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.

Michael Joseph McDonald

2009

ABSTRACT

Experimental microbial populations provide powerful models for testing the most challenging problems in evolutionary biology. In the midst of the genome sequencing revolution microbial evolutionary genetics has flourished; promising high-resolution explanations for the underlying causes of evolutionary phenomena. This thesis describes four investigations into the adaptation of *Pseudomonas fluorescesns* SBW25 to a spatially structured environment. The first builds upon a large body of experimental work characterising the genetic and phenotypic causes of the ability of divergent Wrinkley Spreader (WS) types to colonise the air-liquid interface in spatially structured microcosms. The *mws* and *aws* genetic loci are described, which together with the previously described *wsp* locus, account for the location of the causal mutation for all known WS genotypes. It was found that if these loci were deleted from the *P. fluorescens* genome, it could still evolve the WS phenotype via a previously undiscovered locus (*sws*). This study provides the first explicit evidence that genetic biases can influence the outcome of evolution. The second study used a novel method to sample WS genotypes without the biasing effects of natural selection; the distribution of the fitness effects of these genotypes was measured and analysed from a unique perspective. The distribution of fitness effects of new mutations is found to best fit the normal distribution, facilitating the extension of the mutational landscape model of adaptation to include all possible adaptive walks. The third study investigates the underlying causes of genetic biases on evolution; many WS genotypes are obtained at different time points during colonisation of the air-liquid interface (including WS obtained without selection) and the causal mutations of many of these mutants determined. Together these results allowed the elucidation of the relative effects of natural selection, genetic architecture and mutation rate on evolutionary outcomes. The final study considers the WS mat as the product of cooperative interactions, and uses a group selection experiment to investigate the potential of WS mats to evolve group level adaptations. A novel strategy is developed to overcome cheating types, considered the main barrier to the evolution of group level complexity. Furthermore, WS groups evolved specialised cell types, the first example of a *de novo* evolution of a division of labour, a hallmark of complexity.

ACKNOWLEDGMENTS

Like all things worth having a Ph.D takes a lot of time and work, however, I can say I have enjoyed nearly every minute of it. First I would like to thank Paul for taking me on as a student. On paper I was far from a sure bet and I am not sure how I got past his method for screening potential students. However that happened I am always grateful that I have been able to benefit from his mentorship and gain the new perspective of the world that comes with the study of Evolution.

Secondly, my co-supervisor Xue Xian deserves thanks for his always-sound advice and also for passing on his excellent lab techniques; I soon learnt the way to make it work was do exactly as he did.

The Rainey lab seems to attract excellent post docs. Among the many things taught to me I will remember these lessons: Bertus, for thinking deeply and outside the square; Tim, for perceptive analysis and introducing me to the big problems in evolutionary biology; Dominik, for expounding the history of evolutionary theory and Christian, for demonstrating the benefits of careful experiment design.

Fellow Ph.D students Jenna and Pete, who apart from providing a good source of commiseration, cooperation and information also set high standards for me to keep up with. Also Annabel, Ellen and Andy for technical and other help.

Thanks to my Parents, Bryan and Carol; without their support doing a PhD, getting married and having kids over the same period of time would have been impossible. Also Dad for teaching me how to examine a problem logically, a tool I use everyday.

Finally to my wife Wen-Pei, whose calming influence allowed me to focus my attention on a worth-while endeavour for the first time in my life. Also, by giving me two children before the age of 28, I became the fittest person (evolutionarily speaking) in the lab.

TABLE OF CONTENTS

FIGURES

TABLES

ABBREVIATIONS

WS – Wrinkly spreader

SM- Smooth

DGC- Di-Guanylate Cyclase

PDE- Phosphodiesterase

LSWS- Large Spreading Wrinkly Spreader

AWS- Alternative Wrinkly Spreader

MWS- Mike's Wrinkly Spreader

SWS- Slow Wrinkly Spreader

DFE- Distribution of Fitness Effects

EVT- Extreme Value Theory

CLT- Central Limit Theorem

MSC- Mutation Selection Cassette

1 INTRODUCTION

1.1 DARWIN'S INSIGHT

On the Origin of Species (DARWIN 1859) presented the foundation for our current understanding of the evolutionary process. Until that time species were thought to be the immutable creations of a divine being, the argument from design incontravertable evidence for the existence of god. Darwin proposed that all life's diversity, even the most complex structures found in nature, could be explained by a simple process; the differential survival of organisms based on heritable traits. The massive challenge issued by Darwin to the most established ideas of his time has elicited 150 years of intense scrutiny. Despite this, the core principles of Darwin's theory have remained intact; confirming Darwin's insight as one of the most significant in the history of science. In this section I introduce the merging of Darwinian natural selection and the particulate theory of inheritence that lead to the modern evolutionary synthesis.

1.1.1 THE MODERN SYNTHESIS

The outstanding gap in Darwin's theory was a satisfactory theory of inheritence; this deficiency formented doubt, and sixty years after *Origins,* natural selection was but one of four equally regarded evolutionary mechanisms[1]. The elegant combination of the rediscovered Mendelian genetics (MENDEL 1865) with Darwinian natural selection by R.A Fisher (FISHER 1918) showed the concept of genetic inheritance to be consistent with evolution by natural selection. Subsequent works by J. B. S. Haldane (HALDANE 1932) and Sewall Wright (WRIGHT 1932a) connected the work of Fisher to natural

[1] as well as Lamarkism (the inheritance of acquired traits), Saltationism (evolution in large phenotypic jumps) and Orthogenesis (an inner force driving evolution).

populations, resulting in the creation of the new discipline of population genetics. It seemed a new field was the last thing needed by the biological sciences, which at that time were composed of many quite separate disciplines (MAYR 1988), each with their own prevailing, and sometimes antagonistic, veiws on the evolutionary process. Population genetics was a catalyst for unification, spawning canonical works from Theodosius Dobzhansky (Drosophila Genetics) (DOBZHANSKY 1937), Ernst Mayr (Systematics) (MAYR 1942), George G. Simpson (Paleontology) (SIMPSON 1944), G. Ledyard Stebbins[2] (Botany) (STEBBINS 1950) and Julian Huxley (Biology[3]) (HUXLEY 1942). These works employed studies of natural populations to overcome the divisions of their predecessors, making the somewhat opaque mathematical treatises of Fisher, Haldane and Wright accessible to a wider range of biologists. Together these works comprised the modern synthesis, the tenants of which are listed below:

1. Genetic variation arises by random (with respect to fitness) mutation.

2. Genetic variation is heritable and encoded in discrete units (genes), which can be reorganised through recombination.

3. Gene frequencies within a population change by genetic drift, natural selection and migration.

[2] Sometimes included in this list are Edmund B. Ford and Sergei Chetverikov whose work influenced Dobzhansky, and Bernard Rensch whose work influenced Mayr. Though each made large contribution to evolutionary biology, they were only indirectly involved in the modern synthesis.

[3] Huxley corresponding with many other contributers to the modern synthesis, drew on a wide range of literature to produce its defining work and coined the term "modern synthesis".

4. The accumulation of genetic changes over sufficient time leads to the emergence of new species.

The acceptance of the genetical theory of evolution as proposed by the modern synthesis depended on data gathered during studies of natural populations of flora and fauna. Conducting such studies presents many practical problems; there are limits to what can be inferred by studying only the end result of a process, for instance, quite distinct evolutionary mechanisms could give rise to the same evolutionary outcome[4]. Furthermore, in order to make refinements to evolutionary theory and elucidate the molecular underpinnings of the newly appreciated genetic inheritance, carefully controlled experimental conditions were required. In the next section I introduce the concept and history of using controlled, experimental populations to test evolutionary theories.

[4] For example, parallel evolution could be explained by the action of natural selection, genetic constraints, or a combination of the two.

1.2 EXPERIMENTAL EVOLUTION

Evolutionary experiments have been carried out unintentionally by human breeders of animals for thousands of years. Darwin recognised their utility; a study on the variation of domestic animals comprised the first chapter of his *Origins* (DARWIN 1859). Animal husbandry was a useful model for Darwin; the effects of selection on domestic animals is easier to detect than in natural populations because quantitative traits are often recorded, facilitating comparison between ancestor and descendant. While the environmental conditions to which a wild species has been subjected to are unknown, the habitat of domestic animals is well understood. Moreover, the trait and its adaptive significance can be clearly identified as the selective pressure has been applied by the breeder, something very difficult to do for natural populations.

The juxtaposition of domestic and wild animal populations provides an illustration of the advantages gained by employing experimental instead of natural populations for the investigation of evolutionary processes. Although experimental evolution has been practised intermittently since soon after Darwin's thesis[5], it has only recently been realised that the full power of experimental evolution lies in propogating many replicate lines of evolving populations for many generations. One of the first organisms recruited to such studies was *Drosophilia melanogaster* in which selection for the increased lifespan of experimental populations provided a striking example of how populations could respond to selection in a relatively short amount of time (ENGSTROM *et al.* 1992; LINTS *et al.* 1979; LUCKINBILL and CLARE 1985). Although animals (GIRARD *et al.* 2001; LUCKINBILL and CLARE 1985; WADE 1976) and even entire ecosystems

[5] An early foray into experimental evolution was carried out by W. H. Dallinger; over a period of six years he found that unicellular organisms adapted to gradually increasing temperatures in an incubator (DALLINGER, W. H., 1887 The Presidents Address. Journal of Royal Microscopic Society **10:** 191-192.

(GOODNIGHT 2000; SWENSON and WILSON 2000) have been exploited as experimental evolution models, the practice of using experimental microbial populations to investigate evolutionary phenomena has grown the most rapidly, perhaps because of the ease with which it facilitates the direct testing of evolutionary theories.

1.3 MICROBIAL MODELS OF EVOLUTION

Microbial microcosms provide ideal worlds for the study of evolution (COLEGRAVE and BUCKLING 2005; JESSUP *et al.* 2004; LENSKI *et al.* 1991; TRAVISANO and RAINEY 2000). A single test tube can support billions of individuals and many independent replicate populations can be allowed to evolve in highly controlled conditions. The large supply of mutants within large populations offers the potential to sample the full range of mutational effects possible, with natural selection able to distinguish small differences in fitness. Natural selection is pervasive in large exponentially growing populations; the dominance of a single evolutionary force simplifys the study of evolutionary mechanisms. Bacterial physiology is well understood in many model systems, and robust protocols for their genetic manipulation have been established. This facilitates the tracking down and confirmation of the mutational changes that occur during adaptation. The ability to store samples for indefinite periods of time (LENSKI *et al.* 1991) allows direct genetic and phenotypic comparisons of descendant and ancestor.

1.3.1 THE ESCHERICHIA COLI LONG TERM EVOLUTION EXPERIMENT.

In the longest running evolution experiment ever, Richard E. Lenski and others have propogated 12 populations of *E. coli* since 1988, accumulating over 45,000 generations of evolution. The adaptation of these populations to their environment, minimal medium

with glucose as the sole source of carbon, has provided insight into several aspects of evolutionary theory. Parallel evolution[6] has been pervasive among all twelve populations; initially rapid gains in fitness decreased over time, indicating a diminishing supply of beneficial mutations as populations became better adapted (LENSKI *et al.* 1991; LENSKI and TRAVISANO 1994). Mutations in three of the same genes have been substituted in all populations; however, many mutations are unique to one or two of the populations and mutations accounting for the total observed fitness change during the course of the experiment have not yet been found. Some results have been suprising, for example, four of the 12 lines evolved high mutation rates (the mutator phenotype). High mutation rates are usually postulated to be maladaptive, except for when population fitness is very low, and a greater supply of mutations can facilitate faster adaptation (RAINEY 1999). Strangely, the mutator phenotype has been maintained in these four lines for the entire experiment despite the predicted decreasing need for a high mutation rate as the population becomes better adapted (DE VISSER *et al.* 1999; LENSKI *et al.* 2003; SNIEGOWSKI *et al.* 1997). A recent finding of general interest was that one of the 12 experimental lines had evolved (after 30,500 generations) the ability to utilise citrate as an alternative carbon source (BLOUNT *et al.* 2008). The inability to utilise citrate in oxidizing conditions is a defining characteristic of the *E. coli* species. This trait is difficult to evolve; despite each population sampling billions of mutations, only one of the 12 experimental lines evolved this advantageous phenotype. It was determined that this new trait could only arise in a cell that possesed the specific genotype found in this experimental line after 20,000 generations. Although each line has evolved similar increases in fitness during the long term experiment, different mutations, often in different genes, have by chance led to a potentiated genotype. Only in this genotypic background can a certain

[6] Parallel evolution is the evolution of similar phenotypic characters in independently evolving populations of the same or similar ancestry.

mutation have the effect of conferring the ability to utilise citrate, even if this mutation happened in another genotype, it would not have the same effect. This study concluded that the evolution of innovation sometimes depends on historical contingency (chance events) and not only gradual and cumulative selection. Because natural selection is short-sighted,[7] the accumulation of mutations that led to the potentiated genotype had to be random with respect to citrate utilisation; these mutations can not have been selected for on the basis of future potential. This experiment points to the influence that underlying genetics and chance can have on evolutionary outcomes.

As well as contributing significantly to our understanding of adaptation, this project has inspired the rapid expansion of the use of micro-organisms as experimental models of evolution. However, the fitness changes that have evolved in the *E. coli* populations are subtle, and the explicit connection of the genetic and phenotypic changes to the measured increases in fitness has not yet been achieved. In the next section I introduce *Pseudomonas fluorescens* SBW25, which because of the distinct phenotypic innovation required for its adaptation to a new niche, provides an excellent model for the explicit connection of genotype and phenotype during evolution.

1.4 *PSEUDOMONAS FLUORESCENS* SBW25, A MODEL SYSTEM.

When inoculated into a spatially structured environment (6 ml of KB broth in a static glass vial), wild type *P. fluorescens* SBW25 undergoes an adaptive radiation into a variety of niches (RAINEY and TRAVISANO 1998). Strong selective forces drive this diversification, during which a range of cell types evolve, each adapted to a specific niche. These divergent types comprise three classes, readily distinguishable on an agar plate. The

[7] Natural selection is said to be short-sighted because it cannot increase the frequency of mutations unless they are immediately beneficial in the context of the current environment and genotype, even if they have the potential to be beneficial in the future.

wrinkly spreader (WS) (Figure 1.1) is the most abundant of these types and forms a mat at the air-liquid interface. The smooth type (SM) is similar to the *P. fluorescens* ancestor and lives in the liquid phase, while the fuzzy spreader (FS) colonizes the bottom of the microcosm. A single WS (the Large Spreading WS) has undergone detailed analysis determining its genetic and phenotypic causes (BANTINAKI *et al.* 2007; SPIERS *et al.* 2003; SPIERS *et al.* 2002; SPIERS and RAINEY 2005), these are described in detail in the coming sections.

1.4.1 COOPERATION AND CONFLICT IN WS MATS

The ability of the WS to successfully colonize the interface of liquid and air is due to cooperation between individual cells (RAINEY and RAINEY 2003); each produce large amounts of acetylated cellulose and a proteinaceous attachment factor which allows the cells to adhere to each other and the wall of the glass vial (SPIERS *et al.* 2003; SPIERS *et al.* 2002; SPIERS and RAINEY 2005). The production of these extracelluar materials is metabolically expensive; the growth rate of the LSWS is merely 80% of the ancestor in the microcosm broth phase (RAINEY and RAINEY 2003). However, living

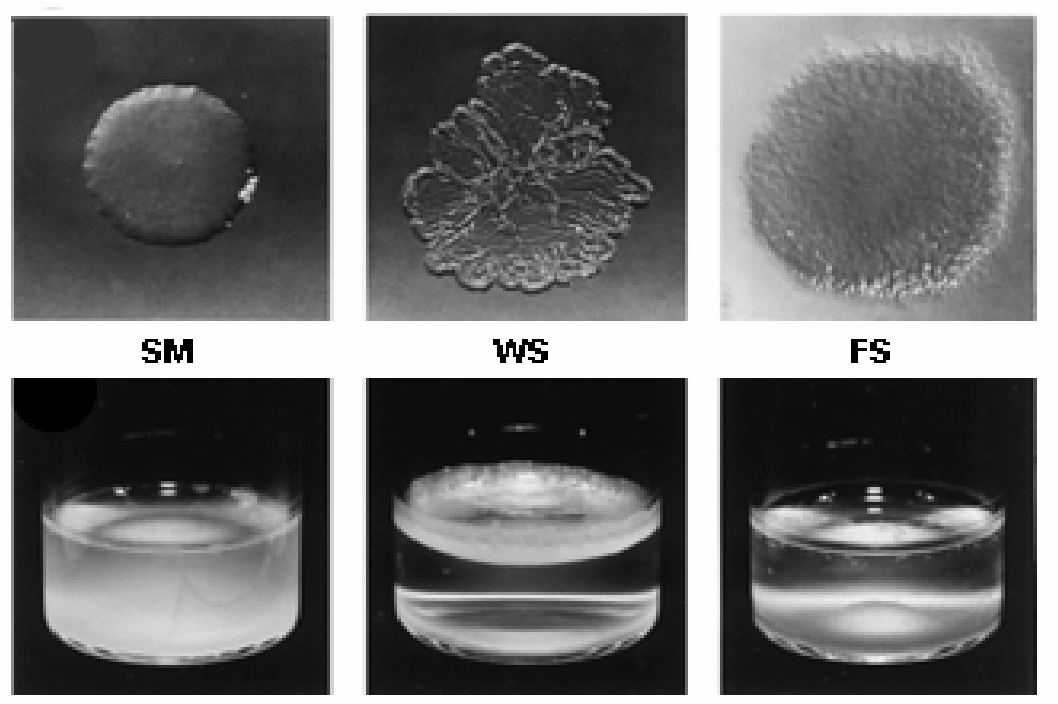


Figure 1.1 Colony morphology and niche specificity among evolved populations of *P. fluorescens*. Three distinct morphs were recognized after 72 hours incubation in a heterogenous environment: the smooth (SM), wrinkly spreader (WS) and the fuzzy spreader (FS). These grew distinctive colonies on agar plates (top row) and displayed individual niche specificities when cultured in microcosms containing King's medium B (bottom row) (Picture from (RAINEY and TRAVISANO 1998)).

at the air-liquid interface is advantageous because of the abundance of oxygen compared to the liquid column. It is this emergent group-level fitness benefit received by the mat which allows individually less fit WS to outcompete the ancestral SM in the static microcosm.

Cooperative groups of cells are thought to be the first step in the transition from single-celled to multicelluar organisms (PFEIFFER and BONHOEFFER 2003). However, groups are susceptible to the invasion of cheats, individual cells that don't pay the cost of cooperation yet benefit from living in the group. Cheats have a high individual fitness relative to cooperators, and increase in frequency within the group. Accordingly, conflict has been shown to arise in WS mats (RAINEY and RAINEY 2003). Non-cellulose

producing cheating types arise by mutation from mat-forming WS and quickly increase in frequency due to their individual selective advantage (RAINEY and RAINEY 2003). The strength of mats containing cheats was measured and compared to mats containing only WS. Cheat infiltrated mats (24% cheaters) collapse under less than a quarter of the weight that a mat composed only of WS can support (RAINEY and RAINEY 2003). Theory predicts that for the cohesive evolution of groups, conflict mediators must evolve to suppress the destructive effect of cheats. The WS model is unique in that it provides one of the only systems for the study of the evolution of cooperation *de novo*, and also because the genetic and phenotypic changes of WS's have been well characterised. In Chapter five of this thesis I conduct a group-selection experiment to investigate WS mat adaptation to the spatially structured environment. The evolution of differentiated cell types in WS mats is observed, and a novel strategy for the suppression of cheats developed and demonstrated. This work examines a different aspect of WS adaptation to chapters two, three and four, as such, a full introduction to this work is provided at the beginning of chapter five.

1.4.2 CAUSES OF THE WS PHENOTYPE

Transposon suppresser studies have identified several loci as necessary for the WS phenotype. The *wss* (wrinkly spreader structural locus) is a ten gene operon encoding biosynthetic machinery for the production of partially acetylated cellulose (SPIERS *et al.* 2003; SPIERS *et al.* 2002) (Figure 1.2). The cellulose produced by the proteins encoded at this locus is essential for the WS phenotype; when *wss* is deleted *P. fluorescens* is unable to form a robust mat at the air-liquid interface, and has a much reduced fitness in the static microcosm. However, sequencing of *wss* in WS revealed no mutations.

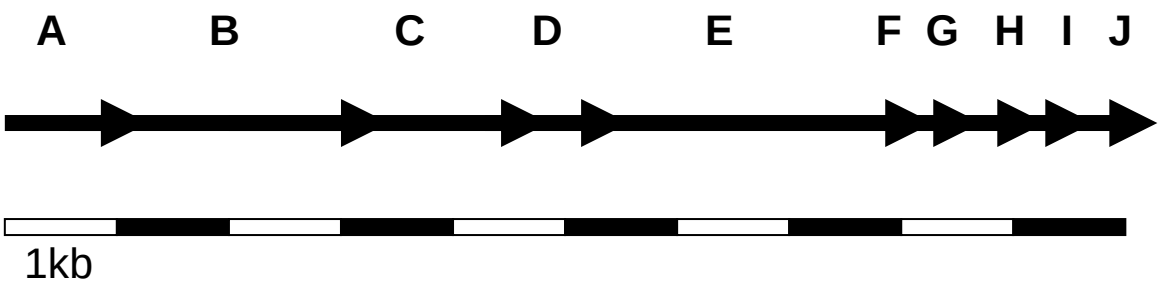


Figure 1.2

Diagrammatic representation of the *wss* operon. The *wss* operon is composed of ten genes, *wssA-J*. The genes *wssB, wssC, wssD* and *wssE* encode the cellulose synthase complex, while *wssF, wssG, wssH* and *wssI* are involved in acetylation of the polymer produced by this complex. The putative function of *wssA* and *wssJ* is for the localization of *wss* gene products to the cell poles have not yet been proven.

The *wsp* (wrinkly spreader phenotype locus) was the first locus implicated as sustaining the mutation which caused the LSWS. A second operon, aws (alternative wrinkly spreader), was found to possess the WS causing mutation in another independantly obtained WS (GEHRIG 2005). The discovery of mutations in these operons, both of which encode di-guanylate cyclase (DGC) enzymes, herald a pattern of causality for the WS phenotype. In this section I describe in detail the *wsp* and *aws* loci and DGC enzymes; the over activation of which confers the ability to colonise the air-liquid interface.

1.4.2.1 THE *WSP* LOCUS.

Functional analysis has shown the *wsp* pathway to be mechanistically similar to the *E. coli* Che pathway, which controls flagellar activity. Deletion of the *wsp* operon has no effect on cell motility in SBW25, and the role of *wsp* in the ancestral cell remains unknown. The model for *wsp* chemosensation puts WspA as a MCP type sensory protein which detects a signal subsequently translocated to kinase WspE which then phosphorylates WspR.

Activation of GGDEF domain protein WspR results in production of secondary signalling molecule, c-di-GMP. WspR, whose overactivation causes the LSWS, has been demonstrated to have DGC activity, and that this activity is elevated in the WS compared to the ancestral SM type (MALONE *et al.* 2007a).

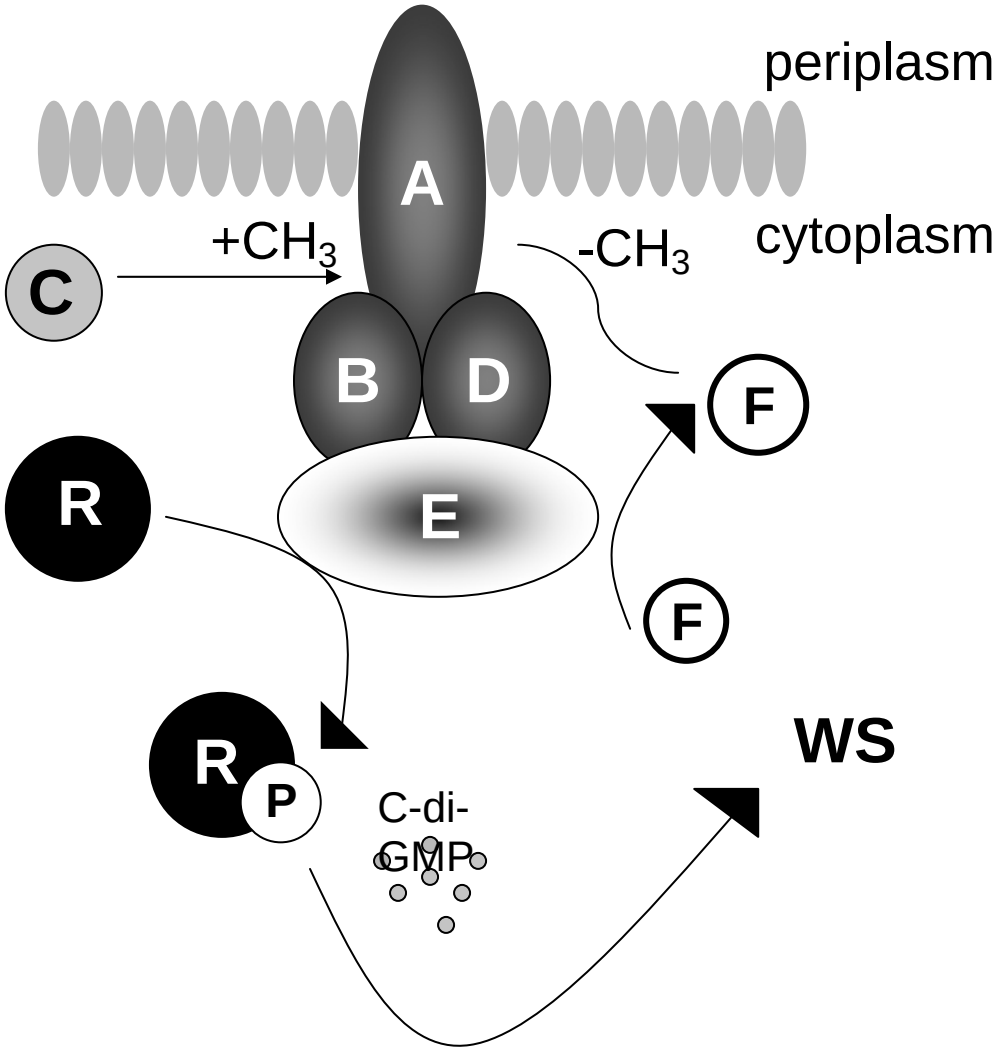


Figure 1.3

A model of the Wsp chemosensory pathway. An unidentified environmental signal is bound by membrane embedded WspA, this induces a conformational change which is translocated through WspC and WspD to kinase WspE. WspE then phosphorylates WspR. WspF and WspB modulate kinase activity of WspA by methylation. Inactivation of WspF leads to constitutive WspR activity and production of large amounts of c-di-GMP, the proximal cause of the WS phenotype.

P= inorganic phosphate.

1.4.2.2 DGC PROTEINS AND C-DI-GMP AS A SECONDARY MESSENGER.

Bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) has long been known as an activator of cellulose synthase in *Gluconacetobacter xylinus*(ROSS *et al.* 1987). Recently it has been recognized as a molecule ubiquitous to bacteria, that acts as a global secondary messenger mediating changes between sessile and motile states (D'ARGENIO and MILLER 2004; ROMLING and AMIKAM 2006; ROMLING *et al.* 2005; SIMM *et al.* 2004). C-di-GMP is synthesized by formation of phosphodiester bonds between the 5' and 3' ends of two GMP molecules. This reaction has been shown to be catalyzed by diguanylate cyclases (DGC), enzymes characterized by the GGDEF motif (RYJENKOV *et al.* 2005). GGDEF domains usually comprise part of two component response regulator proteins along with one or more of a limited repertoire of recognized motifs such as an environmental sensor (PAS, GAF, MCP), a signal transducer (HAMP) and membrane spanning domains, depending on the role of the protein in the cell (GALPERIN 2006). Structural studies of proteins containing GGDEF domains have determined the five amino acids to comprise half the catalytic site for cyclisation of c-di-GMP, suggesting that GGDEF proteins must dimerize to synthesize c-di-GMP (CHAN *et al.* 2004).

Levels of c-di-GMP in the cell are tightly regulated (ROMLING *et al.* 2005). The EAL domain has c-di-GMP phosphodiesterase (PDE) activity (Schmidt, *et al.*, 2005), with competing DGC and PDE activity controlling cellular c-di-GMP levels. DGC containing proteins often feature PDE domains, indicating that some proteins may both synthesize and degrade c-di-GMP. However, recent bioinformatic analysis suggests that the PDE domains of these seemingly bi-functional enzymes may be pseudo-domains in a

state of degradation (SCHMIDT *et al.* 2005), indeed, no bifunctional proteins have been described. In Chapter two I present the discovery of a DGC/PDE domain protein, and implicate it in both the synthesis and degradation of c-di-GMP.

1.4.3 THE GENETIC CAUSES OF WS.

1.4.3.1 THE *WSPF* LOCUS

The mutation that results in the LSWS has been identified as a single non-synonymous nucleotide change in *wspF* (BANTINAKI *et al.* 2007). WspF is a negative regulator of the LSWS phenotype (figure 1.3); an engineered *wspF* deletion in the ancestor resulted in WS, and replacement of the mutant *wspF* allele with the wild type allele, restored the ancestral phenotype (BANTINAKI *et al.* 2007; MALONE *et al.* 2007b). Together, these data are conclusive that a *wspF* SNP is the sole mutational cause of the LSWS. The point mutation is postulated to abolish WspF methylesterase activity, resulting in an unrestrained kinase (WspE), and therefore a constitutively active WspR. The resultant levels of c-di-GMP increases the production of cellulose directly as an allosteric activator of cellulose synthase, and also through increased *wss* transcription (GOYMER *et al.* 2006). C-di-GMP also influences the production of the proteinaceous attachment factor, whether directly or indirectly has not been determined.

Mutations in *wspF* are not the only possible WS causal mutation (BANTINAKI *et al.* 2007). A collection of 26 WS isolates were obtained and designated Independent WS_{A-Z}. Only half of these WS have a mutation in *wspF*, suggesting that there may be an alternative route to WS.

1.4.3.2 THE *AWS* OPERON.

WspR was shown to be necessary for the WS phenotype. However, when the *wsp* deletion strain (phenotypically SM) was inoculated into a static microcosm, it retained the ability to quickly evolve (a *wsp* independent) WS phenotype. The Alternative Wrinkly Spreader (AWS) was selected for using transposon mutagenesis (GEHRIG 2005), resulting in discovery of the *aws* locus. AwsX is a conserved hypothetical protein with a likely signal sequence, indicating that it is translocated across the cytoplasmic membrane, and performs its function in the periplasm. AwsR is a protein that contains two membrane spanning regions, a HAMP domain (signal transduction) and a GGDEF domain, suggesting that overproduction of c-di-GMP may be the cause of the AWS. AwsO shares homology with outer membrane proteins such as OmpA and MotB, specifically their membrane spanning regions; suggesting it is bound to the outer membrane and somehow facilitates transmission of a signal or molecule to AwsXR in the periplasm.

The entire *aws* locus was sequenced and the mutation found to be a 39bp deletion between two repeat sequences. The 26 unique WS isolates (WS_{A-z}) were also sequenced at this locus and a deletion of 33bp was found in WS_T, slightly down stream of the AWS deletion, also between two repeats and in-frame. The presence of an *awsX* deletion in this unique WS confirmed that mutations in the AWS locus represent an alternative route to the WS phenotype. The sequencing of the 26 unique WS at *wsp* and *aws* could not account for all of the WS causal mutations, confirming that there are further alternative genetic routes to the WS phenotype, and that they are probably DGC encoding loci. In chapter two of this thesis I describe the systematic search for these alternative genetic routes, and find that despite there being many possible loci that could sustain WS causal mutations, only three do. I provide evidence that these loci are more likely to sustain WS mutations because the genetic architecture of *P. fluorescens* biases the

production of genetic variation. The results generated in chapter two pave the way for experiments exploring previously experimentally intractable aspects of genetical evolutionary theory. In the next section I introduce the work of R. A. Fisher and the developments leading to our current understanding of the genetics of adaptation. Finally, I introduce the outstanding deficiencies in our current understanding of the genetics of adaptation to be addressed in this thesis.

1.5 THE GENETICAL THEORY OF EVOLUTION

In *The Genetical theory of Natural Selection* (FISHER 1930) R.. A. Fisher expounded the mathematical basis for the increase in frequency of alleles in populations and introduced his geometric model of adaptation. This model considered an adaptive walk, which is the series of mutational steps taken by a population returning to optimal fitness after displacement by environmental change. In a well known diagrammatic representation of his model the optimal phenotype is represented as the 0,0 intersection of cartesian axes (figure 1.4). The genotype currently fixed in the population has been displaced from this position, and each possible mutation alters fitness randomly with respect to direction and magnitude. The question asked of the model is, of all the possible mutations that could occur, what is the expected effect size of the beneficial mutations that will eventually fix in a population?

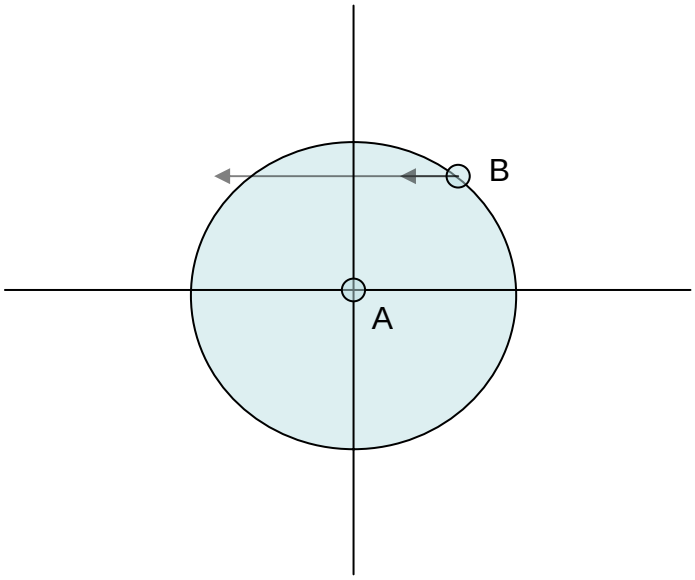


Figure 1.4 A population currently is fixed for genotype of fitness B, which has been displaced from optimal fitness by recent environmental change. Mutations of random effect size and direction will occur and only those that increase fitness (into the area indicated by the large circle) by moving closer to A will go to fixation. Fisher reasoned that for mutations of the same direction small ones (short arrow) were less likely to have a deleterious effect than large (long arrow).

He reasoned that because mutations of large effect are more likely to be harmful in complex organisms, small increments in fitness were those most likely to contribute to adaptive evolution. This lead to the conclusion that the probability of fixation for a new allele falls rapidly as the size of its fitness effect increases (Figure 1.5). As such, evolutionary change could only proceed by the gradual accumulation of tiny increments in fitness, a process referred to as micromutationism.

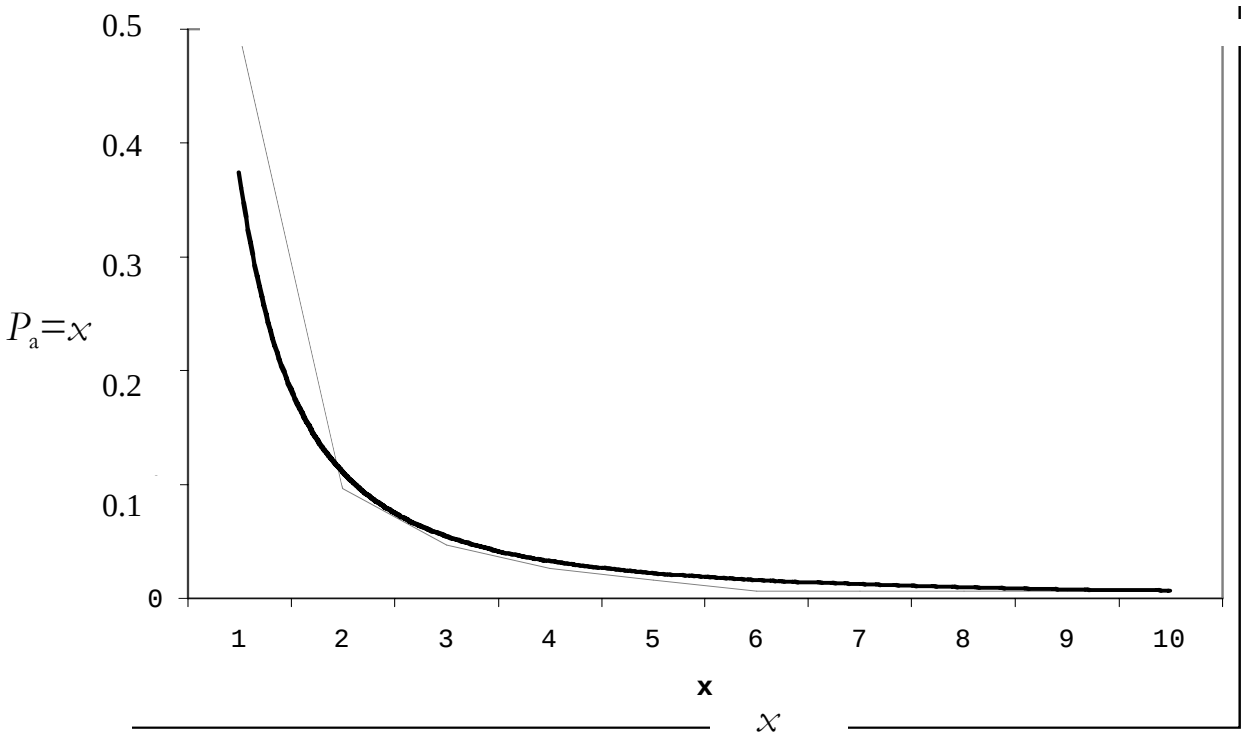


Figure 1.5 Fisher showed that the probability that a random mutation is beneficial ($P_a=x$) falls rapidly as its effect size increases (larger value for x). He concluded that only mutations of small effect size will be important during adaptation.

The key to Fisher's analysis was his applying to organisms the principles of statistical mechanics. He assumed that each trait is caused by an infinite number of genes, each without epistatic[8] interactions and having infintisimally small effects on the phenotype. Fisher saw that by averaging out the varying effects that an allele could have in all possible genetic backgrounds, a value could be assigned to that allele; the effects of all

[8] Epistasis results from the action of one gene being modified by one or several other genes, causing alleles to interact multiplicatively. Fisher assumed that a genes effect on phenotype was independent of the genetic background in which it occurred.

alleles within a single genotype could be combined additively to give the overall organismal fitness.

Fisher's outstanding contribution to the genetics of evolution not only laid the foundation of evolutionary genetics, it also defined the questions asked by those working within the field for a long time. Classical evolutionary genetics defined adaptation as the change of allele frequencies in a population, a mathematically convenient but biologically unrealistic assertion (MAYR 1988). In the next section the developments leading to the eventual rejection of micromutationism and conception of the mutational landscape model of adaptation are presented.

1.5.1 THE DISTRIBUTION OF FITNESS EFFECTS OF NEW MUTATIONS

In one of his many insights, the theoretical geneticist Motoo Kimura (KIMURA 1983) realised that because of the greater selective advantage of beneficial mutation of large effect, they would be more likely to go to fixation than the mutations of small effect favoured in Fisher's model. He calculated that the probability of a beneficial genotype going to fixation was 2 times the selection coefficient (s) of the allele under consideration. However, even if selection strongly favours beneficial mutations, if they are vanishingly rare they are unlikely to play a role in adaptation.

The size of steps taken during an adaptive walk is therefore also dependant on the probability distribution of new mutations from which the fitness effects of beneficial mutations are drawn, known as the distribution of fitness effects (DFE). The x axis of such a distribution shows the fitness values of the discrete set of all possible single step mutations (each resulting in a new possible genotype) that could occur in the current genotypic background. These are presented at random (assuming mutation is random) to natural selection; the frequency of genotypes of any given fitness value is shown on the y

axis. Knowledge of this distribution would be desirable as it could potentially allow the prediction of the size and number of steps taken by an organism during an adaptive walk. Unfortunately, very little is known of the DFE, not even if it varies between species or even within genomes for different environments (ORR 2005; ORR 2006).

Empirical studies measuring the distribution of fitness effects are rare as it is extremely difficult to obtain an unbiased sample of all the mutations presented to natural selection. Most methods for collecting mutations are more likely to sample beneficial mutations of large effect size. The probability that a new allele is not lost from the population depends on its fitness effect (Haldane, 1927). As such deleterious and weakly beneficial mutations are drastically underrepresented in any sample which has experienced selection at all, and the measured distribution will not reflect that presented to natural selection. An alternative to sampling naturally occurring mutations is to construct libraries of mutants created by random mutagenesis and measure the fitness effects of each of these mutations. Such an approach using an *E. coli* transposon mutagenesis library (ELENA *et al.* 1998) revealed the DFE to be best fit by a compound distribution comprised of both gamma and uniform components. Transposon mutations may not provide an accurate approximation of natural mutation, so the relevance of these results to adaptation in real organisms is questionable. In a more realistic approximation of natural mutation, fitness measurements were made on a library of systematically generated single nucleotide polymorphic genotypes in the vesicular stomatis RNA virus (SANJUAN *et al.* 2004). Here the distribution was found to be lognormal; this distribution is known to arise in biological systems where a given variable is characterised by many small multiplicative effects. The DFE is thought to be impossible to determine for all organisms; however this result supports that the DFE can arise by general statistical effects, at least in a very simple organism.

