *aws* and *mws*).

In conjunction with the previous work this study builds upon (Spiers et al. 2002; Goymer et al. 2006; Bantinaki et al. 2007; McDonald et al. 2009), these data provide direct experimental evidence that genetic architecture guides the course of evolution. Previous work on natural populations has provided observational evidence that the production of variation is important in biasing the course of evolution (Schluter 1996). The work presented in this chapter has been able to directly show that the production of variation can strongly guide the course of evolution to a limited number of genetic

pathways, and that this variation is dependent upon the architecture of the loci underpinning adaptive phenotypes.

3.4.3 *fadA-fwsR* as a model for the evolution of new genes

3.4.3.1 The *fadA-fwsR* mutation is a model example of the creative power of modularity

The evolution of *fwsR* provides a real-time illustration of the ability for modular protein domains to fuse together to form a novel protein. Such modularity and the ability for modules to combine together – via fusion mutations or other rearrangements – are important for the diversity of protein families. This is evidenced by the rapidly increasing number of identified protein families with multiple domain architectures, compared to newly recognised domains (Levitt 2009). The promiscuity of domains, combined with the recent ability to mine sequencing data, has resulted in theoretical expectations of the mechanisms that can form new modular gene rearrangements. These arrangements include fusions, exon shuffling, frameshifts, chromosomal recombination and retrotransposition (Long et al. 2003; Kaessmann 2010; Ding et al. 2012; Ranz and Parsch 2012; Bornberg-Bauer and Mar Alba 2013).

Gene fusions provide a unique mechanism for the evolution of new genes. Fusion mutations of neighbouring genes allow the combination of distinct functional protein modules and a concomitant change in expression to result in unique protein functionality and expression. Such unique functionality can extend to novel physiology or behaviour, which may enable a new ecological niche to open, leading to adaptive radiation (Gilbert 1978; Patthy 2003; Ranz and Parsch 2012; Rogers and Hartl 2012). Chimeric genes have been identified, using primarily comparative genomic approaches, across a variety of taxa including zebrafish (Fu et al. 2010), *Caenorhabditis elegans* (Katju and Lynch 2006), *Oryza sativa* (Wang et al. 2006), *Drosophila* (Long and Langley 1993; Nurminsky et al. 1998; Wang et al. 2002; Jones et al. 2005) and mammals (Thomson et al. 2000; Courseaux and Nahon 2001), and in some cases the selective coefficients of the chimeric mutation have been identified (Rogers et al. 2010; Rogers and Hartl 2012). However, the molecular interactions and effects resulting from fusions have rarely been identified, with the notable exceptions of

functional studies of *Drosophila* genes *jingwei* (Zhang et al. 2004) and *sphinx* (Dai et al. 2008).

The *fadA-fwsR* fusion is one of the few adaptive fusion mutations that have been identified and characterised during the course of an evolution experiment. Comparable recombination events in experimental evolution studies with *Escherichia coli* have identified adaptive promoter capture events, resulting from tandem amplification events (Blount et al. 2012), or transposon insertions (Stoebel and Dorman 2010). Attempts using synthetic biology have identified the potential interacting components of chimeric proteins and cached the resulting functionalities within greater cellular interactions (Peisajovich et al. 2010). Peisajovich and colleagues identified the potential for artificial combinations of functional protein domains to provide novel interactions. *In vivo* rearrangements of both spatially determining and enzymatic domains were found in the mating signalling-network of yeast to increase phenotypic variation above gene-duplication controls.

The fadA-fwsR fusion is an example of a naturally generated modular assortment of spatial and enzymatic motifs capable of causing adaptive outcomes. Furthermore, the genetic modules that allow for spatial relocation (from the *fadA* gene) appear to be remarkably interchangeable with other membrane-locating genetic modules, as seen by alteration of the *fwsR* gene with the TMDs encoded by *mwsR*. This supports the theory that modules from separate loci can be easily interchanged, and the resulting interaction of regulatory and catalytic domains may rewire signalling pathways (Pawson and Nash 2003).

3.4.3.2 *FadA-fwsR* provides direct experimental evidence that translational fusion mutations cause adaptive change via protein relocalisation

The characterisation of the *fadA-fwsR* fusion provides direct evidence that a newly formed chimeric gene can confer adaptive benefits via the relocalisation of a protein to an alternate subcellular location. Similar fusions may also confer relocalisation to different tissues in multicellular species. Chimeric genes are able to explore the regulatory functionalities of either parent gene and thus expand, narrow or introduce novelty to the targeting of the resulting protein. In doing so, the relocalised protein is able to operate in a new context and is afforded the opportunity to interact with different proteins and thus generate new phenotypes. Most fusion mutations resulting in such

relocations will have deleterious consequences (Lobo et al. 2009; Rosnoblet et al. 2013), however in some cases these mutations will result in adaptive phenotypic change (Nurminsky et al. 1998; Byun-McKay and Geeta 2007).

The *fadA-fwsR* mutation provides a unique model to investigate the relocalisation of a protein following gene fusion. Such relocalisation, albeit to novel tissues, has been identified as resulting in a spectrum of chimeric genes in *Drosophila melanogaster* (Rogers and Hartl 2012). However, even in model organisms such as *Drosophila*, it is difficult to discern the relative effect that relocalisation has on the adaptive sweep of a given chimeric gene, compared to other effects arising from the chimeric event (such as novel interactions or the effects of altered expression). The amenable model in which *fadA-fwsR* evolves (a model bacterium) has afforded the opportunity to directly assess the phenotypic consequences of relocalisation, as well as exclude other competing hypotheses to explain the adaptive benefit of the chimeric mutation.

Furthermore, the *FadA-FwsR* fusion has provided evidence that adaptation caused by protein relocalisation can be caused by a single deletion event. This suggests more generally that neighbouring genes may provide ample evolutionary 'fuel' – given an appropriate in-frame deletion and redundancy of the constituent genes – in which to cause a change in the location of a protein.

3.4.3.3 Chimeric relocalisation as a novel mechanism behind new genes in the 'neofunctionalisation model'

The relocalisation of chimeric proteins provides a mechanism of generating new proteins from essential genes following gene duplication. Duplication is considered a main source of new genes and protein functions (Kondrashov et al. 2002). The first theorised mechanism that accounted for how duplications could result in new protein functions was Ohno's neofunctionalisation model. In this model, duplicated genes are functionally redundant and accumulate deleterious mutations until the extra copy is lost. However, in a small number of cases, mutations can confer a new function in the encoded protein of the extra gene, and selection acts to maintain the new gene (Ohno 1970). Despite the importance of this model, there are few mutational mechanisms – in particular generalizable mutational mechanisms that may apply to more than one locus – that explain how adaptive phenotypes are generated via duplication and divergence.

The relocalisation of proteins into distinct subcellular compartments has been hypothesised as a mechanism to facilitate the diversification of genes following duplication (Byun-McKay and Geeta 2007). This hypothesis has been supported by subsequent research. Approximately 37% of duplicated gene pairs in *Saccharomyces cerevisiae* encode proteins that are validated as locating to separate cellular compartments (Marques et al. 2008). Of human multi-gene families known to encode proteins predicted to locate to the mitochondria, approximately 64% of families contain a gene predicted to relocate to an alternative subcellular location (Wang et al. 2009). However, point mutations at the 5' genic region (which can encode the *N*-terminal target peptide) is the sole mutational mechanism explicitly described that would facilitate relocalisation of duplicated proteins to specific targets (Byun-McKay and Geeta 2007).

FadA-fwsR has provided an insight into an alternative mechanism to nucleotide substitutions to target peptide sequences – one that would enable the relocation of proteins to an alternative subcellular target in a single mutational step. This mechanism is a fusion mutation that reassigns the location of a protein to the subcellular or tissue target of a neighbouring gene (either through the addition of TMDs or signal peptides). This potential alternative mechanism is given plausibility by the finding that approximately 60% of duplicate gene pairs in *C. elegans* feature structural heterogeneity (Katju and Lynch 2006). Of these heterogeneous pairs, chimeric fusions can be clearly identified in approximately 38% of heterogeneous duplicates. The prevalence of chimeric genes supports a notion that relocalisation driven by chimeric fusions following duplication events may be an important force driving the neofunctionalisation of duplicated genes.

3.4.4 Implications for secondary messenger research and future directions

The discovery and characterisation of *fadA-fwsR* provides a unique model by which to view the evolution of DGCs and similar signalling systems. The protein encoded by *fadA-fwsR* is the smallest characterised DGC complex in *P. fluorescens*, and the variant types identified clearly demonstrate that the predicted FadA-FwsR protein can be as small as 441 amino acids in length and still cause the WS phenotype. This may represent a minimum assemblage of domains (encoding membrane locale, a PAS

domain and DGC domain) required to cause biofilm formation in *P. fluorescens*. However, as a minimum DGC protein, there is little by way of the regulatory units that may negatively regulate DGC activity. Perhaps a recombination event as seen in *fadA-fwsR* is the first mutational step in the evolution of the complex ACP regulators such as Wsp, Aws and MwsR. Following an initial mutational event as the chimera, it is clearly of benefit to bring the nascent signalling system under regulation. However, the ecological conditions that would allow for subsequent evolution of regulation of such a DGC are unknown.

The characterisation of *fadA-fwsR* locality suggests membrane location is critical for DGC activation of the WS phenotype, and provides an important model that may be used to further our understanding of the specificity of bacterial c-di-GMP signalling. The membrane locality for *fadA-fwsR* is consistent with studies demonstrating that membrane locality is important for DGCs active in the regulation of biofilm production across a range of species (Paul et al. 2004; Guvener and Harwood 2007; Malone et al. 2010; O'Connor et al. 2012; Shikuma et al. 2012).

However, we do not know if the relocation of FwsR to the membrane affects DGC activity. Two hypotheses have been suggested to explain the activation of the WS phenotype following relocation of the FwsR DGC to the membrane. Either: a) translocation causes a membrane-based signal to increase activation of the DGC, resulting in a greater level of cellular c-di-GMP capable of activating effectors of the WS phenotype; or b) relocation of the DGC to a subcellular location causes an increase in c-di-GMP levels at the subcellular location of targets such as WSS that are effectors of c-di-GMP signalling. This latter hypothesis is the more plausible, given that known targets such as the WSS complex are predicted to locate to the inner membrane (Spiers et al. 2002). Whilst the sensitive quantification of c-di-GMP levels afforded by the relocation event is beyond the focus of this current study, such biochemical analysis could demonstrate how the spatial location of the DGC to the membrane activates the WS phenotype.

Such a biochemical evaluation of c-di-GMP levels caused by the relocation of FwsR would provide a broader insight into our understanding of c-di-GMP signalling. The c-di-GMP signalling network is recognised as a ubiquitous means across bacteria (Ryjenkov et al. 2005) by which responses to environmental change are mediated by coordinating sessile or motile responses (Hengge 2009). However, bacterial genomes often encode for multiple DGC and PDE domains – as many as 33 GGDEF and 21

EAL domains in *P. aeruginosa*, and 41 GGDEF and 22 EAL domains *Vibrio cholerae* (Galperin et al. 2001). For this reason it is not known to what extent c-di-GMP signalling is specific for effectors that enact the appropriate phenotypic response (such as regulating exopolysaccharide production and motility), and how this specificity may be produced. There is evidence that DGC signalling is specific for certain effectors (Massie et al. 2012), and it has been suggested that signalling may be made specific by the locality of DGC proteins with targeted effectors (Jenal and Malone 2006).

However, despite the expectation that spatial localisation may increase the specificity of DGCs for target effectors, there is little direct evidence that DGCs are located near their effector targets, and at least one report suggests that DGCs are not colocalised with their target transcription effectors (Shikuma et al. 2012). Investigation of c-di-GMP levels from *fwsR* may offer considerable insight as to how the spatial relocation of a signalling DGC from one subcellular location, to a location near an effector (WSS), may increase the specificity of DGCs. If in such an investigation, no c-di-GMP level change is detected, this would provide evidence that, at least in membrane-to-membrane DGC-to-effector relationships, spatial proximity increases the activation of effectors. If c-di-GMP levels are raised, this provides insight that cytoplasm DGCs may receive an activating signal via the membrane that regulates DGC activity. Such research using the *fadA-fwsR* fusion mutant may provide unique insight to the spatial relationship of a given DGC and an effector (such as WSS), and the consequences of this relationship on coordinated cellular processes.

Chapter 4:
The identification of multiple loci that cause WS

4.1 Introduction

The discovery and characterisation of the *fadA-fwsR* fusion, in conjunction with previous work detailing the known paths to wrinkly spreader (WS) (Spiers et al. 2002; Goymer et al. 2006; Bantinaki et al. 2007; McDonald et al. 2009), have provided a rare insight into the role that genetic architecture plays in the translation of mutation into phenotypic change. However, in addition to the four pathways that have been characterised, *Pseudomonas fluorescens* SBW25 encodes many di-guanylate cyclases (DGCs) that are expected to cause WS upon mutation. The genome of SBW25 contains 29 domains containing either GGDEF or GGEEF motifs characteristic of DGC enzymes. The proportion of WS caused by mutations at the *fadA* and *fwsR* loci within the independent slow wrinkly spreader (ISWS) library (eight of 91 WS) indicates additional genes encoding DGCs – hereafter termed 'rare loci' – may mutate to cause the WS phenotype. Similar to *fadA* and *fwsR* loci, these additional rare loci may encode small regions relative to *wsp, aws* and *mws* that can translate mutation into the WS phenotype. This chapter identifies the mutational basis of WS underpinned by these DGCs, and provides evidence of the types of mutations required at these loci to generate WS.

In order to generate a spectrum of mutations at these 'rare loci', a method was employed that was previously utilised to study the adaptive potential of *P. fluorescens* (Beaumont et al. 2009). The reverse evolution experiment (REE) was initiated to investigate Dollo's Law: the extent to which a trait, once lost, was capable of re-evolving (Collin and Miglietta 2008). To investigate this hypothesis, 12 replicate lines of SBW25 were established and subject to rounds of growth in alternating structured (statically incubated) and non-structured (shaken) KB microcosms (see Figure 4.1). Upon detection of a morphologically distinct type, the evolved type was isolated via a bottleneck of a single colony and used to establish a population in the alternative environment. Specialisation in the static environment, typically through the increased production of acetylated cellulosic polymer (ACP) (associated with the WS

morphology), was costly in the shaken environment, as measured by fitness assays (H.J.E. Beaumont and C. Kost, personal communication). As a result of this fitness cost, the evolution of types in the alternative shaken environment typically resulted in the loss of the WS phenotype and the evolution of the ancestral smooth (SM) morphology. The process was then repeated, resulting in multiple bouts of reverse evolution. It is important to note the REE protocol did not exclusively select WS and SM morphotypes. Any morphotype visually distinct from the direct ancestor could be the source of the next population from each round, as long as the morphotype was the earliest arising and numerically dominant type.

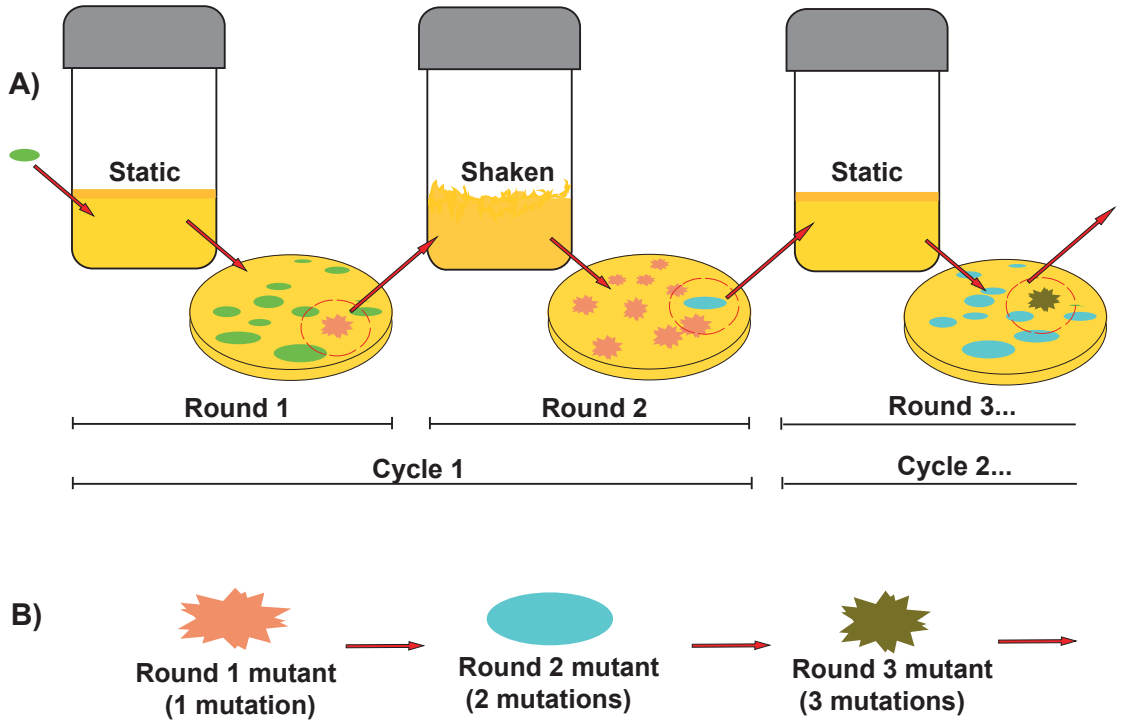

Figure 4.1: Methodology of the reverse evolution experiment. A) Individual experimental lines were initiated in static microcosms inoculated with SBW25. Microcosms were incubated and serially passaged every three days until a visually distinct morphotype was observed. This marked one 'round' of the experiment. The morphotype was then isolated and used to inoculate microcosms in an alternative environment. Types arising in this environment marked the end of a single 'cycle'. B) The experiment results in the accumulation of mutations within a lineage, of which intermediate forms have been stored, with at least one additional mutation expected to underpin each mutant (Beaumont et al. 2009).

Although the experiment was not biased to the selection of WS types, qualitative screening for different morphologies predominately resulted in the evolution of WS

types via alternating mutations to *wsp, aws* and *mws*. Following whole genome sequencing (WGS) of the endpoints of the lines, the total mutations could be identified and then temporally ordered by Sanger sequencing. Adaptive mutations were identified at the *wsp*, *aws*, *mws* and *wss* loci in 126 of the 162 mutants derived from the experiment (P.A. Lind, personal communication). The loss of WS morphotypes (during selection in the shaken environment) was typically caused by predicted loss-of-function mutations within WS-causing DGC-encoding genes or operons. Mutations in *wss* were also associated with loss of the WS phenotype.

The predominance of mutations in the common pathways of *wsp*, *aws* and *mws* of the REE is explained by the presence of negative regulatory structures at these loci, which are more readily able to convert loss-of-function mutations into adaptive WS types (McDonald et al. 2009). What are the other loci outside of *wsp, aws, mws* and *fadA-fwsR* capable of mutating and forming WS types? Do these rare loci encode small regions that may translate mutation into the WS phenotype, unlike the large negative regulators encoded by *wspF*, *awsX* and *mwsR*? Do other factors, such as mutation rate or the fitness of the subsequent mutants, result in a low rate of appearance of these types? To begin to disentangle the relative constraints of genetic architecture from that of selection across these rare loci, we must first identify the genes capable of mutating and causing WS. Identification of these rare pathways and the mutations underpinning the adaptive WS will provide insight to how genetic architecture constrains the production of WS variants at the different loci.

To identify mutations outside of *wsp*, *aws* and *mws* that cause WS via these rare routes, bouts of reverse evolution were initiated with a starting genotype of PBR716 (SM Δ*wsp* Δ*aws* Δ*mws)*. Here, eight lines were inoculated in static incubated microcosms and subjected to four cycles of reverse evolution. By conducting the experiment with the common loci removed, the selection experiment could evolve by alternative pathways outside of *wsp*, *aws* and *mws* – effectively accelerating the REE to progress via rare pathways. This experiment is referred to as the slow reverse evolution experiment (SREE). The mutational history of lines were identified via WGS of the endpoints and Sanger sequencing was used to temporally order this history. The combination of a denuded genotype (PBR716) and the sequencing of multiple lines allowed the identification of multiple rare loci in contrast to the transposon mutagenesis approach utilised to identify the *fadA-fwsR* fusion (see Chapter 3). Of particular interest were the types of mutation at genes encoding GGDEF motifs that are associated with WS types. Mutations associated with the loss of WS phenotypes were also of interest,

as these provide an indication of additional genes underpinning a particular WS phenotype.

4.2 Aims

-To use the REE protocol to generate a spectrum of morphologically distinct types adapted to the air-liquid interface.

-Find the causative mutation of each adaptive step, with particular attention to the mutation of DGC-containing loci.

-Assess to what degree genetic architecture is responsible for the hierarchical order of DGCs, as well as parallel mutations of genes not encoding DGCs.

4.3 Results

4.3.1 The SREE methodology and the subsequent identification of mutations

The SREE was conducted (as per materials and methods Section 2.2.15) using the same protocol as the original REE. This resulted in 16 evolutionary lines, eight of which were subject to alternating selective regimes (in static and shaken microcosms) and eight control lines that were grown in the same shaken environment for the same time period as the paired alternating lines. No lines were deemed to have come to an evolutionary 'dead-end', as distinct morphotypes were identified in each round within 10 successive transfers. This selection protocol resulted in 64 visually distinct morphotypes across the eight experimental lines subjected to alternating environments. For photographs of all morphotypes see Figure 7.4.1 in the Appendix. Details of the number of transfers taken to evolve adaptive types in each round are detailed in Table 7.4.2 (A-H) in the Appendix.

To identify the order of mutations along each of the eight evolutionary lines, the genome of the final genotype of each line was sequenced via whole genome sequencing (WGS) (see Sections 2.2.14 and 2.2.16 for details on genome DNA preparation, sequencing and analysis). As the final genotypes of each line were made genetically homogeneous (via streaking the population to a single CFU before storage), genomic material was extracted from overnight cultures grown directly from the frozen glycerol stock. Point mutations were determined via deviation of sequenced reads from the reference sequence. Indel mutations were unreliably predicted, but could be identified via a search for areas of low sequence coverage. All mutations within the final genotype of each line were confirmed via PCR of the mutation region and Sanger sequencing of the products (for primers and conditions used for each mutation, see Table 7.4.1 in the appendix). Sequence variants that differed from the reference in all lines were excluded from further analyses (such mutations were always found in regions of repetitive sequence).

In order to associate mutations with WS phenotypes, it was important to temporally order the identified mutations along each experimental line. PCR was used to amplify the mutational loci at different evolutionary time points using frozen glycerol stocks of

the isolated genotypes as a template. Mutations were deemed to arise in a specific round of selection if the mutation was identified in a given round and absent in the directly preceding round of selection. Details of the mutations associated with each genotype are shown in Table 7.4.2 (A-H) in the appendix, and a summary of mutations is presented in Table 4.1.

4.3.2 Identification of genotypes that were not identified via sequencing of the endpoints

WGS revealed a smaller set of mutations than the 64 mutations expected to occur in the 64 mutants. The paucity of identified mutations was previously recognised during ordering of mutations arising in the original REE (C. Kost and P.A. Lind, personal communication). In a limited number of REE experimental lines, unknown mutations were revealed to be direct reversals, particularly in *wssB* and *wssE*, which could not be identified via WGS of the line endpoints. All direct reversals involving *wss* initially occurred in shaken environments, following mutations to GGDEF-containing genes. This information was utilised to identify similar direct reversals of mutations.

To determine whether genetic reversals were responsible for the unidentified mutations of the SREE, several regions of the *wss* operon were amplified via PCR and sequenced. This included regions that had been identified during the REE. One such reversal, an indel at nucleotides 3203-3204 of *wssE*, was identified in two independent lines of the SREE (see Table 4.1). Parallelism at this locus could be explained by a suspected elevated mutation rate, as the mutation occurs in a tract of guanine homopolymeric repeats. Four more pairs of reversals were located using a combination of WGS and Sanger sequencing, revealing a genetic reversal in *wssB* in line 2 of the SREE, and a series of three genetic reversals in PFLU2764 (a GGDEF-encoding gene) in line 7 of the SREE. Only two evolutionary steps were not associated with mutations (in lines 4 and 5). However, non-detection of mutations in these genotypes cannot exclude the possible existence of genomic rearrangements that are not detectable via WGS of the genotype.

Round	Environ.	Line 1	Line 2	Line 3	Line 4	Line 5	Line 6	Line 7	Line 8
1	Static	85	*fadA-fwsR*	85	p956	4744 458	4781	p2764, 2764, 2441	p5698, 4414
2	Shaken	*wssE*	*wssB*	85	4939	*wssE*	4782	2764	*wssE*
3	Static	*nlpD*	*wssB*^{r}	4744, 458		*wssE*^{r}	956, 4858	2764^{r}	*nlpD*
4	Shaken	*nlpD*	*fadA-fwsR*	*wssE*	Δ2486-2512, p3409	p4939	4939	2764	*nlpD*
5	Static	478	*fadA-fwsR*	*wssE*^{r}	1739, 956		6004, 5308	2764^{r}	*nlpD*
6	Shaken	Δ1734-1755, Dp3358-3971	*fadA-fwsR*	5748	5034	621	*wssE*	2764	5698
7	Static	476	*nlpD*	3018	3677	4858	*mucA*	2764^{r}	*nlpD*
8	Shaken	477	*nlpD*	3016	85	*wssE*	*algU*	2764	4417

Table 4.1: Temporal order of all identified mutations arising in the SREE. All numbers represent PFLU open reading frame (ORF) designation numbers; otherwise, gene names are provided. Grey boxes represent mutants with identified mutations to ORFs with predicted GGDEF domains or the *wss* operon. The single most common locus of mutation was *wss*. Mutations occurred in the open reading frame of these putative genes, unless marked with the following: 'Δ' represents large deletion mutations; 'Dp' represents large duplication mutations; 'p' represents mutations to a putative promoter region; and '^{r}' represents direct reversals that were not identified by endpoint WGS.

4.3.3 Identity of mutations at GGDEF-containing loci and associated WS phenotypes

4.3.3.1. Mutations to DGC loci during the SREE are rare and diverse, compared to the original REE

Sanger sequencing of the mutations associated with the evolved morphotypes of the SREE uncovered comparatively few mutations at loci associated with the WS phenotype (Table 4.1). This comparison is made to the original reverse evolution experiment, in the first eight lines of the first eight rounds of which, 54 of the 57 known mutations of each evolutionary round were identified as occurring in either *wsp*, *aws*, *mws* or *wss*. By comparison, only 36 of 64 rounds of selection across the SREE involved mutations at loci encoding GGDEF domains, or the *wss* operon.

Despite the reduction in the number of mutational steps associated with loci encoding GGDEF domains, the diversity of such loci was greater in the SREE compared to the REE. Mutations to loci causing WS types during the first eight lines and rounds of the REE were located entirely within *wsp*, *aws* and *mws*. By comparison, the SREE mutational set included mutations to eight loci encoding putative GGDEF domains: PFLU0085, *fadA-fwsR*, PFLU0458, PFLU0956, PFLU4782, PFLU4858, PFLU2764 and PFLU5698.

The rarity of parallel mutations to putative DGCs limits an account for how architecture may constrain the production of variation at each locus. In some cases, only one mutation per locus was associated with the evolution of rare WS types, curbing predictions as to how mutations alter regulation at such loci and generate WS types. However, the types of mutations across these rare loci are different to the types associated with the common loci of *wsp*, *aws* and *mws*. Only in one locus, PFLU0085, were mutations located within the ORF of a putative DGC without additional mutations also arising at other loci. In all other rare loci, mutations were found in combination with mutations at other loci, mutations involved deletions involving two ORFs, or mutation occurred in intergenic regions. Analysis of these mutations strongly suggests these rare loci feature small mutational targets capable of generating WS, indicative of a lack of negative regulation at these rare loci compared to *wsp, aws* and *mws.*

4.3.3.2 Mutations identified at rare WS loci

Presented below are details of the mutations in each locus and a hypothesis of how each mutation causes the associated WS phenotype. As a reconstruction of all mutations was beyond the scope of this study, it remains uncertain whether each mutation is causative of the evolved phenotype, and it is not clear to what degree previous or secondary mutations may influence the mutational trajectory of each line.

4.3.3.2.1 PFLU0085

The locus PFLU0085 is the only predicted GGDEF domain protein with mutations in three independent lines of the SREE (see Table 4.2). PFLU0085 encodes a single predicted GGDEF domain encoded from residues 524 to 680 (Pfam *E*-value 6.9×10^{-47}) (see Figure 4.2). The GGDEF domain features a RXXD-GGEEF motif, indicative of a c-di-GMP binding site (Christen et al. 2006).

Three of the mutations in PFLU0085 are associated with typical WS phenotypes (see Figure 4.3) in terms of morphology, colonisation of the air-liquid interface, and detectable ACP production. Two of these mutations result in L439P. The predicted substitution (leucine to proline) suggests this mutation disturbs the structure of the resulting protein. Mutations immediately following the L439P mutations were located in *wssE* and the GGDEF domain of PFLU0085 (see Table 4.2), which supports the prediction that PFLU0085 regulates c-di-GMP production. The mutation to the GGDEF domain (L3-R2) is clearly associated with loss of the WS phenotype (see Figure 4.3).

Genotype		Locus	Mutation at locus	Mutation position relative to ORF	Amino acid change	Genome location
Line	Round					
1	1	PFLU0085	T>C	1316	L439P	86014
4	8	PFLU0085	T>G	1340	V447G	86038
3	1	PFLU0085	T>C	1316	L439P	86014
3	2	PFLU0085	Δ12 bp	1755-1766	I585M	86453-86464

Table 4.2: Set of mutations at locus PFLU0085 identified across the SREE. Genotypes with an odd round number arose in static microcosms. 'Genome location' is relative to the SBW25 sequence. 'Δ' represents deletion mutations.

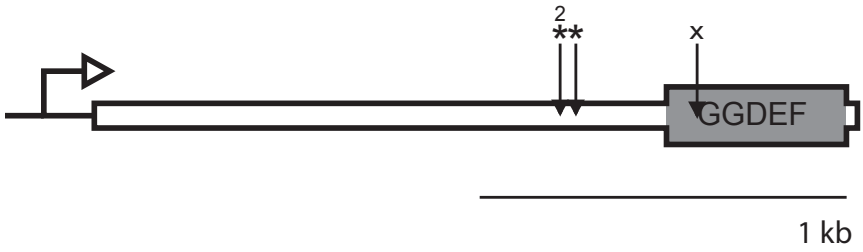

Figure 4.2: Diagram of the PFLU0085 ORF and location of identified mutation. The grey region represents the GGDEF domain containing an RXXD-GGDEF motif. * represents mutational loci associated with WS types. ^{X} represents the position of mutations associated with the loss of WS phenotypes.

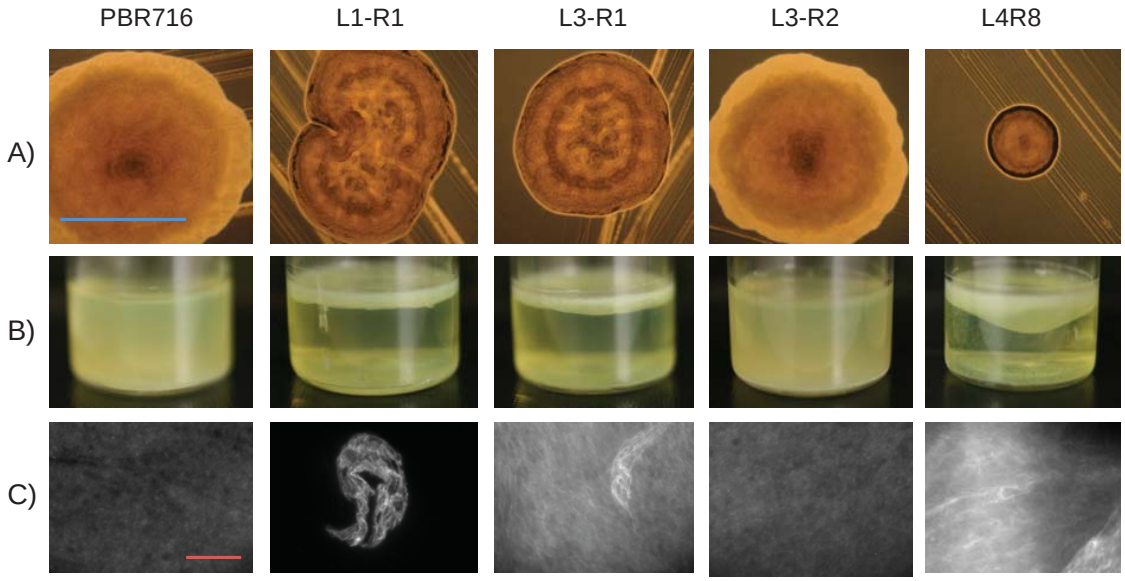

Figure 4.3: Mutations to PFLU0085 are associated with WS phenotypes. Three of the mutations to PFLU0085 were associated with the WS phenotype in: A) morphology; B) colonisation of the air-liquid interface; and C) calcofluor binding as mediated by ACP production. The L3-R2 mutant was associated with loss of the WS phenotype. The blue scale bar is approximately 2 mm; the red scale bar is 20 µm.

The third WS-associated mutation to PFLU0085 – resulting in a V447G substitution – evolved in a shaken environment. This surprising environment for the evolution of WS may be explained by historical contingency. An SM morphotype evolved in the previous round of selection in a static microcosm. This SM direct ancestor may rapidly colonise the broth of the shaken microcosm, and this pre-adaptation to the shaken

environment possibly allowed the limited niche of the shaken microcosm walls to be colonised via the evolution of WS.

The location of all three WS-associated mutations indicates the disruption of this region within PFLU0085 appears to cause WS. Additional mutations were subsequently identified at this region (P.A. Lind, personal communication). Although the structure of the region found at amino acids 439 and 447 is unknown, the identified amino acid changes are suggestive of a region of negative regulation of the encoded DGC. As this region is 5' to the GGDEF domain, mutations that disrupt this negative regulation are likely limited to point mutations or in-frame indels.

4.3.3.2.2 *fadA-fwsR*

The gene *fwsR* encodes a cytoplasmic DGC that can cause WS types following translational fusion of this gene to neighbouring *fadA* (see Chapter 3). Two fusion mutation events occur at this locus in line 2 of the SREE (Table 4.3 provides details of these mutations). The first fusion mutation (L2-R1) consists of a 218 bp deletion (see Figure 4.4B). This fusion is associated with a WS phenotype (see Figure 4.5). Following a directly reversing mutation to *wssB*, the next sequential mutant featured a nonsense mutation to the *fadA-fwsR* fusion, which resulted in an SM morphotype (see Figure 4.4B).

The following mutant (L2-R5) featured a further 672 bp fusion expanding the previous deletion (see Figure 4.4C), which resulted in the removal of the previous nonsense mutation. This mutant was associated with the WS phenotype. This second fusion mutation was confirmed to result in relocalisation of the DGC encoded by *fwsR* to the membrane (see Section 3.3.4.3.2). This fusion mutation was followed by a non-synonymous substitution mutation (L292P), which was associated with the evolution of the ancestral phenotype (see Figure 4.5).

Genotype		Mutation locus	Mutation at locus	Mutation position relative to ORF	Predicted amino acid change	Genome location
Line	Round					
2	1	PFLU0183, PFLU0184	Δ 218bp	1127-(-24)	S376fs: F-8	208965-209182
2	2	PFLU0301	dup24bp	1185-1108	dup I362-A369	329370-329393
2	3	PFLU0301	Δ24bp	1109-1132	ΔI370-A377	329370 - 329393
2	4	PFLU0183, PFLU0184	C>T	693	W231*	209616
2	5	PFLU0183, PFLU0184	Δ890bp	341(PFLU184)-82(PFLU183)	R114fs: L28	208860-209968
2	6	PFLU0183, PFLU0184	A>G	617 (875 FWS)	L292P (FadA-FwsR)	208325

Table 4.3: Set of chimeric *fws* mutations identified across the SREE. Intermediate mutations to *wssB* have been included. 'Genome location' is relative to the SBW25 genome. 'Δ' represents deletion mutations, 'dup' represents duplication mutations, '*' represents nonsense mutations, 'fs' represents frameshift mutations.

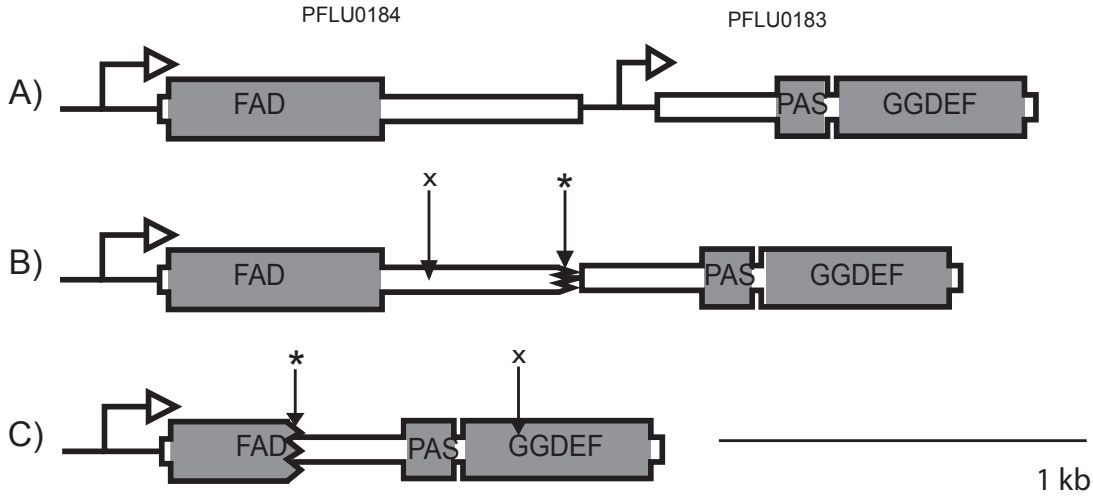

Figure 4.4: The evolution of WS types in line 2 is associated with consecutive chimeric *fws* deletions. A) Ancestral arrangement of *fadA* and *fwsR*. B) The arrangement following the first two mutations to chimeric *fwsR*. C) The arrangement following the last two mutations to *fadA-fwsR*. * represents deletion mutations associated with WS types. ^{X} represents the position of mutations associated with the loss of WS phenotypes.

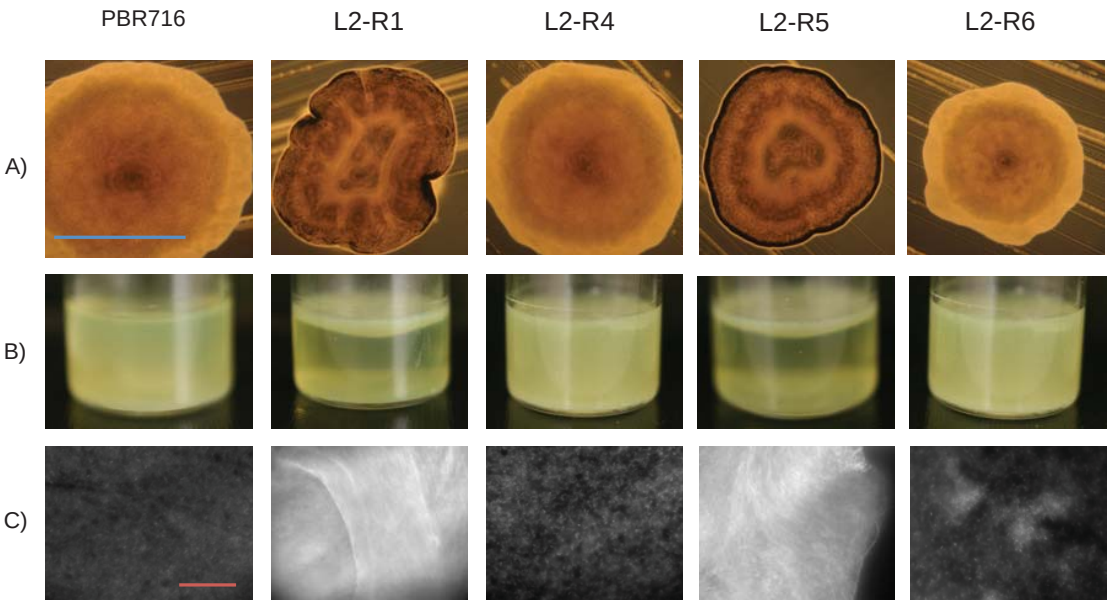

Figure 4.5: Two consecutive *fadA-fwsR* mutations are associated with WS phenotypes. Both deletion mutations resulting in *fadA-fwsR* variants were associated with WS phenotypes in: A) morphology; B) colonisation of the air-liquid interface; and C) calcofluor binding as mediated by ACP production. Mutations L2-R4 and L2-R6 were associated with loss of the WS phenotype. The blue scale bar is approximately 2 mm; the red scale bar is 20 µm.

4.3.3.2.3 PFLU0458 and PFLU4744

Co-occurring mutations to PFLU0458 and PFLU4744 occurred in two separate lines of the SREE (see Table 4.4). The pairs of mutations were found in WS mutants that evolved in static environments. No subsequent mutations occurred in these genes. The independent evolution of the mutation pairs in two lines suggests that both mutational loci are required to generate WS types (see Figure 4.7).

