

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.

Evolution of the Genotype-Phenotype Map

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

Master of Science

In

Genetics

at Massey University, Albany, New Zealand

Michael Barnett

2016

Abstract

The relationship between genotype and phenotype, the genotype-phenotype map (GPM), not only describes the genetic and molecular underpinnings of phenotypes, but also determines their variational properties. That is, it determines how genetic variation maps to phenotypic variation. Because of this, the phenotypic consequence of a random mutation may be highly constrained by properties of the GPM. Motivated by the challenge of understanding the GPM and its effect on the course of evolutionary change I here use a bacterial model to investigate how the GPM itself evolved throughout a previously conducted experiment that selected for lineages adept at cycling between the gain and loss of a simple phenotype. The Wrinkly Spreader (WS) morphotype of *Pseudomonas fluorescens* SBW25 is distinguished from the ancestral type by overproduction of an extracellular cellulose polymer that gives it a wrinkled colony morphology and allows it to colonise the liquid surface of a broth-filled vial, a niche unavailable to the ancestral type. The genes underpinning WS have been previously identified allowing the GPM to be characterized. This formed the basis by which I could compare the GPM of those WS derived from the selection experiment and so determine what changes had occurred throughout the extensive cycling of gain and loss of WS. Suppressor analysis of the derived WS types revealed in some cases a striking difference from the ancestral WS state, including one example of a significant re-wiring of regulatory connections and an expansion of the network of genes underpinning WS. In another case a novel association with a gene encoding a fatty acid desaturase was revealed with possible implications for an unusual switching mechanism. In some derived WS the GPM remained apparently unchanged but these WS were also implicated in switching strategies. By repeatedly re-evolving the same phenotype the GPM is required to find new viable configurations and I show in this thesis that the capacity to do so is vast.

Acknowledgments

Many thanks to my supervisor Professor Paul Rainey for giving me a crazy project that made me think and my co-supervisor Dr Philippe Remigi for his guidance. Also the ever-reliable Yunhao for technical assistance when Philippe wasn't around.

To my mum, dad and brother, thank you.

Also to the two cats in my life, Charles and Milly.

Table of Contents

0.1 Abstract	II
0.2 Acknowledgments	III
0.3 Table of Contents	IV
0.4 Figures	VII
0.5 Tables	VII
0.6 Abbreviations	VIII
1: Introduction	1
1.1 Genetic architecture constrains evolution	1
1.2 The wrinkly spreader phenotype	5
1.2.1 Alternative pathways	6
1.3 The evolution of multicellularity	7
1.4 Research objectives	9
2: Methods and Materials	
2.1 Materials	10
2.1.1 Strains list	10
2.1.2 Plasmids	11
2.1.3 Primers	11
2.1.4 Antibiotics, reagents and enzymes	12
2.1.5 Media and culture conditions	12
2.1.6 Electrophoresis materials	13
2.1.7 Photography and microscopy materials	13
2.2 Methods	13
2.2.1 Cellulose assay	13

2.2.2 Transposon mutagenesis	13
2.2.3 Cre-mediated excision of transposon	14
2.2.4 Tri-parental conjugation	14
2.2.5 Production of electro-competent cells	14
2.2.6 Polymerase chain reaction	15
2.2.7 Arbitrary primed PCR	15
2.2.8 Enzymatic purification	16
2.2.9 Strand overlap extension	16
2.2.10 Extraction and purification	17
2.2.11 Cloning and transformation	17
2.2.12 Allelic exchange	18
2.2.13 Sanger sequencing	18
2.2.14 Identifying gene orthologs, synteny and domain architecture	19
2.2.15 Additional life cycle generations	19

3: Results

3.0.1 Selection of candidate lines from the life cycle experiment	20
3.0.2 Cellulose assay	21
3.1 Interpreting the mutations	22
3.1.1 Overview of mutations	23
3.2 Suppressor Analysis	24
3.2.1 Line 17	25
3.2.1.1 The suppressor loci indicate no change from ancestral WS	26
3.2.2 Line 43	27
3.2.2.1 The suppressor loci indicate no change from ancestral WS	28
3.2.2.2 Suppressor analysis of line 43+3	29

3.2.3 Line 54	30
3.2.3.1 A previously unseen association between <i>aws</i> and a predicted fatty acid desaturase	31
3.2.3.2 The role of pflu0184	31
3.2.3.3 <i>fadA</i>	32
3.2.3.4 The role of <i>fadA</i> in line 54	33
3.2.3.5 Fatty acid desaturases	33
3.2.3.6 pflu5420	34
3.2.3.7 pflu1555	35
3.2.3.8 Mutational causes of altered GPM in line 54	35
3.2.4 Line 57	35
3.2.4.1 Line 57 is underpinned by a complex architecture featuring a known negative regulator of <i>wss</i>	36
3.3 Additional life cycle generations	38
4: Discussion	41
4.1 Evolution of the GPM: a redundancy of pathways	41
4.2 Possible switching mechanisms	44
4.3 Possible switching mechanisms: line 43	45
4.4 Evolution of the GPM: scaffolding	45
4.5 Evolution of the GPM: line 57	45
4.6 Concluding discussion	47
Bibliography	50
Appendix	57

Figures

1.1 A potentially beneficial mutation is unrealised due to an associated deleterious effect	2
1.2 Modularity and pleiotropy. Diagram representing two simples GPMs	3
2.1 AP-PCR amplification of transposon-chromosome junction	16
3.1 Cellulose matrix stained with calcofluor and visualized at 60x magnification	21
3.2 Morphological features of line 17	25
3.3 Morphological features of line 43	27
3.4 Morphological features of line 54	30
3.5 Distribution of transposon inserts affecting <i>fadA</i>	32
3.6 Morphological features of line 57	35
3.7 The fate of line 43 throughout three extra generations of the LCE	39
3.8 The fate of line 54 throughout three extra generations of the LCE	39
3.9 The fate of line 57 throughout three extra generations of the LCE	40

Tables

2.1 Designation and genetic properties of the bacterial strains used in this study	10
2.2 Plasmids used in this study	11
2.3 Primers used in this study	11
3.1 Mutations present in each line as revealed by whole genome sequencing	23
3.2 The distribution of transposon insertions suppressing the WS phenotype in line 17	25
3.3 The distribution of transposon insertions suppressing the WS phenotype in line 43	28
3.4 Suppression loci of line 43+3	29
3.5 The distribution of transposon insertions suppressing the WS phenotype in line 54	30
3.6 The distribution of transposon insertions suppressing the WS phenotype in line 57	35

Abbreviations

GPM: genotype-phenotype map

WS: Wrinkly spreader

SM: Smooth morphology

DGC: diguanylate cyclase

PDE: phosphodiesterase

RE: Re-evolution experiment

1.0 Introduction

Understanding the relationship between genotype and phenotype, the genotype-phenotype map (GPM), is the essential problem of genetics. In its simplest form it requires knowledge of the set of genes and the interactions of their molecular products (development) that gives rise to a particular phenotype (Albercht, 1991). This is no small task as even relatively simple phenotypes are the result of significant coordination among numerous genes and elucidating their molecular interactions is rarely straight-forward. Although advances in sequencing technology have allowed for an incredible level of insight into the content of the genome, deciphering its meaning - that is, how this information articulates phenotype, remains the challenge for geneticists.

The GPM concept is also of great interest to evolutionary biologists as it not only describes the molecular underpinnings of phenotypes but also dictates their variational properties (Albercht, 1991; Wagner & Altenberg, 1996; Hansen, 2006). That is, it concerns how genetic variation maps to phenotypic variation and, consequently, influences the likelihood that random mutation can generate an adaptive phenotype. However in this context I prefer the term 'genetic architecture' and generally reserve GPM for instances when a specific phenotype is being considered.

1.1 Genetic architecture constrains evolution

Organisms exist in a state of tension between the need to change and the need to stay the same. They must be robust to potentially deleterious mutations by suppressing the phenotypic consequences of random mutation. They must also remain evolvable in the face of a changing environment by expressing the phenotypic consequences of random mutation (Kirschner & Gerhart, 1998). Given that the majority of mutations incur a loss-of-function (deleterious effect) to the targeted gene, how this is achieved is not obvious (Kimura, 1968).

The most significant property of genetic architecture mediating this tension is connectivity (Wagner & Altenberg, 1996; Hansen, 2006). This is described by two interrelated principles – that of pleiotropy and epistasis¹. Pleiotropy refers to a single gene that influences multiple phenotypic traits while epistasis occurs when the effect of one gene is determined by its genetic background. In other words epistasis denotes an interaction, either direct or indirect, between genes (Phillips, 1998). The extent of connectivity within a given genetic architecture largely determines the likelihood of a potentially

¹ These terms can also refer to the effect of mutations e.g. a pleiotropic mutation affects more than one trait

beneficial mutation having an accompanying deleterious effect, and so constrains potential paths of evolution (Beaumont *et al.*, 2006).

For instance, in terms of epistasis, mutation to a gene with many interactions is more likely to incur a deleterious effect on one of the connected components, even if the isolated effect of this mutation on certain genes would have been beneficial. A simple example of this is illustrated in Figure 1.1. Mutation to a weakly connected gene on the other hand is likely to have a correspondingly smaller effect and this may allow robustness against deleterious mutations (Jeong *et al.*, 2001). How such patterns are expressed on a genome-wide scale is informed by analysis of protein-protein interaction and gene regulatory networks, which suggest interactions approximate a power-law degree distribution (Proulx *et al.*, 2005). In such networks the majority of nodes (genes) have a single connection while a small subset are highly connected hubs interacting with many others. This arrangement confers robustness in the face of random node removal but is susceptible to targeted attack (i.e. disruption of hubs) and contrasts to a homogeneously connected network which is susceptible to random node removal (Jeong *et al.*, 2001).

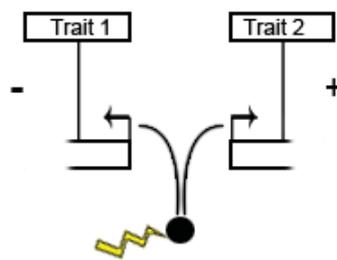


Fig 1.1: A potentially beneficial mutation is unrealized due to an associated deleterious effect. A mutation (lightning bolt) affects a transcription factor (circle). In one gene the altered expression is beneficial (+), but deleterious in the other (-). The mutation also happens to be pleiotropic.

Patterns of pleiotropy can also place significant constraints on evolution (Wagner & Altenberg, 1996; Wagner & Zhang, 2011). Consider an architecture that is universally pleiotropic for instance - where every gene affects every trait in a manner in which alteration to one has a biologically meaningful effect (is visible to selection) on the other. A population of organisms based on this sort of system would be forever required to accommodate contrasting and conflicting selection pressures. Such antagonistic pleiotropy is known (Wright, 1984) but in these cases the number of traits affected is not universal but limited - there is hence some degree of modularity (diminished pleiotropic effects) inherent in the genetic architecture of organisms.

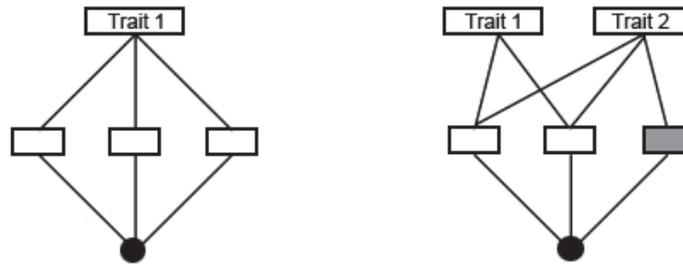


Fig 1.2: Modularity and pleiotropy. Diagram representing two simple GPMs. Circles are regulatory genes and small rectangles structural genes. A) A modular arrangement where each gene influences a single trait. B) In this configuration all genes are pleiotropic with the exception of one structural gene (filled rectangle). Mutation in this gene is hence less constrained in terms of connectivity.

Modularity provides a means of lessening the aforementioned tension between robustness and evolvability (Wagner & Altenberg, 1996). It is achieved, as suggested above, by limiting pleiotropic effects between functionally unrelated complexes. It can therefore facilitate adaptive outcomes by allowing different traits to respond to selection autonomously. In this way it can be seen as providing both robustness (by limiting the phenotypic consequences of mutation to a single trait) and evolvability (by allowing distinct patterns of variation to evolve between traits). Modules also have a number of other features amenable to adaptive change - they can be dissociated through changes in gene regulation via heterochrony – changes in timing, and heterotopy – changes in location of expression (Klingenberg, 1998). This not only aids control of cellular activities but also allows for a more complex development to emerge which can underpin the construction of elaborate phenotypes. A striking example of the, albeit artificial, innovating powers of modularity and heterotopy is the PAX6 gene, a transcription factor that is singularly able to ectopically express the entire developmental module of an eye in *Drosophila* (Halder *et al.*, 1995).

Although emphasizing connectivity in this way is a useful starting point when considering the influence genetic architecture has on evolution it tends to treat genes as equivalent entities and so diminishes the significance of specific functional interactions between genes (the qualitative aspect of epistasis) and properties of the genes themselves (Stern, 2000).

One of the main insights emerging from recent studies of genetic adaptation is the importance of regulatory, versus structural components in accommodating adaptive outcomes (Lozada-Chavez *et al.*, 2006; Wittkopp & Kalay, 2012; Lind *et al.*, 2015). For instance, loss-of-function mutations are indeed most common, but this need not be a loss-of-function to the organism as such a mutation in a negative regulator may activate a pathway which is gain-of-function at the level of phenotype

(McDonald *et al.*, 2009; Lind *et al.*, 2015). The role of *cis*-regulatory regions² in adaptation has also been much touted, particularly in underpinning morphological adaptations of complex organisms, and for good reason (Hoekstra & Coyne, 2007; Wray, 2007; Stern & Orgogozo, 2008; Wittkopp & Kalay, 2012). Because proteins interact with (potentially many) others through the branching network of development, mutation in a protein sequence may affect many or all of these interactions and so is likely to incur deleterious pleiotropic effects. Change to a protein's attendant *cis*-regulatory region however can be limited to the specific temporal or spatial expression of that protein (Hoekstra & Coyne, 2007).

Architectural properties also concern the physical structure and arrangement of genes on a chromosome. As mutation occurs randomly throughout the genome then in a very simple way the physical length of a gene dictates its likelihood of incurring a mutation - a larger gene providing a corresponding larger 'target' for mutation to occur within. Similarly a gene's 'neighbourhood' can expand the spectra of possible mutations to include fusions and promoter capture events generated via an interstitial deletion or similar disruption (Lind *et al.*, 2015).

The final principle to consider is redundancy. This necessarily follows from the means by which genomes 'grow' - through duplication and horizontal gene transfer - meaning components often share homology (Teichmann & Babu, 2004). In some cases this can allow for the masking of a deleterious effect by a functionally equivalent component, such as a duplicated gene or alternative metabolic networks (a means of robustness) (Lenski, *et al.*, 2006). It can also allow for the resurrection of lost phenotypes by co-opting similar components or regulatory pathways (e.g. Taylor *et al.*, 2015).

It should be here noted that robustness and evolvability are not strictly in opposition to each other. Aspects of genetic architecture already described as conferring robustness also allow for the accrual of cryptic variation. This provides a means of traversing genotypic space without cost and can serve as a source of new adaptations (Felix & Wagner, 2008; Hayden *et al.*, 2011).

1.2 The Wrinkly Spreader Phenotype

A useful model for studying the effects of genetic architecture on the way evolution proceeds is the Wrinkly Spreader (WS) system of *Pseudomonas fluorescens* SBW25 (Rainey & Travisano, 1998). Upon inoculation into a static broth-filled microcosm (vial) this aerobic bacterium proliferates and rapidly depletes the dissolved oxygen throughout the liquid phase. Competition for the oxygen replete air-

² *Cis*-regulatory regions are non-coding DNA which regulate the transcription of nearby genes through binding of transcription factors

liquid interface (ALI) is then initiated and this inevitably drives the evolution of the ‘Wrinkly Spreader’ (WS) phenotype, so-called due to its wrinkled appearance when plated. WS are ALI niche-specialists and show a significant negative-frequency dependent fitness advantage over the ancestral ‘Smooth Morphology’ (SM) type. This relationship is reciprocal such that each type can invade the other type when it is rare in the population (Rainey & Travisano, 1998). Colonisation is achieved by the over-production of an extracellular cellulose polymer which acts as a cell-cell glue. By binding cells together, and through attachment to the vial wall, a self-supporting cellulose mat is formed at the ALI.

The correlation between mat-forming WS and their distinctive morphology on plates has allowed for identification of the underlying genes via suppressor analysis (Spiers *et al.*, 2002; Gehrig, 2005; McDonald *et al.*, 2009). This was accomplished with a transposon mutagenesis followed by a screen for those colonies reverted to the SM phenotype. Identifying the precise insertion locus of the transposon within the recipient genome then allowed the characterization of genes underpinning the WS phenotype. This identified a number of common structural and metabolic components as well as a number of distinct activation pathways and has allowed a comprehensive genotype-phenotype map of WS to be constructed.

The primary structural component is *wss* (wrinkly spreader structural locus), a 10 gene operon encoding cellulose, the over-production of which is the proximate cause of the WS phenotype (Spiers *et al.*, 2002). Other structural genes include those responsible for the maintenance of rod-shaped cells (e.g. *mreB*, *rodA*, *pbpA*) as well as others predicted to be involved in cell-wall and membrane biogenesis (pflu1661 - 1667) and modification (pflu0475-pflu0479) (McDonald *et al.*, 2009). These loci effectively act as a scaffold for the production of extracellular cellulose via *wss*. A number of metabolic determinants, particularly those involved in glucose synthesis (cellulose is a glucose polymer) are also required. Such loci are expected to be general to the WS phenotype regardless of activation pathway.