1.5.2 THE MUTATIONAL LANDSCAPE MODEL OF ADAPTATION

Detemining the distribution of fitness effects of new mutations is perceived as an almost experimentally intractable problem (ORR 2003; ORR 2005). Even if knowledge of this distribution was garnered for a single organism, it might well be different for another. GILLESPIE (1984) reasoned that precise knowledge of the nature of this distribution was not necessary. Environmental change is gradual; as such most organisms are already well adapted to their environment. Following a change in conditions, the most common genotype in the population, although relatively fit, is no longer the fittest possible. The ensuing adaptive walk is the process of a new fitter genotype being substituted in the population, followed by another, then another, until the fittest possible genotype is realised. If the adaptive walk starts with a high fitness genotype, the substituting genotypes will be drawn from the right hand tail of the overall distribution of fitness effects (figure 1.6).

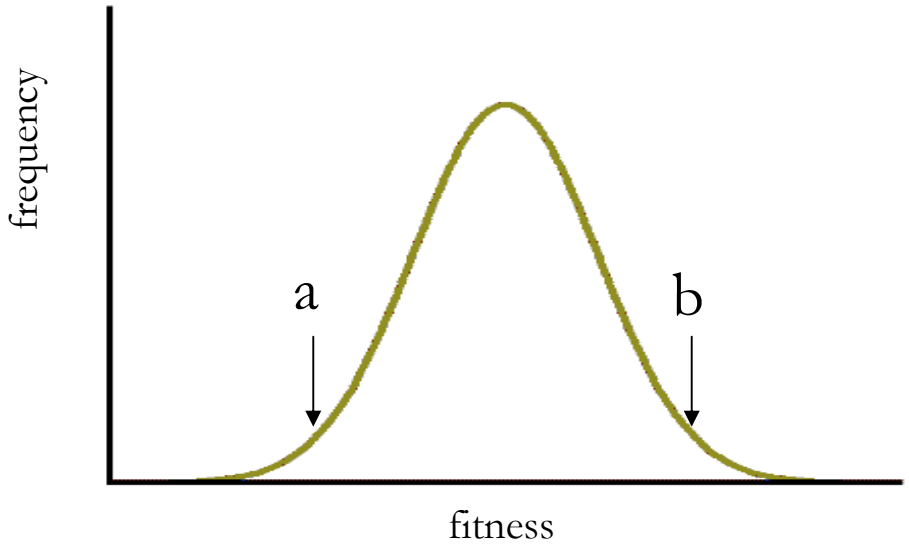


Figure 1.6 The distribution of fitness effects. Following a large environmental change the most common genotype in the population has a low fitness (a). The mutational landscape model assumes that environmental change is usually small, as such the most common genotype in the population at the start of an adaptive walk falls within the right hand tail of the distribution of fitness effects (b). All beneficial mutations therefore are drawn from the the exponentially distributed right hand tail of the distribution.

The right hand tails of many common distributions are exponential; if the distribution of beneficial mutations (those mutations more fit than the ancestor) is exponential then certain properties of mutations within this tail can be described by extreme value theory (EVT)[9], independent of the nature of the overall distribution itself. The model works within parameters determined by the conditions of strong selection weak mutation (SSWM) (MAYNARD-SMITH, 1963). Because strong selection can distinguish between small differences in fitness effects, mutations are either beneficial or deleterious. Weak mutation means that double mutants are too rare to be considered, shrinking the range of explorable genotype space to a discrete set of genotypes. This is an important

[9] Extreme value theory was developed to model extreme deviations from the median of probability distributions. Exponential tails are a property of distributions of the Gumbel type, such as the normal, gamma and Weibull distributions.

difference to Fisher's model in which adaptation occurred in an infinite, continous phenotype space. During an adaptive walk, a sequence of length L mutates recurrently to one of 3L sequences[10]; most mutations are deleterious and are quickly lost from the population. Eventually, a rare beneficial mutation will arise; these to may be lost until eventually one is fixed in the population. This new genotype may then mutate into a new set of 3L genotypes, and the process repeats itself until a genotype is fixed whose set of 3L genotypes contains none fitter than itself.

1.5.2.1 THE MUTATIONAL LANDSCAPE MODEL AND FITNESS LANDSCAPES.

Fitness landscapes are a useful tool for visualising the process of adaptation (WRIGHT 1932b). All possible genotypes are connected in a vast network called genotype space, connections are formed between genotypes that differ at a single nucleotide. Under assumptions of SSWM, each adaptive step can only consider genotypes connected to the current genotype, i.e, that require a single mutation to be realised, other genotypes are not considered as they require an unlikely two or more mutations. Each genotype has a fitness assigned to it; highly fit genotypes are depicted as mountains or peaks; unfit genotypes as valleys. An adaptive walk continues until a genotype is reached that cannot mutate to a more fit genotype, this genotype is at an adaptive peak. A single-peaked landscape is also known as correlated or smooth. In such a landscape, neighbouring genotypes have similar (correlated) fitness values, and adaptation will lead eventually to the attainment of the peak. A multipeaked landscape is called uncorrelated or rugged, and is the complete opposite to smooth landscapes. Characterised by multiple low peaks, the fitness value of each genotype is unrelated (uncorrelated) to that of its neighbours.

[10] Each nucleotide position in the sequence of length L may mutate to one of three alternative nucleotides. Insertions, deletions or double mutations are not considered.

An adaptive walk will lead to the attainment of one of the many peaks, although it may not be the highest possible. The mutational landscape model assumes that each step of the adaptive walk draws from the same distributions of mutational effects, as such the adaptive landscape is highly uncorrelated.

1.5.2.2 EXTENSIONS AND TESTS OF THE MUTATIONAL LANDSCAPE MODEL

Gillispie used his model to generate several important results, firstly, he found that the probability that any given mutation would be the next substituted in an adaptive walk is proportional to its selective advantage. Later he showed that the average number of steps taken during an adaptive walk was two to five steps, indicating that adaptation occurs in bursts. This model was extended by Orr, who examined the size of the average fitness jump taken by the population with each step of the adaptive walk. It was found to be suprisingly large, but diminished in actual size with each new fixed mutation. These results generated predictions, tested explicitly by several experimental studies. Rokyta (ROKYTA *et al.* 2005) sequenced 20 viral populations which had fixed a single mutation. Fitnesses were compared to see if the fitter mutations were the ones most often substituted. It was found that once transition/transversion mutation bias had been accounted for, the distribution of mutations did indeed fit Orr's predictions. Later studies have provided both confirmatory (KASSEN and BATAILLON 2006; ROKYTA *et al.* 2005) and dissenting results (ROKYTA *et al.* 2008).

In Chapter three of this thesis I collect a sample of beneficial mutations that have not experienced natural selection, providing the opportunity to determine the shape of the distribution of beneficial mutations. However, WS adaptation violates a central assumption of the mutational landscape model; the fitness of the ancestral (SM) type is not high, in fact, because it cannot colonise the air-liquid interface, it is effectively zero. I

exploit this fact and show that this set of mutations is representative of the entire DFE, not just beneficial mutations. I measure the shape of this distribution and then discuss the possibility of a dramatically expanded mutational landscpae model using the central limit theorem as opposed to EVT.

1.5.3 GENETIC ARCHITECTURE AND THE GENOTYPE-PHENOTYPE MAP

The models of adaptation described above have had some success explaining the mountain of molecular adaptation data accumulated in the last 20 to 30 years. Some generalities are emerging; the pattern of diminishing returns of beneficial mutations during an adaptive walk was predicted by both Kimura's modification of Fisher's model, the mutational landscape model and is supported by experimental results (COOPER and LENSKI 2000; DE VISSER and LENSKI 2002; HOLDER and BULL 2001; LENSKI and TRAVISANO 1994; NOVELLA *et al.* 1995). However, despite their utility, in order to generate these models the interactions and subtleties within genomes are averaged out, leading to many evolutionary relevent phenomena being overlooked (HANSEN 2006). Our detailed knowledge of biochemical pathways and organised regulation networks tells us that genomes are not homogenous tracts of genes; this raises the possibility that genes and genomes may also vary in their propensity to generate adaptive variation. Such differences could drastically influence evolutionary outcomes.

As defined by Hansen (2006), genetic architecture refers to the pattern of genetic effects that build and control a given phenotypic character and its variational properties. This can be studied by investigating relationship between genotype and phenotype. Specifically, what is required is an understanding of how a mutation alters the function or expression of a protein and then how this altered protein influences cellular processes and structures in such a way that influences fitness (PIGLIUCCI and PRESTON 2004). With

this knowledge comes the understanding of how characters vary in their capacity to produce variation, and thus their ability to evolve and adapt (SCHLICHTING and MURREN 2004). The capacity to produce variation is in itself a trait which can evolve, either towards robustness or evolvability. The difference between these is illustrated on the genotype-phenotype map shown below (figure 1.7).

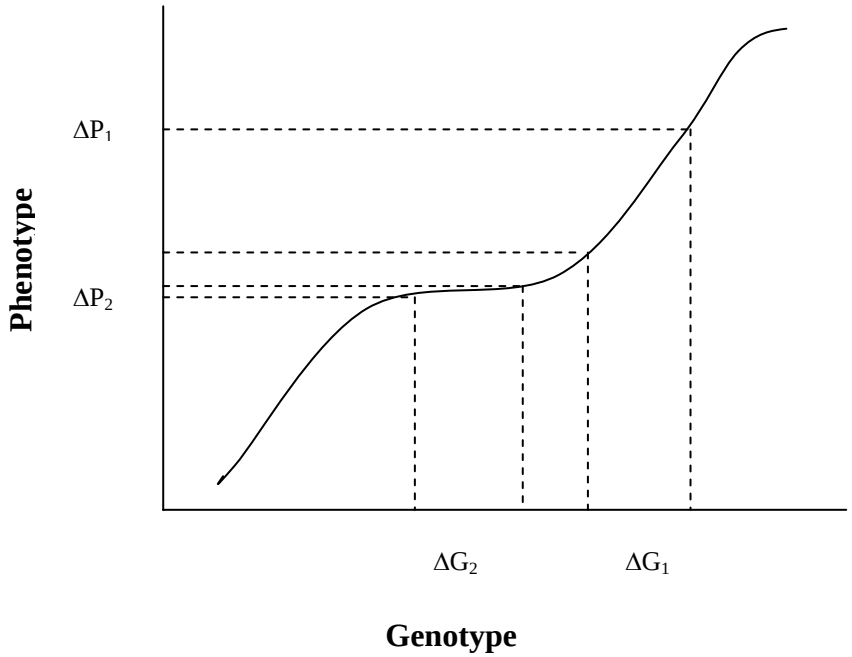


Figure 1.7 The mapping of genetic change to phenotypic change shows how the same magnitude of change in genotype can lead to different magnitudes in the change in phenotype. In the flat (robust) region genetic change has little effect on the phenotype, however in the steep (evolvable) region the same degree of genetic change has a much larger effect on phenotype.

Figure 1.7 shows how differences in the genotype to phenotype map can result in regions of genotype space whose composite genotypes vary in their propensity to convert mutation into phenotypic change. For instance, a genotype found in region 1 has the ability to produce a large amount of phenotypic change per unit of genetic change; we may describe such a genotype as evolvable. Conversely, a genotype in region 2 may sustain the equivalent amount of genetic change without perturbing the phenotype; this

genotype is described as robust. Evolvability and robustness are the two extremes of a continuum, contrasting selective scenarios exist under which either robust or evolvable genotypes are desirable. During periods of prolonged environental stability, once the optimal genotype has been achieved, phenotypic perturbation by mutation can only decrease fitness. In this case a genotype robust to genetic change is predicted to be favoured by natural selection. Conversely, in a frequently changing environment phenotypic variation is desirable, as such evolvable genotypes are those favoured by natural selection (MEYERS *et al.* 2005; PROULX and PHILLIPS 2005; SCHMALHAUSEN 1949; WADDINGTON 1942).

There are several aspects of genetic architecture that influence the degree of evolvability of genomes. Central to evolvability is the production of genetic variation, the ultimate source of which is mutation. The per nucleotide mutation rate can respond to natural selection and settles at a balance between two contrasting selective forces- one to reduce the load of deleterious mutations, the other to increase the supply of variation. Accordingly, increases in mutation rate have been shown to evolve in response to fluctuating conditions (PAL *et al.* 2007), and mutation rates vary widely across genomes and genes (DRAKE *et al.* 1996). However, there is no point having lots of variation if it is not useful. As genomes increase in complexity, mutations that have a beneficial influence on one trait become more likely to have a deleterious effect on the organism as a whole (RAFF and SLY 2000; WAGNER and ALTENBERG 1996; WAGNER *et al.* 2008) but also see[11] (GRISWOLD 2006; HANSEN 2003) for a more subtle perspective. This cost of complexity is hypothesised to be amerliorated by the division of the genome into semi autonomous regions, or modules. Modules contain many internal connections, but few external connections (to other modules). This means that a mutation has a large effect on

[11] These authors argue that strict modularity reduces the variational potential of characters, and that selection for evolvability may lead a genome towards an optimal, lesser degree of modularity.

its own module, without perturbing others. There is mounting evidence that genomes are divided into modules (IHMELS *et al.* 2002; KOMUROV and WHITE 2007; KORCSMAROS *et al.* 2007), although just how many an organism may have or their degree of interconnectedness remain experimentally elusive.

1.5.4 GENETIC ARCHITECTURE AND CONSTRAINTS ON EVOLUTION

Evolutionary constraint is an ambiguously defined topic (ANTONOVICS and VANTIENDEREN 1991). In an early and important review arising from a conference on this topic John Maynard-Smith and others exhaustively listed the possible constraints on evolution (MAYNARD-SMITH ET AL., 1985). They claim there are two types of constraints, absolute and relative. Absolute constraints are the limits imposed by the laws of physics on what can possibly evolve. The limitations imposed on bacterial cell size by the rate of diffusion is an absolute constraint; a cell can only become so big without internal reticulation. Relative constraints are limits imposed on what can possibly evolve in a realistic amount of time, and are interesting to evolutionary biologists because the degree to which they influence evolution is unknown. One possible constraint on evolution is the genetic architecture of an organism. If there are limits to the amount of variation that an organism can produce, this could limit the potential of the species to adapt to and survive environmental change. The plausibility of this argument has been confirmed by analyses of quantitative trait loci (FUTUYMA *et al.* 1995; FUTUYMA *et al.* 1993). Alternatively, if an organism has many potential genetic routes to an adaptive phenotype some of these routes may be more likely to be realised than others (SCHLUTER 1996). Theoretical explorations of the role in genetic constraints in evolution are few in number (BRAKEFIELD 2006; BRAKEFIELD and ROSKAM 2006; MAYNARD-SMITH 1985; SCHLUTER *et al.* 2004), and empirical studies are almost non-existent. Sequence data from repeated

evolutionary events are required to understand the role of chance and genetics on an organisms evolution (BRAKEFIELD 2006; SCHLUTER *et al.* 2004). Chapter four utilises knowledge of the likely location of causal mutations gained in chapter two. Many microcosms are propogated in parallel for varying lengths of time, then sequenced to find the causal mutations. In this study I conclude that evolutionary outcomes depend on a complex interaction of variable mutation rates, the variable ability of genes to translate mutation in ecologically useful variation and the fitness of individual genotypes.

1.6 SUMMARY

The model organism *P. fluorescens* SBW25 provides a unique system with which to study evolutionary phenomena. An array of genetic tools, including an annotated genome sequence (SILBY *et al.* 2008), facilitate the tracking down of the genetic causes of evolution. A substantial body of experimental evidence underlies our understanding of the physiological changes that are required for the WS phenotype. Moreover, the distinct morphological and phenotypic differences between ancestor and WS descendant facilitate the generalisation of discoveries made using WS to evolutionary phenomena more often associated with the animal and plant kingdoms. In this thesis I seek to elucidate the as yet undiscovered genetic routes to the WS phenotype, and then extend the WS system to empirically address important unanswered questions in evolutionary biology. This thesis consists of four results chapters, chapter two has been submitted as a paper to the journal Genetics and has been bound in this thesis as such.

1.7 RESEARCH OBJECTIVES

1. To further characterise the *aws* locus so that the regulatory relationship between AwsX and AwsR is understood. To create an Δ*aws*Δ*wsp* strain and evolve a WS, with the goal of finding all possible alternative genetic routes to the WS phenotype. The results pertaining to these aims are presented in chapter two, which has been submitted to the journal Genetics.

2. To characterise the regulatory mechanisms of any further genetic loci sustaining WS causal mutations, with the aim of understanding how mutation translates into phenotype. The results pertaining to this aim are also included in chapter two.

3. To use the WS system to explore the distribution of mutational effects, with the goal of expanding the mutational landscape model of adaptation. This is presented in Chapter three.

4. To use the extensive genetic parallelism in the 26 independent WS to explore the roles of genetic architecture, mutation rate and natural selection in evolutionary outcomes. The results of this study are presented in this thesis as chapter four.

5. To design and implement an experiment to push WS mat evolution towards increased group level complexity. Chapter five contains the results pertaining to this aim.

1.8 BIBLIOGRAPHY

ANTONOVICS, J., and P. H. VANTIENDEREN, 1991 Ontoecogenophyloconstraints - the chaos of constraint terminology. Trends in Ecology & Evolution **6:** 166-168.

BANTINAKI, E., R. KASSEN, C. G. KNIGHT, Z. ROBINSON, A. J. SPIERS *et al.*, 2007 Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. Genetics **176:** 441-453.

BLOUNT, Z. D., C. Z. BORLAND and R. E. LENSKI, 2008 Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America **105:** 7899-7906.

BRAKEFIELD, P. M., 2006 Evo-devo and constraints on selection. Trends in Ecology & Evolution **21:** 362-368.

BRAKEFIELD, P. M., and J. C. ROSKAM, 2006 Exploring evolutionary constraints is a task for an integrative evolutionary biology. American Naturalist **168:** S4-S13.

CHAN, C., R. PAUL, D. SAMORAY, N. C. AMIOT, B. GIESE *et al.*, 2004 Structural basis of activity and allosteric control of diguanylate cyclase. Proceedings of the National Academy of Sciences of the United States of America **101:** 17084-17089.

COLEGRAVE, N., and A. BUCKLING, 2005 Microbial experiments on adaptive landscapes. Bioessays **27:** 1167-1173.

COOPER, V. S., and R. E. LENSKI, 2000 The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature **407:** 736-739.

D'ARGENIO, D. A., and S. I. MILLER, 2004 Cyclic di-GMP as a bacterial second messenger. Microbiology-Sgm **150:** 2497-2502.

DALLINGER, W. H., 1887 The Presidents Address. Journal of Royal Microscopic Society **10:** 191-192.

DARWIN, C., 1859 *On the Origin of Species*. John Murray, London.

DE VISSER, J., and R. E. LENSKI, 2002 Long-term experimental evolution in *Escherichia coli.* XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. BMC Evolutionary Biology **2:** 19.

DE VISSER, J., C. W. ZEYL, P. J. GERRISH, J. L. BLANCHARD and R. E. LENSKI, 1999 Diminishing returns from mutation supply rate in asexual populations. Science **283:** 404-406.

DOBZHANSKY, T. H., 1937 *Genetics and the Origins of Species*. Columbia University Press, New York.

DRAKE, J., B. CHARLESWORTH, D. CHARLESWORTH and J. F. CROW, 1996 Rates of spontaneous mutation. Genetics **148:** 1667-1686.

ENGSTROM, G., L. E. LILJEDAHL and T. BJORKLUND, 1992 Expression of expression of genetic and environmental variation during ageing. II. Selection for increased life span in *Drosophila melanogaster* Theoretical and Applied Genetics **85:** 26-32.

FISHER, R. A., 1918 The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh **52:** 399-433.

FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.

FUTUYMA, D. J., M. C. KEESE and D. J. FUNK, 1995 Genetic constraints on macroevolution - the evolution of host affiliation in the leaf beetle genus *Ophraella*. Evolution **49:** 797-809.

FUTUYMA, D. J., M. C. KEESE and S. J. SCHEFFER, 1993 Genetic constraints and the phylogeny of insect-plant associations - responses of *Ophraella communa* (*Coleoptera, Chrysomelidae*) to host plants of its congeners. Evolution **47:** 888-905.

GALPERIN, M. Y., 2006 Structural classification of bacterial response regulators: diversity of output domains and domain combinations. Journal of Bacteriology **188:** 4169-4182.

GEHRIG, S. M., 2005 Adaptation of *Pseudomonas fluorescens* SBW25 to the air-liquid interface: a study in evolutionary genetics. Ph.D Thesis. Oxford.

GIRARD, I., M. W. MCALEER, J. S. RHODES and T. GARLAND, 2001 Selection for high voluntary wheel-running increases speed and intermittency in house mice (Mus domesticus). Journal of Experimental Biology **204:** 4311-4320.

GOODNIGHT, C. J., 2000 Heritability at the ecosystem level. Proceedings of the National Academy of Sciences of the United States of America **97:** 9365-9366.

GOYMER, P., S. G. KAHN, J. G. MALONE, S. M. GEHRIG, A. J. SPIERS *et al.*, 2006 Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. II. Role of the GGDEF regulator WspR in evolution and development of the wrinkly spreader phenotype. Genetics **173:** 515-526.

GRISWOLD, C. K., 2006 Pleiotropic mutation, modularity and evolvability. Evolution & Development **8:** 81-93.

HALDANE, J. B. S., 1932 *The Causes of Evolution*. Princeton University Press, Princeton, NJ.

HANSEN, T. F., 2003 Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. Biosystems **69:** 83-94.

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

HOLDER, K. K., and J. J. BULL, 2001 Profiles of adaptation in two similar viruses. Genetics **159:** 1393-1404.

HUXLEY, J., 1942 *Evolution: the Modern Synthesis*. Allen and Unwin, London.

IHMELS, J., G. FRIEDLANDER, S. BERGMANN, O. SARIG, Y. ZIV *et al.*, 2002 Revealing modular organisation in the yeast transcriptional network. Nature Genetics **31:** 370-377.

JESSUP, C. M., R. KASSEN, S. E. FORDE, B. KERR, A. BUCKLING *et al.*, 2004 Big questions, small worlds: microbial model systems in ecology. Trends in Ecology & Evolution **19:** 189-197.

KASSEN, R., and T. BATAILLON, 2006 Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nature Genetics **38:** 484-488.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

KOMUROV, K., and M. WHITE, 2007 Revealing static and dynamic modular architecture of the eukaryotic protein interaction network. Molecular systems biology **3**.

KORCSMAROS, T., I. A. KOVACS, M. S. SZALAY and P. CSERMELY, 2007 Molecular chaperones: The modular evolution of cellular networks. Journal of Biosciences **32:** 441-446.

LENSKI, R. E., M. R. ROSE, S. C. SIMPSON and S. C. TADLER, 1991 Long-term experimental evolution in *Escherichia-Coli*. I. Adaptation and divergence during 2,000 generations. American Naturalist **138:** 1315-1341.

LENSKI, R. E., and M. TRAVISANO, 1994 Dynamics of adaptation and diversification - A 10,000 generation experiment with bacterial populations. Proceedings of the National Academy of Sciences of the United States of America **91:** 6808-6814.

LENSKI, R. E., C. L. WINKWORTH and M. A. RILEY, 2003 Rates of DNA sequence evolution in experimental populations of Escherichia coli during 20,000 generations. Journal of Molecular Evolution **56:** 498-508.

LINTS, F. A., J. STOLL, G. GRUWEZ and C. V. LINTS, 1979 Attempt to select for increased longevity in *Drosophilia melanogaster* Gerontology **25:** 192-204.

LUCKINBILL, L. S., and M. J. CLARE, 1985 Selection for life span in *Drosophilia melanogaster* Heredity **55:** 9-18.

MALONE, J. G., R. WILLIAMS, M. CHRISTEN, U. JENAL, A. J. SPIERS *et al.*, 2007a The structure-function relationship of WspR, a response regulator with a GGDEF output domain. Microbiology-Sgm **153:** 980-994.

MALONE, J. G., R. WILLIAMS, M. CHRISTEN, U. JENAL, A. J. SPIERS *et al.*, 2007b The structure-function relationship of WspR, a Pseudomonas fluorescens response regulator with a GGDEF output domain. Microbiology-Sgm **153:** 980-994.

MAYNARD-SMITH, J., R. BURIAN, S. KAUFFMAN, P.ALBERCH, J. CAMPBELL, B. GOODWIN, R. LANDE, D. RAUP, AND L WOLPERT., 1985 Developmental constraints and evolution. Quarterly Review of Biology **60:** 265-287.

MAYR, E., 1942 *Systematics and the Origin of Species, from the veiw point of a Zoologist*. Harvard University Press, Cambridge.

MAYR, E., 1988 *Towards a New Philosophy of Biology: Observations of an Evolutionist*. Harvard University Press, Boston, MA.

MENDEL, G., 1865 Veruche ueber pflanzenhybride. Verhandlungen des naturforschenden vereines brunn **4:** 3-47.

MEYERS, L. A., F. D. ANCEL and M. LACHMANN, 2005 Evolution of genetic potential. Plos Computational Biology **1:** 236-243.

NOVELLA, I. S., E. A. DUARTE, S. F. ELENA, A. MOYA, E. DOMINGO *et al.*, 1995 Exponential increases of RNA virus fitness during large population transmissions. Proceedings of the National Academy of Sciences of the United States of America **92:** 5841-5844.

ORR, H. A., 2003 The distribution of fitness effects among beneficial mutations. Genetics **163:** 1519-1526.

ORR, H. A., 2005 The genetic theory of adaptation: a brief history. Nature Reviews Genetics **6:** 119-127.

ORR, H. A., 2006 The population genetics of adaptation on correlated fitness landscapes: The block model. Evolution **60:** 1113-1124.

PAL, C., M. D. MACIA, A. OLIVER, I. SCHACHAR and A. BUCKLING, 2007 Coevolution with viruses drives the evolution of bacterial mutation rates. Nature **450:** 1079-1081.

PFEIFFER, T., and S. BONHOEFFER, 2003 An evolutionary scenario for the transition to undifferentiated multicellularity. Proceedings of the National Academy of Sciences of the United States of America **100:** 1095-1098.

PIGLIUCCI, M., and K. PRESTON, 2004 *Phenotypic Integration: Studying the Ecology and Evolution of Genetic Networks*. Oxford University Press, New York.

PROULX, S. R., and P. C. PHILLIPS, 2005 The oppurtunity for canalization and the evolution of genetic networks. American Naturalist **165:** 147-162.

RAFF, R. A., and B. J. SLY, 2000 Modularity and dissociation in the evolution of gene expression territories in development. Evolution & Development **2:** 102-113.

RAINEY, P. B., 1999 Evolutionary genetics: The economics of mutation. Current Biology **9:** R371-R373.

RAINEY, P. B., and K. RAINEY, 2003 Evolution of cooperation and conflict in experimental bacterial populations. Nature **425:** 72-74.

RAINEY, P. B., and M. TRAVISANO, 1998 Adaptive radiation in a heterogeneous environment. Nature **394:** 69-72.

ROKYTA, D. R., C. J. BEISEL, P. JOYCE, M. T. FERRIS, C. L. BURCH *et al.*, 2008 Benefical fitness effects are not exponential for two viruses. Journal of Molecular Evolution **67:** 368-376.

ROKYTA, D. R., P. JOYCE, S. B. CAUDLE and H. A. WICHMAN, 2005 An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nature Genetics **37:** 441-444.

ROMLING, U., and D. AMIKAM, 2006 Cyclic di-GMP as a second messenger. Current Opinion in Microbiology **9:** 218-228.

ROMLING, U., M. GOMELSKY and M. Y. GALPERIN, 2005 C-di-GMP: the dawning of a novel bacterial signalling system. Molecular Microbiology **57:** 629-639.

ROSS, P., H. WEINHOUSE, Y. ALONI, D. MICHAELI, P. WEINBERGEROHANA *et al.*, 1987 Regulation of cellulose synthesis in *Acetobacter-Xylinum* by cyclic diguanylic acid. Nature **325:** 279-281.

RYJENKOV, D. A., M. TARUTINA, O. V. MOSKVIN and M. GOMELSKY, 2005 Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: Insights into biochemistry of the GGDEF protein domain. Journal of Bacteriology **187:** 1792-1798.

SANJUAN, R., A. MOYA and S. F. ELENA, 2004 The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proceedings of the National Academy of Sciences of the United States of America **101:** 8396-8401.

SCHLICHTING, C. D., and C. J. MURREN, 2004 Evolvability and the Raw Materials for Adaptation, pp. 18-29 in *Plant Adaptation: Molecular Genetics and Ecology*, edited by Q. C. B. CRONK. NRC Research Press, Vancouver.

SCHLUTER, D., 1996 Adaptive radiation along genetic lines of least resistance. Evolution **50:** 1766-1774.

SCHLUTER, D., E. A. CLIFFORD, M. NEMETHY and J. S. MCKINNON, 2004 Parallel evolution and inheritance of quantitative traits. American Naturalist **163:** 809-822.

SCHMALHAUSEN, I. I., 1949 *Factors of Evolution: The theory of stabilizing selection*. Blakiston, Philadelphia.

SCHMIDT, A. J., D. A. RYJENKOV and M. GOMELSKY, 2005 The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. Journal of Bacteriology **187:** 4774-4781.

SILBY, M. W., A. M. CERDEÑO-TÁRRAGA, G. VERNIKOS, S. R. GIDDENS, R. W. JACKSON *et al.*, 2008 Genomic and functional analyses of diversity and plant interactions of *Pseudomonas fluorescens*.

SIMM, R., M. MORR, A. KADER, M. NIMTZ and U. ROMLING, 2004 GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. Molecular Microbiology **53:** 1123-1134.

SIMPSON, G. G., 1944 *Tempo and Mode in Evolution*. Columbia University Press, New York.

SNIEGOWSKI, P. D., P. J. GERRISH and R. E. LENSKI, 1997 Evolution of high mutation rates in experimental populations of *E. coli*. Nature **387:** 703-705.

SPIERS, A. J., J. BOHANNON, S. M. GEHRIG and P. B. RAINEY, 2003 Biofilm formation at the air-liquid interface by the Pseudomonas fluorescens SBW25 wrinkly spreader requires an acetylated form of cellulose. Molecular Microbiology **50:** 15-27.

SPIERS, A. J., S. G. KAHN, J. BOHANNON, M. TRAVISANO and P. B. RAINEY, 2002 Adaptive divergence in experimental Populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. Genetics **161:** 33-46.

SPIERS, A. J., and P. B. RAINEY, 2005 The *Pseudomonas fluorescens* SBW25 wrinkly spreader biofilm requires attachment factor, cellulose fibre and LPS interactions to maintain strength and integrity. Microbiology-Sgm **151:** 2829-2839.

STEBBINS, G. L., 1950 *Variation and Evolution in Plants*. Columbia University Press, New York.

SWENSON, W., and D. S. WILSON, 2000 Artificial ecosystem selection. Proc. Natl. Acad. Sci. **97**.

TRAVISANO, M., and P. B. RAINEY, 2000 Studies of adaptive radiation using model microbial systems. American Naturalist **156:** S35-S44.

WADDINGTON, C. H., 1942 Canalization of development and the inheritance of acquired characters. Nature **150:** 563-565.

WADE, M. J., 1976 Group selection among laboratory populations of *Tribolium*. Proceedings of the National Academy of Sciences of the United States of America **73:** 4604-4607.

WAGNER, G. P., and L. ALTENBERG, 1996 Perspective: complex adaptations and the evolution of evolvability. Evolution **50:** 967-976.

WAGNER, G. P., J. P. KENNEY-HUNT, M. PAVLICEV, J. R. PECK, D. WAXMAN *et al.*, 2008 Pleiotropic scaling of gene effects and the 'cost of complexity'. Nature **452:** 470-U479.

WRIGHT, S., 1932a The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proceedings of the 6th International Congress on Genetics **1:** 356-366.

WRIGHT, S., 1932b The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proceedings of the sixth national congress of genetics **1:** 356-366.

2 GENETIC CONSTRAINTS GUIDE EVOLUTIONARY TRAJECTORIES IN A PARALLEL ADAPTIVE RADIATION

Author Contribution

The study described in Chapter Two has been submitted to the journal Genetics. The paper was written by Paul Rainey and Michael McDonald. Michael McDonald performed the following experiments: MWS and SWS transposon mutageneses, creation of the *Δaws*, *Δmws*, *ΔawsΔwsp*, *ΔawsΔmws*, *ΔwspΔmws* and *ΔawsΔwspΔmws* strains. Deletion of the *awsX* gene, the *mwsR* EAL domain and reconstructions of the AWS, WS_T, MWS and WS_R WS genotypes. Fitness assays, and the experiment measuring the time taken for different deletion strains to evolve WS.

Supervisor:

Authors: Michael J. McDonald, Stefanie M. Gehrig, Peter L. Meintjes, Xue-Xian Zhang & Paul B Rainey

2.1 ABSTRACT

The capacity for phenotypic evolution is dependent upon complex webs of functional interactions that connect genotype and phenotype. Wrinkly spreader (WS) genotypes arise repeatedly during the course of a model Pseudomonas adaptive radiation. Previous work showed that the evolution of WS variation was in part explained by simple mutations in *wspF* – a component of the Wsp signaling module –, but drew attention to the existence of unknown mutational causes. Here, we identify two new mutational pathways (Aws and Mws), which in common with the Wsp module contain a di-guanylate cyclase-encoding gene that is subject to negative regulation. Mutations in the Wsp, Aws and Mws regulatory modules account for the spectrum of WS phenotype-generating mutations found among a collection of 26 WS genotypes obtained from independent adaptive radiations. These striking instances of parallel evolution led us to consider the possibility that the genetic architecture underlying the defined regulatory modules might bias the production of genetic variation available to selection – explaining in part the repeated occurrence of mutations in a small number of loci. Mechanistic studies revealed the existence of such biases – a consequence of the propensity of regulatory pathways with specific functionalities and interactions to translate spontaneous mutation into adaptive phenotypic variation at high frequency. Together this work provides insight into genetic / developmental constraints – at multiple levels of organization – and reveals the consequences of these constraints for the evolution of genetic and phenotypic variation.

2.2 INTRODUCTION

Understanding – and importantly, predicting – phenotypic evolution requires knowledge of the factors that affect the translation of mutation into phenotypic variation – the raw material of adaptive evolution. While much is known about mutation rate (e.g., DRAKE *et al.* 1998; HUDSON *et al.* 2002), knowledge of the processes affecting the translation of DNA sequence variation into phenotypic variation is minimal.

Advances in knowledge on at least two fronts suggest that progress in understanding the rules governing the generation of phenotypic variation is possible (STERN and ORGOGOZO 2009). The first stems from increased awareness of the genetic architecture underlying specific adaptive phenotypes and recognition of the fact that the capacity for evolutionary change is likely to be constrained by this architecture (HANSEN 2006; SCHLICHTING and MURREN 2004). The second is the growing number of reports of parallel evolution (e.g., ALLENDER *et al.* 2003; BANTINAKI *et al.* 2007; BOUGHMAN *et al.* 2005; COLOSIMO *et al.* 2004; FFRENCH-CONSTANT *et al.* 1998; KRONFORST *et al.* 2006; MCGREGOR *et al.* 2007; OSTROWSKI *et al.* 2008; PIGEON *et al.* 1997; SHINDO *et al.* 2005; WOODS *et al.* 2006; ZHANG 2006; ZHONG *et al.* 2004) – the independent evolution of similar or identical features in two or more lineages – which suggests the possibility that evolution may follow a limited number of pathways (SCHLUTER 1996). Indeed, giving substance to this idea are studies that show that mutations underlying parallel phenotypic evolution are non-randomly distributed and typically clustered in homologous genes (STERN and ORGOGOZO 2008).

While the non-random distribution of mutations during parallel genetic evolution may reflect constraints due to genetic architecture, many have argued that the primary cause is strong selection (e.g., WICHMAN *et al.* 1999; WOODS *et al.* 2006). Necessary for progress is a means of disentangling the roles of population processes (selection) from genetic architecture (BRAKEFIELD 2006; MAYNARD SMITH *et al.* 1985); also necessary is mechanistic insight into precisely how genetic architecture biases the production of mutations presented to selection.

Despite their relative simplicity microbial populations offer opportunities to advance knowledge. The wrinkly spreader morphotype is one of many different niche specialist genotypes that emerge when experimental populations of *P. fluorescens* are propagated in spatially structured microcosms (RAINEY and TRAVISANO 1998). Previous studies defined, via suppressor analysis, the essential phenotypic and genetic traits that define a single wrinkly spreader (WS) genotype known as LSWS (SPIERS *et al.* 2003; SPIERS *et al.* 2002). LSWS differs from the ancestral SM genotype by a single non-synonymous nucleotide change in *wspF*. Functionally, WspF is a methyl esterase and negative regulator of the WspR di-guanylate cyclase (DGC) (GOYMER *et al.* 2006) responsible for the biosynthesis of c-di-GMP (MALONE *et al.* 2007) – the allosteric activator of cellulose synthesis enzymes (ROSS *et al.* 1987). The net effect of the *wspF* mutation is to promote physiological changes that lead to the formation of a microbial mat at the air-liquid interface of static broth microcosms (RAINEY and RAINEY 2003).

The range of genetic changes in *wspF* was determined by sequencing this gene in a collection of 26 WS genotypes (WS_{A-Z}) obtained from 26 independent adaptive radiations: 13 contained mutations in *wspF* (Table 1) (BANTINAKI *et al.* 2007). The apparent bias of

WS mutations to *wspF* can be explained in part by the specific functional role of WspF and its regulatory interactions: any impairment of WspF function activates WspR, which generates WS. Given that loss of function mutations are more numerous than gain of function mutations, mutations in *wspF* are more likely to generate the WS phenotype than mutations in other components of the Wsp pathway.

In this study we build upon knowledge of mutational routes to the WS phenotype and the influence of genetic architecture by describing two additional pathways (Aws and Mws) that together with the Wsp pathway account for the evolution of 26 independent WS genotypes. Each pathway offers approximately equal opportunity for WS evolution; nonetheless additional – less readily realized – genetic routes producing WS genotypes with equivalent fitness effects exist. Together our data shows that genetic constraints – arising from the fact that certain regulatory interactions have increased propensities to translate spontaneous mutation into adaptive phenotypic variation than others – influence evolutionary trajectories.

2.3 MATERIALS AND METHODS

2.3.1 BACTERIAL STRAINS, GROWTH CONDITIONS AND MANIPULATION

The ancestral (wild-type) "smooth" (SM) strain is *P. fluorescens* SBW25 and was isolated from the leaf of a sugar beet plant grown at the University Farm, Wytham, Oxford, in 1989 (RAINEY and BAILEY 1996). The ancestral strain was immediately stored at -80°C to minimize adaptation to the laboratory environment. The niche-specialist wrinkly spreader (LSWS) genotype (PR1200 (previously named PBR663)) was derived from the ancestral genotype following three days of selection in a spatially structured microcosm (SPIERS *et al.* 2002). 26 independent WS (WS_A-WS_Z) were derived from the ancestral SM genotype following 5 days of selection (a single WS type was isolated from each of 26 independent microcosms) (BANTINAKI *et al.* 2007). LSWS *lacZ* (PBR947) was generated by mobilizing mini-Tn*7*-*lacZ* from *Escherichia coli* (CHOI *et al.* 2005).

P. fluorescens strains were grown in lysogeny-broth (LB) (BERTANI 1951) or King's Medium B (KB) (KING *et al.* 1954) at 28°C. *E. coli* strains were grown in LB at 37°C. Antibiotics and supplements were used at the following concentrations: ampicillin 100 ⌠g ml^{-1}, cycloserine 8 ⌠g ml^{-1}, gentamycin 10 ⌠g ml^{-1}, kanamycin 100 ⌠g ml^{-1}, spectinomycin 100 ⌠g ml^{-1} and tetracyline 12 ⌠g ml^{-1}. X-gal (5-bromo-4-chloro-3-indolyl-®-D-galactopyranoside) was added at a concentration of 40 ⌠g ml^{-1}. *N*-[5-nitro-2-furfurylidene]-1-arminutesohydantoin (NF) was used to counter select *E. coli* (100 ⌠g ml^{-1} in agar plates). Plasmid DNA was introduced to *E. coli* by transformation and *P. fluorescens* by conjugation,

following standard procedures. The helper plasmid pRK2013 was used to facilitate transfer between *E. coli* and *P. fluorescens* (FIGURSKI and HELINSKI 1979).

2.3.2 MOLECULAR BIOLOGY TECHNIQUES

Standard molecular biology techniques were used throughout (SAMBROOK *et al.* 1989). Gateway technology (Invitrogen) – vectors pCR8 and pCR2.1 – were used to clone PCR fragments. The fidelity of all cloned fragments was checked by DNA sequencing. Oligonucleotide primers for allelic replacements and reconstructions were designed based on the SBW25 genome sequence (SILBY *et al.* 2009).