The evolution of WS types associated with PFLU0458 is explained, in part, by the predicted encoded domains (see Figure 4.6). PFLU0458 contains elements typical of a gene encoding composite DGC and PDE activity, including a GAF domain (residues 192 to 328, Pfam *E*-value 2.9×10^{-12}), PAS domain (residues 352 to 454, Pfam *E*-value 4.9×10^{-16}), a GGDEF domain (residues 466 to 619, Pfam *E*-value 1.7×10^{-38}) and an EAL domain (residues 640 to 875, Pfam *E*-value 8.2×10^{-69}). However, the GGDEF domain at this locus does not contain a recognisable RXXD-GGDEF motif characteristic of a GGDEF domain that may encode a functional DGC (Malone et al. 2007). PFLU0458 may thus encode a protein that functions solely as a regulator of

PDE activity (encoded by the EAL domain), with the degenerate GGDEF domain potentially encoding a region of allosteric control over the PDE (Christen et al. 2005).

Genotype		Mutation locus	Mutation at locus	Mutation position relative to ORF	Predicted amino acid change	Genome location
Line	Round					
3	3	PFLU0458	T>G	1801	T601P	513443-513443
3	3	PFLU4744	G>A	77	R26Q	5217873-5217873
5	1	PFLU0458	ΔGTCAG GTTG	1399-1407	ΔN470-T472	513829-513837
5	1	PFLU4744	G>A	97	A33T	5217893

Table 4.4: Mutational details of types with mutations with both PFLU0458 and PFLU4744 identified across the SREE. 'Genome location' is relative to the SBW25 genome. 'Δ' represents deletion mutations.

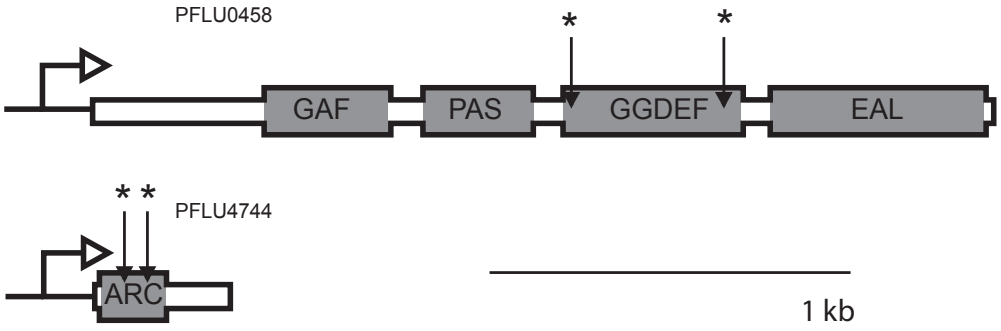

Figure 4.6: Diagram of the predicted domains of PFLU0458 and PFLU4744 and location of identified mutation. Note that only WS-associated mutations were identified at these loci. One mutation at each gene was identified simultaneously in the two WS isolates. Grey regions represent predicted domains. The GGDEF domain does not contain a predicted RXXD-GGDEF motif. * represents mutational loci associated with WS types.

The two mutations found at PFLU0458 are both located at the degenerate GGDEF domain. These mutations involve a substitution (predicted to result in T601P genotype L3-R3), and a deletion mutation (predicted to result in ΔN470-T472). Such mutations are likely to disrupt the putative allosteric activity typically encoded by a degenerated GGDEF domain (Schirmer and Jenal 2009), and decrease the function of the PDE.

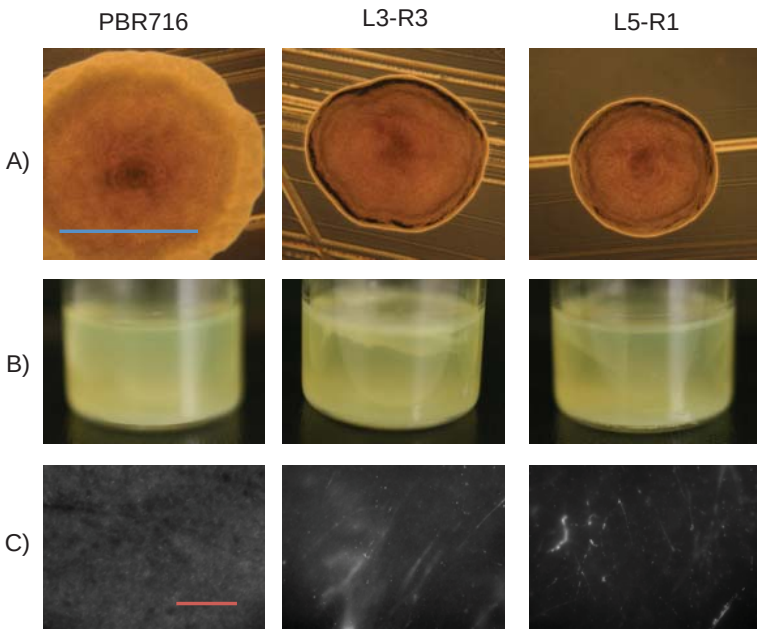

Figure 4.7: Isolates identified with mutations to both PFLU4744 and PFLU0458 are associated with WS phenotypes. These co-occurring mutations result in WS phenotypes in: A) morphology; B) colonisation of the air-liquid interface; and C) calcofluor binding as mediated by ACP production. The observable mat at the air-liquid interface and the level of calcofluor binding were visually weaker than *fadA-fwsR* mutant types. The blue scale bar is approximately 2 mm; the red scale bar is 20 µm.

The mutation to PFLU4744 offers a mechanism by which a PDE mutation may contribute to a WS phenotype. PFLU4744 contains a predicted Arc-like DNA-binding domain (Pfam *E*-value 1.9×10^{-26}), which functions as a negative regulator of *wssE* (Giddens et al. 2007). Loss of this repression following mutation to PFLU4744 is likely to result in a partial WS phenotype. Taken together, the functions of PFLU0458 and PFLU4744 suggest both mutations are complementary in the associated WS phenotype.

4.3.3.2.4 PFLU0621

A mutation to PFLU0621 occurs only once in the SREE, and is associated with the loss of a WS phenotype in a shaken environment in line 5 (see Figure 4.8). PFLU0621 encodes two *N*-terminal response regulator receiver domains (Pfam *E*-value 0.0002 and 3.2×10^{-22}) and a *C*-terminal GGDEF domain (*E*-value 5.2×10^{-40}) (see Figure 4.9). The predicted GGDEF domain features an RXXD-GGEEF motif. The mutation

associated with strain L5-R6 is a predicted L209P mutation located within the *N*-terminal response receiver domain (see Table 4.5). The introduction of the proline would likely disrupt the response regulator secondary structure.

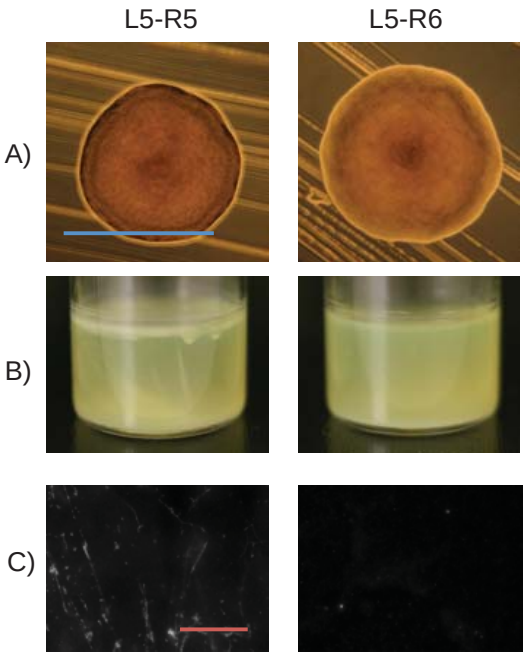

Figure 4.8: Isolates identified with mutations to PFLU0621 are associated with the loss of WS phenotypes. The single mutation identified in type L5-R6 is associated with the loss of the WS phenotype. Images of phenotypes from the L5-R5 direct ancestor are presented as comparison. The phenotype associated with the PFLU0621 is SM in respect to: A) morphology; B) colonisation of the broth; and C) lack of calcofluor binding. The blue scale bar is approximately 2 mm; the red scale bar is 20 µm.

This mutation occurs only in the shaken environment, and produces a characteristically SM phenotype (see Figure 4.9). The only mutation in line 5 in a GGDEF-containing locus, prior to the mutation in PFLU0621, was the L5-R1 mutant involving mutation to PFLU0458 (which co-occurred with the mutation to PFLU4744). This suggests that the likely loss-of-function mutation to PFLU0621 may result in an epistatic interaction that affects the WS phenotype associated with these initial co-occurring mutations. The preceding mutation to PFLU4744 may make the *wss* cellulosic machinery more sensitive to changes in c-di-GMP levels, likely affected by the mutation to PFLU0621. No other mutation occurred in this locus during the SREE, which would have provided information as to whether this locus could, by itself, mutate and result in the WS phenotype.

Genotype		Mutation locus	Mutation at locus	Mutation position relative to ORF	Predicted amino acid change	Genome location
Line	Round					
5	6	PFLU0621	T>G	625	T209P	706478

Table 4.5: Details of PFLU0621 mutation identified in line 5 of the SREE. 'Genome location' is relative to the SBW25 genome.

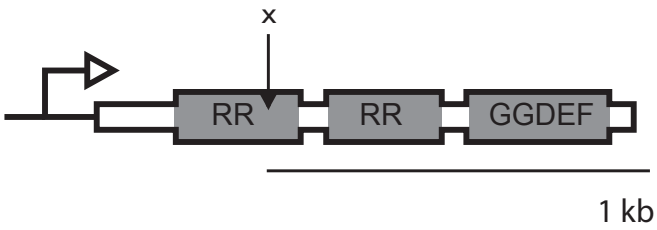

Figure 4.9: Diagram of the predicted encoded domains of PFLU0621 and the identified mutation associated with the loss of WS types. The single mutation identified at this locus is associated with a loss of the WS phenotype. RR represents a predicted response regulator domain, and GGDEF represents a GGDEF domain. ^{X} represents the position of mutations associated with the loss of WS phenotypes.

4.3.3.2.5 PFLU0956

PFLU0956 was mutated in lines 4 and 6 and is associated with the evolution of WS phenotypes (see Table 4.6 and Figure 4.10). PFLU0956 encodes a GGDEF domain (residues 332 to 490, Pfam *E*-value 6.8×10^{-49}) and no other predicted domains (see Figure 4.11). The mutation in line 4 is found directly upstream of PFLU0956 and likely increases transcription of PFLU0956. The production of a WS phenotype from increased expression of *wspR* has been demonstrated in SBW25 (Goymer et al. 2006).

The second mutation to the PFLU0956 locus that is associated with the evolution of the WS phenotype is a point mutation that causes a S21N residue change in the resulting predicted protein. This mutation arises in conjunction with a mutation in PFLU4858 (see below). It is possible the co-occurring mutations in PFLU0956 and PFLU4858 cause epistatic interactions. Without reconstruction of the PFLU0956 mutation found in

L6-R3 in an SM ancestor, it is unknown whether the S21N mutation may cause WS without mutation to PFLU4858.

Genotype		Mutation locus	Mutation at locus	Mutation position relative to ORF	Predicted amino acid change	Genome location
Line	Round					
4	1	Intergenic PFLU0956	A>C	-56	NA	1062817
4	5	PFLU0956	C>T	1057	E353K	1061705
6	3	PFLU0956	C>T	62	S21N	1062700

Table 4.6: Details of PFLU0956 mutation identified in lines 4 and 6 of the SREE. 'Genome location' is relative to the SBW25 genome.

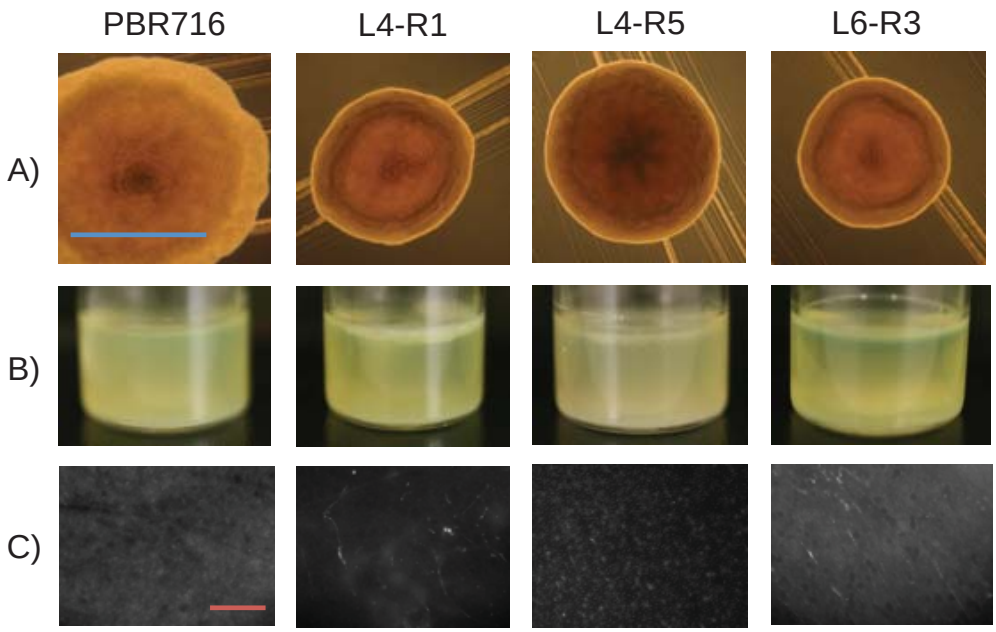

Figure 4.10: Isolates identified with mutations to PFLU0956 are associated with both the gain and loss of WS phenotypes. Both isolates L4-R1 and L6-R3 are associated with a mild WS phenotype in: A) morphology: B) colonisation of the air-liquid interface; and C) calcofluor staining indicative of ACP production. Type L4-R5 is associated with a loss of the mild WS phenotype. The blue scale bar is approximately 2 mm; the red scale bar is 20 µm.

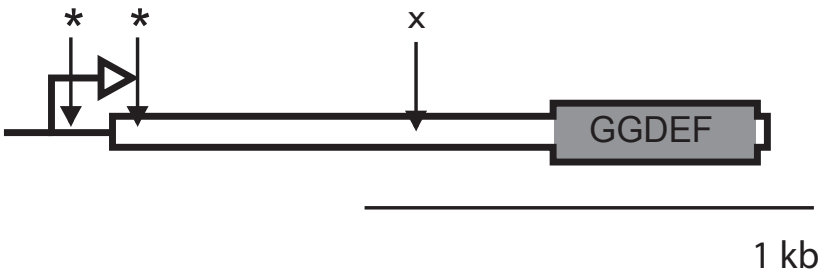

Figure 4.11: Diagram of the PFLU0956 ORF and predicted encoded domain and the identified mutations associated with the loss of WS types. The grey region represents the GGDEF domain containing an RXXD-GGDEF motif. * represents mutational loci associated with WS types. ^{X} represents the position of mutations associated with the loss of WS phenotypes.

4.3.3.2.6 PFLU4781 to PFLU4782

This locus is associated with one WS and may involve two neighbouring genes (see Table 4.7 and Figure 4.12). The transcriptional and functional relationship between the two genes is unknown. PFLU4782 encodes a predicted GGDEF domain (Pfam *E*-value 2.8×10^{-44}), featuring an RXXD-GGDEF motif. The neighbouring PFLU4781 encodes a predicted transcriptional repressor, featuring a tetR family regulatory domain (residues 15-61, Pfam *E*-value 5.3×10^{-18}), and an AefR-like transcriptional repressor (residues 146-200, Pfam *E*-value 1.2×10^{-19}).

The activating mutation associated with PFLU4781 is a 184 bp deletion, which deletes 101 bp from the *C*-terminal-encoding region of the gene, and a further 83 bp from the intergenic region of PFLU4781 and PFLU4782 (leaving the remaining intergenic region 109 bp 5' of PFLU4782). This mutation causes WS (see Figure 4.13) by loss of the predicted transcriptional regulator encoded by PFLU4781, or by capture of the promoter of PFLU4781, thereby increasing the expression of PFLU4782. The deactivation of this WS phenotype (in L6-R2) is associated with a predicted non-synonymous substitution within the predicted GGDEF domain of PFLU4782.

Genotype		Mutation locus	Mutation at locus	Mutation position relative to ORF	Predicted amino acid change	Genome location
Line	Round					
6	1	PFLU4781	Δ184bp	527-83(3')	G176fs	5263626 - 5263809
6	2	PFLU4782	T>G	575	L192R	5264492

Table 4.7: Details of the mutations to PFLU4781 and PFLU4872 identified in line 6 of the SREE. 'Genome location' is relative to the SBW25 genome. 'Δ' represents deletion mutations, 'fs' represents frameshift mutations.

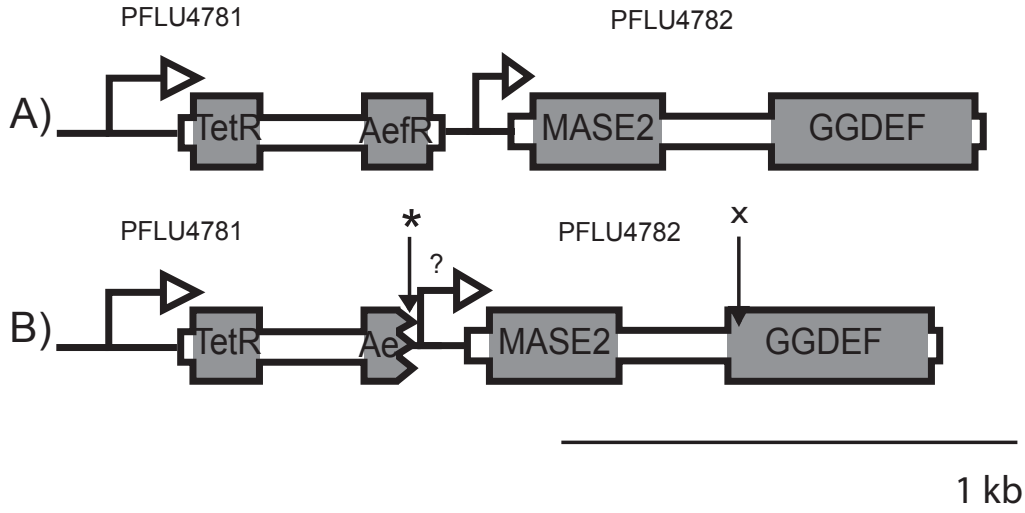

Figure 4.12: The first WS type in line 6 is associated with a deletion at PFLU4781. A) Ancestral SBW25 arrangement of genes at the PFLU4781 and PFLU4782 loci. B) The arrangement following the first two mutations to chimeric *fws*. * represents mutational loci associated with WS types. ^{X} represents the position of mutations associated with the loss of WS phenotypes. ?represents the possible promoter that may be affected by the deletion mutation.

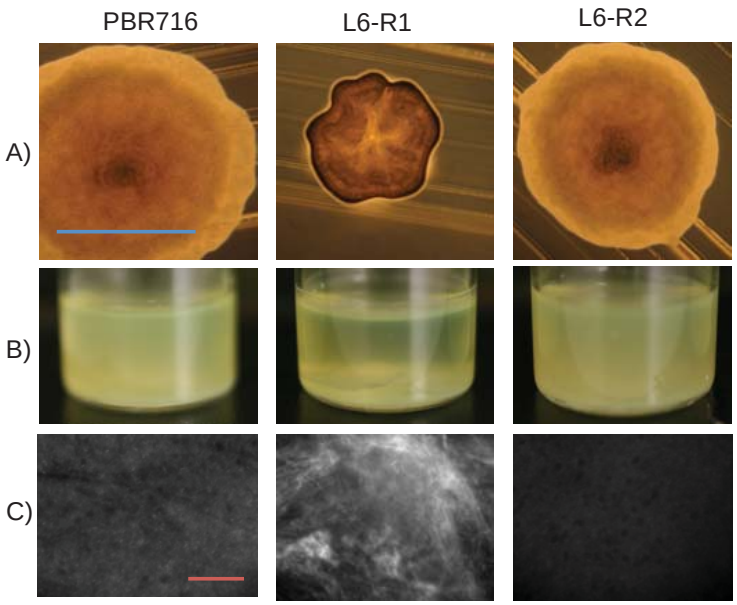

Figure 4.13: The deletion mutation to PFLU4781 is associated with the evolution of a WS type. The first mutation in line 6, a deletion mutation in a gene neighbouring a putative DGC-encoding gene, is associated with the WS phenotype in terms of: A) morphology; B) colonisation of the air-liquid interface (here the mat has collapsed); and C) calcofluor binding as mediated by ACP production. The mutation in L6-R2 is associated with loss of the WS phenotype. The blue scale bar is approximately 2 mm; the red scale bar is 20 μm.

4.3.3.2.7 PFLU4858

This gene is the target of two mutations in the SREE in lines 5 and 6 (see Table 4.8 and Figure 4.14), and both are associated with mild WS phenotypes (see Figure 4.15). The gene encodes a predicted GGDEF domain (residues 250 to 407, Pfam *E*-value 4.1 × 10^{-36}) and an EAL domain (residues 427 to 662, Pfam *E*-value 5.2 × 10^{-69}). However, similar to PFLU458, PFLU4858 contains a degenerate GGDEF motif (GGDQF) and is not expected to encode an active DGC. The mutation to PFLU4858 in line 5, associated with a WS phenotype, is a deletion resulting in a frameshift of the PFLU4858 product (including the EAL domain). This is likely to obliterate the predicted PDE activity of this locus, which could lead to an increase in cellular levels of c-di-GMP. Similar to PFLU0458, such a reduction in PDE activity is not expected to result in WS independently of mutations that increase c-di-GMP activity, and thus requires an epistatic interaction with the previous mutation in the evolutionary series (such as PFLU4744). The mutation in line 6 is an intragenic point mutation, and may require the co-occurring mutation in PFLU0956 to cause WS.

Genotype		Mutation locus	Mutation at locus	Mutation position relative to ORF	Predicted amino acid change	Genome location
Line	Round					
5	7	PFLU4858	ΔCGTG	858-861	C286fs	5335046-5335049
6	3	PFLU4858	G>T	524	G172C	5334702

Table 4.8: Details of the mutations to PFLU4858 identified in lines 5 and 6 of the SREE. 'Genome location' is relative to the SBW25 genome. 'Δ' represents deletion mutations.

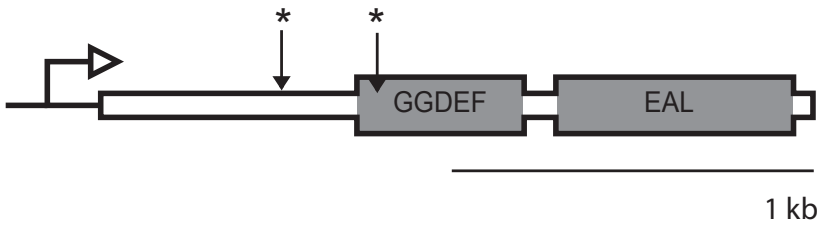

Figure 4.14: Two mutations are identified at PFLU4858 in lines 5 and 6 of the SREE. * represents mutational loci associated with WS types.

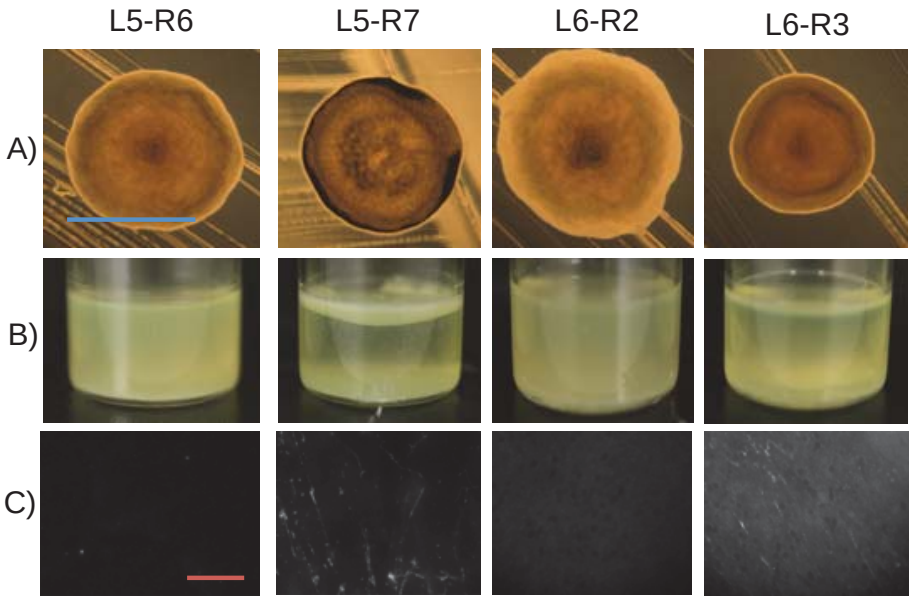

Figure 4.15: Two mutations to PFLU4858 are associated with the evolution of mild WS phenotypes. The phenotypes of mutations to PFLU4858 (isolates L5-R7 and L6-R3) are depicted in relation to their direct ancestors (L5-R6 and L6-R2 respectively). PFU4858 mutants are WS in terms of: A) morphology: B) colonisation of the air-liquid interface; and C) calcofluor binding as mediated by ACP production. ACP production appears low in relation to other mutants such as *fadA-fwsR* mutants. The blue scale bar is approximately 2 mm; the red scale bar is 20 μm.

4.3.3.2.8 PFLU2764

Despite PFLU2764 being the target of mutation in only one line of the SREE (line 7), it is the target of mutation in every round of selection (see Table 4.9 and Figure 4.16). PFLU2764 contains only one predicted conserved domain – a GGDEF domain (featuring a non-degenerate RXXD-GGDEF motif) at the *C*-terminal end of the putative protein (from residues 229 to 383, Pfam *E*-value 9.0×10^{-47}).

Genotype		Mutation locus	Mutation at locus	Mutation position relative to ORF	Predicted amino acid change	Genome location
Line	Round					
7	1	PFLU2441	C>T	446	S149F	2661314
7	1	PFLU2764	C>T	-4	NA	3056567
7	1	PFLU2764	A>T	113	Q38L	3056683
7	2	PFLU2764	dup9bp	208-216	70I-72Vdup	3056778 3056786
7	3	PFLU2764	Δ9bp	208-216	Δ70I-72V	3056778 3056786
7	4	PFLU2764	dup15bp	479-493	T160-L164dup	3057049 3057063
7	5	PFLU2764	Δ15bp	479-493	ΔT160-L164	3057049 3057063
7	6	PFLU2764	ΔT	280	C94fs	3056850
7	7	PFLU2764	insC	281-282	C94fs	3056851- 3056852
7	8	PFLU2764	A>G	665	H222R	3057235

Table 4.9: Details of the mutations to PFLU2764 identified in line 7 of the SREE. 'Genome location' is relative to the SBW25 genome. 'dup' represents a duplication event, 'ins' an insertion event, 'Δ' represents deletion mutations, 'fs' represents frameshift mutations.

The initial WS isolate (L7-R1) contains three independent mutations (see Table 4.9). One mutation arises in PFLU2441, a putative transcriptional regulator, which encodes a LysR family regulatory helix-turn-helix domain (residues 38 to 93, Pfam *E*-value 3.0×10^{-9}). An S149F amino acid substitution arises in the region encoding the LysR substrate-binding domain (residues 115 to 314, Pfam *E*-value 3.0×10^{-8}). LysR family regulators are global regulators in many bacterial species (Maddocks and Oyston 2008), and it is unclear whether this mutation may cause the associated WS

phenotype. The other two mutations arising in strain L7-R1 are both associated with PFLU2764 – a point mutation 4 bp 5' to the start codon of PFLU2764 and another point mutation resulting in a predicted Q38L substitution. The phenotype associated with this triplet of mutations is typically WS in morphology, niche and ACP production (see Figure 4.17). Without further reconstructions, it is unknown whether this locus requires all three mutations of the first round to cause the initial WS phenotype.

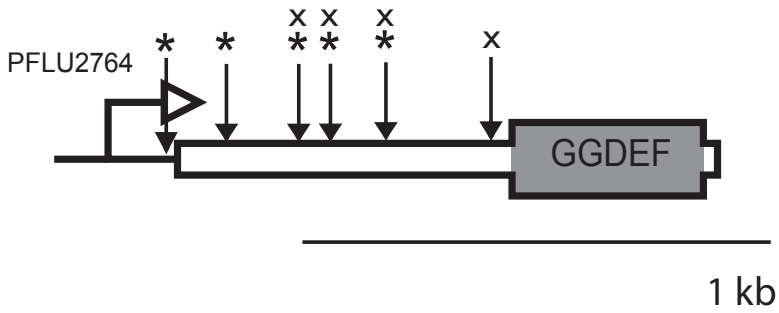

Figure 4.16: Nine mutations are identified at PFLU2764 exclusively in line 7 of the SREE. Three mutations are direct reversals of ancestral mutations and occur at three loci. * represents mutational loci associated with WS types. X represents the position of mutations associated with the loss of WS phenotypes.

The following six rounds of evolution from the evolution of this initial WS isolate involve three consecutive directly reversing mutations. The first reversal involves duplication of a 9 bp region (associated with a SM phenotype), followed by the direct deletion of that duplicate copy (returning the phenotype to WS in the second round of selection in the static environment). The second reversal is another round of duplication and deletion of a 15 bp region from PFLU2764. The third reversal involves a deletion of the nucleotide at position 280, resulting in a frameshift – a frameshift that is reversed by a compensatory insertion in the next selective round. The final mutation in the line is a point mutation resulting in a predicted non-synonymous substitution (H222R) near the GGDEF mutation, which is associated with the final SM mutant. Assuming no possible further reversals occurred in this evolutionary series, these reversals resulted in an identical genotype within strains L7-R1, L7-R3 and L7-R5. It is unknown why direct reversals of mutations occur more readily than other potential WS-causing mutations.

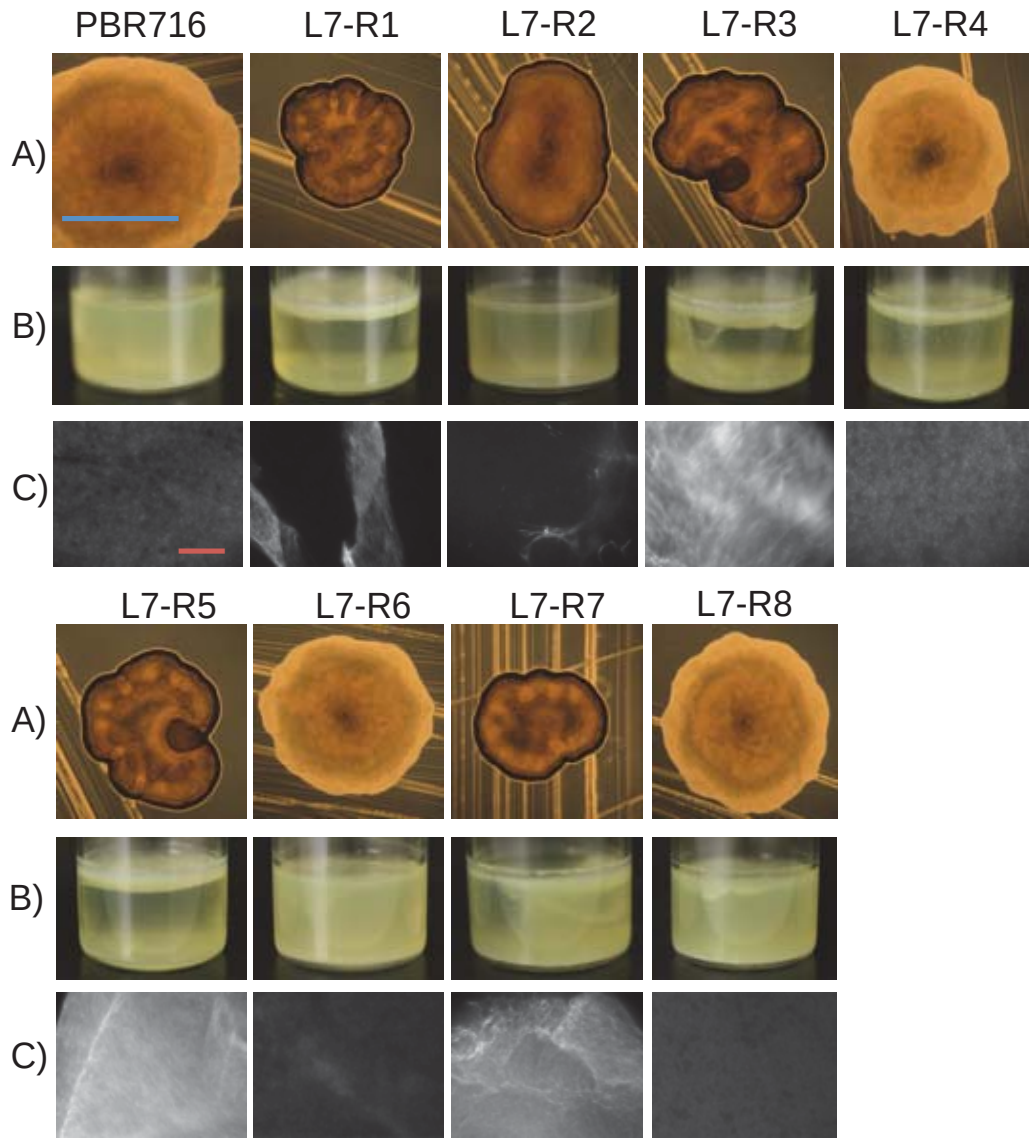

Figure 4.17: Mutations to PFLU2764 are associated with the gain and loss of WS phenotypes. Mutations to PFLU2764 arising in static microcosms (rounds 1, 3, 5 and 7) are associated with strong WS phenotypes in respect to: A) morphology: B) colonisation of the air-liquid interface; and C) calcofluor binding as mediated by ACP production. The blue scale bar is approximately 2 mm; the red scale bar is 20 μm.

4.3.3.2.9 PFLU5698

This locus was the target of mutation in line 8 of the SREE (see Table 4.10 for details). The locus encodes a putative response regulator receiver domain (Pfam *E*-value 2.8 × 10^{-25}), a PAS domain (Pfam *E*-value 3.2 × 10^{-12}), a GGDEF domain (Pfam *E*-value 2.5 × 10^{-43}) and an EAL domain (Pfam *E*-value 1.4 × 10^{-61}) (see Figure 4.18). The GGDEF

features an RXXD-GGDEF motif. The L8-R1 mutant contains a mutation to the upstream promoter region 73 bp 5' to the start codon of PFLU5698. It is likely that this mutation alters the transcription of PFLU5698. This mutation is associated with a WS phenotype in morphology, colonisation of the AL interface, and ACP production (see Figure 4.19), and subsequent nonsense mutation to this locus in strain L8-R6 obliterates the direct ancestor's WS phenotype.

Genotype		Mutation locus	Mutation at locus	Mutation position relative to ORF	Predicted amino acid change	Genome location
Line	Round					
8	1	PFLU4414	ΔTGGTGC	1118-1123	ΔL373-V374	4879311-4879316
8	1	PFLU5698	C>T	-73	NA	6249949
8	6	PFLU5698	G>C	1266	Y422*	6248611

Table 4.10: Details of the mutations to PFLU5698 identified in line 8 of the SREE. 'Genome location' is relative to the SBW25 genome. 'Δ' represents deletion mutations, '*' represents nonsense mutations.

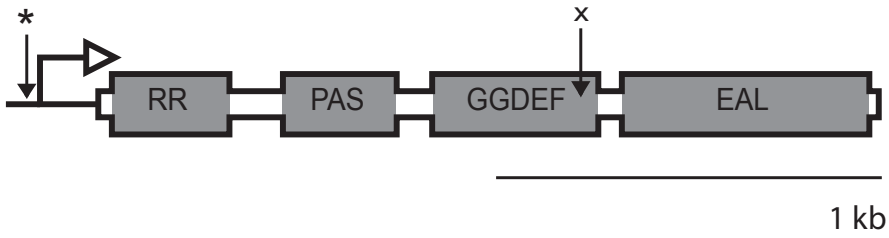

Figure 4.18: The WS-associated mutation is located at the promoter of PFLU5698. * represents mutational loci associated with WS types. ^{X} represents the position of mutations associated with the loss of WS phenotypes.

A co-occurring mutation in strain L8-R1 was also identified in PFLU4414. This locus encodes a multidomain gene homologous to *cheA1* in *Escherichia coli*. The mutation at this locus is an in-frame deletion within a histidine kinase transducing domain (residues 352 to 411, Pfam *E*-value 6.3×10^{-16}). It is unknown to what extent mutations in *cheA* homologs in *P. fluorescens* may result in WS.

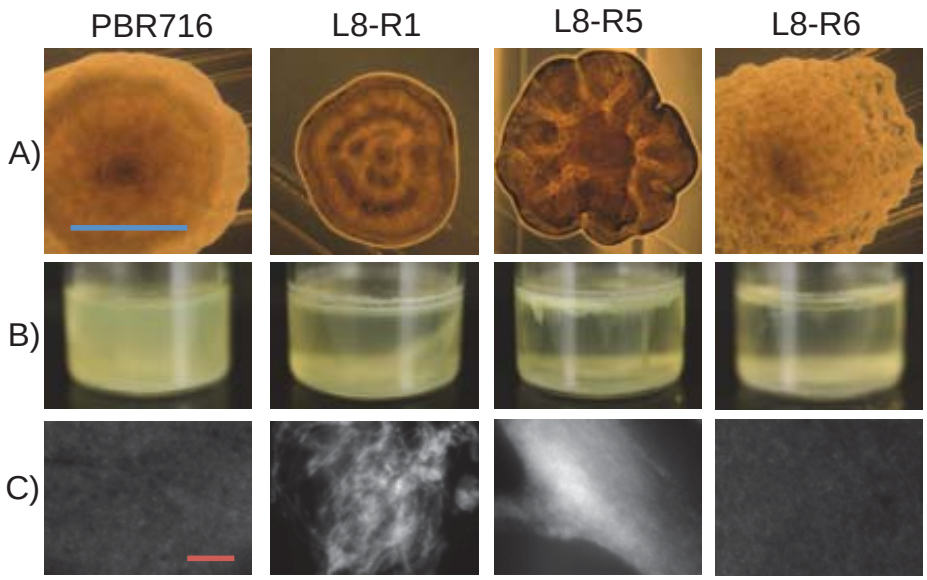

Figure 4.19: The PFLU5698 promoter mutation is associated with the evolution of a WS type. The first mutation in line 8, a substitution mutation to the promoter of PFLU5698, is associated with the WS phenotype in terms of: A) morphology; B) colonisation of the air-liquid interface; and C) calcofluor binding as mediated by ACP production. The mutation in L8-R6, a mutation to the GGDEF domain of PFLU5698, is associated with a loss of calcofluor binding (the direct ancestor L8-R5 is included as a reference). The non-smooth colony morphology of L8-R6 is likely to be related to intermediate mutations to PFLU1301. The blue scale bar is approximately 2 mm; the red scale bar is 20 μm.

4.3.4 The *wss* operon is the most frequent target of mutation in the SREE

In contrast to the original REE, which featured a high level of parallelism to DGC-encoding genes (*wsp, aws* and *mws)*, the *wss* operon was the most frequent target of mutations in the SREE. Deleterious mutations in the *wss* operon are expected, given the common regulation of the *wss* operon by c-di-GMP (for discussion, see Section 4.4.4 below). The *wss* operon is affected by seven mutational events during the SREE. Of those, three mutations were identified as reverting in subsequent rounds of selection, resulting in a return to the ancestral sequence (see Table 4.11 for details).

The identification of the mutations located in the *wss* operon revealed multiple identical mutations. Of the seven mutations to the *wss* operon during the SREE, six mutations were identified within *wssE*, and a further five mutations were located at the same region within *wssE*. Three of these mutations involved the identical 83 bp deletion

mutation also identified in the REE (located at nucleotides 3142 to 3224), and two mutations consisted of an indel mutation at a homopolymeric tract of seven guanosine nucleotides located at coordinates 3197-3203 of *wssE*. Slipped-strand mispairing provides an explanation for the two indel mutations located within the homopolymeric tract of *wssE* (Levinson and Gutman 1987). It is unknown whether this homopolymeric tract may also cause a higher frequency of the 83 bp deletions. A further mutation occurred within the *wssB* locus, consisting of a 24 bp tandem duplication mutation. This duplication, like the indel mutations located in *wssE* in lines 3 and 5, was reverted to the ancestral sequence during selection in the successive static environment. The possible causes of these genetic reversals are discussed below (see Section 4.4.4).

Genotype		Mutation locus	Mutation at locus	Mutation position relative to ORF	Predicted amino acid change	Genome location
Line	Round					
1	2	PFLU0304	Δ83bp	3142-3224	V1048fs	337085-337167
2	2^r	PFLU0301	dup24bp	1185-1108	I362-A369dup	329370-329393
3	4^r	PFLU0304	insG	3203-3204	G1068fs	337146
5	3^r	PFLU0304	ΔG	3203	G1068fs	337146
5	8	PFLU0304	Δ83bp	3142-3224	V1048fs	337085-337167
6	6	PFLU0304	C>T	2213	A738V	336156
8	3	PFLU0304	Δ83bp	3142-3224	V1048fs	337085 337167

Table 4.11: Details of the mutations to PFLU5698 identified across the SREE. '^{r}' represents mutations that undergo subsequent direct reversals back to the WT sequence. 'Genome location' is relative to the SBW25 genome. 'Δ' represents deletion mutations, 'dup' represents duplication mutations, 'fs' represents frameshift mutations, 'ins' represents insertion mutations.