Activation pathways were identified through transposon insertions in diguanylate cyclases (DGC), an enzyme which catalyses the formation of the secondary messenger c-di-GMP, an allosteric activator of cellulose synthase. Initially three separate DGC pathways to WS were discovered: *wsp*, *aws* and *mws* each of which are involved in signal transduction, at least in the ancestral state (McDonald *et al.*, 2009). The first pathway discovered, *wsp*, encodes a two-component system with homology to the *che* chemotaxis pathway in *Escherichia coli* (Bantinaki *et al.*, 2007). Although encoded by seven genes the vast majority of Wsp-activating mutations were found to occur in the methyltransferase *wspF*, a negative regulator of DGC WspR. The other two pathways, *aws* and *mws* are one-component systems in which the input (in DGCs generally a PAS, GAC, HAMP or GAF domain) and output domains are on the same gene. In the case of *aws* (composed of three genes, *awsXRO*), activating mutations fell

almost exclusively in the negative regulator *awsX*. The final pathway, *mws*, is distinct in that mutations occurred within an EAL domain of the DGC-encoding gene. EAL domains encode a phosphodiesterase (PDE) which acts in opposition to DGC activity by degrading c-di-GMP (Tamayo et al. 2005). It was demonstrated that in *mws* the PDE negatively regulates the intragenic DGC (McDonald *et al.*, 2009).

DGCs are characterized by the conserved domain motif GGDEF, which is abundant throughout the SBW25 genome as well as in other bacteria (most commonly residing in one-component systems) (Ryjenkov *et al.*, 2005). This led researchers to investigate the possibility of alternative pathways to the WS phenotype (Lind *et al.*, 2015). Taking advantage of the ability to remove the common pathways (*wsp*, *aws* and *mws*) without incurring any fitness impact, these three route to WS were deleted and the resulting mutant inoculated into 200 microcosms and examined over the course of six days. Alongside this an additional 200 microcosms were inoculated with ancestral SBW25. Of the microcosms inoculated with the mutant type 91 displayed a WS phenotype at the end of the 6 days. In comparison, all 200 of those inoculated with the ancestral strain contained the phenotype after only 3 days. Suppressor analysis of the 91 alternative WS lead to the identification of 13 hidden pathways, all involving DGCs or genes known to relate to c-di-GMP regulation.

1.2.1 Alternative pathways

Knowing why evolution proceeds by a particular genetic path requires knowing why it doesn't proceed by any other. A path might be taken because of a lower fitness of alternative pathways (which would be out-competed), a mutational bias or simply because no alternatives exist. However if none of these explanations can account for the path taken then the reason must lie in that paths propensity to translate random mutation into the phenotype - a property of genetic architecture. The fitness of the WS types arising from mutations in the 13 alternative pathways was found to be similar to that of the common paths (*wsp*, *aws*, *mws*) and no mutational biases were detected. It remains then that properties of the genetic architecture of these alternative pathways are limiting the chances of these WS types being visible to selection from the ancestral state (Lind *et al.*, 2015).

Analysis of the distribution of mutational causes in the alternative pathways, coupled with the knowledge of the importance of negative regulators in the common routes led Lind *et al.* (2015) to propose a hierarchical principle by which evolution tends to proceed (in cases where new phenotypes arise by gene activation). It can be generalized as a 'path of least resistance' - that is, one specified by the fewest and least specific mutations. It involves firstly, the loss-of-function to extragenic negative regulators, followed by loss-of-function to intragenic negative regulators, then gain-of-function: promoter activation, gene fusions and finally intragenic activating mutations. This order then

reiterates but with each mutational class now requiring a double mutation (e.g. mutation in two separate negative regulators), subsequently a triple mutation, and so on.

This 'Re-Evolution' experiment (RE) highlights the significance of a number of architectural elements in determining the spectra of possible mutations able to realise an adaptive outcome. Although at least 16 alternative pathways to WS are now known it's estimated that 99% of mutations that confer the phenotype are via loss-of-function to negative regulators. Moreover, these are far more likely to occur in an extragenic negative regulator (95% of mutations) due to their larger target size – the spectrum of mutations permissible in intragenic negative regulators are limited to those that maintain the unrepressed function of the gene. An example of this can be seen between one of the common pathways *mws* and the most frequented route uncovered in the RE, *pflu0085*. In the case of *mws* the negative regulatory EAL domain (encoding a PDE) is downstream from the cognate GGDEF (encoding a DGC) while in *pflu0085* the negative regulator is upstream of the GGDEF. This arrangement prevents frameshifts and truncations from being permissible in *pflu0085* and so decreases its mutational target size. Beyond this, the study also supports the role of *cis*-regulatory regions (as represented by the class of promoter activation mutants) in being amenable to adaptation and the spatial arrangement of genes in allowing fusions and promoter capture.

In this thesis I take advantage of the wealth of prior knowledge regarding the genetics of the Wrinkly Spreader system (including the characterized GPM, knowledge of activation pathways and mutational causes) to understand a new experiment that I describe below.

1.3 The Evolution of Multicellularity

The WS mat is a product of cooperation - producing the cellulose matrix is costly to individual cells but allows the group of cells to colonise an otherwise unavailable niche (Rainey & Rainey, 2003; Hammerschmidt *et al.*, 2014). This process is driven by the niche-constructing activities of the ancestral SM, which depletes the oxygen throughout the liquid phase and so induces intense competition (and hence strong selection) for colonisation of the ALI (Rainey & Travisano, 1998; Libby & Rainey 2013). In a similar way, once a WS mat is established, it is susceptible to re-invasion by the ancestral SM type that arise within the mat. Gaining the benefit of being localized at the ALI but without paying the cost of cellulose production, these 'cheaters' are bestowed with an energetic advantage over the WS type and their proliferation eventually collapses the mat (Rainey & Rainey, 2003).

How cooperation is maintained in the face of such cheating is a question that concerns the origins of multicellularity (Smith & Szathmary, 1997). One response is to suggest cheater-suppression mechanisms (Michod, 2000). This, after all, is exactly what our own bodies do in response to its cheater, a cancer cell. But such reasoning disregards the level of selection at which cheater-suppression must arise - as a group level adaptation - and so presupposes what it's trying to explain (Rainey, 2007). An alternative hypothesis is to view the cheater as a propagule of the collective (cooperative cells), so long as it has the potential to regenerate the collective phenotype (Rainey, 2007; Rainey & Kerr, 2010). This provides a means of collective-level reproduction - a necessary property for the collective to participate in evolution - and, by uniting the fates of both cell types into a single unit, should allow selection to act upon this new unit.

This was the hypothesis tested in a large scale experiment that aimed to obtain insights into the evolution of collective-level reproduction - a first step in the evolution of multicellular life (Hammerschmidt *et al.*, 2014). Central to the experiment was a life cycle that begins with a SM cell, that produces a WS mat, that gives rise once again to SM cells. More specifically, in the Life Cycle Experiment (LCE), a single WS colony was inoculated into a microcosm and left for six days. The bottlenecked WS colony ensured the WS-mat remained free of within-mat conflict. The mat was required to persist (not collapse) throughout this period as well as produce the alternative SM type at the end of the 6 days. The SM 'propagules' from each microcosm were then pooled and inoculated into another microcosm - the second phase of the life cycle - and examined after three days for the production of WS types. The cycle continued in this manner for 11 generations. Extinction events, which were common, provided opportunity for extant lineages to export their success to other microcosms thereby mimicking selection for lineages adept at maintaining the life cycle.

After 11 generations the derived lineages were shown to have improved in their capacity to transition to the next stage of the life cycle - an increase in fitness at the higher level - while individual cell fitness decreased.

The capacity for SBW25 to maintain the life cycle is impressive. The purpose of this thesis is to examine whether the genotype-phenotype map underpinning the WS phenotype has changed throughout this cycling. This concerns not only the activation pathways, of which we can imagine many have been traversed, but other components of the phenotype. It's also possible we will gain insight into how these lineages have managed to improve in their ability to transition between the two phases of the life cycle. It is an investigation into the evolution of the genotype-phenotype map.

1.4 Research Objectives

1. To identify cellulose-based mats from the derived lineages of the life cycle experiment and select candidate lines for suppressor analysis
2. To perform suppressor analysis on the selected lines using transposon mutagenesis
3. To use the results of the suppressor analysis to characterise the GPM of the derived lines and compare this to the GPM of ancestral WS (those evolved by a single mutational step from ancestral SBW25)
4. To test specific hypotheses resulting from the comparison of the GPMs

Chapter 2: Methods and Material

2.1 Materials

2.1.1 Strains list

All bacterial strains were stored in 33% glycerol saline and stored at -80°C

Strain	Genotype	Reference
<i>Pseudomonas fluorescens</i>		
SBW25 (PBR340)	Wild-type strain isolated from	Rainey & Bailey (1996)
LSWS	WS derived from SBW25 - WS phenotype caused by loss of function to <i>wspF</i>	Spiers et al. (2002)
Line 17 (wt <i>mutS</i>)	Derived line from Life Cycle Experiment. WS isolated at generation 11. <i>mutS</i> restored with wild-type <i>mutS</i>	Hammerschmidt <i>et al.</i> (2014)
Line 43	Derived line from Life Cycle Experiment. WS isolated at generation 11	Hammerschmidt <i>et al.</i> (2014)
Line 54	Derived line from Life Cycle Experiment. WS isolated at generation 11	Hammerschmidt <i>et al.</i> (2014)
Line 57	Derived line from Life Cycle Experiment. WS isolated at generation 11	Hammerschmidt <i>et al.</i> (2014)
17.1 - 17.68	68x WS-suppressed strains derived from line 17 (wt <i>mutS</i>) by transposon mutagenesis with IS-Ω-Km/hah	This study
43.1 - 43.66	66x WS-suppressed strains derived from line 43 by transposon mutagenesis with IS-Ω-Km/hah	This study
54.1 - 54.67	67x WS-suppressed strains derived from line 54 by transposon mutagenesis with IS-Ω-Km/hah	This study
57.1 - 57.67	67x WS-suppressed strains derived from line 57 by transposon mutagenesis with IS-Ω-Km/hah	This study
AWS Δ <i>fadA</i>	WS reconstructed <i>aws</i> mutant from McDonald <i>et al.</i> 2009 with <i>fadA</i> deleted via allelic exchange	This study
Line 54 Δ <i>fadA</i>	Line 54 with <i>fadA</i> (pflu0184) deleted via allelic exchange	This study
Line 57 Δ <i>fadA</i>	Line 57 with <i>fadA</i> (pflu0184) deleted via allelic exchange	This study
Line54 Δpflu0183	Line 54 with pflu0183 deleted via allelic exchange	This study
Line54 Δpflu0185	Line 54 with pflu0185 deleted via allelic exchange	This study
Line43+3	Line 43 taken through an additional 3 life cycle generations	This study
<i>Escherichia coli</i>		
DH5α (λpir)	<i>supE</i> , Δ <i>lacU169</i> (φ 80 <i>lac</i> ZΔM15), <i>hsdR</i> , <i>recA</i> , <i>endA</i> , <i>gyrA</i> , <i>thi</i> , <i>relA</i> (oriR6K replication)	Invitrogen

Table 2.1 Designation and genetic properties of the bacterial strains used in this study

2.1.2 Plasmids

Name	Characteristics	Reference
<i>Transposon</i>		
IS-Ω-Km/hah	<i>ColE1</i> , <i>nptII</i> promoter, <i>LoxP</i> , Km ^R	Giddens <i>et al.</i> 2007
<i>Plasmids</i>		
pGEM-T	High copy number cloning vector; <i>lacZ</i> with MCS, Amp ^R	Promega pGEM-T Easy Vector System
pUIC3	Used for allelic exchange; Tc ^R , <i>mob</i> , <i>oriR6K</i> , <i>bla</i> , Δpromoter- <i>lacZY</i>	Rainey 1999
pRK2013	Helper plasmid for tri-parental mating; Km ^R , <i>tra</i> , <i>mob</i>	Figurski and Helinski 1979
pCre	Used to excise IS-Ω-Km/hah; A derivative of pUT, carrying the <i>cre</i> gene from pRH133, Cm ^R	Manoil 2002

Table 2.2 Plasmids used in this study

2.1.3 Primers

Name	Sequence 5' to 3'	Target
<i>Primers used for transposon mutagenesis</i>		
TnphoAll	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 CEKG2 A:B:C
<i>Universal primers used for verifying cloned inserts</i>		
SP6	ATTTAGGTGACACTATAG	PGEM-T
T7	AATACGACTCACTATAG	PGEM-T
<i>Primers used for deletion constructs</i>		
185_1F	ACTAGTCTTCACCGACAGCTTGCC	pflu0184
185_2R	TGAGCAGATTCAGCATATGGTCAGTAATCCTTACC	3' of pflu0185
185_3F	CCATATGCTGAATCTGCTCAAGACTTTATTTTGC	5' of pflu0185
185_4R	ACTAGTGGATTGCTGGATCTTGACGT	pflu0186
fadA_1F	ACTAGTAGGCTGTGTTCCGGCTAGT	pflu0183
fadA_2R	CCATACAGCCATATGACAACAAGGCGGGAG	3' of pflu0184
fadA_3F	GTTGTCATATGGCTGTATGGCCCTCAGATAA	5' of pflu0184
fadA_4R	ACTAGTAAATTCGACGAATGGCGAC	Pflu0185
183del_1f	ACTAGTACTTTGCCGGTATTCTCTCGAG	Pflu0182
183del_2r	AGTTGTGCCGTGGTCGACTGATCAAATGTGG	3' of pflu0183
183del_3f	CAGTCGACCACGGCACAACCTCTCTGGG	5' of plu0183
183del_4r	ACTAGTCTGTCGGTGAAGAAATGGGAG	Pflu0184

<i>Primers used to verify deletions</i>		
185conF	ATTGGAGCTGGTCTTGAGCC	pflu0184
185conR	GAACACGATGGTCGAGGTG	pflu0186
fadAconF	TTTTCGGTGATGTCTTCGGC	pflu0183
fadAconR	GTGGATAAATCTCTCCGCTTTTGC	pflu0185
183conF	AAGCCCTACGACCTGATCCT	pflu0182
183conR	AAGAACTGCCCGATGACTG	Pflu0184

Table 2.3: Primers used in this study

2.1.4 Antibiotics, reagents and enzymes

Antibiotics (filter-sterilized and dissolved in ddH₂O unless otherwise stated) were purchased from Melford Laboratories and used at the following concentrations: kanamycin (Km) 100 µg ml⁻¹, ampicillin (Amp) 100 µg ml⁻¹, tetracycline (Tet) 12.5 µg ml⁻¹ dissolved in 1:1 ethanol:water, nitrofurantoin (NF) 100 µg ml⁻¹ dissolved in dimethyl sulfoxide (DMSO) and cycloserine (900 µg ml⁻¹). X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) 60 µg ml⁻¹ dissolved in DMSO was used to indicate β-galactosidase activity. A single restriction enzyme, *SpeI* (selected for its low star activity and heat-inactivation ability), was purchased from Thermo Scientific and digestions were carried out at 37°C as per manufacturer's instructions. Standard *Taq* polymerase (Invitrogen), Phusion® High-Fidelity DNA polymerase (NEB), T4 DNA ligase (NEB), Exo I (NEB) and Antarctic Phosphatase (NEB) were used as described in section 2.2.

2.1.5 Media and culture conditions

Unless otherwise specified, all *P. fluorescens* strains were cultured in King's B (KB) medium (In 1L ddH₂O: 10g glycerol, 1.5 g K₂PO₄·3H₂O, 1.5 g MgSO₄·7 H₂O, and either 20g tryptone or proteose peptone for solid or liquid media, respectively). All *E. coli* were cultured in Lyosgeny Broth (LB) medium (In 1L ddH₂O: 10g tryptone, 5g yeast extract and 10g NaCl). Agar was added at 15g L⁻¹ for solid media. Super Optimal Broth (SOB) and Super Optimal Broth with Catabolite repression (SOC) were used during transformations and for making electro-competent cells to increase growth efficiency (for SOB: 20g tryptone, 5g yeast extract, 2ml 0.5M NaCl, 2.5ml 1M KCL, 10ml MgCl₂, 10ml MgSO₄. Mixed in ddH₂O to 1L; for SOC: 20ml 1M glucose added). All overnight cultures were grown from colonies inoculated in glass vials containing 6 ml of media and shaken at 160 rpm for approximately 16 hours. *P. fluorescens* strains were grown at 28°C and *E. coli* strains at 37°C. *P. fluorescens* mat forming ability was assessed visually over 4 days of growth in a static vial (microcosm) inoculated with a single colony and containing 6ml of media. All bacteria were stored indefinitely at -80°C in 33% glycerol saline (70% (v/v) glycerol, 8.5 g NaCl L⁻¹).

2.1.6 Electrophoresis materials

Electrophoresis gels were made with 0.8-1% (^W/_V) ultrapure agarose (Invitrogen), and TBE buffer (90mM Tris-HCL pH 8.0, 0.55% (^W/_V) boric acid, 2mM EDTA) or, for extraction purposes, TAE buffer (40mM Tris-HCL pH 8.0, 20mM acetic acid, and 1mM EDTA) as borate may inhibit downstream enzymatic applications. Addition of 1x SYBR[®] Safe DNA gel stain enabled UV visualisation of DNA. Gels were run in TBE buffer or TAE buffer as necessary. DNA was visualised using a UV transilluminator or during gel excision, a Safe Imager™ 2.0 Blue-Light Transilluminator (Invitrogen).

2.1.7 Photography and microscopy materials

Microcosms were photographed with a Canon Powershot A640 and this same camera used in conjunction with a Zeiss Stemi 2000-C dissection microscope to document colony-level morphology. Cellulose was detected by growing cells in the presence of Calcofluor (Fluorescent Brightener 28, Sigma-Aldrich) and then visualized under UV light with an Olympus BX61 upright fluorescence microscope and F-view II monochrome camera.

2.2 Methods

2.2.1 Cellulose assay

Determining the presence of cellulose involved first growing *P. fluorescens* on KB containing 35µg ml⁻¹ of Calcofluor (fluorescent brightener 28). Cells were grown for 48 hrs and then a single colony picked and placed dry on a microscope slide, fixed with a coverslip, and visualized under UV light with an Olympus BX61 upright fluorescence microscope at 60x magnification. The cellulose signature was unambiguous and appeared as a network of fibres distinct from the background fluorescence produced by the cells themselves.