2.3.3 CONSTRUCTION OF DELETION MUTANTS AND ALLELIC REPLACEMENTS

A two-step allelic exchange strategy was used as described previously (BANTINAKI *et al.* 2007; RAINEY 1999). For the generation of deletion mutants, oligonucleotide primers were used to amplify ~1,000 nucleotide regions flanking the gene(s) of interest. A third PCR reaction was used to SOE flanking sequences together (HORTON *et al.* 1989). The product was cloned into pCR8 (and sequenced to check for errors); the resulting fragment was excised using *Bgl*II and introduced into the suicide vector pUIC3 (RAINEY 1999). Deletion constructs were mobilized into *P. fluorescens* by conjugation and allowed to recombine with the chromosome by homologous recombination. Recombinants from which pUIC3 had been lost were identified from LB agar plates after a brief period of non-selective growth in KB broth. To generate allelic replacements, PCR fragments (~3 kb) centered upon the allele of interest were cloned as above, introduced into pUIC3 and the replacement made by

recombination. Deletion and allelic exchange mutants were confirmed by PCR; allelic exchange mutants were further checked by DNA sequencing.

2.3.4 TRANSPOSON MUTAGENESIS ANALYSIS

Plasmid pCM639 containing IS*phoA*/hah-Tc or IS-∧-kan/hah (GIDDENS *et al.* 2007) was introduced from *E. coli* SM10λ_{pir} into *P. fluorescens* by conjugation. The genomic location of the transposon insertions was determined by sequencing the DNA products of arbitrary primed (AP)–PCR (MANOIL 2000). The precise genomic location of each transposon insertion is recorded in Artemis (RUTHERFORD *et al.* 2000) entry files that can be read into the SBW25 genome (SILBY *et al.* 2009) annotation file (http://www.sanger.ac.uk/Projects/P_fluorescens). The Artemis entry files are available as supplementary files. The SBW25 annotation file is available from the Sanger website.

As the suppressor study progressed (see Results) it became clear that genes encoding di-guanylate cyclases (DGCs) were of primary interest and yet mutations were commonly (and expectedly) found in the *wss* locus (~35 % of insertions). A simple PCR screen was developed to identify these mutants. Each mutant was therefore subjected to a PCR reaction that included a primer specific to the transposon and reading in the outward direction plus a set of 13 *wss*-specific primers spaced ~800 bp (available on request). Mutants from which a PCR fragment was detected were deemed to carry a transposon in the *wss* operon and these mutants were not further characterized.

2.3.5 FITNESS OF GENOTYPES

Competitive fitness of WS genotypes was determined by direct competition between each WS genotype and LSWS *lacZ* (PBR947): the ancestral broth colonizing genotype was included to ensure occupancy of the broth phase (BANTINAKI *et al.* 2007). All strains were grown overnight in shaken KB broth before introduction into spatially structured microcosms (~10^5 cells of each competitor) at a ratio of 1:1:1. Relative fitness was calculated as the ratio of Malthusian parameters of the WS strains being compared (LENSKI *et al.* 1991).

2.4 RESULTS

2.4.1 COMPREHENSIVE SUPPRESSOR ANALYSIS OF LSWS

The strategy devised for identification of *wsp*-independent mutation routes involved: (1), deletion of the *wsp* operon from the ancestral genotype; (2), re-evolution of WS from the *wsp*-deficient starting genotype; (3), suppressor analysis of one selected *wsp*-independent WS and; (4), comparison of the results of the *wsp*-deficient WS suppressor study with those from the suppressor analysis of LSWS. To be successful a comprehensive set of suppressor mutations from both LSWS and the non-*wsp* WS must be available so that differences – and thus the likely identity of the non-*wsp* pathway – could be identified.

The initial suppressor analysis of LSWS reported in Spiers *et al* (2002) was performed prior to the availability of the SBW25 genome sequence; dependence on cloning, mapping and cosmid libraries resulted in a limited set of defined WS mutants (8,000 mutants were screened; 27 unique WS negative mutants were identified and the nature of genetic defect was determined in 11 instances) (SPIERS *et al.* 2002). In order to exploit the comparative suppressor strategy the suppressor analysis of LSWS was extended using IS*phoA*/hah-Tc (BAILEY and MANOIL 2002). A total of 113,000 transposon mutants were screened for loss of the wrinkled colony morphology (assuming a Poisson distribution and a genome containing 6,700 genes, the probability of a gene not being inactivated by a transposon given the screening of 113,000 mutants is less than 0.001). A total of 134 WS suppressor mutants were obtained and the position of the transposon determined in each instance: the results are reported in an accompanying Artemis file (Supplemental Data File 1 LSWS).

All loci previously identified (SPIERS *et al.* 2002) were found in this study: 47 (35 %) insertions were located in *wss* and 26 (19 %) in *wsp*; in addition, insertions were found in genes responsible for maintenance of rod-shaped cells (e.g., *mreB*, *rodA* and *pbpA*) indicating cell shape as an essential determinant of the WS phenotype (SPIERS *et al.* 2002). Two previously undiscovered loci were also identified. The first is an eight-gene locus, unique to SBW25, with a predicted role in cell wall biogenesis (pflu1661 – pflu1668). The locus was termed *wsw* (*w*rinkly *s*preader cell *w*all biogenesis locus). The second locus is comprised of five-genes (pflu0475 – pflu0479) and includes a transcriptional regulator, a membrane bound de-*N*-acetylase and two glycosyl transferases – also with a possible role in modification of cell wall material. This locus was termed *wsm* (*w*rinkly *s*preader cell wall *m*odification locus). Eight transposon insertions were found in *wsw* and nine in *wsm*. Mutations in *wsw*, while eliminating the wrinkled morphology of LSWS, did not abolish ability to form a mat at the air-liquid interface of static broth microcosms (GEHRIG 2005).

2.4.2 AWS: A *WSP*-INDEPENDENT MUTATIONAL ROUTE TO WS

To identify the first *wsp*-independent mutational route to WS the entire *wsp* operon was deleted from the genome of the ancestral SM genotype. This SM ⊗*wspABCDEFR* genotype (PBR210) was phenotypically indistinguishable from the ancestral genotype. When propagated in a spatially structured microcosm a characteristic WS mat was observed within five days at which point the microcosm was destructively harvested and samples plated on KB agar. Morphological diversity was no different to the previously described five day adaptive radiation of the ancestral genotype (RAINEY and TRAVISANO 1998). The three major morph classes were all present, but WS genotypes were the dominant type: a single

WS genotype, designated *a*lternate *w*rinkly *s*preader (AWS (PBR717)), was selected for suppressor analysis using IS*phoA*/hah-Tc as described for LSWS.

Approximately 95,000 colonies were screened for loss of the wrinkled morphology: the position of the transposon was determined in 186 mutants and the results are reported in an accompanying Artemis file (Supplemental Data File 2 AWS). Comparison of the results of AWS and LSWS suppressor studies revealed many loci in common, e.g., 64 insertions were identified in *wss* (35 %), 34 were found in *wsw* (18 %) and 13 in *wsm* (7 %), but also two unique loci. The first was the *muc* locus (BOUCHER *et al.* 1996) comprising genes pflu1467 – pflu1471 (*algU mucABD*) and containing transposon insertions (two each) in *mucA* and *mucD*. The second locus was defined by 11 insertions in a 2 kb region spanning pflu5211 (two insertions) and pflu5210 (nine insertions). Comparison to DNA and protein sequence databases showed that these two genes (and a third CDS, pflu5209), are conserved across *Pseudomonas* and other proteobacteria, but have no known function. Together the genes form a putative operon termed *aws* (pflu5211 *awsX,* plfu5210 *awsR* and pflu5209 *awsO*).

Mutations in the *muc* locus of *P. aeruginosa* have been implicated in expression of the mucoid colony type and while of some considerable interest (see Discussion) our initial attention focused on the *aws* locus because of the presence within this cluster of a di-guanylate cyclase.

Predicted proteins of the Aws locus: AwsX is 171 amino acids in length; with the exception of a proteolytic cleavage site between residues 31 and 32 (suggesting that AwsX is periplasmic) the protein has no defining motifs or database matches. AwsR is 420 amino acids in length and contains two trans-membrane domains at the *N*-terminus (residues 19-41 and 151-173), a HAMP domain (residues 159-229, Pfam *E*-value 1.4 x 10^{-7}) and a GGDEF domain (residues 241-403, Pfam *E*-value 9.2 x 10^{-58}): the GGDEF domain contains all conserved

motifs characteristic of a cyclic-di-GMP (bis-3'-5'-cyclic di-GMP) synthesizing di-guanylate cyclase (DGC). In addition to the invariable DGC motifs there is a potential "I-site": presence of this motif (R**D) indicates that cyclic-di-GMP binds allosterically to AwsR. I-sites have been shown to enable negative feedback control of DGC activity (CHRISTEN *et al.* 2006). AwsO is 163 amino acids and contains an OmpA (MotB) domain (residues 52-147, Pfam *E*-value 4.2 x 10^{-26}) suggesting a porin-like function (SUGAWARA and NIKAIDO 2004).

2.4.2.1 MUTATIONS IN AWSX ARE NECESSARY AND SUFFICIENT FOR THE AWS PHENOTYPE:

DNA sequence analysis of the *aws* locus from AWS revealed a 39 bp deletion spanning nucleotides 100 to 138 of *awsX* (*awsX* ⊗100-138 (AwsX ⊗34-46)). The deletion is in frame and flanked by a five-nucleotide repeat (AGGCG): the intervening sequence plus one copy of the repeat is absent in AWS.

Armed with knowledge of a mutation in *awsX*, the DNA sequence of this locus was determined in the 13 independent WS genotypes lacking mutations in *wspF* (BANTINAKI et al. 2007). In three of these genotypes: WS_S, WS_T and WS_V mutations were found in *awsX*. WS_T contains an in-frame deletion (33 nucleotides) spanning residues 229 – 261 and flanked by a six nucleotide repeat (ACCCAG). As in AWS, the deletion in WS_T removed the intervening nucleotides leaving one copy of the repeated sequence. WS_S and WS_V both contain single nucleotide polymorphisms (a transition and transversion, respectively): see Table 2.1.

To determine whether the *awsX* mutations are alone sufficient to cause the WS phenotype, both the AWS and WS_T mutations were reconstructed in the ancestral genetic background by allelic exchange. Three independent exchanges were made for each allele and in each instance the resulting genotype was WS.

WS genotype	Gene	Nucleotide change	Amino acid change	Source / Reference
LSWS	*wspF*	A901C	S301R	BANTINAKI *et al*. (2007)
AWS	*awsX*	⊗100-138	⊗PDPADLADQRAQA	This study
MWS	*mwsR*	G3247A	E1083K	This study
WS_A	*wspF*	T14G	I5S	BANTINAKI *et al*. (2007)
WS_B	*wspF*	Δ620-674	P206Δ(8)[a]	BANTINAKI *et al*. (2007)
WS_C	*wspF*	G823T	G275C	BANTINAKI *et al*. (2007)
WS_D	*wspE*	A1916G	D638G	This study
WS_E	*wspF*	G658T	V220L	BANTINAKI *et al*. (2007)
WS_F	*wspF*	C821T	T274I	BANTINAKI *et al*. (2007)
WS_G	*wspF*	C556T	H186Y	BANTINAKI *et al*. (2007)
WS_H	*wspE*	A2202C	K734N	This study
WS_I	*wspE*	G1915T	D638Y	This study
WS_J	*wspF*	Δ865-868	R288Δ(3)[a]	BANTINAKI *et al*. (2007)
WS_K	*awsO*	G125T	G41V	This study
WS_L	*wspF*	G482A	G161D	BANTINAKI *et al*. (2007)
WS_M	*awsR*	C164T	S54F	This study
WS_N	*wspF*	A901C	S301R	BANTINAKI *et al*. (2007)
WS_O	*wspF*	Δ235-249	V79Δ(6)[a]	BANTINAKI *et al*. (2007)
WS_P	*awsR*	222insGCCACCGAA	74insATE	This study
WS_Q	*mwsR*	3270insGACGTG	1089insDV	This study
WS_R	*mwsR*	T2183C	V272A	This study
WS_S	*awsX*	C472T	Q158*	This study
WS_T	*awsX*	Δ229–261	⊗YTDDLIKGTTQ	This study
WS_U	*wspF*	Δ823-824	T274Δ(13)[a]	BANTINAKI *et al*. (2007)
WS_V	*awsX*	T74G	L24R	This study
WS_W	*wspF*	Δ149	L49Δ(1)[a]	BANTINAKI *et al*. (2007)
WS_X[b]	?	?	?	This study
WS_Y	*wspF*	Δ166-180	Δ(L51-I55)	BANTINAKI *et al*. (2007)
WS_Z	*mwsR*	G3055A	A1018T	This study

Table 2.1.
The mutational causes of WS. [a]P206Δ(8) indicates a frame shift; the number of new residues before a stop codon is reached is in parentheses
[b]Suppressor analysis implicates the *wsp* locus (17 transposon insertions were found in this locus), however, repeated sequencing failed to identify a mutation.

2.4.2.2 AWSX IS A NEGATIVE REGULATOR

Previous studies showed that the proximate cause of the LSWS phenotype is over-activation of the WspR DGC (BANTINAKI *et al.* 2007; GOYMER *et al.* 2006). The fact that the Aws locus also includes a putative DGC (AwsR) suggested the possibility that over-activation of AwsR might cause the WS phenotype. While there are no clues in the amino acid sequence of AwsR to indicate the means by which activation of the DGC domain might occur, the fact that mutations in *awsX* cause the WS phenotype suggested that AwsX might suppress the activity of AwsR. However, the fact that all *awsX* mutations are in-frame (and in two cases deletions between repeats) left open the possibility that AwsX might be a positive activator – the activated state being realized upon specific in-frame deletions (some support for a positive regulatory role of AwsX came from an earlier study (GEHRIG 2005)).

To test directly the hypothesis that AwsX is a negative regulator a complete *awsX* deletion mutant was made in the ancestral SM genotype. Constructing such a mutant in which the *awsR* open reading frame was fused to the ATG-start codon of *awsX* (replacing *awsX*) proved impossible. However, deletion of *awsX* was achieved when the translational coupling between *awsX* and *awsR* evident in the genome sequence of the ancestral genotype was incorporated into the design of the deletion construct. SM ⊗*awsX* (PBR717) was phenotypically WS and produced copious calcofluor-staining material (the cellulosic product of the *wss* operon). This demonstrates that AwsX is a negative regulator of AwsR; and that AwsR is a DGC capable of causing over-production of the acetylated cellulose synthesizing *wss* operon via the synthesis of c-di-GMP.

2.4.3 MWS: A *WSP*- AND *AWS*-INDEPENDENT ROUTE TO WS

While identification of the *aws* mutational route to WS reduced the number of unknown WS-causing mutations among the 26 independent WS genotypes, the existence of WS genotypes with no mutations in either *wspF* or *awsX* raised the possibility of still further pathways. To identify these the strategy established above was repeated using transposon IS-∧-kan/hah (GIDDENS *et al.* 2007). This time the starting genotype was the ancestral SM genotype from which both the *wsp* and *aws* loci had been cleanly deleted: SM ⊗*wspABCDEFR* ⊗*awsXRO* (PBR712).

PBR712 is phenotypically indistinguishable to the ancestral SM genotype and diversified when propagated in a static broth microcosm in a manner fully in accord with the ancestor. When a sample from a 5 day-old microcosm was diluted and plated on KB WS genotypes were evident. One WS colony was selected for further study and named MWS (*M*ike's *w*rinkly *s*preader: PBR714). Suppressor mutations of MWS that reverted the colony phenotype to smooth were sought as above.

A total of 91 WS defective mutants were obtained. Transposon insertions in *wss* (seven insertions) were eliminated by PCR screening (see Material and Methods). Given that in both LSWS and AWS the mutational route to WS involved a DGC-containing locus the remaining suppressor mutants were analyzed only to the point at which several insertions were found in a previously un-encountered DGC. After determining the insertion point of the transposon in just 19 suppressor mutants, seven were found within a large (3.8kb) gene (Pflu5329) – named *mwsR* – encoding several protein domains including a DGC. The results are reported in an accompanying Artemis file (Supplemental Data File 3 MWS). Flanking

mwsR in SBW25 is *glyA* (a serine hydroxy methyl transferase (SHMT)) and a putative membrane protein. MwsR is present in most of the sequenced *Pseudomonas* genomes: where present it is flanked by *glyA* (always on the downstream side of *mwsR*). A previous study of *P. aeruginosa* and *P. putida* assigned a biofilm phenotype to an orthologue of MwsR named MorA (CHOY *et al.* 2004).

The N terminus of MwsR consists of two trans-membrane regions: residues 477-569 define a PAS_3 domain (Pfam *E*-value 1.3 x 10^{-11}), residues 666-712 and 723-834 both define PAS domains (Pfam *E*-values 5.8 x 10^{-5} and 1.3 x 10^{-10}, respectively); 848-1009 is a GGDEF domain (Pfam *E*-value 7.3 x 10^{-61}) and 1029-1264 is an EAL domain (Pfam *E*-value 3.5 x 10^{-90}). The GGDEF domain contains all the motifs characteristic of a DGC, although lacks the I-site. The EAL domain is a predicted phosphodiesterase (PDE), which facilitates hydrolysation of c-di-GMP to pGpG (TAMAYO *et al.* 2005). PDE domains are often found adjacent to DGC domains leading to the suggestion that DGC/PDE domain proteins may both synthesize and degrade c-di-GMP, although a hybrid protein with both domains functional has not been described (JENAL and MALONE 2006).

2.4.3.1 PREDICTING THE MUTATIONAL CAUSE OF THE MWS

Emerging from the analysis of LSWS (where mutations in *wspF* are the cause of the WS phenotype) and AWS (where mutations in *awsX* cause the WS phenotype) is understanding of the significance of loss of function mutations in negative regulators of DGCs. Applying this trend to MwsR led to the prediction that causal mutations might reside in the EAL domain (if both PDE and DGC activities exist in MwsR then the balance is likely to be

crucial: a mutation that reduced or abolished PDE function would cause over supply of c-di-GMP with ensuing effects on the production of the cellulosic polymer).

DNA sequence analysis of *mwsR* from MWS revealed a single non-synonymous nucleotide substitution in the EAL domain (G3247A) resulting in the substitution of lysine for glutamic acid. Sequence analysis of *mwsR* from the collection of 26 independent WS identified three additional *mwsR* mutations: WS_Q contains a six nucleotide insertion after position 3270 arising from duplication of the preceding six nucleotides; WS_R and WS_Z harbour non-synonymous mutations (Table 2.1).

2.4.3.2 MUTATIONS IN MWSR ARE NECESSARY AND SUFFICIENT FOR THE MWS PHENOTYPE

To test the hypothesis that mutations in *mwsR* are the direct cause of the WS phenotype two mutant alleles (G3247A and T2183C) were recombined into the ancestral genome replacing the wild type sequence. Three independent allelic replacements were made for each allele and in all instances the resulting genotype was WS (each WS genotype tested positive for the production of calcofluor-staining material – the cellulosic product of the *wss* operon).

2.4.3.3 THE EAL DOMAIN NEGATIVELY REGULATES MWSR ACTIVITY

We sought understanding of the relationship between the EAL and GGDEF domains in MwsR: specifically we asked whether the EAL domain negatively regulates the GGDEF domain and whether the GGDEF domain is the source of c-di-GMP that causes the WS phenotype. To address the first question nucleotides 3116 to 3670 (encoding the EAL

domain) were deleted from the ancestral SM genotype: this genotype (PBR935) was morphologically WS. While the proximate cause of WS is, in every case so far examined, a consequence of over production of c-di-GMP, the source of c-di-GMP caused by deletion of the EAL domain of MwsR need not be the GGDEF domain of MwsR. The fact that no DGC activity has been detected in proteins containing both EAL and GGDEF domains (JENAL and MALONE 2006) raised the possibility that the activated GGDEF domain might not reside within MwsR. To explore this further the remainder of the *mwsR* open reading frame was removed from PBR935: the resulting genotype (SM ⊗*mwsR*; PBR771) was morphologically smooth. If the primary function of MwsR is as a PDE (i.e., the GGDEF domain is non-functional) whose activity balanced the DGC activity encoded by some other protein, then deletion of the GGDEF domain of MwsR in PBR935 would not have altered the WS phenotype. Together these data provide evidence that both the EAL and GGDEF domains of MwsR are functional and that the PDE activity of the EAL domain negatively regulates the DGC activity of the GGDEF domain. Moreover, the fact that this and the allelic exchange mutants (*mwsR* G3247A and T2183C) produce cellulosic material indicates that the net effect of the over-active DGC domain is once again over-production of acetylated cellulose due to the over production of c-di-GMP.

2.4.4 FITNESS OF AWS AND MWS

In all cases replacement of the wild type allele by each mutant allele (in the ancestral genetic background) resulted in colony morphology and mat forming ability indistinguishable from that of the original WS mutant. In order to assess the fitness consequences of these mutations, the fitness of mutants generated by allelic exchange was determined relative to a

reference WS (the LSWS genotype) and compared with the fitness of the original mutants (again relative to LSWS containing a *lacZ* marker (PBR947)). These fitness assays included the ancestral SM genotype whose superior growth in the broth phase insures that differences between the competing WS strains are due to competitive ability at the air-liquid interface (BANTINAKI *et al.* 2007). In all cases the fitness of the reconstructed mutant was indistinguishable from the fitness of the original mutant indicating that the mutation identified in each derived genotype accounts fully for both phenotypic and fitness differences between ancestral and derived types. The spectrum of fitness effects is shown in Figure 1 and the caption includes the results of the statistical analyses.

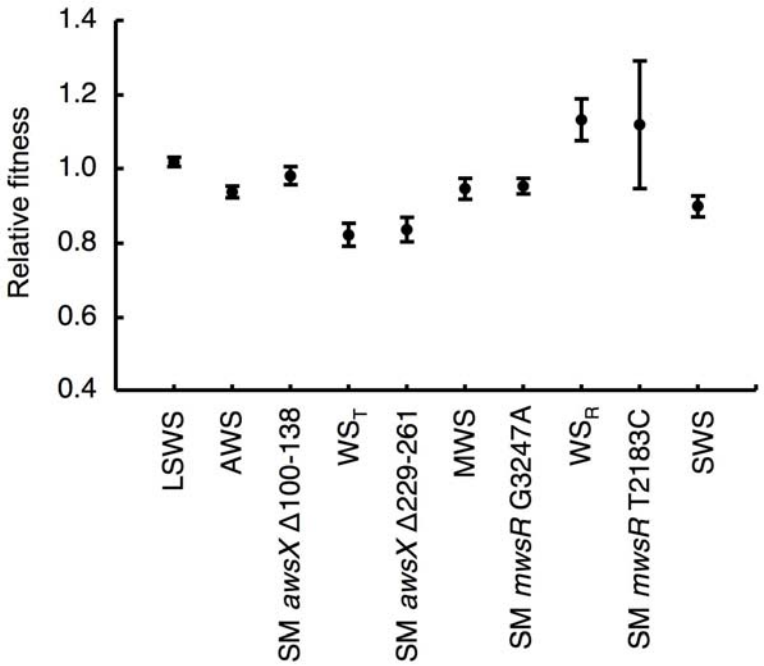


Figure 2.1. Relative fitness of WS genotypes. Fitness was determined relative to the LSWS genotype carrying a neutral *lacZ* marker. The experiment was designed to test the hypothesis that the mutations identified in WS genotypes AWS and MWS are causal, and to test the hypothesis that the fitness of SWS differs significantly from the known *wspF*, *awsX* and *mwsR* generated WS genotypes. *t*-Tests revealed no significant differences between derived WS and reconstructed genotypes for AWS vs. SM *awsX* ⊗100-138 (P = 0.081), WS_T vs. SM *awsX* ⊗229-261 (P = 0.381), MWS vs. SM *mwsR* G3260A (P = 0.424), and WS_X vs. SM *mwsR* T2183C (P = 0.497). One way ANOVA showed a significant difference among the means for the entire data set ($F_{[8, 52]} = 10.12$, $P < 0.0001$). Tukey's HSD showed that the fitness of SWS was not significantly different to LSWS, MWS or AWS. Data are means and standard errors of six replicates.

2.4.5 THE RELATIVE CONTRIBUTION OF *WSP, AWS* AND *MWS* TO THE EVOLUTION OF WS VARIATION

By the fifth day of propagation WS genotypes were detected in spatially structured microcosms founded by the ancestral genotype from which the *wsp* operon had been deleted (SM ⊗*wsp*); within the same time period WS genotypes also arose from the ancestral genotype from which both *wsp* and *aws* were deleted (SM ⊗*wsp* ⊗*aws*). To systematically investigate the contribution of each locus (*wsp*, *aws*, and *mws*) to the production of WS variation a set of deletion mutants was generated in which all possible single, double and triple locus mutants were constructed (SM ⊗*wsp* (PBR210), SM ⊗*aws* (PBR681), SM ⊗*mws* (PBR711), SM ⊗*wsp* ⊗*aws* (PBR712), SM ⊗*wsp* ⊗*mws* (PBR713), SM ⊗*mws* ⊗*aws* (PBR718), SM ⊗*wsp* ⊗*aws* ⊗*mws* (PBR716); replicated assays in KB broth showed no differences in growth dynamics among genotypes (see supplementary data)). Each genotype was inoculated into a series of independent broth-filled microcosms (~5 x 10^6 cells per microcosm) and one microcosm from each series was destructively harvested on a daily basis through the course of nine days (each assay was replicated three-fold). Cells were diluted, plated, and following 48 h growth on KB agar, the phenotype of 500 colonies from each microcosm was checked for the presence of WS genotypes. With the exception of microcosms founded with the triple deletion genotype (PBR716) WS genotypes were detected in each and every two-day old microcosm: WS genotypes were detected in microcosms founded by PBR716 at day five.

2.4.6 SWS: A *WSP*- *AWS*- AND *MWS*-INDEPENDENT ROUTE TO WS:

A single WS genotype derived from SM ⊗*wsp* ⊗*aws* ⊗*mws* (PBR716) was selected and subjected to suppressor analysis as described above. A total of 41 suppressor mutants were identified (the results are reported in an accompanying Artemis file (Supplemental Data File 4 SWS)): transposon insertions were found in previously identified loci including *wss* (three were identified by PCR screen). As in the analysis of MWS, characterization of suppressor mutants was halted once a previously un-encountered GGDEF domain-encoding gene was identified. In this instance, two insertions were found in a single open reading frame (pflu1349) predicted to encode a GGDEF and EAL domain-containing protein. The predicted product of this gene – designated *sws*R (*s*low *w*rinkly *s*preader) – is a protein of 759 amino acids and has two repeat regions and two domains: residues 76-137 and 139-200 define two MHYT *N*-terminal repeats (GALPERIN *et al.* 2001) (Pfam *E*-value 1.2 x 10^{-15} and 3.5 x 10^{-13}, respectively) with seven probable membrane spanning regions between residues 32 and 262; residues 295-452 define a GGDEF domain (Pfam *E*-value 1.6 x 10^{-56}), and residues 472-708 define an EAL domain (Pfam *E*-value 3.5 x 10^{-79}).

Deletion of *sws*R from SWS caused the phenotype to revert to the ancestral smooth type confirming its causal role, however repeated sequencing of pflu1349 (including promoter region) revealed no mutations. The fitness of SWS was determined relative in LSWS *lacZ* as described above and found to be statistically indistinguishable from the fitness of LSWS, SWS, MWS and AWS (Figure 2.1). Furthermore, with a fitness 0.89 relative to LSWS, the SWS genotype is at the mid-point of the fitness measures determined for WS genotypes (BANTINAKI *et al.* 2007).

2.4.7 THE MUTATIONAL ORIGINS OF THE REMAINING INDEPENDENT WS GENOTYPES

The data thus far indicate the existence of three – and only three – readily achievable single-step mutational routes to WS: *wsp*, *aws*, and *mws*. Clearly, the discovery of SWS indicates that WS genotypes may arise by additional routes, however the fact that SWS – with a fitness equivalent to other characterized WS genotypes – took three days longer to arise from the 'ancestral' SM ⊗*wsp* ⊗*aws* ⊗*mws* (PBR716) genotype than from any of the single or double locus 'ancestral' mutants indicates that these other mutational routes, while obtainable, are less likely to be realized. Assuming equivalent per locus mutation rates we conclude that the mutational route involving *swsR*, and numerous other potential non-*wsp*, non-*aws*, and non-*mws* genetic routes to WS (see discussion), is rarely available to selection because of a reduced capacity to translate mutation into WS variation. If this is true, then the mutational origins of the 26 WS genotypes from independent adaptive radiations should be accounted for by mutations in the three known loci (*wsp*, *aws*, or *mws*). However, sequencing of *wspF*, *awsX* and *mwsR* from the 26 WS genotypes shows that mutation in these loci accounts for just 19 of the 26 WS genotypes. One possibility is that these genotypes do contain mutations in *wsp*, *aws*, or *mws*, but in components other than *wspF*, *awsX* and *mwsR*. This led us to return to the independent WS genotypes from which mutations had not been identified.

WS genotypes WS_D, WS_H, WS_I, WS_K, WS_M, WS_P, and WS_Z were each subjected to suppressor analysis using transposon IS-∧-kan/hah. The position of the transposon was mapped in each suppressor mutant: in each instance transposon insertions were found in either *wsp*, or *aws*: WS_D, *wsp*; WS_H *wsp*; WS_I, *wsp*; WS_K *aws*; WS_M, *aws*; WS_P, *aws*; and WS_z *wsp* (failure to find insertions in *mws* is consistent with *mwsR* being the sole component of this locus). This strongly implicated the *wsp* and *aws* loci as mutational targets. The DNA

sequence of the entire *wsp* or *aws* locus was therefore obtained from each of the independent WS genotypes (the locus sequenced was that indicated by the results of the suppressor analysis) and compared with the ancestral sequence. In six of the seven WS genotypes a mutation was identified: WS_D, WS_H, WS_I, WS_K and WS_M each contain a single non-synonymous substitution (WS_D,WS_H, WS_I harbor a mutation in the *wspE* kinase, WS_K contains a mutation in *awsO* and WS_M contains a mutation in *awsR*). WS_P also carries a defect in *awsR*, but the mutation is caused by a nine nucleotide (in-frame) insertion (GCCACCGAA) after position 222 (details are given in Table 2.1). Repeated sequencing of *wsp* from WS_X revealed no mutation in any component of this locus, even though this locus is firmly implicated by the results of the suppressor analysis. The sequence of *mwsR* and *awsXRO* was also wild type for WS_X.

2.5 DISCUSSION

Systematic change in the genetic structure of populations is brought about by the action of natural selection. But the opportunity for selection to act is dependent upon the presence of adaptive phenotypic variation. The rate at which changes in nucleotide sequence arise (mutation rate) has received (e.g., DRAKE *et al.* 1998; HUDSON *et al.* 2002) – and continues to receive (e.g., LIND and ANDERSSON 2008) – considerable attention, but poorly understood are the processes that connect variation in nucleotide sequence with phenotypic variation (WAGNER and ALTENBERG 1996) – the raw material for adaptive evolution. Despite important theoretical advances (reviewed in RICE (2008)), knowledge of the rate (per generation, per locus) at which DNA sequence change (mutation) translates into phenotypic variation is scant. Nonetheless, there is no doubt that the value of such knowledge – a measure of an organism's capacity for evolution – would be considerable (HANSEN 2006).

In this study we have defined in some detail two new loci (*aws* and *mws*) within which mutation can generate the WS phenotype; these loci, along with the previously uncovered *wsp* locus, comprise three approximately equally achievable single-step mutational routes to the adaptive WS phenotype. The three pathways are remarkably different to one another, but share two features in common: each contains a gene that encodes a protein with di-guanylate cyclase activity (*wspR*, *awsR* and *mwsR*); in each case the protein is subject to negative regulation. That these pathways do comprise the primary routes by which WS variation arises during the course of the model Pseudomonas radiation is strongly supported by the fact that 25 of 26 WS genotypes from independent radiations harbor mutations in one of the three identified loci (Table 1). WS_X, the one WS genotype with no mutation in *wsp*, *aws* or *mws* nonetheless requires a functional Wsp pathway for expression of the phenotype

indicating that the causal mutation in WS_X is accessory to – but part of – the Wsp regulatory pathway.

Based on previously published genetic and biochemical evidence (in the case of *wsp* (GOYMER *et al.* 2006; MALONE *et al.* 2007)), and the results of the analyses presented here for *aws* and *mws*, we infer that in each instance the proximate cause of WS is over production of c-di-GMP, which is itself a consequence of constitutive activation of the corresponding DGC (WspR, AwsR or MwsR). Interestingly, despite the diverse nature of the *wsp*, *aws* and *mws* regulatory modules, the discovery of transposon insertions in the acetylated cellulose polymer-encoding *wss* operon from suppressor analysis of LSWS, AWS and MWS (also SWS) indicates that the primary structural determinant of WS is in each instance the acetylated cellulose polymer (SPIERS *et al.* 2002). This raises a number of questions as to the nature of the regulatory connections between the DGC-encoding genes and the product of the Wss locus in the ancestral genotype. At this stage it is unclear whether the primary role of the numerous DGC-encoding genes in the genome (of which there are 39 (SILBY *et al.* 2009)) is, in each instance, to regulate the production of cellulose (presumably in response to different input signals), or whether the many-to-one mapping revealed in our study of WS genotypes is a consequence of the co-option (GERHART and KIRSCHNER 1997) of DGC-encoding modules whose regulatory targets have been altered as a consequence of constitutive activation such that they gratuitously activate production of the cellulose polymer. This latter possibility is not improbable given increasing awareness of the tight spatial and temporal regulation of DGC activity (JENAL and MALONE 2006).

A related issue concerns the degree of overlap and connectivity (pleiotropy and epistasis) of the architecture that defines the WS phenotype in WS genotypes whose phenotype arises from over-activation of different DGCs (KNIGHT *et al.* 2006). A comparison of the results

of the comprehensive suppressor studies of LSWS and AWS indicates that the underlying architectures are subtley different: expression of the AWS phenotype requires the *algU mucABD* locus, but this locus appears to have no role in expression of the LSWS phenotype. Just what role the *muc* locus plays in AWS is unclear, however studies in *P. aeruginosa* show that AlgU is an alternate sigma factor (σ^{22}) (YU *et al.* 1995), MucB an antisigma factor (of AlgU), and MucC a negative regulator (of alginate biosynthesis) (SCHURR *et al.* 1996). In the ancestral genotype of *P. fluorescens* SBW25 it was previously shown that AlgR is a positive activator, and AmrZ (AlgZ) a negative regulator, of *wss* transcription (GIDDENS *et al.* 2007), so the discovery of genes that in *P. aeruginosa* are linked with alginate regulation does not necessarily imply the involvement of alginate in the WS phenotype. Nonetheless, the regulatory nature of this locus and its specific involvement in expression of the wrinkled morphology of AWS points to further complexity and a network of overlapping regulatory connections – all fertile material for the evolution of phenotypic innovation (LANDRY *et al.* 2007).

2.5.1 PARALLEL GENETIC EVOLUTION DUE TO GENETIC CONSTRAINTS

Discovery that the WS phenotype is repeatedly generated by mutations in a small number of loci provides a further example of parallel genetic (and phenotypic) evolution: for example, 13 of the 26 independent WS genotypes have mutations in *wspF* (WS_N is identical to LSWS), three are in *wspE*, three are in *awsX*, four are in *mwsR* and two have mutations in *awsR* (*awsO* was the only singleton). Some have argued that such striking parallelism is evidence of selection (e.g., WICHMAN *et al.* 1999; WOODS *et al.* 2006); indeed, measurements of the fitness effects of specific mutations following allelic replacements in the ancestral

background support this claim. However, our data show that genetic constraints – a consequence of specific interactions and functionalities within the *wsp*, *aws* and *mws* regulatory modules – bias the spectrum of WS-generating mutations upon which natural selection acts.

Clearest evidence of this bias comes from the fact that the ancestral genotype devoid of *wsp*, *aws* and *mws* (SM ⊗*wsp* ⊗*aws* ⊗*mws* (PBR716)), can generate WS variation via mutations that activate previously unencountered DGCs (in this instance SwsR). This demonstrates that *wsp*, *aws*, and *mws* are not the only mutational routes to WS and yet the mutational causes of the 26 independent WS genotypes are confined to *wsp*, *aws*, and *mws*. A clue to the cause of this bias comes from the fact that in experimental adaptive radiations WS genotypes arising from PBR716 take 3 days longer to reach detectable frequencies compared to WS genotypes arising from the ancestor (or the ancestral genotype devoid of any one or two (but not three) of *wsp*, *aws*, or *mws*). Given that the fitness of WS generated from PBR716 (SWS) is equivalent to LSWS, AWS and MWS, and assuming the mutation rate underlying all possible WS-generating mutations is not different to the background spontaneous mutation rate (there is no indication of mutable loci among the spectrum of mutations in Table 1), then the most likely explanation for the bias stems from some feature of the genetic architecture underlying the *wsp*, *aws* and *mws* pathways that affects the rate at which spontaneous mutations in these loci are translated into WS variation. If this rate is greatest in *wsp*, *aws* and *mws* then selection will only rarely see WS variation caused by mutation in other loci. Parallel genetic evolution therefore, in this instance, is in part a product of the genetic constraints that bias the production of WS variation.

Just why the other mutational routes to WS are less readily achievable given that the ancestral genome would appear to offer numerous potential routes for WS evolution is not

clear. It is possible that some, such as *swsR* (the proximate cause of the SWS genotype), require very specific mutations to generate WS, others may require two mutations, while some DGCs may be non-functional; others, although readily activated, may be unable to interact with the cellulose synthases encoded by Wss and thus be unable to deliver c-di-GMP to the cellulose synthases and related structural components at the appropriate time.

Further and more detailed evidence of bias in the production of WS variation exists within the *wsp*, *aws* and *mws* pathways, although it is most pronounced within *wsp*. The *wsp* operon is comprised of seven genes, all, in principle, capable of acquiring mutations leading to constitutive activation of the WspR DGC. GOYMER *et al.* (2006), for example, showed that it was possible, through in vitro manipulation, to generate constitutively active WspR alleles; it is easy to imagine mutations in WspA (the methyl-accepting chemotaxis protein) that might mimic the ligand-bound state, and BANTINAKI *et al.* (2007) showed that over-activation of the WspC methyl transferase results in over-activation of WspR. However, despite the existence of seven genetic targets – and thus a large potential mutational target size – mutations causing WS were found in just *wspF* and *wspE*, with the majority (13/26) being in *wspF*. Assuming that the mutation rate is constant across the seven-gene operon, then the conclusion that each gene differs in its capacity to translate mutation into WS variation must hold. As we have previously argued (BANTINAKI *et al.* 2007), this can be explained (and was previously predicted) in terms of the function and regulatory connections among components of the Wsp pathway. WspF is a negative regulator, which when inactivated by mutation (or has its activity reduced) results in activation of WspR (BANTINAKI *et al.* 2007). This fact, combined with knowledge that most mutations result in a loss of function, explains why most *wsp*-generated WS arise from mutations in *wspF*: *wspF*,

on account of its function and role in the Wsp signalling pathway, has a high propensity to translate DNA sequence change into WS phenotypic variation.

Arguments with regard to the capacity of *wspF* to translate mutation into WS variation can be equally applied to *awsX* and the EAL domain of the protein encoded by *mwsR*, which negatively regulate the activity of the AwsR DGC and the DGC domain of MwsR, respectively. In fact it is highly likely that negative regulation of DGC activity – a feature of all three pathways – explains the prevalence of WS-causing mutations in these pathways.

2.5.2 MUTATIONAL TARGET SIZE

Mutational target size is an important factor governing the evolution of phenotypic variation. A trait with a complex functional architecture is likely to be influenced by many loci and thus by definition affords a large mutational target (HANSEN 2006; HOULE 1998). While in general terms this makes sense, our work shows that the concept can be refined given a more detailed understanding of the specific connections between genotype and phenotype.

Based firstly on the results of the comprehensive suppressor studies (for example of LSWS) we concluded that the WS phenotype is underpinned by a complex architecture of ~100 genes, however further analysis has allowed a distinction to be drawn between direct structural determinants of the WS phenotype (e.g., the cellulose-encoding *wss* operon and related adhesive factors (SPIERS *et al.* 2003; SPIERS *et al.* 2002; SPIERS and RAINEY 2005), determinants such as cell shape (SPIERS *et al.* 2002) and various cell wall modifying genes (as shown in this study) that provide a necessary scaffold for correct production of adhesive factors, and genes that define regulatory modules within which mutations generating WS

arise (BANTINAKI *et al.* 2007). It is this last set of genes that define the mutational target. But further dissection shows that target size as a function of the number of genes within regulatory modules is overly simplistic. Of the 17 genes in the combined *wsp*, *aws* and *mws* modules, 19 of the 26 causal WS mutations were found in the three negative regulators (*wspF*, *awsX* and the EAL domain of *mwsR*). What matters then, is not just the number of genes, but more importantly their function and regulatory connections: this explains why certain functional components – in the context of their appropriate set of interactions – are better able to translate mutation into phenotypic variation than others.