4.3.5 Parallel evolution at non-DGC regulatory loci of *mvaT* and *nlpD*

The original REE featured convergent mutations to several loci encoding DGC activity – *wsp*, *aws* and mws. The potential to evolve via these pathways was removed in the SREE and the most frequent parallelism occurred at novel loci. It appears the parallel mutations occurring within the SREE are facilitated by the regulation encoded at these

loci. PFLU4939 and *nlpD* encoded predicted post-translational functionalities – particularly negative-regulation – that may elevate the production of phenotypic variants upon mutation. The analysis of mutations alone does not allow the specific causes of this parallelism to be established; however, it does allow the development of hypotheses as to the likely causes of such parallelism.

Below are details of mutational loci targeted in three independent events during the SREE (apart from *wss* and PFLU0085) and the predicted regulatory functions these loci encode. These putative functions, and the associated mutations, provide an explanation as to why these loci were the target of multiple mutations during the SREE.

4.3.5.1 PFLU4939

This locus was the subject of three independent mutations during the SREE, with mutations located in the ORF or promoter region of PFLU4939 (encoding the transcriptional regulator MvaT) in lines 4, 5 and 6 (see Table 4.12). The three mutations all arose in shaken environments and have WS types as direct ancestors. The phenotypes associated with all three mutations involved an opaque morphology of a visually similar size to the directly preceding WS ancestor (see Figure 4.20). One of the three mutations is a nonsense mutation, suggesting that all three mutations cause the opaque phenotype via loss of function of MvaT.

Genotype		Mutation locus	Mutation at locus	Mutation position relative to ORF	Predicted amino acid change	Genome location
Line	Round					
L4-R2		PFLU4939	T>G	161	I54A	5419678
L5-R4		intergenic PFLU4939	A>T	-51	NA	5419467
L6-R4		PFLU4939	G>T	31	E11*	5419548

Table 4.12: Details of the mutations to PFLU1301 identified across the SREE. 'Genome location' is relative to the SBW25 genome, '*' represents nonsense mutations.

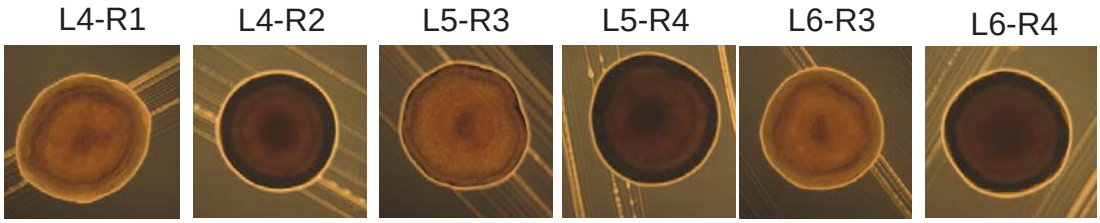

Figure 4.20: The mutations to PFLU4939 produce an opaque WS morphology. The *mvaT* mutants are strains L4-R2, L5-R4 and L6-R4, as compared with the respective direct ancestors. These strains are visually opaque compared to the direct ancestors.

The predicted function of PFLU4939 in *Pseudomonas* spp. provides insight as to how deleterious mutations to this locus may cause an opaque morphotype. PFLU4939 shares 77% amino acid identity with MvaT in *P. aeruginosa* PA01. MvaT is a transcriptional regulator with homologs across *Pseudomonas* spp., which share structural similarity to the H-NS class of transcriptional regulators (Tendeng et al. 2003). MvaT functions in *P. aeruginosa* as a transcriptional regulator of ~350 genes, including extracellular enzymes, quorum-sensing molecules and toxins (Diggle et al. 2002; Brencic et al. 2009). MvaT regulates expression by suppressed transcription of regulatory small RNAs – via binding AT-rich regions (Brencic et al. 2009). This capacity to selectively suppress expression of AT-rich regions allows MvaT to suppress the transcription of foreign DNA in *Salmonella typhimurium* (Navarre et al. 2006) and *P. aeruginosa* (Castang et al. 2008).

The function of MvaT as a negative regulator provides an explanation for the multiple independent mutations identified at this locus. Loss-of-function mutations at *mvaT* result in the increase of multiple exoproducts, and an associated mucoid morphology, in *P. fluorescens* CHA0 (Baehler et al. 2006). A similar mucoid product is associated with transposon mutants of PFLU4939 in other studies (Gallie 2010). However, it is unknown what specific extracellular products *mvaT* mutations may induce, and why these phenotypic features may provide a selective advantage in shaken microcosms (see discussion 4.4.5.2).

Interestingly, for two PFLU4939 mutants, mutations were not identified in the immediate successive mutants. However, this subsequent isolate was associated with a heritable loss of the mucoid morphology. The lack of an associated mutation may possibly be explained by an epigenetic change that alters the production of the

exoproducts determining the mucoid phenotype. Alternatively, mutational events may be associated with the loss of mucoidy that cannot be easily identified with WGS, such as genomic rearrangements including unstable duplications. The possible role of MvaT in silencing foreign DNA (Castang et al. 2008) may result in increased transposon events following deleterious mutations to MvaT. WGS of isolates L4-R3, L5-R5 and L6-R5 indicated a 2-3-fold coverage of regions PFLU5306 to PFLU5309 and PFLU2486 to PFLU2487 (see Table 7.4.2 in the appendix). PFLU5306 encodes an HTH_Tnp_1 transposase (Pfam *E*-value 1.1×10^{-14}), and PFLU5308 encodes a predicted multidomain protein, including an IS66 family transposase (residues 178 to 459, Pfam *E*-value 1.1×10^{-99}). An increase in coverage in these regions was not seen in other lines. Such an increase in coverage following the mutation to a region suspected to supress transposase expression may be due to the transposition of the region of PFLU5305 to PFLU5309.

4.3.5.2 PFLU1301 (*nlpD*)

PFLU1301 was mutated in three independent lines of the SREE and was the target of extreme molecular parallelism (see Table 4.13 for details). Three pairs of mutation, which resulted in truncation of the predicted protein product reversion of this truncation, were identified at a single region within in PFLU1301. Such parallelism is matched only by the three deletions within *wssE* (see Section 4.3.5.1).

NlpD was the target of three independent C to T transitions at nucleotide 565 of the ORF (see Figure 4.21). All C565T point mutations arose in static microcosms. This mutation is predicted to result in a truncation at residue Q189. A fourth mutation arising during selection in a static microcosm occurred in strain L8-R5, which occurred at another locus (nucleotide 709 of the *nlpD* ORF, resulting in another truncation mutation); however, this mutation occurred subsequent to a mutation to the recurring C565T mutation. Both truncations were associated with the chaining of cells at the cell poles (see Figure 4.22). The previous reconstruction of this C565T mutation in SBW25 (PBR923) has shown that this truncation mutation causes the mutant cell-division phenotype (see Figure 4.22). Time-lapse microscopy indicated that cell growth occurs at individual cells within the chain (data not shown). The associated colony morphology is SM with a subtle opacity (see Figure 4.22). All *nlpD* mutants are associated with a

subtle biofilm when grown in statically incubated microcosms. It is not known how cell chaining may provide an adaptive benefit to colonisation of the air-liquid interface.

Genotype		Mutation locus	Mutation at locus	Mutation Position relative to ORF	Predicted amino acid change	Genome location
Line	Round					
1	3	PFLU1301	C>T	565	Q189*	1439177
2	4	PFLU1301	A>C	566	*189S	1439178
2	7	PFLU1301	C>T	565	Q189*	1439177
2	8	PFLU1301	A>G	566	*189W	1439178
8	3	PFLU1301	C>T	565	Q189*	1439177
8	4	PFLU1301	G>T	567	*189Y	1439179
8	5	PFLU1301	C>T	709	Q237*	1439321
8	7	PFLU1301	A>G	710	*237W	1439322

Table 4.13: Details of the mutations to PFLU1301 identified across the SREE. 'Genome location' is relative to the SBW25 genome. '*' represents nonsense mutations.

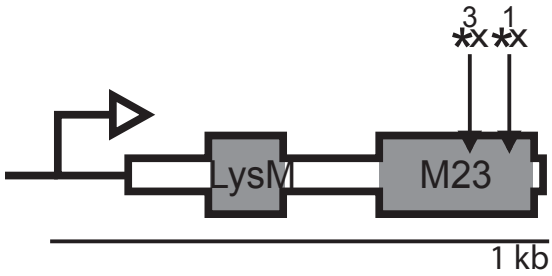

Figure 4.21: The predicted peptidase domain of PFLU1301 is the target of eight mutations, six of which were at the same codons. * Represents the first of consecutive mutation pairs, X represents the second mutation found in pairs.

The predicted function of *nlpD* readily explains the cell-chaining phenotype associated with truncation mutations within this ORF. A homologous *nlpD* gene in *E. coli* has previously been shown to encode an activating regulator of AmiC, an enzyme responsible for cleavage of the murein layer of the cell wall (Uehara et al. 2010). Insufficient activation of the AmiC enzyme, via loss of function of the NlpD activator, is thus likely to result in incomplete cell division and the formation of cell chains. Such a negative regulatory feature provides an explanation for the repeated number of mutations at this locus (similar to the normal pathways containing DGCs). However, the parallelism of mutations to nucleotide 565 of *nlpD* is strongly suggestive of a

constraint to this particular mutation. To further understand this extreme parallelism, this mutation to *nlpD* was subject to further study (see Chapter 5).

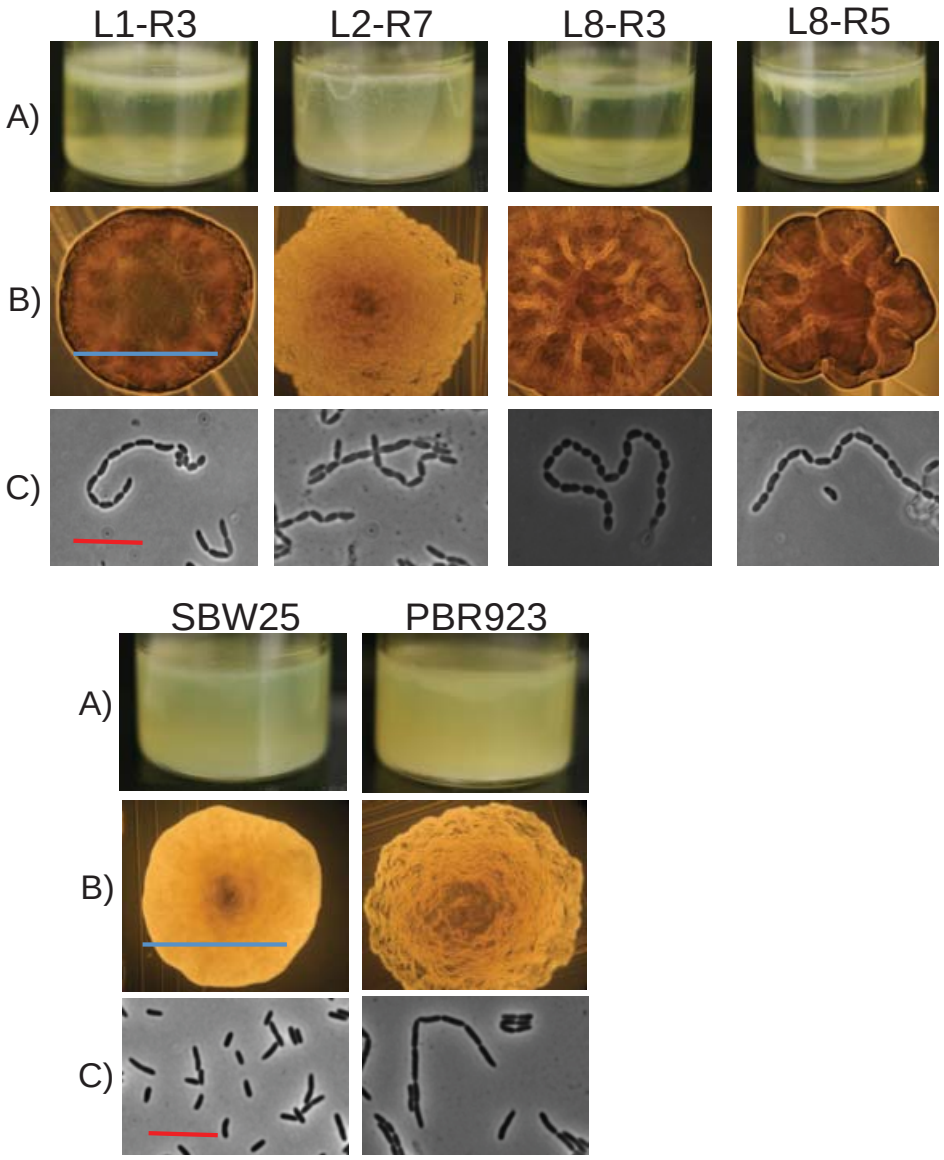

Figure 4.22: Nonsense mutations to *nlpD* are associated with deficiency in cell division. Top) Phenotypes of SREE mutant types containing nonsense mutations to *nlpD* codon 189 of *nlpD* (L1-R3, L2-R7, L8-R3) or codon 237 (L8-R5), respective to: A) colonisation of the air-liquid interface; B) colony morphology; or C) cellular morphology. There is a diversity of colony morphologies, indicative of interactions with other mutations, in most cases controlling DGC activity. Bottom) The C565T mutation has been previously reconstructed in SBW25, and the phenotype of this type shows the effect on phenotype of the mutation without further epistatic mutations of cellular morphology. The blue scale bars are 2 mm; the red scale bars are 5 µm. Cellular microscopy was performed using a 100x objective.

4.4 Discussion

4.4.1 General summary of putative DGC-regulatory mutations

This chapter has demonstrated that multiple routes to WS exist outside of *wsp*, *aws, mws* and *fwsR.* These routes were identified through parallel bouts of reverse evolution initiated with the PBR716 (Δ*wsp* Δ*aws* Δ*mws*) genotype. A lower frequency of putative DGC-encoding targets were identified compared to the original REE. Approximately 54 of 64 identified mutations from the first eight rounds and lines of the REE contained mutations to c-di-GMP regulators (*wsp*, *aws* or *mws)*. By comparison, 37 of 64 mutations of the SREE were identified at putative c-di-GMP regulators. However, despite this lower frequency, at least 10 loci were identified that are likely to regulate c-di-GMP production. This represents a far greater diversity of loci than that previously identified as capable of causing WS types (McDonald et al. 2009).

The rarity of mutations at each identified DGC locus limited insight of the genetic architectures regulating these DGCs. In previous studies, mutational parallelism has identified post-translational negative regulatory regions encoded by *wsp*, *aws* and *mws* as being readily able to produce adaptive changes upon mutation (Bantinaki et al. 2007; McDonald et al. 2009). As a consequence of these regulatory regions, most mutations of the original REE involved mutations at these loci*. C*ompeting explanations may account for the lower frequency and greater diversity of these mutant types compared to the frequency of *wsp*, *aws* and *mws* mutants in the REE. It is possible that the fitness of mutants caused by these rare pathways – compared to the fitness of PFLU4939 or *nlpD* mutants – prevents the greater frequency of repeated evolution by these rare pathways. Alternatively, low mutation rates within these rare paths (compared to within *wsp*, *aws* and *mws*) may reduce the available mutants that can be subject to selection in static microcosms. However, the reconstructions and fitness assays required to disentangle the relative influence of selection in determining the relative frequency of these rare WS mutants in the SREE await future study.

Despite the paucity of mutation to each locus, the mutational spectrum and predicted function of these rare loci suggest that genetic architecture constrains the production of variation relative to the common routes. Most of these GGDEF domain-encoding targets, with the possible exception of PFLU0085, do not appear to cause WS with

single loss-of-function mutations. The distinctive mutational spectra identified across these rare loci strongly suggest a requirement for gain-of-function mutations to activate WS at these loci. This suite of rare mutations further support the findings demonstrated in Chapter 3: that genetic architecture can bias the frequency in which mutation can be translated into phenotypic change. For the rare paths to WS identified in the SREE, the spectrum of mutations at each locus appears to be greatly limited compared to the typical routes to WS. These potential adaptive types are seen only via mutations at promoter regions, by rare fusion events or by epistatic interactions with preceding or co-occurring mutations.

4.4.2 Mutational set at PFLU0085 suggests post-translational negative regulation encoded at this locus

One of the nine mutated loci encoding predicted regulators of c-di-GMP, PFLU0085 is unique in that it may encode a region of negative regulation. Only PFLU0085 shared the characteristics of mutations found in *wsp*, *aws* and *mws* in the REE. Mutations to PFLU0085 were not associated with secondary mutations at other loci, and were exclusively identified within the ORF. These point mutations are similar to mutations that occur to known negative regulators in *wspF*, *awsX* and *mwsR*, and likely change the structure of the predicted protein encoded by PFLU0085. Although few in number, these mutations are suggestive of negative regulation encoded by PFLU0085, possibly at the post-translational level. This finding was confirmed in another study featuring a larger spectrum of mutations affecting PFLU0085, in which a large number of in-frame deletions and point mutations at residues ~400 to 500 result in WS phenotypes (P.A. Lind, personal communication). The identification of intragenic deleterious mutations at regions associated with WS is strongly suggestive of a region of negative regulation encoded within PFLU0085 that controls DGC activity.

The structure of PFLU0085 explains why this gene is not a common cause of WS. The order of regulators encoded in PFLU0085, and the predicted structure of the protein, contrasts to the common pathways to WS. PFLU0085 encodes a predicted central region of negative regulation and consists of a single gene. Both *wsp* and *aws* are arranged in operons, with regions of negative regulation (*wspF* and *awsX*) in separate genes from the encoded DGCs. Loss-of-function mutations via frameshifts/truncations

or altered protein structure in the negative regulators of *wspF* and *awsX* are unlikely to reduce DGC activity, given the separation of these regions into distinct proteins.

This separation of the negative regulatory region to the GGDEF domain is similar to the arrangement in *mwsR*, in which the EAL domain is the last encoded domain. Positioning of the EAL domain to a *C*-terminal region and within a larger protein may enable a spectrum of loss-of-function mutations to be permissible without affecting the stability of the GGDEF domain encoded 5' to the EAL domain. In contrast to *wsp*, *aws* and *mws*, PFLU0085 is encoded on a single gene (of ~2 kb in length), and the encoded region of negative regulation is in the middle of this gene. The small size of this mutational target, and the position of this region transcriptionally upstream of the predicted GGDEF domain, may limit the spectra of mutations at this region to those that do not affect the structure or function of the GGDEF domain. The presence of a post-translational region of negative regulation encoded by PFLU0085 awaits further experimental validation.

4.4.3 The remaining putative c-di-GMP regulators require co-occurring mutations, fusions or promoter mutations to cause WS

The majority of DGC-encoding genes associated with WS during the SREE had not previously been found to cause WS. Aside from mutations in PFLU0085, intragenic mutations at these rare loci are associated with additional mutations at other loci, or rare deletion mutations spanning neighbouring genes, which explains the rare observation of WS types caused by mutations to these genes. Intragenic mutations occurring at PFLU0458, PFLU4858, PFLU0956 and PFLU2764 appear to require the presence of additional mutations to cause WS types. Both PFLU0458 and PFLU4858 are unlikely to contain functional DGC-active sites, and thus required epistatic interactions with functional DGCs to cause WS. The intragenic mutations to PFLU0956 and PFLU2764 are associated with mutations either in other loci such as PFLU4858 or the promoter region of PFLU2764 respectively. Assuming these putative DGCs require such additional mutations to cause WS, such double mutations would be rare compared to the single mutations required at the common loci. This provides an explanation as to why loci such as PFLU0458, PFLU4858, PFLU956 and PFLU2764 have not been previously associated with WS evolved from SBW25.

Mutations to the intergenic operator regions of genes encoding predicted GGDEF domains are associated with the evolution of the remaining set of WS types. Single SNP mutations associated with the evolution of WS occur within 80 bp of the predicted start codon of PFLU0956, PFLU2764 and PFLU5698. In considering both mutation locations and the presence of a conserved RXXD-GGDEF motif in each gene, it is likely that such mutations confer an increased level of transcription. Quantitative PCR analysis of mutations to the upstream regions of PFLU0956 and PFLU5698 identified a 17- and 15-fold increase in transcription of these genes respectively (P.A. Lind, personal communication).

Large deletions involving the intergenic operator regions occurred at PFLU4782 and *fwsR*. The requirement for a spectrum of deletions resulting in in-frame translational fusions has been demonstrated for the *fadA-fwsR* locus (see Chapter 3). However, no clear explanation is provided for the evolution of WS via the mutation upstream of PFLU4782, and mutations to this locus have not been identified in any other studies. At least in the case of the *fadA-fwsR* fusion, this target is clearly limited in size compared to the large regions of negative regulation in *wsp*, *aws* and *mws*, providing an explanation of the rarity of this gene as a target, both in the REE and in other cases of experimental evolution of SBW25 in static microcosms. The second grouping of rare pathways (PFLU0956, PFLU2764, PFLU5698, PFLU0183 and possibly PFLU4782) thus represent targets that appear to require mutations to a smaller target, either increasing transcription, or by causing translational fusions.

In summary, these rare pathways, in contrast with the readily trod pathways to WS, appear to require specific mutations or combinations of mutations to cause WS types. The specificity of these mutations are likely to be of a lower probability of occurring compared to loss-of-function mutations to *wspF, awsX* or *mwsR,* accounting for the low observed frequency of these WS types. The SREE thus provides a broad indication that genetic architecture across multiple DGC-encoding genes biases the frequency of production of WS types.

4.4.4 The *wss* operon is the source of the greatest parallelism in the SREE

The molecular functions of the proteins encoded by *wss* provide a likely target of mutation during the SREE and are the targets of multiple parallel mutations. These

repeated mutations are explained by both the functionality and target size of the *wss* operon. The product of *wss* expression – ACP – is a requirement for the WS phenotype (Spiers et al. 2002; Spiers et al. 2003). Furthermore, the length of the operon, at ~10 kb (compared to 1-2 kb of an average rare DGC-encoding gene), provides ample target size for deleterious mutations that may 'deactivate' the WS phenotype.

Surprisingly, such deleterious mutations to the *wss* operon are associated with only seven transitions in the SREE (three of these mutations are reverted back to ancestral sequence in the successive evolutionary step). The limited set of mutations to this large target may be explained by pleiotropic effects. Mutations to the *wss* operon may cause a decrease in cellulose production and a return to the SM phenotype (Spiers et al. 2002). However, deleterious mutations to *wss* are not expected to affect c-di-GMP levels that have been altered in the ancestral WS type, which have a large number of potential consequences (Knight et al. 2006). The deleterious pleiotropic consequences of c-di-GMP up-regulation may decrease the relative fitness of a *wss* mutant in a shaken environment. Thus, a deleterious mutation to the c-di-GMP regulator that was mutated in the direct ancestor is likely to produce a fitter mutant that is selected over *wss* mutants. Therefore, the target size of *wss* may be balanced by selection in determining the frequency in which evolution proceeds via this pathway. However, the consequences to fitness of activated c-di-GMP regulators in the context of a *wss* mutation awaits further investigation.

4.4.5 Parallel evolution during the SREE is associated with negative regulators

In strong contrast to the REE, the most frequent targets of mutation outside of *wss* or PFLU0085 in the SREE were not genes encoding putative DGCs. The removal of *wsp*, *aws* and *mws* appears to have increased the relative frequency of non-WS types arising from this selective regime. These types were either unique to the SREE, as in the case of mutations to PFLU4939, or occurred at a higher frequency than in the REE, as in the case of the mutation to *nlpD*.

Similar to the common pathways in the REE – *wsp, aws* and *mws* – these convergent pathways in the SREE appear to cause an adaptive trait following a loss-of-function

mutation. This provides a possible example of loci, additional to *wsp*, *aws* and *mws,* at which encoded regulation increases the translation of mutation into adaptive phenotypic change. An explanation of the functionality of each pathway, and how this functionality interacts with deleterious mutations to provide adaptive outcomes in the SREE, is presented below.

4.4.5.1 Parallelism at locus PFLU4939 is associated with loss-of-function mutations

Mutations to PFLU4939 arose in three separate lines of the SREE and in all cases arose in shaken environments following mutations to c-di-GMP regulators. As described in Section 4.3.5.1 above, this locus encodes a transcription factor homologous to MvaT - a regulator for multiple exoproducts across *Pseudomonas* spp. The identification of three independent mutations at this locus, one of which is a nonsense mutation, suggests these three mutants all involve the loss of function of MvaT, which results in the increased expression of these exoproducts. The regulatory nature of the regulator encoded at this locus appears to make this gene a target of repeated deleterious mutations. This again appears to provide an example of the role of negative regulation in biasing the rate at which loci cause phenotypic change.

Mutations to PFLU4939 were not identified in the original REE. This is surprising given that these mutants arose in environments similar to that which *wss* mutants evolved in both the original REE and the SREE. As with mutations to *wss*, mutations to PFLU4939 arose in shaken environments with WS types as direct ancestors. The adaptive advantage of mutations in *mvaT* in the SREE – as compared to the REE – can be explained via three, non-exclusive, mechanisms.

1) Molecular epistasis. The mutated product of PFLU4939 may produce specific epistatic interactions with preceding mutations, which may increase the fitness of the PFLU4939 mutants. These interactions result either from the three deletions in the ancestral genotype, or from mutations that have arisen during the course of the SREE. This is supported by the identification of PFLU0956 as the direct ancestor of two PFLU4939 mutants. MvaT may act as a transcriptional regulator of DGCs such as PFLU0956, in which case loss of function of MvaT may be adaptive via reduction of expression of these DGCs. This hypothesis may be interrogated by introducing the PFLU4939 mutation in genotypes with and without DGC mutations (including mutations

of PFLU0956). The fitness of genotypes featuring PFLU4939 mutations in combination with PFLU0956 mutations should be significantly higher than the fitness caused by PFLU4939 mutations in combination with other DGC mutations.

2) Target size of the ancestral WS-causing genes. WS types may be reverted to SM by mutations at non-GGDEF domain loci that regulate the WS phenotype (such as *wss*) or by the mutation of genes that reduce the phenotypic changes caused by c-di-GMP production (here called 'masking mutants'). In the genotype in which PFLU4939 mutation arose, the preceding GGDEF domain-encoding genes have relatively small target sizes. This small size may preferentially increase the chance evolution of 'masking mutants' such as *mvaT* mutants, compared to mutants that target the preceding GGDEF domain-encoding genes. In lines 4 and 6, mutations to PFLU0956 were identified in the direct ancestor to the PFLU4939 mutation. PFLU0956 is a relatively small gene (1494 bp) compared to the operons such as *wsp* (~8.4 kb), *aws* (~2.3 kb) and *mwsR* (3.9 kb). In line 5, the preceding mutations are loss-of-function PFLU458 and PFLU4744 mutants – these genes are thus unlikely to be the target of further reverting mutations. The small target size of these preceding WS-causing genes may reduce the rate at which loss-of-function mutations to these c-di-GMP regulators arise in shaken environments. This reduced target size may allow the evolution of alternative mutants, via mutations to targets such as PFLU4939 that may 'mask' the effects of the WS phenotype.

3) The relatively high fitness of ancestral WS types in shaken environments. The ancestral WS prior to PFLU4939 mutants may exhibit a high degree of fitness in the shaken environment (compared to *wsp, aws* or *mws* mutants). A reduced cost of the WS phenotype in the shaken environment will decrease the frequency in which loss-of-function mutations are selected and may also result in the evolution of alternative phenotypes that increase fitness in the shaken environment, without affecting cellulose production. In such a historical context, PFLU4939 mutants may arise that would not be adaptive with a preceding mutation to *wsp, aws* or *mws*. This hypothesis could be investigated by comparative fitness assays of different WS types, including PFLU0956, when grown in shaken microcosms.

Without further detailed work exploring the molecular interactions and the phenotype caused by the PFLU4939 mutations, such hypotheses explaining the repeated evolution of PFLU4939 mutants remain untested.

4.4.5.2 Accounting for parallelism in the *nlpD* gene

The propensity for the *nlpD* gene to affect a phenotype via a loss-of-function mutation provides an explanation as to why this gene is the target of mutation in multiple instances in the SREE. It is interesting to note that *nlpD* mutants arose in two of the three cases following deleterious mutation to the *wss* operon, suggesting that these cell-division mutants have an increased chance of arising provided an absence of the WS type. This is consistent with the REE, in which *nlpD* mutation arose in five independent lines, with deleterious mutations to the *wss* operon occurring in the ancestors of four of these *nlpD* mutants. However, despite *nlpD* appearing as a target for loss-of-function mutations, there is extreme molecular parallelism at this locus, with mutations occurring in three independent instances in the SREE at an identical nucleotide (C565T). The only alternative loss-of-function mutation to *nlpD* was found in line 8, and only after a mutation to C565T had previously occurred. This strongly suggests a secondary feature biases either the mutation rate to this region or constrains the possible spectrum of adaptive mutations to this particular locus. Given the highly unusual nature of this molecular parallelism, this mutation was the subject of further study (see Chapter 5).

Chapter 5:
The causes of parallel molecular evolution of *nlpD* mutations

5.1 Introduction

Identical mutations in *nlpD* occurred in multiple independent lines during the reverse evolution experiment (REE) and slow reverse evolution experiment (SREE) (see Chapter 4). These mutations provide an example of parallel molecular evolution[6] – identical or very similar genetic changes occurring in independent lineages. In the eight independent lines in which *nlpD* was first mutated, a cytosine to thymine transition occurred at base pair 565 of *nlpD* – hereafter referred to as C565T.

Such molecular parallelism is not unique – many examples of parallel evolution are known to be caused by identical nucleotide (Gerstein et al. 2012; Manske et al. 2012; Zhen et al. 2012) or amino acid substitutions (Ffrenchconstant 1994; Wichman et al. 1999; Davies et al. 2012a; Davies et al. 2012b; Shen et al. 2012). These examples of molecular parallelism have been interpreted as resulting from strong selection for particular amino acid substitutions and resulting changes in protein function (Christin et al. 2010; Martin and Orgogozo 2013; Stern 2013).

However, it is possible that localised mutation rates may result in a high frequency of particular point mutations, resulting in parallel molecular evolution. An example of high rates of mutation at particular nucleotides has been identified within the *E. coli* gene *thyA* (Viswanathan et al. 2000; Dutra and Lovett 2006). Approximately 50% of all mutations in this gene are at one nucleotide. This molecular parallelism is explained by a mutational hotspot caused by the switching of the template used for DNA repair from the normal opposing DNA strand to a mismatching strand generated by a palindromic secondary structure. Despite these studies, mutational hotspots are rarely invoked as

[6] It is important to distinguish parallel molecular evolution (identical mutations evolving in independent lineages) from the common phenomenon of parallel genetic evolution (mutation in identical or similar genes evolving in independent lineages).

explanations for molecular parallelism, other than cases in which repeat regions may cause similar recombination events in independent lineages (Martin and Orgogozo 2013; Stern 2013).

High local mutation rates are only infrequently evoked and tested as a cause of molecular parallelism, possibly due to the rarity of identical parallel mutations in model species that are highly amenable to genetic experimentation. Molecular parallelism is rarely observed in experimental evolution studies with bacteria (Dettman et al. 2012) – in instances where identical nucleotide changes have been observed in independent lines, such mutations rarely occur over more than two independent lines (Woods et al. 2006; Conrad et al. 2010) – although more observations of molecular parallelism have followed the capacity to sequence greater numbers of replicate lines (Tenaillon et al. 2012). Consequently, there have been few identical parallel mutations observed in appropriate species in which to investigate the causes of parallel molecular evolution. The repeated evolution of the identical C565T *nlpD* mutation – in three independent lines of the SREE – provided an opportunity to investigate the causes of molecular parallelism in a genetically manipulable model system.

5.1.1 The observed parallelism to *nlpD* and the associated phenotype change

The C565T mutation is predicted to result in an amber stop mutation (Q189*) and truncation of NlpD. The aberrant cell-division (cell-chaining) phenotype associated with this mutation (see Section 4.3.5.2) is consistent with the known function of NlpD across several bacterial species. The *nlpD* gene encodes two predicted domains – a LysM domain from residues 59 to 100 (Pfam *E*-value 4.8 x 10^{-11}), which allows binding of the NlpD protein to the peptidoglycan layer (Buist et al. 2008), and a putative zinc metallopeptidase domain between residues 178 and 271 (Pfam *E*-value 3.3 x 10^{-30}). In *Escherichia coli*, NlpD activates AmiC (Uehara et al. 2010), which is required for the cleavage of the murein layer during cell division (Heidrich et al. 2001). Spatial and temporal regulation of AmiC is essential to prevent degradation of the cell wall, which can result in cell lysis (Peters et al. 2011). Thus, in the case of *Pseudomonas fluorescens* SBW25, the C565T mutation results in truncation of NlpD, which likely results in a lack of activation of AmiC and incomplete cell division.

5.1.2 The parallel evolution of C565T mutants may have several causes

The function of NlpD activity does not explain the extreme molecular parallelism of the C565T mutation. Nonsense or frameshift mutations across most of *nlpD* are expected to cause loss of function of NlpD. The *nlpD* gene has 74 codons at which nonsense mutations may occur via a single point mutation. Assuming an equal chance of these mutations occurring, that a nonsense mutation causes an increase in fitness, and that the fitness of these mutants have a Gumbel distribution (as per Orr (2005)), the probability of observing parallel nonsense mutations in two replicate lines is approximately 0.027. Furthermore, codon bias or standard mutational biases are unlikely to cause the C565T parallelism – there are six additional codons in *nlpD* of sequence CAG, of which a C to T transition in the first base would result in truncation of NlpD. The extreme molecular parallelism to only one such codon invokes an additional constraint that may restrict the observed spectrum of *nlpD* mutations.

There are three possible, and not mutually exclusive, explanations for the molecular parallelism of C565T. Firstly, although many similar mutations may occur at *nlpD* resulting in a similar phenotype, only the C565T mutation causes a high fitness and outcompetes alternative *nlpD* mutants. Secondly, there is a functional constraint whereby only one mutation (C565T) can cause the observed cell-division deficient phenotype, resulting in one observed mutant of *nlpD*. Thirdly, the mutation rate at base pair 565 is extremely high, particularly in regard to the rate of C to T transitions, resulting in many C565T mutants that can be subject to selection.

The presence of a predicted promoter at the site of the C565T mutation is likely to be important for these explanations. This promoter controls expression of the downstream gene *rpoS.* RpoS is a highly conserved sigma factor and flanks *nlpD* in many gram-negative bacteria (Venturi 2003). Studies in *P. putida* have identified the promoter within *nlpD* from which *rpoS* is expressed (Kojic et al. 2002). This promoter is found within the predicted M23 peptidase region of *nlpD* and the sequence at this region is 95% identical to the same region within *P. fluorescens* SBW25. The C565T substitution is 7 bp upstream of the homologous starting residue for *RpoS* transcription in *P. putida*, suggesting this mutation may alter transcription of *RpoS.*

RpoS encodes a global transcriptional regulator expressed during stationary phase (Venturi 2003). In *P. fluorescens* Pf-5, RpoS allows survival under osmotic and

oxidative stress (Sarniguet et al. 1995), UV irradiation, freezing, desiccation and starvation (Stockwell and Loper 2005). It has been theorised these functions trade off against nutrient acquisition (Ferenci 2005). Possibly as a result of this tradeoff, mutations to *rpoS* have been identified across multiple evolution experiments within bacteria (Maharjan et al. 2006; Conrad et al. 2009; Kinnersley et al. 2009; Wang et al. 2010).

The function of RpoS transcribed from this predicted promoter supports the first two explanations: the parallelism of C565T is due to the high fitness of the C565T mutants as a result of altered transcription of *rpoS*, or the *nlpD* nonsense mutation may require simultaneous changes to *rpoS* transcription in order to cause the cell-chaining phenotype. There is no literature to suggest that the presence of a nested promoter should raise local mutation rates. These hypotheses that explain the molecular parallelism of C565T were tested, as detailed in this chapter.

5.2 Aim

Identify whether the molecular parallelism of C565T is due to the fitness of derived mutants, a functional constraint requiring changed expression of *rpoS*, or a high mutation rate of the C565T mutation compared to other mutations of *nlpD*.

5.3 Results

5.3.1 A guide to the nomenclature of mutations presented in this chapter

Unlike many genetic studies, both the nucleotide and predicted amino acid sequence may have important effects on the *nlpD* mutant phenotype. For this reason, mutants are generally described by the nucleotide change, followed by both the amino acid change and cellular phenotype in brackets. For example, the C565T mutant is described as '*nlpD* C565T (Q189*, chaining)'.

5.3.2 Identification of the possible spectrum of *nlpD* mutants evolved from SBW25 Δ*wss*

The parallelism of the C565T mutation may be caused by the high relative fitness of this mutant relative to other *nlpD* loss-of-function mutants. In order to identify the phenotypes and fitness caused by alternative mutants, an experimental evolution approach was used that was intended to generate different *nlpD* mutants.

Prior to this evolution experiment, ancestral SBW25 was evaluated as an appropriate genotype to evolve a set of *nlpD* mutants. In both the REE and SREE, most *nlpD* mutations were identified in experimental lines that already featured predicted loss-of-function mutations to the *wss* operon (a necessary gene for the wrinkly spreader (WS) phenotype). This temporal order of mutations suggests that the evolution of WS types may prevent the growth of *nlpD* mutants to observable frequencies. This competitive advantage of WS types was confirmed by fitness assays between *nlpD* C565T (reconstructed in SBW25) and two WS types of known fitness (Bantinaki et al. 2007). SBW25 *nlpD* C565T was significantly less fit (*t*-test, *p*-value 1.3 x 10^{-5}) than 'large spreading wrinkly spreader' (LSWS) with a mean relative fitness of 0.530 ± 0.103 and WS_G (*t*-test, *p*-value 2.6 x 10^{-5}) with a mean relative fitness of 0.651 ± 0.085 (see Section 7.5.1 in the appendix). Informed by the level of fitness of the C565T mutant in SBW25, a SBW25 Δ*wss* strain was made that would prevent the evolution of WS types (due to the absence of the WSS complex), enabling the observation of evolved *nlpD* mutants. This SBW25 Δ*wss* type was utilised for further experiments (see Figure 5.1).

To generate a set of different *nlpD* mutants, SBW25 Δ*wss* was inoculated in 24 static microcosms that were incubated statically for 5 days. Aliquots of incubated culture were spread on KB agar and the evolution of *nlpD* mutants was detected via assessment of colony morphology (see Figure 5.1 for an example of colony morphology). A minimum of 500 colonies (maximum 2000 colonies) were visually scanned from each plated microcosm for the *nlpD* mutant morphology. At least one characteristic *nlpD* mutant morphotype was found in all 24 microcosms and were checked via PCR and Sanger sequencing for mutations to *nlpD*.

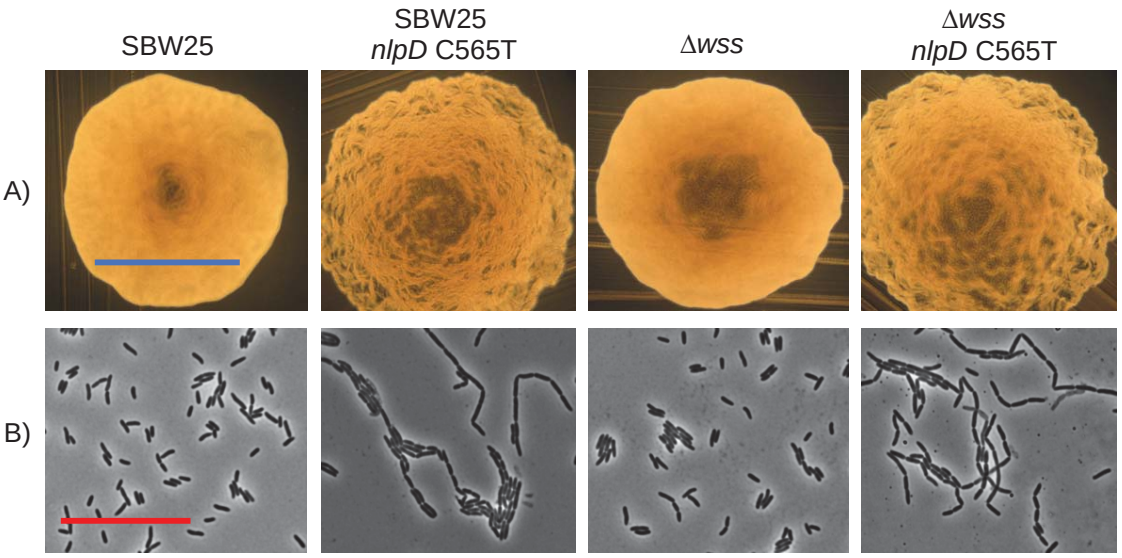

Figure 5.1: The C565T mutant evolved from SBW25 Δ*wss* is similar to SBW25 C565T (PBR923). The C565T types were similar to PBR923 in terms of: A) colony morphology; and B) associated cellular morphology (viewed under a 100x objective). This demonstrates the *wss* locus is not required for this chaining phenotype. The blue scale bar is ~2 mm; the red scale bar is 20 µM.

Sequencing of the isolated mutants revealed a C565T mutation in 23 of the 24 isolated *nlpD* mutants (no mutation was identified in the remaining mutant). No other mutations in *nlpD* were identified. Instead of providing insight as to the spectrum of alternative mutations, this experiment further demonstrated extreme molecular parallelism.

Assuming the six alternative CAG codons all produce nonsense mutations of equal fitness, and that each nonsense mutation has the same mutation rate, the probability of

finding the identical mutation in 23 replicate microcosms can be calculated as per equation 5.1 (where by P = the probability of parallel evolution, n = the number of similar possible mutations and r = the total number of replicate lines):

P = (1/n)^r

$P = (1/7)^{23} = 3.65 \times 10^{-20}$ Equation 5.1

The repeated finding of C565T mutants thus has an extremely low probability of being observed by chance.