2.2.2 Transposon Mutagenesis

This involved a tri-parental conjugation (see section 2.2.4) between the recipient and an *E. coli* donor containing the IS-Ω-Km/hah transposon (Giddens *et al.*, 2007) as well as *E. coli* containing the helper plasmid pRK2013 (which encodes the transfer genes). The resulting inoculum was plated on KB with Kan to select for transconjugants and NF to counter-select *E. coli*. After two days growth the plates were inspected for transconjugant colonies exhibiting a loss of the WS morphology and so having the appearance of an ancestral SM type. These colonies were grown overnight with kanamycin and stored at -80°C. The site of insertion was determined by amplifying the transposon-chromosome junction through AP-PCR (section 2.2.7) and the resulting product purified and sent for sequencing. Sequences

were then mapped to the precise location on the genome using BLAST (NCBI) and this position recorded in Artemis 16.0.0.

2.2.3 Cre-mediated excision of transposon

Transposon excision was achieved through introduction of a Cre recombinase-containing plasmid via bi-parental conjugation. The Cre-recombinase acts on the direct repeat LoxP sites at either end of IS- Ω -Km/hah and deletes the intervening sequence. This leaves a 189bp scar and leads to the loss of kanamycin resistance. Bi-parental conjugation involved 1 ml of overnight culture of the recipient heat-shocked for 20 minutes at 45°C before pelleting and resuspension in 200 μ L while 400 μ L of donor culture was pelleted and resuspended in 200 μ L. Thereafter the re-suspensions were mixed, pelleted, and resuspended in 30 μ L LB which was spread onto the centre of a pre-warmed KB plate. This was incubated for 24 hours and 50 μ L plated on KB+NF at 10^{-5} and 10^{-6} . Excision was confirmed by demonstrating kanamycin sensitivity.

2.2.4 Tri-parental conjugation

For conjugations where the donor was *E. coli* DH5 α (which doesn't contain the *tra* and *mob* genes necessary for conjugation) an additional strain containing the helper plasmid pRK2013 was required. Overnight cultures of the donor and helper *E. coli* were grown in LB with appropriate antibiotics and recipient *P. fluorescens* grown in KB. A 2ml aliquot of recipient culture was heat-shocked at 45°C for 20 minutes then pelleted and resuspended in 500 μ L LB while 1ml each of the donor and helper cultures were pelleted and resuspended together in 1ml. The donor and helper mix had 1 ml of the resuspended recipient cells added and this was pelleted once again and resuspended in 60 μ L. If the conjugation was to introduce pUIC3 then at this stage the entirety of the concentrated resuspension was spread onto a pre-warmed KB plate. If introducing IS- Ω -Km/hah during the transposon mutagenesis then four separate conjugations consisting of 15 μ L of inoculum were spread to 1cm². The transposon-introducing conjugation was left for five hours and the inoculum resuspended in 150 μ L LB and 50 μ L of a 10-fold dilution plated on KB containing Kan and NF. The pUIC3 conjugation was left overnight, resuspended in 2ml of ddH₂O and 100 μ L of this plated on KB containing Tet and NF.

2.2.5 Production of electro-competent cells

E. coli DH5 α (λ pir) was used for the transformation of ligated plasmids via electroporation. Production of electro-competent *E. coli* consisted of the following. A 10ml overnight starter culture grown in SOB was used to inoculate 1L of SOC and grown until an OD₆₀₀ ~0.35, at which time 50ml aliquots were dispensed into pre-chilled falcon tubes and stored on ice. The cells were then pelleted (4000 x g, 10 minutes, 4°C) and resuspended in 50ml chilled ddH₂O. This process was repeated twice more but with

the final pellet resuspended in a chilled 10% glycerol solution which was itself pelleted once again. The resulting pellet was gently resuspended in 1ml chilled 10% glycerol solution and 50 μ L aliquots dispensed into eppendorf tubes on ice and transferred to storage at -80°C.

2.2.6 Polymerase Chain Reaction (PCR)

All standard reactions for electrophoresis or general sequencing purposes contained 0.125 μ L *Taq* polymerase, 2.5 μ L 10x PCR buffer, 0.75 μ L 25mM MgCl₂, 0.5 10mM dNTP mix and 0.625 μ L of each 10 nmol ml⁻¹ primer, topped up to 25 μ L with DNase/RNase-free water. Cells taken directly from colonies were used as a template for the reaction. Cycling conditions consisted of a 5 min initial denaturation period at 94°C, followed by 35 cycles of 30 sec denaturation at 94°C; 30 sec annealing at primer-specific temperature (calculated using the online NEB *t_m* calculator); and a 72°C elongation period for 60 sec per kilobase of target DNA amplified. This was followed by a final elongation period of 5 mins at 72°C.

2.2.7 Arbitrary Primed PCR (AP-PCR)

AP-PCR (Manoil, 2000) was used to amplify the transposon-chromosome junction of IS- Ω -Km/hah insertions. This was achieved using a two-step PCR. The first reaction involved a primer specific to IS- Ω -Km/hah and a semi-degenerate primer mixture designed to hybridise with many sites in the genome. The template for this first reaction was prepared by resuspending a colony scraping in 50 μ L ddH₂O. A 3 μ L aliquot of this was combined with 0.5 μ L *Taq* polymerase, 2.5 μ L 10x PCR buffer, 0.80 25mM MgCl₂, 1 μ L 10mM dNTP mix, 2 μ L of each primer and topped up to 20 μ L with ddH₂O. Cycling conditions were as follows. An initial denaturation period of 94°C for 10 min, then 5 cycles consisting of denaturation at 94°C for 30 s; an annealing period at 42°C for 30 s and then decreased by 1°C for each subsequent cycle; and elongation at 72°C for 3 min. This was followed by 25 cycles of 94°C for 30 sec; 65°C for 30 s; 72°C for 3 min and a final 5 min elongation at 72°C. The product of this first PCR reaction was then diluted with 80 μ L and 2 μ L of this was taken as template for the second PCR. This second PCR used a primer specific to IS- Ω -Km/hah and a primer specific to the 5' tail of the semi-degenerate primer mix used in the first reaction. Quantities of other reagents were the same as the first PCR but instead topped up to 25 μ L with ddH₂O. Cycling consisted of a 94°C initial denaturation for 5 min, then 30 cycles of 94°C for 30 s; 65°C for 30 s and 72°C for 3 min. The samples were then stored at 4°C prior to enzymatic purification.

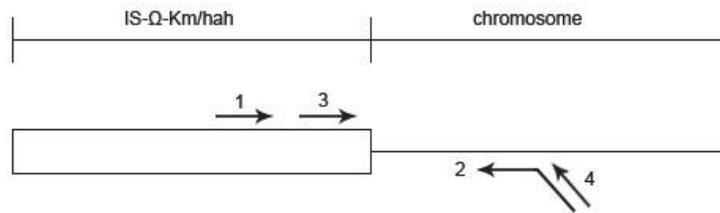


Figure 2.1: AP-PCR amplification of transposon-chromosome junction: amplification of the transposon-chromosome junction involved a two-step PCR. The first reaction generated an amplicon between transposon-specific primer 1 (*TnphoA-II*) and primer 2 (*CEKG2 A:B:C*) which is a mix of semi-degenerate primers designed to insert throughout the genome and each containing an identical 5' tail. The product of this first reaction is used to seed a second PCR with transposon-specific primer 3 (*hah-1*) and primer 4 (*CEKG4*) designed to hybridize with the 5' tail of primer 2. Figure adapted from Manoil (2000).

2.2.8 Enzymatic purification

Enzymatic purification was performed on all PCR products intended for sequencing. This involved the addition of 0.2µL Antarctic Phosphatase and 0.1µL Exo I to each 20-25µL PCR product which was then incubated for 30 min at 37°C and then 15 min at 85°C to inactivate the enzymes.

2.2.9 Strand Overlap Extension (SOE-PCR)

Strand Overlap Extension was used to generate the site-directed deletion constructs. In this process, two partially overlapping PCR products are separately amplified and then annealed together, producing a single amplicon. Firstly, primers are designed to amplify a ~800bp region either side of the gene to be deleted with those primers directly flanking the gene sharing a complementary 20bp region and the outermost primers having a restriction site (*SpeI*) added to the 5' end. This complementary region anneals during a second PCR effectively 'stitching' the two separate amplicons together which is then amplified as one. The restriction sites allow subsequent ligation into plasmids during cloning.

Reactions employed Phusion® High-Fidelity DNA polymerase (NEB) in order to minimize copying errors. The first two reactions (which amplified the flanking regions of the gene to be deleted) consisted of 10µL 5x GC buffer (Phusion-specific buffer used for GC-rich templates), 1µL 10mM dNTP, 2.5µL of each primer, 1.5µL of DMSO, 0.5 µL of Phusion® High-Fidelity DNA polymerase, 1µL of gDNA template and topped up to 50µL with ddH₂O. Cycling conditions involved initial denaturation at 98°C for 30 s then 35 cycles of 98°C for 10 s, annealing for 30 s at primer-specific temperature, elongation at 72°C for 30 s and, following the last cycle, a final elongation for 10 min. The resulting PCR products were then verified using gel electrophoresis. The product of both these reactions formed the template of the SOE-PCR in which strand overlap at the complementary regions would join the two products

together in a single amplicon. This reaction involved 1 μ L of each product (1:1 DNA concentration) and used the same quantities of reactants as above. Cycling conditions differed in that the first five cycles contained only the template and used an annealing temperature specific to the overlapping region. After this the outermost primers used in the first reactions were added and the annealing temperature altered accordingly before an additional 30 cycles.

The final PCR product was run on a 0.08% (^{w/v}) agarose gel with TAE buffer and the individual DNA bands were illuminated under a Safe-imagerTM 2.0 Blue-Light Transilluminator and excised using a sterile razor blade. The DNA was then extracted from the extirpated gel and stored at -20° to await cloning.

2.2.10 Extraction and purification

Plasmid DNA was extracted from overnight cultures using the E.Z.N.A.[®] Plasmid Mini Kit 1 (Omega Bio-Tek), and stored in TE elution buffer at -20°C. Plasmid DNA and SOE-PCR products used in cloning were isolated via gel extraction using a QIAEX II Gel Extraction Kit (Qiagen). The extracted DNA was purified with SureClean Plus (Bioline) prior to ligation.

2.2.11 Cloning and transformation

Before ligation into an intermediate vector, pGEM-T, the purified SOE-PCR product underwent a final PCR in order to add 3' adenine overhangs and allow TA cloning. This reaction consisted of 8 μ L of gel-extracted SOE-PCR product, 0.5 μ L 10mM dATP, 1 μ L 10x PCR buffer and 0.2 μ L *Taq* polymerase incubated for 30 min at 72°C. Ligation was then performed with T4 DNA ligase according to the manufacturer's instructions (pGEM-T Easy Vector Systems, Promega). Transformation was achieved via electroporation into *E. coli* DH5 α -*λpir* electro-competent cells. Aliquots of 50 μ L *E. coli* competent cells were thawed on ice and 4 μ L of ligation product added and mixed gently. The mixture was then transferred into pre-chilled electroporation cuvettes and shocked with 1.8kV after which the cells were quickly transferred into 1ml of pre-warmed (37°C) SOB medium and incubated for 1 hour at 37°C. This was then plated on LB containing X-gal, amp and IPTG, with the successful transformants showing ampicillin resistance and the successfully ligated transformants being unable to cleave X-gal (the insert disrupts *lacZ*) and so appearing white. Inserts were confirmed with a PCR screen and sent for sequencing. Colonies with plasmids containing the insert were grown overnight in LB + amp and the plasmid extracted and digested with *SpeI* restriction enzyme according to manufacturer's instructions. The digested insert was then gel extracted in preparation for ligation into the suicide vector pUIC3.

Ligation between the extracted insert and a linearized pUIC3 (digested with *SpeI* and dephosphorylated with Antarctic Phosphatase) plasmid was performed. A 10 μ L reaction consisted of

1 μ L T4 ligase, 1 μ L 10x T4 ligase buffer, 50ng linearized pUIC3 and 150ng of insert, topped up with ddH₂O and was incubated overnight at 16°C. The resulting ligation mix was purified using SureClean Plus to remove excess salts and so improve the efficiency of electroporation. The final step of cloning the deletion constructs required transformation of pUIC3 into electro-competent *E. coli* DH5 α - λ pir and this was performed via electroporation as above but with transformants screened for tetracycline resistance and the presence of an insert checked via PCR screen.

2.2.12 Allelic exchange

Once cloned into pUIC3 the deletion constructs were then introduced into the recipient genome via a two-step allelic exchange. The first step requires homologous recombination between the deletion construct and the wildtype allele which results in the integration of the entire pUIC3 plasmid into a single strand of the genomic DNA. This was achieved by a tri-parental conjugation between the recipient *P. fluorescens*, donor *E. coli* with pUIC3 and helper *E. coli* containing pRK2013. Successful transconjugants are tetracycline resistant and since pUIC3 is unable to replicate in *P. fluorescens*, must be partially integrated into the genome in the manner described above. The next step requires a second rarer homologous recombination between the now-integrated deletion construct and the remaining wild-type allele. In order to enrich for cells having undergone this event 20 μ L of overnight transconjugant culture was inoculated into 400ml LB without antibiotics and grown for 24 hours at 28°C. This period allowed for loss of the chromosomally-integrated pUIC3 leaving cells with either the wildtype or the desired mutant genotype. Next 400 μ L of this culture was added to 20ml LB and incubated for 30 min at which time 12.5 μ g ml⁻¹ was added and the culture incubated for a further 2 hours. This temporarily selected for the growth of bacteria that maintained the pUIC3 vector and so the associated tet_R gene (i.e. the original transconjugants). After this period 900 μ g ml⁻¹ of the bacteriostatic D-cycloserine was added and incubation continued for 5 hours. During this step addition of the cycloserine kills only growing bacteria due to it disrupting the formation of peptidoglycans. This in turn selects for bacteria that aren't growing - and so enriches for those that have lost the pUIC3 vector via the rare second homologous recombination event. A 1ml aliquot of this culture was taken then pelleted and resuspended in 500 μ L ddH₂O. The resuspension was plated on LB containing X-gal and grown for two days at 28°C. A number of white cells (those without pUIC3 and so either wildtype or deletion mutants) were then picked and the deletion detected via PCR. This was then confirmed by sequencing.

2.2.13 Sanger Sequencing

Sanger sequencing was conducted on purified PCR product and isolated plasmids by Macrogen Inc. (Seoul) and the resulting sequence aligned to the SBW25 genome using Geneious 6.1.7.

2.2.14 Identifying gene orthologs, synteny and domain architecture

The *Pseudomonas* Genome Database (Windsor, 2015) was used extensively to identify gene orthologs and synteny. NCBI Conserved Domain Database (Marchler-Bauer *et al.*, 2015) was used to identify conserved domains.

2.2.15 Additional life cycle generations

The additional three life cycle generations were conducted according to the 'cheat-embracing' regime of Hammerschmidt *et al.*, (2014). Firstly, eight microcosms were inoculated with a single WS colony of each line. These were grown over six days during which time each microcosm had to maintain a mat at the ALI. This constituted phase 1. At the end of the six days the microcosms were vortex mixed and 50µL of culture plated at 8×10^{-6} . After two days growth the plates were then examined for SM colonies, each of which was inoculated into a 96-well plate containing 200µL KB and grown for 24 hours. Following this each well was visually assessed for growth in the liquid phase (to confirm their status as SM types). The confirmed SM from each microcosm were then pooled and 6µL from this inoculated into fresh microcosms. Three days of growth followed (phase 2) and the microcosms vortex mixed and 50µL plated at 3.35×10^{-5} . After two days of growth the plates were inspected for WS types of which a single representative of the most dominant type was taken and used to initiate phase 1. Microcosms failing to produce the required cell type or in which the mat collapsed were randomly replaced by a viable line. Representative WS from each generation were selected from the plate which contained the highest frequency of WS.

3.0 Results

3.0.1 Selection of candidate lines from the Life Cycle Experiment

The life cycle experiment (LCE) initially propagated 15 replicate populations consisting of eight individual microcosms (lines) each. Any line that became extinct was randomly replaced by a viable line from within the same population, while an extinction event of all eight microcosms meant replacement by a line from a random extant population (Hammerschmidt *et al.*, 2014).

After 11 generations of the LCE, 14 populations were present (one population failed to produce any WS) and from each of these a single representative WS colony was selected for further analysis. It was these representative WS from which I selected the candidates to take forward for suppressor analysis in order to identify the genes underpinning the WS phenotype.

The criteria for inclusion as a candidate line included, firstly, the ability to form a cellulose-based mat at the ALI. Cellulose production was determined by calcofluor staining and fluorescent microscopy. Out of the 14 derived lines of the LCE only one (line 56) displayed a negative result for cellulose and so was disregarded. Additionally, lines that showed an evolutionary relationship at the level of replicates (i.e. where one replicate population was used to found a new replicate in the case of extinction of an entire replicate) throughout the LCE were also disregarded: each of the lines analysed here is hence independent.

Based on this criteria, four lines were selected: 17, 43, 54 and 57.

Line 17 is of special significance in that it was the only one for which a specific switching mechanism has been discovered (Hammerschmidt *et al.*, 2014). This switch relies on mutated *mutS*, a mismatch repair protein that corrects replicative errors. In line 17 a dysfunctional *mutS* results in an elevated mutation rate and, importantly, a shift in the spectrum of mutations. Specifically, there was found to be an increase in frameshift mutations - the vast majority of which were associated with homopolymeric tracts of guanine residues. The *wspR* response regulator contains a tract of 7 guanine residues and in line 17 this was found to expand and contract between WS and SM transitions, alternately shifting the alignment of the reading frame and so switching WspR on and off. The present study uses a derivative of Line 17 in which the wild-type *mutS* has been restored in order to avoid false positives during the suppressor analysis.