But genetic target size as refined above is not sufficient to account for the full spectrum of mutations observed here. In particular it does not explain the relatively large number of mutations in *wspF* relative to *awsX*: *wspF* and *awsX* are both negative regulators; *awsX* is approximately half the size of *wspF*; mutation in *wspF* was the cause of WS on 13 occasions, whereas mutation in *awsX* caused WS in just three instances – half the number expected. To some extent this might reflect the small sample size, but it is likely that there are additional constraints operating on *awsX* (and *mwsR*). In terms of transcriptional organization the negative regulatory components of both *aws* and *mws* precede the DNA sequence that encodes the DGC and in both instances a functional DGC is critical for expression of the WS phenotype. Mutations that disrupt function of the negative regulator, but also impact on the downstream DGC function, will not generate WS, and indeed in the case of *aws*, appear to be lethal. No frameshift mutations were detected in either *awsX*, or *mwsR*, which most likely reflects the fact that frameshift mutations in these regulators have deleterious pleiotropic effects. Frameshift mutations are however tolerated in *wspF* – the downstream DGC-encoding gene (*wspR*) is separated from *wspF* by 50 nucleotides and contains its own ribosome binding site (BANTINAKI *et al.* 2007). Of the 13 mutations in *wspF*,

half (six) are frameshift mutations. When the number of substitutions in *wspF* and *awsX* are compared – and corrected for gene size – the numbers are equivalent.

While it is understood that loss of function mutations are more common than gain of function mutations (KIMURA 1968) it is surprising to find that ~25% of the causal WS mutations (six out of 26) are predicted to be of this kind. The three mutations within the WspE kinase, and the two within the AwsR DGC cannot – given their phenotypic effects – be loss of function mutations. In both cases these mutations must activate the WspE kinase and AwsR DGC, respectively. The mutation in the putative outer membrane component *awsO* most likely mimics the active state of the transporter leading to release of repression of AwsR by AwsX. Strictly speaking however these activating mutations might be more appropriately considered a special case of loss of function mutation, because the activity of the regulator is no longer inducible. Either way, the prevalence of these mutations is worthy of note and raises questions concerning the relative importance of both kinds of mutations – and opportunity for their occurrence – in phenotypic evolution.

2.5.3 CONCLUSION:

The complexities of genetic architecture determine extant phenotypes, but such architectures also underpin the capacity of lineages to generate new genetic and phenotypic variation. Statistical population genetics has traditionally studied the evolution of phenotypes by averaging gene interactions (FISHER 1930) and in so doing has treated the complex network of connectivities that link genotype and phenotype as 'noise'. While the power of quantitative genetics is undisputed, the complexities arising from the genotype-phenotype map do have evolutionary implications that need to be understood (HANSEN 2006;

SCHLICHTING and MURREN 2004). Indeed, as we have shown here, parallel genetic evolution of the WS phenotype is driven by selection, but the mutations available to selection are a biased subset of all possible WS-generating mutations: this subset is comprised of mutations in loci that, on account of specific interactions and functionalities, translate mutation into adaptive phenotypic variation at the highest rates.

Unequivocal experimental demonstration of the existence of such constraints, combined with insight into their mechanistic bases and consequences, confirms that developmental biases do indeed – along with natural selection – affect the course of phenotypic evolution (BRAKEFIELD 2006; GOULD 2002; MAYNARD SMITH *et al.* 1985; SCHLUTER 1996).

2.6 BIBLIOGRAPHY

ALLENDER, C. J., O. SEEHAUSEN, M. E. KNIGHT, G. F. TURNER and N. MACLEAN, 2003 Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. Proc. Natl. Acad. Sci. USA **100:** 14074-14079.

BAILEY, J., and C. MANOIL, 2002 Genome-wide internal tagging of bacterial exported proteins. Nat. Biotechnol. **20:** 839-842.

BANTINAKI, E., R. KASSEN, C. G. KNIGHT, Z. ROBINSON, A. J. SPIERS *et al.*, 2007 Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. Genetics **176:** 441-453.

BERTANI, G., 1951 Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J. Bacteriol. **62:** 293-300.

BOUCHER, J. C., J. MARTINEZ-SALAZAR, M. J. SCHURR, M. H. MUDD, H. YU *et al.*, 1996 Two distinct loci affecting conversion to mucoidy in *Pseudomonas aeruginosa* in cystic fibrosis encode homologs of the serine protease HtrA. J. Bacteriol. **178:** 511-523.

BOUGHMAN, J. W., H. D. RUNDLE and D. SCHLUTER, 2005 Parallel evolution of sexual isolation in sticklebacks. Evolution **59:** 361-373.

BRAKEFIELD, P. M., 2006 Evo-devo and constraints on selection. Trends Ecol. Evol. **21:** 362-368.

CHOI, K.-H., J. B. GAYNOR, K. G. WHITE, C. LOPEZ, C. M. BOSIO *et al.*, 2005 A Tn*7*-based broad-range bacterial cloning and expression system. Nat. Methods **2:** 443-448.

CHOY, W. K., L. ZHOU, C. K. C. SYN, L. H. ZHANG and S. SWARUP, 2004 MorA defines a new class of regulators affecting flagellar development and biofilm formation in diverse *Pseudomonas* species. J. Bacteriol. **186:** 7221-7228.

CHRISTEN, B., M. CHRISTEN, R. PAUL, F. SCHMID, M. FOLCHER *et al.*, 2006 Allosteric control of cyclic di-GMP signaling. J. Biol. Chem. **281:** 32015-32024.

COLOSIMO, P. F., C. L. PEICHEL, K. NERENG, B. K. BLACKMAN, M. D. SHAPIRO *et al.*, 2004 The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLOS Biol. **2:** 635-641.

DRAKE, J. W., B. CHARLESWORTH, D. CHARLESWORTH and J. F. CROW, 1998 Rates of spontaneous mutation. Genetics **148:** 1667-1686.

FFRENCH-CONSTANT, R. H., B. PITTENDRIGH, A. VAUGHAN and N. ANTHONY, 1998 Why are there so few resistance-associated mutations in insecticide target genes? Philos. T. R. Soc. B **353:** 1685-1693.

FIGURSKI, D. H., and D. R. HELINSKI, 1979 Replication of an origin-containing derivative of plasmid RK2 dependant on a plasmid function provided in *trans*. Proc. Natl. Acad. Sci. USA **76:** 1648-1652.

FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.

GALPERIN, M. Y., T. A. GAIDENKO, A. Y. MULKIDJANIAN, M. NAKANO and C. W. PRICE, 2001 MHYT, anew integral membrane sensor domain. FEMS Microbiol. Lett. **205:** 17-23.

GEHRIG, S. M., 2005 Adaptation of *Pseudomonas fluorescens* SBW25 to the air-liquid interface: a study in evolutionary genetics, pp. 238 in *Department of Plant Sceinces*. University of Oxford, Oxford.

GERHART, J., and M. KIRSCHNER, 1997 *Cells, Embryos, and Evolution*. Blackwell, Malden.

GIDDENS, S. R., R. W. JACKSON, C. D. MOON, M. A. JACOBS, X. X. ZHANG *et al.*, 2007 Mutational activation of niche-specific genes provides insight into regulatory networks and bacterial function in a complex environment. Proc. Natl. Acad. Sci. USA **104:** 18247-18252.

GOULD, S. J., 2002 *The Structure of Evolutionary Theory*. Belknap Press, Cambridge MA.

GOYMER, P., S. G. KAHN, J. G. MALONE, S. M. GEHRIG, A. J. SPIERS *et al.*, 2006 Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. II. Role of the GGDEF regulator WspR in evolution and development of the wrinkly spreader phenotype. Genetics **173:** 515-526.

HANSEN, T. F., 2006 The evolution of genetic architecture. Annu. Rev. Ecol. Syst. **37:** 123-157.

HORTON, R. M., H. D. HUNT, S. N. HO, J. K. PULLEN and L. R. PEASE, 1989 Engineering hybrid genes without the use of restriction enzymes - gene-splicing by overlap extension. Gene **77:** 61-68.

HOULE, D., 1998 How should we explain variation in the genetic variance of traits? Genetica **102-103:** 241-253.

HUDSON, E. R., U. BERGTHORSSON, J. R. ROTH and H. OCHMAN, 2002 Effect of chromosome location on bacterial mutation rates. Molecular Biology and Evolution **19:** 85-92.

JENAL, U., and J. MALONE, 2006 Mechanisms of cyclic-di-GMP signaling in bacteria. Annu. Rev. Genet. **40:** 385-407.

KIMURA, M., 1968 Evolutionay rate at the molecular level. Nature **217:** 624-626.

KING, E. O., M. K. WARD and D. C. RANEY, 1954 Two simple media for the demonstration of pyocyanin and fluorescin. J. Lab. Clin. Med. **44:** 301-307.

KNIGHT, C. G., N. ZITZMANN, S. PRABHAKAR, R. ANTROBUS, R. DWEK *et al.*, 2006 Unravelling adaptive evolution: how a single point mutation affects the protein co-regulation network. Nat. Genet. **38:** 1015-1022.

KRONFORST, M. R., D. D. KAPAN and L. E. GILBERT, 2006 Parallel genetic architecture of parallel adaptive radiations in mimetic *Heliconius* butterflies. Genetics **174:** 535-539.

LANDRY, C. R., B. LEMOS, S. A. RIFKIN, W. J. DICKINSON and D. L. HARTL, 2007 Genetic properties influencing the evolvability of gene expression. Science **317:** 118-121.

LENSKI, R. E., M. R. ROSE, S. C. SIMPSON and S. C. TADLER, 1991 Long-Term Experimental Evolution in *Escherichia-Coli*. I. Adaptation and Divergence During 2,000 Generations. Am. Nat. **138:** 1315-1341.

MALONE, J. G., R. WILLIAMS, M. CHRISTEN, U. JENAL, A. J. SPIERS *et al.*, 2007 The structure-function relationship of WspR, a response regulator with a GGDEF output domain. Microbiol. **153:** 980-994.

MANOIL, C., 2000 Tagging exported proteins using *Escherichia coli* alkaline phosphatase gene fusions. Method Enzymol **326:** 35-47.

MAYNARD SMITH, J., R. BURIAN, S. KAUFFMAN, P. ALBERCH, J. CAMPBELL *et al.*, 1985 Developmental constraints and evolution. Q. Rev. Biol. **60:** 265-287.

MCGREGOR, A. P., V. ORGOGOZO, I. DELON, J. ZANET, D. G. SRINIVASAN *et al.*, 2007 Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature **448:** 587-590.

OSTROWSKI, E. A., R. J. WOODS and R. E. LENSKI, 2008 The genetic basis of parallel and divergent phenotypic responses in evolving populations of *Escherichia coli*. Proc. R. Soc. B **275:** 277-284.

PIGEON, D., A. CHOUINARD and L. BERNATCHEZ, 1997 Multiple modes of speciation involved in the parallel evolution of sympatric morphotypes of lake whitefish (*Coregonus clupeaformis*, Salmonidae). Evolution **51:** 196-205.

RAINEY, P. B., 1999 Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. Environ. Microbiol. **1:** 243-257.

RAINEY, P. B., and M. J. BAILEY, 1996 Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. Mol. Microbiol. **19:** 521-533.

RAINEY, P. B., and K. RAINEY, 2003 Evolution of cooperation and conflict in experimental bacterial populations. Nature **425:** 72-74.

RAINEY, P. B., and M. TRAVISANO, 1998 Adaptive radiation in a heterogeneous environment. Nature **394:** 69-72.

ROSS, P., H. WEINHOUSE, Y. ALONI, D. MICHAELI, P. WEINBERGEROHANA *et al.*, 1987 Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. Nature **325:** 279-281.

RUTHERFORD, K., J. PARKHILL, J. CROOK, T. HORSNELL, P. RICE *et al.*, 2000 Artemis: Sequence visualisation and annotation. Bioinformatics **16:** 944-945.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbour Laboratory, New York, New York.

SCHLICHTING, C. D., and C. J. MURREN, 2004 Evolvability and the raw material for adaptation, pp. 18-29 in *Plant Adaptation: Molecular Genetics and Ecology*, edited by Q. C. B. CRONK, J. WHITTON, R. H. REE and I. E. P. TAYLOR. NRC Research Press, Ottawa.

SCHLUTER, D., 1996 Adaptive radiation along genetic lines of least resistance. Evolution **50:** 1766-1774.

SCHURR, M. J., H. YU, J. M. MARTINEZ-SALAZAR, J. C. BOUCHER and V. DERETIC, 1996 Control of AlgU, a member of the sigma E-like family of stress sigma factors, by the negative regulators MucA and MucB and *Pseudomonas aeruginosa* conversion to mucoidy in cystic fibrosis. J. Bacteriol. **178:** 4997-5004.

SHINDO, C., M. J. ARANZANA, C. LISTER, C. BAXTER, C. NICHOLLS *et al.*, 2005 Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. Plant Physiol **138:** 1163-1173.

SILBY, M. W., A. M. CERDENO-TARRANGA, G. VERNIKOS, S. R. GIDDENS, R. W. JACKSON *et al.*, 2009 Genomic and functional analyses of diversity and plant interactions of *Pseudomonas fluorescens*. Genome Biology **Accepted pending revision**.

SPIERS, A. J., J. BOHANNON, S. M. GEHRIG and P. B. RAINEY, 2003 Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. Mol. Microbiol. **50:** 15-27.

SPIERS, A. J., S. G. KAHN, J. BOHANNON, M. TRAVISANO and P. B. RAINEY, 2002 Adaptive divergence in experimental Populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. Genetics **161:** 33-46.

SPIERS, A. J., and P. B. RAINEY, 2005 The *Pseudomonas fluorescens* SBW25 wrinkly spreader biofilm requires attachment factor, cellulose fibre and LPS interactions to maintain strength and integrity. Microbiol. **151:** 2829-2839.

STERN, D. L., and V. ORGOGOZO, 2008 The loci of evolution: how predictable is genetic evolution? Evolution **62:** 2155-2177.

STERN, D. L., and V. ORGOGOZO, 2009 Is genetic evolution predictable? Science **323:** 746-751.

SUGAWARA, E., and H. NIKAIDO, 2004 OmpA/OprF: Slow porins or channels produced by alternative folding of outer membrane proteins, pp. 119-138 in *Bacterial and Eukaryotic Porins: Structure, Function, Mechanism*, edited by R. BENZ. Wiley-VCH, Weinheim.

TAMAYO, R., A. D. TISCHLER and A. CAMILLI, 2005 The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase. J. Biol. Chem. **280:** 33324-33330.

WICHMAN, H. A., M. R. BADGETT, L. A. SCOTT, C. M. BOULIANNE and J. J. BULL, 1999 Different trajectories of parallel evolution during viral adaptation. Science **285:** 422-424.

WOODS, R., D. SCHNEIDER, C. L. WINKWORTH, M. A. RILEY and R. E. LENSKI, 2006 Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. Proc. Natl. Acad. Sci. USA **103:** 9107-9112.

YU, H., M. J. SCHURR and V. DERETIC, 1995 Functional equivalence of Escherichia coli sigma E and *Pseudomonas aeruginosa* AlgU: E. coli rpoE restores mucoidy and reduces sensitivity to reactive oxygen intermediates in algU mutants of *P. aeruginosa*. J. Bacteriol. **177:** 3259-3268.

ZHANG, J. Z., 2006 Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. Nat. Genet. **38:** 819-823.

ZHONG, S., A. KHODURSKY, D. E. DYKHUIZEN and A. M. DEAN, 2004 Evolutionary genomics of ecological specialization. Proc. Natl. Acad. Sci. USA **101:** 11719-11724.

3 THE DISTRIBUTION OF FITNESS EFFECTS OF NEW MUTATIONS.

3.1 INTRODUCTION

The distribution of the fitness effects (DFE) of naturally occurring mutations is of fundamental importance to a complete genetic theory of evolution. The shape of this distribution determines the average size and number of steps taken during an adaptive walk; knowledge of this would allow predictions to be made for organisms adapting to new environments. The paucity of empirical data quantifying this distribution is due primarily to the difficulty of obtaining unbiased samples of mutations. Beneficial mutations of larger effect size are more likely to survive loss by drift when they are rare and to out compete any co-occurring mutations (HALDANE 1924; KIMURA 1983), decreasing the probability of collecting a representative sample of all possible mutations.

The apparent experimental intractability of this problem led Gillespie to develop the mutational landscape model (GILLESPIE 1983; GILLESPIE 1984). He noted that many common distributions have an exponential right hand tail regardless of the shape of the entire distribution. He reasoned that because the typical change in environment experienced by natural populations is small, the ancestral genotype, although displaced as the fittest, is still highly fit and is thus part of the extreme right hand tail of the overall DFE. Under strong selective conditions (as assumed by the model) only beneficial mutations can fix; because these are exponentially distributed certain properties of these genotypes can be adequately described by extreme value theory (section 1.5.2). Accordingly, the mutations comprising the unknown portion of the distribution can be ignored and only those in the right hand tail, whose distribution we know, will contribute to the adaptive walk. As such, tests of the mutational landscape model cannot speak to the nature of the fitness effects of all possible mutations (the DFE), but only a subset of these, beneficial mutations.

The mutational landscape model has been lauded as the model providing the most realistic account of adaptation (ORR 2005; ORR 2006b). Empirical studies have lent support to several of the predictions made by this model (IMHOF and SCHLOTTERER 2001; KASSEN and BATAILLON 2006; ROKYTA *et al.* 2005), although some recent studies have claimed that in some cases the DFE will not have a right hand tail that is exponentially distributed (MARTIN and LENORMAND 2008; ROKYTA *et al.* 2008).

The key assumption of the model is that the typical environmental change is small; therefore the fitness of the most common genotype at the start of an adaptive walk is already very high. This limits the explanatory power of the model as it discounts adaptive walks starting from a low or intermediate fitness, such as those occurring during the evolution of antibiotic resistance and adaptive radiations. For instance, when an antibiotic is introduced to an antibiotic-sensitive culture, it is lethal to most of the population; the ancestral phenotype thus has a fitness of zero in the new antibiotic contaminated environment. Clearly, the mutational landscape model will not be of use here (unless the entire DFE is exponential); the size of steps taken during this adaptive walk will be determined by the overall DFE, not its right hand tail.

A second assumption is that each step of an adaptive walk draws from the same DFE. The first adaptive step consists of the fixation of a genotype; this genotype is one of the ancestor genotype's "one step neighbours"[1] that arise by mutation. The next adaptive step requires the mutation of this newly fixed genotype, the mutational landscape model assumes that the DFE of genotypes produced by mutation of this newly fixed allele is the same as the DFE of genotypes produced by mutation of the ancestor. This is regarded as counter intuitive (ORR 2006b); an adaptive landscape in such a scenario is maximally rugged, with the fitness of each genotype having no

[1] The one step neighbours of genotype 1 are those genotypes which can arise by the occurrence of a single mutational event in genotype 1.

correlation with the fitness of its neighbouring genotypes. The alternative to this is that landscapes are correlated, specifically, that mutation of the new fitter genotype should produce a DFE with an higher average fitness than the DFE of the one step neighbours of a less fit genotype. Despite what may or may not seem intuitive, there is very little evidence indicating the degree to which the fitness of neighbouring genotypes on fitness landscapes are correlated; the validity of this assumption is therefore in question.

In this study a sample of WS mutations unbiased by drift or natural selection is obtained. The distribution of fitness effects of these genotypes is measured with the aim of determining general properties of the overall DFE and testing the validity of assumptions of the mutational landscape model.

3.2 RESULTS

3.2.1 GATHERING AN UNBIASED SAMPLE OF WS MUTATIONS.

The WS reporter construct consists of a promoter less kanamycin resistance (*nptII*) gene fused to the *wss* promoter; *wss* encodes the cellulose biosynthetic machinery and is constitutively active in WS (SPIERS *et al.* 2002; SPIERS and RAINEY 2005), but silent in the ancestral genotype. Upon mutation to WS, transcription of *wss* increases such that a strain carrying a *wss-nptII* fusion becomes kanamycin resistant. The ability to identify WS genotypes on the basis of a switch from kanamycin sensitivity to resistance provides an opportunity to identify WS genotypes in the absence of selection for the WS phenotype *per se*. For example, in a spatially unstructured microcosm (a shaken tube) WS are not selectively favoured, and while presumably arising by mutation, WS genotypes are not detected because of their fitness disadvantage relative to the ancestral type. Use of the *wss-nptII* fusion means that these selectively disadvantaged WS genotypes (which will be a minor component of the population) can be detected by plating cells onto media containing kanamycin. The WS obtained in such a manner represent as close to an unbiased set of WS genotypes as is possible to obtain.

3.2.1.1 ARGUMENT THAT WS WILL NOT BE SELECTED FOR IN SHAKEN MICROCOSMS INCUBATED FOR ONLY 16 HOURS.

A central assumption of the studies described in this chapter as well as chapter four is that the WS collected from shaken microcosms incubated for 16 hrs are not more likely to be recovered that any other WS. It is known that some WS perform better than others in the shaken tube, and so it is possible that if there was more than one WS in a shaken vial, one would be more likely to be recovered than the other. Here I show that it is unlikely that in the short time shaken tubes are most likely to arise very late, and that when the do arise they will not experience selection strong enough to influence their detection.

An overnight culture of the SM^{MSC} WS reporter strain was diluted so that approximately 10^3 cells were used to inoculate 6mls of KB in a shaken glass vial. Tubes inoculated in such a manner were shaken for 16 hrs at 28 degrees Celsius and then plated on KB agar with 30 μg/ml Kanamycin. These conditions were maintained in order to minimise the effects of natural selection on the sample of WS mutations obtained. WS have differences in their fitness in the shaken environment; as such, one could speculate that some WS would be more likely to be recovered than others, thereby biasing the sample of WS mutations towards those that are fitter in the shaken microcosm. The following arguments have been constructed to address the validity of this claim.

1. Assuming a division time of 40 minutes, 24 cell divisions can take place in 16hrs. With a starting population of 10^3 it will take 21 cell divisions to achieve 10^8 per ml of culture, the point after which after log phase growth will cease to be sustainable. Until this point all WS types that arise by mutation will not

experience selection because there will be sufficient resources for all cells to divide, even those that divide at a slower rate.

2. Mutations arise during genome replication, so the greater the number of cells the greater the number of mutations within the population. At any time, half of the total population are daughter cells that arose from the last round of replication; therefore half of all mutations arose in that last division. It follows then that over 99% of mutations in the population have arisen within the last 5 cell divisions. Even if two WS mutations with the potential for differential survival did arise in a single culture, they would be most likely to arise in the late stages. Even if there is strong selection against some WS, the simply need to remain viable for the short time until they are plated.

3.2.1.2 EXPERIMENTS TO EXCLUDE BIAS IN THE COLLECTION OF 100 NON SELECTED WS.

A total of 100 independent Kanamycin resistant mutants were isolated from the progenitor SM^{MSC} strain. In order to demonstrate that our identification method did not lead to spurious non-WS Kanamycin resistant mutants, all clones were inoculated into separate microcosms and incubated statically. In each case robust WS mats were quickly formed, demonstrating that each recovered mutant was a WS. There was also the possibility that not all WS are kanamycin resistant, thus biasing the sample to exclude some WS. Accordingly, in a further control experiment, a set of 100 independent microcosms inoculated with SM^{MSC} were incubated for 36 hours, sufficient time for WS to increase in frequency without requiring detection by plating on kanamycin. All 100 WS

obtained in this manner were kanamycin resistant despite being selected for their WS morphology, not their kanamycin resistance. A further factor that could skew the sample of recovered WS mutants is a correlation between WS fitness and the level of *wss* transcription. If the most (or least) fit WS have a higher rate of *wss* expression, this would manifest as greater kanamycin resistance, meaning that some WS may grow very slowly on kanamycin and their colonies may not be visible when WS are picked from the plate. In order to test this, 31 of the 100 WS were randomly selected and plated at a low dilution on kanamycin plates. The diameter of four colonies for each WS genotype was measured and the average plotted against its fitness value in a static microcosm (as measured in 3.2.1 below).

Regression of fitness in the static environment against colony size revealed no correlation (r^2= 2e-05; $F_{1,7}$=0.278; p=0.614) (Figure 3.1) showing that the fittest WS were not more likely to be detected because of faster colony growth. Based on these results it is reasonable to conclude that the screening method allowed the identification of an unbiased sample of WS mutants.

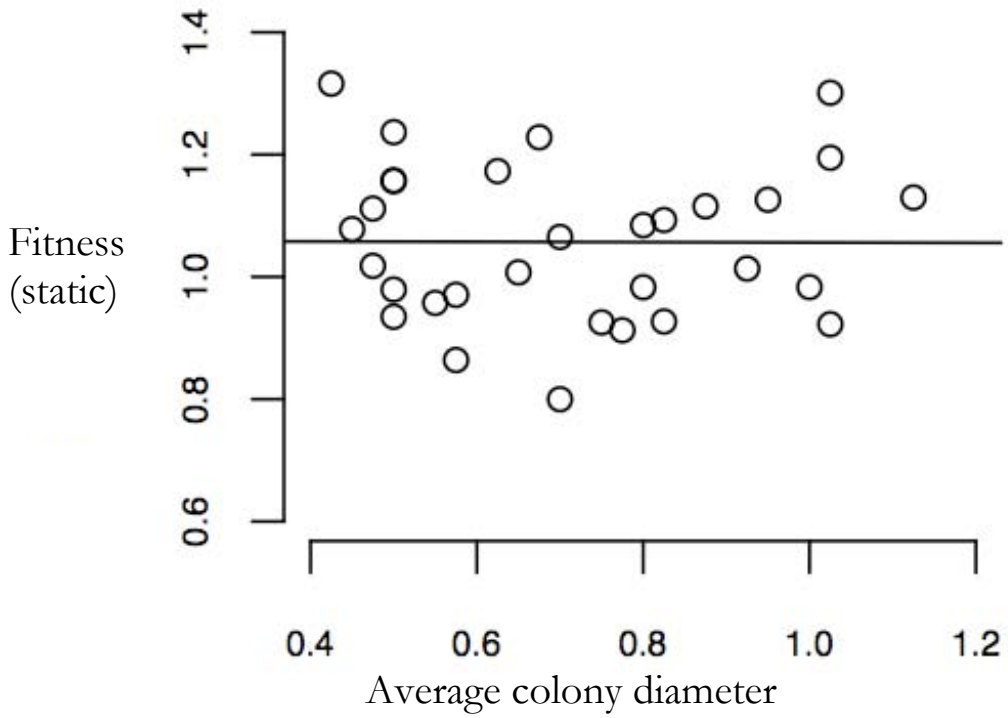


Figure 3.1

Average colony diameter was regressed against fitness in the static microcosm. The straight line was fitted to the data with R^2= 2e-05.

An important condition for the veracity of studies measuring the fitness effects of single mutations is that the measured fitness effects are indeed due to a single mutation. Previous studies have attempted to increase the likelihood that only a single mutation is present in the genotype of interest by recovering clones after a short incubation; thereby minimising the number of mutations that could have occurred. In order to ensure that the sample used in this study consisted of clones with only a single mutation, the WS mutants were recovered after short period of incubation (16 hours).

3.2.2 FITNESS OF THE 100 WS

Fitness measurements were obtained by competing each WS against a marked reference strain, the LSWS*lacYZ*, in the static microcosm. As each of the 100 WS genotypes contained the MSC WS reporter construct, the LSWS*lacYZ* was also transformed with this construct, to avoid any cost conferred by carrying the construct biasing the fitness assays. All 100 WS fitness's were measured in a single block, with two replicates, for each assay care was taken with plating densities so that at least 30

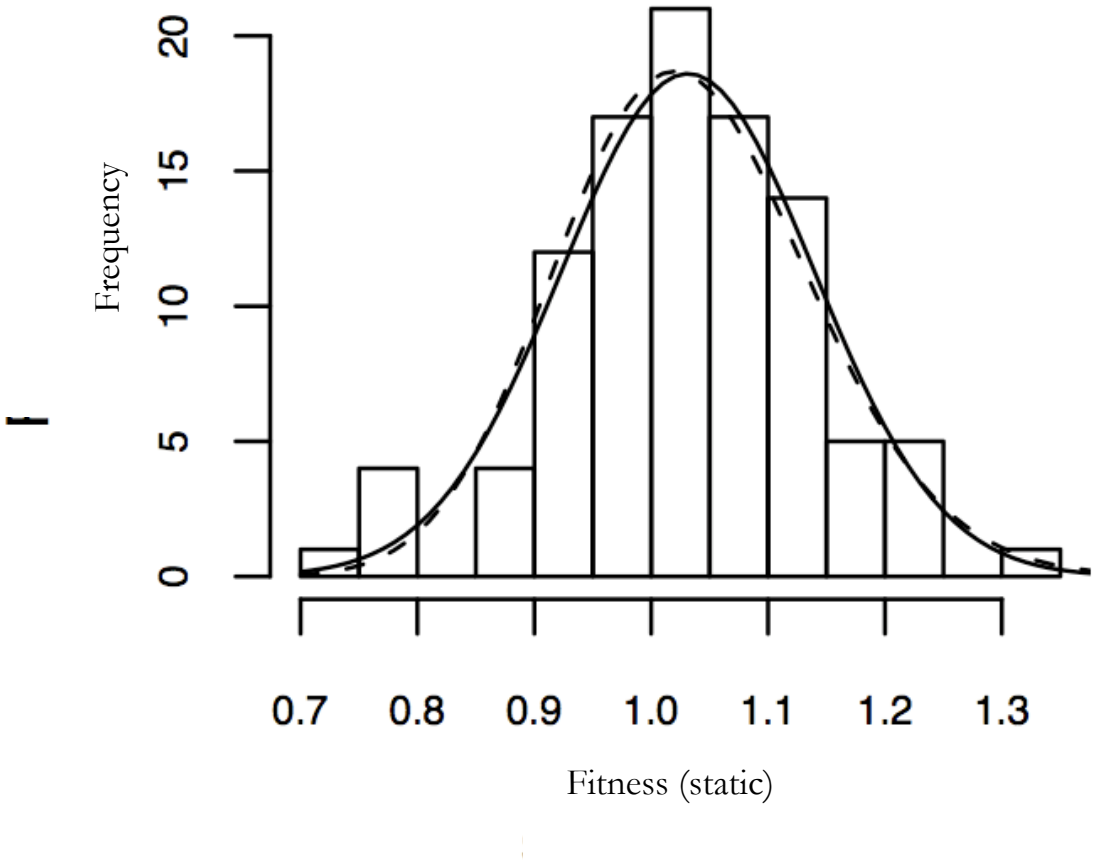


Figure 3.2

The gamma and exponential distributions were rejected in favour of the normal distribution; the comparison of theoretical and sample quantiles shows the goodness of fit (shown in appendix 8.4).

colonies were counted for each competitor; minimising random effects caused by low counts. In previous studies fitness assays of large collections of clones have been carried out by comparing initial and final densities of cell cultures with only a single replicate, with between group controls when all samples cannot be measured as a block (KASSEN AND BATAILLON, 2006). The employing of biological replicates and a competitive fitness assay is an improvement of these standards. An analysis of variance revealed significant variance among strains ($F_{100,201}$= 1.632; p= 0.007). Maximum likelihood methods were used to compare the observed distribution to the gamma, exponential and normal distributions. The distribution of beneficial mutations under the mutational landscape model is predicted to be exponential, yet the normal distribution was found to better fit the data than both the exponential and gamma (for maximum likelihood table see appendix 8.4.2). This may be because the experiment violates an important assumption of the mutational landscape model; the fitness of the *P. fluorescens* ancestor is very low at the air-liquid interface. The WS niche is the air-liquid interface, the wild type cannot colonise this niche and so effectively has a fitness of zero. This violation of mutational landscape model parameters has interesting implications for what this set of 100 WS mutations represents. The set of beneficial mutations is a subset of all possible mutations, specifically those conferring a higher fitness than the currently most common genotype. As the fitness of the current genotype approaches zero, the distribution of beneficial mutations and the overall set of mutations conflates; if the fitness of the founding genotype is zero, then all viable genotypes are beneficial. The result of this is that the 100 WS are not only an unbiased sample of beneficial mutations, but also a sample of the overall DFE.

Previous estimates have found distributions other than the normal to best fit the DFE (ELENA *et al.* 1998; SANJUAN *et al.* 2004); indeed, it is possible that the DFE may differ for each trait and environment under consideration. In order to test if the normal

distribution was specific to this experiment the DFE for the same 100 genotypes was determined in a different environment; one in which the performance of each genotype was not correlated with its fitness in the first (the static glass tube). If the normal distribution is a universal descriptor of the DFE, then when measured in different environments the genotypes should re-assort into the normal distribution even though they would have different ranks. The performance of each genotype must be assayed in an environment where the fitness of any genotype in the first environment is uncorrelated with its performance in the second. To test the universality of the normal distribution, the fitnesses of the 100 WS were measured in an unstructured environment (a shaken glass tube). It is known that WS are significantly less fit than the ancestral type in a shaken vial; this environment favours the ancestral type over most WS because WS mutations come at the cost of reduced carbon catabolism. It has been shown that secondary mutations can ameliorate the effects of the WS mutations, indicating that the costs associated with WS mutations are not the cause of adaptation (MACLEAN *et al.* 2004). This experiment points to the issue of importance in the comparison of these fitness assays; whether or not the fitness of a WS in a static tube is indicative of its fitness in a shaken tube. In another study, a comparison of the fitnesses of 26 unique WS in both the shaken and static environment found no correlation between the performances of the same genotype in the two environments (MEINTJES 2008) (figure 3.3).

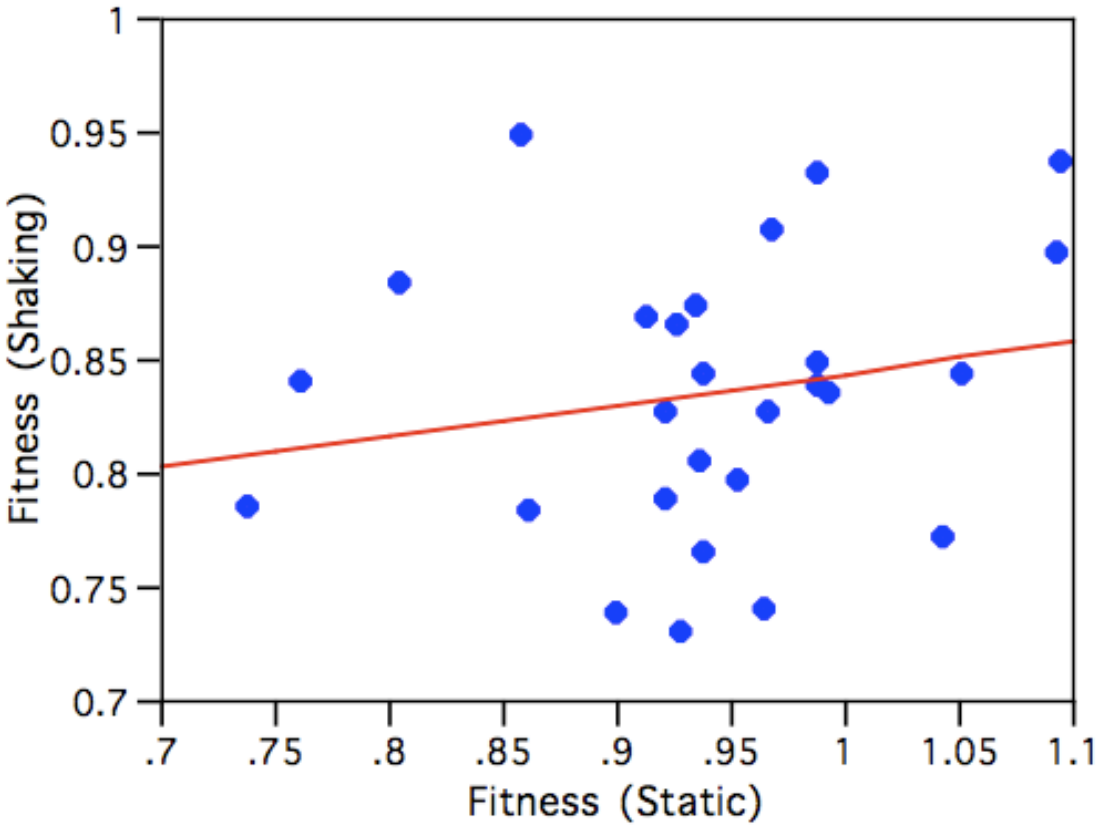


Figure 3.3

A straight line was fit to the data with r^2=0.037 ($F_{1,25}$ = 0.974, p = 0.333). Genotypes that are further from the line of best fit are relatively fitter in one environment that the other (figure taken from Meintjes, 2008).

The fitness of all 100 genotypes obtained without selection were measured, and the DFE of the 100 WS genotypes in this alternative (shaking tube) environment were found to be best described by the normal distribution. The significant lack of correlation in the rank order of genotype fitnesses measured in the different environments (Spearman's rank correlation, S = 165166, p-value = 0.7051) further supports that the fitness's of WS in shaken and static vials are not correlated; the repeated finding of the normal distribution is not due to the similarity of genotypes fitnesses between the two environments.

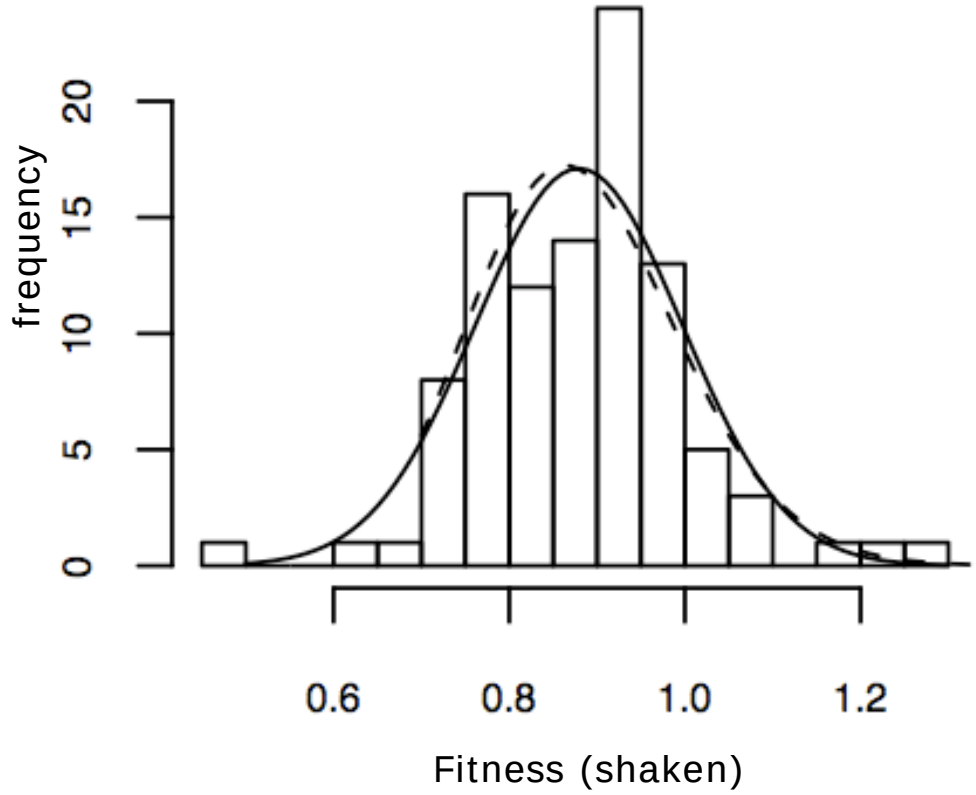


Figure 3.4

The curve was fit using maximal likelihood methods, the normal distribution was found to best fit the measured DFE. The comparison of theoretical and sample quantiles is shown in appendix 8.4.

3.2.2 DETERMINING THE UNIQUENESS OF CLONES WITHIN THE 100 WS

A weakness of previous studies of beneficial mutants has been the lack of knowledge of the causal mutations. Variable per nucleotide mutation rates can lead to many of the mutations in the collection not being unique. We sequenced 20 of the 100 WS at the four genes known to sustain most WS mutations; *wspF*, *awsX, awsR* and *mwsR*. We found 13

of the 20 mutations in these genes, 5 of these 13 (or 38.5%) mutations were not unique[2]. This degree of homogeneity does not diminish the power of the conclusions reached in this chapter, as unlike other studies we do not focus on individual mutations but on the distribution of mutational effects.

No[a]	Nucleotide change	Gene
1	C608A	*wspF*
3	DEL3066-3074	*mwsR*
6	Del 228-261	*awsX*
8	DEL3068-3076	*mwsR*
9	DEL3066-3074	*mwsR*
11	Del 228-261	*awsX*
12	Del 99-138	*awsX*
13	C160G	*awsR*
15	Del 140-185	*awsX*
16	Del 228-261	*awsX*
17	Del 155-173	*awsX*
18	Del 99-138	*awsX*
20	Del 99-138	*awsX*

Table 3.1.

The mutations found by sequencing WS isolated from microcosms 1-20 of the 100 WS genotypes used in this study. [a]13 mutations were found, the number refers to which of the 20 microcosms this genotype was isolated from.

[2] These mutations are also used in a study described in chapter four (4.2), in the day zero set of mutations.

3.3 DISCUSSION

The mutational landscape model of adaptation considers instances for which beneficial mutations are a small subset of the entire DFE. One of the novel insights of the work described in this chapter is to interpret the distribution of the 100 WS beneficial mutations as equivalent to the overall DFE. This is possible because the ancestor cannot live in the niche colonised by the WS, all the viable genotypes that arise by mutation of the ancestral genotype in this case are beneficial, and so the DFE for all genotypes and that of beneficial mutations is equivalent. The fitnesses of all 100 genotypes were measured in static microcosms and the gamma and exponential distributions were rejected in favour of the normal; in order to test if this distribution was specific to this trait in the static environment the fitnesses of all 100 WS genotypes were measured in shaken vials also. In both cases the normal distribution best fit the data, despite the performance of genotypes in the shaken tube not being correlated with performance in the static tube. It has been predicted by theoretical study of several models of adaptation that the DFE is normal. Here, I apply our knowledge of WS genetics to present a biological basis for the normal distribution as a universal descriptor of the DFE.