5.3.3 The chaining phenotype caused by *nlpD* mutations does not require transcription of *rpoS*

The parallel evolution of the C565T mutation may be explained by a constraint in the number of mutations in *nlpD* that may cause the cell-chaining phenotype (the second hypothesis in Section 5.1.2). Simultaneous changes in transcription of *rpoS* may be required to observe aberrations in cell division. In *E. coli,* RpoS is a global regulator and affects the transcription of approximately 10% of genes (Weber et al. 2005). These genes include *bolA,* the expression of which can cause rod-like morphologies (Aldea et al. 1988). Altered regulation of similar genes in SBW25 – by changes in expression of *rpoS* – may be required for the cell-division phenotype, or to allow the viability of this mutant phenotype.

In order to test whether NlpD truncation and altered *rpoS* expression are simultaneously required to produce the chaining phenotype, *nlpD* and *rpoS* were genetically disentangled. This disentangling separated the predicted promoter of *rpoS* and the open reading frame (ORF) of *nlpD*, and allowed the consequences of the C565T mutation to be assessed independently of *rpoS*. To do this, two pUIC3 plasmids were constructed to encode ~800 bp of the 3' region of *nlpD*, which featured either the ancestral sequence or the sequence with the C565T mutation. These plasmids were recombined by a single crossover at the *nlpD* locus to construct four recombinant types in which *nlpD* and the predicted promoter for *rpoS* (P_{rpoS}) were separated by pUIC3 (see Figure 5.2A). Four distinct constructs were made in which the separated *nlpD* and P_{rpoS} featured: i) the ancestral sequence; ii) the C565T mutation only in *nlpD*; iii) the C565T mutation only in $P_{rpoS;}$ and iv) the C565T mutation in both nlpD and P_{rpoS}. The

sequences and position of each construct was validated via PCR and Sanger sequencing.

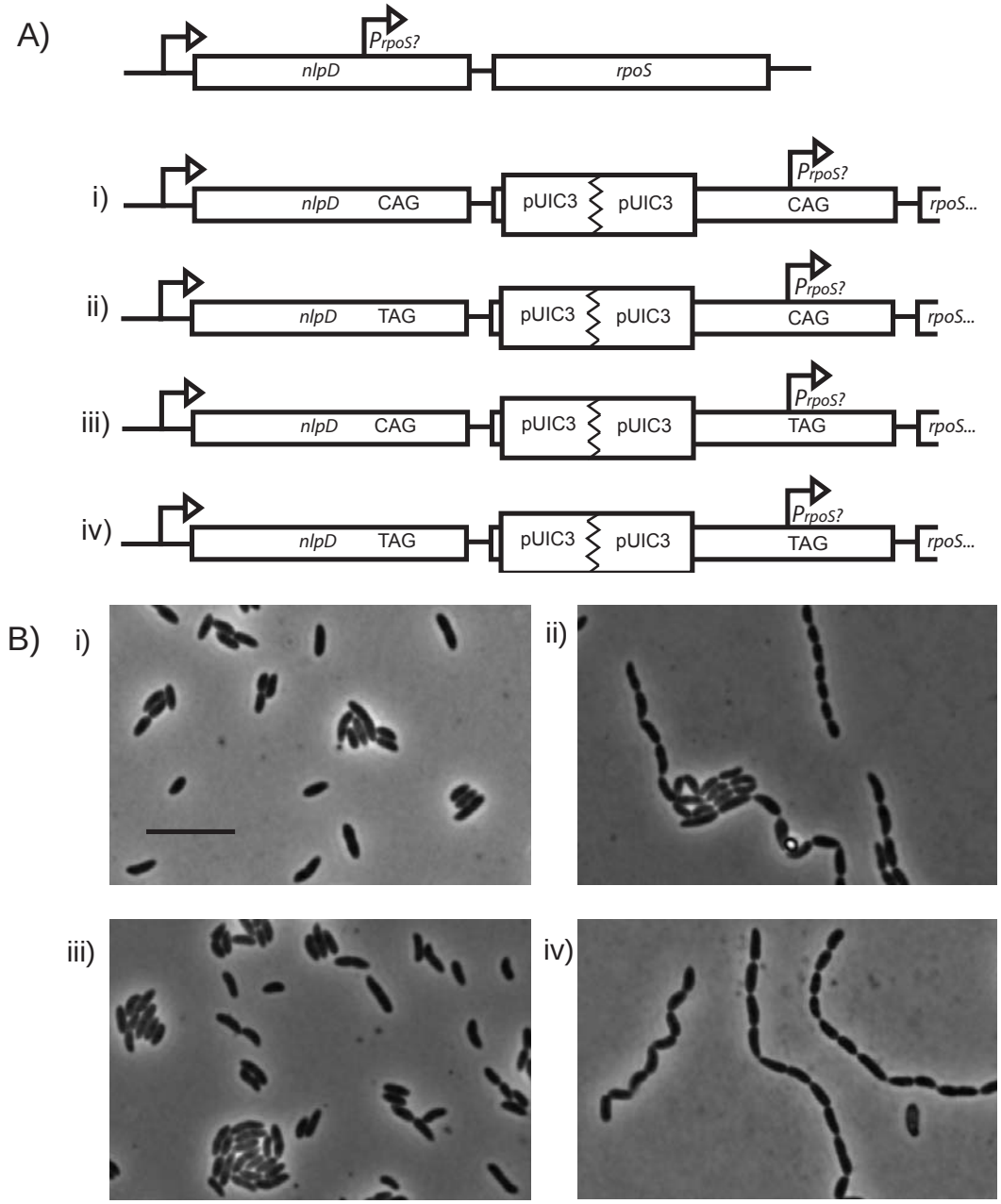

Figure 5.2: Nonsense mutations at C565T at C565T in *nlpD* are sufficient to cause cell-division mutants independently of possible effects on *rpoS.* A) The four different transconjugants of pUIC3 encoding a homologous region of *nlpD* and *rpoS.* The four different mutants consist of either the ancestral CAG sequence at codon 189, or the mutant TAG nonsense mutation, either 5' or 3' to the pUIC3 plasmid, in all four possible combinations (i to iv). B) Microscopy reveals cell-division mutants are caused by nonsense mutations to the ORF of *nlpD*. Constructs ii) and iv), which both involved nonsense mutations to *nlpD*, cause cell-division mutants. Microscopy performed using phase contrast via a 100x objective. Black scale bar is ~5 µM.

The chaining phenotype was caused exclusively by mutations to *nlpD*, as indicated by the cellular appearance of constructs ii and iv (see Figure 5.2B). This suggests that the cell-chaining phenotype does not require altered expression of *rpoS* and that cell-division mutants are viable without altered *rpoS* expression. This strongly suggests the parallelism of C565T is not caused by a requirement to change the transcription of *rpoS* in order to observe cell-division mutants.

5.3.4 Alteration of the *nlpD* sequence causes a lower frequency of molecular parallelism

Two causes remained that may account for the parallelism of C565T: C565T mutants may outcompete alternative *nlpD* mutants or the mutation rate at nucleotide 565 is higher than at other regions of *nlpD* (the first and third hypotheses in Section 5.1.2). Assuming this parallelism is caused by a high mutation rate, alteration of the *nlpD* sequence may destroy a sequence-specific 'mutational hotspot' and result in a decreased frequency of C565T mutants.

To test whether a sequence change of *nlpD* may break a mutational hotspot, a pre-existing *nlpD* mutation was utilised that occurred in the original REE. This mutation – *nlpD* C565T A566G – caused a TGG sequence at codon 189, and was still capable of mutating to the TAG codon by a single G566A nucleotide substitution (see Figure 5.3). The *nlpD* C565T A566G mutant expressed an ancestral (WT) morphology and did not cause the cell-division deficient phenotype (see Figure 5.4).

The *nlpD* C565T A566G mutation was reconstructed in SBW25 Δ*wss,* resulting in SBW25 Δ*wss nlpD* C565T A566G (Q189W, non-chaining) (see Figure 5.4). This mutant was then allowed to evolve to determine the spectrum of resulting *nlpD* mutants. SBW25 Δ*wss nlpD* C565T A566G (Q189W, non-chaining) was inoculated in 54 static microcosms for a period of 5 days. Following incubation, the cultures were plated on KB agar, and resulting colonies were visually inspected for *nlpD* mutant morphologies (screening through a minimum of 500 colonies, and a maximum of 2000 colonies per plated microcosm). PCR and Sanger sequencing of mutant morphotypes was used to identify *nlpD* mutations.

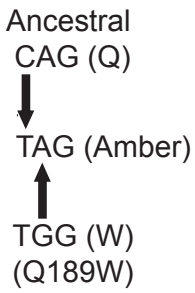

Figure 5.3: TGG at codon 189 can mutate to TAG in a single step. The *nlpD* gene, subject to nucleotide substitutions of C565T A566G, results in a tryptophan (W) at residue 189 of the encoded protein. A single base pair substitution at position 566 (G>A) is sufficient to generate a TAG (amber) stop codon.

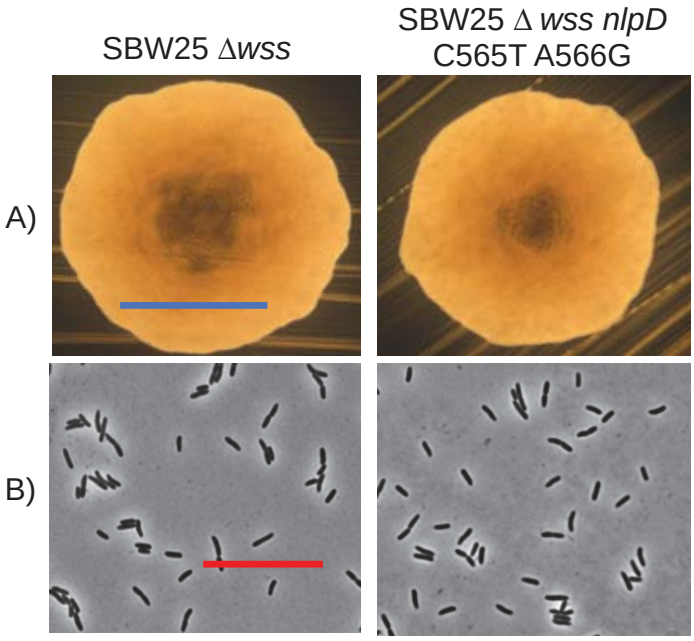

Figure 5.4: The *nlpD* gene carrying a *nlpD* C565T A566G (Q189W) mutation is SM. The *nlpD* C565T A566G mutation, reconstructed in a Δ*wss* genotype, is phenotypically ancestral in both: A) colony morphology; and B) cellular morphology. The blue scale bar is ~2 mm; the red scale bar is ~10 μM.

Cell-chaining morphotypes were identified in 38 microcosms, and mutations within *nlpD* were identified for 37 mutants (see Table 5.1). Of the 37 *nlpD* mutants, eight contained the G566A mutation (resulting in the W189* mutation). This finding is in stark contrast to the spectrum of *nlpD* mutants evolved from SBW25 Δ*wss* (where *nlpD* mutants from 23 of 24 microcosms contained the C565T mutation). A degree of

parallelism was nonetheless observed. The low frequency of the *nlpD* G566A chaining mutant suggests the altered sequence (C565T A566G) may reduce the local rate of mutation (supportive of hypothesis three).

The remaining 29 mutations were identified in a variety of nucleotides along *nlpD* – including base-pair substitutions (resulting in non-synonymous amino acid substitutions), deletions (resulting in either frameshifts or the deletion of amino acids) or nonsense mutations. Predicted amino acid substitutions were observed to result in fewer cells per chain (see Section 7.5.2 in the appendix). This spectrum of alternative *nlpD* mutants suggests many loss-of-function mutations of *nlpD* may cause a cell-chaining phenotype.

Isolate	Genome position	Nucleotide change[a]	Amino acid change[b]
1	1439165	G553A	G185R
2	1439095-1439178	Δ483-566	Δ161-189 (Δ 28aa)
3	1439215	C603G	Y201*
4	1439292	A680G	H227R
5	1439178	G566A	Q189*
6	1439178	G566A	Q189*
7	1439178	G566A	Q189*
8	1439235-1439246	Δ623-634	Δ208-212 (Δ ATAS)
9	1439178	G566A	Q189*
10	1439387	C775G	H259N
11	1439095	G483A	W161*
12	1439232	Δ620	ΔR207(146)
13	1439178	G566A	Q189*
14	1439292	A680G	H227R
15	1439074	Δ462	P154(13)
16	1439178	G566A	Q189*
17	1439289	G677A	G226D
18	1438824	C212T	A71V
19	1439399	C787T	R263C
20	1439178	G566A	Q189*
21	1439214	A602G	Y201C
22	1439178	G566A	Q189*
23	1439423	C811T	P271S
24	1439429	C817T	Q273*
25	1439151	G539A	G180D
26	1439214	A602G	Y201C
27	1439292	A680G	H227R

Isolate	Genome position	Nucleotide change[a]	Amino acid change[b]
28	-	-	-
29	1439105	G493A	G165R
30	1439151	G539A	G180D
31	1439214	A602G	Y201C
32	1438703	T91C	C31R
33	1439008-1439009	Δ396-397	P132(35)
34	1438753-1438756	Δ141-144	A47(6)
35	1439096	C484T	P162S
36	1439207	G595T	V199L
37	1438661	C49T	R17*
38	1439421	A809G	D270G

Table 5.1: The diversity of mutations arising in *nlpD* when microcosms are founded by SBW25 Δ*wss nlpD* C565T A566G (Q189W, non-chaining). [a]Nucleotide changes to *nlpD* relative to the first encoded nucleotide. [b]The resulting encoded amino acid change in NlpD. Brackets enclose the number of predicted residues from the frameshift mutation until an encoded nonsense mutation. '-' Represents unknown mutations.

5.3.5 Assessing whether the fitness of *nlpD* mutants explains parallel evolution

The evolution of different loss-of-function mutations of *nlpD* could be used to assess whether the high fitness of mutants may cause the molecular parallelism of the C565T mutation. The collection of mutants evolved from SBW25 Δ*wss nlpD* C565T A566G (Q189W, non-chaining) was used to measure the fitness of the codon 189 amber nonsense mutant compared to other alternative *nlpD* mutants.

It is important to appreciate that the fitness of alternative *nlpD* mutants evolving from SBW25 Δ*wss nlpD* C565T A566G (Q189W, non-chaining) may not be identical to mutants arising from the ancestral *nlpD* sequence – the presence of the C565T A566G mutation may alter the fitness of these alternative *nlpD* mutants. However, assuming the molecular parallelism of the G566A mutation seen in Section 5.3.4 is caused by the high fitness of the codon 189 amber nonsense mutation, these mutants are expected to be significantly fitter than other *nlpD* mutants.

The competitive fitness of several *nlpD* mutants (as isolated in Section 5.3.4) was measured using flow cytometry. To conduct these assays, a C565T mutant in a Δ*wss*

background was marked with GFP expression to generate a reference strain for fitness assays. This reference was used to measure the relative fitness of isolates 2 *nlpD* Δ458-566 (Δ161-189), 11 *nlpD* G483A (residue W161*), 24 *nlpD* C817T (Q273* nonsense mutation), 34 *nlpD* Δ141-144 (A47 frameshift) and 37 *nlpD* C49T (R17*). Two of the parallel *nlpD* mutants (Q189*, chaining) – which evolved by either C565T or G566A mutations – were included to enable correction for the cost of GFP expression. The resulting competitions were performed over 48 hours in statically incubated microcosms of KB media (see materials and methods 2.2.10.2 for details of this competition experiment and generation of cell counts).

The resulting measures of fitness are expressed as the selection coefficient – a measure of the change in the ratio of the tested and reference strains per generation of competition (see materials and methods 2.2.10.2 for details of this coefficient), following correction for the cost of GFP. The results of this competition assay are presented in Table 5.2 (raw data and details of statistical analyses are presented in Tables 7.5.6. and 7.5.7 in the appendix).

Genotype	Mutation	Mean CSC±CI[a]	*P*-value of *t*-test[b]
Control a	C565T (Q189*)	0.000 ± 0.0344	1
Isolate 2	Δ458-566 (Δ161-189)	-0.1534 ± 0.0562	0.00035
Isolate 11	G483A (W161*)	-0.0354 ± 0.0664	0.247
Isolate 24	C817T (Q273*)	-0.0788 ± 0.0385	0.0019
Isolate 34	Δ141-144 (A47 FS)	-0.0969 ± 0.0599	0.0064
Isolate 37	C49T (R17*)	0.0525 ± 0.0381	0.0139
Control b	G566A (Q189*)	0.0551 ± 0.0332	0.0057

Table 5.2: The mean selection coefficients of alternative *nlpD* mutants relative to a C565T chaining reference. [a]Mean corrected selection coefficient (CSC) for each competition ± 95% confidence interval (CI). [b]Student's *t*-test *p*-value for one-sample, two-tailed, *t*-test that the relative frequency differs from 1. All competitions were made with eight biological replicates.

Isolates 37 (C49T, R17*) and 11 (G483A, W161*) were found to have a statistically higher or similar fitness to the reference strain (see Table 5.2). Considering the C49T (R17*) and G483A (W161*) mutants only arose once, despite a higher or similar selection coefficient, it is unlikely the fitness caused by the codon 189 amber mutation

explains the observed molecular parallelism of this type arising from SBW25 Δ*wss* *nlpD* C565T A566G.

Isolates 2 (Δ458-566, Δ161-189), 24 (C817T, Q273*) and 34 (Δ141-144, A47FS) were associated with significantly lower mean selection coefficients relative to the reference (see Table 5.2). This lower selection coefficient may be due to differences in phenotype caused by the different *nlpD* mutations. A significant difference in selection coefficient was also found between the two controls (see Table 5.2). The significant differences between the controls may be due to additional pre-existing mutations. Alternatively, many random events may alter the fitness of mutants, including mat collapse, the evolution of other mat-forming mutants, or the rapid evolution of non-chaining revertant types. This difference is important, as it provides uncertainty as to the cost of the GFP marker. However, even with correction for the cost of the marker with the fitter G566A (Q189*) control, the C49T (R17*) mutant has a similar mean selection coefficient relative to the reference (see Table 7.5.8 in the appendix). In summary, despite the variation in mean selection coefficient between many of the genotypes and the reference, there is no evidence that the codon 189 amber mutation arose in parallel due to a high selection coefficient associated with this type.

5.3.6 The reduced parallelism of *nlpD* subtle mutants is associated with reduced *rpoS* transcription

Rather than being a cause of high fitness of the C565T mutants, the nested promoter of *rpoS* may cause a high substitution mutation rate that results in the molecular parallelism of C565T. This mutation rate may be caused by a mechanism involving either the attachment of initiation factors, or conformational changes of the DNA resulting from the initiation of transcription of *rpoS*.

The association of transcription and the mutagenesis at C565 – resulting in the repeated observation of the C565T mutation – may be explained by a phenomenon similar to ‘transcription associated mutagenesis’ (TAM). TAM describes the increased mutation rate of genes with high rates of transcription (Beletskii and Bhagwat 1996; Hudson et al. 2003). However, were TAM to operate within *nlpD*, the mutation rate would be expected to be elevated downstream of the *rpoS* promoter, and not elevated at only nucleotide C565. Similarly, other mutation mechanisms such as mismatch

repair template switching (see Section 5.1 above) do not explain the repeated observation of the C565T mutation, as this nucleotide is not situated at the unpaired position of a stem-loop structure (see Figure 7.5.4 in the appendix).

The molecular parallelism of the C565T mutation – if it is caused by a high mutation rate – is possibly caused by the initiation of transcription and is hereafter termed 'transcription initiation associated mutation' (TIAM). Consistent with this hypothesis, the ancestral *nlpD* sequence – which is associated with the extreme molecular parallelism of C565T – should be associated with higher transcription from the nested *rpoS* promoter compared to the *nlpD* C565T A566G (Q189W, non-chaining) mutated sequence.

To test whether the *nlpD* C565T A566G (Q189W, non-chaining) mutant differs from the ancestor in transcription from P_{rpoS}, reverse transcription quantitative real-time PCR (RT-qPCR) was performed. To do this, relative mRNA abundance across two regions of *nlpD* (see Figure 5.5) were measured across two strains: SBW25 Δ*wss nlpD* C565T A566G (Q189W, non-chaining) and SBW25 Δ*wss* (non-chaining). These strains were grown in overnight cultures of KB shaken at 28°C, and then subcultured and grown to an optical density (OD) of 2.5-3.0 to induce expression of *rpoS*. If the TIAM hypothesis is correct, the SBW25 Δ*wss* (non-chaining) genotype should have a higher level of transcription from the predicted nested promoter, compared with the *nlpD* C565T A566G (Q189W, non-chaining) genotype. The raw data of the levels of mRNA were normalised to constitutively expressed *recA* and are presented in the appendix (see Table 7.5.9 for raw data, Figures 7.5.5 and 7.5.6 for boxplots of this data, and Table 7.5.11 for statistical analyses).

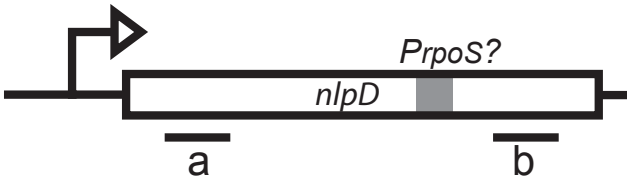

Figure 5.5: Regions a (5') and b (3') were assessed for levels of transcript via RT-qPCR across strains.

RT-qPCR results are consistent with the hypothesis that the WT *nlpD* sequence encodes an active P_{rpoS} compared to *nlpD* C565T A566G (see Table 5.3 – data is presented as ratios to correct for changes in transcription from the promoter for *nlpD* across the strains). The ancestral *nlpD* genotype has a higher ratio of approximately 2.3 times the relative expression of the 3' region of *nlpD* compared to the 5' region (this ratio is not equal to 1.0, *t*-test *p*-value 3.44×10^{-6}). In genotypes with a C565T A566G substitution mutation, the mean relative mRNA levels are lower at the 3' *nlpD* region relative to the 5' *nlpD* region.

Genotype 1	Region of mRNA quant.	Mean Cq relative to *nlpD* (5') ± SE	CI[a]	*P*-value of *t*-test[b]
Δ*wss nlpD (WT)*	*nlpD* 3'	2.330 ± 0.101	2.039 - 2.569	3.44×10^{-6}
Δ*wss nlpD* C565T A566G (Q189W)	*nlpD* 3'	0.700 ± 0.0301	0.629 - 0.771	2.19×10^{-5}

Table 5.3: Ancestral *nlpD* has an increase in transcription downstream of the *rpoS* promoter compared to *nlpD* C565T A566G mutants. Mean Cq have already been made relative to *recA* Cq values. SE stands for standard error about the mean. [a]95% confidence intervals about the mean as calculated by Student's *t*-tests. [b]Tests the means are the same, suggesting similar levels of transcription. Cq values were measured from two biological replicates (four technical replicates per biological replicate).

This data is consistent with the ancestral sequence of *nlpD* containing within it (likely near nucleotide 565) an active transcription start site for *rpoS*. When the *nlpD* sequence is subtly altered by C565T A566G (Q189W, non-chaining), transcription is greatly reduced. This finding is consistent with the TIAM hypothesis – the *nlpD* sequence with a greater rate of transcription from P_{rpoS} also evolves a great frequency of mutants at codon 189, and may have a greater mutation rate at nucleotide 565.

At the same time as the above RT-qPCR measures were taken, mRNA quantities in the SBW25 Δ*wss nlpD* C565T (Q189*, chaining) mutant were also measured to be ~19 times higher at the 3' region relative to the 5' region of *nlpD* (*t*-test that the ratio is equal to 1.0, *p*-value 1.79×10^{-5}). By comparison, the alternative mutant SBW25 Δ*wss* C49T (R17*, chaining) – which also contained the *nlpD* C565T A566G mutation – had significantly lower mRNA quantities transcribed in the 3' region of *nlpD* (see Table 5.4).

This observation provided an opportunity to indirectly assay the frequency of mutations within *nlpD* of different mutants.

Genotype 1	Region of mRNA quant.	Mean Cq relative to *nlpD* (5') ± SE	CI[a]	*P*-value of *t*-test[b]
Δ*wss nlpD* C565T (Q189*)	*nlpD* 3'	18.905 ± 1.840	14.554 – 23.256	1.79×10^{-5}
Δ*wss nlpD* C49T (R17*) and C565T A566G	*nlpD* 3'	0.650 ± 0.032	0.574 – 0.724	1.21×10^{-5}

Table 5.4: The *nlpD* 565T mutation was identified as having high levels of transcription from the *rpoS* promoter. Mean Cq have already been made relative to *recA* Cq values. SE stands for standard error about the mean. [a]95% confidence intervals about the mean as calculated by Student's *t*-tests. [b]Tests the means are the same, suggesting similar levels of transcription. Cq values were measured from two biological replicates (four technical replicates per biological replicate).

5.3.7 Increased transcription from P_{rpoS} correlates with the frequency of mutations within *nlpD*

The hypothesis that the C565T parallel mutation is caused by a high mutation rate, possibly due to transcription from P_{rpoS}, requires validation by a measurement of the rate of mutation. There is no direct way of measuring the rate of mutation of the ancestral sequence of *nlpD*. Attempts to generate a construct that would measure the rate of mutation within *nlpD* were unsuccessful[7].

However, the insight that the *nlpD* C565T mutation was associated with a high level of transcription enabled an indirect assessment of whether high transcription from P_{rpoS} may cause a high frequency of mutation at nucleotide 565 of *nlpD*. Considering that the C565T mutant has a high level of transcription from P_{rpoS} (see Section 5.3.6), then according to the TIAM hypothesis, C565T mutants should also have a high mutation rate at nucleotide 565, and a high rate of reversion back to the ancestral sequence and phenotype.

[7] This construct involved in frame fusions of the predicted *rpoS* promoter to *lacI* that repressed transcription of a kanamycin resistance marker. The resulting cassette was intended to perform fluctuation tests that would measure the rate of C565T mutations; however, repression of *kanR* by LacI was not possible with the translational fusions.

These rare reverting mutants are observable in a population dominated by *nlpD* mutants when such populations are plated on semi-solid agar. Rare revertants with the ancestral phenotype are able to spread further on semi-solid agar compared to cell-division mutants, as observed in spreading sectors of plated populations (see Figure 5.6 for an example). Using this technique, cultures of cell-division mutants, caused by different mutations in *nlpD*, can be tested for the relative production of reverting mutations.

Figure 5.6: Example of a revertant spreading section of colony founded predominately with C565T mutants.

The frequency of reversion was assessed for multiple *nlpD* mutants by the plating of cultures on semi-solid agar. The genotypes included two identical codon 189 nonsense mutants (evolved by either C565T or G566A mutations) that – consistent with the TIAM hypothesis – are expected to revert at a high rate. Alternative *nlpD* mutant types (see Section 5.3.4) were also measured for the frequency of reversion. These types included isolate 37 C49T (R17*, chaining), isolate 11 G483A (W161*, chaining) and isolate 24 C817T (273*, chaining). These types have mutations distal to the predicted transcription start of P_{rpoS} – that was also measured or expected to operate with lower activity in these strains – and consequently are predicted to evolve revertants at a low frequency. A negative control was included of isolate 2 Δ483-566 (Δ161-189) that should not be able to revert back to WT cellular morphology via mutation to *nlpD*.

Each strain was cultivated in 115 replicate overnight cultures (grown in 200 µL of KB, shaken for 16 h). For each culture, 1 µL of culture was then spotted on semi-solid KB agar plates and incubated at 28°C for 48 hours. Resulting ‘spots’ were assessed visually for ‘spreading’ sections and the frequency of colonies with spreading sectors

was recorded. The spreading sections were confirmed by microscopy as having the ancestral (non-chaining) morphology, indicative of mutations that returned the *nlpD* nonsense mutation to the ancestral (or similar) sequence. Samples of such revertants were also confirmed via PCR and Sanger sequencing as involving reverting mutations.

Measures of the rate of rifampicin resistance were also obtained to demonstrate that variances in the frequency of sectoring were not as a result of a genome-wide increase in mutation rate (see Methods and Materials 2.2.17 for details). Both the estimated global mutation rate for each genotype and the frequency of colonies with visually observed sectors are presented in Table 5.5.

***nlpD* Mutant**	**Global mutation rate (MSS-MLE)**	**Global mutation rate 95% CI**	**Colonies with WT sectors**
G566A (W189*)	4.28×10^{-9}	7.04×10^{-9} to 2.08×10^{-9}	55%
C565T (Q189*)	4.44×10^{-9}	8.57×10^{-9} to 1.43×10^{-9}	38%
G483A (W161*)	4.19×10^{-9}	6.93×10^{-9} to 2.01×10^{-9}	1.7%
C817T (R273*)	5.40×10^{-9}	8.97×10^{-9} to 2.56×10^{-9}	4.4%
C49T (R17*)	1.17×10^{-8}	1.87×10^{-8} to 5.98×10^{-9}	0%
Δ483-566 (Δ161-189)	5.44×10^{-9}	8.45×10^{-9} to 2.95×10^{-9}	0%

Table 5.5: The frequency of cultures with revertant sections is higher when initiated with the Q189* mutant. The frequency of colonies with at least one 'reverting' sector is presented (from a total of 115 cultures per genotype). Estimates of the global mutation rate is also presented to indicate the rate of *nlpD* revertants is not caused by genome wide mutation rates. CI is the confidence interval about the mean of the predicted global mutation rate per strain calculated from 20 cultures.

Independent populations founded by the identical G566A or C565T mutant genotypes more frequently feature sectored colonies when grown in semi-solid agar, than those founded by the alternative *nlpD* mutants. Of the 1 µL spotted colonies founded from the two genotypes encoding codon 189 amber nonsense mutants of *nlpD*, 55% and 38% displayed at least one spreading sector. Populations founded with genotypes featuring

alternative *nlpD* mutations present a lower frequency of representative colonies with sectoring. These alternative mutants have a frequency of sectored colonies between ~5% and 0%. The *nlpD* mutant encoding a large deletion Δ483-566 (Δ161-189, chaining) was not expected to evolve revertants. Consistent with this expectation, no sectored colonies were observed associated with this genotype. Measures of rifampicin resistance were approximately equal across all genotypes (at ~4 x 10^{-9} mutants per CFU), implying that the global mutation rates do not explain the observed differences in the frequency of revertants.

The difference in the frequency of revertants between the various *nlpD* mutants is consistent with the hypothesised TIAM mechanism – that the C565T genotype has high transcription from P_{rpoS} and a high mutation rate that reverts the C565T mutation. Other *nlpD* mutants expressing low transcription within *nlpD* have a low frequency of reversion. This reversion assay strongly suggests that the ancestral genotype (with an active P_{rpoS} promoter) would likely have a high mutation rate at nucleotide 565 of *nlpD*, which may be responsible for the molecular parallelism of C565T mutants.

5.4. Discussion

5.4.1 Summary of the causes of parallel mutations to *nlpD*

The independent evolution of *nlpD* C565T provides an exceptional example of molecular parallelism arising within a microbial evolution experiment. Across experimental lines, initial mutations to *nlpD* exclusively consist of a C565T mutation. This mutation results in a nonsense mutation and a predicted truncation of the NlpD protein, resulting in a deficiency in cell division.

One possible explanation for the parallel evolution of C565T stems from the overlap of this mutation with the promoter of *rpoS* (P_{rpoS}). Mutations at this promoter – via altered transcription of *rpoS* – might cause a beneficial change in *rpoS* transcription, delivering a single, high-fitness genotype that would outcompete alternative loss-of-function *nlpD* mutants. However, the C565T mutation is not associated with a significant increase in fitness compared to other *nlpD* mutants (see Section 5.3.5). Fitness assays presented no evidence that the parallelism of C565T is derived from the high fitness of this mutant, and an alternative explanation was required to account for this parallelism.

There is indirect evidence that the molecular parallelism of the C565T mutation is caused by a mutational 'hotspot', possibly resulting from transcription initiation, here termed 'transcription initiation associated mutation' (TIAM). Replacement of the ancestral *nlpD* sequence with the Q189W mutation results in a genotype that evolves *nlpD* mutants with a diverse spectrum of mutations and a lower frequency of the Q189* mutation (see Section 5.3.4). This suggests the alteration of the *nlpD* sequence may partially destroy a 'hotspot' for mutations, which reduces the frequency of molecular parallelism.

Further evidence supports the TIAM hypothesis as a plausible cause of the parallel evolution of C565T. Genotypes with this suspected mutation rate 'hotspot' are associated with a high level of transcription from P_{rpoS}. In contrast, the Q189W mutation is associated with a low transcription from P_{rpoS}, as well as a reduced frequency of the codon 189 amber nonsense mutant. The association between transcription and the frequency of mutation extends to loss-of-function mutants of *nlpD*. The C565T mutation was associated with a relatively high level of transcription from P_{rpoS}, as well as a high

frequency of reversion of the *nlpD* mutation back to the ancestral genotype. Thus, an association exists between transcription levels from P_{rpoS} and the frequency of mutation at this promoter. Transcription from P_{rpoS} appears to cause a high local mutation rate, which may be responsible for the molecular parallelism of C565T.

The hypothesis that the molecular parallelism of C565T is caused by high mutation rates requires a direct measurement of mutation rates in order to be validated, if such a method can be devised. Such validation is important because a confirmation of high mutation rate being caused by transcription initiation has important consequences for our understanding of bacterial evolution.

5.4.2 A mutational mechanism may result in the parallelism of C565T

The predicted high mutation rate of the C565T mutation appears to be associated with the positioning and activity of the *rpoS* promoter within *nlpD*. Considering this relationship of transcription and a predicted high mutation rate, the parallel evolution of C565T mutants may be caused by localised hyper-mutation via a mechanism similar to transcription associated mutation (TAM). TAM is a general term that describes the increased mutation rate associated with the transcription of genes (Hendriks et al. 2010; Kim and Jinks-Robertson 2012). This phenomenon has been observed across species as diverse as Bacteriophage T7 (Beletskii et al. 2000; Klapacz and Bhagwat 2002), *E. coli* (Yoshiyama and Maki 2003), *Salmonella enterica* (Hudson et al. 2003), *Saccharomyces cerevisiae* (Lippert et al. 2011) and vertebrates (Polak and Arndt 2008; Polak et al. 2010). In *E. coli,* the rate of cytosine deamination (which results in a thymine nucleotide) on the non-transcribed strand (NTS) is approximately four times higher provided induced transcription of the tested gene (Beletskii and Bhagwat 1996). This C to T mutation on the NTS is identical to the parallel mutation identified in *nlpD*.

Molecular mechanisms have been proposed to explain the higher rate of mutation of transcribed genes, with the deamination of cytosine nucleotides on the NTS of particular relevance in accounting for the C565T mutation. The advancement of RNA polymerase along a transcribed gene is matched by a temporary separation of the non-coding strand, resulting in a single-stranded NTS known as the R-loop (Kim and Jinks-Robertson 2012). Whilst in the single-stranded state, the NTS is sensitive to mutation. *In vitro* assays have demonstrated that single-stranded DNA has a ~140-fold increase

in the rate of cytosine deamination (Frederico et al. 1990). It has been proposed that in highly transcribed genes, the NTS is more frequently made single stranded, which results in a ~4-fold increase in point mutations in some bacterial species (Beletskii and Bhagwat 1996; Hudson et al. 2003).

Further refined models have also been proposed to explain the higher mutation rate of the NTS. The movement of RNAP can create regions of negative supercoiling (Liu and Wang 1987), which are predicted to form stem-loop structures (SLS) (Dayn et al. 1992). A higher mutation rate has been demonstrated at unpaired compared to paired regions with these structures (Burkala et al. 2007). The correlation of SLS structures and mutation frequency has even been extended to the tumour-suppressor gene *p53* in cancer cells (Wright et al. 2002; Wright et al. 2011).

An alternative mechanism explains the high mutation rate of only some nucleotides within a coding sequence (Ripley 1982). Enzymes such as MutS detect the mutation of nucleotides by recognition of the mismatch of mutated nucleotides with the corresponding template sequence. These enzymes then instigate repair of this mismatch. However, in the case of stem-loop structures (SLS), the recognition of unpaired nucleotides in the stem can be erroneously recognized as a mutation, and repaired according to the alternative template provided by the opposing stem. In the *E. coli* gene *thyA*, this mechanism causes parallel molecular mutations of nucleotides within palindromic sequences (Viswanathan et al. 2000; Dutra and Lovett 2006).

These existing mechanisms of mutagenesis do not sufficiently explain the extreme parallelism of C565T mutants. A TAM mechanism of mutagenesis would be expected to result in the mutagenesis of the five CAA or CAG codons downstream of P_{rpoS} into nonsense mutations of *nlpD*. The availability of these alternative targets for mutation, and the lack of observation of such *nlpD* mutants evolving (with WT *nlpD*), suggests mutagenesis within an R-loop is not sufficient to explain the parallelism of C565T. Furthermore, predictions of secondary structure around C565 suggest nucleotide 565 is paired in a stem-loop structure (see Figure 7.5.4 in the appendix). This pairing suggests the parallelism of C565T is not caused by the mutagenesis of unpaired nucleotides within a SLS, nor by mutagenesis by template switching. In summary, these known mechanisms do not readily account for the suspected high mutation rate within *nlpD*.

The extreme localisation of mutation to C565T may involve a novel mutational mechanism involving the initiation of transcription (here termed 'transcription initiation associated mutagenesis', TIAM). Given the novelty of this finding, there is no clear mechanism that describes how transcription initiation may cause this suspected mutagenesis. There are no known examples of nested promoters causing a high rate of mutation in other species, and it is unknown whether nested promoters cause high mutation rates in other genes. Mutagenesis in active promoters is not readily measured using standard reversion assays, and measurement of the mutation rate of the P_{rpoS} would require a novel construct. A measurement of the rate of mutation across *nlpD* awaits future studies. Assuming such a measurement is possible, this may reveal the existence of a mutagenesis mechanism in promoter regions that has been overlooked.

5.4.3 Implications of the potential mutability of nested promoters at *nlpD* and other loci

This chapter presents evidence that the *nlpD*-nested promoter of *rpoS* may cause a high rate of mutation. This is a unique finding considering the paucity of information on how mutation rates may be biased across loci and how these rates my increase the production of variants subject to selection. There is scant discussion of how local mutation rates can bias the production of variation and cause instances of parallel evolution (Stern 2013). An experimental evolution study with 115 replicates identified multiple instances of molecular parallelism, especially of non-synonymous substitutions, however variances in mutation rate are not suggested as causal of these parallel evolution events (Tenaillon et al. 2012).

The rare invocation of local mutation rates may reflect a genuine lack of mutation rates as causal of parallel mutational events. However, the absence of mutation rates as an explanation of molecular parallelism may also reflect an assumption that mutation rates are identical across any given gene. Mutational parallelism to *nlpD* provides evidence that high mutation rates may be caused by sequences that are not typically associated with high mutation rates, such as non-repetitive coding regions. This study thus provides a cautionary conclusion for other experimental evolution studies – local mutation rates may cause parallelism through mechanisms that are not readily identifiable by the type of mutation.

The predicted high mutation rate of *nlpD* may not be limited to SBW25 and may occur in other species. However, expectations of which species may have a 'mutational hotspot' in *nlpD* will depend upon the elucidation of a molecular mechanism that may cause this high local mutation rate. It is possible that the regulation of transcription initiation at P_{rpoS} may affect this mutation rate, and that due to this regulation, this hotspot is limited to certain species. Despite the conservation of the nesting of P_{rpoS} within *nlpD* across many bacterial species (Paesold and Krause 1999; Hengge-Aronis 2002; Kojic et al. 2002; Wang et al. 2014), there is variation across species in how P_{rpoS} is regulated. In *E. coli*, cAMP receptor protein (CRP) positively regulates transcription from P_{rpoS}, as opposed to the transcription factor PsrA in *Pseudomonas putida* (Venturi 2003). If the specific regulation of P_{rpoS} causes the mutational hotspot in *nlpD*, then this hotspot may be limited across species. However, if the nesting of P_{rpoS} causes a high mutation rate regardless of the type of regulation, then a hotspot may be predicted within *nlpD* across a broad number of bacterial species (especially when propagated at stationary phase).

The discovery of a TIAM mechanism – should it be confirmed – may have wide applicability for predicting evolution at loci that are not homologs of *nlpD*. The absence of a molecular mechanism accounting for TIAM prevents a prediction of whether this phenomenon is relevant at promoters that are not nested, or are nested but the regulation at which is unknown. However, even if TIAM is restricted to nested promoters, knowledge of such a hotspot is highly relevant for predicting the locus of evolution. Promoters can be readily identified by sequence homology, enabling the identification of promoters nested within ORFs. It is estimated that in *Salmonella Typhimurium*, nested (or internal) promoters comprise approximately 16% of identifiable transcription start sites (Kroeger et al. 2012). The identification of this high degree of nesting may allow identification of genes with a high mutation rate. The identification of genes that may have a mutational hotspot (due to such promoter nesting), in combination with an improved estimation of the selection coefficient of genotypes in a given environment, may allow accurate predictions of how organisms may evolve.

5.4.4 What is the benefit of a nested *rpoS* promoter?

This study has provided a rare insight on the functionality of *nlpD* and how this gene interacts with *rpoS*. The high conservation of *nlpD* and *rpoS* as neighbouring genes has not previously been explained. Although the transcription start of *rpoS* has been frequently identified within *nlpD* across diverse species, the benefit of this nested promoter has not been identified.