3.0.2 Cellulose assay

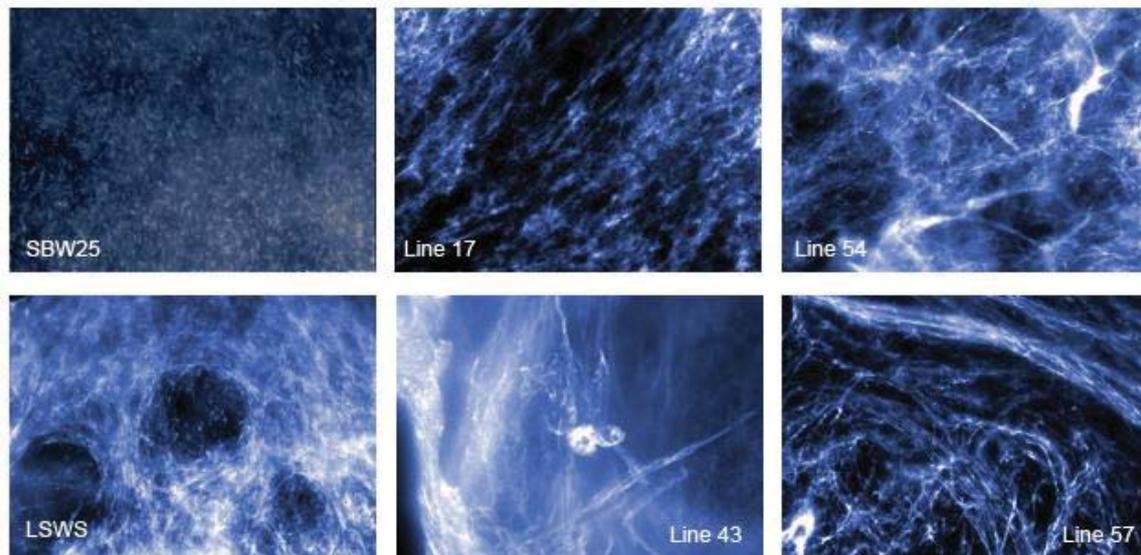


Figure 3.1: Cellulose matrix stained with calcofluor and visualized at 60x magnification under UV light. LSWS is a WS caused by a mutation in *wspF* - a single mutational event from the ancestral SM genotype. SBW25 is the ancestral SM type and required >10ms exposure time for visualisation; WS types used <1ms exposure. The assay was simply to determine the presence of cellulose and was not further qualified. The cellulose signature was unambiguous and appeared as a network of fibres distinct from the background fluorescence produced by the cells themselves.

3.1 Interpreting the mutations

Previous whole-genome sequencing (WGS) of each of the derived lines provided a basis for which to understand any differences in the GPM between the ancestral and derived WS phenotypes. Interpreting the mutational history of the lines is of obvious importance to understanding how these differences might have come about. Although the order of the mutations has not been established in this experiment, findings from previous work make it possible to infer this and the likely effect of a number of the mutations. In some cases, due to high parallelism between studies, it is possible to correlate the exact mutation with a transition toward either of the two phenotypic states.

The following table presents the mutations as identified by WGS.

Locus	Gene symbol	Annotation	Gene coordinate	Mutation	Amino acid change	Effect of mutation
Line 17						
Pflu 1164	<i>mutS</i>	DNA mismatch repair protein MutS	1489	T>G	T497P	substitution
Pflu 1219	<i>wspA</i>	putative methyl accepting chemotaxis protein	757	G>T	D253Y	substitution
Pflu 1220	<i>wspB</i>	putative chemotaxis-like protein	109-117	delGCCGAA GTG		deletion in-frame
Pflu 1224	<i>wspF</i>	chemotaxis specific methyl esterase	884	T>G	I295S	substitution
Pflu 4306		putative GGDEF/GAF domain sensory box protein	403	G>C	A135P	substitution
Pflu 5210	<i>awsR</i>	DGC	79	T>G	T27P	substitution
Pflu 5210	<i>awsR</i>	DGC	620	G>A	P207M/L	substitution
Pflu 5329	<i>mwsR</i>	DGC	2908	T>C	F970L	substitution
Pflu 5329	<i>mwsR</i>	DGC	3580	C>T	Q1194*	truncation
Line 43						
Pflu 1219	<i>wspA</i>	putative methyl accepting chemotaxis protein	1362-1421	del 73bp		deletion
Pflu 1224	<i>wspF</i>	chemotaxis specific methyl-esterase	937-939	del 3bp	313del	deletion in-frame
Pflu 1768		hypothetical protein		G>A	none	synonymous
Pflu 3027		putative short-chain dehydrogenase	262	G>T	G88C	substitution
Pflu 5211	<i>awsX</i>	hypothetical protein	170	T>C	Y57C	substitution
Line 54						
Pflu 0185		putative sensory box GGDEF/EAL domain-containing protein	26	A>G	N9S	substitution
Pflu 0185		putative sensory box GGDEF/EAL domain-containing protein	551	T>G	I184S	substitution
Pflu 0414	<i>gltB</i>	glutamate synthase	1649	T>C	V550A	substitution
Pflu 1224	<i>wspF</i>	chemotaxis specific methyl esterase	937-939	del 3bp		frameshift
Pflu 1225	<i>wspR</i>	DGC	663-883	del 234bp		frameshift
Pflu 3027		putative short-chain dehydrogenase	235-237	del 3bp	G88del	deletion in-frame
Pflu 5210	<i>awsR</i>	DGC	79	T>G	T27P	substitution

Pflu 5329	<i>mwsR</i>	DGC	3071-3079	del 9bp	R1024-E1026del	deletion in-frame
Pflu 5329	<i>mwsR</i>	DGC	3439-3444	del 6bp	E1150-L1151del	deletion in-frame
Pflu 5691	<i>dbpA</i>	exhibits an RNA-dependent ATPase activity, specifically stimulated by bacterial 23S rRNA	539	G>T	R180L	substitution
Line 57						
pflu0185		putative GGDEF/EAL domain-containing protein	557	A>G	N186S	substitution
pflu0185		putative GGDEF/EAL domain-containing protein	601	C>T	R201C	substitution
Pflu 0458	<i>dipA</i>	PDE	1870	A>G	F624L	substitution
Pflu 4197/ Pflu 4198		putative LysR family transcriptional regulator, putative GGDEF/EAL domain protein	chimera	920bp deletion		chimera
Pflu 4439	<i>fliF</i>	flagellar M-ring protein	1000	ins TGGTGG		duplication
Pflu 4744	<i>algZ</i> , <i>amrZ</i>	putative DNA-binding protein, negative regulator of wss	236	T>C	L79P	substitution
Pflu 5210	<i>awsR</i>	DGC	1097	G>A	T366I	substitution
Pflu5211	<i>awsX</i>	negative regulator of awsR	223-255	33bp deletion		frameshift
Pflu5329	<i>mwsR</i>	DGC	3071-3079	9bp deletion	R1024-E1026del	deletion in-frame
Pflu5329	<i>mwsR</i>	DGC	3442-3447	del TGGAGC		deletion in-frame
Pflu5960		putative GGDEF/EAL domain-containing protein	1072	G>A	P358S	
Pflu5960		putative EAL/GGDEF domain-containing protein	822-836	15bp deletion		deletion in-frame

Table 3.1: Mutations present in each line as revealed by whole-genome sequencing at generation 11.

Highlighted mutations indicate a predicted involvement in the underlying GPM of the line or some other significance noted throughout the results. Due to the elevated mutation rate in line 17 only those mutations with a known or predicted relevance to the WS phenotype are shown, the other 38 mutations can be found in the appendix. DGC: Diguanylate Cyclase – synthesizes the secondary messenger c-di-GMP, an allosteric activator of cellulose synthase (characterized by the conserved domain motif GGDEF). PDE: Phosphodiesterase – degrades the secondary messenger c-di-GMP (characterized by the conserved domain motif EAL).

3.1.1 Overview of mutations

One feature that is marked throughout each of the lines is an activation-inactivation pattern represented by two mutations within a single DGC-encoding pathway. For instance, a mutation, most

commonly a loss-of-function to a negative regulator, activates the pathway responsible for the over-production of c-di-GMP which is subsequently inactivated via loss-of-function to the cognate DGC. Another overt feature is the lack of recorded mutations in accounting for the 20 transitions between phenotypic states required to reach generation 11 of the life cycle - assuming each transition requires a unique mutation.

3.2 Suppressor Analysis

Suppressor analysis was achieved by transposon mutagenesis with IS- Ω -Km/hah (Giddens *et al.*, 2007). Upon integration into the recipient genome the IS- Ω -Km/hah transposon generally disrupts function at the site of insertion and confers the cell with kanamycin resistance by which it can be selected. Depending on the orientation and placement of the transposon a gene may alternatively be transcriptionally activated by virtue of the encoded neomycin phosphotransfer (*nptII*) promoter. This allows identification of both genes which support the phenotype and genes which repress it. However in the same way loss-of-function mutations are most prevalent the requirement for a specific insertion means it is rarely activating. Another useful feature is the inclusion of two LoxP sites within IS- Ω -Km/hah which enable Cre-mediated excision of the majority of the transposon leaving only a 189bp scar at the site of insertion. This allows more localized mutational effects within the gene to be discerned while excluding the possibility of polar effects and *nptII* promoter activity.

The mutagenesis produced a variety of transconjugant colony morphologies. The reverted type were however easily distinguished and appeared as ancestral SM. Selection of these WS-suppressed transconjugant colonies was biased toward those of larger size. This was due to previous work suggesting insertions in the cellulose synthase operon *wss* resulted in smaller colonies. In addition, smaller colonies are likely to harbour transposons in loci causing metabolic defects that disrupt cellulose indirectly and so are likely to be uninformative. As the study progressed it became evident this was indeed the case so in order to enrich for those transconjugants of interest fewer small colonies were taken. Unique insertion sites indicated all transconjugants were independent.

Each line will be considered in turn. Connections between the results of the suppressor analysis and the mutational history as represented by the previous whole-genome sequencing will be made. Comparison with the GPM of the ancestral WS is based on the findings of the comprehensive suppressor analyses conducted on two WS evolved by one mutational step from SBW25: Large Spreading Wrinkly Spreader (LSWS) and Alternative Wrinkly Spreader (AWS) (Gehrig, 2005; McDonald, 2009). These suppressor analyses respectively uncovered the *wsp* and *aws* pathways and were designed to ensure full genome saturation. Those loci common to both WS mutants were hence

deemed necessary for WS in general, while those found in only one deemed specific to the activation pathway. This will be noted throughout.

3.2.1 Suppressor analysis: Line 17

Approximately 68,000 colonies were screened for the loss of the WS phenotype from 40 independent conjugations. From this, 68 transconjugants were retrieved and the position of the transposon insertion site determined by AP-PCR and Sanger sequencing. The WS-suppression frequency observed on plates was 0.11%.

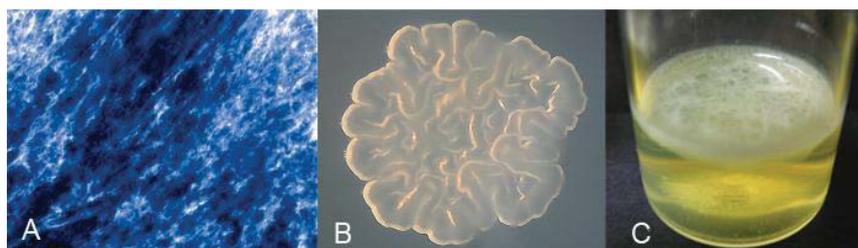


Figure 3.2: Morphological features of line 17. A) UV fluorescent image of extracellular cellulose matrix at 60x magnification. B) Colony morphology on plate after 48 hours growth C) Mat formed in microcosm

Insertion locus	Function	Notes
<i>wspA</i> (14)	methyl-accepting protein	D253Y mutation during LCE, Activation pathway
<i>wspD</i> (7)	scaffold protein - signal transduction	Activation pathway
<i>wspE</i> (13)	sensor kinase	Activation pathway
<i>wspR</i> (11)	response regulator, GGDEF domain	Activation pathway
<i>wspC</i> (2)	methyltransferase	Activation pathway
<i>wssA</i> (4)	Cellulose synthase subunit	Primary structural component of WS
<i>wssB</i> (3)	Cellulose synthase subunit	Primary structural component of WS
<i>wssC</i> (5)	Cellulose synthase subunit	Primary structural component of WS
<i>wssD</i> (1)	Cellulose synthase subunit	Primary structural component of WS
<i>wssE</i> (3)	Cellulose synthase subunit	Primary structural component of WS
pflu0411 ' <i>aroK</i> ' (1)	Shikimate kinase	Metabolic defect
pflu0670 (1)	short chain dehydrogenase	<i>aws</i> mutant suppression in pflu0673 and pflu0674
pflu2985 ' <i>galU</i> ' (1)	UTP--glucose-1-phosphate uridylyltransferase	Small colony. Metabolic component of WS

Table 3.2: The distribution of transposon insertions suppressing the WS phenotype in line 17. Activation pathway refers to the source of c-di-GMP production; scaffold components refer to those genes affecting structural aspects of the cell that were required in the ancestral state for WS production; transposon excision refers to the Cre-mediated excision of IS- Ω -Km/hah which leaves a 189bp scar, excluding the possibility of polar effects and activity of the transposon *nptII* promoter.

3.2.1.1 The suppressor loci indicate no change from an ancestral WS

There are no differences apparent in the suppression loci for Line 17 (wt *mutS*) compared to those of the ancestral *wsp* pathway mutants. Large Spreading Wrinkly Spreader (LSWS) was the first WS for which the GPM was identified through a comprehensive suppressor analysis (Gehrig, 2005). The mutational cause in LSWS was a loss-of-function to *wspF* and the sequence data for line 17 shows this same pattern (table 1.0). Transposons were found inserted in all genes of the *wsp* operon with the exception of *wspB* and *wspF*. After excision of each transposon by Cre-mediated recombination the corresponding mutant maintained suppression indicating that all were still necessary (all genes except *wspF* are required in LSWS) for the generation of WS. Insertions in *wspF* are not expected due to its role as a negative regulator of DGC-encoding *wspR*. The lack of insertions in the scaffold protein *wspB*, along with its 9bp in-frame deletion, might indicate its redundancy in line 17 although this might be due to the genes small size (513bp within the 8.3kb operon). I also found three insertion loci with a single representative; pflu0412 '*aroK*', pflu0670, and pflu2985 '*galU*'. This latter locus is responsible for the synthesis of UDP-glucose, a precursor for cellulose synthesis (Cannon & Anderson, 1991) and was seen to suppress both ancestral *wsp* and *aws* mutants indicating its role as an essential metabolic determinant of the WS phenotype. This transconjugant also displayed a small colony size. The putative short-chain dehydrogenase (pflu0670) has not been previously seen however two transposon insertions in upstream pflu0673 and pflu0674 suppressed *aws* mutants. The remaining locus, a shikimate synthase *aroK*, is involved in aromatic amino acid biosynthesis. All three of these loci are likely to represent metabolic defects.

A departure of the dual activation-inactivation mutation pattern can be seen in the *wsp* operon with three separate mutations situated in *wspA*, *wspB* and *wspF*. It is known that mutations in *wspF* are only capable of WS activation given its negative regulatory role. Mutations in *wspA* on the other hand have been associated with both gain and a loss of the WS phenotype in previous studies. There was however no pattern evident in their distribution nor any similarity with the *wspA* mutation (D523Y) found in line 17. There is no previous instance of *wspB* mutations either causing the gain or loss of WS. It is most likely that either the mutation in *wspA* or *wspB* is compensatory and facilitated the reactivation of the *wsp* pathway through the *mutS*-dependent switching mechanism.

3.2.2 Suppressor analysis: Line 43

Approximately 198,000 colonies were screened for the loss of the WS phenotype from 40 independent conjugations. From this, 66 transconjugants were retrieved and the positions of the transposons determined by AP-PCR and Sanger sequencing. The WS suppression frequency on plates was 0.04%, the lowest of any of the lines analysed. Unique insertion sites indicated all transconjugants were independent.

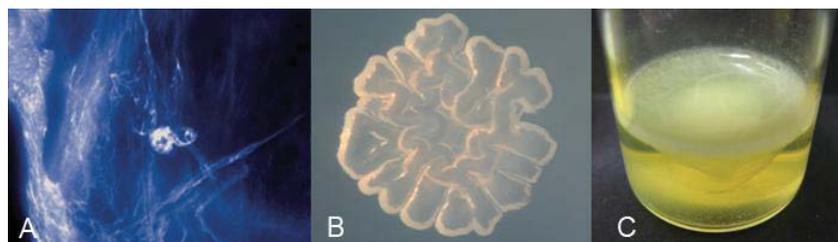


Figure 3.3: Morphological features of line 43. A) UV fluorescent image of extracellular cellulose matrix at 60x magnification. B) Colony morphology on plate after 48 hours growth C) Mat formed in microcosm

Insertion locus	Function	Notes
<i>awsR</i> (41)	DGC	Activation pathway
<i>awsX</i> (9)	putative negative regulator of AwsR	Activation pathway. Transposon excision relieves suppression
<i>wssA</i> (1)	Cellulose synthase subunit	Primary structural component of WS
<i>wssC</i> (2)	Cellulose synthase subunit	Primary structural component of WS
<i>wssD</i> (1)	Cellulose synthase subunit	Primary structural component of WS
<i>wssE</i> (2)	Cellulose synthase subunit	Primary structural component of WS
<i>pflu1687</i> (1)	methyl accepting chemotaxis protein, HAMP domain	
<i>pflu1545 'norM'</i> (1)	multidrug resistance protein	
<i>pflu0200</i> (1)	putative ABC sulfur transporter membrane protein	
<i>pflu0134</i> (1)	putative membrane protein	
<i>pflu5684 'betI'</i> (1)	HTH-type transcriptional regulator TetR-family	
<i>pflu2294</i> (1)	levansucrase	
<i>pflu0888</i> (1)	putative ABC transport system, ATP-binding protein,	<i>wsp</i> mutant suppression locus
<i>pflu3158</i> (1)	putative aminotransferase	
<i>pflu2463</i> (1)	putative metallothionein	
<i>pflu1603</i> (1)	putative two-component system sensor kinase	flanked by EAL domain response regulator. Transposon excision maintains suppression

pflu0841 '*petA*' (1) ubiquinol-cytochrome C reductase iron-sulfur subunit

Table 3.3: The distribution of transposon insertions suppressing the WS phenotype in line 43. Activation pathway refers to the source of c-di-GMP production; scaffold components refer to those genes affecting structural aspects of the cell that were required in the ancestral state for WS production; transposon excision refers to the Cre-mediated excision of IS- Ω -Km/hah which leaves a 189bp scar, excluding the possibility of polar effects and activity of the transposon *nptII* promoter.