3.3.1 THE NORMAL DISTRIBUTION AND MODULAR SYSTEMS

A robust prediction of Fisher's and other models of adaptation is that the DFE is normally distributed (KAUFFMAN and LEVIN 1987; ORR 2006a; ORR 2006b; PERELSON and MACKEN 1995). These predictions are all dependant on the assumption that the organism under consideration has a large number of weakly correlated traits each independently contributing towards overall fitness. This assumption has been considered

unrealistic, although it is hard to see why; almost no empirical studies have been able to shed light on the degree of independence of traits or how many any organism has.

Complex organisms, like any other, require mutations in order to adapt, however if a mutation has a beneficial effect on one trait, the mutation is likely to have a deleterious effect on other traits (FISHER 1918). This problem is overcome by separating gene and protein networks into modules; mutations have a large effect on the module in which they occur but little effect on other modules. Organisms are postulated to have evolved modular systems to limit the detrimental pleiotropic effects of mutations on the organism as a whole (HANSEN 2003; RAINEY and COOPER 2004; WAGNER and ALTENBERG 1996; WAGNER *et al.* 2008). The pleiotropic changes effected by WS mutations are some of the best characterised of any naturally arising adaptive mutations (KNIGHT *et al.* 2006; MACLEAN *et al.* 2004). Comparison of the protein expression profiles of seven unique WS genotypes led to the conclusion that WS mutations effect a core module of co-regulated genes (KNIGHT *et al.* 2006). Importantly, although all WS genotypes affect the same core set of genes, each also had an effect on a small number of genes outside of the module, indicating that the WS module is not totally independent of, but instead weakly correlated with other modules.

The importance of the independence of traits pertains to the central limit theorem (CLT), the statistical phenomenon that leads to the emergence of the normal distribution. The central limit theorem considers a set of variables that could form any distribution or compound distribution. A random sample of these variables is taken, and the average calculated; this value is a single data point. If many data points are generated in this manner, then the CLT predicts that the distribution of these data points will be normal; regardless of the underlying distribution from which variables are randomly drawn. An organism can be considered as having many modules, each corresponding to a trait; the fitness values of each of these traits is averaged to give the overall fitness of

the organism. Each genotype is thus analogous to a data point obtained by averaging a sample of random variables. As such, the distribution of 100 of these genotypes (or averages) will be normal. This can be so under the conditions of weak correlation between modules described above; the genetic change that distinguishes each of the 100 genotypes will mainly influence a single module (the WS module), but each genotype will have some random effects on a number of other modules. The fitness of each genotype is the average of this sample of modules, because each mutation affects different modules in a different manner, the fitness value of each genotype is different.

3.3.2 THE NORMAL DISTRIBUTION AND THE FITNESS LANDSCAPE.

Each step of an adaptive walk considers a new set of genotypes. The mutational landscape model assumes that these genotypes are drawn from the same DFE. Under this assumption the fitness landscape is perfectly uncorrelated. Smooth (correlated) landscapes predict that genotypes differing by a single mutation should have correlated fitness's. This means that genotypes that are next to each other in genotype space should have more similar fitnesses than more genetically different genotypes. In this case, the DFE should have a mean skewed either towards (a positive correlation) or away from (a negative correlation) the fitness value of the ancestor. This is because if the ancestor has a low fitness, then most the 100 genotypes should also have a low fitness, after all, they all differ by only a single mutation from the ancestral genotype. The distributions presented in this study suggest this is not the case. The peak of a normal distribution is by definition precisely in between the maximum and minimum fitness values. In the static tube the fitness of the ancestor is zero, and in the shaken environment the ancestor has fitness near the maximum. In both environments the ancestral fitness value is at the very extreme of the DFE of its one step neighbour genotypes, with no correlation

between the ancestral and one-step neighbour fitness. That the genotypes produced by single mutations are neither skewed towards a high or low fitness demonstrates that they are neither positively nor negatively correlated with the fitness of their founding genotype. The lack of correlation between the fitness of a genotype and its one step neighbour genotypes supports the conjecture that fitness landscapes are indeed highly uncorrelated.

The apparent highly uncorrelated nature of the local adaptive landscape may be specific to the WS system, however empirical studies of adaptive walks consistently find that the first mutation often accounts for the largest increase in fitness (COOPER and LENSKI 2000; DE VISSER and LENSKI 2002; HOLDER and BULL 2001; LENSKI and TRAVISANO 1994; NOVELLA *et al.* 1995). This robust empirical result is counter-intuitive considering the vastness of genotype space. The proportion of genotype space explorable within one mutational step is tiny compared to that explored in three steps. For a gene 1000 nucleotides in length, one mutational step can access 3000 genotypes; any given nucleotide could mutate into one of the three alternative nucleotides. The accessible genotype space in three steps is (3000 x 3000 x 3000) 27 billion genotypes. That these mutations of large effect are available within one mutational step suggest that each step of the entire adaptive walk samples from a similar range of genotypic fitness's, even if landscapes are not perfectly rugged as assumed by the mutational landscape model (PERELSON and MACKEN 1995). On correlated landscapes this would not be possible, low fitness genotypes would be most likely to have low fitness one step mutant neighbours and adaptation would move in small steps (as predicted by Fisher for different reasons, described in section 1.5).

3.3.3 OTHER EMPIRICAL SUPPORT FOR THE GENERALITY OF THE NORMAL *DFE*.

The possible link between the normal *DFE* and organismal complexity requires confirmation by experiments in different environments and organisms. Two experiments have sets of mutations relatively untouched by natural selection. Kassen and Bataillon (KASSEN and BATAILLON 2006) obtained a set of nalidixic acid resistant *P. fluorescens* mutants while minimising the effects of selection. The fitness effects of these 665 mutants were measured in both the permissive and selective environment. Although it was not the focus of their analysis, they found the DFE to be superficially normal. The data set in Kassen and Bataillon's study was similar in many ways to that presented in this chapter, except that no sequence data were available. Despite this, they did not interpret the data as it has been interpreted here, and instead focused on 18 of the nalidixic acid resistant mutants that were fitter than the ancestor in the permissive environment.

Sanjuan and co-workers (SANJUAN *et al.* 2004) generated a set of random mutations by site directed mutagenesis in a RNA virus, and measured their fitness's. They found the DFE to fit the lognormal distribution- a distribution known to arise when a given variable is determined by multiplicative small effects. They went on to suggest that because of the compact genomes and overlapping functions of genes with in RNA viruses, single nucleotide changes may have strong pleiotropic effects through out the viral genome. This explanation illustrates an important difference between viral and bacterial models, that is, viruses are not complex enough to be modular. These contrasting results support the conclusion reached in this study, that the nature of the distribution of fitness effects may be solely determined by organismal complexity.

3.3.4 CONCLUDING COMMENTS

If each step of an adaptive walk in a complex organism can be characterised as drawing from the same normal *DFE*, this would facilitate the modelling of adaptation from a low starting fitness, thus overcoming the main shortcoming of the current mutational landscape model. The results here complement rather than contradict the application of EVT in the mutational landscape model, adding a new statistical framework to extend its reach. The finding that the fitness landscape is highly rugged suggests that the parameters of the mutational landscape model are sound, but restrictive assumptions can be removed as the nature of the variation presented to natural selection is better understood.

3.4 BIBLIOGRAPHY

COOPER, V. S., and R. E. LENSKI, 2000 The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature **407:** 736-739.

DE VISSER, J., and R. E. LENSKI, 2002 Long-term experimental evolution in *Escherichia coli*. XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. BMC Evol. Biol. **2:** 19.

ELENA, S. F., L. EKUNWE, N. HAJELA, S. A. ODEN and R. E. LENSKI, 1998 Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. Genetica **103:** 349-358.

FISHER, R. A., 1918 The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh **52:** 399-433.

GILLESPIE, J. H., 1983 Some properties of finite populations experiencing strong selection and weak mutation Am. Nat. **121:** 691-708.

GILLESPIE, J. H., 1984 Molecular evolution over the mutational landscape. Evolution **38:** 1116-1129.

HALDANE, J. B. S., 1924 A Mathematical theory of natural and artificial selection, part 1. Transactions of the Cambridge Philosophical Society **23:** 19-23.

HANSEN, T. F., 2003 Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. Biosystems **69:** 83-94.

HOLDER, K. K., and J. J. BULL, 2001 Profiles of adaptation in two similar viruses. Genetics **159:** 1393-1404.

IMHOF, M., and C. SCHLOTTERER, 2001 Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. Proc. Natl. Acad. Sci. USA **98:** 1113-1117.

KASSEN, R., and T. BATAILLON, 2006 Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nature Genet. **38:** 484-488.

KAUFFMAN, S., and S. LEVIN, 1987 Towards a general theory of adaptive walks on rugged landscapes J. Theor. Biol. **128:** 11-45.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

KNIGHT, C. G., N. ZITZMANN, S. PRABHAKAR, R. ANTROBUS, R. DWEK *et al.*, 2006 Unraveling adaptive evolution: how a single point mutation affects the protein coregulation network. Nature Genetics **38:** 1015-1022.

LENSKI, R. E., and M. TRAVISANO, 1994 Dynamics of adaptation and diversification - A 10,000 generation experiment with bacterial populations. Proc. Natl. Acad. Sci. USA **91:** 6808-6814.

MACLEAN, R. C., G. BELL and P. B. RAINEY, 2004 The evolution of a pleiotropic fitness tradeoff in *Pseudomonas fluorescens*. Proc. Natl. Acad. Sci. USA **101:** 8072-8077.

MARTIN, G., and T. LENORMAND, 2008 The distribution of beneficial and fixed mutation fitness effects close to an optimum. Genetics **179:** 907-916.

MEINTJES, P. J., 2008 *Pseudomonas fluorescens* SBW25 WS groups as a model of cooperation. Ph.D Thesis. School of Biological Sciences. Auckland, Auckand.

NOVELLA, I. S., E. A. DUARTE, S. F. ELENA, A. MOYA, E. DOMINGO *et al.*, 1995 Exponential increases of RNA virus fitness during large population transmissions. Proc. Natl. Acad. Sci. USA **92:** 5841-5844.

ORR, H. A., 2005 Theories of adaptation: what they do and don't say. Genetica **123:** 3-13.

ORR, H. A., 2006a The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. J. Theor. Biol. **238:** 279-285.

ORR, H. A., 2006b The population genetics of adaptation on correlated fitness landscapes: The block model. Evolution **60:** 1113-1124.

PERELSON, A. S., and C. A. MACKEN, 1995 Protein evolution of partially correlated landscapes. Proc. Natl. Acad. Sci. USA **92:** 9657-9661.

RAINEY, P. B., and T. E. COOPER, 2004 Evolution of bacterial diversity and the origins of modularity. Res. Microbiol. **155:** 370-375.

ROKYTA, D. R., C. J. BEISEL, P. JOYCE, M. T. FERRIS, C. L. BURCH *et al.*, 2008 Benefical fitness effects are not exponential for two viruses. J. Mol. Evol. **67:** 368-376.

ROKYTA, D. R., P. JOYCE, S. B. CAUDLE and H. A. WICHMAN, 2005 An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nature Genet. **37:** 441-444.

SANJUAN, R., A. MOYA and S. F. ELENA, 2004 The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc. Natl. Acad. Sci. USA **101:** 8396-8401.

SPIERS, A. J., S. G. KAHN, J. BOHANNON, M. TRAVISANO and P. B. RAINEY, 2002 Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. Genetics **161:** 33-46.

SPIERS, A. J., and P. B. RAINEY, 2005 The *Pseudomonas fluorescens* SBW25 wrinkly spreader biofilm requires attachment factor, cellulose fibre and LPS interactions to maintain strength and integrity. Microbiology-(UK) **151:** 2829-2839.

WAGNER, G. P., and L. ALTENBERG, 1996 Perspective: complex adaptations and the evolution of evolvability. Evolution **50:** 967-976.

WAGNER, G. P., J. P. KENNEY-HUNT, M. PAVLICEV, J. R. PECK, D. WAXMAN *et al.*, 2008 Pleiotropic scaling of gene effects and the 'cost of complexity'. Nature **452:** 470-U479.

4 THE CAUSES AND CONSEQUENCES OF THE BIASED PRODUCTION OF WS VARIATION.

4.1 INTRODUCTION

In chapter two of this thesis it was demonstrated that genes vary in their capacity to convert DNA sequence change into phenotypic variation. Mutations causing the WS phenotype were always found in one of three DGC encoding loci; *wsp*, *aws* or *mws*. Only when these three loci were deleted did WS genotypes arise by mutation in one of the other 36 DGC-encoding loci contained in the *P. fluorescens* genome. Genes within the *aws*, *wsp* and *mws* loci also varied in their propensity to convert mutation into the WS phenotype; the most frequent sites of WS causal mutations were the genes encoding MwsR (three genotypes) and the negative regulators WspF (13 genotypes) and AwsX (three genotypes) (table 2.1), setting them apart from other *wsp* and *aws* genes. Loss of function mutations are more common than those conferring a gain of function (KEIGHTLEY and LYNCH 2003); more WS genotypes are caused by mutation in *awsX* and *wspF* than other *aws* and *wsp* genes because only in these genes are loss of function mutations sufficient to cause the WS phenotype. All WS-causing mutations in *awsX* and *wspF* are loss of function mutations, facilitating the comparison of the capacity of these two genes to produce WS causing mutations; *mwsR* is quite different. Unlike the *aws* and *wsp* pathways, *mws* is a single gene locus; *mwsR* encodes all the enzymatic functions required of a signal transduction pathway. As such, mutations in different locations within *mwsR* cause the WS phenotype different ways. Two of the three WS causing mutations found in *mwsR* were in the PDE encoding EAL domain, these mutations are similar to those causing a loss of function in AwsX or WspF because they result in a loss of the negative regulatory function of the PDE domain of MwsR (see section 2.5.2). WS causing mutations are also found in the signal transduction domains and c-di-GMP synthesizing domain of MwsR; that the observed number of *mwsR* mutations isolated from the 26 independent WS is comparable to *awsX* is thus due to a combination of its larger size (3.5kb

compared to ~0.5kb for *awsX* and ~1kb for *wspF*) and number of functional domains as well as the capacity for loss of function mutations within its PDE domain to confer the WS phenotype.

The work reported in Chapter two showed that genetic constraints do bias the production of variation, but that this can be understood in light of knowledge of the particular functionalities and regulatory connections of the components involved. While a significant step forward, there is room for a yet more detailed understanding of the factors leading to biases in the production of genetic variation and to find how these biases interact with natural selection and drift to produce evolutionary outcomes. Consider two genetic loci within a single organism; locus A, for which there is a single possible mutation that confers an adaptive phenotype (phenotype X) and locus B for which there are 100 such mutations- each capable of generating phenotype X. Assuming that the per nucleotide mutation rate at the two loci is identical, then the adaptive phenotype is 100 times more likely to arise as a consequence of mutation at locus B compared to locus A: mutations at locus B are therefore most likely to go to fixation. Next consider the possibility that the rate of mutation at locus A is 100 times that at locus B. Under such circumstances the likelihood that the phenotype of interest arises as a result of a mutation in locus A or B is approximately the same. Thus far I have assumed that the selective value of mutants in locus A or B is equivalent, but if this unrealistic assumption is relaxed then predicting the mutational origins of phenotype X would require knowledge of fitness effects, combined with knowledge of the mutation rate per locus, and knowledge of the rate at which DNA sequence variation is translated into phenotypic variation.

The example above shows how the capacity of different genes to produce adaptive mutation, variation in the per-nucleotide mutation rate and the fitness differences between mutations can influence which alleles will eventually fix in a population. Just as for the hypothetical loci above, *awsX* and *wspF* differ from other genes in their capacity to produce adaptive mutation. We know why this is; both proteins normally repress the WS phenotype.

Mutations inactivating these proteins therefore result in constitutive derepression of DGC proteins, leading to overproduction of c-di-GMP and the WS phenotype. Assuming the rate of mutation is invariant among genes, and given that loss of function mutations are more common than gain of function mutations (KEIGHTLEY and LYNCH 2003), it follows that a larger proportion of mutations in *wspF* and *awsX* will be translated into the ecologically useful WS phenotype than mutations in other components of the *wsp*, *aws* or *mws* pathways. According to this reasoning, there should be approximately the same number of mutations in *awsX* as in *wspF*; however the 26 independent WS were found to have 13 *wspF* -generated WS genotypes compared to three found in *awsX*. After adjusting for *awsX* being half the length of *wspF,* six *awsX* mutations (compared to the three observed) are expected among the 26 independent WS. In chapter two it was proposed that the difference between the expected and observed frequencies of *awsX* -generated WS mutations is due to constraints on the types of mutations in *awsX* that can result in the WS phenotype. All WS causing mutations found in *awsX* are in frame, this may be because *awsX* is located immediately upstream (and overlapping) the DGC encoding gene of the *aws* operon (*aws*R). Mutations that alter the reading frame have polar effects on *aws*R that appear to be lethal (section 2.5.2). This is in stark contrast to *wspF*; located downstream and separated by 50 nucleotides from the rest of the *wsp* operon, any type of loss of function mutation is tolerated, resulting in a wider range of mutations that can cause the WS phenotype. While this appears to explain the differences between *awsX* and *wspF*, further aspects of *awsX* mutations are intriguing- the AWS mutation (*ΔawsX* 100-138) and one of the 26 independently obtained WS, WS_T (*ΔawsX* 229-261) (table 2.1), result from a deletion between two identical sequences five nucleotides in length. Mutations of this kind are similar to those that occur at a high frequency in tandem repeat sequences in many bacteria. Because the 26 independently obtained WS were sampled after the influence of 5 days of natural selection, our view of some of

the possible genetic constraints on the production of variation may have been obscured. The likelihood that hidden complexity underlies the production of variation at *awsX* and that insight could be gained by lessening the influence of natural selection suggest a more comprehensive analysis of *awsX* and *wspF* WS mutations is warranted.

In the study described in this chapter I sought a more precise understanding of the genetic influences on the outcome of the *P. fluorescens* adaptive radiation and how they interact with the force of natural selection. To that end I propagated 80 replicate populations in static microcosms and harvested a set of microcosms at a different time points during three days of adaptive radiation of the ancestral genotype. I then obtained the DNA sequence of *awsX* and *wspF*, and measured the competitive fitness of WS containing mutations in these genes. This approach allowed the tracking of the fitness and prevalence of *awsX* and *wspF* WS genotypes during successive days of the *P. fluorescens* adaptive radiation, and facilitates the determination of how genetic biases and natural selection shape the outcome of WS evolution.

4.2 RESULTS

4.2.1 THE EXPERIMENTAL DISSECTION OF AN ADAPTIVE RADIATION.

The ancestral genotype from which all WS genotypes obtained in this study evolved was the wild type *P. fluorescens* modified with a WS reporter construct, SM^{MSC} (FUKAMI *et al.* 2007). The WS reporter construct consists of the *nptII* kanamycin resistance gene fused to the *wss* promotor; upon mutation to the WS genotype the *wss* promoter is up-regulated, conferring resistance to kanamycin and allowing the detection of a single WS genotype among a large population of ancestral SM *P. fluorescens* cells. Three sets of 20 spatially structured (unshaken) microcosms were inoculated with the SM^{MSC} strain. Each set of 20 microcosms was allowed to evolve for either one, two, or three days before a single WS clone was isolated from each microcosm (Figure 3.1). In addition to these, a further set of 20 microcosms was harvested after a 16-hour incubation in shaken (spatially unstructured) microcosms in which WS do not have a selective advantage (RAINEY and TRAVISANO 1998). These 20 "non-selected" microcosms were part of the 100 WS obtained without selection described in chapter three. The 26 sequenced WS described in chapter two were also included in the analysis, providing a set of WS obtained after 5 days incubation.

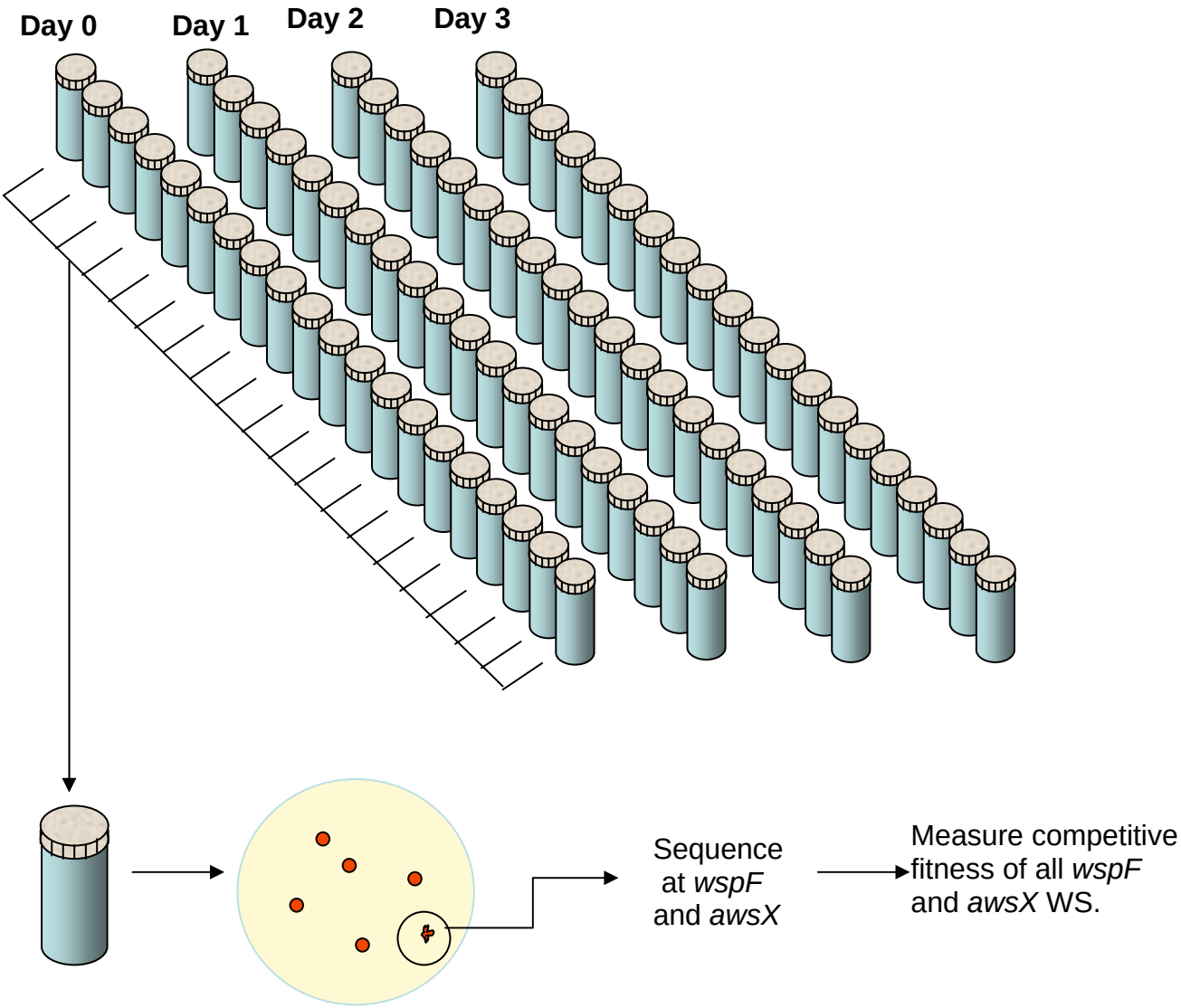


Figure 4.1

Flow diagram showing how WS genotypes were obtained, isolated and analysed in this study. Eighty microcosms in total were harvested for this experiment: 20 at day one, 20 at day two and 20 at day three of incubation in a spatially structured (unshaken) microcosm and 20 after 16 hrs of incubation in spatially unstructured (shaken) conditions. Each microcosm was homogenised by vortexing, plated on selective media and a single WS chosen: the nucleotide sequence of *wspF* and *awsX* was determined for each genotype. For each WS genotype determined by sequencing to possess a *wspF* or *awsX* mutation the competitive fitness of that type was assayed relative to a *lacZ* marked LSWS.

In total, 80 WS clones were recovered, one from each of the 80 microcosms, and the DNA sequences of *wspF* and *awsX* were obtained (the WS genotypes from day 5 are the 26 "independent WS genotypes" whose WS mutations had been determined in chapter two; see Table 4.1). Not all WS genotypes carried a mutation in *awsX* or *wspF*, however, for those that did, the competitive fitness of that genotype was determined relative to the *lacZ* marked LSWS genotype (data presented in Figure 4.2).

Day[a]	WS Mutations	
	awsX	*wspF*
0	8 / 20	1 / 20
1[b]	8 / 20[b]	2 / 20
2	5 / 20	0 / 20
3	1 / 20	4 / 20
5	3 / 26	13 / 26

Table 4.1

The distribution of *awsX* and *wspF* mutations among 106 microcosms incubated for 0, 1, 2, 3 or 5 days. After varying lengths of static incubation, microcosms were destructively harvested, a single WS isolated and then the DNA sequence obtained for both *awsX* and *wspF*. The day five results are from (BANTINAKI *et al.* 2007) and chapter two (see table 2.1). [a]Day: the number of days for which microcosms were statically incubated before WS were harvested. [b]*awsX,* day 1, 8/20: eight of the twenty WS clones recovered from the twenty microcosms incubated after one day of selection had mutations in *awsX*.

As a control, the LSWS genotype was competed against the LSWS::*lacZ*, which by definition has a fitness of one (see methods). Because all evolved strains had been modified with the WS reporter construct, LSWS and LSWS::*lacZ* were similarly modified to carry the *wss*::*nptII* fusion, thus ensuring an absence of bias in the fitness assays due to carriage of genetic markers.

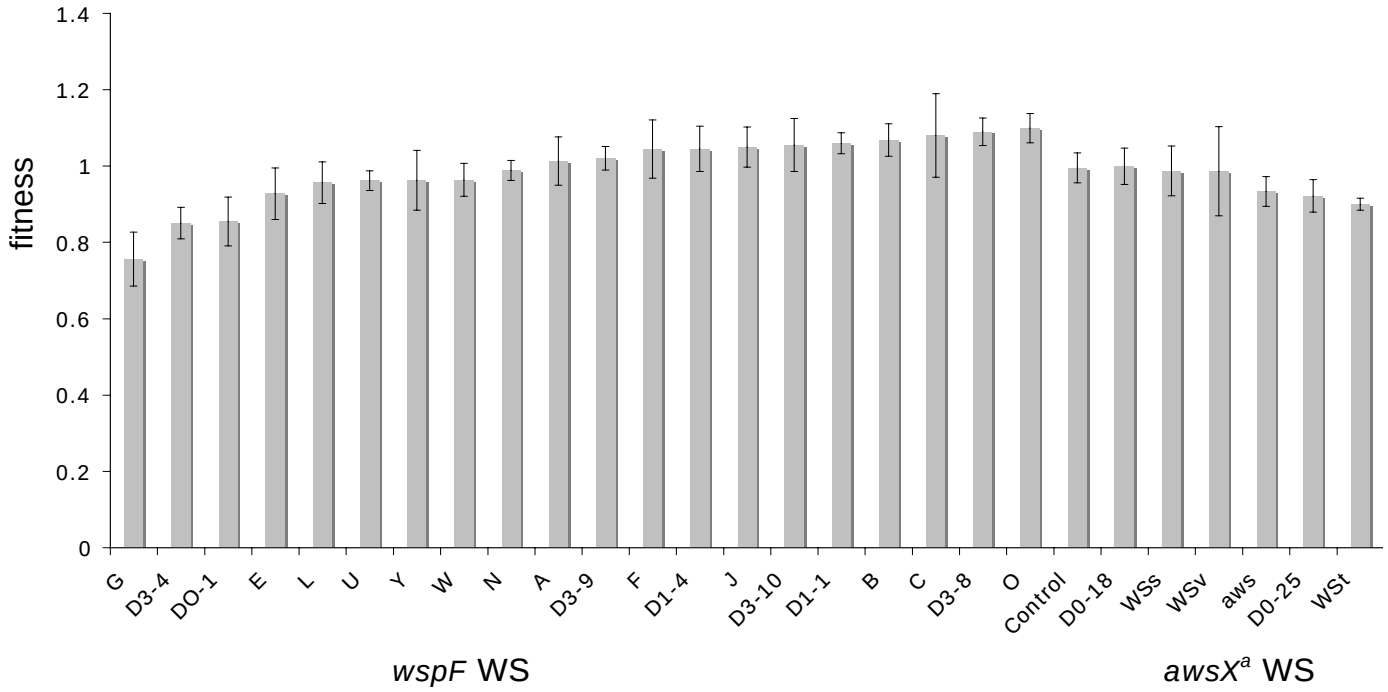


Figure 4.2

Fitness measurements performed in this study for all unique WS genotypes carrying mutations in *awsX* and *wspF*. Each genotype fitness is the average of six replicates, the error bars show 95% confidence intervals. A *t*-test showed that there was no significant difference between the means of the *wspF* and *awsX* WS's ($p = 0.168$). Clones to the left of the control all contain *wspF* mutations, and those to the right mutations in *awsX*. Those WS designated letters (A – Z) are from the 26 independent WS (BANTINAKI *et al.* 2007)), WS obtained in this experiment are named for the number of days for which the are incubated; for example D3-4 is the clone isolated from the fourth out of twenty microcosms incubated for three days. [a]The AWS and WS_T mutation were isolated from multiple microcosms, see table in appendix 8.3.2.

Of the set of 20 microcosms incubated in shaking microcosms for 16 hours (without selection), one WS with a mutation in *wspF* was recovered, while eight microcosms were found to have WS caused by mutations in *awsX*. Table 4.1 shows how the proportion of WS mutations in *awsX* is high at days zero and one, and decreases with increasing length of static incubation. Conversely, *wspF*-generated WS mutations comprise a small

proportion of day zero microcosms, but then are recovered with an increasing frequency from microcosms incubated for a greater length of time.

4.2.2 WS CAUSED BY MUTATION IN *WSPF* COVER THE GREATEST RANGE OF WS FITNESS EFFECTS.

WS genotypes caused by mutation in *wspF* were rarely recovered from microcosms incubated for 16 hours with shaking or from microcosms incubated statically for 24 hours. However, WS genotypes arising due to mutations in *wspF* were found in 13 of the 26 microcosms that were incubated for five days (BANTINAKI *et al.* 2007). WS genotypes caused by mutation in *awsX* exhibited the opposite pattern; many were recovered from microcosms incubated in shaken microcosms (without selection for WS), but were increasingly rare in samples taken from microcosms incubated statically for longer periods of time. These phenomena require explanation.

Fitness effects seem to explain the prevalence of *wspF*-generated WS in microcosms incubated for three or five days; but even though the fittest WS are caused by mutations in *wspF*, the mean fitness of WS with mutations in *wspF* (0.99 n=20) and *awsX* (0.95 n=25) are not significantly different (*t*-test, $p = 0.168$). Because 13 of 20 *wspF* mutations found were obtained from microcosms incubated for five days, and had experienced a longer duration of natural selection, we expect their fitness values to be representative of the fittest *wspF*-generated WS genotypes, but not of all possible *wspF*-generated WS genotypes. The reason why *wspF*-generated WS genotypes eventually become dominant must stem from either the fact that they are inherently more fit (Figure 4.3a), or from the fact that the *wspF*- generated WS have a greater range of fitness effects (Figure 4.3b).

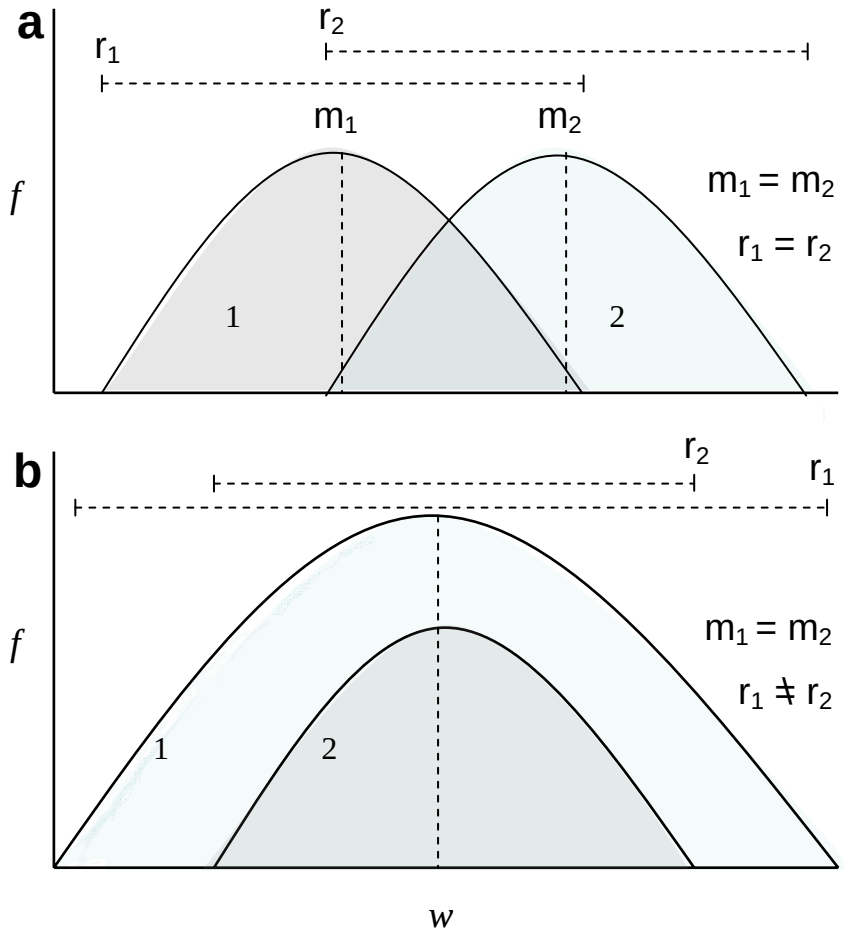


Figure 4.3

Two possible scenarios for the relationship of the distribution of fitness effects of mutations for two genes. (a), mutations in gene 2 are more likely to go to fixation because even the though the range and shape of the distribution of fitness effects are identical, the mean fitness effect of a mutation in gene 2 is higher. (b), alternatively, the two genes have an identical mean, however gene 1 has a wider range of genotype fitness effects. With sufficient mutation supply and under conditions for strong selection alleles from the right hand tails of the distributions will go to fixation, the fittest of which are gene 1 mutants.

To differentiate between these two possibilities, the genotypes with the highest and lowest fitness values for each gene were used to estimate the range of fitness's for WS caused by mutation in *wspF* and *awsX*. The WS genotypes caused by mutations in *wspF* have a much wider range of fitness's (0.342, n=20) than the *awsX*- generated WS (0.1, n=25). Similar to the means of *awsX* and *wspF*- generated WS, the mid points of the range of fitness effects for both genes are more similar; 0.927 (n=20) for *wspF*- generated WS and 0.949 (n=25) for *awsX*- generated WS. These data support the second of the proposed hypotheses (Figure 4.3b); the fittest WS are *wspF*- generated WS because even though they have a mean fitness indistinguishable from *awsX*- generated WS, *wspF*- generated WS genotypes cover a wider range of fitness's. That mutations in *wspF* produces more WS genotypes than *awsX* has already been accounted for (section 2.5.2)*;* the *wspF* gene encodes a negative regulator in which any loss of function mutation suffices to make a WS (BANTINAKI *et al.* 2007). AwsX is also a negative regulator, however, loss of function mutations in this gene appear constrained to be in-frame; missense mutations have severe polar effects negating the benefit of the WS phenotype (section 2.5.2). These differences result in *wspF* being more likely to produce the fittest of all WS genotypes; with increasing length of static incubation natural selection has had more time to act and thus the fittest (usually *wspF*) WS are more frequently isolated from static (spatially structured) microcosms incubated for three or more days.

4.2.3 THE AWS AND WS_T ALLELES ARE PRODUCED AT A HIGHER RATE THAN OTHER WS GENOTYPES.

The organization of the *aws* operon constrains *awsX*- generated WS genotypes to be caused by in frame mutations only. The previous section provides an explanation for how this contributes to the observed low number of *awsX*- generated WS compared to *wspF*- generated WS recovered from microcosms incubated for 3 days or more; however, it does not explain why *awsX*- generated WS comprise such a significant proportion of WS obtained from microcosms incubated (both with and without shaking) for only one day. Most of the *awsX*- generated WS isolated in this study were either a 39 bp deletion between two copies of the sequence AGGCG (*awsX*Δ100-138, isolated 4 times in this study, called AWS) or a 33bp deletion between two copies of the sequence ACCCAG (*awsX*Δ229-261, isolated 16 times in this study, called WS_T). The complete list of mutations is available in appendix 8.3.4.

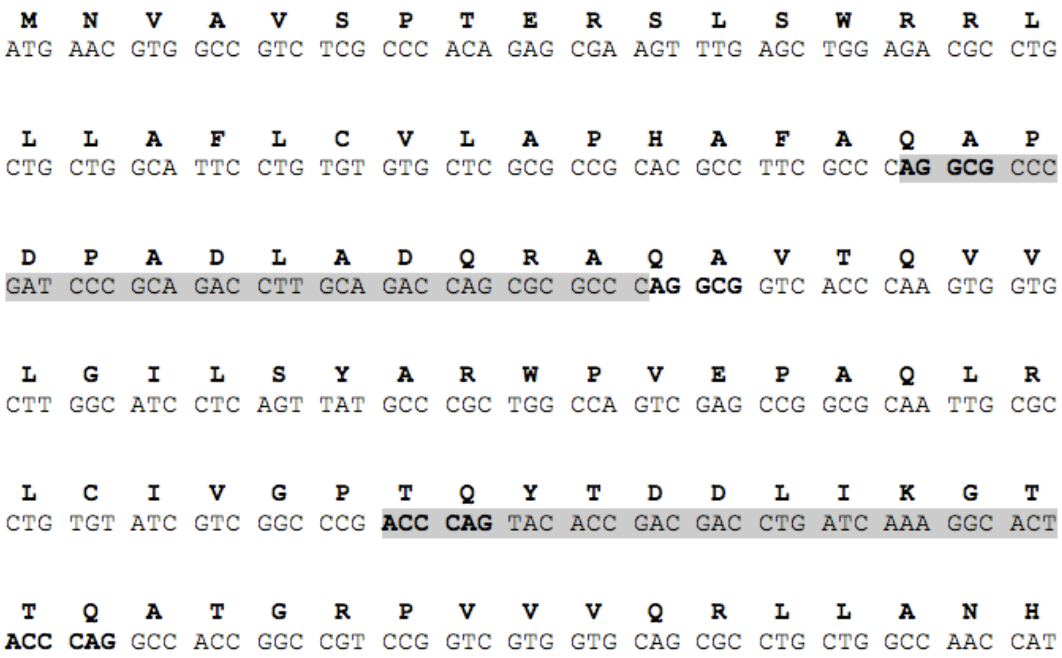


Figure 4.4

Details of the AWS (*awsX*⊗100-138) and WS_T (*awsX*Δ229–261) mutations. The first highlighted region shows the AWS 39 bp deletion, and the second the WS_T 33 bp deletion. Bold type indicates the repeat sequences between which the deletion occurred. These deletions result in a loss of function of AwsX, a cause of the WS phenotype.

Fitness measurements confirm that the AWS and WS_T *awsX* genotypes were not repeatedly sampled from microcosms incubated for one day because of their superior competitive ability; both fall into the lower-intermediate range of overall WS fitness's (AWS = 0.93 and WS_T = 0.89, range of WS fitness's in this study, 0.75 – 1.098). Furthermore, these genotypes are isolated less frequently from microcosms incubated statically for a greater number of days. With an increasing length of incubation fitter genotypes are expected to increase in frequency relative to the less fit; that the AWS and WS_T genotypes are rarely recovered after 3 or more days of static incubation is inconsistent with their early prevalence being due to fitness effects. An alternative

explanation is that the WS causing deletions in *awsX* occur at a high frequency relative to other WS mutations.

The instability of repetitive DNA sequence due to DNA polymerase strand slippage has been well documented (KUNKEL 1990; STREISINGER *et al.* 1966); in bacteria, "contingency" loci arise due to selection for high rates of switching between two phenotypes, each suited to one of the two environmental conditions to which the population is frequently subjected (MOXON *et al.* 1994). The exact type of deletion between repeats found here in *awsX* have yet to be described in bacteria, although deletions of this kind occur in humans (KRAWCZAK and COOPER 1991). It is possible that the deletions occur following DNA replication if, during re-annealing of the two complementary DNA strands, the second repeat of one strand re-anneals with the first repeat on the complementary strand (instead of the complementary second repeat). The ensuing repair reaction will result in the deletion of one of the repeats and the section of DNA between the two repeats (see figure 4.6). The repeats flanking the DNA sequence deleted in the AWS and WS_T WS causing mutations do not conform to classic contingency loci structure, consisting of widely spaced repeats rather than tracts of many tandem repeats (see Figure 4.4). Despite this, the sequence resulting from these deletions is as expected from a strand slippage mechanism; as well as the intervening sequence one of the repeat sequences is also deleted (Figure 4.4).