The lack of an explanation for the nesting of P_{rpoS} is surprising given the detailed research describing the complex regulation of *rpoS,* especially in *E. coli*, at both the transcriptional and translational levels (Landini et al. 2014). The 5' untranslated region (UTR) of *RpoS* in *E. coli* is the longest described bacterial UTR. This length seems necessary to incorporate regions encoding antagonistic binding sites that promote or reduce translation. Deregulation of such a global regulator may have strongly deleterious consequences, and complex regulation of this gene is consistent with the expected pleiotropic consequences of deregulation of *rpoS*.

The benefit derived from nesting this complex UTR within the coding frame of *nlpD* has not been addressed from an evolutionary perspective. Disentanglement of the nested *rpoS* promoter from the ORF of *nlpD* is achieved by a simple duplication event (similar to that constructed in Section 5.3.3). Considering such mutational events could readily occur, why is the nesting of P_{rpoS} within *nlpD* so widely conserved? The phenotypic consequences of mutations to *nlpD* suggests two, alternative, explanations for the conserved nesting of P_{rpoS} within *nlpD*.

5.4.4.1 The nesting of P_{rpoS} in *nlpD* may preserve the promoter

The nesting of P_{rpoS} may prevent the loss of this promoter in environments where it remains inactive and *rpoS* is not expressed. This study has clearly demonstrated that a nonsense mutation to *nlpD* may have dramatic consequences at the cellular level. The deficiency in cell separation may confer a low fitness, especially in environments supporting exponential growth. Such a low fitness caused by loss of function of *nlpD* will prevent the increased frequency of genotypes with loss-of-function mutations to the *rpoS* promoter. This protection from loss-of-function P_{rpoS} mutations would occur even in conditions in which the promoter is not expressed, allowing preservation of the

promoter without direct selection on the expression of *rpoS*. Nesting of P_{rpoS} in *nlpD* may help conserve the complex regulation of *rpoS* from mutation in times of relaxed selection on *rpoS* expression, such as during exponential-phase growth.

While it is difficult to directly test this hypothesis, it may be possible to gain support for it by a bioinformatic search for similar transitively induced genes that have promoters nested in neighbouring genes. The nesting of these promoters may similarly prevent degeneration of such promoters during times of relaxed selection for the functionality of the downstream gene.

5.4.4.2 The nesting of P_{rpoS} may cause a stochastic change in cell length

The evidence presented in this chapter provides an alternative explanation for the conserved nesting of P_{rpoS} within *nlpD*. Assuming the transcription from the promoter results in a high mutation rate at the promoter of *rpoS*, the concomitant high frequency of loss-of-function mutations to *nlpD* – and the high frequency of reversions – may confer an adaptive benefit in causing the stochastic production of sessile chaining types. The benefit of stochastic switching in fluctuating environments has been well described (Acar et al. 2008; Veening et al. 2008). The stochastic generation of cell-chaining and motile types has been identified in populations of *Bacillus subtilis* (Kearns and Losick 2005). These chaining (sessile) types cause the biofilm to be initiated, which allows protection from environmental hazards such as antimicrobials or UV exposure (Hall-Stoodley et al. 2004) and colonisation of specific niches such as the plant surface (Vlamakis et al. 2013), whilst the stochastic generation of motile sub-populations helps disperse the biofilm. In the case of *B. subtilis*, the stochastic switching between biofilm chaining types and motile types is caused by an epigenetic switch (Norman et al. 2013). Perhaps the suspected high mutation rate of *nlpD* causes a similar stochastic switch between motile and sessile biofilm-forming types, via a reversible genetic switch.

A high mutation rate of *nlpD*, and a high frequency of cell-division mutants, may also allow protection from microfaunal predation. Predation by a variety of species, such as nematodes and protozoa, is believed to be a major selective force of diversity in bacterial species (Jousset 2012). Bacterial populations respond to the selective pressure of predation via a number of anti-grazing strategies. These strategies include

evasion of detection (Wildschutte et al. 2004) and the production of toxins (Vaitkevicius et al. 2006; Mazzola et al. 2009; Jousset et al. 2010). Predation can also be prevented by changes in cell shape and by aggregation, which prevent ingestion by predators. Such morphological changes can be produced by the production of biofilms, and exopolymers have been shown to limit ingestion (Hahn et al. 2004; Matz et al. 2004; Weitere et al. 2005). The production of cellular filaments has been shown to limit ingestion across diverse bacterial species (Hahn et al. 1999; Corno and Jurgens 2006; Queck et al. 2006). These filaments are visually similar to those caused by loss-of-function mutations of *nlpD* in SBW25.

The filamentous morphology of *nlpD*, combined with a high mutation rate at this locus, may be a stochastic switching mechanism to prevent grazing by microfaunal predators. The location of the nested promoter, which may also confer a high rate of mutation of *nlpD*, may provide a means of producing filamentous mutant types at a high rate, which are resistant to grazing. These filamentous types may be able to revert back to a non-filamentous WT morphotype, and not be locked in an evolutionary 'dead end' as a filamentous type. Predation may thus provide a means of selecting the nested promoter, by direct selection for genotypes that are able to stochastically switch between filamentous and non-filamentous types. Direct measurement of mutation rate at this locus in SBW25, and association of this mutation rate with fitness in predator-prey coevolutionary studies, may provide evidence that microfaunal predation selects for a high mutation rate in *nlpD*. This hypothesis may also apply to species outside of *P. fluorescens*, assuming the predicted high mutation rate in *nlpD* is found in other species.

Chapter 6:
Concluding remarks

6.1. Project background

Natural selection requires a supply of heritable phenotypic variation, which is ultimately caused by mutation. Whilst the random nature of mutation suggests that the production of phenotypic variation is similarly random, features encoded at different loci may alter the production of variation, thus altering the course of the evolution of species. Understanding and identifying these features may prove important in allowing predictions of how organisms evolve.

Recent studies have identified regulatory loci within larger developmental networks as the frequent target of parallel mutations (catalogued in Martin and Orgozozo (2013)). The effect of mutations at these regulators, in particular the low pleiotropic effect of these mutations, has been used to explain this parallelism (Stern 2013). This thesis does not contradict these claims, but suggests, along with previous works (Bantinaki et al. 2007; McDonald et al. 2009), that genetic features may also affect the production of phenotypic variation that is then subject to selection, resulting in observations of parallel evolution. These genetic features include the genetic architecture underpinning phenotypes, and the presence of high local mutation rates.

The findings presented in this thesis would not have been possible without previous studies detailing the genes of *Pseudomonas fluorescens* SBW25 which mutate and cause the niche specialist wrinkly spreader (WS) phenotype (Spiers et al. 2002; Goymer et al. 2006; Bantinaki et al. 2007). This series of studies led to the identification by McDonald et al. (2009) of three genetic loci (*wsp*, *aws* and *mws*) that could mutate to cause WS, despite a multitude of similar loci encoding predicted di-guanylate cyclases (DGCs). McDonald and colleagues suggested that the negative regulation encoded by these three loci may, upon mutation, more readily cause WS types than other genes encoding DGCs. This hypothesis provided an initial motivation for the studies here presented – to identify the mutation causal of the slow wrinkly spreader (SWS) phenotype (which evolved in the absence of *wsp*, *aws* and *mws*) and determine whether genetic architecture prevents the causal locus from regularly

generating WS. This causal mutation was found to be an in-frame fusion mutation between *fadA* and *fwsR*, which is rare in comparison to loss-of-function mutations at *wsp, aws* and *mws*.

Following this investigation was a study of how genetic architecture limits the evolution of the WS phenotype at other 'rare loci' encoding DGCs. This study resulted in multiple independent observations of the identical mutation within *nlpD* that causes a cell-division deficient phenotype. An examination of the causes of this molecular parallelism revealed local mutation rates may also alter the production of variation across loci.

6.2 Summary of findings

Identification of the mutational cause of SWS shows very directly that genetic architecture can affect the rate at which mutation is translated into phenotypic variation (Chapter 3). The mutational cause of SWS is an in-frame deletion that fuses two functionally unrelated genes: one encoding a membrane localised fatty acid desaturase (*fadA*) and the other encoding a cytosolic DGC (*fwsR*). Of the spectrum of possible mutations, only in-frame deletions within specific length limits can generate the necessary DGC-activating fusion. This stands in stark contrast to mutational pathways involving Wsp, Aws or Mws – each of which contain a negative regulator – and each of which can generate WS via any loss-of-function mutation in the negative regulator.

That genetic architecture explains the infrequent evolution of SWS (compared to WS generated via mutations in *wsp*, *aws* or *mws*) – and not selection – was further bolstered by showing that the *fadA-fwsR* mutation generated a WS type whose fitness was not significantly higher compared to WS arising from a mutation in *wspF*.

The occurrence of an adaptive fusion mutation in a genetically amenable model system (SBW25) provided opportunity to determine how such a fusion results in an adaptive phenotype. The *fadA-fwsR* fusion causes SWS by relocation of FwsR to the periphery of the cell (most likely the inner cell membrane – the predicted location of FadA), as demonstrated by fluorescent microscopy. Additionally, the translational fusion of transmembrane domains encoded by *mwsR* to *fwsR* was sufficient to cause the SWS phenotype. These findings suggest that the relocalisation of FwsR, achieved by in-

frame fusions of *fadA* and *fwsR,* is necessary for this DGC to cause the WS phenotype.

The *fadA-fwsR* fusion is one of the few adaptive gene fusion mutations identified in a real-time evolution experiment. This is significant given the importance of domain shuffling for the evolution of new genes (Patthy 2003; Vogel et al. 2004; Bashton and Chothia 2007; Peisajovich et al. 2010). The *fadA-fwsR* fusion demonstrates that a fusion can confer an immediate adaptive benefit by relocalisation of a protein domain. Few studies have been able to identify how the evolution of chimeric proteins may result in an adaptive benefit (Zhang et al. 2004; Dai et al. 2008).

Chapter 4 presented evidence that pathways to WS other than *wsp, aws* and *mws* encode either small mutational targets or require multiple mutations. By subjecting independent lines of PBR716 (SM Δ*wsp* Δ*aws* Δ*mws*) to selection in alternating static and shaken microcosms during the slow reverse evolution experiment (SREE), WS were derived with mutations at several loci aside from *wsp*, *aws* and *mws*. Similar to the loci of *fadA* and *fwsR*, these newly identified loci were the targets of a different type of mutation to those that cause WS at *wsp*, *aws* and *mws*.

These 'rare loci' can be categorised into three non-exclusive groups – loci featuring mutations to the promoter regions of genes (PFLU0956, PFLU2764 and PFLU5698), loci that involved deletion mutations across two putative genes (*fadA-fwsR* and PFLU4781), and loci associated with secondary mutations to other loci (PFLU0458, PFLU4858, PFLU0956). Most of these loci did not appear to cause WS with single mutations within the coding region of the gene (see Figure 6.1). This suggests that, like *fadA-fwsR,* these loci are not able to cause WS types by loss of function of a post-translational negative regulator. All of these loci are predicted to encode small targets (such as promoter regions) that may be mutated to cause WS, and the resulting WS are thus infrequent compared to adaptive *wsp*, *aws* or *mws* or mutants. Similarly, WS that are associated with multiple mutations have a lower chance of occurring and are less likely to colonise the air-liquid interface than are readily occurring WS mutants.

The only ‘rare locus’ that was found to contain a single intragenic mutation was PFLU0085. Mutations at this locus were identified at an intragenic region, suggestive that this region may encode a negative regulator. Mutations to this regulator must preserve the activity of the downstream DGC, and are likely converted into the WS phenotype at a lower frequency at PFLU0085 compared to *wsp*, *aws* or *mws.*

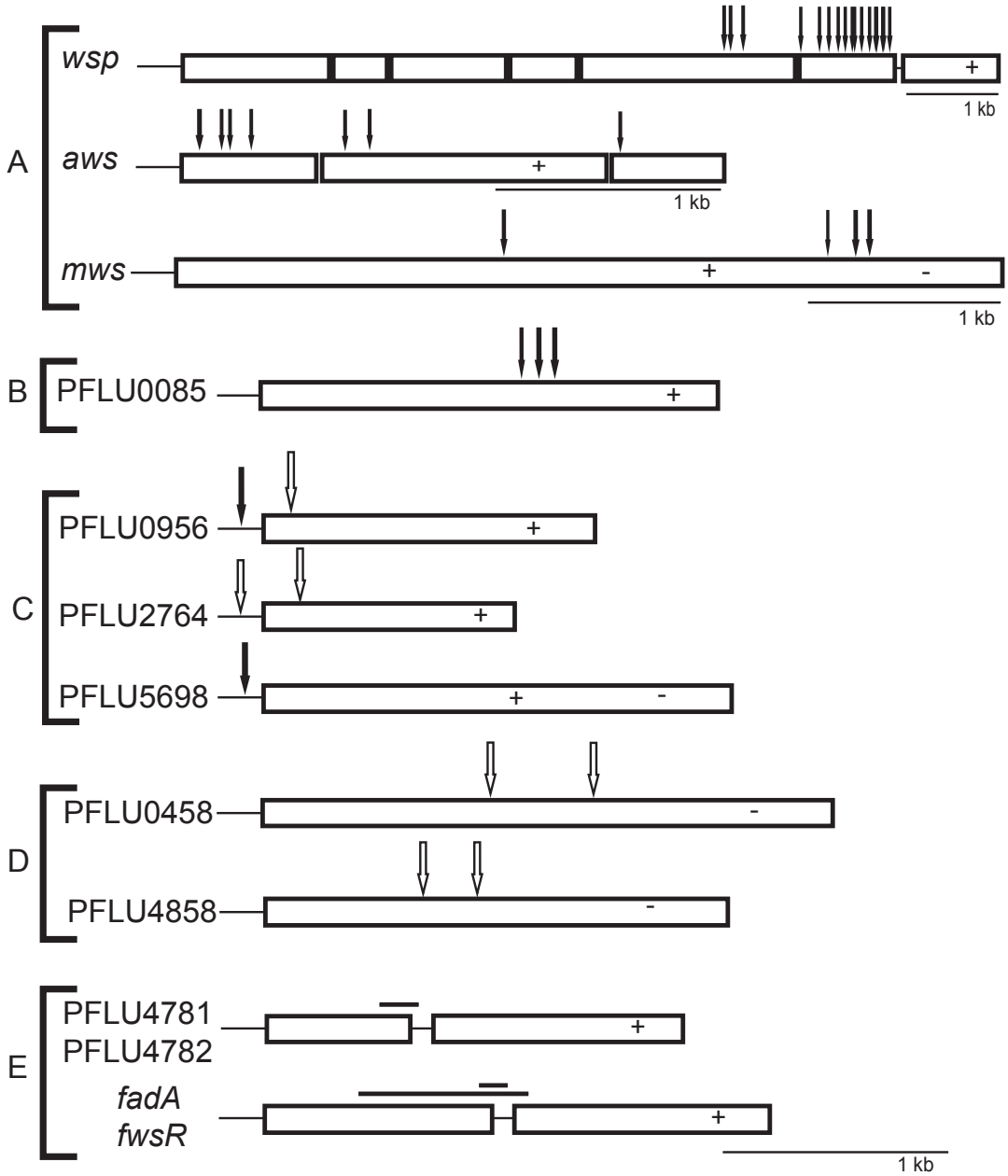

Figure 6.1: The location of mutations in genes encoding DGCs arising during the SREE. Loci are grouped by: A) mutations identified in McDonald et al. (2009); and WS-associated mutations identified during the SREE within B) an open reading frame, C) promoter regions of genes, D) open reading-frames (ORFs) of genes that are associated with secondary mutations to other genes, and E) genes that are targets of large deletions. Only groups A and B feature loci associated with single mutations within a single coding region. Black arrows indicate mutations that did not occur with other mutations; outlined arrows indicate mutations that were associated with at least one other mutation in a single mutant; Black bands above ORFs indicate deletion mutations; '+' represent predicted GGDEF domains; '-' represent predicted EAL domains.

In summary, the SREE demonstrates that multiple rare pathways to WS can be identified following the propagation of PBR716 in static microcosms, and these rare pathways appear to encode small targets for mutations that cause WS.

The results of Chapter 5 suggest that in some instances difference in local mutation rate may affect the production of variation, to the extent that high mutation rates within genes (in particular the C565T mutation within *nlpD*) may be responsible for instances of extreme molecular parallelism. Instances of parallel evolution at the nucleotide level have been identified by phylogenetic (Christin et al. 2008; Li et al. 2010; Liu et al. 2010) or experimental evolution approaches (Woods et al. 2006; Conrad et al. 2010; Gerstein et al. 2012; Tenaillon et al. 2012). Constraints on protein structure and functionality have been invoked to explain this molecular parallelism (Arendt and Reznick 2008; Christin et al. 2010; Martin and Orgogozo 2013; Stern 2013).

However an explanation may account for this molecular parallelism that invokes more than the functionality of encoded proteins. Localised high mutation rates may result in the early evolution of these mutations compared to other adaptive mutants elsewhere in the gene, resulting in the repeated evolution of one mutant type across independent lines. The discovery of repeated instances of the C565T point mutation within *nlpD* in genetically amenable *P. fluorescens* SBW25 provided an opportunity to disentangle the relative influences of selection and mutation rate as a cause of this molecular parallelism.

The likely cause of this extreme molecular parallelism at nucleotide 565 of *nlpD* is a localised high mutation rate. No evidence suggested the C565T mutation was selected in independent lines because it caused a significantly high fitness. Furthermore, there was a high frequency of reverting mutants at this locus in genotypes with a high level of transcription. This supports the putative hypothesis that the presence and activity of the *rpoS* promoter (nested within the coding frame of *nlpD*) may result in a high local mutation rate, resulting in the independent evolution of the C565T mutation.

This finding requires validation via mutation-rate assays. If such assays can be developed, this will allow a further investigation as to whether a high mutation rate is located in homologs of *nlpD*, or in other nested promoters. Importantly, this study suggests that local mutation rates may result in the parallel evolution of point mutations, which have previously been explained by the high relative fitness of such mutants.

In summary, the functional and regulatory connectivities of genetic loci bias the production of phenotypic variants that are then subject to selection. Across the set of known DGCs, there appears to be a clear bias in the ready production of WS variants, and this bias is associated with the genetic architecture encoded at these loci. Within *nlpD,* the likely existence of a highly mutable locus appears to result in the repeated finding of one *nlpD* variant over other possible variants. Whilst the functionalism of these types is necessary for the observation of these mutants, formalistic features – genetic architecture and mutation rates – appear to create a bias in the types of phenotypes that are observed in these studies.

6.3 Implications of these findings for predicting evolution

This work, and the body that it contributes to (Spiers et al. 2002; Goymer et al. 2006; Bantinaki et al. 2007; McDonald et al. 2009), has identified formalist features encoded by genomes that cause instances of parallel evolution at the regulatory and nucleotide levels. This parallel evolution was observed to occur either because some loci are able to translate mutations into phenotypic variation, or because some loci are likely to have a higher mutation rate. The identification of these formalist elements may be useful in generating meaningful predictions of the locus underpinning adaptive mutants. Systems biology approaches have successfully taken advantage of the wealth of knowledge of metabolomics to make accurate predictions of how metabolic pathways may evolve in serial propagation experiments (Ibarra et al. 2002; Lewis et al. 2010). It is possible that a deeper understanding of the production of variation may allow predictions of the genetic locus of adaptive evolution.

Several syntheses have suggested that an advanced understanding of an organism may allow predictions of the loci underpinning adaptive change (Lobkovsky and Koonin 2012; Schenk and de Visser 2013; Stern 2013; de Visser and Krug 2014). These claims have been informed by several empirical observations. Some genetic ‘hotspots’ are repeatedly found to cause phenotypic variation because they cause phenotypic change whilst minimising pleiotropic effects in developmental cascades (Stern and Orgogozo 2008; Stern and Orgogozo 2009; Stern 2013). Additionally, some genetic pathways are more likely to be followed in the evolution of a phenotype than others (Weinreich et al. 2006; Lozovsky et al. 2009), possibly due to sign epistasis preventing

alternative pathways from being utilised. In the syntheses above, the repeatability of evolution – due to pleiotropic effects or epistatic interactions – is interpreted as resulting from aspects of the genotype-fitness map. It is possible that an improved knowledge of this mapping may enable predictions of evolution (de Visser and Krug 2014).

The *P. fluorescens* model system has enabled a unique insight that predictions of the genes in which adaptive mutations arise may be aided by knowledge of the production of phenotypic variation. Assuming the existence of multiple adaptive loci underpinning a given phenotype, the existence of biases in the rate at which each locus can cause adaptive mutants may bias the course of evolution. Consequently, it may be possible to assess the likelihood at which different mutational paths are taken, particularly if those different paths contain features that may increase the production of adaptive mutants. These features may include negative regulators of an enzyme that is adaptive upon activation, or localised high mutation rates of potentially adaptive genes.

Predicting the mutational pathways that evolution takes is not possible without a good knowledge of the genotype-phenotype map and the fitness landscape of the organism of interest in a particular environment. Predictions of evolution therefore require an *a priori* assessment of the potential adaptive phenotypes, and the possible genes that may mutate and cause such phenotypes. Such predictions would also require knowledge of the rate at which each adaptive locus may produce adaptive phenotypic variation. These requirements suggest accurate predictions of evolution may only be made of well-studied organisms (or highly related organisms), about which knowledge may be ascertained.

The ability to predict the locus of evolution may allow specific predictions of the genetic changes that enable the emergence of antibiotic resistant bacteria to be made. Antibiotic resistance may be caused by a diverse range of mechanisms – such as efflux pumps, catalysis of antibiotics, and reduced permeability of the membrane (Palmer and Kishony 2013). Currently, the loci through which antibiotic resistance emerges may be assessed retrospectively by the serial passage of bacteria in a medium containing an antibiotic (Martinez et al. 2011). Assuming resistance mechanisms are encoded by genes featuring negative regulation, or higher mutation rates, the identification of these features may allow predictions of the mechanisms by which resistance may emerge. The identification of architectural features that enable

the ready emergence of antibiotic resistance may inform decisions concerning the clinical use of new antibiotics.

However, predictions of the mutational pathways of evolution based upon the architecture or mutation rates of adaptive alleles may be complicated by the innovative nature of mutation. The *fadA-fwsR* fusion provides an elegant example of the innovative capacity of genomes to generate adaptive types. Despite a relatively advanced understanding of the WS phenotype, it was difficult to predict that a fusion mutation of such disparate post-translational elements would cause SWS. Clearly, more research is required before accurate predictions of the locus of adaptation are possible. Perhaps recent claims that our understanding of evolutionary processes may enable predictions of evolution reflect the improving resolution of evolutionary biology.

Regardless of the utility of the study of genetic architecture or mutation rates, the study of both these formalistic features confers a deeper understanding of the production of variation. With this knowledge, we gain a richer comprehension of the processes that result in the origin of new species.

Reference list

Acar M, Mettetal JT, van Oudenaarden A. 2008. Stochastic switching as a survival strategy in fluctuating environments. *Nature Genetics* **40**(4): 471-475.

Aguilera A. 2002. The connection between transcription and genomic instability. *EMBO Journal* **21**(3): 195-201.

Aldea M, Hernandezchico C, Delacampa AG, Kushner SR, Vicente M. 1988. Identification, cloning, and expression of *bola*, an *ftsz*-dependent morphogene of *Escherichia coli*. *Journal of Bacteriology* **170**(11): 5169-5176.

Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution* **23**(1): 26-32.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* **2**: 2006.0008.

Badyaev AV, Foresman KR. 2000. Extreme environmental change and evolution: stress-induced morphological variation is strongly concordant with patterns of evolutionary divergence in shrew mandibles. *Proceedings of the Royal Society B - Biological Sciences* **267**(1441): 371-377.

Baehler E, de Werra P, Wick LY, Pechy-Tarr M, Mathys S, Maurhofer M, Keel C. 2006. Two novel MvaT-like global regulators control exoproduct formation and biocontrol activity in root-associated *Pseudomonas fluorescens* CHAO. *Molecular Plant-Microbe Interactions* **19**(3): 313-329.

Bailey J, Manoil C. 2002. Genome-wide internal tagging of bacterial exported proteins. *Nature Biotechnology* **20**(8): 839-842.

Bantinaki E, Kassen R, Knight CG, Robinson Z, Spiers AJ, Rainey PB. 2007. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. *Genetics* **176**(1): 441-453.

Bao Y, Lies DP, Fu H, Roberts GP. 1991. An improved tn7-based system for the single-copy insertion of cloned genes into chromosomes of gram-negative bacteria. *Gene* **109**(1): 167-168.

Barrick JE, Lenski RE. 2013. Genome dynamics during experimental evolution. *Nature Reviews Genetics* **14**(12): 827-839.

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**(7268): 1243-1247.

Bashton M, Chothia C. 2007. The generation of new protein functions by the combination of domains. *Structure* **15**(1): 85-99.

Baurle I, Dean C. 2006. The timing of developmental transitions in plants. *Cell* **125**(4): 655-664.

Beaumont HJE, Gallie J, Kost C, Ferguson GC, Rainey PB. 2009. Experimental evolution of bet hedging. *Nature* **462**(7269): 90-U97.

Beletskii A, Bhagwat AS. 1996. Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **93**(24): 13919-13924.

Beletskii A, Grigoriev A, Joyce S, Bhagwat AS. 2000. Mutations induced by bacteriophage T7 RNA polymerase and their effects on the composition of the T7 genome. *Journal of Molecular Biology* **300**(5): 1057-1065.

Bergmiller T, Ackermann M, Silander OK. 2012. Patterns of evolutionary conservation of essential genes correlate with their compensability. *PLOS Genetics* **8**(6): e1002803.

Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**(7417): 513-518.

Boell L. 2013. Lines of least resistance and genetic architecture of house mouse (*Mus musculus*) mandible shape. *Evolution & Development* **15**(3): 197-204.

Bolte S, Cordelieres FP. 2006. A guided tour into subcellular colocalization analysis in light microscopy. *Journal of Microscopy - Oxford* **224**: 213-232.

Bornberg-Bauer E, Mar Alba M. 2013. Dynamics and adaptive benefits of modular protein evolution. *Current Opinion in Structural Biology* **23**(3): 459-466.

Bowler PJ. 1979. Theodor Eimer and orthogenesis: evolution by 'definitely directed variation'. *Journal of the History of Medicine and Allied Sciences* **34**(1): 40-73.

Bowler PJ. 1992. *The Eclipse of Darwinism: Anti-Darwinian evolution theories in the decades around 1900*. London, England: The John Hopkins University Press.

Brencic A, McFarland KA, McManus HR, Castang S, Mogno I, Dove SL, Lory S. 2009. The GacS/GacA signal transduction system of *Pseudomonas aeruginosa* acts exclusively through its control over the transcription of the RsmY and RsmZ regulatory small RNAs. *Molecular Microbiology* **73**(3): 434-445.

Buckling A, Maclean RC, Brockhurst MA, Colegrave N. 2009. The Beagle in a bottle. *Nature* **457**(7231): 824-829.

Buist G, Steen A, Kok J, Kuipers OR. 2008. LysM, a widely distributed protein motif for binding to (peptido)glycans. *Molecular Microbiology* **68**(4): 838-847.

Bull JJ, Badgett MR, Wichman HA, Huelsenbeck JP, Hillis DM, Gulati A, Ho C, Molineux IJ. 1997. Exceptional convergent evolution in a virus. *Genetics* **147**(4): 1497-1507.

Burkala E, Reimers JM, Schmidt KH, Davis N, Wei P, Wright BE. 2007. Secondary structures as predictors of mutation potential in the *lacZ* gene of *Escherichia coli*. *Microbiology-SGM* **153**: 2180-2189.

Byun-McKay SA, Geeta R. 2007. Protein subcellular relocalization: a new perspective on the origin of novel genes. *Trends in Ecology & Evolution* **22**(7): 338-344.

Castang S, McManus HR, Turner KH, Dove SL. 2008. H-NS family members function coordinately in an opportunistic pathogen. *Proceedings of the National Academy of Sciences of the United States of America* **105**(48): 18947-18952.

Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U, Schirmer T. 2004. Structural basis of activity and allosteric control of diguanylate cyclase. *Proceedings of the National Academy of Sciences of the United States of America* **101**(49): 17084-17089.

Chanut-Delalande H, Fernandes I, Roch F, Payre F, Plaza S. 2006. *Shavenbaby* couples patterning to epidermal cell shape control. *PLOS Biology* **4**(9): 1549-1561.

Chevin LM, Martin G, Lenormand T. 2010. Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution* **64**(11): 3213-3231.

Choi K-H, Schweizer HP. 2006. Mini-Tn7 insertion in bacteria with single *att*Tn7 sites: example *Pseudomonas aeruginosa*. *Nature Protocols* **1**(1): 153-161.

Christen B, Christen M, Paul R, Schmid F, Folcher M, Jenoe P, Meuwly M, Jenal U. 2006. Allosteric control of cyclic di-GMP signaling. *Journal of Biological Chemistry* **281**(42): 32015-32024.

Christen M, Christen B, Folcher M, Schauerte A, Jenal U. 2005. Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *Journal of Biological Chemistry* **280**(35): 30829-30837.

Christin PA, Salamin N, Muasya AM, Roalson EH, Russier F, Besnard G. 2008. Evolutionary Switch and genetic convergence on *rbcL* following the evolution of C-4 photosynthesis. *Molecular Biology and Evolution* **25**(11): 2361-2368.

Christin PA, Weinreich DM, Besnard G. 2010. Causes and evolutionary significance of genetic convergence. *Trends in Genetics* **26**(9): 400-405.

Collin R, Miglietta MP. 2008. Reversing opinions on Dollo's Law. *Trends in Ecology & Evolution* **23**(11): 602-609.

Conrad TM, Frazier M, Joyce AR, Cho BK, Knight EM, Lewis NE, Landick R, Palsson BO. 2010. RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. *Proceedings of the National Academy of Sciences of the United States of America* **107**(47): 20500-20505.

Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, Palsson BO. 2009. Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biology* **10**(10): R118.

Cooper TF, Rozen DE, Lenski RE. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **100**(3): 1072-1077.

Corno G, Jurgens K. 2006. Direct and indirect effects of protist predation on population size structure of a bacterial strain with high phenotypic plasticity. *Applied and Environmental Microbiology* **72**(1): 78-86.

Courseaux A, Nahon JL. 2001. Birth of two chimeric genes in the Hominidae lineage. *Science* **291**(5507): 1293-1297.

Crozat E, Winkworth C, Gaffe J, Hallin PF, Riley MA, Lenski RE, Schneider D. 2010. Parallel genetic and phenotypic evolution of DNA superhelicity in experimental populations of *Escherichia coli*. *Molecular Biology and Evolution* **27**(9): 2113-2128.

Dai HZ, Chen Y, Chen SD, Mao QY, Kennedy D, Landback P, Eyre-Walker A, Du W, Long MY. 2008. The evolution of courtship behaviors through the origination of a new gene in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **105**(21): 7478-7483.

Darwin C. 1859. *On the Origin of Species* (1st ed.). London, England: John Murray.

Darwin C. 1872. *The Origin of Species* (6th ed.). London, England: John Murray.

Davies KTJ, Cotton JA, Kirwan JD, Teeling EC, Rossiter SJ. 2012a. Parallel signatures of sequence evolution among hearing genes in echolocating mammals: an emerging model of genetic convergence. *Heredity* **108**(5): 480-489.

Davies WIL, Collin SP, Hunt DM. 2012b. Molecular ecology and adaptation of visual photopigments in craniates. *Molecular Ecology* **21**(13): 3121-3158.

Dayn A, Malkhosyan S, Mirkin SM. 1992. Transcriptionally driven cruciform formation in vivo. *Nucleic Acids Research* **20**(22): 5991-5997.

De N, Navarro M, Raghavan RV, Sondermann H. 2009. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. *Journal of Molecular Biology* **393**(3): 619-633.

De Visser JAGM, Krug J. 2014. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics* **15**(7): 480-490.

De Vries H. 1909. *The Mutation Theory: Experiments and observations on the origin of species in the vegetable kingdom*. Chicago, IL: Open Court Publishing Company.

Deitsch KW, Lukehart SA, Stringer JR. 2009. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nature Reviews Microbiology* **7**(7): 493-503.

Dettman JR, Rodrigue N, Melnyk AH, Wong A, Bailey SF, Kassen R. 2012. Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Molecular Ecology* **21**(9): 2058-2077.

Diggle SP, Winzer K, Lazdunski A, Williams P, Camara M. 2002. Advancing the quorum in *Pseudomonas aeruginosa*: MvaT and the regulation of N-acylhomoserine lactone production and virulence gene expression. *Journal of Bacteriology* **184**(10): 2576-2586.

Ding Y, Zhou Q, Wang W. 2012. Origins of new genes and evolution of their novel functions. *Annual Review of Ecology, Evolution, and Systematics* **43**: 345-363.

Dutra BE, Lovett ST. 2006. Cis and trans-acting effects on a mutational hotspot involving a replication template switch. *Journal of Molecular Biology* **356**(2): 300-311.

Dykhuizen DE. 1990. Experimental studies of natural-selection in bacteria. *Annual Review of Ecology and Systematics* **21**: 373-398.

Eimer GHT. 1890. *Organic Evolution as the Result of the Inheritance of Acquired Characters According to the Laws of Organic Growth*; translated by J. T. Cunningham. London, England: Macmillan.

Elde NC, Child SJ, Eickbush MT, Kitzman JO, Rogers KS, Shendure J, Geballe AP, Malik HS. 2012. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. *Cell* **150**(4): 831-841.

Elena SF, Lenski RE. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics* **4**(6): 457-469.

Falke JJ, Bass RB, Butler SL, Chervitz SA, Danielson MA. 1997. The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annual Review of Cell and Developmental Biology* **13**: 457-512.

Ferenci T. 2005. Maintaining a healthy SPANC balance through regulatory and mutational adaptation. *Molecular Microbiology* **57**(1): 1-8.

Ferguson GC, Bertels F, Rainey PB. 2013. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. V. Insight into the niche specialist fuzzy spreader compels revision of the model *Pseudomonas* radiation. *Genetics* **195**(4): 1319-1335.

Ffrenchconstant RH. 1994. The molecular and population-genetics of cyclodiene insecticide resistance. *Insect Biochemistry and Molecular Biology* **24**(4): 335-345.

Figurski DH, Helinski DR. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proceedings of the National Academy of Sciences of the United States of America* **76**(4): 1648-1652.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K et al. 2010. The Pfam protein families database. *Nucleic Acids Research* **38**: D211-D222.

Fisher RA. 1930. *The Genetical Theory of Natural Selection*. Oxford, England: Clarendon Press.

Foster PL, Hanson AJ, Lee H, Popodi EM, Tang HX. 2013. On the mutational topology of the bacterial genome. *G3-Genes Genomes Genetics* **3**(3): 399-407.

Frederico LA, Kunkel TA, Shaw BR. 1990. A sensitive genetic assay for the detection of cytosine deamination - determination of rate constants and the activation-energy. *Biochemistry* **29**(10): 2532-2537.

Fu BD, Chen M, Zou M, Long MY, He SP. 2010. The rapid generation of chimerical genes expanding protein diversity in zebrafish. *BMC Genomics* **11**: 657.

Futuyma DJ, Keese MC, Scheffer SJ. 1993. Genetic constraints and the phylogeny of insect-plant associations - responses of *Ophraella Communa* (Coleoptera, Chrysomelidae) to host plants of its congeners. *Evolution* **47**(3): 888-905.

Gallie J. 2010. *Evolutionary and Molecular Origins of a Phenotypic Switch in Pseudomonas fluorescens SBW25.* A thesis submitted in partial fulfilment of the requirements for the degree of Ph.D. in evolutionary genetics at Massey University, Auckland, New Zealand.

Galperin MY. 2004. Bacterial signal transduction network in a genomic perspective. *Environmental Microbiology* **6**(6): 552-567.

Galperin MY. 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiology* **5**: 35.

Galperin MY, Nikolskaya AN, Koonin EV. 2001. Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiology Letters* **203**(1): 11-21.

Galton F. 1869. *Hereditary Genius*. London, England: Macmillan.

Gawel D, Fijalkowska IJ, Jonczyk P, Schaaper RM. 2014. Effect of dNTP pool alterations on fidelity of leading and lagging strand DNA replication in *E. coli*. *Mutation Research* **759**: 22-28.

Gerstein AC, Lo DS, Otto SP. 2012. Parallel genetic changes and nonparallel gene-environment interactions characterize the evolution of drug resistance in yeast. *Genetics* **192**(1): 241-252.

Giddens SR, Jackson RW, Moon CD, Jacobs MA, Zhang X-X, Gehrig SM, Rainey PB. 2007. Mutational activation of niche-specific genes provides insight into regulatory networks and bacterial function in a complex environment. *Proceedings of the National Academy of Sciences of the United States of America* **104**(46): 18247-18252.

Gilbert W. 1978. Why genes in pieces. *Nature* **271**(5645): 501-501.

Goldschmidt R. 1940. *The Material Basis of Evolution*. New Haven, CT: Yale University Press.

Gould SJ. 2002. *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press.

Gould SJ, Eldredge N. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* **3**(2): 115-151.

Gould SJ, Lewontin RC. 1979. Spandrels of San-Marco and the Panglossian paradigm - a critique of the adaptationist program. *Proceedings of the Royal Society B - Biological Sciences* **205**(1161): 581-598.

Goymer P, Kahn SG, Malone JG, Gehrig SM, Spiers AJ, Rainey PB. 2006. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. II. Role of the GGDEF regulator WspR in evolution and development of the wrinkly spreader phenotype. *Genetics* **173**(2): 515-526.

Guvener ZT, Harwood CS. 2007. Subcellular location characteristics of the *Pseudomonas aeruginosa* GGDEF protein, WspR, indicate that it produces cyclic-di-GMP in response to growth on surfaces. *Molecular Microbiology* **66**(6): 1459-1473.

Haacke W. 1893. *Gestaltung und Vererbung*. Leipzig, Germany: Weigel.

Hahn MW, Lunsdorf H, Janke L. 2004. Exopolymer production and microcolony formation by planktonic freshwater bacteria: defence against protistan grazing. *Aquatic Microbial Ecology* **35**(3): 297-308.

Hahn MW, Moore ERB, Hofle MG. 1999. Bacterial filament formation, a defense mechanism against flagellate grazing, is growth rate controlled in bacteria of different phyla. *Applied and Environmental Microbiology* **65**(1): 25-35.

Hall BM, Ma CX, Liang P, Singh KK. 2009. Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* **25**(12): 1564-1565.

Hall-Stoodley L, Costerton JW, Stoodley P. 2004. Bacterial biofilms: from the natural environment to infectious diseases. *Nature Reviews Microbiology* **2**(2): 95-108.

Hansen TF. 2006. The evolution of genetic architecture. *Annual Review of Ecology, Evolution, and Systematics* **37**: 123-157.

Heineman RH, Bull JJ, Molineux IJ. 2009. Layers of evolvability in a bacteriophage life history trait. *Molecular Biology and Evolution* **26**(6): 1289-1298.

Heidrich C, Templin MF, Ursinus A, Merdanovic M, Berger J, Schwarz H, de Pedro MA, Holtje JV. 2001. Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. *Molecular Microbiology* **41**(1): 167-178.

Hendriks G, Jansen JG, Mullenders LHF, de Wind N. 2010. Transcription and replication: far relatives make uneasy bedfellows. *Cell Cycle* **9**(12): 2300-2304.

Hengge-Aronis R. 2002. Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiology and Molecular Biology Reviews* **66**(3): 373-395.

Hengge R. 2009. Principles of c-di-GMP signalling in bacteria. *Nature Reviews Microbiology* **7**(4): 263-273.

Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**(1): 51-59.

Houle D. 1998. How should we explain variation in the genetic variance of traits? *Genetica* **102-3**: 241-253.

Huangyutitham V, Guvener ZT, Harwood CS. 2013. Subcellular clustering of the phosphorylated WspR response regulator protein stimulates its diguanylate cyclase activity. *mBio* **4**(3): e00242-13.

Hudson RE, Bergthorsson U, Ochman H. 2003. Transcription increases multiple spontaneous point mutations in *Salmonella enterica*. *Nucleic Acids Research* **31**(15): 4517-4522.

Huxley J. 1942. *Evolution. The Modern Synthesis*. London, England: George Allen and Unwin.

Ibarra RU, Edwards JS, Palsson BO. 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420**(6912): 186-189.

Jenal U, Malone J. 2006. Mechanisms of cyclic-di-GMP signaling in bacteria. *Annual Review of Genetics* **40**: 385-407.

Jones CD, Custer AW, Begun DJ. 2005. Origin and evolution of a chimeric fusion gene in *Drosophila subobscura*, *D. madeirensis* and *D. guanche*. *Genetics* **170**(1): 207-219.

Jousset A. 2012. Ecological and evolutive implications of bacterial defences against predators. *Environmental Microbiology* **14**(8): 1830-1843.

Jousset A, Rochat L, Scheu S, Bonkowski M, Keel C. 2010. Predator-prey chemical warfare determines the expression of biocontrol genes by rhizosphere-associated *Pseudomonas fluorescens*. *Applied and Environmental Microbiology* **76**(15): 5263-5268.

Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Research* **20**(10): 1313-1326.

Katju V, Lynch M. 2006. On the formation of novel genes by duplication in the *Caenorhabditis elegans* genome. *Molecular Biology and Evolution* **23**(5): 1056-1067.

Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. 2012. Experimental evolution. *Trends in Ecology & Evolution* **27**(10): 547-560.

Kearns DB, Losick R. 2005. Cell population heterogeneity during growth of *Bacillus subtilis*. *Genes & Development* **19**(24): 3083-3094.