3.2.2.1 The suppressor loci indicate no change from an ancestral WS

Suppressor analysis of this line revealed it to be underpinned by the common pathway *aws* with no clear alterations to the GPM apparent. Transposon insertions in the negative regulator *awsX* are expected to have a polar effect on the downstream *awsR* which encodes the DGC responsible for the upregulation of cellulose. Hits in *awsO*, a predicted sensory porin, are not expected to suppress the WS phenotype as, according to the current model, AwsO is necessary only for repression via AwsX and so is involved in the negative regulation of AwsR (McDonald *et al.*, 2009).

The low suppression frequency is unusual. Previous suppressor analysis of the ancestral *aws* WS had a suppression frequency almost three times higher. There are however a number of issues in comparing this value between studies - namely there is no measure of error. There is also the somewhat subjective nature of what constitutes a SM on plates and the possibility of misidentified clonal transconjugants (the previous analyses screened only 6 conjugations (Gehrig, 2005) or didn't report the number (McDonald, 2009)). But it's unlikely this can completely account for such a discrepancy. A comparison of target size (the sum of those loci specific to a line) between line 54 (*aws* + *fadA* + pflu5420 = 4.5kb) and line 43 (*aws* = 2.3kb), a difference of 51%, does align somewhat closer to the 70% difference in suppression frequency. It was also noted that line 43 cells had a tendency to clump together which may indicate the transconjugants were more likely to be clonal, effectively lowering the number of colonies screened. In fact in this line alone were patches of WS-suppressed transconjugants found when screening on plates and upon sequencing these were revealed to be clonal.

Line 43 also contains a number of anomalous insertion loci with only a single representative. With the exception of pflu0888, none of these loci have been previously reported to suppress WS. That there was only a single insertion could not be explained by small target size as many of these genes were large. One of the loci, pflu1603 was upstream of an EAL-domain containing protein and the transposon was in the correct orientation for *nptII* promoter activity to have caused its over-expression. The resulting increase in transcription of the phosphodiesterase could then act to degrade c-di-GMP and so decrease cellulose production. However the corresponding Cre-mediated excision mutant

maintained suppression. It is possible these insertions are not causative of the suppression but simply reside in a cell which switched on its own accord. However, in light of the low suppression frequency in line 43 it seems particularly odd: if this line is a proficient switcher then it makes the apparent low suppression frequency even lower.

Line 43 is also anomalous in its distinct paucity of mutations – whereas all other lines of the LCE see an average of 11 (excluding the hyper-mutable line 17), line 43 totals five with only three in known WS-generating pathways. The paired mutations in *wsp* suggest this was the first path taken: activation via loss-of-function to *wspF* followed by inactivation via loss-of-function to *wspA*. A mutation in *awsX*, the negative regulator of the *aws* pathway, then seemingly leads us to the current architecture.

Two other mutations have an unknown effect. One, a synonymous substitution in a small (291bp) hypothetical protein pflu1768 with no annotated domains or informative orthologs should have no effect. The second, a G88C substitution in pflu3027; a putative short-chain dehydrogenase, is of interest in that a similar mutation is seen in line 54 in which the same glycine residue is deleted. Such specific molecular parallelism is indicative of an adaptation to the some aspect of the life cycle regime, although its mode of action is currently obscure. This mutation will be further examined in the results of line 54 and discussed in chapter 4.

3.2.2.2 Suppressor analysis of line 43+3

The low number of mutations suggested line 43 may have arrived upon a switching strategy. In order to examine this idea further line 43 was subject to a second suppressor analysis at generation 14 (an additional three generations amounting to six transitions).

After an additional 3 generations of the LCE (see section 2.3) a representative WS was selected from the derived lineages of line 43, confirmed for the presence of cellulose by calcofluor staining, and subjected to transposon mutagenesis with IS- Ω -Km/hah. From this 33,000 colonies were screened and a total of 8 SM transconjugants retrieved. This amounts to a suppression frequency of 0.024%.

Insertion locus	Function
<i>awsR</i> (2)	DGC
<i>awsX</i> (4)	negative regulator of AwsR
pflu1732A (1)	Putative transposase

Table 3.4: Suppression loci of line 43+3

Although with a limited number of transconjugants the insertion loci indicate the six additional transitions have not altered the underlying DGC-activating locus. The putative transposase is unlikely to be a genuine suppressor locus.

3.2.3 Suppressor analysis: Line 54

A total of 70,000 colonies were screened for the loss of the WS phenotype from 45 independent conjugations. From this, 67 transconjugants were retrieved and the positions of the transposons determined. The WS suppression frequency on plates was 0.13%.

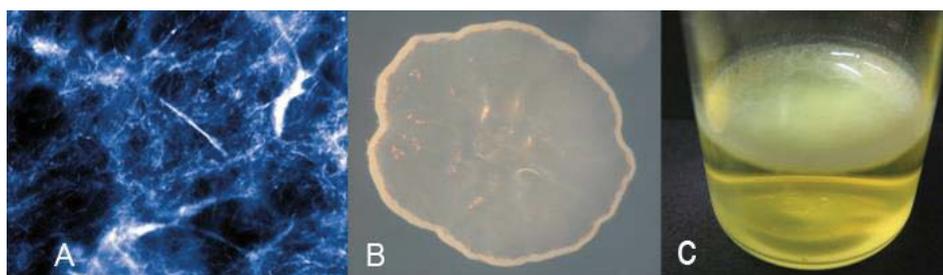


Figure 3.4: Morphological features of line 54. A) UV fluorescent image of extracellular cellulose matrix at 60x magnification. B) Colony morphology on plate after 48 hours growth C) Mat formed in microcosm

Insertion locus	Function	Notes
<i>awsR</i> (15)	putative signalling-related membrane protein, GGDEF domain	Activation pathway
<i>awsX</i> (6)	negative regulator of AwsR	Activation pathway
<i>pflu0184</i> (14)	putative fatty acid desaturase	
<i>pflu0183</i> (1)	GGDEF domain-containing	Suppression relieved upon transposon excision
<i>pflu0185</i> (1)	GGDEF domain-containing	Suppression relieved upon transposon excision
<i>wssB</i> (3)	Cellulose synthase subunit	Primary structural component of WS
<i>wssC</i> (3)	Cellulose synthase subunit	Primary structural component of WS
<i>wssE</i> (1)	Cellulose synthase subunit	Primary structural component of WS
<i>pflu3542</i> (1)	putative serine/threonine-protein kinase	<i>aws</i> mutant suppression locus
<i>pflu3541</i> (1)	conserved hypothetical protein	<i>aws</i> mutant suppression locus
<i>pflu5420</i> (5)	putative lipoprotein	Suppression maintained upon transposon excision
<i>pflu1555</i> (1)	hypothetical protein	Adjacent gene involved fatty acid metabolism
<i>pflu1667</i> ' <i>wagE</i> ' (1)	putative aminotransferase	Scaffold component of WS

pflu4186 (1)	dihydrofolate:folylpolyglutamate synthetase	Small colony
--------------	--	--------------

Table 3.5: The distribution of transposon insertions suppressing the WS phenotype in line 54. Activation pathway refers to the source of c-di-GMP production; scaffold components refer to those genes affecting structural aspects of the cell that were required in the ancestral state for WS production; transposon excision refers to the Cre-mediated excision of IS- Ω -Km/hah which leaves a 189bp scar, excluding the possibility of polar effects and activity of the transposon *nptII* promoter.

3.2.3.1 Suppressor analysis of line 54 reveals a previously unseen association between *aws* and a predicted fatty acid desaturase

Line 54 is underpinned by one of the previously ancestral WS pathways, *aws*. In contrast to the suppressor analysis however, an apparently novel association with a predicted fatty acid desaturase (pflu0184) was revealed. Transconjugants with inserts in pflu0184 were phenotypically SM on plates (but see fig 1.3 for further details), stained negatively for cellulose and colonised the broth phase of the microcosm. However, the ability to form a mat at the ALI was maintained, though it took longer to do so, being apparent after 3 days whereas line 54 mats are visible after 24 hours and maintain a clear broth throughout.

The predicted function of pflu0184 and the partial suppression of WS in the microcosm environment suggests an additive and physiological effect on the expression of *aws*. Nevertheless, this gene was not seen during the previous two suppressor analyses of the one-step *aws* type which in total screened approximately 200,000 colonies with the insertion locus determined in 300 transconjugants (Gehrig, 2005; McDonald, 2009). It is thus a product of evolution through the life cycle experiment and so a candidate for further analysis.

3.2.3.2 The role of pflu0184

Pflu0184 was seen during the RE to be involved in the generation of the WS phenotype via fusion with an adjacent DGC-encoding gene, pflu0183. This occurred at a fairly high frequency (8 out of the 91 WS mutants). The resulting chimera, dubbed *fwsR-fadA* (for Farr's Wrinkly Spreader and Fatty Acid Desaturase A), has been previously scrutinized and the precise mechanism revealed: association with the predicted fatty acid desaturase results in the re-localization of FwsR to the cell membrane where the cellulose synthase complex resides (Farr, 2015). Increased transcription of *fwsR* via promoter capture of *fadA* (pflu0184) alone was unable to cause the phenotype. As can be seen in the whole-genome sequencing of line 54 no evidence of such a fusion exists and this was confirmed by re-sequencing across the two genes.

3.2.3.3 *fadA*

The *fadA* gene encodes a predicted protein 394 residues in length and contains a predicted fatty acid desaturase at residues 11 to 246 (Pfam *E*-value $1.4e^{-14}$), a predicted transposase element at residues 261 to 389 (Pfam *E*-value 2.3×10^{-9}) and two predicted transmembrane domains (TMD) between residues 10-32 and 135-157. Transposon insertions appeared to cluster within the fatty acid desaturase and between the TMDs. Two anomalous insertions were located near to *fadA* but in the adjacent genes *pflu0185* and *pflu0183* which curiously both encode DGCs. However upon Cre-mediated excision from these loci WS suppression was relieved. Downstream effects from the insertions in *pflu0185* and *pflu0183* were discounted as *pflu0182* is transcribed in the opposite direction and *pflu0186* is bridged by a 300bp intergenic region which contains a predicted rho-independent transcriptional terminator. It was thus assumed the transposons at these locations were interfering with *fadA* somehow, possibly through altered transcriptional activity by the transposon-associated *nptII* promoter in the case of the insertion 5' of *fadA*.

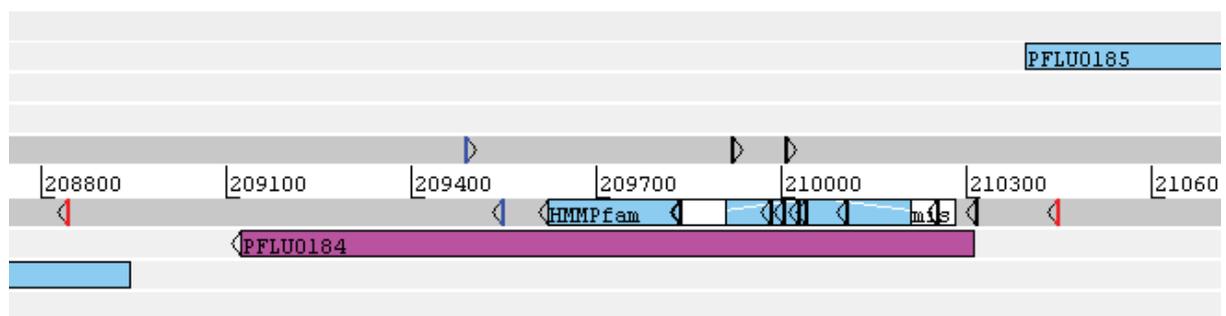


Fig 3.5: Distribution of transposon inserts affecting *fadA*. This figure illustrates the location and orientation of transposons (triangles) found within and near *fadA* (*pflu0184*) during the suppressor analysis. The colours correspond to the different effects upon excision. Excision of IS- Ω -Km/hah leaves a 189bp scar allowing more localized mutational effects to be discerned while excluding the possibility of *nptII* promoter activity. Red triangles indicate WS suppression being relieved upon excision. Blue triangles correspond to those transconjugants that displayed the induced WS phenotype on LB without salt. The majority of transposons (black triangles) can be seen clustered within the trans-membrane helices (white boxes) of the fatty acid desaturase domain (light blue box).

Little is known about *fadA* in *P. fluorescens* although an ortholog *desA* in *P. aeruginosa* with which it shares 83% of its amino acid sequence functions as a fatty acid desaturase (Zhu *et al.*, 2006). Fatty acid desaturases alter membrane composition to affect membrane fluidity in response to environmental factors. In *P. aeruginosa* *DesA* introduces double bonds into the $\Delta 9$ position of acyl chains resulting in an increase in membrane fluidity. Interestingly, *desA* transcription was shown to be negatively

correlated with oxygen levels while its enzymatic activity is oxygen-dependent (Zhu *et al.*, 2006). The potential significance of this will be discussed in chapter 4.

3.2.3.4 The role of *fadA* in line 54

As mentioned above *fadA* is positioned between two DGCs. Given the importance of these enzymes in establishing the WS phenotype this is a rather suspicious arrangement and suggested the possibility of a functional or regulatory relationship. Evidence of synteny throughout a number of orthologs also supported this. Of possible relevance here are the two mutations recorded in pflu0185 (N9S and I184S), although this gene is transcribed in the opposite direction to *fadA*. To gain insight into this relationship both pflu0183 and pflu0185 were deleted via allelic exchange from the Line 54 background. Both deletions resulted in an unchanged colony morphology indicating they were not interacting with *fadA*. Additionally, *fadA* was deleted from the line 54 background and an ancestral *aws* mutant background: the former to confirm its necessity and the latter to confirm its novelty. As was expected the line 54 Δ *fadA* mutant lost the WS phenotype while the AWS Δ *fadA* mutant remained unaffected.

3.2.3.5 Fatty acid desaturases

Connections between fatty acid composition and biofilm formation in *Pseudomonas* have been reported elsewhere (Blanka *et al.*, 2015). Of particular relevance is the proposed effect of membrane fluidity on activation of the Wsp signalling pathway in *P. aeruginosa*. Blanka *et al.* (2015) found parallel mutations affecting fatty acid synthesis in Small Colony Variants (SCV, an aggregative growth form similar to WS) of *P. aeruginosa* isolated from cystic fibrosis patients. The mutations were shown to increase the proportion of shorter-chain fatty acids in the plasma membrane and in turn activate the Wsp, leading to the SCV phenotype. A necessary precursor mutation was a loss-of-function mutation in the negative regulator *wspF* which, as in *P. fluorescens*, results in constitutive c-di-GMP production and overproduction of an exopolysaccharide and generation of the WS morphology. In this case the significance of the fatty-acid-affecting mutation was in establishing a reversible-switch which acted indirectly on Wsp through the physical properties of the membrane, although the reversibility of this was only inferred. It is thought that Wsp in *P. aeruginosa* responds to mechanical perturbation of the membrane, such as that which might be experienced during contact with a surface or from osmotic pressure, though the exact nature of the signal is obscure (Blanka *et al.*, 2015). The suggestion is that alterations in fatty acid membrane composition can incur conformational changes in the membrane-bound sensor proteins, simulating the effect of external stimuli and so affect the output of the pathway.

A similar sensing role for the AwsXRO ortholog YfiB_{NR} (notated in opposite direction to AwsXRO) in *P. aeruginosa* has also been proposed. Malone *et al.* 2012 suggest that the YfiB (AwsO) lipoprotein which spans the outer membrane and the peptidoglycan might monitor changes between these two layers. Changes in the relationship between the outer membrane and peptidoglycan would expose or obscure the hydrophobic YfiR (AwsX) binding site, controlling the amount of YfiR (AwsX) bound to YfiB (AwsO) and hence the level of YfiN (AwsR) activation. With this model it is easy to see how alterations to the physical properties of the outer membrane by FadA could influence the activity of the Aws pathway in line 54.

Both Aws and Wsp signal transduction pathways in *P. aeruginosa* can reportedly be induced to some extent in high salt environments (Blanka *et al.*, 2015). To explore whether such an effect exists in line 54, *fadA*-inactivated transconjugants were plated on LB containing varying concentrations of NaCl to examine whether the WS phenotype could be rescued under high salt conditions. There was no effect seen with increasing salt concentration, however two of the transconjugants showed a WS phenotype on LB without salt (details in Figure 3.5) although this wasn't apparent on salt-less KB, suggesting a simple absence of salt wasn't the specific factor. The precise insertion point of these transconjugants within *fadA* is at residues 252 and 274, placing the first just outside the transposase element and the second within it (Figure 3.5). It was also noted that wild-type line 54 appeared 'wrinklier' on salt-less LB, suggesting that those transconjugants at residues 252 and 274 were only partially impairing *fadA* activity and that this was overcome by some inducing property of the salt-less LB environment.

3.2.3.6 Pflu5420

A putative lipoprotein, pflu5420, was also found at high frequency during the suppressor analysis. Transconjugants appeared SM on plates but developed wrinkled edges after 4 days growth. Ability to form a mat was also maintained, although retarded in a similar way to *fadA* mutants. This gene was not identified in previous suppressor analyses suggesting its association with *aws* is also novel. It is however situated in a probable operon including cell-shape determinants *mrdB*, *pbpA* and *rodA* which each suppress WS by turning cells spherical (Spiers, 2002; McDonald *et al.*, 2009). However pflu5420-inactivated transconjugants remained rod-shaped and the corresponding excision mutant maintained suppression. Pflu5420 contains a RlpA-like double-psi beta barrel domain (Pfam *E*-value $9.0e^{-19}$) and a Sporulation related domain (Pfam *E*-value $8.2e^{-12}$). Its ortholog in *P. aeruginosa*, *rlpA* (rare lipoprotein A), is a septal ring protein with a role in morphogenesis and peptidoglycan metabolism (Jorgenson *et al.*, 2014).

3.2.3.7 Pflu1555

This small gene (236bp) gene is of unknown function but intriguingly is positioned adjacent to a putative 3-ketoacyl-CoA thiolase, an enzyme involved fatty acid beta oxidation. This arrangement shows strong synteny throughout all 42 orthologs in the Pseudomonas Genome Database (Windsor *et al.*, 2015). Although only represented by a single transposon insertion, this could be accounted for by its small size.