The inability of fitness effects to explain the high frequency of these *awsX* deletions and the similarity of these sites to known mutable loci in other species led me to consider the possibility that these loci may have a higher relative rate of mutation.

4.2.3.1 COMPARING THE OBSERVED TO EXPECTED DISTRIBUTION OF MUTATIONS UNDER THE ASSUMPTION OF RANDOM MUTATION.

When determining the rates at which different mutations occur, the null hypothesis is that each possible mutation is just as likely to occur as any other. Previous studies have attempted to detect mutational hotspots by asking if the observed number of times any given base substitution occurs deviates from that expected from binomial sampling (eg. LANG and MURRAY 2008). However, detection of variable mutation rates is instead analogous to what is known as the "birthday problem"[1]. The birthday problem considers a group of people and the probability that *any* two of these people have the same birthday. This is distinct from the binomial distribution, which considers what is known as "the same birthday as mine problem"; that *one* of these people has the same birthday as a given person- a much slimmer probability. The incorrect use of binomial sampling may have led studies to incorrectly conclude that the frequency of the occurrence of some mutants exceeded that expected if all mutations occur with an equal frequency.

Above I have determined that the repeated sampling of the *awsX* AWS and WS_T genotypes are not likely to be due to their high fitness, however this possibility can only be excluded for genotypes sampled without the biasing effects of natural selection. The set of WS genotypes obtained from shaking microcosms after 16 hours using the WS reporter construct are such a sample and therefore represent a collection of independently obtained genotypes suitable for statistical analysis. As such, only these genotypes were used to determine if the mutations causing the AWS and WS_T genotypes occurred more often than expected. The application of the birthday problem framework

[1] This formulation of this problem is widely attributed to Harold Davenport, although he denied the honor; he claimed it was so simple some else must have already thought of it.

facilitates the calculation of the probability that any mutation will occur more than once, or if any two mutations occurs more than once, and so on.

A program was written[2] that exploited the birthday problem framework to calculate the probability that any given mutation should occur more than a given number of times. The WS isolated from the non shaking microcosms (without selection) were evolved from the wild type genetic background, so all possible WS mutations were available as candidates as the mutations causing these WS genotypes. There exists a total of 40 unique WS genotypes discovered in published and unpublished experiments in the Rainey laboratory; in WS microcosms incubated for three days or more the same genotype has been recovered only twice. This indicates that there are many more unique WS genotypes than 40, and so 100 is used as a conservative estimate for the total number of WS causing mutations that could possibly occur in the ancestral genotype. In the set of 20 WS obtained without selection, the AWS and WS_T mutations were found three times each. The programme is then asked, if 20 samples are taken randomly from a pool of 100, how many times is it expected that we see any two of those mutations drawn three times. From 10,000 simulations, only 35 times were any two of the 20 individual genotype drawn more than two times, confirming that is significantly unlikely that our observed results were due to chance ($p = 0.0035$).

4.2.4 ELEVATED MUTATION RATES ARE ALSO FOUND IN *MWSR*

Compared to other genes within the *aws*, *wsp* and *mws* loci, *wspF* and *awsX* are predicted to translate DNA sequence change into the WS phenotype at a higher rate per mutational event. Despite this, the sequencing of *wspF* and *awsX* revealed less than half

[2] Program written by Fredrich Bertels, see appendix 8.3.1.2 for details.

of the causal mutations of the WS obtained in this experiment. A possible explanation for this is that one or some of the non-negative regulator genes of *wsp*, *aws* and *mws* have a higher relative mutation rate. Even if these genes have only a few possible mutations that confer the WS phenotype, their relatively high rate of occurrence could have resulted in them in comprising a significant proportion of the WS recovered in this experiment.

Alternatively, the relative amount of gain to loss of function mutations could be higher than commonly thought. The expectation is that as proteins are generally well adapted, the ratio of mutations conferring a loss of function is high relative to those that could contribute a gain of function (KEIGHTLEY and LYNCH 2003). However, although this assumption has been strongly supported by experimental data (KEIGHTLEY and LYNCH 2003), the precise ratio is unknown and it is possible that a greater proportion of all possible WS genotypes could be produced by mutation in genes requiring a gain of function mutation than expected. Indeed, an estimation of this ratio using the 26 independent WS (see section 2.5.2) reveals that ~25% of mutations were due to gain of function mutations, higher than even the most tentative estimates (KEIGHTLEY and LYNCH 2003).

The *mwsR* gene provides an ideal model for determining the relative rate of gain to loss of function mutations. Unlike the *wsp* or *aws* loci, MwsR possesses a DGC domain for producing c-di-GMP and a PDE domain that catalyses its degradation (see section 2.4.3.3). Gain of function mutations have been identified that over activate the DGC, causing overproduction of c-di-GMP and the WS phenotype (Rainey Lab, unpublished data). Loss of function mutations in the PDE domain abrogate PDE mediated breakdown of c-di-GMP, also causing the WS phenotype. Both domains are confined to the same protein so all mutations are subject to the same constraints; for example, all WS causing *mwsR* mutations found by members of the Rainey lab have been in frame. This is probably because the DGC enzymatic component of the mutated MwsR must still be

able to fold and function properly in order to produce c-di-GMP and thus the WS phenotype. It was known from our previous comparison of *wspF* and *awsX* that the constraint for mutations to be in frame does not apply to all genes. In MwsR, gain of function and loss of function mutations can occur in the same gene; differences in the number of possible mutations in either domain that can cause the WS phenotype will be due to the relative number of gain to loss of function mutations rather than other subtle differences in the types of mutations that can occur in different genes.

It was known from studies presented in chapter two that WS causal mutations preferentially occur in *aws*, *wsp* or *mws*; exploiting this knowledge, the *aws* and *wsp* loci were deleted from the SM^{MSC} genotype, resulting in a strain most likely to attain the WS phenotype by mutation in *mwsR* (SM^{MSC}*ΔwspΔaws*). This strain contains the WS reporter construct integrated into its genome; WS were isolated without the biasing effects of natural selection by incubating microcosms with shaking for 16 hours before isolation of a single WS clone. WS are not favoured by natural selection under these conditions, however, the SM^{MSC} strain contains the *wss* promotor fused to the *nptII* kanamycin resistance gene, upon mutation to the WS genotype the *wss* promoter is up-regulated, conferring resistance to kanamycin and allowing the detection of a single WS genotype among a large population of ancestral SM *P. fluorescens* cells.

A total of 25 microcosms were harvested; the DNA sequence of the regions of the *mwsR* gene pertaining to the PDE and DGC domains of *mwsR* was obtained. Mutations conferring the WS phenotype are known to occur in other regions of the *mwsR* gene; of the 25 WS, mutations were found for 16 of them; seven in the DGC and nine in the PDE domains of *mwsR*. The set of 20 WS from the experiment described in section 4.2 that were isolated without selection (the day zero set of 20 WS) were also sequenced at *mwsR* PDE and DGC domains and three of these found to contain *mwsR*

PDE mutations, giving a total of 19 independent *mwsR*– generated WS genotypes obtained without the influence of selection.

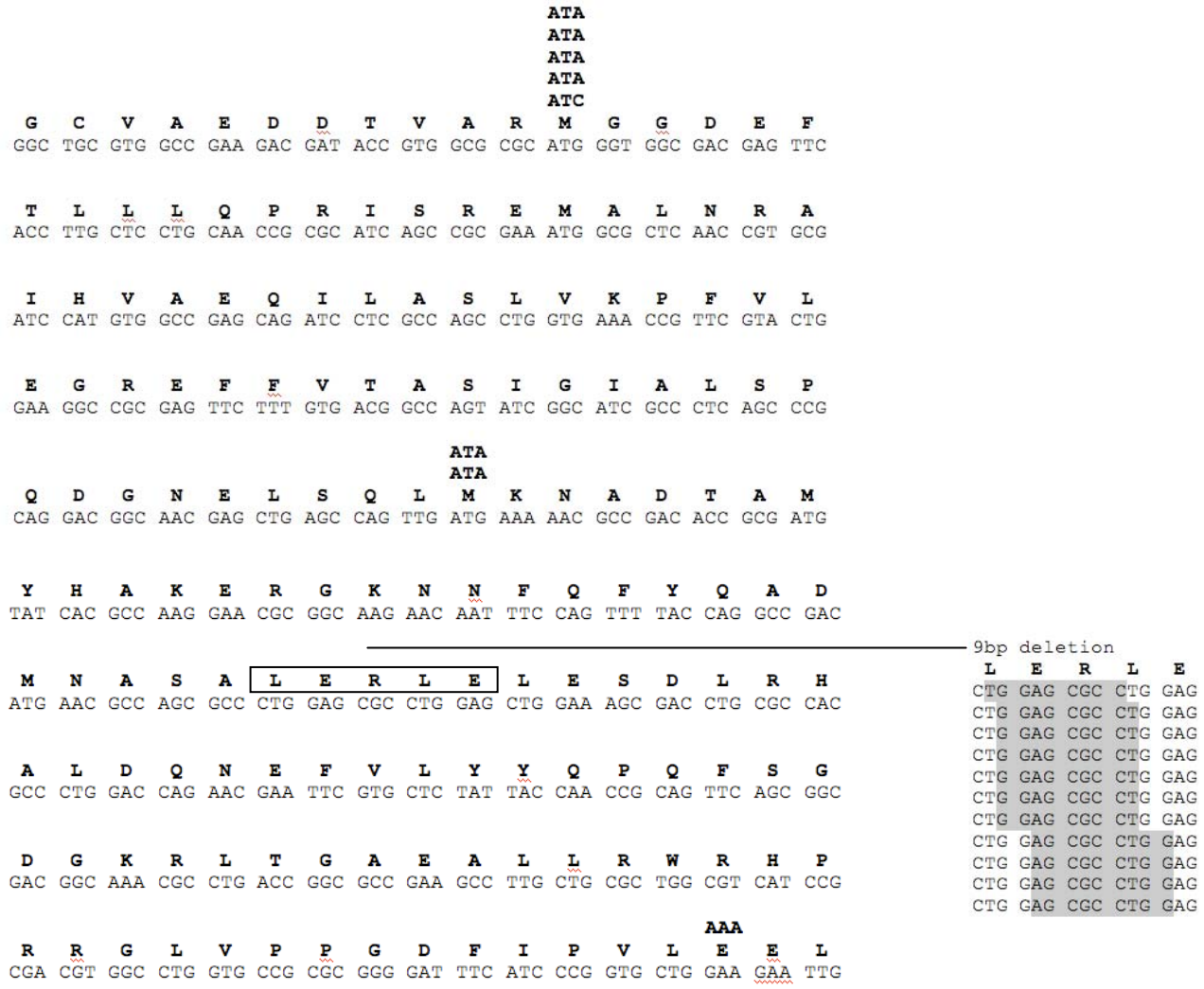


Figure 4.5

Nineteen *mwsR*- generated WS mutations were obtained from non-shaken microcosms using the *wss::nptII* WS reporter system, thereby minimising the effects of selection. Base pair changes are indicated in bold above the wild type sequence. The mutated sequence is shown in bold above the original. The 9 bp deleted for each of the "9 bp deletions" are highlighted. A highlighted or bold entry is shown as many times as that mutation was found among the *mwsR*- generated WS (a table of all mutations found in this study is in appendix 8.3.2).

4.2.4.1 COMPARISON OF THE OBSERVED AND EXPECTED FREQUENCIES OF *MWSR* MUTATIONS.

The *mwsR*- generated WS genotypes obtained from shaking microcosms without selection were evolved from a strain with the *aws* and *wsp* loci deleted. Because WS could only arise by mutation in *mwsR,* the possible number of WS mutations that could occur is much less than in the ancestral strain. There have been, in total, 15 unique *mwsR*-generated WS causal mutations found by workers in the Rainey laboratory. As an extremely conservative estimate, 15 is taken as the total pool of *mwsR*- generated WS mutants available to cause the WS phenotype in the SM^{MSC} *ΔwspΔaws* strain. The birthday program (described and implemented above in section 4.2.3.1) was then utilised in order to compare the observed to expected frequencies of WS causal mutations in *mwsR*. The birthday problem programme was asked; if 20 samples are taken from a pool of 15 mutants, how many times is it expected that any three of those mutations will be drawn more than three times. This is more conservative than the observed recovery of mutations; one mutation was found 6 times, another 5 times and another four times. This more conservative observed frequency of *mwsR* mutations was used because site 2778 (five *mwsR*- generated WS genotypes had mutations at this site) was inherently more fragile to mutation than most other sites (see 4.3.4.2 below). The programme ran 10,000 simulations and found that only in 25 instances would any three mutations be recovered three times in the same set of 20 genotypes, it is therefore highly unlikely ($p = 0.0025$) that the observed distribution of *mwsR*– generated WS mutations would arise if all mutations occurred at equal frequencies.

4.3 DISCUSSION

Previous studies have allowed only speculation about genetic constraints, a rather vague, all encompassing concept referring to the influence an organism's genome may have on its own evolution. The second chapter in this thesis provided the first *prima facie* evidence for the existence of genetic biases on evolutionary trajectories. The work described in this chapter goes further to define more precisely the sources of the bias and how these interact with the sole directional force in evolution, natural selection, to produce evolutionary outcomes.

4.3.1 THE CAUSES OF THE GENETICALLY BIASED PRODUCTION OF VARIATION.

At the outset of this study it was known that *awsX* and *wspF* were the main source of WS causal mutations within the 26 defined WS (BANTINAKI *et al.* 2007). The reason for this is that they encode negative regulators; mutations in *awsX* or *wspF* result in a loss of function for the AwsX or WspF proteins, but confer the WS genotype. Because mutations causing a loss in function are more common than those causing a gain in function (KEIGHTLEY and LYNCH 2003), a greater proportion of the mutations occurring in *awsX* and *wspF* than mutations in other genes of the *aws*, *wsp* and *mws* loci will be translated into the WS phenotype. The analysis of the 26 Independent WS was insufficient to provide a satisfactory explanation for the observed distribution of causal mutations; natural selection had acted on these WS for 5 days, so that some aspects of the genetically biased production of variation, especially in *awsX*, may be obscured.

A striking pattern emerged during the analysis of the 80 WS microcosms; *wspF*-generated WS dominate microcosms incubated statically for 3 or more days, while *awsX*-generated WS are isolated more frequently from microcosms incubated for one day. Although both *awsX* and *wspF* are negative regulators, *wspF* does not seem to be restricted in the types of mutations it can sustain (section 2.5.2). Any types of deletion, insertion or substitution, regardless of a change in reading frame, are tolerable. Consequently *wspF* can produce the greatest number of WS genotypes; these genotypes cover the greatest range of fitness effects (Figure 4.3). As microcosms are incubated for more time, more distinct WS genotypes arise by mutation, and natural selection has more time to distinguish between them, causing an increase in the frequency of the fittest types; such genotypes are most likely to arise by mutation in *wspF*.

The repeated isolation of the AWS and WS_T *awsX* WS genotypes during this study led to speculation that these two genotypes are caused by mutations that occur at a rate faster than other WS causing mutations. The observed distribution of WS causing mutations in the set of 20 microcosms incubated without selection (16 hrs in a shaken microcosm) was compared to the distribution expected under the assumption that all mutations occur at an equal rate; simulations using the birthday program confirmed that the AWS and WST genotypes arise a higher frequency than expected if all mutations occur with an equal frequency (P=0.0035).

Despite having a mid to low fitness (AWS = 0.93 and WS_T = 0.89, range of WS fitness's in this study, 0.75 – 1.098), the AWS and WS_T WS genotypes were repeatedly isolated from microcosms incubated for a short time, strongly suggesting that the higher rate of mutation to these WS instead of others led them to be the most likely to fix in WS mats and thus be isolated from microcosms during this study. In the next section I discuss how genotypes arising by mutation at a high frequency can be more likely to fix than fitter genotypes arising by mutation at a lower frequency.

4.3.2 THE RATE OF MUTATION INFLUENCES EVOLUTIONARY OUTCOMES

It has been previously shown (YAMPOLSKY and STOLTZFUS 2001), that for two alleles of different selective value and arising by mutation at different rates, if the difference in mutation rate is larger than the difference in the fitness, then the more frequently arising allele will be more likely to fix. This effect emerges in cases where mutations must arise *de novo;* because WS genotypes have a very low fitness in the environment inhabited by its ancestor, in the ancestral environment WS genotypes are subject to strong purifying selection and are not maintained as standing variation in the population. The probability of fixation of a WS genotype is thus the combined effects of the rate of allele origin and the selection coefficient. When the difference in mutation rate is greater than the difference in selective values, mutation (and thus contingency) plays a relatively larger role in determining which molecular path evolution will take. In microcosms incubated for a short time, only a few WS genotypes will have had time to arise by mutation; in this case the high rate at which the WST or AWS genotypes arise by mutation is enough to out weigh any fitness disadvantage they may have compared to competing WS genotypes. When microcosms are incubated for a longer period of time, a greater range of WS genotypes have had time to arise, covering a wider range of fitness; the difference in selective value between the best available WS genotypes and the WS genotypes that arise at a higher frequency (such as AWS and WS_T) is greater than the differences in rate of mutation. This leads to the displacement of the AWS and WS_T genotypes as the most common WS genotypes in microcosms exposed to longer periods of natural selection.

4.3.3 THE *MWSR*- GENERATED WS OBTAINED WITHOUT SELECTION.

The impetus for gathering a collection of *mwsR*- generated WS using the WS reporter containing strain (SM^{MSC}*ΔwspΔaws*) was to obtain the ratio of gain to loss of function mutations unbiased by natural selection or other differences caused by comparing sets of mutations from two different genes. The DGC domain of MwsR produces the secondary signalling molecule c-di-GMP; MwsR also encodes a PDE domain that facilitates the degradation of this molecule. Mutations in either domain are capable of resulting in the over production of c-di-GMP, the proximate cause of the WS phenotype. Firstly, gain of function mutations can occur in the DGC, causing an increase in c-di-GMP production to levels sufficient to cause the WS phenotype (despite the c-di-GMP degrading effects of the still active PDE domain). Secondly, mutations conferring a loss of function in the PDE domain result in the WS phenotype by inactivating the PDE and therefore allowing the DGC domain to produce c-di-GMP unchecked. Of the 19 mutations recovered, 7 were in the region of *mwsR* encoding the DGC and 12 in the region encoding the PDE domain. At first glance it seems to be a dramatic contradiction of the received wisdom that loss of function mutations are much more common than gain of function mutations (KEIGHTLEY and LYNCH 2003). However unexpected differences in the relative mutation rates skewed this result. Five of the seven mutations found in the DGC domain were due to substitutions at a single site (*mwsR*2778), and 11 of the PDE domain mutations were due to 9 bp deletions within a 12 nucleotide tract (*mwsR*3065-3076); all except one mutation within the set of 19 was found at least twice. The reason for the repeated discovery of the same mutations is not due to a paucity of possible WS causal mutations in *mwsR*. There are at least 15 confirmed unique *mwsR* WS mutations (Rainey lab), only two of which were found in the 19 sequenced here. Using extremely

conservative parameters it is clear that the repeated finding of the same mutations violates expectations under the assumption of random mutation (P= 0.0025).

4.3.4 CHARACTERISTICS OF MUTABLE DNA SEQUENCE.

Some DNA sequences are known to be highly mutable; tandem repeats or long tracts of a single nucleotide are susceptible to mispairing during DNA strand reannealing (figure 4.6). Such misalignments are repaired and in the process the tandem repeats either gain or lose repeats, depending on whether the leading or complementary strand misaligned. The loci at which such mechanisms have been characterised are known as contingency loci. Contingency loci are presumed to be an adaptation evolved in response to an environment that frequently yet unpredictably changes between two distinct sets of conditions (MOXON *et al.* 1994; MOXON *et al.* 2006). The expansion or contraction of these repeats can alternatively activate or inactivate a promoter or an enzyme, influencing a trait essential in one environment and undesirable in another. These loci are maintained over evolutionary time because the tracts of repeat DNA sequence can both increase and decrease in length and both states are alternatively favoured by selection.

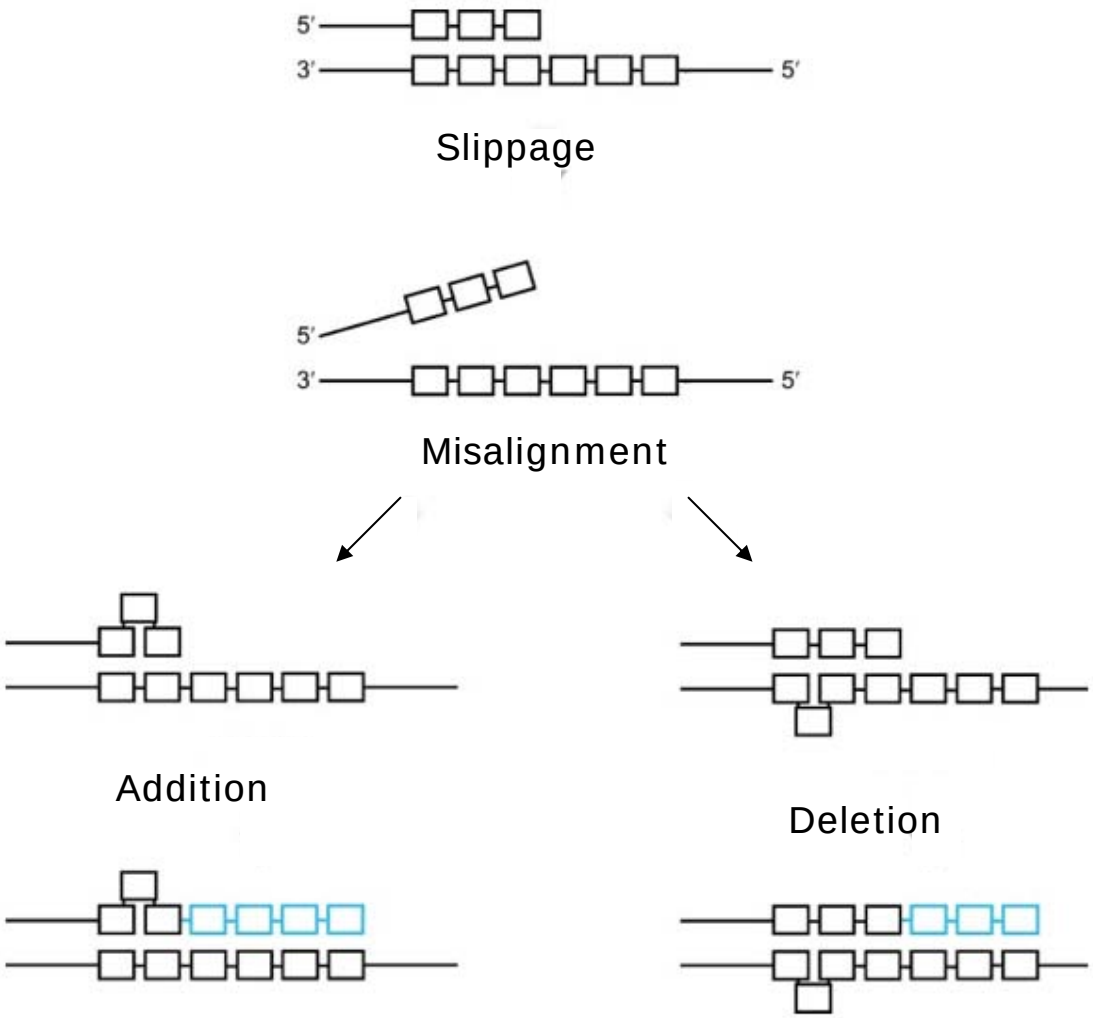


Figure 4.6

The basic mechanism of DNA strand slippage which can lead to either an insertion or deletion of the repeated unit. Each block represents a repeat sequence. Following mis-pairing, DNA repair mechanisms facilitate the insertion or deletion of a repeat and thus either the expansion or contraction of the chain of repeats (adapted from MOXON *et al.* 2006).

The mutation rates at such loci are up to one million times higher than the background mutation rate; although the rate at which WS causing mutations occur has not been explicitly measured, even the WS genotypes shown in this study to occur at a higher frequency clearly do not attain mutations rates this high. If there was such a WS

causing mutation, all WS genotypes other than this hypothetical "contingency WS locus" WS would be too rare to be detected.

The reason that the more frequently arising WS do not attain such high rates of mutation is perhaps because the sequence characteristics of these loci are reminiscent of, but not the same as "classic" contingency loci. For instance, the *awsX* deletions almost certainly arise by a strand slippage mechanism, however in all described contingency loci, mutable repeat sequences consist of many tandem repeats. The *awsX* mutations happen between two widely separated repeats and can result in deletion but not expansion; because the sequence deleted from *awsX* by these mutations is lost and cannot be recovered by further mutation, this locus has no potential to become a contingency site.

The *mwsR* WS causing mutations G2976A and G3244A were part of short homopolymeric tracts of adenine residues (poly(A) tracts); guanine nucleotides within or next to these sequences are susceptible to becoming an extension of the poly(A) tract by (G-A) transition mutation. While homopolymeric adenine tracts are known to be mutable, the loci described here are different and seemed to be recombined less efficiently (as evidenced by mutation frequencies clearly beneath the levels of genuine contingency loci) than those described in the literature (ALCALA *et al.* 2004; VAN DER ENDE *et al.* 2000). The inability of both the *awsX* and *mwsR* loci to switch back and forth and their low relative mutability firmly excludes the sites found here as being contingency loci. It is unlikely that the presence of such loci in the *P. fluorescens* genome can be the result of selection; while the advantage conferred by the mutation is heritable, the mutability of the sequence is not. While the mutant allele can fix in a population, in these cases mutability as a trait *per se* is lost with the mutation and cannot increase by selection. The mutable loci described in this study can arise by chance mutational events, and could be the kind of sequence that has the potential to evolve into a contingency locus.

4.3.4.1 THE BIRTH OF A CONTINGENCY LOCUS

The 9 bp deletions found to be so prevalent in *mwsR* accounted for 11 of the *mwsR* mutations; all occur within two imperfect 9 bp tandem repeats, each excision leaves one of the imperfect 9 bp repeats intact. This locus is different from the others in that it leaves one of the repeats intact, and thus has the potential to mutate back to having two, however the imperfect nature of the repeats makes this unlikely. The TGGAGC sequence (see figure 4.7, box 1, the top highlighted sequence) has been shown to be associated with mutation in humans and features in this *mwsR* 9 bp deletion locus. In *mwsR,* two copies of this sequence are separated by the arginine (R) encoding codon (CGC) (figure 4.7). This sequence has been described previously as the Krawczak–Cooper consensus sequence (TG G/A G/A G/T A/C); this sequence acts as an arrest site for human polymerase and was shown to be in the vicinity of 41% of sporadic gene deletions in humans as well as being associated with many larger human gene deletions and translocations (KRAWCZAK and COOPER 1991). This is the first instance in which the Krawczak-Cooper consensus sequence has been associated with mutable loci in prokaryotes, and this case provides the first evidence that this sequence may influence bacterial polymerases.

I submit that this site in *mwsR* has the potential to become a contingency locus; what follows is a possible scenario by which this could happen. If the codon encoding the arginine (R) residue was deleted a six base pair tandem repeat $(CTGGAG)^2$ would result. The arginine encoding codon is the only one deleted in all of the *mwsR* 9 bp WS mutations (figure 4.5), so its deletion is likely to be sufficient to cause the WS phenotype. This $(CTGGAG)^2$ sequence may then be subject to expansion and contraction (figure 4.7). If the expansion of the repeat sequence to three copies $(CTGGAG)^3$ was capable of

restoring the ancestral phenotype then having either two or three copies of these repeats could confer the WS or ancestral phenotype respectively.

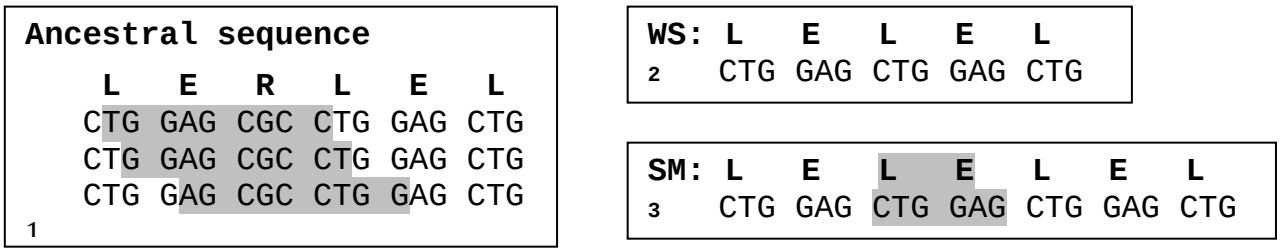


Figure 4.7

Sequence of possible steps of the ancestral sequence to a contingency locus. Highlighted segments of the ancestral sequence are already subject to relatively high rates of deletion (box 1). The only site consistently removed is the arginine (R) residue. If this was removed by deletion two tandem CTGGAG repeats and the WS phenotype would result (box 2). Extension of this to a triplet repeat may confer the ancestral phenotype, and a contingency locus would be born (box3); extension and contraction of the CTGGAG sequence would alternatively result in the WS and ancestral genotypes.

The presence of the Krawczak-cooper consensus sequence makes it likely that the switching between these two genotypes would occur at a high frequency. This scenario is plausible; the arginine (R) in question is not part of a conserved motif and the loss of function caused by the 9 bp deletions may be due to spacing rearrangements, also glutamine (E) (GAG) and arginine (R) are both polar residues. If selective conditions (such as rapid switching between environments alternatively favouring the WS and ancestral phenotypes) then mutations at this locus could facilitate this transition at a high frequency without mutations irreversibly destroying the function of *mwsR*.

4.3.4.2 MUTABLE OR FRAGILE

Two of the mutated sites in the DGC domain (G2778A/C and G2976A) resulted in conversion of methionine to isoleucine. Methionine is one of two amino acids coded by a single codon, any nucleotide change results in a change of amino acid; rendering motifs containing methionine fragile to non-synonymous change by mutation. Such fragile codons have been postulated to confer upon sequences containing them the greater capacity to evolve (MEYERS *et al.* 2005). Although site 2778 is still likely to have a higher per nucleotide mutation rate, fragile sites give the perception of a higher mutation rate; all nine of the nine possible snp's within that codon result in a non-synonymous change, where as in most other codons only six of the nine will.

4.3.5 CONCLUDING COMMENTS

The results presented here show that the range of mutations presented to natural selection is strongly influenced by two factors: firstly, variation in the capacity of genes to translate DNA sequence change into phenotypic variation; and secondly, variation in the relative rates with which certain mutational changes arise. The relative importance of these two effects depends on the supply of mutations within the evolving population. Genes in which DNA sequence change most often translates into a novel adaptive phenotype will most likely also produce the fittest genotypes; the fitness effects of these genotypes will cover a wider range than genotypes produced by mutation in other genes less able to convert sequence change into adaptive variation. If natural selection has the opportunity to choose between a wide range of adaptive genotypes, genotypes produced

by mutation in genes best able to convert mutation to variation are more likely to fix than others. Conversely, if the supply of mutations is lower, perhaps because of a low population size, then the genotypes that arise first are also most likely to fix. Under these conditions, sites in which DNA sequence change happens at a higher frequency than others will tend to contribute the next genotype to fix in the population. These results suggest also that the traditional view of mutation rate as a constant may be more accurately replaced with the idea that there is a distribution of mutation rates across a genome; this will further complicate models of adaptation.

4.4 BIBLIOGRAPHY

ALCALA, B., C. SALCEDO, L. ARREAZA, R. ABAD, R. ENRIQUEZ *et al.*, 2004 Antigenic and/or phase variation of PorA protein in non-subtypable *Neisseria meningitidis* strains isolated in Spain. J. Med. Microbiol. **53:** 515-518.

BANTINAKI, E., R. KASSEN, C. G. KNIGHT, Z. ROBINSON, A. J. SPIERS *et al.*, 2007 Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. Genetics **176:** 441-453.

FUKAMI, T., H. J. E. BEAUMONT, X. X. ZHANG and P. B. RAINEY, 2007 Immigration history controls diversification in experimental adaptive radiation. Nature **446:** 436-439.

KEIGHTLEY, P. D., and M. LYNCH, 2003 Toward a realistic model of mutations affecting fitness. Evolution **57:** 683-685.

KRAWCZAK, M., and D. N. COOPER, 1991 Gene deletions causing human genetic disease- mechanisms of mutagenesis and the role of the local DNA sequence environment. . Hum. Genet. **86:** 425-441.

KUNKEL, T. A., 1990 Frameshift mutagenesis by eukaryotic DNA polymerases *in vitro*. J. Biol. Chem. **261:** 13581-13587.

LANG, G. I., and A. W. MURRAY, 2008 Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. Genetics **178:** 67-82.

MEYERS, L. A., F. D. ANCEL and M. LACHMANN, 2005 Evolution of genetic potential. PLoS Comput. Biol. **1:** 236-243.

MOXON, E. R., P. B. RAINEY, M. A. NOWAK and R. E. LENSKI, 1994 Adaptive Evolution of Highly Mutable Loci in Pathogenic Bacteria. Curr. Biol. **4:** 24-33.

MOXON, R., C. BAYLISS and D. HOOD, 2006 Bacterial contingency loci: The role of simple sequence DNA repeats in bacterial adaptation. Annu. Rev. Genet. **40:** 307-333.

RAINEY, P. B., and M. TRAVISANO, 1998 Adaptive radiation in a heterogeneous environment. Nature **394:** 69-72.

STREISINGER, G., Y. OKADA, J. EMRICH, J. NEWTON, A. TSUGITA *et al.*, 1966 Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31:** 77-84.

VAN DER ENDE, A., C. T. P. HOPMAN and J. DANKERT, 2000 Multiple mechanisms of phase variation of PorA in *Neisseria meningitidis*. Infect. Immun. **68:** 6685-6690.

YAMPOLSKY, L. Y., and A. STOLTZFUS, 2001 Bias in the introduction of variation as an orienting factor in evolution. Evol. Dev. **3:** 73-83.

5 THE EVOLUTION OF COMPLEXITY IN WS MATS.

5.1 INTRODUCTION

Rainey and Rainey (2003) showed WS mats to be cooperative groups, establishing a microbial model for the *de novo* evolution of cooperation. It has been recognised that the formation of such undifferentiated groups of cells is a likely first step towards multicellularity (PFEIFFER and BONHOEFFER 2003). Such groups exist in a precarious position between two levels of the evolutionary hierarchy, corresponding to two possible units of selection[1]; the individual or the group. Several challenges must be overcome if such nascent groups are to become a fully integrated multicellular individual; in the volvacine algae 12 steps has been postulated as a minimum for a complete transition (proposed by KIRK, 2005 and then developed by HERRON and MICHOD, 2008). In this chapter I describe experiments designed to investigate the potential of WS mats to go beyond the step they have already taken and evolve complexity by way of group level selection. In this introduction I present what are considered the main challenges faced by nascent multicellular individuals as well as some of the prevailing hypothesised solutions.

A newly evolved group of undifferentiated cells is composed of reproductively independent individuals. During the transition from unicellular to multicellularity, these cells cooperate to form a higher-level unit of selection (the group); if group level traits are to evolve, selection must act to facilitate adaptation of groups, despite the interests of the composite individuals. Conflict can arise if the interests of the individual and group are not aligned. Because cooperation between cells by definition requires between cell interactions; the only way that unicellular organisms can interact is by secreted products. Such interactions are vunerable to exploitation; producers cannot control the distribution of the secreted goods and cheats may arise that use or benefit from the product, while

[1] A biological entity subject to natural selection.

not paying the cost of production (BUSS 1987; MICHOD 1998; SOBER and WILSON 1998). Not surprisingly then, cheats arise in WS mats. Mats are formed by the cooperative production of copious extracellular material by WS; non-mat forming cheating types arise by mutation that are phenotypically indistinguishable from the ancestor (RAINEY and RAINEY 2003). These cheating types do not secrete mat forming cellulose, LPS or produce attachment factor and so have a high individual fitness because they can expend the saved resources on reproducing at a faster rate than their WS progenitors. Fitness of the group however, decreases; the cheats eventually attain such high numbers relative to cooperators that the mat cannot be supported and collapses into the broth (RAINEY and RAINEY 2003). This problem is universal to all cooperating groups of cells; for the group to have the potential to evolve further complexity, the lower level units of selection (cells) must be suppressed and the higher level of selection (groups) enhanced.

The hypothesised resolution to the problem is conflict mediation (FRANK 1995; MICHOD and HERRON 2006; MICHOD *et al.* 2006; RAINEY and RAINEY 2003). Conflict mediators exist in many forms, including germ-line sequesterisation, controlled apoptosis, and self-policing. In each case the goal is the same, to suppress lower level individuals from reproducing independently of the group. Such systems are easily recognisable in contemporary multicellular organisms; cancer cells are cheating types, and the targeted destruction of these by apoptosis, and autophagy can be seen as a form of conflict mediation. The challenge though is not finding examples of conflict mediation, but answering the question, how did they evolve? The examples mentioned involve complex mechanisms of self-recognition and/or selective destruction. A catch 22 emerges: in order for group level complexity to evolve conflict mediation is required, yet conflict mediation is in itself a complex group level trait.

Conflict mediation is but one of the requirements for a truly multicellular organism; a second hallmark of complexity is the division of labour. Posited as the

greatest benefit of being multicellular (MAYNARD-SMITH and SZATHMARY 1995), a division of labour requires the evolution of specialised cell types, each performing a subset of the tasks required by the group as a whole. Life as part of a team provides opportunities for specialization not available to free living unicellular organisms. The principle of allocation asserts that the resources that an organism may devote to non-essential processes are limited (LEVINS 1968). Groups of specialised cells have an advantage; a specialised cell can devote more genetic and physical resources to a task than a unicellular generalist that must divide its resources between many jobs. It is postulated that tradeoffs must exist, whereby the performance of each task cannot be improved as much in a generalist as in a specialist (ARNOLD 1992). If two specialised cells cannot perform both tasks as well as two generalists then the evolution of differentiated cell types would not occur. An important form of the division of labour is the separation of germ and soma, postulated to be an adaptation to the life history trade-off between reproduction and growth (MICHOD *et al.* 2006).

5.1.1 A MODEL FOR THE EVOLUTION OF A GERM-SOMA SEPARATION.

The volvocine green algae contain both unicellular and multicellular members, and as such provides a useful model for the study of the intermediate steps between unicellularity and multicellularity (SACHS 2008). Michod and co-workers (2006) proposed a model by which conflict mediation, separation of germ and soma (and thus a division of labour) could arise in a single step, overcoming the apparent catch 22 posed above. In a mathematical analysis using known biological parameters they found that the life history trade-off between reproductive and vegetative survival could drive the evolution of specialised germ and soma cells in groups of sufficiently large size. In the case of

volvocine algae the cost of producing "embryos" (clumps of cells that are dispensed as offspring) increases for parent colonies of larger size; the evolution of distinct single cell propagules greatly reduces this cost. When cells specialize into germ and soma they automatically become interdependent, this transfers fitness and individuality from the single cell to the group, obviating the need for another conflict mediation system. This single step facilitates the transfer to a new program of group selection and the subsequent development of group level traits. Interestingly, J.T Bonner has independently described a similar correlation between the size of volvacine algae and complexity as defined by the number of differentiated cell types (BONNER 2004).

5.1.2 GROUP SELECTION

Most previous investigations into the evolution of group level complexity have been limited to studies of organisms at putative intermediatory evolutionary stages during the evolution of multicellularity (BONNER 1999; BONNER 2003; MICHOD 2007; MICHOD *et al.* 2003) but see (BORAAS *et al.* 1998). Elucidation of the required conditions for the evolution of such a complex phenotype as multicellularity is difficult when they are obscured by millions of years of evolution[2]. An ideal experiment would facilitate the observation of this process rather than its end result. The evolution of group level traits requires group level selection; that is, the differential survival of groups of organisms. Groups of individuals have been shown to respond to selection (GOODNIGHT and STEVENS 1997; WADE 1976); in these experiments the demonstration of group selection was an end in itself, here I employ group selection as a means for pushing groups of WS to become more multicellular like.

[2] The volvacine algae is the most recently evolved multicellular lineage, having diverged from a unicellular ancestor ~35 million years ago.

5.1.3 THE EVOLUTION OF BACTERIAL MULTICELLULARITY

Despite having evolved half a billion years before the first multicellular eukaryotes, bacterial multicellularity is rare and clearly less diverse. The soil proteobacterium *Myxococcus xanthus* provides a well known model; cells cooperate to form a predatory slug and upon starvation a fruiting body, with up to 80% of cells sacrificing themselves to form the stalk which supports it. Bacterial systems and, as perhaps the sole model for the *de novo* evolution of cooperation, WS mats provide a unique opportunity to explore the transition from the uni- to the multicellular. In this chapter a group selection experiment is implemented with the aim of promoting the evolution of increased group level complexity, and a strategy developed to overcome the problem of cheats. I find that WS groups can evolve complex traits characteristic of multicellular organisms, but the persistence of these groups as evolutionary individuals is dependant on the application of an artificial conflict mediator.

5.2 RESULTS

5.2.1 GROUP SELECTION FOR STRONGER WS MATS

The evolution of the WS mat is dependant on the strong selective advantage derived from living in a group of cells located at the interface of liquid and air. In order to promote the evolution of group level traits, a selection experiment was designed in which the reproductive success of the group was solely dependant on a group trait. This was enforced by choosing the best groups, instead of the best individuals, to found successive generations. The group trait chosen was mat strength, as defined by the number of glass beads that the WS mat under consideration could support. Mat strength was thought an appropriate group trait, as stronger mats should be so because of superior cooperative interactions between cells.