Kim N, Jinks-Robertson S. 2012. Transcription as a source of genome instability. *Nature Reviews Genetics* **13**(3): 204-214.

Kimura M. 1968. Evolutionary rate at molecular level. *Nature* **217**(5129): 624-626.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**(4184): 107-116.

Kinnersley MA, Holben WE, Rosenzweig F. 2009. *E Unibus Plurum*: genomic analysis of an experimentally evolved polymorphism in *Escherichia coli*. *PLOS Genetics* **5**(11).

Kirisits MJ, Prost L, Starkey M, Parsek MR. 2005. Characterization of colony morphology variants isolated from *Pseudomonas aeruginosa* biofilms. *Applied and Environmental Microbiology* **71**(8): 4809-4821.

Kirschner M, Gerhart J. 1998. Evolvability. *Proceedings of the National Academy of Sciences of the United States of America* **95**(15): 8420-8427.

Kitten T, Kinscherf TG, McEvoy JL, Willis DK. 1998. A newly identified regulator is required for virulence in *Pseudomonas syringae*. *Molecular Microbiology* **28**(5): 917-929.

Klapacz J, Bhagwat AS. 2002. Transcription-dependent increase in multiple classes of base substitution mutations in *Escherichia coli*. *Journal of Bacteriology* **184**(24): 6866-6872.

Knight CG, Zitzmann N, Prabhakar S, Antrobus R, Dwek R, Hebestreit H, Rainey PB. 2006. Unraveling adaptive evolution: how a single point mutation affects the protein coregulation network. *Nature Genetics* **38**(9): 1015-1022.

Kojic M, Aguilar C, Venturi V. 2002. TetR family member PsrA directly binds the *Pseudomonas rpoS* and *psrA* promoters. *Journal of Bacteriology* **184**(8): 2324-2330.

Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene duplications. *Genome Biology* **3**(2): 0008.1.

Koza A, Moshynets O, Otten W, Spiers AJ. 2011. Environmental modification and niche construction: developing O2 gradients drive the evolution of the Wrinkly Spreader. *ISME Journal* **5**(4): 665-673.

Kroeger C, Dillon SC, Cameron ADS, Papenfort K, Sivasankaran SK, Hokamp K, Chao Y, Sittka A, Hebrard M, Haendler K et al. 2012. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proceedings of the National Academy of Sciences of the United States of America* **109**(20): E1277-E1286.

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**(3): 567-580.

Landini P, Egli T, Wolf J, Lacour S. 2014. sigmaS, a major player in the response to environmental stresses in *Escherichia coli*: role, regulation and mechanisms of promoter recognition. *Environmental Microbiology Reports* **6**(1): 1-13.

Lee H, Popodi E, Tang HX, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **109**(41): E2774-2783.

Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. 1. Adaptation and divergence during 2,000 generations. *American Naturalist* **138**(6): 1315-1341.

Levinson G, Gutman GA. 1987. Slipped-strand mispairing - a major mechanism for DNA-sequence evolution. *Molecular Biology and Evolution* **4**(3): 203-221.

Levitt M. 2009. Nature of the protein universe. *Proceedings of the National Academy of Sciences of the United States of America* **106**(27): 11079-11084.

Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D et al. 2010. Omic data from evolved *E*. *coli* are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology* **6**.

Lippert MJ, Kim N, Cho JE, Larson RP, Schoenly NE, O'Shea SH, Jinks-Robertson S. 2011. Role for topoisomerase 1 in transcription-associated mutagenesis in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **108**(2): 698-703.

Li Y, Liu Z, Shi P, Zhang J. 2010. The hearing gene *Prestin* unites echolocating bats and whales. *Current Biology* **20**(2): R55-R56.

Liu LF, Wang JC. 1987. Supercoiling of the DNA-template during transcription. *Proceedings of the National Academy of Sciences of the United States of America* **84**(20): 7024-7027.

Liu Y, Cotton JA, Shen B, Han X, Rossiter SJ, Zhang S. 2010. Convergent sequence evolution between echolocating bats and dolphins. *Current Biology* **20**(2): R53-R54.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}{}_{T}$ Method. *Methods* **25**(4): 402-408.

Lobkovsky AE, Koonin EV. 2012. Replaying the tape of life: quantification of the predictability of evolution. *Frontiers in Genetics* **3**: 246.

Lobo GP, Waite KA, Planchon SM, Romigh T, Nassif NT, Eng C. 2009. Germline and somatic cancer-associated mutations in the ATP-binding motifs of PTEN influence its subcellular localization and tumor suppressive function. *Human Molecular Genetics* **18**(15): 2851-2862.

Long M, Betran E, Thornton K, Wang W. 2003. The origin of new genes: glimpses from the young and old. *Nature Reviews Genetics* **4**(11): 865-875.

Long MY, Langley CH. 1993. Natural-selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila*. *Science* **260**(5104): 91-95.

Lozovsky ER, Chookajorn T, Brown KM, Imwong M, Shaw PJ, Kamchonwongpaisan S, Neafsey DE, Weinreich DM, Hartl DL. 2009. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proceedings of the National Academy of Sciences of the United States of America* **106**(29): 12025-12030.

Maddocks SE, Oyston PCF. 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology-SGM* **154**: 3609-3623.

Maharjan R, Seeto S, Notley-McRobb L, Ferenci T. 2006. Clonal adaptive radiation in a constant environment. *Science* **313**(5786): 514-517.

Malone JG, Jaeger T, Manfredi P, Dotsch A, Blanka A, Bos R, Cornelis GR, Haussler S, Jenal U. 2012. The YfiBNR signal transduction mechanism reveals novel targets for the evolution of persistent *Pseudomonas aeruginosa* in cystic fibrosis airways. *PLOS Pathogens* **8**(6): e1002760.

Malone JG, Jaeger T, Spangler C, Ritz D, Spang A, Arrieumerlou C, Kaever V, Landmann R, Jenal U. 2010. YfiBNR mediates cyclic di-GMP dependent small colony variant formation and persistence in *Pseudomonas aeruginosa*. *PLOS Pathogens* **6**(3): e1000804.

Malone JG, Williams R, Christen M, Jenal U, Spiers AJ, Rainey PB. 2007. The structure-function relationship of WspR, a *Pseudomonas fluorescens* response regulator with a GGDEF output domain. *Microbiology-SGM* **153**: 980-994.

Mao XZ, Ma Q, Zhou C, Chen X, Zhang HY, Yang JC, Mao FL, Lai W, Xu Y. 2014. DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic Acids Research* **42**(D1): D654-D659.

Manoil C. 2000. Tagging exported proteins using *Escherichia coli* alkaline phosphatase gene fusions. *Applications of Chimeric Genes and Hybrid Proteins, Pt A* **326**: 35-47.

Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, O'Brien J, Djimde A, Doumbo O, Zongo I et al. 2012. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487**(7407): 375-379.

Marques AC, Vinckenbosh N, Brawand D, Kaessmann H. 2008. Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. *Genome Biology* **9**(3): R54.

Martin A, Orgogozo V. 2013. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**(5): 1235-1250.

Martinez JL, Baquero F, Andersson DI. 2011. Beyond serial passages: new methods for predicting the emergence of resistance to novel antibiotics. *Current Opinion in Pharmacology* **11**(5): 439-445.

Massie JP, Reynolds EL, Koestler BJ, Cong JP, Agostoni M, Waters CM. 2012. Quantification of high-specificity cyclic diguanylate signaling. *Proceedings of the National Academy of Sciences of the United States of America* **109**(31): 12746-12751.

Matz C, Bergfeld T, Rice SA, Kjelleberg S. 2004. Microcolonies, quorum sensing and cytotoxicity determine the survival of *Pseudomonas aeruginosa* biofilms exposed to protozoan grazing. *Environmental Microbiology* **6**(3): 218-226.

Maynard Smith J, Burian R, Kauffman S, Alberch P, Campbell J, Goodwin B, Lande R, Raup D, Wolpert L. 1985. Developmental constraints and evolution. *Quarterly Review of Biology* **60**(3): 265-287.

Mazzola M, de Bruijn I, Cohen MF, Raaijmakers JM. 2009. Protozoan-induced regulation of cyclic lipopeptide biosynthesis is an effective predation defense mechanism for *Pseudomonas fluorescens*. *Applied and Environmental Microbiology* **75**(21): 6804-6811.

McDonald MJ, Gehrig SM, Meintjes PL, Zhang XX, Rainey PB. 2009. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. IV. Genetic constraints guide evolutionary trajectories in a parallel adaptive radiation. *Genetics* **183**(3): 1041-1053.

McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, Payre F, Stern DL. 2007. Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* **448**(7153): 587-590.

Meyer JR, Dobias DT, Weitz JS, Barrick JE, Quick RT, Lenski RE. 2012. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335**(6067): 428-432.

Michaels SD, Amasino RM. 1999. *Flowering locus C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**(5): 949-956.

Mitra A, Kesarwani AK, Pal D, Nagaraja V. 2011. WebGeSTer DB - a transcription terminator database. *Nucleic Acids Research* **39**: D129-D135.

Mivart SGJ. 1871. *On the Genesis of Species*. New York, NY: D Appleton.

Moxon R, Bayliss C, Hood D. 2006. Bacterial contingency loci: The role of simple sequence DNA repeats in bacterial adaptation. In *Annual Review of Genetics* **40**: 307-333.

Navarre WW, Porwollik S, Wang YP, McClelland M, Rosen H, Libby SJ, Fang FC. 2006. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* **313**(5784): 236-238.

Norman TM, Lord ND, Paulsson J, Losick R. 2013. Memory and modularity in cell-fate decision making. *Nature* **503**(7477): 481-486.

Nurminsky DI, Nurminskaya MV, De Aguiar D, Hartl DL. 1998. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature* **396**(6711): 572-575.

O'Connor JR, Kuwada NJ, Huangyutitham V, Wiggins PA, Harwood CS. 2012. Surface sensing and lateral subcellular localization of WspA, the receptor in a chemosensory-like system leading to c-di-GMP production. *Molecular Microbiology* **86**(3): 720-729.

Ohno S. 1970. *Evolution by Gene Duplication*. New York, NY: Springer.

Orr HA. 2000. Adaptation and the cost of complexity. *Evolution* **54**(1): 13-20.

Orr HA. 2005. The probability of parallel evolution. *Evolution* **59**(1): 216-220.

Paesold G, Krause M. 1999. Analysis of *rpoS* mRNA in *Salmonella dublin*: identification of multiple transcripts with growth-phase-dependent variation in transcript stability. *Journal of Bacteriology* **181**(4): 1264-1268.

Palmer AC, Kishony R. 2013. Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nature Reviews Genetics* **14**(4): 243-248.

Park C, Qian WF, Zhang JZ. 2012. Genomic evidence for elevated mutation rates in highly expressed genes. *EMBO Reports* **13**(12): 1123-1129.

Patrick WM, Quandt EM, Swartzlander DB, Matsumura I. 2007. Multicopy suppression underpins metabolic evolvability. *Molecular Biology and Evolution* **24**(12): 2716-2722.

Patthy L. 2003. Modular assembly of genes and the evolution of new functions. *Genetica* **118**(2-3): 217-231.

Paul R, Abel S, Wassmann P, Beck A, Heerklotz H, Jenal U. 2007. Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization. *Journal of Biological Chemistry* **282**(40): 29170-29177.

Paul R, Weiser S, Amiot NC, Chan C, Schirmer T, Giese B, Jenal U. 2004. Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes & Development* **18**(6): 715-727.

Pavlicev M, Hansen TF. 2011. Genotype-phenotype maps maximizing evolvability: modularity revisited. *Evolutionary Biology* **38**(4): 371-389.

Pawson T, Nash P. 2003. Assembly of cell regulatory systems through protein interaction domains. *Science* **300**(5618): 445-452.

Peisajovich SG, Garbarino JE, Wei P, Lim WA. 2010. Rapid diversification of cell signaling phenotypes by modular domain recombination. *Science* **328**(5976): 368-372.

Peters NT, Dinh T, Bernhardt TG. 2011. A fail-safe mechanism in the septal ring assembly pathway generated by the sequential recruitment of cell separation amidases and their activators. *Journal of Bacteriology* **193**(18): 4973-4983.

Polak P, Arndt PF. 2008. Transcription induces strand-specific mutations at the 5' end of human genes. *Genome Research* **18**(8): 1216-1223.

Polak P, Querfurth R, Arndt PF. 2010. The evolution of transcription-associated biases of mutations across vertebrates. *BMC Evolutionary Biology* **10**.

Queck SY, Weitere M, Moreno AM, Rice SA, Kjelleberg S. 2006. The role of quorum sensing mediated developmental traits in the resistance of *Serratia marcescens* biofilms against protozoan grazing. *Environmental Microbiology* **8**(6): 1017-1025.

Rainey PB. 1999. Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. *Environmental Microbiology* **1**(3): 243-257.

Rainey PB, Bailey MJ. 1996. Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. *Molecular Microbiology* **19**(3): 521-533.

Rainey PB, Rainey K. 2003. Evolution of cooperation and conflict in experimental bacterial populations. *Nature* **425**(6953): 72-74.

Rainey PB, Travisano M. 1998. Adaptive radiation in a heterogeneous environment. *Nature* **394**(6688): 69-72.

Rajon E, Plotkin JB. 2013. The evolution of genetic architectures underlying quantitative traits. *Proceedings of the Royal Society B - Biological Sciences* **280**(1769): 20131552.

Ranz JM, Parsch J. 2012. Newly evolved genes: moving from comparative genomics to functional studies in model systems. *BioEssays* **34**(6): 477-483.

Rausher MD. 2008. Evolutionary transitions in floral color. *International Journal of Plant Sciences* **169**(1): 7-21.

Renaud S, Auffray JC, Michaux J. 2006. Conserved phenotypic variation patterns, evolution along lines of least resistance, and departure due to selection in fossil rodents. *Evolution* **60**(8): 1701-1717.

Ripley LS. 1982. Model for the participation of quasi-palindromic DNA sequences in frameshift mutation. *Proceedings of the National Academy of Sciences of the United States of America* **79**(13): 4128-4132.

Rogers RL, Bedford T, Lyons AM, Hartl DL. 2010. Adaptive impact of the chimeric gene *Quetzalcoatl* in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* **107**(24): 10943-10948.

Rogers RL, Hartl DL. 2012. Chimeric genes as a source of rapid evolution in *Drosophila melanogaster*. *Molecular Biology and Evolution* **29**(2): 517-529.

Romling U. 2005. Characterization of the rdar morphotype, a multicellular behaviour in Enterobacteriaceae. *Cellular and Molecular Life Sciences* **62**(11): 1234-1246.

Rosnoblet C, Legrand D, Demaegd D, Hacine-Gherbi H, de Bettignies G, Bammens R, Borrego C, Duvet S, Morsomme P, Matthijs G et al. 2013. Impact of disease-causing mutations on TMEM165 subcellular localization, a recently identified protein involved in CDG-II. *Human Molecular Genetics* **22**(14): 2914-2928.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**(10): 944-945.

Ryan RP, Fouhy Y, Lucey JF, Dow JM. 2006. Cyclic Di-GMP signaling in bacteria: recent advances and new puzzles. *Journal of Bacteriology* **188**(24): 8327-8334.

Ryjenkov DA, Tarutina M, Moskvin OV, Gomelsky M. 2005. Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. *Journal of Bacteriology* **187**(5): 1792-1798.

Sarniguet A, Kraus J, Henkels MD, Muehlchen AM, Loper JE. 1995. The sigma factor σ^S affects antibiotic production and biological control activity of *Pseudomonas fluorescens* Pf-5. *Proceedings of the National Academy of Sciences of the United States of America* **92**(26): 12255-12259.

Schenk MF, de Visser J. 2013. Predicting the evolution of antibiotic resistance. *BMC Biology* **11**:14.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**(7): 676-682.

Schirmer T, Jenal U. 2009. Structural and mechanistic determinants of c-di-GMP signalling. *Nature Reviews Microbiology* **7**(10): 724-735.

Schluter D. 1996. Adaptive radiation along genetic lines of least resistance. *Evolution* **50**(5): 1766-1774.

Segre AV, Murray AW, Leu JY. 2006. High-resolution mutation mapping reveals parallel experimental evolution in yeast. *PLOS Biology* **4**(8): 1372-1385.

Shen YY, Liang L, Li GS, Murphy RW, Zhang YP. 2012. parallel evolution of auditory genes for echolocation in bats and toothed whales. *PLOS Genetics* **8**(6).

Shikuma NJ, Fong JCN, Yildiz FH. 2012. Cellular levels and binding of c-di-GMP control subcellular localization and activity of the *Vibrio cholerae* transcriptional regulator VpsT. *PLOS Pathogens* **8**(5).

Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C. 2005. Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of Arabidopsis. *Plant Physiology* **138**(2): 1163-1173.

Silby MW, Cerdeno-Tarraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SAC et al. 2009. Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome Biology* **10**(5): R51.

Spiers AJ, Bohannon J, Gehrig SM, Rainey PB. 2003. Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. *Molecular Microbiology* **50**(1): 15-27.

Spiers AJ, Kahn SG, Bohannon J, Travisano M, Rainey PB. 2002. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics* **161**(1): 33-46.

Srikanth A, Schmid M. 2011. Regulation of flowering time: all roads lead to Rome. *Cellular and Molecular Life Sciences* **68**(12): 2013-2037.

Stern DL. 2013. The genetic causes of convergent evolution. *Nature Reviews Genetics* **14**(11): 751-764.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* **62**(9): 2155-2177.

Stern DL, Orgogozo V. 2009. Is genetic evolution predictable? *Science* **323**(5915): 746-751.

Stockwell VO, Loper JE. 2005. The sigma factor RpoS is required for stress tolerance and environmental fitness of *Pseudomonas fluorescens* Pf-5. *Microbiology-SGM* **151**: 3001-3009.

Stoebel DM, Dorman CJ. 2010. The effect of mobile element IS10 on experimental regulatory evolution in *Escherichia coli*. *Molecular Biology and Evolution* **27**(9): 2105-2112.

Streisfeld MA, Rausher MD. 2011. Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. *Evolution* **65**(3): 629-642.

Sucena E, Delon I, Jones I, Payre F, Stern DL. 2003. Regulatory evolution of *shavenbaby*/*ovo* underlies multiple cases of morphological parallelism. *Nature* **424**(6951): 935-938.

Sucena E, Stern DL. 2000. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of *ovo*/*shaven-baby*. *Proceedings of the National Academy of Sciences of the United States of America* **97**(9): 4530-4534.

Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012. The molecular diversity of adaptive convergence. *Science* **335**(6067): 457-461.

Tendeng C, Soutourina OA, Danchin A, Bertin PN. 2003. MvaT proteins in *Pseudomonas* spp.: a novel class of H-NS-like proteins. *Microbiology-SGM* **149**: 3047-3050.

Thomson TM, Lozano JJ, Loukili N, Carrio R, Serras F, Cormand B, Valeri M, Diaz VM, Abril J, Burset M et al. 2000. Fusion of the human gene for the polyubiquitination coeffector *UEV1* with *Kua*, a newly identified gene. *Genome Research* **10**(11): 1743-1756.

Toomajian C, Hu TT, Aranzana MJ, Lister C, Tang CL, Zheng HG, Zhao KY, Calabrese P, Dean C, Nordborg M. 2006. A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. *PLOS Biology* **4**(5): 732-738.

Uehara T, Parzych KR, Dinh T, Bernhardt TG. 2010. Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis. *EMBO Journal* **29**(8): 1412-1422.

Vaitkevicius K, Lindmark B, Ou GW, Song TY, Toma C, Iwanaga M, Zhu J, Andersson A, Hammarstrom ML, Tuck S et al. 2006. A *Vibrio cholerae* protease needed for killing of *Caenorhabditis elegans* has a role in protection from natural predator grazing. *Proceedings of the National Academy of Sciences of the United States of America* **103**(24): 9280-9285.

van Steensel B, van Binnendijk EP, Hornsby CD, van der Voort HTM, Krozowski ZS, de Kloet ER, van Driel R. 1996. Partial colocalization of glucocorticoid and mineralocorticoid receptors in discrete compartments in nuclei of rat hippocampus neurons. *Journal of Cell Science* **109**: 787-792.

Veening JW, Stewart EJ, Berngruber TW, Taddei F, Kuipers OP, Hamoen LW. 2008. Bet-hedging and epigenetic inheritance in bacterial cell development. *Proceedings of the National Academy of Sciences of the United States of America* **105**(11): 4393-4398.

Venturi V. 2003. Control of *rpoS* transcription in *Escherichia coli* and *Pseudomonas*: why so different? *Molecular Microbiology* **49**(1): 1-9.

Viswanathan M, Lacirignola JJ, Hurley RL, Lovett ST. 2000. A novel mutational hotspot in a natural quasipalindrome in *Escherichia coli*. *Journal of Molecular Biology* **302**(3): 553-564.

Vlamakis H, Chai Y, Beauregard P, Losick R, Kolter R. 2013. Sticking together: building a biofilm the *Bacillus subtilis* way. *Nature Reviews Microbiology* **11**(3): 157-168.

Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. 2004. Structure, function and evolution of multidomain proteins. *Current Opinion in Structural Biology* **14**(2): 208-216.

Wagner GP, Zhang JZ. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics* **12**(3): 204-213.

Wang HX, Ayala JC, Benitez JA, Silva AJ. 2014. The LuxR-type regulator VpsT negatively controls the transcription of *rpoS*, encoding the general stress response regulator, in *Vibrio cholerae* biofilms. *Journal of Bacteriology* **196**(5): 1020-1030.

Wang L, Spira B, Zhou Z, Feng L, Maharjan RP, Li X, Li F, McKenzie C, Reeves PR, Ferenci T. 2010. Divergence involving global regulatory gene mutations in an *Escherichia coli* population evolving under phosphate limitation. *Genome Biology and Evolution* **2**: 478-487.

Wang W, Brunet FG, Nevo E, Long M. 2002. Origin of *sphinx*, a young chimeric RNA gene in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* **99**(7): 4448-4453.

Wang W, Zheng HK, Fan CZ, Li J, Shi JJ, Cai ZQ, Zhang GJ, Liu DY, Zhang JG, Vang S et al. 2006. High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* **18**(8): 1791-1802.

Wang XJ, Huang Y, Lavrov DV, Gu X. 2009. Comparative study of human mitochondrial proteome reveals extensive protein subcellular relocalization after gene duplications. *BMC Evolutionary Biology* **9**: 275.

Wassmann P, Chan C, Paul R, Beck A, Heerklotz H, Jenal U, Schirmer T. 2007. Structure of BeF3-modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. *Structure* **15**(8): 915-927.

Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R. 2005. Genome-wide analysis of the general stress response network in *Escherichia coli*: σ^S-dependent genes, promoters, and sigma factor selectivity. *Journal of Bacteriology* **187**(5): 1591-1603.

Weinreich DM, Delaney NF, DePristo MA, Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**(5770): 111-114.

Wessinger CA, Rausher MD. 2012. Lessons from flower colour evolution on targets of selection. *Journal of Experimental Botany* **63**(16): 5741-5749.

Weitere M, Bergfeld T, Rice SA, Matz C, Kjelleberg S. 2005. Grazing resistance of *Pseudomonas aeruginosa* biofilms depends on type of protective mechanism, developmental stage and protozoan feeding mode. *Environmental Microbiology* **7**(10): 1593-1601.

Whiteside MD, Winsor GL, Laird MR, Brinkman FSL. 2013. OrthologeDB: a bacterial and archaeal orthology resource for improved comparative genomic analysis. *Nucleic Acids Research* **41**(D1): D366-D376.

Whitman CO, Riddle O. 1919. *Orthogenetic Evolution in Pigeons*. Washington, DC: Carnegie Institution.

Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ. 1999. Different trajectories of parallel evolution during viral adaptation. *Science* **285**(5426): 422-424.

Wielgoss S, Barrick JE, Tenaillon O, Cruveiller S, Chane-Woon-Ming B, Medigue C, Lenski RE, Schneider D. 2011. Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. *G3-Genes Genomes Genetics* **1**(3): 183-186.

Wildschutte H, Wolfe DM, Tamewitz A, Lawrence JG. 2004. Protozoan predation, diversifying selection, and the evolution of antigenic diversity in *Salmonella*. *Proceedings of the National Academy of Sciences of the United States of America* **101**(29): 10644-10649.

Winkel-Shirley B. 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology* **126**(2): 485-493.

Wong A, Rodrigue N, Kassen R. 2012. Genomics of adaptation during experimental evolution of the opportunistic pathogen *Pseudomonas aeruginosa*. *PLOS Genetics* **8**(9): e1002928.

Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. 2006. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **103**(24): 9107-9112.

Wright BE, Reimers JM, Schmidt KH, Reschke DK. 2002. Hypermutable bases in the *p53* cancer gene are at vulnerable positions in DNA secondary structures. *Cancer Research* **62**(20): 5641-5644.

Wright BE, Schmidt KH, Hunt AT, Lodmell JS, Minnick MF, Reschke DK. 2011. The roles of transcription and genotoxins underlying *p53* mutagenesis in vivo. *Carcinogenesis* **32**(10): 1559-1567.

Yoshiyama K, Maki H. 2003. Spontaneous hotspot mutations resistant to mismatch correction in *Escherichia coli*: transcription-dependent mutagenesis involving template-switching mechanisms. *Journal of Molecular Biology* **327**(1): 7-18.

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ et al. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**(13): 1608-1615.

Zhang JM, Dean AM, Brunet F, Long MY. 2004. Evolving protein functional diversity in new genes of Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **101**(46): 16246-16250.

Zhang X-X, Rainey PB. 2007a. Construction and validation of a neutrally-marked strain of *Pseudomonas fluorescens* SBW25. *Journal of Microbiological Methods* **71**(1): 78-81.

Zhang X-X, Rainey PB. 2007b. Genetic analysis of the histidine utilization (hut) genes in *Pseudomonas fluorescens* SBW25. *Genetics* **176**(4): 2165-2176.

Zhang YM, Rock CO. 2008. Membrane lipid homeostasis in bacteria. *Nature Reviews Microbiology* **6**(3): 222-233.

Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012. Parallel molecular evolution in an herbivore community. *Science* **337**(6102): 1634-1637.

Zhu K, Choi KH, Schweizer HP, Rock CO, Zhang YM. 2006. Two aerobic pathways for the formation of unsaturated fatty acids in *Pseudomonas aeruginosa*. *Molecular Microbiology* **60**(2): 260-273.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**(13): 3406-3415

Appendix

Appendix for Chapter 3

Strain	Working name[a]	Gene target	Genome position[b]
SWSTN1	SWSTn3 A	Unknown	Unknown
SWSTN2	SWSTn3 C	PFLU183	208832-208833
SWSTN3	SWSTn3 D	Unknown	Unknown
SWSTN4	SWSTn3 G	PFLU3346	3702846-3702847
SWSTN5	SWSTn3 H	Unknown	Unknown
SWSTN6	SWSTn3 J	PFLU0184	209763-209764
SWSTN7	SWSTn3 K	Unknown	Unknown
SWSTN8	SWSTn3 L	PFLU0184	209924-209925
SWSTN9	SWSTn3 M	PFLU0184	209917-209918
SWSTN10	SWSTn3 N	PFLU0183	208832-208833
SWSTN11	SWSTn3 P	Unknown	Unknown
SWSTN12	SWSTn4 1A	PFLU0184	210179-210180
SWSTN13	SWSTn4 1B	PFLU0184	210179-210180
SWSTN14	SWSTn4 4B	PFLU0184	210030-210031
SWSTN15	SWSTn4 5A	PFLU0184	209650-209651
SWSTN16	SWSTn4 6B	PFLU0183	208266-208267
SWSTN17	SWSTn4 7A	PFLU0183	208217-208218
SWSTN18	SWSTn4 7B	PFLU0184	209825-209826
SWSTN19	SWSTn4 7C	PFLU0184	209783-209784
SWSTN20	SWSTn4 8A	PFLU0184	209995-209996
SWSTN21	SWSTn4 9A	PFLU0184	210199-210200
SWSTN22	SWSTn4 9B	PFLU0184	210005-210006
SWSTN23	SWSTn4 9C	PFLU0184	209833-209834
SWSTN24	SWSTn4 10A	PFLU0184	209959-209960
SWSTN25	SWSTn4 13A	PFLU0183	208161-208162
SWSTN26	SWSTn4 14A	PFLU0184	210218-210219
SWSTN27	SWSTn4 15A	PFLU0183	208396-208397
SWSTN28	SWSTn4 15B	PFLU0183	208210-208211
SWSTN29	SWSTn4 15C	PFLU0183	208831-208832
SWSTN30	SWSTn4 16A	PFLU0184	209664-209665
SWSTN31	SWSTn4 16B	PFLU0183	208008-208009
SWSTN32	SWSTn4 16C	PFLU0184	209517-209518
SWSTN33	SWSTn4 17A	*wssE*	336300-336301
SWSTN34	SWSTn4 18A	PFLU0183	208635-208636
SWSTN35	SWSTn4 20A	PFLU0183	208838-208839
SWSTN36	SWSTn4 21A	PFLU1667	1830251-1830252

Table 7.3.1: Location of the transposon insertions of SWS. [a] The allocated name given to strains during isolation and sequencing. [b]The genome location of the two flanking nucleotides at the transposon insertion.

A) Raw data for competitions of SBW25 *fadA-fwsR lacZ* (WTfwsRZ) vs LSWS

Rep	0 h count			72 h count			Malthusian fitness		Relative frequency
	Vol	WTfwsRZ	LSWS	Vol	WTfwsRZ	LSWS	WTfwsRZ	LSWS	
1	25	20	79	50	36	147	9.105	9.138	0.996
2	25	24	79	50	39	144	9.003	9.118	0.987
3	25	72	60	50	88	105	8.718	9.077	0.960
4	25	41	54	75	71	156	8.661	9.173	0.944
5	75	39	166	75	27	130	8.843	8.966	0.986
6	2.5	33	122	75	46	142	8.444	8.264	1.022
7	2.5	73	153	75	59	100	5.596	5.384	1.039
8	2.5	123	121	75	132	140	5.880	5.955	0.987
9	2.5	122	164	75	70	179	5.254	5.897	0.891
10	2.5	87	190	75	50	200	5.255	5.860	0.897
11	2.5	75	153	75	36	112	5.075	5.497	0.923
Mean relative frequency (± 95% CI)									0.966 ± 0.033

B) Raw data for competitions of SBW25 *fadA-fwsR* vs *fadA-fwsRlacZ* (WTfwsRZ)

Rep	0 h count			72 h count			Malthusian fitness		Relative frequency
	Vol	WT fwsR	WT fwsRZ	Vol	WT fwsR	WT fwsRZ	WT fwsR	WT fwsRZ	
1	75	95	71	75	57	48	8.700	8.819	0.986
2	75	36	45	75	31	33	9.061	8.900	1.018
3	75	74	69	75	30	45	8.307	8.783	0.946
4	75	32	119	75	25	66	8.963	8.621	1.040
5	75	35	36	75	33	35	9.151	9.182	0.997
6	25	37	42	75	39	41	8.164	8.088	1.009
7	2.5	56	71	75	118	110	6.554	6.247	1.049
8	2.5	108	90	75	34	23	4.653	4.445	1.047
9	2.5	119	133	75	117	148	5.792	5.916	0.979
10	2.5	96	117	75	23	24	4.380	4.225	1.037
11	2.5	54	67	75	55	41	5.827	5.318	1.096
Mean relative frequency (± 95% CI)									1.019 ± 0.028

C) Raw data for competitions of SBW25 *fadA-fwsR lacZ* (WTfwsRZ) vs IWSG

Rep	0 h count			72 h count			Malthusian fitness		Relative frequency
	Vol	WTfwsRZ	IWSG	Vol	WTfwsRZ	IWSG	WTfwsRZ	IWSG	
1	75	34	74	50	43	34	9.851	8.838	1.115
2	25	33	39	50	46	84	8.849	9.284	0.953
3	75	65	89	50	33	43	8.938	8.888	1.006
4	25	94	34	75	91	42	8.079	8.323	0.971
5	75	168	63	75	100	30	8.692	8.468	1.026
6	75	41	112	75	37	44	9.108	8.276	1.100
7	5	129	72	150	103	52	5.584	5.484	1.018
8	5	178	43	150	120	24	5.415	5.226	1.036
9	5	117	59	75	182	31	6.944	5.859	1.185
10	5	142	51	150	99	29	5.448	5.245	1.039
11	5	169	36	150	141	22	5.628	5.317	1.059
Mean relative frequency (± 95% CI)									1.046 ± 0.045

Table 7.3.2 (A to C): The raw data of volumes, CFU counts and resulting relative fitness of competition experiments with SBW25 *fadA-fwsR*. Counts were made from plates featuring various μL volumes (Vol) of 10^{-6} dilutions. Counts for each genotype are of colony forming units (CFU). Counts were taken from two trials. Malthusian fitness is a calculation of (final CFU/mL / initial CFU/mL).

Tested genotype 1	Tested genotype 2	Mean RF ± CI	S-W test	*t*-test
WTfwsRZ	LSWS	0.966 ± 0.033	0.605	0.046
WTfwsRZ	WTfwsR	1.019 ± 0.028	0.979	0.166
WTfwsRZ	IWSG	1.046 ± 0.045	0.678	0.044

Table 7.3.3: Statistical tests of the raw competition data in Table 7.3.2 (A to C). RF (relative frequency), CI (confidence interval), S-W test (Shapiro-Wilk test), *t*-test (Student's *t*-test). Values for both Shapiro-Wilk and Student's *t*-test represent *p*-values. For the Student's *t*-test, test is of a one-sample, two-tailed test that the mean frequency differs from 1. Parametric *t*-tests were performed following the Shapiro-Wilk test of normality of the raw data.

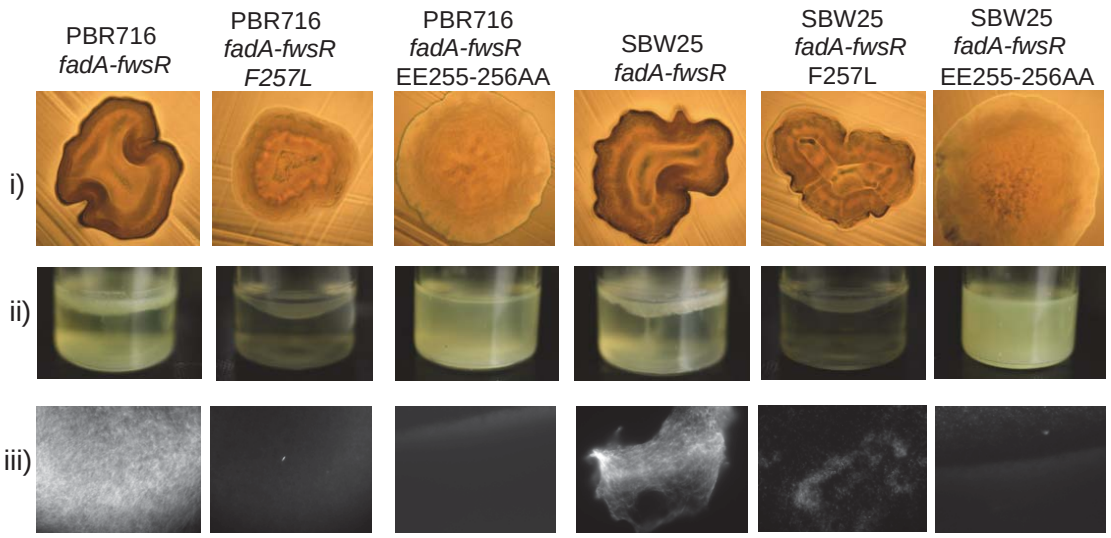

Figure 7.3.1: The mutagenesis of the GGEEF motif to GGEEL in SBW25 *fadA-fwsR* does not cause SM morphotype. Allelic replacement of F257 for L257 (in the context of the native *fwsR*, which alters the GGEEF motif to GGEEL in the putative DGC domain) in the *fwsR* gene reconstructed in both background PBR716 and ancestral SBW25, gives only partially deactivated WS phenotypes. A qualitative change occurs in both WS morphology, observable mat thinkness and calcofluor binding as mediated by ACP biosynthesis compared to the reconstructed *fwsR* mutants. However, this change is not to the same extent as that seen for the reconstructed EE255-AA255. For details of each phenotype image, see Figure 3.3.

Section 7.3.1 Transcription assays suggest *fadA* and *fwsR* are independently transcribed

The prediction that the genes *fadA* and *fwsR* are transcriptionally separated by a transcription terminator 3' to *fadA* was confirmed by *lacZ* fusions. Evidence that this transcription terminator is present suggests the two genes are independently transcribed, suggesting the *fadA-fwsR* fusion results in a promoter capture event. To assess the presence of the terminator, levels of transcription of *fadA* and *fwsR* were measured in genotypes with a *fadA-fwsR* mutation (without the terminator), and in SBW25 (with the terminator), by making promoterless '*lacZ* fusions to *fadA* or *fwsR* using pUIC3 (Rainey 1999). See Figure 7.3.2 for details of these constructs.

Measures of transcription of types expressing the *fadA-fwsR* fusion were complicated by the change in optical density (OD) associated with the WS phenotype. Because of this, another fusion genotype that was phenotypically SM was used instead of the

fadA-fwsR fusion. This type involved replacement of the intergenic region (removing the predicted transcription terminator) with stop codons (*fadA-3X-fwsR*), and was originally intended for another assay (see Section 3.3.4.2).

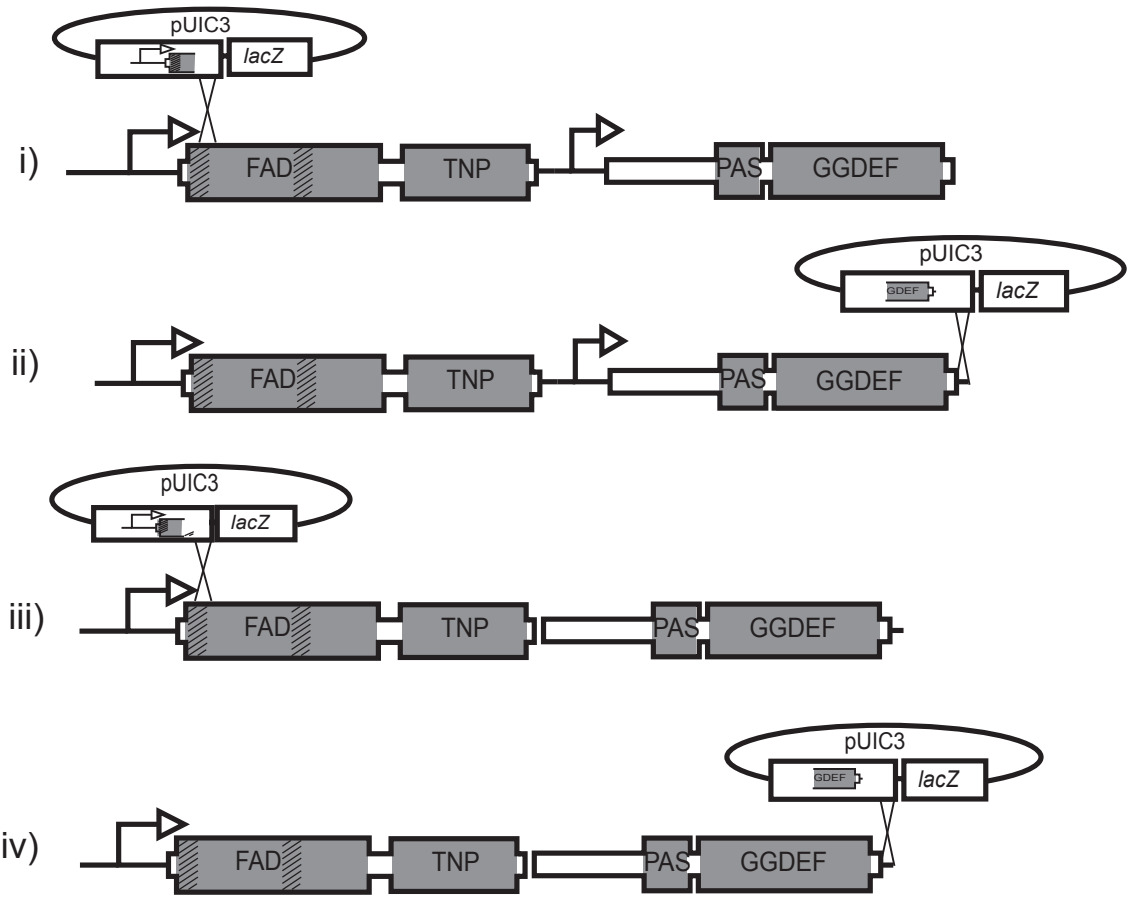

Figure 7.3.2: Diagram of the four independent pUIC3 recombinants used to measure transcription of *fwsR* with and without the fusion mutation. To make the recombinants, homologous DNA from the 3' region of *fadA* and 5' of *fwsR* was cloned into two pUIC3 constructs. Constructs were used to make four promoterless '*lacZ* fusions to the 5' region of *fadA* (i and iii) or the 3' region of *fwsR* (ii and iv). Constructs i and ii were made in SBW25, and iii and iv in a *fadA-fwsR* fusion genotype (in which the intergenic region was replaced with stop codons). These four '*lacZ* fusions would be used to demonstrate whether *fadA* and *fwsR* have independent operators.

The levels of transcription across ancestral *fadA* and *fwsR*, and the transcriptionally fused mutation (*fadA-3X-fwsR*) were determined by β-galactosidase assay. This assay was performed using cultures grown to mid-log phase in shaken microcosms of KB (see Figure 7.3.3 for results, and Tables 7.3.4 and 7.3.5 for raw data and details of statistics).