3.2.3.8 Mutational causes of altered GPM in line 54

What mutation or mutations have caused the changes we see in the GPM of Line 54 WS is uncertain. However the deletion of the same glycine residue as line 43, pflu3027 is of interest. This glycine constitutes the final residue of a NAD(P) binding site in a NADP_Rossmann superfamily (Pfam E-value $4.47e^{-122}$). This gene also contains a fabG Domain giving it a predicted role as a 3-ketoacyl-(acyl-carrier-protein) reductase (Pfam E-value $7.52e^{-67}$), an enzyme involved in fatty acid synthesis.

3.2.4 Suppressor analysis: Line 57

A total of 30,000 colonies were screened for the loss of the WS phenotype from 32 independent conjugations. From this, 67 transconjugants were retrieved and the positions of the transposons determined. The WS suppression frequency on plates was 0.3%, the highest of all lines.

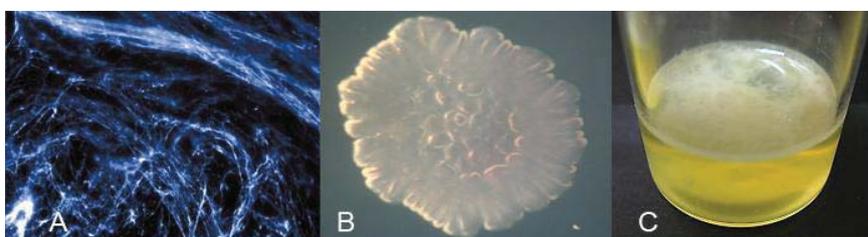


Figure 3.6: Morphological features of line 57. A) UV fluorescent image of extracellular cellulose matrix at 60x magnification. B) Colony morphology on plate after 48 hours growth C) Mat formation in microcosm

Insertion locus	Function	Notes
pflu4418 ' <i>fleN</i> ' (3)	ParA family ATPase, fleQ antagonist	Excision maintains suppression
pflu4419 (1)	Flagellar biosynthesis regulator FlhF	Excision relieves suppression
pflu1667 ' <i>wagE</i> ' (1)	putative aminotransferase	Scaffold component of WS
pflu4443 ' <i>adNa</i> ' ' <i>fleQ</i> ' (5)	fleQ domain (flagellum regulator), Sigma-54 interaction domain, HTH domain	

pflu4783 (1)	putative MCP-type signal transduction protein	biofilm dispersion gene ' <i>bdlA</i> ' ortholog in <i>P. aeruginosa</i>
<i>wssA</i> (4)	Cellulose synthase subunit	Primary structural component of WS
<i>wssB</i> (11)	Cellulose synthase subunit	Primary structural component of WS
<i>wssC</i> (13)	Cellulose synthase subunit	Primary structural component of WS
<i>wssD</i> (1)	Cellulose synthase subunit	Primary structural component of WS
<i>wssE</i> (7)	Cellulose synthase subunit	Primary structural component of WS
pflu5625 (3)	hypothetical protein	
Pflu2985 ' <i>galU</i> ' (2)	UTP--glucose-1-phosphate uridylyltransferase	Metabolic component of WS

Table 3.6: The distribution of transposon insertions suppressing the WS phenotype in line 57. Activation pathway refers to the source of c-di-GMP production; scaffold components refer to those genes affecting structural aspects of the cell that were required in the ancestral state for WS production; transposon excision refers to the Cre-mediated excision of IS- Ω -Km/hah which leaves a 189bp scar, excluding the possibility of polar effects and activity of the transposon *nptII* promoter.

3.2.4.1 Line 57 is underpinned by a complex architecture featuring a known negative regulator of *wss*

Suppressor analysis revealed an unusual architecture underpinning WS that was unlike any known pathway to the WS phenotype hitherto uncovered. Although with only a limited number of non-*wss* transconjugants at this time the main components appear to be *fleQ* (also known as *adnA*), *fleN*, pflu4783 and pflu5625. Cre-mediated excision of transposons from each gene maintained suppression, excluding polar effects and the activity of the *nptII* promoter. The pattern of *wss*-inactivated transconjugant colonies being relatively smaller as seen in the other lines was not adhered to in line 57 and so a visual screen against these was untenable. It is therefore possible that those loci listed in table 1.2 represent an incomplete picture of the GPM of this line.

The regulatory connections between these genes are complex but there exists significant literature regarding their function in *P. fluorescens* SBW25, as well as in closely related orthologs. This allows for some inferences to be made.

FleQ is a c-di-GMP-responsive transcriptional factor and master regulator of flagellum synthesis. It contains an N-terminal FleQ domain, an AAA σ 54 interaction domain, and a helix–turn–helix DNA binding domain. It was previously identified as a negative regulator of *wss* transcription in *P. fluorescens* SBW25 (Giddens *et al.*, 2007). Its current role as a positive regulator of the WS phenotype suggests a significant re-wiring of its regulatory connection with *wss* has occurred. This might be

explained by the possibility of FleQ exhibiting a bifunctional mode of regulation, in which case its role as an activator of *wss* has been induced. Such bi-functionality of FleQ has been observed to occur in regulation of the *pel* operon in *P. aeruginosa* PAO1 and requires the activity of *fleN*, a gene we also see in the suppressor analysis, which encodes an ATPase co-factor (Matsuyama *et al.*, 2016). The FleQ:DNA complex ordinarily represses *pel* expression. Under high cyclic-di-GMP conditions however a FleQ:DNA:FleN complex not only relieves repression but potently increases expression - leading to the production of extracellular polysaccharide and an aggregative colony morphology (Matsuyama *et al.*, 2016). The *pel* operon is also activated by the *wsp* pathway in *P. aeruginosa* PAO1 (Matsuyama *et al.*, 2016). Both FleQ and FleN in *P. fluorescens* SBW25 share high sequence homology with their PAO1 counterparts at 84% and 91%, respectively, suggesting a conserved functionality. It's possible this bi-functionality exists for the *wss* operon in SBW25 although *wss* has an additional layer of regulation missing in *pel*, that of the negative regulator *algZ* (also known as *amrZ*). In PAO1, AlgZ operates on a separate exopolysaccharide-encoding operon, *psl*, which doesn't display the bi-functionality of FleQ, although is nonetheless repressed by it (Baraquet and Harwood, 2016). AlgZ is also a transcriptional repressor of *fleQ*. A mutation to *algZ* is evident in line 57 and this is examined in detail below.

A third suppressor locus, *pflu4783*, also shares a well-studied ortholog in *P. aeruginosa*, *bdlA* (biofilm dispersion locus A) (Morgan *et al.*, 2006). This gene, as its name implies, encodes a protein involved in inducing dispersion that acts in response to high c-di-GMP levels. Its homology with *pflu4783* is fairly low at 52% so it's plausible it has a different function in SBW25. *Pflu4783* is a putative MCP-type signal transduction protein and contains two sensory PAS domains.

No suppressor loci with a direct role in c-di-GMP production were found. However a regulatory connection between *BdlA* and a specific phosphodiesterase 'DipA' (which is mutated in line 57) and a specific DGC 'GcbA' is known (Morgan *et al.*, 2006).

A key mutation with relevance to the loci of the suppressor analysis appears to be the loss-of-function to *algZ* (*pflu4474*, also known as *amrZ*). This gene is a known negative regulator of both *wss* and *fleQ* in *P. fluorescens* SBW25 (Malone *et al.*, 2007) which, if disrupted, causes an increase in cellulose production and a wrinkled colony morphology on plates, but fails to colonise the ALI due to an inability to adhere to the vial wall (Gehrig, 2005). It hence requires an additional mutation for full expression of the WS phenotype. This appears to have come through the F624L substitution in *pflu0458*. Both *algZ* and *pflu0458* constituted the most common 'double negative regulator' mutational pathway in the Re-Evolution experiment (RE) (Lind *et al.*, 2015). *Pflu0458* contains three sensory domains, an EAL domain (phosphodiesterase) and a degenerate GGDEF domain, with a catalytically impaired ASNEF active site. Its ortholog *DipA* (Dispersion-induced phosphodiesterase A) in PAO1 which shares the

same degenerate active site, has no DGC activity but instead functions solely as a PDE (Roy *et al.*, 2012). Interestingly the second double mutational pathway to WS seen in Lind *et al.*, (2015) involves a loss-of-function to *algZ* once again, but this time coupled with promoter activation of the putative DGC, pflu0621. This latter gene is known as *gcbA* in *P. aeruginosa* PAO1 and its encoded DGC appears to act in opposition to DipA, the dynamic of which modulates the activity of BdlA and so the transition between sessile and planktonic growth modes (Petrova *et al.*, 2015). This suggests *dipA* is a negative regulator of *gcbA* (the DGC) in *P. fluorescens*. BdlA activity is dependent on elevated c-di-GMP levels and involves an unusual post-translational proteolytic cleavage following which it oligomerizes with DipA, c-di-GMP levels are lowered, and dispersion initiated (Petrova & Sauer, 2012). What role *bdla* plays in supporting the WS phenotype in line 57 is not clear.

Although only speculative, what this suggests is that in line 57 the source of cyclic-di-GMP for which up to this point there is means of accounting for, is likely to be pflu0621 (*gcbA*), which appears to be negatively regulated by *dipA*. It is thus expected further transposon mutagenesis will reveal this locus. Alternatively, an additional mutation in *fliF*, which encodes a MS-ring protein crucial for the construction of flagella has led to the current genetic architecture, following activation via the *AlgZ-DipA* loss-of-function path just described. The relevance of the *fliF* mutation is essentially unknown however.

The final locus of the suppressor analysis is pflu5625, a gene which has no known function in SBW25 nor any informative orthologs. Of relevance is its single annotated cyclic nucleotide-binding domain (Pfam *E*-value $6.2e^{-8}$).

3.3 Additional life cycle generations

Although the derived lines from the LCE were examined in terms of their ability to produce the next stage of the life cycle this was assessed only over a single generation (with the exception of line 17, here excluded, that was taken through further generations in order to elucidate the *mutS*-dependent switch) (Hammerschmidt *et al.*, 2014). In order to examine the longer-term evolutionary fates of the other lines as well as have another time point at which to compare the evolution of the GPM I took each line through an additional three generations of the LCE. Due to its interest as a potential switcher a third generation line 43 descendent was taken and subjected to suppressor analysis (see section 3.2.2.2).

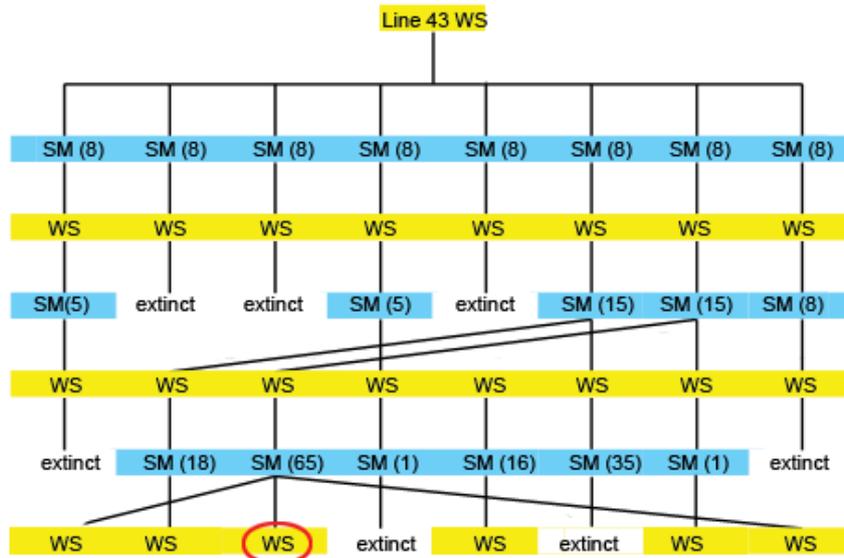


Figure 3.7: The fate of line 43 throughout 3 extra generations of the life cycle experiment

Numbers in parentheses denote the number of pooled SM colonies contributing to each microcosm. The transition from WS->SM represents 6 days of growth, while the transition from SM->WS represents 3 days. Extinction could be due to a lack of SM being produced (black line to missing blue bar “extinct”); a collapsed mat (no black line to missing yellow bar “extinct”) or a failure to produce WS (black line to missing yellow bar “extinct”). A third generation descendent (circled in red) was subject to suppressor analysis.

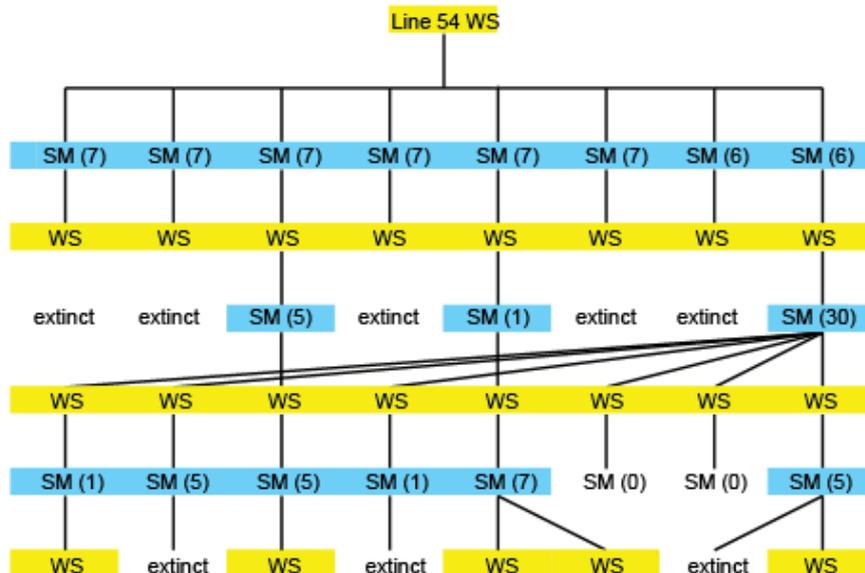


Figure 3.8: The fate of line 54 throughout 3 extra generations of the life cycle experiment

Numbers in parentheses denote the number of pooled SM colonies contributing to each microcosm. The transition from WS->SM represents 6 days of growth, while the transition from SM->WS represents 3 days. Extinction could be due to a lack of SM being produced (black line to missing blue bar “extinct”); a collapsed mat (no black line to missing yellow bar “extinct”) or a failure to produce WS (black line to missing yellow bar “extinct”).

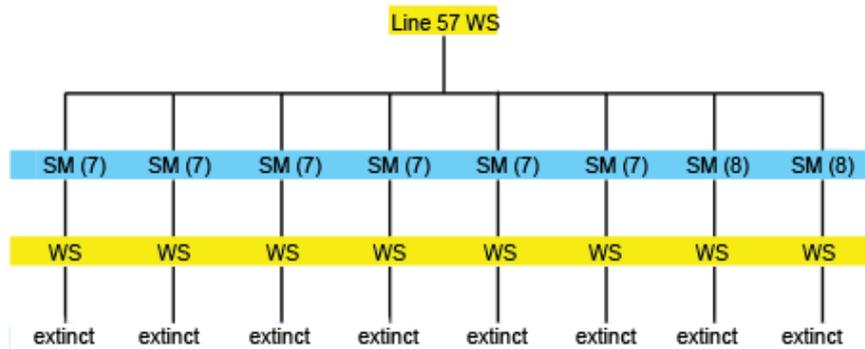


Figure 3.9: The fate of line 57 throughout 3 extra generations of the life cycle experiment

Numbers in parentheses denote the number of pooled SM colonies contributing to each microcosm. The transition from WS->SM represents 6 days of growth, while the transition from SM->WS represents 3 days. Extinction could be due to a lack of SM being produced (black line to missing blue bar “extinct”); a collapsed mat (no black line to missing yellow bar “extinct”) or a failure to produce WS (black line to missing yellow bar “extinct”).

4.0 Discussion

Motivated by the challenge of understanding the genotype-phenotype map and its effect on the course of evolutionary change I have here examined how the GPM itself has changed throughout an experiment which reiteratively selected for the gain and loss of the WS phenotype. That the production of cellulose can still be reached after so many successive disruptions is no trivial thing and, although knowledge of a number of pathways to WS from the ancestral state were known, this didn't necessarily entail they could be accessed sequentially, or how this process would be reflected as changes in the GPM. While the WS phenotype has stayed much the same (in terms of an ability to produce a cellulose mat) the genes underpinning it have in a number of cases drastically altered from the ancestral state. In other cases the genotype has also remained apparently unchanged.

In the following discussion I highlight the properties of the genetic architecture of ancestral SBW25 that have allowed for the maintenance of the WS phenotype and examine how these have manifested in the particular mutations and changes (or a lack thereof) we see to the GPM of WS. Together with the results of the extra life cycle generations a description of the evolutionary trajectory of each line is given, including speculation as to how some lineages have improved in their ability to passage through the life cycle. I conclude with a brief discussion on the current state of research regarding the GPM and its relevance to evolution.

4.1 Evolution of the GPM: a redundancy of pathways

The ability to transition between the phases of the life cycle has evidently been achieved, at least at first, by mutation to various DGC-encoding pathways. In each line we see a general pattern of activation-inactivation across multiple pathways, all represented by loci that have been previously associated with the production of WS (with the exception of the putative *lysR*-*pflu4198* fusion in line 57). That the activating and inactivating mutations are so tightly coupled (either in the same or a closely associated gene) attests to the modular nature of these regulatory pathways. Such modularity is a general feature of signal transduction which needs to be highly specific and insular so as to avoid unwanted cross-talk (Laub & Goulian, 2007). This likely poses significant challenges for c-di-GMP signalling in SBW25 given the vast number of DGC-encoding genes, many of which are likely to be paralogs. It is presently unknown quite how the c-di-GMP network coordinates the various signals, effectors and targets (Jenal & Malone, 2006; Hengge, 2009). Distinct c-di-GMP circuits could be

separated in time, through differential expression, and in space, through compartmentalized signalling (Jenal & Malone, 2006). Alternatively the signals may contribute to a global pool of c-di-GMP or contain local gradients as well (Jenal & Malone, 2006; Hengge, 2009). Regardless, the relevance to the present study is that the large number of DGCs provides significant evolvability (by way of redundancy) in respect to WS. This is not to be confused with the kind of redundancy that confers robustness - the classic example being a duplicated gene that masks the effect of a deleterious mutation to its paralog (gene duplications are also sources of evolvability through the potential of the paralog to diverge while maintaining function) (Ohno, 1970).