Eight microcosms were inoculated with the Alternative Wrinkly Spreader (AWS) and incubated statically at 28^0C for 7 days. The selection of this incubation period was based on the known tendency of WS mat strength to peak at days three and four (RAINEY and RAINEY 2003). Mat strength wanes significantly by day seven, due to the invasion of cheats. By focusing on this stage of WS mat development, selection would challenge mats to overcome the problem of cheats. Also, because mat strength was much lower at this stage than maximum attainable mat strength, it was thought that WS mats had the potential to improve.

Following 7 days of incubation, each of the eight mats was assayed for mat strength then the microcosm homogenized by vortexing and a sample plated on KB agar plates. The subsequent generation was founded by inoculation of fresh microcosms with

6μl of the homogenized culture from the microcosm containing the strongest mat. Three selection lines were propagated in this manner. A further three control lines followed identical treatment except the weakest mat was chosen to found the next generation. The results of five weeks of this selection regime are shown in figure 5.1.

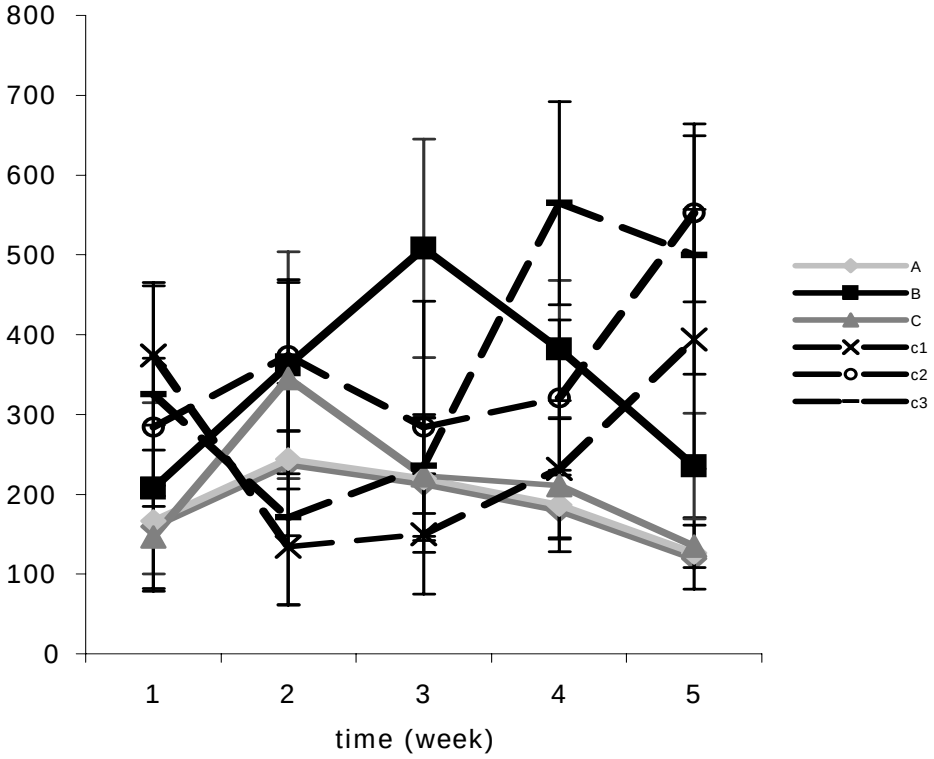


Figure 5.1

Fitness trajectories of experimental lines A, B, and C (solid lines) during group selection in a spatially structured environment. Lines B and C have an initial response to group selection, however by week 4 group fitness of all three lines had returned to initial fitness's; values are means +/- 95% C.I. (n = 8). Control lines c1, c2, c3 (dashed lines) exhibited mat strength that varied greatly between weeks.

After the full course of the experiment the group selected lines did not increase mat strength significantly higher than the first measurement (Tukeys HSD, $\alpha < 0.05$) (also see table 5.1), although one of the experimental lines (b) did significantly increase its fitness between weeks one (207.2 mg; C.I 100.2-314.2 mg) and three (507.8 mg; CI 371-

644.6 mg). The control lines fitness values oscillated quite randomly with respect to the selective regime for weaker mat strength.

5.2.2 DIMINISHING MAT STRENGTH CORRELATES WITH AN INCREASED FREQUENCY OF CHEATS

The strength of WS mats is known to diminish with the increasing proportion of cheats (Rainey and Rainey, 2003). To obtain a more rigorous measure of the relationship between mat strength and the invasion of cheats the strength of WS mats was measured and the proportion of cheats scored over ten days of incubation (figure 5.2). One hundred microcosms were inoculated with the AWS; mat strength and the proportion of cheating types to mat formers were measured in 10 of these each day. Cheating types have the same colony morphology as ancestral (or SM) cells. This allows for ready scoring of WS and cheating types on KB agar. Only mats declining in strength were included in analyses. This is because mats from microcosms incubated for only one or two days are weak due their incomplete development, not because of the invasion of cheating types.

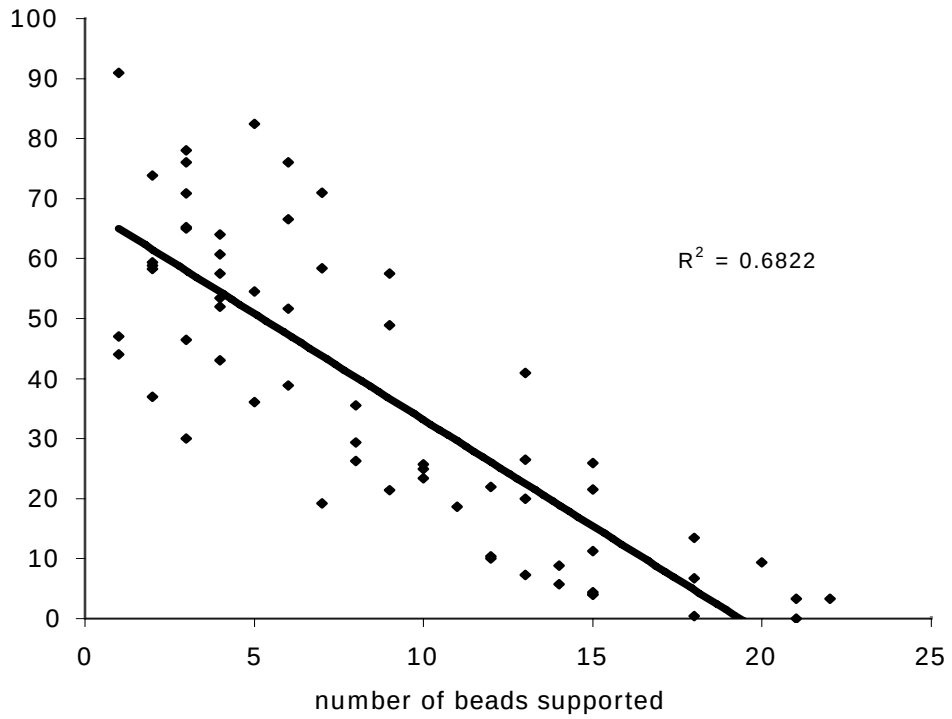


Figure 5.2

Relationship between mat strength and the proportion of the mat made up of cheats.
This data comprises measurements from 62 individuals ($R^2 = 0.6822$) ($F_{1,62} = 130.94$, $p = 8e^{-17}$).

Having established a clear relationship between the proportion of cheating types comprising a WS mat and mat strength, the next step was to establish the cause of the failure of all group selected lines to respond to selection. The most likely explanation for this is the evolutionary emergence of cheating types during the course of the group selection experiment. Even though at each step of the group selection experiment the strongest groups were selected for propagation, powerful individual selective forces can undermine the evolution of group-selected traits. A measure of this was obtained by plating out samples from each microcosm at each stage of the selection experiment, and scoring WS and cheating types as described above.

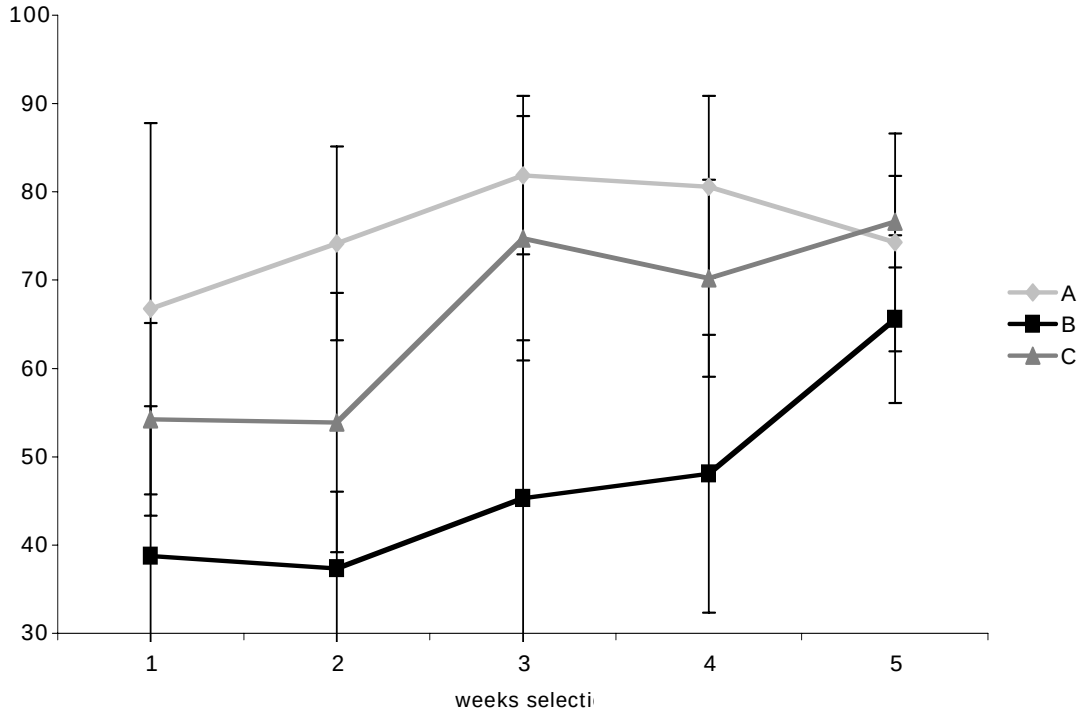


Figure 5.3

For the group selected lines, the percentage of cheating types in the microcosm with the strongest mat in each microcosm, the proportion of cheats is approximately inversely proportional to the strength of the mat, however the large variance of mat strength measurements (shown in figure 5.1) results in a large variance in the proportion of cheats when averaged over all eight replicates.

Although the mean percentage of cheats correlated approximately with the points within the group selection experiment at which mat strength decreased (figure 5.1), the large spread of error means that no significant conclusion can be drawn.

5.2.3 CONFLICT MEDIATION FACILITATES EVOLUTION OF THE GROUP.

The known correlation of diminishing mat strength and increasing prominence of cheating types was supported by the scoring of cheats in the group selected lines. The high degree of variation within experimental lines in the group selection experiment may have obscured the link between the invasion of cheating types and the failure of group selected lines to improve in fitness. In order to rescue the evolution of group level complexity within WS mats, a novel conflict mediation strategy was designed, the goal of which was to demonstrate the facilitation of group level adaptation by conflict mediation.

In work described in earlier chapters of this thesis the SM^{MSC} strain has been used to identify WS types as soon as they arise by mutation. This strain contains a kanamycin resistance gene fused to the *wss* promoter in the otherwise ancestral *P. fluorescens* genetic background. It has been reported that the *wss* promoter is expressed at higher levels in WS than the SM ancestor (SPIERS *et al.* 2002; SPIERS and RAINEY 2005); mutations conferring the WS phenotype also confer resistance to kanamycin. Conversely, WS containing this construct lose kanamycin resistance when they mutate into the cheating type. If WS mats are propagated in broth containing kanamycin, as cheating types arise by mutation they will die immediately. The implementation of the *wss*::kanamycin construct in this context fulfils the requirement of a conflict mediator, that is the ability to suppress the defection of individuals within the group.

In order to rescue the group selected lines, archived strains from the group selection experimental lines were recovered from the time during the experiment at which mat strength peaked (week 2 for lines A and C, week 3 for line B). These strains were transformed with the pUIC3 plasmid containing the *wss*::*kanamycin* fusion. This plasmid is unable to replicate in SBW25; sequence homologous to the *wss* operon facilitates insertion into the genome upstream from the *wss* operon by recombination.

Following insertion of the conflict mediator construct into the selected strains from experimental lines A, B and C, each was used to found eight microcosms, the microcosms were incubated for seven days and the mat's strength measured. As before the strongest mat from each line was used to found the next generation of mats.[3]

	Mean Mat Strength (mg)					
			w/o kanamycin		w/kanamycin	
Line	**Initial**		**Final**		**Final**	
A	166.5	A	125.9	A	381.9	B
B	207.2	A	235.6	A	544.375	B
C	146.2	A	134.6	A	418.4	B

Table 5.1

The strength of mats at the first and last weeks of the group selected (w/o kanamycin) and conflict mediated (w/kanamycin) group selection experiments. The value is the average of eight mats measured (mg), Tukeys HSD were performed; levels with different letters are significantly different ($\alpha < 0.05$).

All group-selected lines attained a fitness (as defined by mat strength) significantly higher than mat strength measured at week one (tukeys HSD, $\alpha < 0.05$) (table 5.1). At each time point, all mats were scored for cheats, however none were ever detected, providing confirmation that the conflict mediation strategy succeeded in suppressing cheating types and that the reason for the maintenance of high mat strength indeed was the absence of cheating types. These data strongly suggest that the inability of two of the three groups to respond to group selection and the failure of all to maintain a high level of group fitness was due to the invasion of cheating types. The presence of a conflict mediator construct facilitated the maintenance of high group level fitness.

[3] Line B had two morphologically distinct types (discussed later in this chapter) both types were modified with the conflict mediator construct and used to found the conflict mediator "line B".

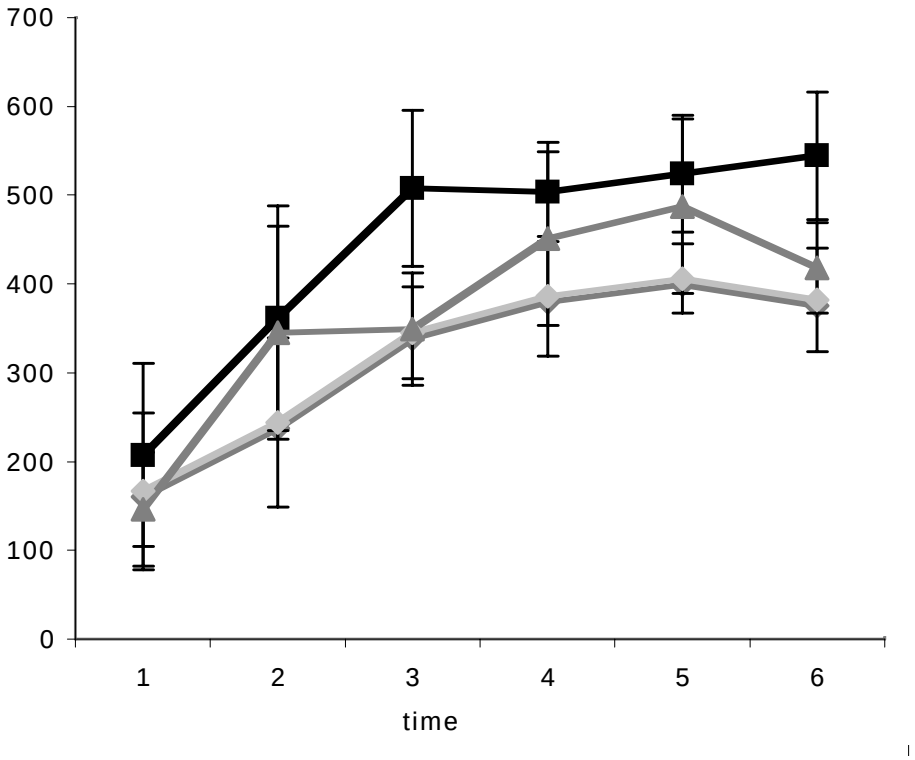


Figure 5.4

Clones from each group-selected line were taken from the strongest microcosm at the peak of the response to group selection and modified with the genetic conflict mediator construct. This construct rescued each line from invasion by cheating types and lead to the maintenance of mat strength gained by group selection. Values are means +/- 95% C.I. (n = 8).

5.2.4 THE EMERGENCE OF DIFFERENTIATED CELL TYPES WITHIN WS MATS.

During the original group selection experiment a single line (B), increased in mat strength each week until week four when, like all lines, gains in mat strength were lost. During scoring of plates for the frequency of cheats it was noted that two WS types with distinct morphology had arisen that were maintained at a stable frequency in the mats. These types were designated WS_1 and WS_2. Inoculation of a microcosm with only the WS_2 type

resulted in the formation of a mat that developed in a unique way and appeared much thicker than the WS_1 type mat. In order to show that this was not a case of clonal succession (a new WS invading and taking over another) reciprocal invasion assays were conducted over one week. Starting at a ratio of 100:1 both types were maintained for 7 days, in both cases the lower frequency type was able to increase in frequency, but not significantly invade or displace the other (n=6).

The stable maintenance of this diversity led to the formation of the hypothesis that the two WS types each required the presence of the other to produce fitter groups.

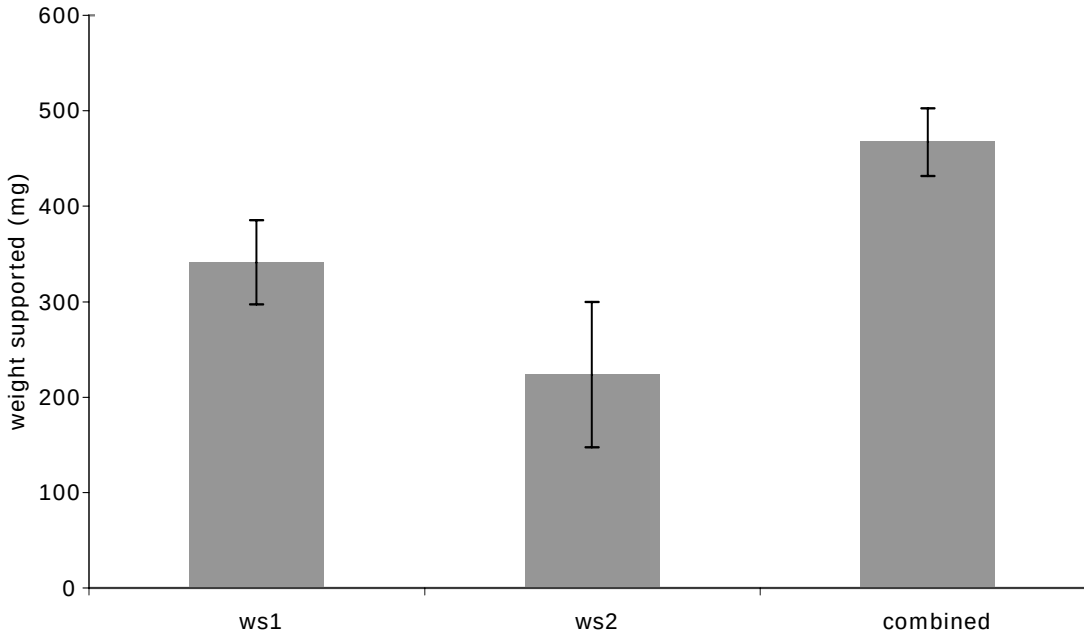


Figure 5.5

Each column shows the average value of eight mats, 95% confidence intervals shown. Mats formed by WS_1 and WS_2 combined could support 467 mg (95% CI: 424-510), significantly more than either the WS_1, (341 mg; 95% CI: 265-417), or the WS_2, (223 mg; 95% CI: 188-258).

Following the demonstration that the strength of the mats is due to the combined presence of these two cell types, the basis of this interaction was explored by testing the

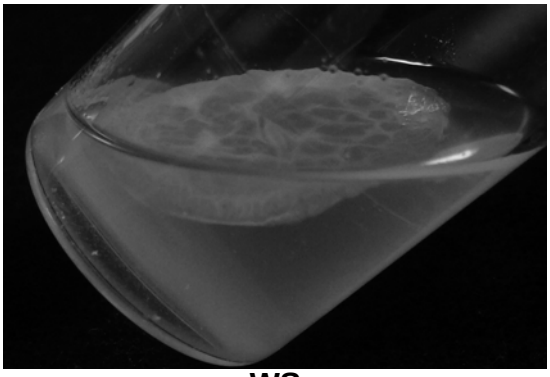
two aspects of the WS phenotype that facilitate exploitation of the air liquid interface: production of cellulose and proteinaceous attachment factor.

5.2.5 THE WS_1 AND WS_2 HAVE SPECIALISED AND COMPLEMENTARY FUNCTIONS WITHIN THE WS MAT.

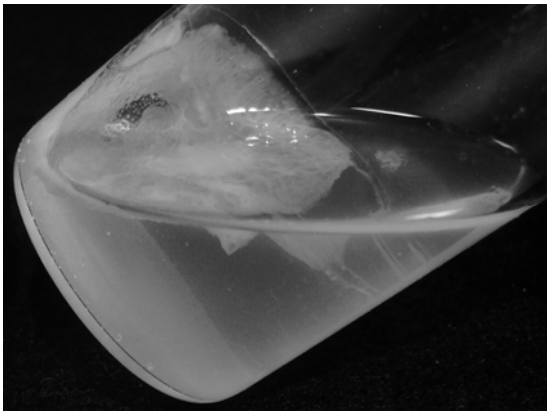
Cellulose facilitates attachment between cells and the proteinaceous attachment factor attachment to glass (SPIERS and RAINEY 2005). In order to compare the relative abilities of the WS_1 and WS_2 types to perform these two tasks, the time taken for mats formed by either type to detach from the glass wall was measured, then after detachment the time taken to sink. This assay was performed by carefully placing microcosms to be assayed on a shaking platform. Microcosms were consistently shaken at 100 rpm, and the time taken for the mat to detach from the glass wall recorded. Following detachment, shaking continued and the time taken for the detached mat to break up and sink to the bottom of the microcosm was also noted.

The difference between the two types was striking. The WS_1 types stuck longer to the glass (500.16s C.I 372.8-627.52s), but quickly sank to the bottom of the vial once detached (3.16s C.I 0.47-5.85s). Conversely, the WS_2 type required almost no agitation to detach from the glass (9s C.I 0-18.08s), yet the mats were more stable and would float on the surface until the microcosm was severely disrupted (485s C.I 318.81-651.19s). A two-sampled t-test assuming unequal variances showed a highly significant difference between the two types both for detachment ($p = 0.00065$) and sinking ($p = 0.0023$).

WS_1



WS_2



WS_1+WS_2

Figure 5.6

Phenotypic differences of mats composed of the WS_1, WS_2 or WS_1+WS_2 types. WS_1 mats attach strongly to the glass wall and leave a residue on the surface when dislodged, however they sink almost immediately after detachment (see text). WS_2 mats require little agitation to be detached from the glass surface and leave no residue, however they float persistently despite extended vigorous shaking. Mats composed of both WS_1 and WS_2 types are attached firmly to the glass surface and leave a residue as well as floating persistently at the surface.

5.3 DISCUSSION

Much of this thesis has been devoted to the study of the first step of *P. fluorescens* adaptation to the air-liquid interface. The work presented in this chapter explores the continued adaptation of WS mats to the static microcosm and more generally the evolution of complexity at the level of the group. Group selection for increased mat strength was seen to fail due to the invasion of cheating types, a problem common to nascent cooperative groups. This problem was solved by the implementation of a novel conflict mediation strategy; the subsequent suppression of cheats facilitated the maintenance of high group fitness. In one of the group selected lines, two specialised cell types evolved, each able to perform a different task allowing mats comprised of both types to be fitter than mats with only one type. In this section I interpret and explore the significance of these results.

5.3.1 THE FAILURE OF GROUP SELECTION WAS DUE TO THE INVASION OF CHEATS.

During the group selection experiment, all lines were characterised by the eventual decline in fitness. This result makes sense in light of the corresponding increase in the proportion of cheats in each line. During this experiment, groups were propagated by vortexing of the entire microcosm followed by a 1000 fold dilution to seed the next generation. Each group "propagule" then contains not only the superior clones that conferred the high mat strength, but also a number of cheats. Cheats that had arisen in

one generation would be transferred to the next, with each generation higher numbers of cheats were introduced; more cheats had a chance to join the mat. When individual and group selection acts in opposing directions, individual selection has the advantage. Individuals cells have a much faster generation time (40 minutes) compared to that of a group (7 days), meaning that a selfish trait conferring an adaptive edge over cooperative cells could spread to a high frequency before a single group generation has taken place.

The control lines (weak mat selection) displayed wildly fluctuating mat strengths, this differs from the seemingly obvious prediction; weaker groups should have more cheats, the cheats should increase in frequency with each generation and mat strength should consistently decrease. However, cheats are only fit in the context of one of the niches of the static microcosm (the WS mat) or during the period immediately following inoculation, when the liquid column still has sufficient oxygen. It is possible that if a weak mat was chosen and a large amount of cheats were transferred to the next generation, initial WS growth would be slow in the broth as they start from low numbers and most cells in the broth are cheating types, which thrive in the newly inoculated microcosm. As oxygen runs out, cheat growth slows, and only now may WS have the density to form the WS mat. This delay in mat formation may mean that a mat measured at 7days might be typical in strength to a four or five day mat, and therefore be stronger than expected. Also, at this point cheats have not had the chance to invade the mat, and numbers of cheats in the broth are relatively low due to the WS mat reducing the diffusion of oxygen into the broth (RAINEY and RAINEY 2003). A propagule taken from this mat may then found a weak or average strength mat because the founding population had less cheats, so that the formation of the mat was not delayed. The results of the control line shows the complicated interactions that can arise when multiple cell types are growing in multiple niches and selection is acting at both the individual and group level.

5.3.2 AN ARTIFICIAL CONFLICT MEDIATOR RESCUED THE EVOLUTION OF THE GROUP.

The *wss* operon is constitutively active in WS (SPIERS *et al.* 2003), I exploited this to apply a novel conflict mediation construct comprised of the *wss* promoter fused to a gene conferring kanamycin resistance. While WS containing the construct are resistant to kanamycin, mutation to the ancestral (cheating) type results in a reduction of *wss* expression (FUKAMI *et al.* 2007) and the relegation of kanamycin resistance. In order to rescue the group selected populations from the invasion of cheats the genetic conflict mediator was introduced into WS recovered from each experimental line at the time point at which they had achieved maximum fitness. By continuing the group selection experiment in media containing kanamycin, the death of any cheating types that arose was ensured. The weak selection groups were not chosen for modification with the conflict mediator, but because there weakness was caused by cheats, it could be that they would sustain some maintained improvement in mat strength too. We saw that each experimental line was rescued from decline and able to maintain high group fitness. This result is the first empirical confirmation that conflict mediation is required for the evolution of cooperative groups. Although group level traits were able to evolve in the absence of conflict mediation, their evolutionary future was extremely limited.

The emergence and spread of cheats within the group-selected lines had a two-fold effect on the fitness. Firstly, it greatly reduced the probability that group level adaptation could occur at all. Only one of the group-selected lines (B) consistently increased in mat strength for more than one week. Secondly, even if high mat strength was attained it could not be maintained as the invasion by cheating types was inevitable.

Conflict mediation did not enhance the mat strength *per se*, indeed the average mat strength of group-selected lines with conflict mediation did not exceed that attained by group-selected line B in the non-conflict mediated experiment. Rather the benefit of conflict mediation is that it provides the group with the temporal cohesiveness required for the maintenance of complex traits.

5.3.3 THE EVOLUTION OF A DIVISION OF LABOUR.

Specialists may be favoured within groups if they can allocate more resources to a specific function, while a second cell type specializes in the function neglected by the first. This division of labour overcomes physiological and genetic constraints that could restrict the evolutionary potential of unicellular generalists. Unique to one of the group selected experimental lines (B) was the emergence and stable maintenance of a second morphologically distinct mat forming cell type. The complementary behaviour of these two cell types under identical conditions supports that the combinatorial beneficial effects are due to specialisation of the WS_1 to glass attachment and the WS_2 to cell-cell glue production. This comprises the first example of the evolution, *de novo,* of a division of labour.

The regulation of the production of cell-cell glue and attachment factor in the WS is known to be dependant upon constitutive production of secondary signalling molecule c-diGMP (MALONE *et al.* 2007). The genetic basis for the division of labour was not investigated, however the underlying cause of the WS phenotype is well understood. It may be that the mutational cause of the division of labour in the cell-cell glue specialist blocks c-diGMP from activating attachment factor. This would facilitate the devotion of c-diGMP to its function as an allosteric and transcriptional activator of cellulose biosynthesis (ROSS *et al.* 1987). This would both increase regulatory stimulation and free

up resources previously devoted to attachment factor production for the increased production of cellulose.

5.3.4 PROKARYOTIC AND EUKARYOTIC POTENTIAL TO EVOLVE MULTICELLULARITY.

WS adaptation to the air liquid interface is achieved by the production of an attachment factor and cellulose; providing cell-glass and cell-cell adhesion respectively. Selection for increased mat strength led to the evolution of cell types specialised to perform either one or the other of these tasks, demonstrating that natural selection acting at the level of the group can overcome the tradeoffs that limit the evolutionary potential of individuals. It has been postulated that selection on the life history trade off between reproduction and growth can lead to the evolution of germ-soma separation (MICHOD *et al.* 2006), in a single step achieving conflict mediation and also a division of labour- the hallmarks of true multicellularity. The reason that the experiment described here could not lead to that precise evolutionary step is that selection was acting on a group level trait (mat strength) independent from reproduction. The method by which groups were propagated from one generation to the next was artificial; by homogenising and diluting the cultures it was ensured that all characteristics of a group would be passed on, including genetically distinct types. If groups were propagated using a single bacterial cell then it is very unlikely that two distinct phenotypes could have evolved within a genetically identical population.

Eukaryotes have large genomes with the ability to regulate gene expression by methylation; information for many cell types may lie within a single genome. As bacteria do not possess such mechanisms bacterial multicellularity may be inherently limited to one or two cell types, which could be produced by high frequency random switching

between two phenotypes, as seen in bet hedging strategies. Indeed, even though bacterial multicellularity is thought to have evolved approximately 500 mya before the first multicellular eukaryotes (MICHOD 2007), they have never evolved comparable complexity.

Evolution by natural selection requires the production of heritable variation. Eukaryotic genomes differ in several important ways from those of prokaryotes; they are larger, abundant in non-coding DNA, undergo frequent recombination and can silence genes by methylation. Within these differences we may find two alterative, although not mutually exclusive, explanations for the advantage eukaryotes have over prokaryotes in the evolution of multicellular complexity. First, prokaryotes may not be able to generate sufficient variation for the evolution of complex traits. Secondly, prokaryotes may not have the ability to encode multiple distinct cell types from a single genome, required for the evolution of specialised cell types.

Following the application of conflict mediator construct to the group selected lines the increase of group fitness plateaued, with no sign of increasing. This suggested that there were no further beneficial mutations being produced, so that the groups could not respond to selection. This happened despite the groups not being restricted to a single cell for propagation; the group could evolve distinct genetic types, so that they got the benefit of multiple cell types without the requirement of the multiple cell types being encoded by a single genome. This provides some support for the hypothesis that the lack of the ability to produce heritable group level variation limits the evolution of prokaryotic multicellularity. The alternative hypotheses may be tested with the development of eukaryotic models of cooperation.

5.3.5 CONCLUDING COMMENTS.

Groups may draw on variation inaccessible to individuals; that of the genetically determined interactions between individuals within groups (GOODNIGHT 2000; GOODNIGHT and STEVENS 1997). This increased range of selectable traits may provide groups with new evolutionary solutions and potentially be the driving force for individuals to unite and make the transition to a new, higher level of selection. The *de novo* evolution of a division of labour described here show that group complexity can quickly evolve; however the evolutionary future of such groups is dependent on the suppression of cheats.

5.4 BIBLIOGRAPHY

ARNOLD, S. J., 1992 Constraints on phenotypic evolution. Am. Nat. **140:** S85-S107.

BONNER, J. T., 1999 The history of the cellular slime moulds as a "model system" for developmental biology. J. Biosci. **24:** 7-12.

BONNER, J. T., 2003 On the origin of differentiation. J. Biosci. **28:** 523-528.

BONNER, J. T., 2004 Perspective: The size-complexity rule. Evolution **58:** 1883-1890.

BORAAS, M. E., D. B. SEALE and J. E. BOXHORN, 1998 Phagotrophy by a flagellate selects for colonial prey: A possible origin of multicellularity. Evol. Ecol. **12:** 153-164.

BUSS, L. W., 1987 *The Evolution of Individuality*. Princeton University Press, Princeton.

FRANK, S. A., 1995 Mutual policing and repression of competition in the evolution of cooperative groups. Nature **377:** 520-522.

FUKAMI, T., H. J. E. BEAUMONT, X. X. ZHANG and P. B. RAINEY, 2007 Immigration history controls diversification in experimental adaptive radiation. Nature **446:** 436-439.

GOODNIGHT, C. J., 2000 Heritability at the ecosystem level. Proc. Natl. Acad. Sci. USA **97:** 9365-9366.

GOODNIGHT, C. J., and L. STEVENS, 1997 Experimental studies of group selection: what do they tell us about group selection in nature? Am. Nat. **150:** S59-S79.

HERRON, M. D., and R. E. MICHOD, 2008 Evolution of complexity in the volvocine algae; transitions in individuality through Darwin's eye. Evolution **62:** 436-451.

KIRK, D. L., 2005 A twelve-step program for evolving multicellularity and a division of labour. Bioessays **27:** 299-310.

LEVINS, R., 1968 *Evolution in Changing Environments: Some Theoretical Explorations*. Princeton University Press, Princeton.

MALONE, J. G., R. WILLIAMS, M. CHRISTEN, U. JENAL, A. J. SPIERS *et al.*, 2007 The structure-function relationship of WspR, a response regulator with a GGDEF output domain. Microbiology-(UK) **153:** 980-994.

MAYNARD-SMITH, J., and E. SZATHMARY, 1995 *Major transitions in evolution*. Freeman, Oxford.

MICHOD, R. E., 1998 Evolution of individuality. J. Evol. Biol. **11:** 225-227.

MICHOD, R. E., 2007 Evolution of individuality during the transition from unicellular to multicellular life. Proc. Natl. Acad. Sci. USA **104:** 8613-8618.

MICHOD, R. E., and M. D. HERRON, 2006 Cooperation and conflict during evolutionary transitions in individuality. J. Evol. Biol. **19:** 1406-1409.

MICHOD, R. E., A. M. NEDELCU and D. ROZE, 2003 Cooperation and conflict in the evolution of individuality IV. Conflict mediation and evolvability in *Volvox carteri*. Biosystems **69:** 95-114.

MICHOD, R. E., Y. VIOSSAT, C. A. SOLARI, M. HURAND and A. M. NEDELCU, 2006 Life-history evolution and the origin of multicellularity. J. Theor. Biol. **239:** 257-272.

PFEIFFER, T., and S. BONHOEFFER, 2003 An evolutionary scenario for the transition to undifferentiated multicellularity. Proc. Natl. Acad. Sci. USA **100:** 1095-1098.

RAINEY, P. B., and K. RAINEY, 2003 Evolution of cooperation and conflict in experimental bacterial populations. Nature **425:** 72-74.

ROSS, P., H. WEINHOUSE, Y. ALONI, D. MICHAELI, P. WEINBERGEROHANA *et al.*, 1987 Regulation of cellulose synthesis in *Acetobacter-Xylinum* by cyclic diguanylic acid. Nature **325:** 279-281.

SACHS, J. L., 2008 Resolving the first steps to multicellularity. Trends Ecol. Evol. **23:** 245-248.

SOBER, E., and D. S. WILSON, 1998 *Unto Others: The Evolution and Psychology of Unselfish Behaviour*. Harvard University Press, Cambridge, MA.

SPIERS, A. J., J. BOHANNON, S. M. GEHRIG and P. B. RAINEY, 2003 Biofilm formation at the air-liquid interface by the Pseudomonas fluorescens SBW25 wrinkly spreader requires an acetylated form of cellulose. Mol. Microbiol. **50:** 15-27.

SPIERS, A. J., S. G. KAHN, J. BOHANNON, M. TRAVISANO and P. B. RAINEY, 2002 Adaptive divergence in experimental Populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. Genetics **161:** 33-46.

SPIERS, A. J., and P. B. RAINEY, 2005 The *Pseudomonas fluorescens* SBW25 wrinkly spreader biofilm requires attachment factor, cellulose fibre and LPS interactions to maintain strength and integrity. Microbiology-(UK) **151:** 2829-2839.

WADE, M. J., 1976 Group selection among laboratory populations of *Tribolium*. Proc. Natl. Acad. Sci. USA **73:** 4604-4607.

6 Concluding Discussion

6.1 INTRODUCTION

The adaptation of *Pseudomonas fluorescens* SBW25 to the spatially structured static microcosm has proven an excellent model for the study of evolutionary processes (BANTINAKI *et al.* 2007; BUCKLING *et al.* 2000; BUCKLING and RAINEY 2002; FUKAMI *et al.* 2007; GAL *et al.* 2003; GIDDENS *et al.* 2007; GOYMER *et al.* 2006; KASSEN *et al.* 2004; KNIGHT *et al.* 2006; MACLEAN *et al.* 2004; MALONE *et al.* 2007; RAINEY 2007; RAINEY and RAINEY 2003; RAINEY and TRAVISANO 1998; SPIERS *et al.* 2003; SPIERS *et al.* 2002; SPIERS and RAINEY 2005). The WS system possesses all of the benefits of microbial models of evolution (see section 1.3); moreover the distinct niche preference and morphological difference between the WS and its ancestor are reminiscent of macro-evolutionary change. The striking WS phenotype combined with the relative ease with which its genetic causes are pursued facilitate the explicit experimental connection of genotype and phenotype. The initial aim of this thesis was to discover and characterise all genetic pathways to the WS phenotype. However, insight gained during the elucidation of these pathways provided the opportunity to conduct experiments determining the genetic dynamics underlying the first stage of adaptation to the air liquid interface, as well as the further evolution of WS group level complexity. In this chapter I develop and discuss ideas seeded by the studies described in this thesis, and then make my final remarks.

6.2 DGC DOMAIN PROTEIN NETWORKS AS A MODEL OF EVOLVABILITY.

The *P. fluorescens* DGC network normally modulates the precise physiological response of the cell to the change between sessile and motile living conditions. Although robust to physiological perturbation, a simple mutation de-couples production of c-di-GMP from environmental signals and produces the adaptive WS phenotype, a state that the ancestor cannot attain by mere physiological regulation. The WS morphotype, also known as rugose or rdar, has been documented in *V. paraheamolyticus E. coli, P. aeruginosa, E. sakazaki, A. actinomycemcomitus, V. cholerae* and throughout *Salmonella.* WS have enhanced survival in harsh environments (ROMLING 2005), forming stable biofilms with elaborate secondary structure. Extensive studies, especially on *V. cholerae* and *S. typhimurium*, have revealed that like *P. fluorescens,* the WS of these respective species form mats in static microcosms and are caused by genetic or regulatory changes in c-di-GMP production (BEYHAN and YILDIZ 2007; KADER *et al.* 2006; LIM *et al.* 2006; LIM *et al.* 2007; RASHID *et al.* 2003; SIMM *et al.* 2007).

The mode of regulation between the WS and SM states differs between species. *V. cholerae* undergoes phase variation, a strategy by which contingency loci mutate at a high frequency (see chapter 4) so that each generation has significant proportions of both WS and SM morphotypes. *S. typhimurium* undergoes a non-genetic physiological change into the WS state induced by environmental conditions. *A. actinomycetemcomitans*, a coloniser of the human mouth, is permanently in WS mode (FINE *et al.* 1999). For many of the species able to generate WS, environmental conditions alternatively favour either the WS or SM phenotype. As the rate of environmental change varies, it is predicted that the species will evolve either mechanisms that allow switching between the two phenotypes quickly, or the robust maintenance of one of the phenotypic states. Species that utilise the DGC network to produce WS cover a wide range of rates at which the

phenotypic change is brought about. Populations utilising contingency loci always have a significant proportion of the population in either the WS of SM state; such populations thus respond instantly to conditions favouring either the SM or WS phenotype. Populations requiring environmental signals to induce physiological change respond slightly more slowly, however, species such as *P. fluorescens* react to change the slowest, requiring a mutation and then fixation of the rare allele. DGC proteins provide an ideal example of how a network of proteins has responded to selection for evolvability (the ability to respond to selection), resulting in different rates of switching between the WS/SM phenotype across a wide range of species. Indeed, observing the rate of adaptation to WS/SM favouring conditions could provide clues as to the life histories of a wide range of bacteria. The idea that the production of variation has been fine-tuned in order to maximize the evolutionary potential of populations is controversial, if not new (SNIEGOWSKI and MURPHY 2006). The driving force behind the degree of evolvability of a population in WS/SM selective conditions is the rate of change between environments favouring either type, and the large apparent tradeoffs in the fitness of each type. A trade-off would prohibit the evolution of generalist able to compete in both environments. Although DGC networks may not have evolved to be evolvable *per se*, their evolvability certainly has been co-opted and preserved as an adaptive phenotype.