Compared to the levels of transcription reported from '*lacZ* integrated 102 nucleotides from the ATG start of *fadA*, levels reported from '*lacZ* integrated 15 nucleotides upstream of the stop codon of *fwsR* were approximately 3.8-fold higher in genotypes with the *fadA-3X-fwsR* construct compared to the ancestral SBW25 (see Figure 7.3.3). These β-galactosidase measures between genotypes are different, with a high degree of significance (Welch's *t*-test *p*-value = 0.004). The notable difference in these two genotypes is the absence of the intergenic region between *fadA* and *fwsR* in the SBW25 *fadA-3X-fwsR* fusion, which results in the higher relative measures of *fwsR* transcription. This difference strongly suggests the presence of a transcriptional terminator situated between the two genes, indicating these genes have independent operators, and that the *fadA-fwsR* fusion mutation may cause SWS because of altered transcription of the *fwsR* gene.

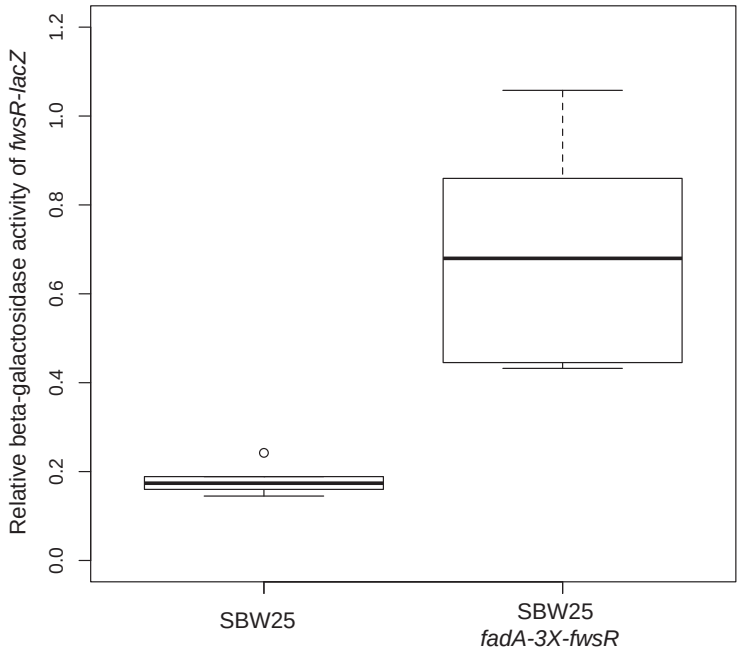

Figure 7.3.3 Transcription levels of *fadA* and *fwsR* with and without the intergenic region. β-galactoside activity (in nmol 4MU/min/cell) was determined for '*lacZ* fusions at four *in situ* positions: i) the beginning of the ORF of WT *fadA*; ii) the end of the ORF of WT *fwsR*; iii) the beginning of the ORF of the *fadA* (as *fadA*::3X-*fwsR*); and iv) the end of the ORF of *fwsR* (as *fadA*::3X-*fwsR*). The expression from the *fwsR* gene (i) is significantly lower than the expression of the *fwsR* when the intergenic region between *fadA* and *fwsR* is removed (iv) (df = 10, Student's *t*-test *p*-value = 4.2×10^{-7}). Boxplots represent six replicates. See Tables 7.3.4 and 7.3.5for raw data and statistical details.

Genotype	Measure of *fadA-lacZ*	Measure of *fwsR-lacZ*	Relative measure	Mean ± SE	Shapiro-Wilk *p*-value
SBW25	0.721, 0.983, 1.199, 1.044, 1.256, 1.166	0.125, 0.238, 0.209, 0.167, 0.237, 0.169,	0.173, 0.242, 0.174, 0.160, 0.189, 0.145	0.181 ± 0.0137	0.417
SBW25 *fad-3X-fwsR*	0.832, 1.176, 1.483, 0.891, 1.489, 0.848	0.621, 0.721, 0.641, 0.766, 0.663, 0.897,	0.746, 0.613, 0.432, 0.860, 0.445, 1.058	0.692 ± 0.100	0.452

Table 7.3.4 Raw data of the β-galactosidase activity as a proxy for transcription either 5' or 3' of the stated constructs. Data indicates measures of β-galactosidase activity in respective genotypes with *lacZ* fusion to *fadA* or *fwsR.* The relative measure is used for further analysis. Measures of β-galactosidase activity were taken as per materials and methods (2.2.11). SE (standard error).

	Variance test		*t*-test	
Comparison	***F*-stat**	***p*-value**	**Type**	***p*-value**
SBW25 vs SBW25 (fadA-3X-fwsR)	0.0188	0.0005	Welch	0.0035

Table 7.3.5: Statistical tests of the raw data in Table 7.3.4. The Student's *t*-test is of a two-sample, two-tailed test that the mean differs between tested types. The use of the Welch *t*-test was dependent upon unequal variance as determined by the *F*-test of variance.

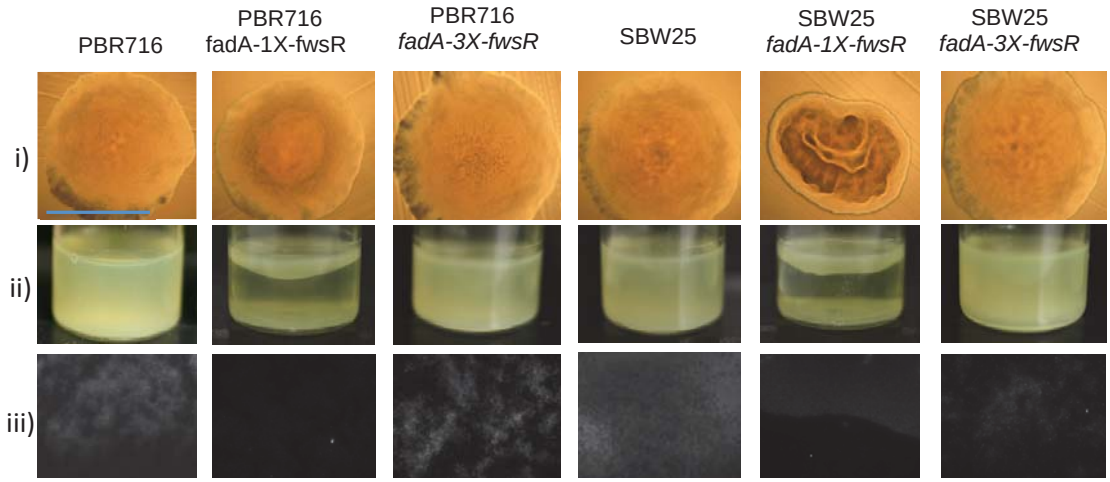

Figure 7.3.4: Single stop codon between fadA and fwsR ORFs cause a slight WS morphotype. Phenotype of the *fadA-1X-fwsR* construct reconstructed in both PBR716 (SM Δ*wsp* Δ*aws* Δ*mws*) and SBW25, as compared to the direct ancestors and the construct with three stop codons in different coding frames at the 3' end of *fadA*. The *fadA-1X-fwsR* has characteristic WS features of colonisation of the air-liquid interface and distinctive WS morphology when reconstructed in SBW25 backgrounds. In comparison to the construct made with three stop codons, the exhibition of WS traits is indicative of reduced, but not obliterated, translation of the whole genic product (for details of each phenoype image, see Figure 3.3).

Genotype	Raw data	Mean ± SE	Shapiro-Wilk *p*-value
GFP(-)	266.3, 280.4, 260.6, 283.5, 297.5, 303.1, 239.4, 333.6, 300.0, 316.7, 288.0, 327.6, 370.5, 360.2, 287.4, 330.2, 345.8, 363.5, 306.3, 316.6, 394.4,	312.9 ± 8.6	0.986
GFP(+)	1499.8, 1597.5, 1311.6, 1627.4, 1682.5, 1569.8, 1441.1, 2375.2, 2718.3, 1901.2, 2426.8, 2482.5, 2482.8, 1899.7, 2315.2, 2549.3, 2300.5, 1975.9, 1937.6, 2263.2, 2270.0	2029.0 ± 92.1	0.199
fadA-gfp	970.7, 926.5, 1153.0 , 932.8, 908.2, 829.1, 1036.4, 1258.8, 924.0, 1073.6, 917.7, 894.1, 979.6, 921.0, 1134.2, 1164.6, 937.6, 1037.0, 1225.3, 851.1, 952.0,	1001.3 ± 26.8	0.053
fwsR-gfp	984.3, 701.1, 994.0, 499.9, 856.0, 754.3, 646.1, 718.8, 910.6, 602.7, 767.8, 651.1, 643.0, 752.1, 623.3, 583.4, 450.4, 746.5, 649.0, 594.1, 509.0	697.0 ± 32.3	0.402
fadA-fwsR-gfp	901.5, 1118.7, 1194.0, 768.5, 1071.8, 1200.6, 1090.2, 855.9, 982.0, 801.6, 906.1, 954.6, 807.6, 958.0, 1389.3, 847.3245, 921.6, 920.8, 1058.1, 1011.8, 1055.6	991.2 ± 33.5	0.380
fadA::3X-fwsR-gfp	631.7, 521.0, 667.7, 678.0, 715.1, 607.7, 589.0, 708.8, 712.7, 729.1, 741.9, 665.1, 738.2, 487.7, 722.5, 693.1, 714.0, 485.8, 569.8, 659.2, 735.6	655.9 ± 18.0	0.007

Table 7.3.6: The raw data of the average level of fluorescence of genotypes. 'Raw data' represents the average level of fluorescence (corrected for the background fluorescence) along lateral transects of individual cells expressing the genotype stated in an inducible cassette. Cells were randomly picked from the corresponding phase contrast image. SE (standard error). Seven measures were taken per biological replicate (21 replicates in total).

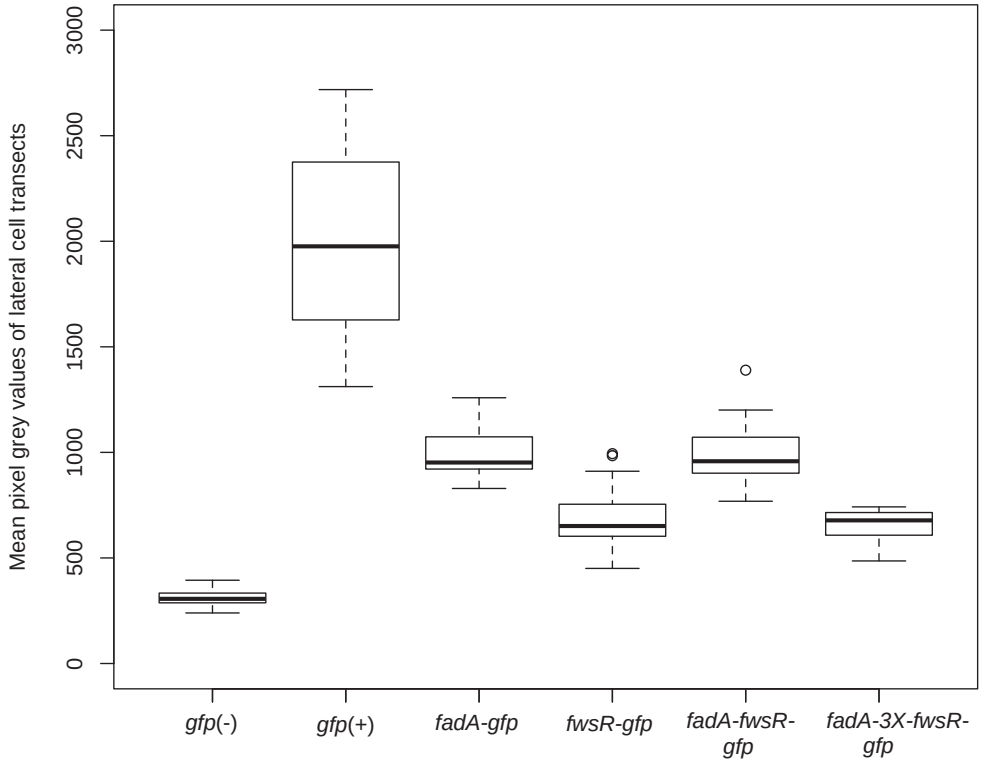

Figure 7.3.5: Cytoplasmic levels of fluorescence from protein fusions are above background levels. Both *fwsR* and *fadA-3X-fwsR* have a dispersed cytoplasmic signal. Expression of these *gfp* translational fusions required validation via analysis of fluorescence profiles of lateral transects of cells relative to negative controls. Approximately 21 randomly picked cells representative of three separate cultures were used in analysis. Increases in the levels of fluorescence from the negative control demonstrate induced expression of genes fused with GFP. Data is presented as standard boxplots. Dark bars represent median values, boxes are the range from first to third quartile, whiskers represent minima and maxima, and circles represent values outside the 1.5 interquartile range.

	Variance test		**Test**	
Comparison	***F*-stat**	***p*-value**	**Type**	***p*-vaue**
GFP(+) vs GFP(-)	115.6	4.4e-16	Welch *t*	3.2e-14
fadA-gfp vs GFP(-)	9.77	4.0e-06	Welch *t*	2.2e-16
fwsR-gfp vs GFP(-)	14.2	1.6e-07	Welch *t*	5.9e-11
fadA-fwsR-gfp vs GFP(-)	15.3	8.53e-08	Welch *t*	1.1e-15
fadA::3X-fwsR-gfp vs GFP(-)	4.4	0.0017	M-W-W	2.2e-16

Table 7.3.7: Details of statistical analyses applied to the raw data in Table 7.3.6. Table depicts comparisons of the mean values of cellular fluorescence for six different genotypes (the exception is the Mann-Wilcoxon-Whitney (M-W-W) tests, which compares the ranking of the values). The use of the Welch *t*-test was dependent upon unequal variance, and the use of the M-W-W test on the non-normality of the raw data, as tested by the Shapiro-Wilk *p*-value of the raw data.

Section 7.3.2 FadA-FwsR-GFP fusions have greater foci intensities than FadA-GFP fusions

During analyses of the localisation images (Figure 3.12), it was noticed that cells expressing *fadA-gfp* and *fadA-fwsR-gfp* presented a qualitative difference in the intensity of fluorescent foci visualised under fluorescent light (see Figure 7.3.6 for examples). This intensity was measured by recording the maximum pixel intensity of cells expressing both *fadA-gfp* and *fadA-fwsR-gfp* (see Figure 7.3.7). The intensity of the brightest foci of cells expressing *fadA-fwsR-gfp* was significantly higher than cells expressing *fadA-gfp* (M-W-W-test, *p*-value = 0.015) (see Figure 7.3.7).

The intensity of foci in cells expressing *fadA-fwsR-gfp* may be explained by the association of the encoded DGC with a membrane-located protein that would increase the cluster size of the DGC, resulting in an increase in the associated foci intensity. Such a scaffold could be the WSS complex, through which the DGC encoded by *fadA-fwsR* is predicted to regulate ACP production. To test the intensity of *fadA-fwsR-gfp* foci without the WSS complex, both *fadA-gfp* and *fadA-fwsR-gfp* expression cassettes were transformed into an SBW25 strain with the entire *wss* operon deleted. There was no significant reduction in intensity of foci of cells expressing *fadA-fwsR-gfp* to the levels expressed by *fadA-gfp* in the Δ*wss* genotype (see Figures 7.3.6 and 7.3.7, for raw data and details of statistics see Tables 7.3.8 and 7.3.9. This suggests that the WSS complex has no role in the location of the DGC encoded by *fadA-fwsR*. Multiple other factors may account for the increased intensity of the *fadA-fwsR*-encoded proteins compared to the *fadA*-encoded proteins. Such a factor may be the oligomeric nature of active DGCs, with localisation studies demonstrating WspR forms active clusters following phosphorylation (Huangyutitham et al. 2013). Further work on the clustering or activation of the DGC encoding by *fadA-fwsR* is beyond the scope of this study.

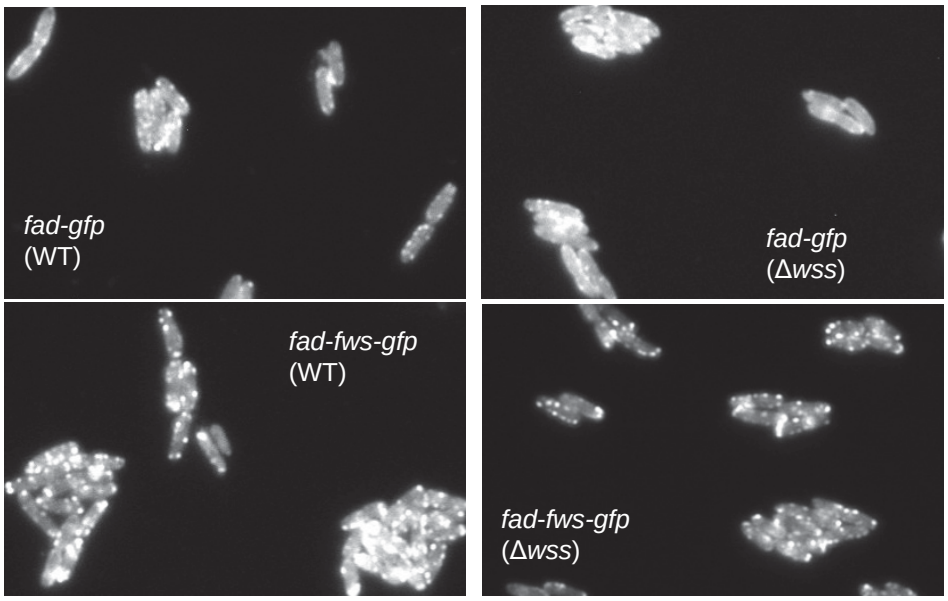

Figure 7.3.6: The foci for *fadA-fwsR-gfp* compared to *fadA-gfp* appear brighter and larger in both SBW25 and Δ*wss* genotypes. Images presented are of fluorescent images of the stated genotypes. Fluorescent microscopy was performed as per materials and methods 2.2.13.

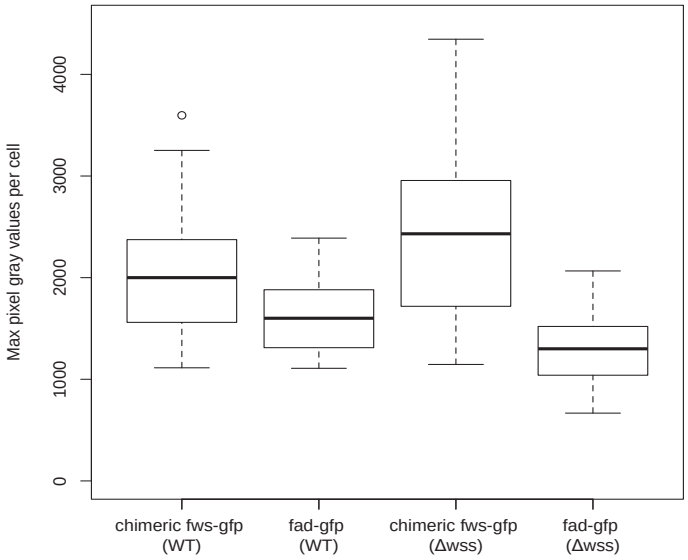

Figure 7.3.7: The maximum pixel intensities per cell are higher for *fadA-fwsR-gfp* compared to *fadA-gfp*. Measures of the maximum pixel grey values for genotypes photographed under fluorescence are depicted. Boxplots represent the distribution of the maximum pixel values for 30 cells from each genotype across three biological replicates (10 cells per replicate). Pixel values have been corrected for background fluorescence. The median values of *fadA-fwsR-gfp* (expressed in SBW25) are significantly higher than the median values for *fadA-gfp* (expressed in SBW25) (M-W-W-test, *p*-value = 0.015), and the mean values of *fadA-fwsR-gfp* (expressed in SBW25 Δ*wss*) are significantly higher than the mean values for *fadA-gfp* (expressed in SBW25 Δ*wss*) (Welch *t*-test , *p*-value $= 1.7 \times 10^{-8}$).

Genotype	Raw data	Mean ± SE	Shapiro-Wilk *p*-value
FadA-fwsR-gfp (WT)	1764, 1565, 2011, 1996, 2712, 2148, 1432, 1432, 1113, 1735, 1523, 2039, 1560, 2374, 1246, 1670, 1460, 3191, 2935, 1404, 3252, 3597, 2004, 2527, 2130, 3046, 2036, 1619, 2020, 1762,	2043 ± 286	0.02128
fadA-gfp (WT)	1513, 2389, 1282, 1719, 1311, 1649, 1285, 1283, 1399, 1176, 1864, 1733, 2340, 1751, 1131, 2197, 1792, 1548, 1178, 2139, 1405, 1464, 1881, 1887, 1366, 1597, 1604, 1108, 2206, 2041,	1641 ± 68	0.1622
FadA-fwsR-gfp (Δ*wss*)	1628, 2742, 1993, 3127, 2555, 2036, 1716, 2419, 2742, 4069, 2882, 2444, 2365, 2622, 3540, 4346, 2956, 2766, 1718, 1290, 3694, 2280, 3152, 3151, 1406, 2114, 1610, 1242, 1146, 2114,	2462 ± 151	0.6307
fadA-gfp (Δ*wss*)	1101, 1174, 1163, 1734, 1007, 1510, 1057, 2066, 1219, 1972, 921, 782, 667, 788, 1288, 854, 872, 1573, 1344, 1312, 1532, 1470, 1083, 1520, 1378, 1926, 1854, 1393, 1040, 1373,	1299 ± 68	0.569

Table 7.3.8: The raw data of the maximum pixel grey values of cells expressing genes tagged with GFP. 'Raw data' represents the maximum pixel grey values for a given cell. Ten cells were picked per biological replicate, and three biological replicates were used per genotype (resulting in 30 samples). Cells were picked at random from corresponding phase contrast images, and the pixel grey values measured from the entire cell. Maximum values were then identified, and corrected for the background fluorescence. SE (standard error). Shapiro-Wilk tests were used to ascertain normality.

	Variance test		Test	
Comparison	***F*-stat**	***p*-value**	**Type**	***p*-value**
FadA-fwsR-gfp (WT) vs *fadA-gfp* (WT)	3.0131	0.0040	M-W-W	0.01501
FadA-fwsR-gfp (Δwss) *fadA-gfp* (Δwss)	4.9659	4.52 e-05	*Welch T*	1.669e-08

Table 7.3.9: Details of statistical analyses applied to the raw data in Table 7.3.8. Table depicts comparisons of genotypes measured for the maximum pixel grey values of individual cells expressing a given GFP-tagged protein. Compared genotypes are shown in the column 'Comparison'. The Mann-Wilcoxon-Whitney (M-W-W) test compares the ranking of the values for genotypes, which were not normally distributed, as tested by the Shapiro-Wilk test in the above table. The use of the Welch *t*-test was dependent upon unequal variance (which would otherwise be parametric).

Appendix for Chapter 4

Line	Round	Working strain name	Forward primer	Reverse primer	WT product size	Temp
1	1	1w0	PF0085_1010_F PL	PF0085_1730_R ADF	700	57.3
1	2	1s1	PFLU0304_3060 (PL)	PFLU0304_DR (PL)	900	56.3
1	3	1w1	PF1302F	PF1302R	700	57.9
1	4	1s2	PF1302F	PF1302R	700	57.9
1	5	1w2	PFLU0478_F	PFLU0478_R	790	59.3
1	6	1s3	1s4Del_F	1s4del_R	NA	59
1	7	1w3	PFLU0476_F	PFLU0476_R	680	59
1	8	1s4	PFLU0477_F	PFLU0477_R	660	59.3
2	1	2w0	PF183F	PF183R	1000	57.8
2	2	2s1	PFLU0301_632F	PFLU0301_1629R	1000	55.8
2	3	2w1	PFLU0301_632F	PFLU0301_1629R	1000	55.8
2	4	2s2	PF184F	PF184R	1000	58
2	5	2w2	PF184F	PF183R	1500	57.8
2	6	2s3	PF183F2	PF183R2	1000	57.2
2	7	2w3	PF1302F	PF1302R	700	57.9
2	8	2s4	PF1302F	PF1302R	700	57.9
3	1	3w0	PFLU0085_1010_F (PL)	PFLU0085_D_R (PL)	1100	57.1
3	2	3s1	PFLU0085_1010_F (PL)	PFLU0085_D_R (PL)	1100	57.1
3	3	3w1	3-13F	3-13R	640	56.2
3	3	3w1	PFLU4744_DR	PFLU4744_UF	700	55.7
3	4	3s2	PFLU0304_3060	PFLU0304_DR	900	56.3
3	5	3w2	PFLU0304_3060	PFLU0304_DR	900	56.3
3	6	3s3	PF5748_3300_F	PF5748_4000_R	680	58.5
3	7	3w3	PF3018_U_F	PF3018_600_R	700	58.7
3	8	3s4	PF3016U_F	PF3016D_R	750	60.4
4	1	4w0	PFLU0956_U_F PL	PFLU0956_250_R	300	56.3
4	2	4s1	PF4939_40U_F	PF4939_D_R	500	59.9
4	3	4w1	NK	NK		
4	4	4s2	PFLU2485_U_F	PF2512_D_R	NA	57.4
4	4	4s2	PFLU3409_U_F	PFLU3409_R	620	60.1
4	4	4s2	PFLU0956_870_F	PFLU0956_D_R	950	55
4	5	4w2	PFLU1739_2300_F	PFLU1739_2900_R	650	58
4	6	4s3	PF5034_660_F	PF5034_1210_R	580	58.8
4	7	4w3	PFLU3677_F	PFLU3677_R	630	59
4	8	4s4	PF0085_1010_F PL	PF0085_1730_R ADF	700	57.1
5	1	5w0	PFLU4744_UF PL	PFLU4744_DR PL	650	55.7
5	1	5w0	3-13F	3-13R	640	56.2
5	2	5s1	PFLU0304_3060	PFLU0304_DR	900	56.3

Line	Round	Working strain name	Forward primer	Reverse primer	WT product size	Temp
5	3	5w1	PFLU0304_3060	PFLU0304_DR	900	56.3
5	4	5s2	PFLU4939_340U_F	PFLU4939_150_R	510	59.7
5	5	5w2	NK	NK		
5	6	5s3	PF0621_400_F	PF0621_850_R	475	58.1
5	7	5w3	PFLU4858_530F	PFLU4858_1070_R	550	57.5
5	8	5s4	PFLU0304_3060	PFLU0304_DR	900	56.3
6	1	6w0	PFLU4781_330_F	PFLU4782_443R	730	57.5
6	2	6s1	PFLU4782_250_F	PFLU4782_900_R	630	57.1
6	3	6w1	PFLU0956_U_F PL	PFLU0956_250_R	300	56.3
6	3	6w1	PFLU4858_350F	PFLU4858_1070_R	680	57.5
6	4	6s2	PFLU4939_340U_F	PFLU4939_150_R	520	59.7
6	5	6w2	PFLU6004_100_F	PFLU6004_750_R	680	56.2
6	6	6s3	wss21Pf	PFLU0304_DR	2000	56.3
6	7	6w3	PFLU1468_F	PFLU1468_R	635	57.1
6	8	6s4	PFLU1467_U_F	PFLU1467_400_R	530	57.5
7	1	7w0	PFLU2441_150_F	PFLU2441_600_R	500	59.3
7	1	7w0	PFLU2764_100U_F	PFLU2764_900_R	1000	56.7
7	1	7w0	PFLU2764_100U_F	PFLU2764_900_R	1000	56.7
7	2	7s1	PFLU2764_100U_F	PFLU2764_900_R	1000	56.7
7	3	7w1	PFLU2764_100U_F	PFLU2764_900_R	1000	56.7
7	4	7s2	PFLU2764_100U_F	PFLU2764_900_R	1000	56.7
7	5	7w2	PFLU2764_100U_F	PFLU2764_900_R	1000	56.7
7	6	7s3	PFLU2764_100U_F	PFLU2764_900_R	1000	56.7
7	7	7w3	PFLU2764_100U_F	PFLU2764_900_R	1000	56.7
7	8	7s4	PFLU2764_100U_F	PFLU2764_900_R	1000	56.7
8	1	8w0	PFLU5698_400_UF	PFLU5698_270_R	685	57.6
8	1	8w0	PFLU4414_560_F	PFLU4414_1330_R	789	57.2
8	2	8s1	PFLU0304_3060	PFLU0304_DR	900	56.3
8	3	8w1	PF1302F	PF1302R	700	57.9
8	4	8s2	PF1302F	PF1302R	700	57.9
8	5	8w2	PF1302F	PF1302R	700	57.9
8	6	8s3	PFLU5698_950_F	PFLU5698_1570_R	670	58.7
8	7	8w3	PF1302F	PF1302R	700	57.9
8	8	8s4	PFU4417_400_F	PFLU4417_300_DR	640	57.6

Table 7.4.1: Primer pairs used to order mutations in each line in the SREE. 'NK' represents an unknown mutation, 'WT product size' is in base-pairs, 'Temp' is the annealing temperature of primers (in °C).

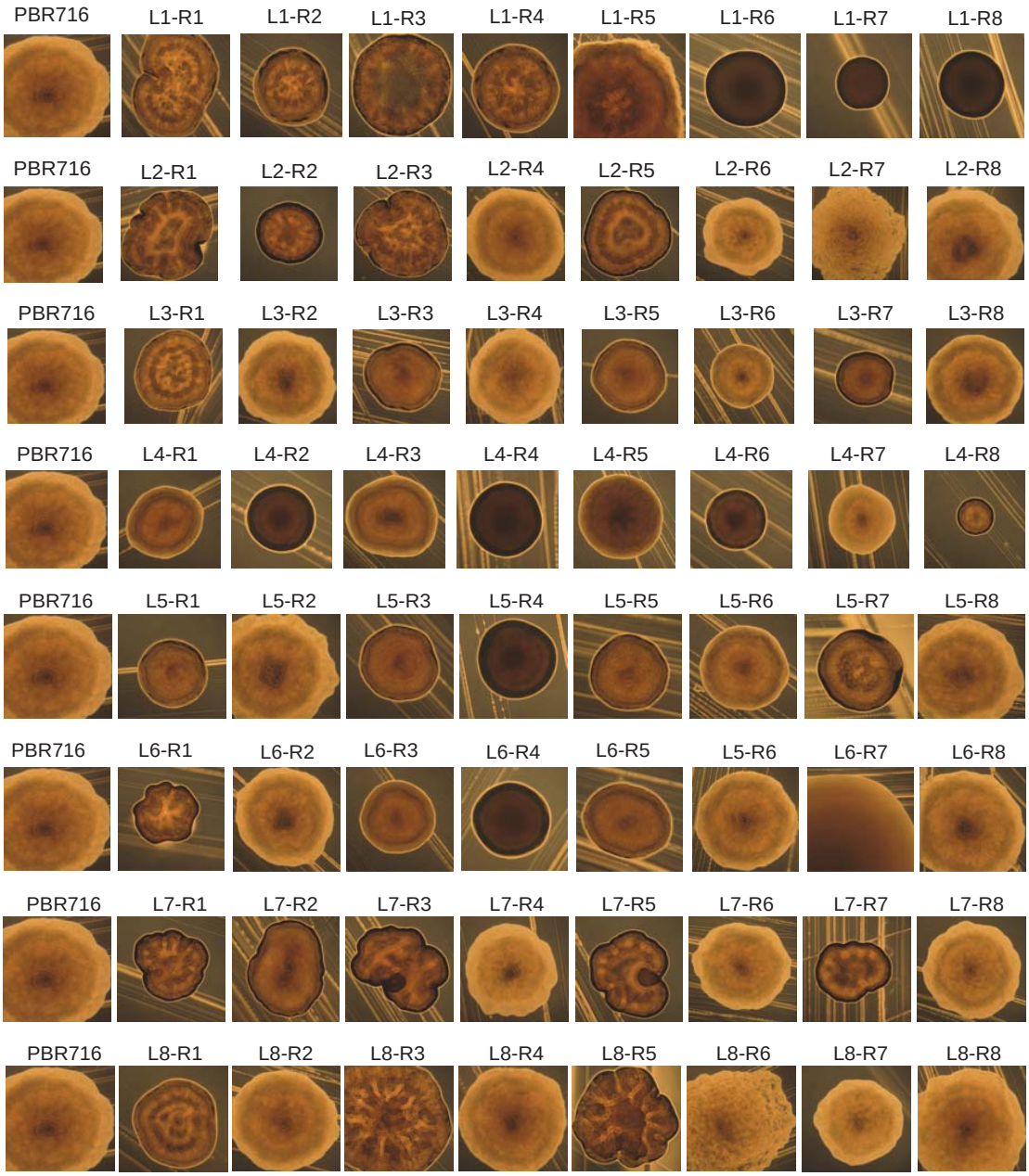

Figure 7.4.1: Morphologies of the mutants isolated during the SREE. Colonies are grown in King's B agar for approximately 30 h (at 28°C). Pictures are approximately 3mm in width.

A) Line 1

Round	Selective transfers	Mutation locus	Nucleotide change	Amino acid change	Coordinates	Mutation effect
1	4	PFLU0085	T>C, 1316	L439P	86014	Substitution
2	2	wssE	del 83bp, 3142-3224	V1048fs	337085-337167	Frameshift
3	2	nlpD	C>T, 565	Q189*	1439177	Truncation
4	1	nlpD	A>C, 566	*189S	1439178	Elongation
5	2	PFLU0478	del 6bp, del 874-879	L292-E293	542057-542062	Deletion in-frame
6	3	PFLU1734-PFLU1755			~1901200 to ~1923000	Chromosomal deletion
6	-	PFLU3358-PFLU3971			~3715600 to ~4387700	Chromosomal duplication
7	3	PFLU0476	T>A, 290	L97Q	539120	Substitution
8	2	PFLU0477	A>C, 723	K241N	540419	Substitution

B) Line 2

Round	Selective transfers	Mutation locus	Nucleotide change	Amino acid change	Coordinates	Mutation effect
1	2	PFLU0183, PFLU0184	del 218bp, 1127-(-24),	S376fs: F-8	208965-209182	Gene fusion in-frame
2	2	PFLU0301	dup24bp, 1185-1108,	I362-A369dup	329370-329393	Insertion in-frame
3	1	PFLU0301	del24bp, 1109-1132,	I370-A377del	329370 - 329393	
4	2	PFLU0183, PFLU0184	C>T, 693	W231*	209616	Truncation
5	1	PFLU0183, PFLU0184	del890bp, 341(PFLU184) -82(PFLU183)	R114fs: L28	208860-209968	Gene fusion in-frame
6	2	PFLU0183, PFLU0184	A>G, 617 (875 FWS)	L206P(PFLU183) L292P(FWS)	208325	Substitution
7	1	PFLU1301	C>T, 565	Q189*	1439177	Truncation
8	2	PFLU1301	A>G, 566	*189W	1439178	Elongation

C) Line 3

Round	Selective transfers	Mutation locus	Nucleotide change	Amino acid change	Coordinates	Mutation effect
1	2	PFLU0085	T>C, 1316	L439P	86014-86014	Substitution
2	2	PFLU0085	del12bp, 1755-1766	I585M	86453-86464	Deletion in-frame
3	2	PFLU4744	G>A, 77	R26Q	5217873-5217873	Substitution
3	-	PFLU0458	T>G, 1801	T601P	513443-513443	Substitution
4	4	PFLU0304	insG, 3203-3204	G1068fs	337146	Frameshift
5	1	PFLU0304	delG, 3204	G1068fs	337146	Frameshift
6	2	PFLU5748	T>G, 3685	T1229P	6301004	Substitution
7	1	PFLU3018	delA, 299	K100R	3289101	Frameshift
8	3	PFLU3016	del268bp, 80-347	E27fs	3284718-3284985	Frameshift

D) Line 4

Round	Selective transfers	Mutation locus	Nucleotide change	Amino acid change	Coordinates	Mutation effect
1	1	Intergenic PFLU0956	A>C, -56		1062817-1062817	NA
2	5	PFLU4939	T>G, 161	I54A	5419678	Substitution
3	2	PFLU5306 - PFLU5309			5827900-5832500	Duplication
3	-	PFLU2486–PFLU2487			2700700-2703000	Duplication
4	4	PFLU2486 – PFLU2512	del 26,900bp		2698940-2725800	Deletion
4	-	PFLU3409	G>A, -88		3771356-3771356	NA
5	6	PFLU0956	C>T, 1057	E353K	1061705	Substitution
5	-	PFLU1739	A>G, 2560	F854L	1906507	Substitution
6	3	PFLU5034	G>A, 1007	S336F	5529040	Substitution
7	1	PFLU3677	T>C, 1742	D581G	4071838	Substitution
8	10	PFLU0085	T>G, 1340	V447G	86038	Substitution

E) Line 5

Round	Selective transfers	Mutation locus	Nucleotide change	Amino acid change	Coordinates	Mutation effect
1	2	PFLU4744	G>A, 97	A33T	5217893	Substitution
1	-	PFLU0458	delGTCAGG TTG, 1399-1407	N470-T472del	513829-513837	Deletion in-frame
2	2	PFLU0304	delG, 3203	G1068fs	337146	Frameshift
3	1	PFLU0304	insG, 3202-3204	NA	337145-337147	Genetic reversal
4	4	intergenic PFLU4939	A>T, -51	NA	5419467	NA
5	3	PFLU5306 - PFLU5309	NA	NA	5827900-5832500	Duplication
5	-	PFLU2486 – PFLU2487	NA	NA	2700700 - 2703000	Duplication
6	2	PFLU0621	T>G, 625	T209P	706478	Substitution
7	1	PFLU4858	delCGTG, 858-861	C286fs	5335046-5335049	Frameshift
8	2	PFLU0304	del 83 bp, 3142-3224	V1048fs	337085-337167	Frameshift

F) Line 6

Round	Selective transfers	Mutation locus	Nucleotide change	Amino acid change	Coordinates	Mutation effect
1	3	PFLU4781	del 184bp, 527-83(3')	G176fs	5263626 - 5263809	Deletion
2	3	PFLU4782	T>G, 575	L192R	5264492	Substitution
3	2	PFLU0956	C>T, 62	S21N	1062700	Substitution
3		PFLU04858	G>T, 514	G172C	5334702	Substitution
4	5	PFLU4939	G>T, 31	E11*	5419548	Truncation
5	3	PFLU5306-PFLU5309	NA	NA	5827900 5832500	Duplication
5	-	PFLU2486 – PFLU2487	NA	NA	2700700 - 2703000	Duplication
5	-	PFLU6004	delCTG, 394-396	L132del	6559923 - 6559925	Deletion in-frame
6	5	PFLU0304	C>T, 2213	A738V	336156	Substitution
7	6	PFLU1468	G>T, 340	E114*	1614050	Truncation
8	4	PFLU1467	delT, 99	I33fs	1613195	Frameshift

G) Line 7

Round	Selective transfers	Mutation locus	Nucleotide change	Amino acid change	Coordinates	Mutation effect
1	3	PFLU2441	C>T, 446	S149F	2661314	Substitution
1	-	PFLU2764	C>T, -4	NA	3056567	NA
1	-	PFLU2764	A>T, 113	Q38L	3056683	Substitution
2	2	PFLU2764	dup9bp, 208-216	70I-72Vdup	3056778 3056786	Insertion in-frame
3	1	PFLU2764	del9bp, 208-216	70I-72Vdel	3056778 3056786	Genetic reversal
4	2	PFLU2764	dup15bp, 479-493	T160-L164dup	3057049 3057063	Duplication in-frame
5	1	PFLU2764	del15bp, 479-493	T160-L164dup	3057049 3057063	Genetic reversal
6	1	PFLU2764	delT, 280	C94fs	3056850	Frameshift
7	2	PFLU2764	insC, 281-282	C94fs	3056851-3056852	Genetic reversal
8	3	PFLU2764	A>G, 665	H222R	3057235	Substitution

H) Line 8

Round	Selective transfers	Mutation locus	Nucleotide change	Amino acid change	Coordinates	Mutation effect
1	2	PFLU4414	delTGGTGC, 1118-1123	L373-V374del	4879311-4879316	Deletion in-frame
1	-	PFLU5698	C>T, -73		6249949	NA
2	2	PFLU0304	83bp del, 3142-3224	V1048fs	337085-337167	NA
3	1	PFLU1301	C>T, 565	Q189*	1439177	Truncation
4	2	PFLU1301	G>T, 567	*189Y	1439179	Elongation
5	4	PFLU1301	C>T, 709	Q237*	1439321	Truncation
6	2	PFLU5698	G>C, 1266	Y422*	6248611	Truncation
7	3	PFLU1301	A>G, 710	*237W	1439322	Elongation
8	1	PFLU4417	A>C, 659	V220G	4881876	NA

Table 7.4.2 (A-H): The temporally ordered set of all identified mutations from each line of the SREE. Selective transfers account for the number of passages in static incubated KB microcosms until the observation of morphologically distinct types. Nucleotide change is numbered relative to the first base-pair of the ORF (of the mutational locus) and amino acid change is relative to the predicted start codon of the gene. Coordinates are genome coordinates, and mutation effect is the predicted effect on the protein.