Consider the first transition from the ancestral SM type to WS. The mutational target in this case is the sum of all known pathways and likely others as-yet identified. The activation-inactivation pattern evident in each line strongly implies that once a pathway has been traversed its re-activation is unlikely (an interesting exception is line 17 in which re-activation of *wsp* was necessary for the *mutS* switch). This makes sense given that the mutation required is necessarily a gain-of-function and the likelihood of this occurring would be strongly biased by the kind of loss-of-function mutation that happened to disrupt the pathway. For example a large deletion is unlikely to be reverted or compensated. Thus the spectrum of mutations able to generate WS, which is at first vast, progressively diminishes - making each subsequent transition more improbable. Additionally, this transition needs to take place within the three day period provided by phase 2 of the LCE. This is a small window of time given that WS in microcosms founded by SM cells in which the common pathways were deleted took five days to appear, while those founded by ancestral SBW25 took only two, granted the number of microcosms assayed in this study was small (McDonald, 2009). Nevertheless the LCE has evidently provided ample opportunity for the traversal of alternative pathways (those other than *wsp*, *aws* and *mws*) as we can see in line 57.

Once the WS phenotype has been established the next requirement is to then revert to SM. The mutational target in this case is similarly large - in fact we can view all suppressor loci as analogous to a means of transitioning to SM via a loss-of-function mutation. As to why there is such a narrow representation of this target in the mutational history of the lines seems clear: those lines that did happen to transition via loss-of-function to an essential structural or metabolic component of WS likely became extinct. This isn't necessarily so of course - direct reversion, compensatory mutation or even the use of alternative components are all possibilities. Indeed the cellulose assay revealed one of the lines to have a non-cellulose mat, indicating alternative structural components have been utilised during the LCE. In addition to this, previous studies on SBW25 give evidence that direct reversions, particularly in *wss*, are more common than expected, suggesting mutational biases may be responsible for a number of the transitions (P.Rainey, personal communication). This could also

help explain the low number of mutations recorded throughout the lines. Alternatively the low number of mutations is explained by the evolution of a switching strategy, or perhaps some mutations simply weren't identified in sequencing. Given the nature of selection during the LCE it's also likely that lineages transitioning by direct reversions were favoured due to their maintained genetic potential: the WS mutational target does not diminish if transitions are by direct reversion.

With their known role in the degradation of c-di-GMP it's also possible phosphodiesterases, which are prevalent throughout the SBW25 genome and most commonly associated with GGDEF domains, might also provide a potential target for the transition to the SM phenotype. That no PDEs appear to have been activated may be explained by the requirement for rare gain-of-function mutations or alternatively speaks to the idea that cellular c-di-GMP levels are compartmentalized within the cell rather than contributing to a global pool (Jenal & Malone, 2006). In which case if a particular PDE was to inactivate WS via degradation of c-di-GMP, it would need to be co-localized.

The redundancy of pathways to WS, along with the possibility of direct reversions, confers a simple means of transitioning through the life cycle but as a longer-term strategy becomes increasingly difficult to maintain and, ultimately, is an evolutionary dead-end (as the fate of line 57 attests to). Perhaps the real advantage of these multiple WS pathways is that it allows time for an exploration of genotypic space - specifically, the space which consists of improvements in transitioning ability i.e. a switching strategy. This may be a general architectural feature influencing an organism's ability to arrive at strategies to deal with alternating environments - if it is a matter of probability as to whether a mutation or mutations conferring this sort of benefit arise, then an ability to respond at first by readily achievable mutations may buy enough time for the evolution of a more sophisticated mechanism.

Mutations that improve the ability to persist throughout the life cycle, even by a slight degree, are likely to have been selected in the LCE. A candidate for this sort of effect is a mutation found targeting the same residue of pflu3027 in both lines 43 and 54. In line 43 this takes the form of a G>T substitution resulting in the existing glycine residue being exchanged, while in line 54 this same glycine residue is removed by an in-frame deletion. Such specific molecular parallelism is the hallmark of adaptive evolution and requires explanation, although what this explanation might be is not certain. Pflu3027 does however have a predicted role as a 3-ketoacyl-(acyl-carrier-protein) reductase, an enzyme involved in fatty acid biosynthesis. This potentially relates to changes in the GPM revealed in line 54, specifically the association with *fadA*. The connection at this point is tenuous at best but nonetheless *fadA* has other sources of evidence which implicate it in a possible switching mechanism.

4.2 Possible switching mechanisms

As described earlier (section 3.2.3.3) *fadA* shares 83% homology with the *P. aeruginosa* fatty acid desaturase *desA*. A connection between fatty acid composition of the periplasmic membrane and activation of embedded sensory proteins (specifically WspA in *P. aeruginosa*) was established from the work of Blanka *et al.* (2015). The proposal that *aws* (the activation pathway underpinning line 54) might be amenable to such a process was suggested by the signalling mechanism of the close AwsXRO ortholog YfiBNR which, by homology, suggested that AwsO might sense mechanical perturbation of the cell wall through changes in the distance between peptidoglycan and the outer membrane (Malone *et al.*, 2012). This led to an attempt to induce the WS phenotype in line 54 through osmotic pressure by increasing NaCl concentration, which had little effect. The significance of all this in terms of a potential transitional mechanism for line 54 relies on the idea that *fadA* expression and activity may be dependent on oxygen levels. Evidence for this comes again from its ortholog *desA*, the expression of which is known to be negatively correlated with oxygen levels while its enzymatic activity is oxygen-dependent (Zhu *et al.*, 2006). Oxygen levels are an essential environmental component of the microcosm - the impetus for the evolution of WS and so driver of the entire system. A speculative model of *fadA*-mediated transitioning follows: SM cells proliferate throughout the liquid phase of the microcosm gradually depleting the dissolved oxygen contained within. As the oxygen depletes *fadA* expression increases, although its activity remains low. However when the SM cell happens to find itself at the oxygen replete ALI, FadA activity is suddenly increased, fatty acid membrane composition adjusted, and AwsR (the DGC) is activated: the WS phenotype is established. Abundant oxygen at the ALI would then decrease *fadA* expression in offspring cells which, lacking activity of the desaturase, would develop membranes that no longer induce AwsR, and therefore transition back to SM.

Whether such a mechanism would actually function in the LCE and then whether it actually exists in line 54 is perhaps doubtful. The number of extinctions throughout the additional life cycle generations doesn't seem to attest any precise transitioning mechanism - although it's of interest to note that a major early extinction was due to the collapse of mats, rather than the failure to produce SM or WS types. Additionally, in this model the deletion of *fadA* should completely abolish the mat forming ability of line 54 if it is alone responsible for activation of AwsR - which it does not. However there are two other suppressor loci in line 54 that are relevant. The putative lipoprotein pflu5420 which has a predicted role in peptidoglycan metabolism and morphogenesis and pflu1555, a small protein of unknown function adjacent to a putative 3-ketoacyl-CoA thiolase, an arrangement with strong synteny throughout orthologs suggestive of a functional connection.

4.3 Possible switching mechanisms: Line 43

That the GPM of the two lines harbouring the pflu3027 mutation show no similarities other than *aws* and *wss* suggests whatever effect pflu3027 might have had on the GPM of line 54 it is not the sole cause of changes we see. The most intriguing result from line 43 comes from the suppressor analysis of its 3rd generation descendent. That it was revealed to still be underpinned by *aws* is strongly suggestive of the evolution of a transition mechanism other than random mutation and, given the rate of extinction, almost certainly rules out direct reversions. Along with the few mutations recorded in this line (two in *wsp*, one in *awsX*, one synonymous change in an unknown protein) this further implicates pflu3027 as a mutation adapted to transitioning through the LCE.

4.4 Evolution of the GPM: scaffolding

An alternative explanation for the novel loci associated with *aws* in line 54 is that they represent a change to the scaffolding required for WS i.e. those components that maintain structural aspects of the cell necessary for cellulose production and export. Ancestral WS requires a number of genes related to cell-shape, cell wall biogenesis and membrane modification. An example from the suppressor analysis are the transposon inserts in pflu1667 in both line 54 and line 57. This gene constitutes part of an operon (pflu1668-pflu1661) encoding acyl carrier proteins (ACP), again involved in the synthesis of fatty acids. As in the line 54 Δ *fadA* mutant, transconjugants within this operon eliminated the wrinkled morphology on plates but did not abolish ability to form a mat at the ALI. (Gehrig, 2005). The suggestion is that for whatever reason - likely compensating for some unknown effect (possibly an undetected mutation) - an alteration to this scaffolding has occurred. There are also two other mutations in line 54 with an unaccounted affect - one in *gltB*, a glutamate synthase, and another in *dpbA*, an RNA helicase with specificity for 23s rRNA. Neither of these have any predicted relevance to the current GPM of line 54 WS. Assuming the adaptive mutation hypothesis for pflu3027, the presence of these other, likely deleterious or neutral, mutations are tolerated due to their association with pflu3027.

4.5 Evolution of the GPM: Line 57

If the aim is to find striking differences in the genetic architecture underpinning WS it is perhaps best to look at a line which quite clearly didn't evolve any strategy other than successive mutations, as it is only through mutation that genetic architecture is altered. Although evidently quite successful by making it to generation 11, line 57 appears to be of this type.

That line 57 has evolved no particular switching strategy is suggested firstly by the accordance of the mutational history with that predicted by the results of the re-evolution experiment (Lind *et al.*, 2015).

As expected, we see mutations in the common routes characterized by negative regulators (*aws* and *mws*), a rarer gene fusion/promoter capture event (pflu4197/4198), a putative intragenic activating mutation in pflu5960 and finally a double mutation to the extragenic negative regulators *algZ* and *dipA*. Both the intragenic activating and double negative regulator mutants were of the same loci as found by Lind *et al.* (2015) and occurred in 3 of the 91 microcosms in that experiment. Due to the connection between the double mutant loci and the extant GPM in line 57 it is reasonable to infer that these mutations occurred most latterly, which aligns with the prediction that its occurrence is less probable than either a fusion event or intragenic activation in activating WS. The wholesale extinction after two additional transitions in the life cycle also corroborates the view that this line is relying on mutation. So how has this whittling down of pathways manifested itself in the GPM of line 57?

Line 57 displays a vastly different GPM compared to the ancestral WS and any of the other lines. It is characterized by an apparent co-option of the flagellum synthesis master regulator *fleQ* and related loci. *fleQ* in ancestral SBW25 is a known negative regulator of *wss*, although a bifunctional mode of regulation may also be possible (see section 3.2.4.1). That it appears to now positively regulate *wss* attests to a rewiring of the regulatory circuits controlling its expression. The extent of connectivity and complexity of the regulatory relationships means a mechanistic account of how it is generating the WS phenotype is not possible at this time, nor is the exact mutational route to the present GPM fully understood. It is reasonable to surmise however that the loss-of-function mutations to *algZ* (a negative regulator of *wss* and *fleQ*) and *dipA* (a phosphodiesterase and possible negative regulator of the DGC, *gcbA*) are a necessary precursor to the current GPM. An additional mutation to a crucial structural component of the flagellum, the MS-ring protein FliF, is also likely to be involved. Disruption of FliF in *P. aeruginosa* leads to loss of flagella (Arora *et al.*, 1996). It's likely the mutation to *fiF* also incurs a loss-of-function and so loss of flagella. In this case the re-wiring of *fleQ* as positive regulator of *wss* has seemingly resulted from at least two, or possibly three loss-of-function mutations. The GPM of line 57 is also a demonstration of the creative re-wiring of regulatory components

The known and predicted functions of both *fleQ* (as master flagellum regulator and negative regulator of *wss*) and *bdIA* (interacts with the DGC-encoding *gcbA* and the PDE-encoding *dipA*) suggests they sit at the core of the c-di-GMP network, where regulation of flagella synthesis and exopolysaccharide (EPS) synthesis converge. In general, high cellular c-di-GMP levels induce EPS synthesis and low c-di-GMP levels induce flagellum synthesis (Matsuyama *et al.*, 2016), although perhaps this is no longer the case in line 57 - no locus with a direct involvement in c-di-GMP production was found in the suppressor analysis (although as stated above DipA may negatively regulate the DGC GcbA). In comparison to the modular pathways characteristic of ancestral WS these loci can hence be considered hubs of the c-di-GMP network, with an associated pleiotropic relationship to flagella and

EPS. The suppression frequency of this line was also found to be over twice as high as any other, suggesting an expansion of the network of genes underlying WS.

Because it is known mutation to *algZ* and *dipA* are together sufficient to form the WS phenotype, this perhaps implicates the *fliF* mutation as responsible for an additional transition, assuming it came after. Whether this was a transition for the gain or loss of WS cannot be said at this time. In either context however the significance of *fliF* in being a mutational target for that transition was determined by the genetic background created by the preceding mutations to *algZ* and *dipA*. This implicates these two mutations not simply as superficial 'scars' as is the case with other mutational pathways, which could presumably be traversed in any order, but as epistatic determinants of a particular course. It's of interest to note that line 57 shows no evidence of having traversed the *wsp* pathway - the most common route and taken in all other lines - indicating that perhaps it is irrevocably committed to another path, as is suggested by the *fliF* mutation. This might be expected following mutation to more highly connected genes: it's no longer a simple matter of disrupting the output of a single module as epistatic effects create a more complicated mutational target, with each subsequent mutation further altering this target.

4.6 Concluding discussion

The connection between genotype and phenotype has been an important question in evolution since the re-discovery of Mendel's work at the beginning of the 20th century. It wasn't until relatively recently however that a more sophisticated framework for understanding this relationship was developed. This required realization that genes did not act in isolation on a single trait, but that the construction of phenotypes was most often governed by multiple interacting genes and further, that genes could influence multiple phenotypic traits. Since first described by Albercht (1991), the GPM concept has laid the way for copious literature (Wagner & Altenberg, 1996; Hansen, 2003; Carlbord *et al.*, 2006; Crombach & Hogeweg, 2008; Pigliucci, 2010; Pavlicev, 2011), most of it theoretical in nature (for example, Wagner & Altenberg, 1996), regarding those properties of genetic architecture that allow organisms to evolve. Albercht's conception of the GPM emphasized the developmental aspect of genes - genes weren't simply the determinants of form but instead controlled developmental processes which, in turn, generate form. Subsequent work has led to the GPM being largely described in terms of connectivity between components; modularity, pleiotropy and epistasis (Wagner & Altenberg, 1996) and used to frame questions regarding evolvability, robustness and even the 'evolution of evolvability' (Dawkins, 1989; Pigliucci, 2008; Wagner & Altenberg, 1996). This latter

notion concerns whether the GPM itself has evolved to improve its ability to generate phenotypically useful variation (Wagner & Altenberg, 1996).

Another aspect of the GPM that has been widely studied is its potential for predicting the path of evolution. Mutation may be random, but as the present study and many previously have demonstrated, the phenotypic consequences of it are not (Gompel & Prud'homme, 2009; Lind *et al.*, 2015). Rather, they depend entirely on the GPM. Understanding the GPM and how it constrains the phenotypic consequences of mutation can hence go some ways to establishing a predictive element to the study of evolution. The field of evolutionary development (evo-devo) has made some headway in this regard (Gompel & Prud'homme, 2009; Stern & Orgogozo, 2008).

Evo-devo generally focuses on morphological adaptations of complex multicellular organisms and recent studies have attempted to explain patterns of mutational bias in terms of constraints imposed by the GPM. One of the most interesting findings from this has been that mutational biases differ from within-species to between species. For instance loss-of-function mutations, which arise frequently and often cause pleiotropic effects, seem to contribute more to phenotypic variation within species than between species, whereas for mutations to *cis*-regulatory genes, the opposite is true (Gompel & Prud'homme, 2009; Wittkopp & Kalay, 2012). This is interpreted to mean that loss-of-function mutations, which arise more readily due to their larger mutational target, are the first mutational step in differentiation. Subsequent change however is mediated by *cis*-regulatory mutations which are able to maintain or refine this initial variation and lessen pleiotropic effects (Gompel & Prud'homme, 2009).

However a prediction of genetic evolution via properties of the GPM has been never so explicitly stated as in the work of Lind *et al.* (2015) which has been heavily referred to throughout this thesis.

My work has done something entirely different. I have taken advantage of that fact that a simple phenotype is able to sequentially re-evolve. This phenotype can then act as a reference point by which the map is fixed, allowing the genotype to shift beneath it and the dynamics of this be observed. This concerns the fluidity of genetic architecture and has enabled me to address questions not only of how many ways there are to generate the same phenotype (as in the Lind *et al.* (2015) study) but also what sort of re-wiring of linkages between regulators and structural components does this entail, to what extent is re-wiring even possible and how this comes about. Additionally this approach is able to address questions of contingency such as that posed by the possible epistatic *fliF* mutation being reliant on the preceding *algZ* and *dipA* mutations and the irreversibility of that course. With the various stages of the evolution of each line frozen as revivable 'fossils', the ability to deduce the order of mutations and correlate these directly with changes in the GPM is significant.

It mustn't be forgotten that this study also concerned the genetic basis of a very particular experiment. The life cycle experiment is a unique means of exploring evolution of the GPM in the context of a major evolutionary transition. Potential switching mechanisms uncovered in this study require further scrutiny but the possibility of an oxygen-responsive fatty acid desaturase transducing a signal through the biophysical properties of the cell membrane is an exciting prospect.

Finally, this study has a general relevance to microbial evolution in contexts where repeated gain and loss of a trait is adaptive. Such conditions might be experienced in fluctuating environments or in response to ephemeral threats such as antibiotics. My findings demonstrate that bacteria can repeatedly return to a particular phenotype via a variety of different genetic routes.