6.3 THE RELATIVE ROLE OF GENETIC STRUCTURE: MUTATIONAL DISTANCE.

Chapter two describes the characterisation of the *aws* and *mws* loci which, like *wsp*, sustain mutations resulting in the constitutive activation of DGC proteins. Previous work identifying the *wsp* and *aws* loci had led to the formation of two expectations at the undertaking of this project. Firstly, loci able to sustain WS causal mutations should be DGCs, secondly, as there are 43 putative DGC proteins in SBW25 (SILBY *et al.* 2008), its

potential to adapt to the air liquid interface would remain unperturbed until all DGC loci had been deleted from its genome. The demonstration that all WS evolved from the wild type strain were caused by mutations in *wsp*, *aws* or *mws*, even though other equally fit WS were possible, comprised the first empirical confirmation that genetic structure biases the course of evolution. Previously the vague term genetic constraints had been used to describe the putative characteristics of a genome that may bias the production of variation (ANTONOVICS and VANTIENDEREN 1991; BRAKEFIELD 2006; MAYNARD-SMITH 1985; SCHLUTER *et al.* 2004). Here I have defined what elements of genetic structure cause the observed patterns of WS evolution and have generalised them to three characteristics that differ within genomes. Studies in this thesis have focussed on the increased likelihood of some genes and nucleotide sites to produce adaptive variation, here I discuss in depth the significance of how the number of mutations required to attain the adaptive phenotype separates the *aws*, *wsp* and *mws* loci from all other DGC proteins.

Mutational routes that can achieve an adaptive phenotype with one mutation have a massive temporal advantage over those requiring two mutations- especially if the two mutations do not individually confer any selective advantage. The probability of two individually neutral mutations coinciding in the same organism is extremely low, compared to the rate at which a single beneficial mutation can increase to fixation in a population. Such a scenario can be easily thought of in the case of the WS causal mutations. The only adaptive mutations we see are in the *aws*, *mws* and *wsp* loci, which disrupt DGC post translational regulation (for example, inactivation of WspF causes derepression of WspR) indicating that there are sufficient levels of expression independent of the WS mutation. The other DGC containing loci superficially appear to have the same regulatory mechanisms as *aws*, *wsp* and *mws*- there are 11 other putative proteins with the same DGC/PDE domain structure as MwsR (such as SwsR).

However, if these loci are not expressed, they may first require a mutation activating expression before a mutation causing constitutive post-translational activation could have any phenotypic effect. Even if such an expression activating mutation is adaptive in its own right the likelihood that it will occur in tandem with the second mutation is low until the first mutation increases to a high frequency. In future work, a simple experiment to test this theory would be to over express a DGC/PDE protein in the *ΔwspΔawsΔmws* mutant, the gene encoding this protein should then become the most likely target of a WS causal mutation.

6.4 THE PREDICTABILITY OF EVOLUTION

It has been said that "A theory can predict to the extent that it can describe and explain" (BUNGE 1959). Some (MAYR 1988) have claimed that this statement (made by a physicist) does not apply to evolutionary biology, whose theories can describe and explain, but can not make predictions for the future. Despite this, predicting the path of adaptation is the holy grail of any evolutionist, and with a greater molecular understanding of macro-evolutionary processes there is no reason why prediction cannot become as much the realm of biologists as it is of physicists[1]. In this thesis progress was made towards two kinds of predictions; the genetic location of the next mutation to fix in an adaptive walk and the fitness effects of that genotype. Chapters two and four identified several elements that can render genes more likely than others to sustain adaptive mutations. Genetic architecture and mutable characteristics of sequence were seen to increase the likelihood of certain genes and even nucleotide sites contributing the

[1] Max Delbruck said of the fact that the behavior of quantum particles can only be assigned probabilities "now we cannot account for the behaviour of even a single hydrogen atom" DELBRUCK, M., 1949 A physicist looks at biology. Trans. Conn. Acad. Arts Sci. **33:** 173-190.

next allele to fix in an adaptive walk. Knowledge of an organism's physiology and genome sequence could theoretically be used to predict sites most likely to sustain mutations. Accurate predictions may not be possible without a much deeper understanding of microbial physiology than currently held, however the WS system could provide an ideal system to generate and test predictions. The mutable loci discovered in *awsX* and *mwsR* provide some indication of how an elevated rate of mutation could have occurred. The mutable sequences do not stand out as unusual in the genome; rather it is the coinciding of such sequences with conserved motifs that may provide the clue to determining where the mutations will happen.

Chapter three took the first steps towards constructing a general model describing the size of steps taken during an adaptive walk. The possibility that the normal distribution could describe the distribution of fitness effects for adaptive walks for all complex organisms comprises a significant extension to previous theory. The next step is a formal mathematical model describing the likelihood that a mutation of a given effect size will fix, dependant on the shape of the underlying distribution; such models will be constructed within the robust framework of the Central Limit theorem.

In these studies of the genetics of adaptation demonstrate that precise predictions are impossible; predictions can only be assigned a probability. If studies of *P. fluorescens* had led to the *a priori* prediction that WS recovered from microcosms incubated for many days would be caused by *wspF*, and that the first WS would be caused by *awsX* mutations, these would have been right only half the time. These results may have utility however. In pathogenic microbial populations selection is very strong and mutations may be restricted to fewer loci. In these populations the average effect will be a more accurate predictor of what actually may happen during adaptation.

6.5 Final Comments

The studies in this thesis comprise several significant advances in our understanding of evolutionary genetics and the evolution of group level complexity. Genetic constraints have been demonstrated and defined for the first time, the mutational landscape model extended and conflict mediation and the *de novo* evolution of group level complexity observed in action. These findings break new ground, however, none of them contradict theory. Rather they contribute by providing a more detailed account of the evolutionary process. The veracity of these claims can only be confirmed by further experimentation, and will soon be submitted to the academic community for peer review.

6.6 BIBLIOGRAPHY

ANTONOVICS, J., and P. H. VANTIENDEREN, 1991 Ontoecogenophyloconstraints - the chaos of constraint terminology. Trends Ecol. Evol. **6:** 166-168.

BANTINAKI, E., R. KASSEN, C. G. KNIGHT, Z. ROBINSON, A. J. SPIERS *et al.*, 2007 Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. Genetics **176:** 441-453.

BEYHAN, S., and F. H. YILDIZ, 2007 Smooth to rugose phase variation in *Vibrio cholerae* can be mediated by a single nucleotide change that targets c-di-GMP signalling pathway. Mol. Microbiol. **63:** 995-1007.

BRAKEFIELD, P. M., 2006 Evo-devo and constraints on selection. Trends Ecol. Evol. **21:** 362-368.

BUCKLING, A., R. KASSEN, G. BELL and P. B. RAINEY, 2000 Disturbance and diversity in experimental microcosms. Nature **408:** 961-964.

BUCKLING, A., and P. B. RAINEY, 2002 The role of parasites in sympatric and allopatric host diversification. Nature **420:** 496-499.

BUNGE, M., 1959 *Causality*. Harvard University Press, Cambridge.

DELBRUCK, M., 1949 A physicist looks at biology. Trans. Conn. Acad. Arts Sci. **33:** 173-190.

FINE, D. H., D. FURGANG, H. C. SCHREINER, P. GONCHAROFF, J. CHARLESWORTH *et al.*, 1999 Phenotypic variation in *Actinobacillus actinomycetemcomitans* during laboratory growth: implications for virulence. Microbiology-(UK) **145:** 1335-1347.

FUKAMI, T., H. J. E. BEAUMONT, X. X. ZHANG and P. B. RAINEY, 2007 Immigration history controls diversification in experimental adaptive radiation. Nature **446:** 436-439.

GAL, M., G. M. PRESTON, R. C. MASSEY, A. J. SPIERS and P. B. RAINEY, 2003 Genes encoding a cellulosic polymer contribute toward the ecological success of *Pseudomonas fluorescens* SBW25 on plant surfaces. Mol. Ecol. **12:** 3109-3121.

GIDDENS, S. R., R. W. JACKSON, C. D. MOON, M. A. JACOBS, X. X. ZHANG *et al.*, 2007 Mutational activation of niche-specific genes provides insight into regulatory networks and bacterial function in a complex environment. Proc. Natl. Acad. Sci. USA **104:** 18247-18252.

GOYMER, P., S. G. KAHN, J. G. MALONE, S. M. GEHRIG, A. J. SPIERS *et al.*, 2006 Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. II. Role of the GGDEF regulator WspR in evolution and development of the wrinkly spreader phenotype. Genetics **173:** 515-526.

KADER, A., R. SIMM, U. GERSTEL, M. MORR and U. ROMLING, 2006 Hierarchical involvement of various GGDEF domain proteins in rdar morphotype development of *Salmonella enterica* serovar *Typhimurium*. Mol. Microbiol. **60:** 602-616.

KASSEN, R., M. LLEWELLYN and P. B. RAINEY, 2004 Ecological constraints on diversification in a model adaptive radiation. Nature **431:** 984-988.

KNIGHT, C. G., N. ZITZMANN, S. PRABHAKAR, R. ANTROBUS, R. DWEK *et al.*, 2006 Unraveling adaptive evolution: how a single point mutation affects the protein coregulation network. Nature Genet. **38:** 1015-1022.

LIM, B., S. BEYHAN, J. MEIR and F. H. YILDIZ, 2006 Cyclic-diGMP signal transduction systems in Vibrio cholerae: modulation of rugosity and biofilm formation. Mol. Microbiol. **60:** 331-348.

LIM, B., S. BEYHAN and F. H. YILDIZ, 2007 Regulation of Vibrio polysaccharide synthesis and virulence factor production by CdgC, a GGDEF-EAL domain protein, in *Vibrio cholerae*. J. Bacteriol. **189:** 717-729.

MACLEAN, R. C., G. BELL and P. B. RAINEY, 2004 The evolution of a pleiotropic fitness tradeoff in Pseudomonas fluorescens. Proc. Natl. Acad. Sci. USA **101:** 8072-8077.

MALONE, J. G., R. WILLIAMS, M. CHRISTEN, U. JENAL, A. J. SPIERS *et al.*, 2007 The structure-function relationship of WspR, a response regulator with a GGDEF output domain. Microbiology-(UK) **153:** 980-994.

MAYNARD-SMITH, J., R. BURIAN, S. KAUFFMAN, P.ALBERCH, J. CAMPBELL, B. GOODWIN, R. LANDE, D. RAUP, AND L WOLPERT., 1985 Developmental constraints and evolution. Quarterly Review of Biology **60:** 265-287.

MAYR, E., 1988 *Towards a New Philosophy of Biology: Observations of an Evolutionist*. Harvard University Press, Boston, MA.

RAINEY, P. B., 2007 Unity from conflict. Nature **446:** 616-616.

RAINEY, P. B., and K. RAINEY, 2003 Evolution of cooperation and conflict in experimental bacterial populations. Nature **425:** 72-74.

RAINEY, P. B., and M. TRAVISANO, 1998 Adaptive radiation in a heterogeneous environment. Nature **394:** 69-72.

RASHID, M. H., C. RAJANNA, A. ALL and D. K. R. KARAOLIS, 2003 Identification of genes involved in the switch between the smooth and rugose phenotypes of Vibrio cholerae. FEMS Microbiol. Lett. **227:** 113-119.

ROMLING, U., 2005 Characterization of the rdar morphotype, a multicellular behaviour in Enterobacteriaceae. Cell. Mol. Life Sci. **62:** 1234-1246.

SCHLUTER, D., E. A. CLIFFORD, M. NEMETHY and J. S. MCKINNON, 2004 Parallel evolution and inheritance of quantitative traits. Am. Nat. **163:** 809-822.

SILBY, M. W., A. M. CERDEÑO-TÁRRAGA, G. VERNIKOS, S. R. GIDDENS, R. W. JACKSON *et al.*, 2008 Genomic and functional analyses of diversity and plant interactions of *Pseudomonas fluorescens*.

SIMM, R., A. LUSCH, A. KADER, M. ANDERSSON and U. ROMLING, 2007 Role of EAL-containing proteins in multicellular behavior of *Salmonella enterica* serovar typhimurium. J. Bacteriol. **189:** 3613-3623.

SNIEGOWSKI, P. D., and H. A. MURPHY, 2006 Evolvability. Curr. Biol. **16:** R831-R834.

SPIERS, A. J., J. BOHANNON, S. M. GEHRIG and P. B. RAINEY, 2003 Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. Mol. Microbiol. **50:** 15-27.

SPIERS, A. J., S. G. KAHN, J. BOHANNON, M. TRAVISANO and P. B. RAINEY, 2002 Adaptive divergence in experimental Populations of Pseudomonas fluorescens. I. Genetic and phenotypic bases of wrinkly spreader fitness. Genetics **161:** 33-46.

SPIERS, A. J., and P. B. RAINEY, 2005 The *Pseudomonas fluorescens* SBW25 wrinkly spreader biofilm requires attachment factor, cellulose fibre and LPS interactions to maintain strength and integrity. Microbiology-(UK) **151:** 2829-2839.

7 MATERIALS AND METHODS

7.1 MATERIALS

7.1.1 MEDIA AND GROWTH CONDITIONS

P. fluorescens cultures were grown at 28°C in 30 ml glass vials containing 6 ml of King's Medium B (KB: 10 g/l glycerol; 20 g/l Proteose Peptone No.3; 1.5 g/l Mg_2SO_4; 1.5 g/l K_2HPO_4 (King *et al.*, 1954)) or Lysogeny Broth (LB: 10 g/l NaCl, 20 g/L bacto-peptone, 10 g/L yeast extract). Cultures were incubated statically or were shaken at 150 rpm.

E. coli cultures were grown at 37°C and incubated in LB (supplemented with antibiotics and substrates where appropriate) in plastic 30 ml vials.

All bacterial strains created in this work are archived at -80°C in 70% glycerol saline solution (8.5 g NaCl, 300 ml H_20, glycerol to 1 litre).

Agar plates were prepared by adding 1.5% w/v agar to the broth media before autoclaving, cooling to 50^0C then pouring into petri dishes (also supplemented with antibiotics or substrates where appropriate).

7.1.2 BACTERIAL STRAINS

Strain	Description	Reference
SBW25	Ancestral, wild type, environmental isolate, smooth SM)	(Rainey and Bailey, 1996)
SBW25-LacZ	SBW25 with a lacZ transcriptional fusion	Zhang, X. X. unpublished
IWS_A	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007
IWS_B	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007
IWS_C	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007
IWS_E	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007
IWS_F	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007
IWS_G	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007
IWS_J	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007
IWS_L	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007
IWS_N	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007
IWS_O	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007
IWS_S	Independent WS from day 7 of static incubation	McDonald *et al.*, 2008
IWS_T	Independent WS from day 7 of static incubation	McDonald *et al.*, 2008
IWS_U	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007
IWS_V	Independent WS from day 7 of static incubation	McDonald *et al.*, 2008
IWS_W	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007

IWS_Y	Independent WS from day 7 of static incubation	Bantinaki *et al*, 2007
LSWS	Large Spreading Wrinkly Spreader derived from SBW25	
AWS	Alternative Wrinkley Spreader	McDonald et al., 2008
SM^{MSC}	WS reporter strain, TetR fused to the wss promoter And inserted together with the entire pUIC3 plasmid into the *P. fluorescens* genome upstream of the native *wss* operon.	Fukami *et al*, 2007

7.1.3 PLASMIDS AND TRANSPOSONS

Table -1: Names and characteristics of plasmids and transposons

Strain	Description	Reference
Plasmids		
pCR8TOPO	Spe^R, pUC *ori*, 2.8kb sequencing plasmid	Invitrogen
pUIC3	Tc^R, *mob*, *oriR6K*, *bla*, Δpromoter-lacZY	(Rainey, 1999)
pRK2013	Km^R, incP4, tra, mob, mobilization plasmid used as a helper in tri-parental mating	(Figurski and Helinski, 1979)
Transposons		
IS-Ωkm/hah	Km^R, ColE1, *ori*, *npt* promoter, *loxP*	

7.1.4 PRIMERS

Name	Sequence	Target
Aws1	CGA GAT CTG GAG CGT CTG GTA CTG G	External 5' *awsXRO* deletion
Aws4	GCA GAT CTC CAT CGT TAT GTA CAT AGG	External 3' *awsSRO* deletion
Aws2	CAG CAT GCG GAT CCG TTG ACG GGCA TGA ACG TGG CCG TCT C	Internal 5' *awsXRO* deletion
Aws3	CAG CAT GCG GAT CCG TTG ACG GGTG GTG ATG TCG GAC TGA GC	Internal 3' *awsXRO* deletion
awsXrv	GCA GAT CTG TCG TGG TTG GCC TGG TGG	External 3' *awsX* deletion
awsXfw	AGC GTC TTG CAA CTG TCC CGT C	Internal 3' *awsX* deletion
Tn*phoA*-II	GTG CAG TAA TAT CGC CCT	IS-Ωkm/*hah* (seq. tn. Inserts)
CEKG2A	GGC ACG CGT CGA CTA GTA CAN NNN NNN NNN AGA G	SBW25 degenerate primer (seq. tn. Inserts)
CEKG2B	GGC CAC GCG TCG ACT AGT CAN NNN NNN NNN ACG CC	SBW25 degenerate primer (seq. tn. Inserts)
CEKG2C	GGC CAC GCG TCG ACT AGT CAN NNN NNN NNN GAT AT	SBW25 degenerate primer (seq. tn. Inserts)
hah-1	ATC CCC CTG GAT GGA AAA CGG	5' end of CEGK2A,B & C
CEKG4	GGC CAC GCG TCG ACT AGT AC	IS-Ωkm/*hah*(seq. tn. Inserts)
AxFw	GGC CGT CTC GCC CAC AGA GC	Seq. awsX
AxRv	CGG CGA CGG GAC AGT TGC AA	Seq. awsX
Wss1	GCC GTT TCT GAC CAC TCA GC	Screening wss insertion muts
Wss2	CGT CAA CAG TAT CGG CAG CC	Internal *mws* primer
Wss3	CGC CAC CTT CTT CTG TGG CT	Internal *mws* primer
Wss4	CGC GTG GTG TTC AGC AAG G	3' *mws*
Wss5	GGC AAC GCC ATC GTG CTG GT	Reverse internal *mws* primer
Wss6	CCG CTG CCA ACC AAC CGC T	3' *aws* primer
Wss7	CGC AAC CTG ATC GTG CGT GT	5' *aws* primer
Wss8	GCG CCA GGC CTT GCA ACT G	Reverse internal *aws* primer

Wss9	CGG ATG CCT TGA GCG GCT T	Reverse internal *aws* primer
Wss10	CCG TGG CTT GGC GCA AAG C	Reverse internal *aws* primer
Wss11	CGC GTC GTT GCA TGA GTT GC	Reverse internal *aws* primer
Wss12	CCG ACA CCG GCA CCA AGT G	Internal *aws* primer
Wss13	CCG AGC TGG CGC AAC GTG AT	Internal *aws* primer
Wss14	CGG CAC TGA CCA TGT ACG C	5' *wspA*
Wss15	GCG CTG ATC GGC ACG TCG T	Reverse Internal *wspA*
18mer	CGA TGA CGC CCT GCT GGA	5' *wspF*
zrwspF2	TAT TTT CTT CAT GGG CCA GG	3' *wspF*
Mws1fw	CCG CGC CTG CAG CAC GCG TA	5' *mwsR*
rv	GGA CGA GCG TCG CCA ATG CG	3' *mwsR*
Mws2fw	CGC ATT GGC GAC GCT CGT CC	5' *mwsR*
Rv	GCG ACA GGC GCA GCG CCC	3' *mwsR*
Mws3fw	GGG CGC TGC GCC TGT CGC	5' *mwsR*
Rv	CCT CGG TGA TGC GCA GTT CC	3' *mwsR*
Mws4fw	GGA ACT GCG CAT CAC CGA GG	5' *mwsR*
Rv	CCT GGC GTT GTT GCT GGC GC	3' *mwsR*
Mws5fw	GCG CCA GCA ACA ACG CCA GG	5' *mwsR*
Rv	GCG ACC GTC AGC GCG TAG GC	3' *mwsR*
Mws6fw	GCC TAC GCG CTG ACG GTC GC	5' *mwsR*
Rv	CGC TGC TCG CTG GCC TTG CG	3' *mwsR*
Mws7fw	CGC AAG GCC AGC GAG CAG CG	5' *mwsR*
rv	GGC TCA GCT CGT TGC CGT CC	3' *mwsR*
Mws8fw	GGA CGG CAA CGA GCT GAG CC	5' *mwsR*
rv	GGG CTT CGT TGA CTT CGC GC	3' *mwsR*
Mws9fw	GCG CGA AGT CAA CGA AGC CC	5' *mwsR*
rv	GGC TGT GAG TGC CTA ATC GG	3' *mwsR*
Aaw1	AGA TCT GCG TCA TGT CTG CAC TCG CG	5' external *mwsR* deletion
Aaw2	CAG CAT GCG GAT CCG TTG ACG GGG CGG CGT GAC TTT GGA CAA	5' internal *mwsR* deletion
Aaw3	CCG TCA ACG GAT CCG CAT GCT GCG GTT CGA GGC GCA GTT CAG	3' interna *mwsR* deletion
Aaw4	AGA TCT GGC CGT TGG CAT CGA TAC CG	3'external *mwsR* deletion
M13F	GTA AAA CGA CGG CCA G	5' PCR8TOPO MCS
M13R	CAG GAA ACA GCT ATG AC	3' PCR8TOPO MCS
Bla	CAG GGT TAT TGT CTC ATG AGC G	5' pUIC3 MCS
P-lacZ	TGG GAT TAA CTG CGC GTC GCC	3' pUIC3 MCS

7.1.5 ANTIBIOTICS AND MARKERS

Unless otherwise specified, all chemicals were obtained from Sigma. Antibiotics were purchased from Melford Laboratories and used in the following concentrations: kanamycin 50 μg/ml (15 μg/ml for the group selection experiments), gentamycin

40μg/ml, spectinomycin 100 μg/ml, tetracycline 10 μg/ml, N-[5-Nitro-2-furfurylidene]-1-aminutesohydantoin (NF) 100 μg/ml (dissolved in dimethyl formamide – DMF) and 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside (X-gal; dissolved in DMF) 40 μg/ml.

7.2 METHODS

7.2.1 DNA PREPARATION

Genomic template DNA was prepared by centrifugation of 200 ml of overnight culture followed by resuspension in 500 µl of water. Cells would lyse due to osmotic pressure and lysate was used for PCR.

Plasmid DNA was prepared using the Qiagen plasmid preparation kit.

7.2.2 POLYMERASE CHAIN REACTION (PCR)

Polymerase chain reactions were performed using a DNA Engine DYAD Peltier Thermal Cycler (Bio-Rad). Reactions were performed in 200 µl tubes with a total volume of 25 µl containing: 5 µl 10x PCR Buffer, 1 µl dNTP mix (200 µM), 1 µl 50 mM $MgCl_2$ and 0.5 µl Taq polymerase, 1 µl of each primer (10 pmol) and the remainder and deionised water added to make up 50 µl. Template DNA was either cell lysate or plasmid DNA. For each template-primer combination, the thermal cycler program was optimised. Each cycling program started with an initial denaturation step (95°C) for 3 minutes followed by 25-30 cycles consisting of a denaturation step (94°C for 30 s), an annealing step (50-60°C for 30-45 s specific to each primer pair) and an extension step at 72°C for an appropriate time depending on the desired product length. The reaction completed with an additional extension step at 72°C for 10 minutes to finish partially synthesised products. Typically, the PCR was followed by analysis using gel electrophoresis.

7.2.3 ELECTROPHORESIS

Electrophoretic gels were prepared with 1% w/v agarose and 0.5 x TBE (90 mM Tris-HCl pH 8.0, 0.55% boric acid, 2 mM EDTA). Ethidium bromide (10 μg/ml) was incorporated into the gel for visualisation of DNA under ultraviolet light. Samples were loaded with a loading buffer containing 0.25% bromophenol blue and 30% glycerol. Gels were submerged in 0.5x TBE in a gel electrophoresis tank and electrophoressed at a constant voltage of ~110 mV until bands had attained sufficient resolution. A 1 kb DNA ladder (Invitrogen) was used as a size marker suitable for sizing DNA from 500 bp-10 kb.

7.2.4 DNA EXTRACTION

In the case where desired fragments appeared as clear single bands of DNA on agarose gels, DNA was recovered directly from the PCR reaction using a QIAquick PCR purification kit (QIAGEN). If multiple fragments were present, desired fragments were isolated by cutting them out of the agarose gel with a sterile razor blade while been visualised under UV light. DNA was recovered from the agarose using the QIAquick gel extraction kit (QIAGEN). DNA samples were eluted in sterilised de-ionised water and DNA concentration measured using a nanodrop (Nanodrop Technologies, Inc).

7.2.5 DNA SEQUENCING

DNA was purified as described then sent to Macrogen Incorporated (Korea) for sequencing. Confirmation of a mutant sequence required two-fold coverage of the mutated region.

7.2.6 ALLELIC REPLACEMENT

Oligonucleotide primers were designed to amplifly ~1000bp regions flanking the gene(s) of interest. A third PCR reaction was used to fuse these two pcr fragments, this product was topo cloned into pCR8TOPO and sequenced to check for errors. This fragment was excised from the pCR8 vector via a *Bgl* II digest, then cloned into *Bgl* II digested suicide vector pUIC3. Recombination with the wildtype allele is selected for by plating on LB with tetracycline and X-gal. Enrichment for double recombinants leads to a 1:1 ratio of wild type to recombinant strains. Deletion mutants consisted of a clean replacement of the wild type allele with the introduced allele. Mutants were confirmed by pcr.

7.2.7 TRANSFORMATION

DNA and competent cells were chilled on ice for 10 minutes. 2 µl of DNA was added to 40 µl of chemically competent cells, mixed and then incubated on ice for 5-30 minutes. The cells were heat-shocked at 42°C for exactly 30 seconds, then 500µl of SOC medium was added (previously warmed to 37°C). This mixture was incubated at 37°C shaking for 1 hour to assist recovery of cells before dilution and plating on and LB plates containing appropriate selective and/or indicative reagents.

7.2.8 RESTRICTION ENZYME CLEAVAGE

The restriction enzymes used were from Invitrogen or New England Biolabs. 3.5 µl of DNA was digested with 0.5 µl of restriction enzyme with 1µl of 10x Buffer 3 and water added to make up to 10 µl reaction volume. Reactions were left at 37°C for one hour. The product of the digest was run on a gel. If the fragment was to be used in a cloning reaction, CIP (calf intestinal phosphatase, NEB) was added at the end of the reaction

for 30 minutes at 37°C to avoid re-ligation of compatible ends. CIP was deactivated by heating to 70°C for 10 minutes.

7.2.9 BI-PARENTAL CONJUGATION

P. fluorescens (recipient) and *E. coli* containing ISΩkm/hah (donor) were grown overnight. 1 ml of each of the cultures was pelleted at 13,000 rpm and resuspended in 800 µl of pre-warmed LB to ensure that no antibiotics were included in the conjugation mixture. 200 µl of *P. fluorescens* was heat-shocked for 20 minutes at 45°C. 800 µl of *E. coli* was added to the heat-shocked cells and the mixture was spun at 12,000 rpm for 1 minute. The supernatant was removed and the cell mixture was resuspended in 50 µl of ¼ Ringer's Solution. The suspension was gently pipetted onto a sterile, 25 mm-diameter nitrocellulose filter (0.22 µm pores, Millipore) on an LB plate and incubated for 2-4 hours at 28°C. After incubation, cell were covered by placing the filter in 1 ml of deionised water and vortexed for 15 seconds. The filter was removed and the suspension appropriately diluted and plated on LB plates supplemented with kanamycin (to select for the transposon) and NF (to counterselect the *E. coli* donor).

7.2.10 TRI-PARENTAL CONJUGATION

As donor cells were *mob* deficient *E. coli* DH5α, a tri-parental strategy was implemented. In addition the acceptor *P. fluorescens* strain a third strain (DH5α containing plasmid pRK2013 expressing *mob* transfer proteins) was used to assist in facilitate transfer of genetic material from donor to recipient. DH5α, pRK2013 and the recipient *P. fluorescens* strain were grown up overnight. 200 µl of the recipient strain was heat-shocked at 42°C and mixed with 200 µl of DH5α and 200 µl of pRK2013 cells. The cells were spun at

13,000 rpm for 1 minute to form a pellet. The supernatant was removed and the cells were resuspended in 50 µl of LB and applied to an LB plate. The plate was incubated overnight at 28°C. The cells were scraped off the LB plate and suspended in a 1.5ml Eppendorf tube containing 1 ml of deionised water. Dilutions were spread on LB plates containing appropriate antibiotics and NF to counterselect the *E. coli* cells. The plates were incubated for 2 days at 28°C and then screened for transconjugants.

7.2.11 TRANSPOSON MUTAGENESIS

Transposon mutagenesis was performed using the *E. coli* strain ISKm/hah according to a method developed in the Rainey laboratory; following a bi-parental conjugation (see Section 7.2.9). The conjugation mix was diluted and plated on LB plates supplemented with kanamycin (to select for the transposon) and NF (to counterselect the *E. coli* donor). After 2 days of incubation at 28°C the colonies that were picked were those that exhibited a phenotypic change from WS to SM indicating that the phenotype had been altered by the insertion of the transposon. Colonies were grown up overnight with kanamycin and stored at -80°C. Transformants were sequenced at the site of insertion and by using blast comparisons exact transposon insertion positions were mapped to their position in the *P. fluorescens* genome. A description of and the rationale behind further optimisation of this procedure is included in the appendix. An Artemis (see section 0.2.13) file containing the exact sites of transposon mutants described in this study is included as supplementary data.

7.2.12 FITNESS ASSAYS

A fitness measure was obtained for each WS genotype by direct competition with the LSWS; genetically modified to contain the *lacZY* genes. Competing strains were

inoculated into KB broth and shaken overnight, then 50ul was taken to found six replicate overnight cultures also in KB. No antibiotics or selective media were used during the preculture stage to ensure that all strains are at comparative stages of growth. Each of the six replicates for each strain were taken to found a microcosm together with equal amounts of the LSWS *lacZY* reference strain. The ancestral strain was also added to the microcosm to minimise competition between WS in the broth phase, thereby enhancing reproducibility of the experiments. The *lacZY* marked strain was distinguished from its competitor by plating cells on LB plates supplemented with X-gal (on this medium *lacZY* marked strains are readily distinguished by their blue colour). The competitive fitness of WS genotypes was determined as the ratio of the Malthusian parameters of the strains being compared (Lenski, 1991).

7.2.13 ARTEMIS

Artemis (Rutherford *et al.*, 2000) is free software available online and for download at http://www.sanger.ac.uk/Software/Artemis/. This software was used to view the entire annotated *P. fluorescens* genome available for private use at http://www.sanger.ac.uk/Projects/P_fluorescens/private/.

7.2.14 GROUP SELECTION EXPERIMENT.

The three experimental lines consisted of eight microcosms each inoculated with picked colonies of the AWS WS. Microcosms were incubated statically for 7 days at 28ºC, mats were assayed for strength by lowering glass beads (each weighing 32.5mg) carefully onto the mat until the mat detached from the glass or beads passed through the mat. The strongest mat in each line was used to found the next generation of eight microcosms. In the three control lines the weakest microcosms were chosen to found the next

generation of microcosms. After the measurement of mat strength, each microcosm was vortexed for 30 seconds then 700µl mixed with an equal volume of glycerol saline and stored at -80ºC. In order to estimate the proportion of cheats comprising each microcosm, microcosms were diluted and plated on solid KB media.

7.2.15 RECIPROCAL INVASION ASSAYS.

The WS_1 and WS_2 types were incubated overnight with shaking at 28ºC, and the cultures used to inoculate six microcosms in a 100:1 ratio WS_1:WS_2 and six more the opposite WS_2:WS_1. The ratio of WS_1:WS_2 was scored before and after incubation for 24hours and the ratio of each type scored.

7.2.16 ASSAY FOR MAT PERSISTENCE AT THE BROTH SURFACE.

A single microcosm was set upon the surface of an orbital shaker, the shaker was set to 100 rpm and then the time until the mat completely detached from the glass wall determined. Following detachment, the time until the mat sunk was measured.

8 APPENDICES

Appendices are organised by chapter for example, appendices for chapter 2 are prefixed 8.2.

8.2 CHAPTER TWO

8.2.1 OPTIMISATION OF TRANSPOSON MUTAGENESIS

Chapter two describes four transposon mutageneses, the first two were carried out by Stephanie Gehrig on the LSWS and Δ*wsp*WS as part of her Ph.D thesis (Gehrig, 2007) the third and fourth by myself of the Δ*aws*Δ*wsp*WS and Δ*aws*Δ*wsp*Δ*mws*WS respectively. The LSWS and Δ*wsp*WS mutageneses were exceedingly comprehensive in order to ensure that all possible WS suppressor loci were recovered. Following mutagenesis of the Δ*aws*Δ*wsp*WS strain it was considered desirable to avoid sequencing transposon insertions in genes that had already been discovered in previous studies, especially *wss*, which had comprised 35% of sequenced transposon insertions in both the LSWS and Δ*wsp*WS mutageneses. Two strategies were employed. Firstly, only WS^- revertants that formed healthy fast growing colonies were picked. It is known that transposon insertion in *wss* revert the WS phenotype by disrupting cellulose biosynthetic machinery; these mutants would continue to produce large amounts of c-di-GMP thus inducing physiologically expensive cellular processes. Presumably these WS^- revertants sustain a large fitness cost compared to those with insertions in genes resulting in suppression of c-di-GMP production; the proximal cause of the WS phenotype. Secondly, primers were designed to cover the entire *wss* operon. Before undertaking to sequence transposon insertion mutants PCRs were performed using a 3' primer designed to anneal with the transposon and 15 5' *wss* primers. Using this method only three from 45 mutants were noted as having *wss* insertions. Twenty of the remaining transposon mutants were sequenced; no insertions in *wss* were recovered. Seven of the twenty were found to have insertions in a

novel GGDEF/EAL locus was later designated *mwsR*, indicating that the sequenced strains were enriched for transposon mutants of interest.

8.3 CHAPTER THREE

8.3.1 THEORETICAL AND SAMPLE QUANTILES FOR FITTED CURVES.

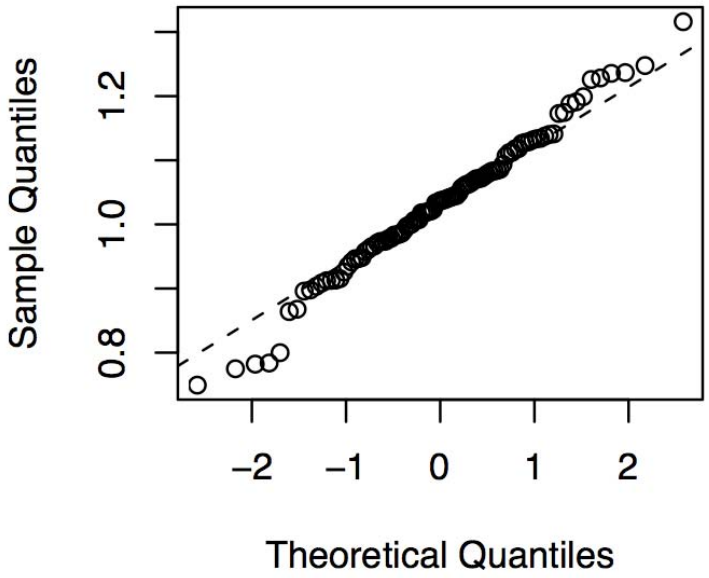


Figure 8.1 Regression of sample and theoretical quantiles showing the goodness of the fit of the normal distribution for the static environment.

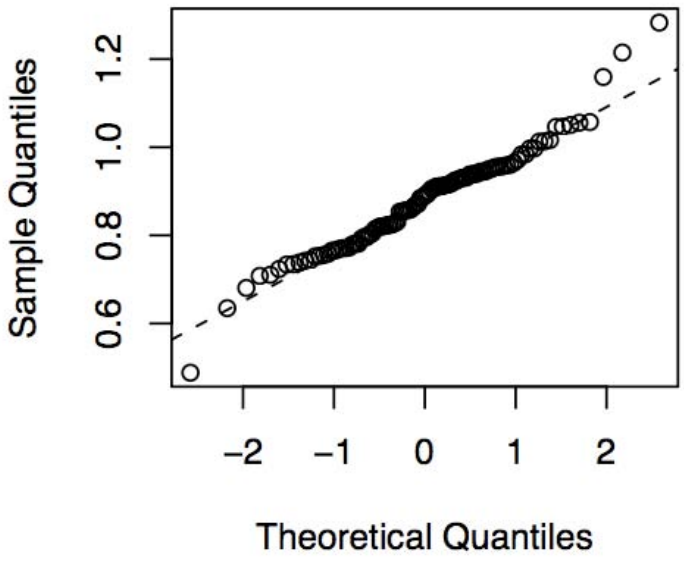


Figure 8.2 Regression of the sample and theoretical quantiles showing the goodness of fit of the normal distribution for the shaken environment.

8.3.2 MAXIMUM LIKELIHOOD TABLE FOR STATIC AND SHAKEN FITNESS MEASUREMENTS.

	Log likelihood		-2log likelihood		Dif. -2LL norm.-x		Chi^2
	static	shaken	static	shaken	static	shaken	static
gamma	-59.28	36.60	118.5	73.20	44.87	73.24	$2.09e^{-11}$
exponential	-36.60	-6.12	73.20	-12.25	90.23	158.7	
normal	-81.72	-73.22	163.4	146.4			

Table 8.1The normal distribution fit better with less parameters that the gamma and exponential.

8.4 CHAPTER FOUR

8.4.1 THE BIRTHDAY PROBLEM PROGRAM

```
package mini;
import util.*;
// simulates the birthday problem 10000 to calculate the likelihood
that more than one person has its bday on one day
//numberSites=number of days a year has
//number of trials is the number of people in a room
//number of simulations=number of rooms where the experiment is made
//lower bound= minum number of people whose bday is on the same day
//Event = the event has to take place at least that many times to be
counted
//the question of the current program is, how often dow we observe a
room where at least lowerBound number of people
// have their bday at the same day and this event has to occur Event
many times in the room to be counted
public class MikeSimulation {
      public static void main(String[] args){
            int numberSites=Integer.parseInt(args[0]);
            int numberTrials=Integer.parseInt(args[1]);
            int numberSimulations=Integer.parseInt(args[2]);
            int lowerBound=Integer.parseInt(args[3]);
            int Event=Integer.parseInt(args[4]);
            RandomDistributionSimulation                        rd=new
RandomDistributionSimulation(numberTrials,1,0,numberSites);
            int numberExceeded=0;
            Histogram<Integer>                                   h=new
Histogram<Integer>(rd.posList.toArray(new Integer[0]));
            for(int i=1;i<numberSimulations;i++){
      numberExceeded+=checkExceed(h.getValues(),lowerBound,Event);
                  rd=new
RandomDistributionSimulation(numberTrials,1,0,numberSites);
                  h=new     Histogram<Integer>(rd.posList.toArray(new
Integer[0]));
            }
            System.out.println(numberExceeded);
      }
      private  static  int  checkExceed(Double[]  freq,int  lower,int
Event){
            int x=0;
            for(int i=0;i<freq.length;i++){
                  if(freq[i]>lower) x++;
            }
            if(x>=Event)return 1;
            else return 0;

      }
```

8.4.2 MUTATIONS DISCOVERED IN CHAPTER FOUR.

TABLE 8.2

Mutation	Gene
MwsR WS	
G2778A	*mws*R
DEL3068-3076	*mws*R
DEL3068-3076	*mws*R
G2976A	*mws*R
DEL3065-3073	*mws*R
A3391C	*mws*R
G3244A	*mws*R
G2778A	*mws*R
DEL3066-3074	*mws*R
DEL3066-3074	*mws*R
G2778A	*mws*R
G2778A	*mws*R
DEL3066-3074	*mws*R
G2976A	*mws*R
DEL3066-3074	*mws*R
DEL3068-3076	*mws*R
G2778C	*mws*R
DEL3066-3074	*mws*R
DEL3068-3076	*mws*R
DEL3066-3074	*mws*R

G2778A	*mwsR*
Day zero	
C608A	*wspF*
Del 228-261	*awsX*
Del 99-138	*awsX*
Del 140-185	*awsX*
Del 155-173	*awsX*
Del 99-138	*awsX*
Del 99-138	*awsX*
Day1	
7bp insert428	*wspF*
Del 99-138	*awsX*
A290C	*wspF*
Del 228-261	*awsX*
Del 228-261	*awsX*
Del 228-261	*awsX*
Del 228-261	*awsX*
Del 228-261	*awsX*
Del 228-261	*awsX*
Del 228-261	*awsX*
Day2	
Del 228-261	*awsX*
Del 228-261	*awsX*
Del 228-261	*awsX*
Del 228-261	*awsX*

Del 228-261	*awsX*
Day3	
T613A	*wspF*
T540G	*wspF*
13bp insert695	*wspF*
C624T	*wspF*
Del 228-261	*awsX*