Appendix for Chapter 5

Section 7.5.1: SBW25 *nlpD* C565T is less fit than WS types

In order to determine whether WS types may exclude the evolution of *nlpD* mutants, the relative fitness of the C565T *nlpD* mutants relative to known WS was measured. To determine this, a pre-existing strain was utilised (PBR923 – used with permission from C. Kost) that featured the C565T mutation reconstructed in SBW25. The fitness of this strain was measured relative to both LSWS (marked with *lacZ*), a high-fitness WS, and WS^G, a WS of low fitness relative to the IWS collection (Bantinaki et al. 2007), in competitions in static microcosms over 48 hours (see Table 7.5.1 and Figures 7.5.1 for results and Table 7.5.2 for raw data and Table 7.5.3 for details of statistics). Relative to LSWS, PBR923 (C565T reconstructed in SBW25) is of a lower fitness with high significance (*t*-test *p*-value 1.3×10^{-5}). The relative fitness of PBR923 relative to IWSG was also lower, with high significance (*t*-test *p*-value 2.6×10^{-5}). This strongly indicates that if an *nlpD* mutant were to arise simultaneous to a WS mutant, competition would likely reduce *nlpD* mutants to a low observable frequency.

Competition	Competitor 1	Competitor 2	Mean RF ± 95% CI[a]	d.f[b]	*t*-test *p*-value[c]
A	WT *nlpD*(C565T)	LSWS-*lacZ*	0.530 ± 0.103	7	1.29×10^{-5}
B	WT *nlpD*(C565T)	IWSG	0.651 ± 0.085	7	2.62×10^{-5}
C	IWSG	LSWS-*lacZ*	0.899 ± 0.058	7	0.0047
D	LSWS	LSWS-*lacZ*	1.007 ± 0.051	7	0.771

Table 7.5.1: Statistical tests on competition data of PBR923 directly compared to WS. Competitions of LSWS and IWSG were included to check the relative fitness of these strains, and LSWS-*lacZ* with LSWS to check neutrality of the marker. [a]Mean relative frequency (RF) ± 95% confidence interval (CI). [b]Degrees of freedom (*n* - 1). [c]Student's *t*-test *p*-value for one-sample, two-tailed, *t*-test that the relative frequency differs from 1.

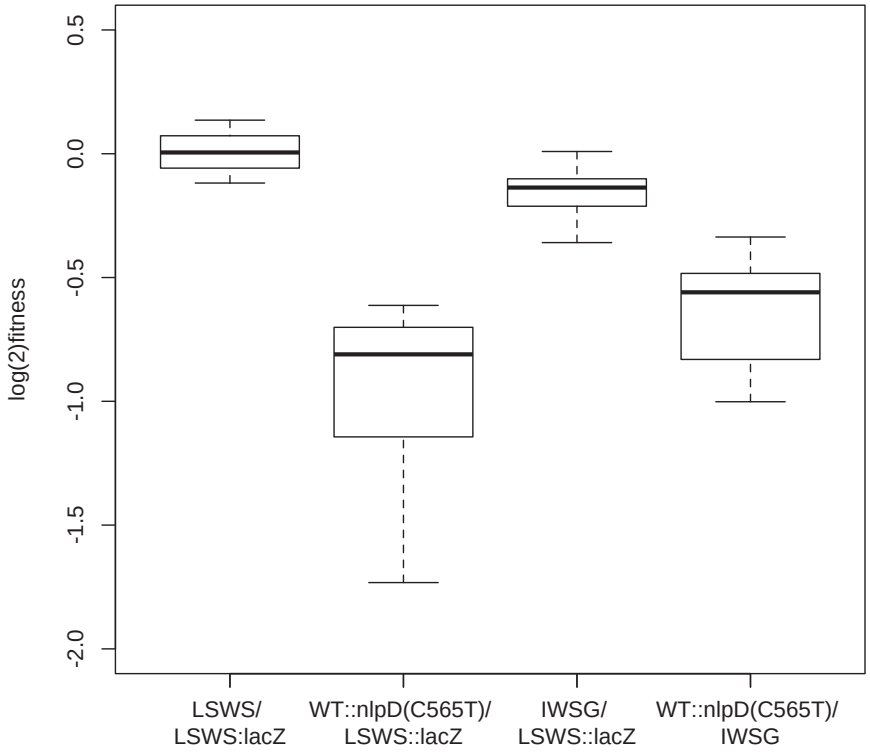

Figure 7.5.1: Fitness of WT *nlpD*(C565T) relative to LSWS-*lacZ* and WSG indicates that the fitness of *nlpD*(C565T) mutants are less than of typical WS types. The LSWS-*lacZ* strain was used to measure the relative fitness of strains and so a competition of LSWS-*lacZ* with LSWS was included. The other three competitions consisted of SBW25 *nlpD* (C565T) against LSWS *lacZ*, IWSG against LSWS *lacZ*, and WT *nlpD*(C565T) against IWSG. The test of IWSG against LSWS *lacZ* was included to confirm the relative fitness of these WS. See Table 7.5.3 for details of statistical tests. Boxplots represent eight replicates.

A) Test of SBW25 (C565T) vs LSWS *lacZ* over 72 hours of competition

Rep	0 h count			72 h count			Malthusian fitness		Relative frequency
	Dil.	C565T	LSWS *lacZ*	Dil.	C565T	LSWS *lacZ*	C565T	LSWS *lacZ*	
1	7.5	34	59	50	3	69	4.885	7.469	0.654
2	7.5	44	41	50	10	277	4.733	8.125	0.582
3	7.5	31	57	50	6	162	4.977	7.664	0.649
4	7.5	35	43	50	1	132	3.064	7.741	0.395
5	25	102	189	75	3	223	1.589	5.281	0.300
6	7.5	43	47	50	4	157	3.839	7.420	0.517
7	7.5	39	42	75	8	263	4.630	8.049	0.575
8	25	82	152	50	6	104	2.906	5.141	0.565
Mean relative frequency (± SE)									0.530 ± 0.0437

B) Test of SBW25 (C565T) vs WS^G over 72 hours of competition

Rep	0 h count			72 h count			Malthusian fitness		Relative frequency
	Dil.	C565T	IWSG	Dil.	C565T	IWSG	C565T	IWSG	
1	25	25	142	75	75	24	3.338	5.616	0.594
2	25	25	115	75	75	13	2.936	5.521	0.531
3	25	25	100	75	75	38	4.148	6.085	0.681
4	25	25	86	75	75	19	3.606	5.287	0.681
5	25	25	99	75	75	23	3.656	4.873	0.750
6	25	25	92	75	75	41	4.307	5.438	0.792
7	25	25	96	75	75	18	3.442	5.097	0.675
8	25	25	55	75	75	5	2.718	5.442	0.499
Mean relative frequency (± SE)									0.651 ± 0.0361

C) Test of LSWS vs LSWS *lacZ* over 72 hours of competition

Rep	0 h count			72 h count			Malthusian fitness		Relative frequency
	Dil.	LSWS	LSWS *lacZ*	Dil.	LSWS	LSWS-*lacZ*	LSWS	LSWS-*lacZ*	
1	7.5	53	55	50	74	84	6.95	6.48	1.07
2	7.5	77	59	50	104	59	6.92	7.33	0.944
3	7.5	47	41	50	96	92	7.33	6.68	1.10
4	7.5	84	55	50	93	69	6.72	6.82	0.984
5	7.5	48	102	75	21	43	5.39	5.85	0.921
6	7.5	78	88	50	73	71	6.55	6.41	1.02
7	7.5	72	44	75	144	72	6.91	6.71	1.03
8	7.5	51	32	50	85	63	7.13	7.30	0.977
Mean relative frequency (± SE)									1.01 ± 0.0216

D) Test of WS^G vs LSWS *lacZ* over 72 hours of competition

Rep	0 h count			72 h count			Malthusian fitness		Relative frequency
	Dil.	IWSG	LSWS *lacZ*	Dil.	IWSG	LSWS *lacZ*	IWSG	LSWS *lacZ*	
1	25	87	118	50	39	144	4.719	5.720	0.825
2	25	45	148	75	45	220	5.116	5.512	0.928
3	25	38	147	50	30	204	5.285	5.849	0.904
4	25	46	124	50	29	126	5.060	5.537	0.914
5	25	88	173	75	28	169	3.971	5.092	0.780
6	25	45	189	75	44	179	5.094	5.061	1.01
7	7.5	31	52	75	19	47	5.720	6.113	0.936
8	25	26	53	75	29	102	5.225	5.770	0.905
Mean relative frequency (± SE)									0.899 ± 0.0248

Table 7.5.2(A-D): Raw counts, dilutions (Dil.), Malthusian fitness and relative fitness for indicated competition experiments. Raw counts indicate the CFU of each indicated genotype plated at 7.5 µL, 25 µL, 50 µL or 75 µL volumes of a 10^{-6} dilution. Relative frequency is converted to three significant values.

Genotype 1	Genotype 2	Mean RF ± CI[a]	Shapiro-Wilk test[b]	*t*-test[c]
SBW25 C565T	LSWS *lacZ*	0.530 ±0.103	0.880	1.29e-05
SBW25 C565T	IWSG	0.651 ±0.085	0.944	2.62e-05
LSWS	LSWS *lacZ*	1.007 ±0.051	0.888	0.771
IWSG	LSWS *lacZ*	0.899 ±0.058	0.930	0.00468

Table 7.5.3: Statistical tests on raw data for each competition assay in Table 7.5.2(A-D). [a]Mean relative frequency (RF) ± 95% confidence interval (CI). [b]Shapiro-Wilk test for normality *p*-values. [c]Student's *t*-test *p*-value for one-sample, two-tailed, *t*-test that the relative frequency differs from 1.

Section 7.5.2: The degree of change to NlpD is associated with cell-division deficiency

The mutants evolved from SBW25 Δ*wss nlpD* C565T A566G (W189W) were observed to vary in the length of cell chains. To assess the affect of different types of mutation on chaining, colonies from each isolate were grown and two colonies of each were suspended in water and visualised under 100x magnification. Resulting images were

used to select 10 cell chains per image, and cells were counted per chain (see Figure 7.5.2).

There is a significant difference in the median number of cells per chain between isolates that have non-synonymous substitution mutations and those that have deletions (M-W-W test, *p*-value <2.2 x 10^{-16}) and truncation mutations (M-W-W test, *p*-value <2.2 x 10^{-16}) (for raw data and details of statistics, see Tables 7.5.4 and 7.5.5). The variation in chaining associated with mutation severity is further evidence that the chaining phenotype results from loss of NlpD functionality. Furthermore, there was no statistical difference in the median number of cells per chain between nonsense mutations that occurred at codon 189 and at alternative positions in *nlpD* (M-W-W test, *p*-value 0.258). This suggests nonsense mutations at residue 189 do not confer a unique effect on cell division compared to other nonsense mutations in *nlpD*.

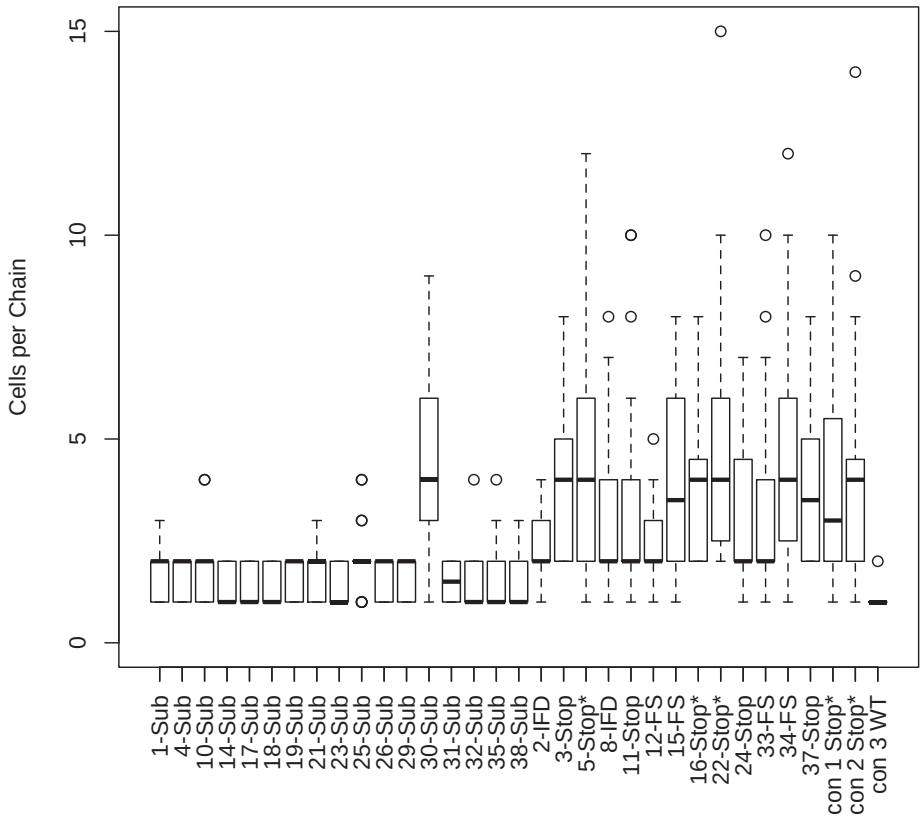

Figure 7.5.2: The separation of cells is associated with the class of mutation within *nlpD.* 'Sub' represents amino acid substitution mutation, 'IFD' represents Inframe deletion, 'Stop*' represents a typical (C565T) nonsense mutation, and 'Stop' represents an alternative (non-C565T) nonsense mutation (all these types evolved from a Δ*wss* C189W background), 'con Stop*' represents a C565T typical nonsense mutation evolved in a Δ*wss* background. Data is presented as standard boxplots representing data for 20 individual cell chains. Substitution mutants have been grouped for visual clarity.

***nlpD* mutant**	**Protein mutation**	**Cells per chain**
1	Substitution	1,1,2,1,1,1,2,2,1,2,2,3,3,3,1,2,2,2,2,1
2	Deletion (IF)	1,2,1,4,2,2,2,2,2,2,4,4,2,4,4,2,2,2,2,2
3	Alt Nonsense	2,6,4,4,7,2,6,2,8,4,6,2,2,4,4,2,4,4,4,4
4	Substitution	2,2,1,2,2,1,2,2,1,2,2,1,2,1,1,1,2,1,2,2
5	Typical Nonsense	1,4,9,6,1,2,1,4,2,2,4,4,2,6,9,12,5,7,2,2
8	Deletion (IF)	2,2,2,7,2,4,4,6,2,1,2,2,2,4,7,4,8,2,2,3
10	Substitution	1,2,1,4,1,4,2,1,2,1,2,2,1,2,2,2,2,2,2,2
11	Alt Nonsense	10,1,4,2,4,2,4,2,4,1,2,2,10,2,2,8,6,2,2,4
12	Deletion (FS)	2,2,2,5,3,2,2,2,2,1,2,3,2,2,2,2,4,4,1,4
14	Substitution	2,2,2,1,1,1,2,1,2,1,2,1,1,1,2,1,2,2,1,1
15	Deletion (FS)	6,3,4,1,8,2,2,1,8,2,2,4,2,8,3,6,4,16,4,2
16	Typical Nonsense	4,2,7,4,4,2,2,2,2,4,4,4,4,5,3,8,3,8,2,8
17	Substitution	1,2,1,1,1,2,2,1,1,2,1,2,1,2,1,2,1,2,1,2
18	Substitution	2,1,2,1,1,1,2,1,1,1,1,1,1,2,1,1,2,2,1,1
19	Substitution	2,2,1,1,2,2,2,2,1,2,1,2,2,1,2,1,2,2,1,2
21	Substitution	1,2,1,3,1,2,1,2,2,3,1,1,1,1,2,2,1,2,2,2
22	Typical Nonsense	4,3,4,2,2,3,6,2,15,2,6,4,6,3,4,2,5,10,6,4
23	Substitution	2,2,2,1,1,2,2,1,2,1,1,1,1,1,1,2,1,2,1,2
24	Alt Nonsense	6,2,2,4,2,2,2,5,4,3,1,7,2,2,4,2,5,7,2,2
25	Substitution	1,2,2,2,3,2,2,1,1,2,2,3,2,4,1,2,2,4,2,2
26	Substitution	2,2,2,1,1,2,1,2,1,1,1,2,2,2,2,2,1,1,2,1
29	Substitution	1,2,2,2,2,2,1,1,2,1,1,1,2,2,1,2,2,1,2,1
30	Substitution	8,4,6,2,6,4,6,6,9,4,4,9,6,2,8,1,2,4,4,2
31	Substitution	2,2,1,1,1,1,1,2,2,2,1,2,2,2,2,1,1,1,2,2,1
32	Substitution	1,2,1,1,2,1,2,1,2,2,1,1,2,4,1,1,1,1,1,2
33	Deletion (FS)	2,4,2,2,1,8,2,2,2,2,4,4,2,2,4,7,2,10,6,4
34	Deletion (FS)	4,2,12,10,1,2,4,5,4,1,4,4,4,2,4,3,6,10,8,6
35	Substitution	1,1,1,2,1,1,1,3,4,1,1,2,2,1,1,1,2,2,1,1,
37	Alt Nonsense	4,3,3,5,2,4,2,3,4,2,2,8,2,4,7,5,5,6,2,2
38	Substitution	1,2,1,1,1,1,1,3,2,2,2,1,1,2,2,1,2,1,2,1
∆wss *nlpD* C565T 1	Typical Nonsense	10,3,2,4,4,8,2,2,8,2,4,2,5,2,3,1,6,8,3,1
∆wss *nlpD* C565T 2	Typical Nonsense	8,2,4,5,14,2,2,1,2,4,6,2,2,4,4,9,4,4,2,4
∆wss	WT	1,1,1,2,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1

Table 7.5.4: Raw data counts of the numbers of cells per chain in types evolved from mutant type *nlpD* C565T G566A (Q189W).

Comparison	M-W-W test, *p*-value
Substitutions vs Deletions	$<2.2 \times 10^{-16}$
Substitutions vs Alternative Nonsense	$<2.2 \times 10^{-16}$
Substitutions vs Typical Nonsense	$<2.2 \times 10^{-16}$
Alternative Nonsense vs Typical Nonsense	0.258
Typical Nonsense vs Typical Nonsense evolved from Δ*wss*	0.579

Table 7.5.5: Statistical tests of the grouped data from Table 7.5.4 using M-W-W tests.

A) C565T (Q189*) (control a) tested against *nlpD* C565T

Rep	0 h count			48 h count			Competitive fitness	
	CFU	Tested	Ref	CFU	Tested	Ref	Selection coefficient	Corrected coefficient
1	37	54211	52103	9	67839	31898	0.0760	-0.0163
2	22	52863	47137	27	71289	28709	0.0845	-0.00789
3	17	53957	46043	21	74710	25290	0.0982	0.00588
4	31	53193	46807	18	76318	23681	0.1108	0.0184
5	15	52367	47633	7	75056	24943	0.1070	-0.0218
6	19	52376	47623	6	68847	31152	0.0741	-0.0547
7	60	50455	49545	9	88044	11953	0.2103	0.128
8	20	53249	46751	16	78524	21476	0.1239	-0.00490
Corrected mean selection coefficient (± SE)								0.000 ± 0.0190

B) C49T (R17*) tested against *nlpD* C565T

Rep	0 h count			48 h count			Competitive fitness	
	CFU	Tested	Ref	CFU	Tested	Ref	Selection coefficient	Corrected coefficient
1	24	62451	37549	22	86498	13502	0.1357	0.0433
2	24	62089	37911	14	90470	9530	0.1768	0.0844
3	40	61586	38414	57	89517	10482	0.1683	0.0759
4	16	64151	35849	5	92895	7103	0.2001	0.107
5	16	62555	37445	8	81811	18189	0.0996	-0.0218
6	18	62544	37456	8	89705	10295	0.1662	-0.054701
7	29	59996	40004	12	89881	10114	0.1790	0.1288
8	33	63796	36203	6	87815	12184	0.1417	-0.00490
Corrected mean selection coefficient (± SE)								0.0525± 0.0232

C) C817T (Q273*) tested against *nlpD* C565T

Rep	0 h count			48 h count			Competitive fitness	
	CFU	Tested	Ref	CFU	Tested	Ref	Selection coefficient	Corrected coefficient
1	8	53517	46483	11	54895	45105	0.00635	-0.0860
2	52	47732	52268	18	45589	54410	-0.00985	-0.1022
3	52	52473	47526	13	73402	26598	0.1048	0.0124
4	41	55164	44836	12	52925	47075	-0.0103	-0.1027
5	8	49700	50300	7	57241	42759	0.0347	-0.0941
6	-	45302	54698	5	55395	44604	0.0463	-0.0825
7	28	47661	52339	4	50061	49939	0.0109	-0.1178
8	32	48230	51770	8	68299	31700	0.0959	-0.0329
Corrected mean selection coefficient (± SE)								-0.0788± 0.0154

D) G483A (W161*) tested against *nlpD* C565T

Rep	0 h count			48 h count			Competitive fitness	
	CFU	Tested	Ref	CFU	Tested	Ref	Selection coefficient	Corrected coefficient
1	8	53517	46483	11	54895	45105	0.00635	-0.0860
2	52	47732	52268	18	45589	54410	-0.00985	-0.1022
3	52	52473	47526	13	73402	26598	0.1048	0.0124
4	41	55164	44836	12	52925	47075	-0.0103	-0.1027
5	8	49700	50300	7	57241	42759	0.0347	-0.0941
6	-	45302	54698	5	55395	44604	0.0463	-0.0825
7	28	47661	52339	4	50061	49939	0.0109	-0.1178
8	32	48230	51770	8	68299	31700	0.0959	-0.0329
Corrected mean selection coefficient (± SE)								-0.0354± 0.0154

E) Deletion 141-144 (A47FS) tested against *nlpD* C565T

Rep	0 h count			48 h count			Competitive fitness	
	CFU	Tested	Ref	CFU	Tested	Ref	Selection coefficient	Corrected coefficient
1	-	54549	45451	28	31932	68066	-0.0913	-0.1837
2	-	55993	44007	38	34244	65754	-0.0868	-0.1792
3	-	53687	46313	11	61714	38286	0.0320	-0.0603
4	-	52154	47846	8	39040	60960	-0.0516	-0.1440
5	20	48248	51752	15	63398	36602	0.0602	-0.0686
6	11	50661	49338	9	75057	24939	0.1045	-0.0243
7	34	46816	53184	14	59526	40474	0.0498	-0.0789
8	3	50661	49339	6	79242	20758	0.1276	-0.00124
Corrected mean selection coefficient (± SE)								-0.0969± 0.0244

F) Deletion 458-566 (deletion 161-189) tested against *nlpD* C565T

Rep	0 h count			48 h count			Competitive fitness	
	CFU	Tested	Ref	CFU	Tested	Ref	Selection coefficient	Corrected coefficient
1	21	49344	50656	20	63941	36050	0.0615	-0.0309
2	21	48190	51809	29	37248	62743	-0.0461	-0.138
3	21	50994	49006	31	34687	65313	-0.0690	-0.161
4	85	47242	52758	25	34304	65683	-0.0553	-0.148
5	12	46581	53418	20	25676	74318	-0.0950	-0.224
6	42	49308	50692	14	26317	73680	-0.102	-0.232
7	6	43563	56436	22	38250	61748	-0.0226	-0.151
8	36	47738	52262	13	56387	43609	0.0356	-0.0932
Corrected mean election coefficient (± SE)								-0.1534± 0.0231

G) G566A (Q189*) (control b) tested against *nlpD* C565T

Rep	0 h count			48 h count			Competitive fitness	
	CFU	Tested	Ref	CFU	Tested	Ref	Selection coefficient	Corrected coefficient
1	-	58729	41271	15	81292	18708	0.118	0.0265
2	-	54655	45345	19	78911	21089	0.121	0.0282
3	-	50387	49613	19	82287	17713	0.162	0.0695
4	-	51164	48836	28	87418	12581	0.202	0.109
5	10	50040	49960	34	90069	9930	0.235	0.105
6	14	49780	50220	31	87808	12192	0.211	0.0824
7	30	50310	49690	17	78145	21854	0.134	0.00553
8	67	50348	49652	47	90361	9639	0.237	0.108
Corrected mean selection coefficient (± SE)								0.0551± 0.0147

Table 7.5.6(A-G): Raw data of the competition assays of mutants evolved from the *nlpD* C565T A566G (Q189W) mutant vs a C565T reference. '0 h CFU' represents total counts of CFU from 50 µL volumes of a 10^{-6} dilution of the initial mixed culture, which was then diluted 1000 times to start the competition assay. '48 h CFU' represents CFU of 50 µL volumes of the competition culture diluted by 10^{-6}. 'Tested' and 'Ref' represent counts of tested and reference cells as determined by levels of fluorescence from a total of 100,000 cells. 'Selection coefficient' is measured by the model $s = [\ln(R(t)/R(0))]/[t]$, where R is the ratio of measured to marked cell counts, and t is the number of generations calculated by ln(final CFU/initial CFU)/ln(2) (as per Dykhuizen 1990). Selection coefficients were corrected relative to the cost of GFP expression measured in competition A (C565T (control a) tested against nlpD C565T marked wth GFP).

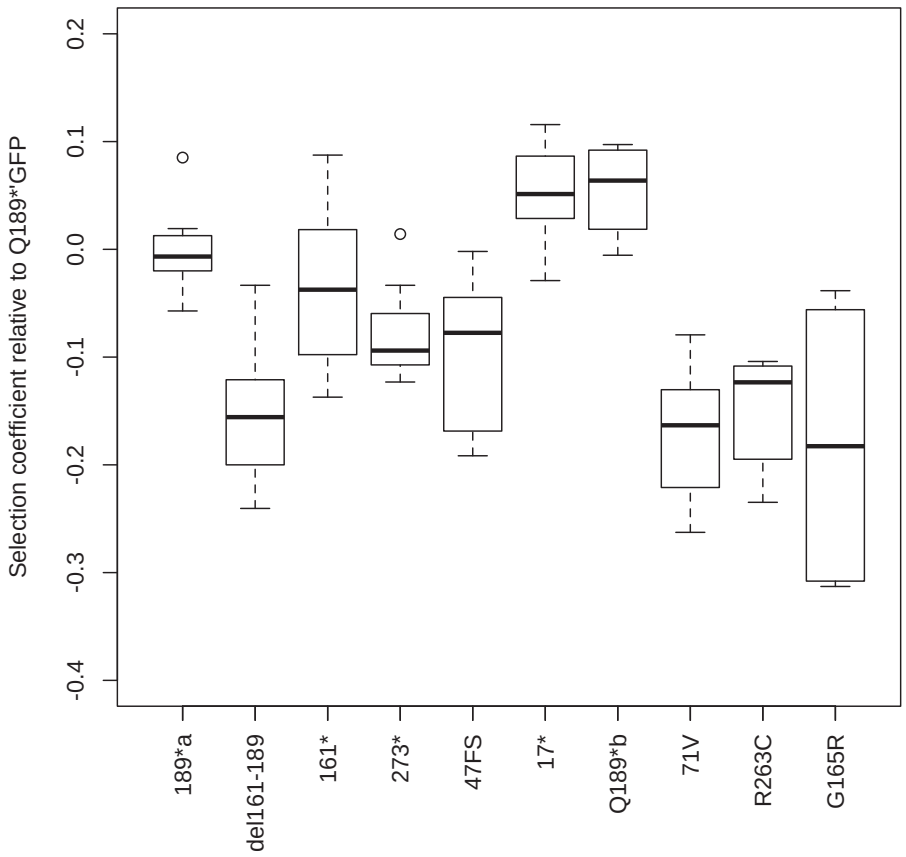

Figure 7.5.3: Boxplots of raw data in Tables 7.5.6 of the relative fitness of alternative mutant types in competition with a C565T mutant. Boxplots represent measures of the selection coefficient from eight replicates.

Comp	Genotype 1	Genotype 2	Shapiro-Wilk test[a]	Mean CSC ± CI[b]	*t*-test[c]
A	Q189* (a)	C565T GFP	0.3412	0.000 ± 0.0344	1
B	2 Δ161-189	C565T GFP	0.5921	-0.1534 ± 0.0562	0.000347
C	11 W161*	C565T GFP	0.6779	-0.0354 ± 0.0664	0.2476
D	24 Q273*	C565T GFP	0.05422	-0.0788 ± 0.0385	0.001872
E	34 A47(6)	C565T GFP	0.4077	-0.0969 ± 0.0599	0.006492
F	37 R17*	C565T GFP	0.9347	0.0525 ± 0.0381	0.01388
G	Q189* (b)	C565T GFP	0.2269	0.0551 ± 0.0332	0.00572

Table 7.5.7: Statistical tests on raw data for each competition assay in Table 7.5.6(A-G). 'Comp' represents competition. [a]Shapiro-Wilk test for normality *p*-values. [b]Mean corrected selection coefficient (CSC) for each competition ± 95% confidence interval (CI). [c]Student's *t*-test *p*-value for one-sample, two-tailed, *t*-test that the relative frequency differs from 1.

Comp	Genotype 1	Genotype 2	Shapiro-Wilk test[a]	Mean CSC ± CI[b]	*t*-test[c]
A	Q189* (a)	C565T GFP	0.9716	-0.0670 ± 0.0338	0.002243
B	2 Δ161-189	C565T GFP	0.6201	-0.214 ± 0.058	5.385×10^{-05}
C	11 W161*	C565T GFP	0.77	-0.0973 ± 0.0619	0.00748
D	24 Q273*	C565T GFP	0.08701	-0.143 ± 0.038	4.931×10^{-05}
E	34 A47(6)	C565T GFP	0.5285	-0.160 ± 0.052	0.0001701
F	37 R17*	C565T GFP	0.952	-0.0191 ± 0.042	0.3187
G	Q189* (b)	C565T GFP	0.6222	0.000 ± 0.0340	1

Table 7.5.8: Identical data to Table 7.5.7, with the selection coefficients corrected by the second C676T mutant. 'Comp' represents competition. [a]Shapiro-Wilk test for normality *p*-values. [b]Mean corrected selection coefficient (CSC) for each competition ± 95% confidence interval (CI). [c]Student's *t*-test *p*-value for one-sample, two-tailed, *t*-test that the relative frequency differs from 1.

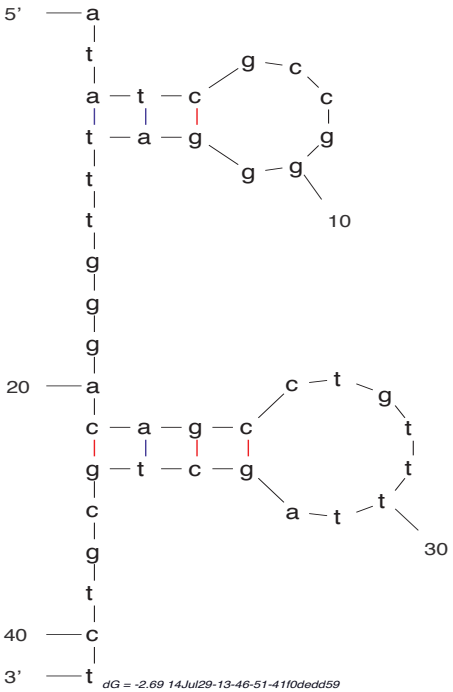

Figure 7.5.4: The secondary structure of the region of the C565T mutation. Depicted is the most stable structure (as predicted in the *mfold* web-program (Zuker 2003)) in the immediate region of the C565T mutation, simulating structures that may form provided negative-supercoiling in a transcribed gene (Burkala et al. 2007). The cytosine of nucleotide 565 is depicted in nucleotide 21 of the diagram, as part of a stem of a secondary structure.

Genotype	Region of mRNA quant.	Raw data	Mean ± SE	Overall mean ± SE
Δ*wss* *nlpD* WT	*nlpD* 5'	1:0.584,0.639, 0.856, 0.724, 2:0.506, 0.472, 0.524, 0.489,	0.701 ± 0.059 0.498 ± 0.011	0.599 ± 0.047
	nlpD 3'	1:1.510, 1.711, 1.860, 1.834 2:1.171, 1.123, 1.155, 0.869,	1.729 ± 0.08 1.080 ± 0.071	1.404 ± 0.132
	rpoS 5'	1:0.516,0.613,0.553, 0.892, 2:0.569, 0.558,0.506,0.554,	0.644 ± 0.085 0.547 ± 0.014	0.595 ± 0.044
Δ*wss* *nlpD* W189*	*nlpD* 5'	1:0.245, 0.224, 0.240,0.240, 2:0.280,0.252,0.276,0.267,	0.237 ± 0.005 0.269 ± 0.006	0.253 ± 0.007
	nlpD 3'	1:4.891, 5.897,5.540,5.816, 2:3.713,3.979,3.662,4.206,	5.536 ± 0.228 3.890 ± 0.126	4.713 ± 0.334
	rpoS 5'	1:1.602,1.580,1.682,1.682, 2:1.081,1.016, 1.002,1.242,	1.637 ± 0.027 1.085 ± 0.055	1.361 ± 0.108
Δ*wss* *nlpD* R17*	*nlpD* 5'	1:0.361, 0.570, 0.483, 0.486 2:0.201,0.188,0.187,0.194,	0.475 ± 0.043 0.193 ± 0.003	0.333 ± 0.057
	nlpD 3'	1:0.212,0.301,0.289,0.275, 2:0.140,0.147,0.132,0.141,	0.269 ± 0.020 0.140 ± 0.003	0.204 ± 0.026
	rpoS 5'	1:0.171,0.171,0.159,0.174, 2:0.074, 0.072, 0.080, 0.083,	0.169 ± 0.003 0.0773 ± 0.003	0.123 ± 0.017
Δ*wss* *nlpD* Q189W	*nlpD* 5'	1:0.375,0.375,0.362,0.391, 2:0.328,0.328,0.308,0.339,	0.376 ± 0.006 0.326 ± 0.006	0.351 ± 0.010
	nlpD 3'	1:0.307,0.265,0.241,0.253, 2:0.272,0.199,0.219,0.209,	0.267 ± 0.014 0.225 ± 0.016	0.246 ± 0.013
	rpoS 5'	1:0.122,0.099,0.112, 0.185, 2:0.089,0.080,0.076,0.113,	0.130 ± 0.019 0.0895 ± 0.008	0.110 ± 0.012

Table 7.5.9: Raw data measures of the levels of mRNA expressed from each region of each strain. Data represents Cq values measured at each region (already made relative to measures of *recA*). Raw data from '1' and '2' indicate the four technical replicates from each biological replicate (combined into the overall mean).

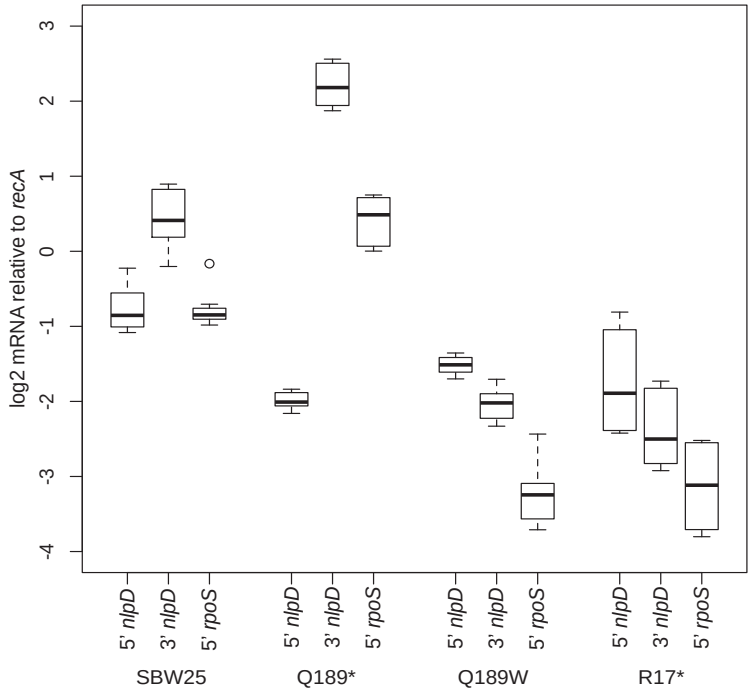

Figure 7.5.5: Boxplots of the raw data of Cq measures from each region of *nlpD* or *rpoS* across all strains, made relative to *recA*. Boxplots represent eight replicates.

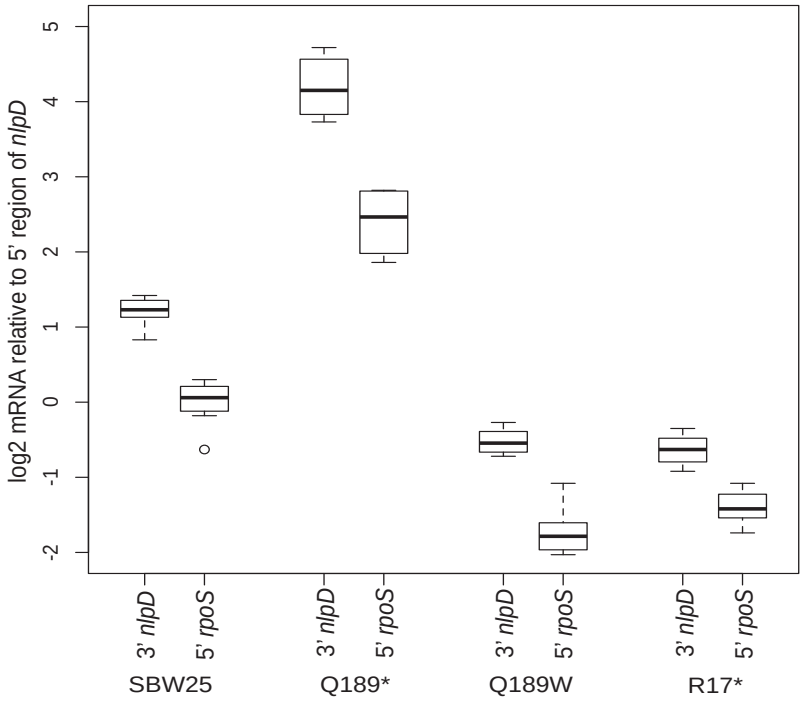

Figure 7.5.6: Boxplots of the relative levels of transcription across *nlpD* and *rpoS*. The relative levels of mRNA are presented relative to the 5' region of *nlpD*. Boxplots represent data from eight replicates per region.

Genotype	Region of mRNA quant.	Raw data	Mean ± SE	Overall mean ± SE
Δ*wss* (*nlpD* WT)	*nlpD* 3'	1:2.585,2.676,2.173,2.532, 2: 2.313,2.378,2.204,1.778	1: 2.491 ± 0.110 2: 2.168 ± 0.135	2.330 ± 0.101
	rpoS 5'	1:0.883,0.959,0.646,1.231, 2: 1.125,1.181,0.966,1.133	1: 0.930 ± 0.121 2: 1.101 ± 0.047	1.016 ± 0.0681
Δ*wss* *nlpD*W189*	*nlpD* 3'	1:19.973,26.355,23.103, 24.251, 2: 13.269,15.78,13.269,15.242	1: 23.421 ± 1.332 2: 14.390 ± 0.656	18.905 ± 1.840
	rpoS 5'	1:6.543,7.062,7.013,7.013, 2: 3.864,4.028,3.63,4.659	1: 6.908 ± 0.122 2: 4.045 ± 0.22	5.477 ± 0.553
Δ*wss* *nlpD*R17*	*nlpD* 3'	1:0.586,0.529,0.599,0.566, 2: 0.697,0.785,0.707,0.727	1: 0.570 ± 0.015 2: 0.729 ± 0.020	0.650 ± 0.032
	rpoS 5'	1:0.473,0.299,0.33,0.358, 2: 0.366,0.382,0.426,0.429	1: 0.365 ± 0.038 2: 0.401 ± 0.016	0.383 ± 0.0202
Δ*wss* *nlpD*Q189W	*nlpD* 3'	1:0.818,0.707,0.664,0.646, 2: 0.829,0.607,0.712,0.616	1: 0.709 ± 0.039 2: 0.691 ± 0.052	0.700 ± 0.0301
	rpoS 5'	1:0.325,0.264,0.31,0.473, 2: 0.272,0.245,0.248,0.332	1: 0.343 ± 0.045 2: 0.274 ± 0.02	0.309 ± 0.0264

Table 7.5.10: The relative Cq from each region downstream of the putative nested promoter, made relative to the Cq measures from the 5' region of *nlpD*. Raw data from '1' and '2' indicate the four technical replicates from each biological replicate (combined into the overall mean).

Genotype 1	Region of mRNA quant.	Shapiro-Wilk test[a]	Mean Cq relative to *nlpD*(5') ± CI[b]	*t*-test[c]
Δ*wss* (*nlpD* WT)	*nlpD* 3'	0.6249	2.330 ± 0.239	3.44×10^{-6}
	rpoS 5'	0.4069	1.016 ± 0.161	0.826
Δ*wss* (W189*)	*nlpD* 3'	0.2453	18.905 ± 4.351	1.79×10^{-5}
	rpoS 5'	0.02791	5.477 ± n.a.	0.00781
Δ*wss* (Q189W)	*nlpD* 3'	0.2291	0.700 ± 0.071	2.19×10^{-5}
	rpoS 5'	0.03602	0.309 ± n.a.	0.00781
Δ*wss* (R17*)	*nlpD* 3'	0.5357	0.650 ± 0.076	1.21×10^{-5}
	rpoS 5'	0.9403	0.383 ± 0.047	1.037×10^{-8}

Table 7.5.11: Statistical tests of whether the relative Cq means across *nlpD* (*nlpD* 3') or *nlpD* to *rpoS* (*rpoS* 5') differ from 1.0. A test of the relative change in expression across the *nlpD* (or *rpoS*) gene was made relative to 1.0 to detect possible changes in transcription along the gene, as may be caused by the nested promoter. [a]Shapiro-Wilk test for normality *p*-values. [b]Mean Cq values measured or each region (from Table 7.5.10) ± 95% confidence interval (CI). [c]Student's *t*-test *p*-value for one-sample, two-tailed, *t*-test that the mean Cq (relative to the Cq of 5' nlpD) differs from 1.