Bibliography

- Altenberg, L. (1995). Genome growth and the evolution of the genotype-phenotype map. In *Evolution and biocomputation* (pp. 205-259). Springer Berlin Heidelberg.
- Arora, S. K., Ritchings, B. W., Almira, E. C., Lory, S., & Ramphal, R. (1996). Cloning and characterization of *Pseudomonas aeruginosa* *fliF*, necessary for flagellar assembly and bacterial adherence to mucin. *Infection and immunity*, *64*(6), 2130-2136.
- Bantinaki, E., Kassen, R., Knight, C. G., Robinson, Z., Spiers, A. J., & Rainey, P. B. (2007). Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. *Genetics*, *176*(1), 441-453.
- Barabasi, A. L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature reviews genetics*, *5*(2), 101-113.
- Beaumont, H. J., Gehrig, S. M., Kassen, R., Knight, C. G., Malone, J., Spiers, A. J., & Rainey, P. B. (2006, January). The genetics of phenotypic innovation. In *SYMPOSIA-SOCIETY FOR GENERAL MICROBIOLOGY* (Vol. 66, p. 91). Cambridge; Cambridge University Press; 1999.
- Blanka, A., Düvel, J., Dötsch, A., Klinkert, B., Abraham, W.R., Kaefer, V., Ritter, C., Narberhaus, F. and Häussler, S., 2015. Constitutive production of c-di-GMP is associated with mutations in a variant of *Pseudomonas aeruginosa* with altered membrane composition. *Science signaling*. *8*(372), ra36
- Cannon, R. E., & Anderson, S. M. (1991). Biogenesis of bacterial cellulose. *Critical reviews in microbiology*, *17*(6), 435-447.
- Carlborg, Ö., Jacobsson, L., Åhgren, P., Siegel, P., & Andersson, L. (2006). Epistasis and the release of genetic variation during long-term selection. *Nature genetics*, *38*(4), 418-420
- Cooper, T. F., Rozen, D. E., & Lenski, R. E. (2003). Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *100*(3), 1072-1077.
- Crombach, A., & Hogeweg, P. (2008). Evolution of evolvability in gene regulatory networks. *PLoS Comput Biol*, *4*(7), e1000112.
- Dawkins, R. (1989). The evolution of evolvability. In: *Artificial Life*, C. Langton, (ed.). Addison-Wesley

- Félix, M. A., & Wagner, A. (2008). Robustness and evolution: concepts, insights and challenges from a developmental model system. *Heredity*, *100*(2), 132-140
- Figuerski, D. H., & Helinski, D. R. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proceedings of the National Academy of Sciences*, *76*(4), 1648-1652.
- Gehrig, S. (2005). *Adaptation of Pseudomonas fluorescens SBW25 to the air-liquid interface: a study in evolutionary genetics* (Doctoral dissertation, University of Oxford).
- Giddens, S. R., Jackson, R. W., Moon, C. D., Jacobs, M. A., Zhang, X. X., Gehrig, S. M., & Rainey, P. B. (2007). Mutational activation of niche-specific genes provides insight into regulatory networks and bacterial function in a complex environment. *Proceedings of the National Academy of Sciences*, *104*(46), 18247-18252.
- Gompel, N., & Prud'homme, B. (2009). The causes of repeated genetic evolution. *Developmental biology*, *332*(1), 36-47.
- Gu, Z., Steinmetz, L. M., Gu, X., Scharfe, C., Davis, R. W., & Li, W. H. (2003). Role of duplicate genes in genetic robustness against null mutations. *Nature*, *421*(6918), 63-66.
- Halder, G., Callaerts, P., & Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science*, *267*(5205), 1788-1792.
- Hansen, T. F. (2006). The evolution of genetic architecture. *Annual Review of Ecology, Evolution, and Systematics*, 123-157.
- Hayden, E. J., Ferrada, E., & Wagner, A. (2011). Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature*, *474*(7349), 92-95.
- Hedrick, P. W. (1999). Antagonistic pleiotropy and genetic polymorphism: a perspective. *Heredity*, *82*(2), 126-133.
- Hengge, R. (2009). Principles of c-di-GMP signalling in bacteria. *Nature Reviews Microbiology*, *7*(4), 263-273.
- Hoekstra, H. E., & Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution*, *61*(5), 995-1016.
- Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., & Pease, L. R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, *77*(1), 61-68.

- Jenal, U., & Malone, J. (2006). Mechanisms of cyclic-di-GMP signaling in bacteria. *Annual Review Genetics*, *40*, 385-407.
- Jeong, H., Mason, S. P., Barabási, A. L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41-42.
- Jorgenson, M. A., Chen, Y., Yahashiri, A., Popham, D. L., & Weiss, D. S. (2014). The bacterial septal ring protein RlpA is a lytic transglycosylase that contributes to rod shape and daughter cell separation in *Pseudomonas aeruginosa*. *Molecular microbiology*, *93*(1), 113-128.
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, *217* (5129), 624-626.
- Kirschner, M., & Gerhart, J. (1998). Evolvability. *Proceedings of the National Academy of Sciences*, *95*(15), 8420-8427.
- Klingenberg, C. P. (1998). Heterochrony and allometry: the analysis of evolutionary change in ontogeny. *Biological Reviews of the Cambridge Philosophical Society*, *73*(01), 79-123.
- Laub, M. T., & Goulian, M. (2007). Specificity in two-component signal transduction pathways. *Annual Review Genetics*, *41*, 121-145.
- Lenski, R. E., Barrick, J. E., & Ofria, C. (2006). Balancing robustness and evolvability. *PLoS biology*, *4*(12).
- Libby, E., & Rainey, P. B. (2013). Eco-evolutionary feedback and the tuning of proto-developmental life cycles. *PLoS one*, *8*(12), e82274.
- Lind, P. A., Farr, A. D., & Rainey, P. B. (2015). Experimental evolution reveals hidden diversity in evolutionary pathways. *Elife*, *4*, e07074.
- Lozada-Chavez, I., Janga, S. C., & Collado-Vides, J. (2006). Bacterial regulatory networks are extremely flexible in evolution. *Nucleic acids research*, *34*(12), 3434-3445.
- Malone, J. G., Jaeger, T., Manfredi, P., Dötsch, A., Blanka, A., Bos, R., ... & Jenal, U. (2012). The YfiBNR signal transduction mechanism reveals novel targets for the evolution of persistent *Pseudomonas aeruginosa* in cystic fibrosis airways. *PLoS Pathology*, *8*(6), e1002760.
- Matsuyama, B. Y., Krasteva, P. V., Baraquet, C., Harwood, C. S., Sondermann, H., & Navarro, M. V. (2016). Mechanistic insights into c-di-GMP-dependent control of the biofilm regulator FleQ from *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences*, *113*(2), E209-E218.

- Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., ... & Gwadz, M. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic acids research*, *39*(suppl 1), D225-D229.
- McDonald, M. J. (2009). The genetics of *Pseudomonas fluorescens* SBW25: adaptation to a spatially structured environment: a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Genetics at Massey University, Auckland Campus.
- McDonald, M. J., Gehrig, S. M., Meintjes, P. L., Zhang, X. X., & Rainey, P. B. (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.
- Mezey, J. G., Cheverud, J. M., & Wagner, G. P. (2000). Is the genotype-phenotype map modular?: a statistical approach using mouse quantitative trait loci data. *Genetics*, *156*(1), 305-311.
- Michod, R. E. (2000). *Darwinian dynamics: evolutionary transitions in fitness and individuality*. Princeton University Press.
- Morgan, R., Kohn, S., Hwang, S. H., Hassett, D. J., & Sauer, K. (2006). BdlA, a chemotaxis regulator essential for biofilm dispersion in *Pseudomonas aeruginosa*. *Journal of bacteriology*, *188*(21), 7335-7343.
- Ohno, S. (1970). The enormous diversity in genome sizes of fish as a reflection of nature's extensive experiments with gene duplication. *Transactions of the American Fisheries Society*, *99*(1), 120-130.
- Pavlicev, M., Norgard, E. A., Fawcett, G. L., & Cheverud, J. M. (2011). Evolution of pleiotropy: epistatic interaction pattern supports a mechanistic model underlying variation in genotype–phenotype map. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *316*(5), 371-385
- Petrova, O. E., Cherny, K. E., & Sauer, K. (2015). The diguanylate cyclase GcbA facilitates *Pseudomonas aeruginosa* biofilm dispersion by activating BdlA. *Journal of bacteriology*, *197*(1), 174-187.
- Petrova, O. E., & Sauer, K. (2012). Dispersion by *Pseudomonas aeruginosa* requires an unusual posttranslational modification of BdlA. *Proceedings of the National Academy of Sciences*, *109*(41), 16690-16695.
- Phillips, P. C. (2008). Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, *9*(11), 855-867.

- Pigliucci, M. (2008). Is evolvability evolvable?. *Nature Reviews Genetics*, 9(1), 75-82.
- Pigliucci, M. (2010). Genotype–phenotype mapping and the end of the ‘genes as blueprint’ metaphor. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1540), 557-566.
- Proulx, S. R., Promislow, D. E., & Phillips, P. C. (2005). Network thinking in ecology and evolution. *Trends in Ecology & Evolution*, 20(6), 345-353.
- Prud'homme, B., Gompel, N., & Carroll, S. B. (2007). Emerging principles of regulatory evolution. *Proceedings of the National Academy of Sciences*, 104 (suppl 1), 8605-8612.
- Rainey, P. B. (2007). Unity from conflict. *Nature*, 446(7136), 616-616.
- Rainey, P. B., & Kerr, B. (2010). Cheats as first propagules: a new hypothesis for the evolution of individuality during the transition from single cells to multicellularity. *Bioessays*, 32(10), 872-880.
- Rainey, P. B., & Rainey, K. (2003). Evolution of cooperation and conflict in experimental bacterial populations. *Nature*, 425(6953), 72-74.
- Rainey, P. B., & Travisano, M. (1998). Adaptive radiation in a heterogeneous environment. *Nature*, 394(6688), 69-72.
- Rainey, P. B., & Cooper, T. F. (2004). Evolution of bacterial diversity and the origins of modularity. *Research in microbiology*, 155(5), 370-375.
- Rockman, M. V. (2008). Reverse engineering the genotype–phenotype map with natural genetic variation. *Nature*, 456(7223), 738-744.
- Roy, A. B., Petrova, O. E., & Sauer, K. (2012). The phosphodiesterase DipA (PA5017) is essential for *Pseudomonas aeruginosa* biofilm dispersion. *Journal of bacteriology*, 194(11), 2904-2915.
- Ryjenkov, D. A., Tarutina, M., Moskvina, O. V., & 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.
- Shao, H., Burrage, L. C., Sinasac, D. S., Hill, A. E., Ernest, S. R., O'Brien, W., ... & Daly, M. J. (2008). Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. *Proceedings of the National Academy of Sciences*, 105(50), 19910-19914.
- Smith, J. M., & Szathmari, E. (1997). *The Major Transitions in Evolution*. Oxford University Press.

- Spiers, A. J., Kahn, S. G., Bohannon, J., Travisano, M., & Rainey, P. B. (2002). Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics*, *161*(1), 33-46.
- Stern, D. L. (2000). Perspective: evolutionary developmental biology and the problem of variation. *Evolution*, *54*(4), 1079-1091.
- Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics*, *14*(11), 751-764.
- Stern, D. L., & Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution?. *Evolution*, *62*(9), 2155-2177.
- Stern, D. L., & Orgogozo, V. (2009). Is genetic evolution predictable?. *Science*, *323*(5915), 746-751.
- Tamayo, R., Tischler, A. D., & Camilli, A. (2005). The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase. *Journal of Biological Chemistry*, *280*(39), 33324-33330.
- Taylor, T. B., Mulley, G., Dills, A. H., Alsohim, A. S., McGuffin, L. J., Studholme, D. J., ... & Jackson, R. W. (2015). Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system. *Science*, *347*(6225), 1014-1017.
- Teichmann, S. A., & Babu, M. M. (2004). Gene regulatory network growth by duplication. *Nature genetics*, *36*(5), 492-496.
- Wagner, G. P., & Altenberg, L. (1996). Perspective: complex adaptations and the evolution of evolvability. *Evolution*, 967-976.
- Wagner, G. P., & Zhang, J. (2011). The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics*, *12*(3), 204-213.
- Whitacre, J., & Bender, A. (2010). Degeneracy: a design principle for achieving robustness and evolvability. *Journal of Theoretical Biology*, *263*(1), 143-153.
- Winsor, G. L., Griffiths, E. J., Lo, R., Dhillon, B. K., Shay, J. A., & Brinkman, F. S. (2016). Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic acids research*, *44*(D1), D646-D653.
- Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, *13*(1), 59-69.

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in *cis* and *trans* gene regulation. *Nature*, 430(6995), 85-88.

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, 8(3), 206-216.

Wright, S. (1984). *Evolution and the Genetics of Populations, Volume 3: Experimental Results and Evolutionary Deductions* (Vol. 3). University of Chicago press.

Zhu, K., Choi, K. H., Schweizer, H. P., Rock, C. O., & Zhang, Y. M. (2006). Two aerobic pathways for the formation of unsaturated fatty acids in *Pseudomonas aeruginosa*. *Molecular microbiology*, 60(2), 260-27

Full list of line 17 mutations

Gene locus	Gene coordinate	Annotation	Type of mutation	Effect of mutation	Mutation	Amino acid substitution
Pflu 0046	426	putative hydrolase	substitution	none	A>G	
Pflu 0270	492	cell filamentation protein	substitution	none	T>C	
Pflu 0337	898	catalyzes the interconversion of 2-phosphoglycerate and 3-phosphoglycerate	substitution		C>G	A>P residue 300
Pflu 0348	928	glutamine synthetase	substitution		C>T	G>S residue 310
Pflu 0541	993	hypothetical protein	insertion	frame shift	insG	
Pflu 0582	265	urease subunit gamma; UreA, with UreB and UreC catalyzes the hydrolysis of urea into ammonia and carbon dioxide; nickel...	substitution		C>T	G>S residue 89
Pflu 0616	777	ribosomal protein L11 methyltransferase	substitution	none	G>A	
Pflu 0823	1481	dipeptide ABC transporter substrate-binding protein	substitution		T>C	D>G residue 494
Pflu 0917	518	putative exported peptidase	insertion	frame shift	insG	
Pflu 0920	717	Catalyzes the reduction of hydroxypyruvate to form D-glycerate, using NADH as an electron donor	substitution	none	A>G	
Pflu 0954; Intergenic		between Pflu 0953 and 0954; hypothetical protein	deletion		delT	
Pflu 1093	1980	putative fimbrial outer membrane usher protein	insertion	frame shift	insG	
Pflu 1164	1489	DNA mismatch repair protein MutS; This protein performs the mismatch recognition step during the DNA repair process	substitution		T>G	T>P residue 497
Pflu 1219	757	putative methyl accepting chemotaxis protein	substitution	Transversion	G>T	D>Y
Pflu 1220	109-117	putative chemotaxis-like protein	deletion	Deletion in frame	delGCCGAAGTG	
Pflu 1224	884	chemotaxis specific methyl esterase	substitution		T>G	I>S residue 295
Pflu 1490	243	GntR family transcriptional regulator	substitution	none	A>G	
Pflu 1593	98	hypothetical protein	insertion	frame shift	insG	
Pflu 1668	1540	putative polysaccharide biosynthesis-related membrane protein	insertion	frame shift	insG	
Pflu 1732A		putative transposase	high coverage			
Pflu 1990	1111	putative transporter-like membrane protein	substitution		C>T	G>S residue 371
Pflu 2000; Intergenic		between Pflu 1999 and 2000; putative hydrolase and hypothetical protein	substitution	none	A>G	

Pflu 2079	259	putative polysaccharide exported-related lipoprotein	substitution		C>T	D>N residue 87
Pflu 2295	1088	putative methyl-accepting chemotaxis protein	substitution		A>G	D>G residue 363
Pflu 2381	1092	hypothetical protein	insertion	frame shift	insG	
Pflu 2732	762	hypothetical protein	substitution	none	C>T	
Pflu 2743	150	putative ABC transporter permease	substitution	none	G>A	
Pflu 2807	500	putative nicotinamide mononucleotide transporter	substitution		A>G	L>P residue 167
Pflu 2947	847	putative ABC transporter periplasmic binding protein	substitution		C>T	G>R residue 283
Pflu 3051A; Intergenic		between Pflu 3050 and 3051A ; hypothetical proteins	insertion		insC	
Pflu 3409	561	methyl-accepting chemotaxis protein	substitution	none	T>C	
Pflu 3426	1924	putative bifunctional reductase	substitution		G>A	E>K residue 642
Pflu 3488	305	hypothetical protein	insertion	frame shift	insCC	
Pflu 3728	104	putative ribose ABC transporter ATP-binding protein	substitution		C>T	S>L residue 35
Pflu 3951	628	substrate-binding periplasmic protein	insertion	frame shift	insC	
Pflu 3971A		putative transposase	low coverage (transposase gone)			
Pflu 4017	75	amino acid ABC transporter permease	insertion	frame shift	insC	
Pflu 4042	708	pyridine nucleotide-disulfide oxidoreductase	insertion	frame shift	insC	
Pflu 4306	403	putative GGDEF/GAF domain sensory box protein	substitution		G>C	A>P residue 135
Pflu 4466	991	patatin-like phospholipase	insertion	frame shift	insC	
Pflu 4572A		putative transposase	high coverage			
Pflu 4600	470	putative regulatory protein	substitution		T>C	Y>C residue 157
Pflu 4639A		putative transposase	low coverage (transposase gone)			
Pflu 4769	1726	hypothetical protein	substitution		A>T	G>A residue 576
Pflu 4801; Intergenic		between Pflu 4800 and 4801; LysR family transcriptional regulator	substitution		T>C	
Pflu 4976	304	putative HlyD family secretion protein	deletion	frame shift	delC	
Pflu 5210	79	putative signaling-related membrane protein CDS	substitution		T>G	T>P residue 27
Pflu 5210	620	putative signaling-related membrane protein CDS	substitution		G>A	P>m/L residue 207
Pflu 5329	2908	putative signaling-related membrane protein CDS	substitution		T>C	F>L residue 970

Pflu 5329	3580	putative signaling-related membrane protein CDS	substitution	Truncation	C>T	Q>* residue 1194
Pflu 5883	154	hypothetical protein	substitution		G>A	D>N residue 52
Pflu 5883	313	hypothetical protein	substitution		G>A	G>S residue 105
Pflu 5978	106	hypothetical protein	substitution		T>C	Y>H residue 36
Pflu 6104	1896	hypothetical protein	substitution	none	G>A	