Evolutionary and molecular origins of a phenotypic switch in *Pseudomonas fluorescens* SBW25

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

Ph.D.

in

Evolutionary Genetics

at Massey University, Auckland, New Zealand.

Jenna Gallie

2010

Abstract

Survival in the face of unpredictable environments is a challenge faced by all organisms. One solution is the evolution of mechanisms that cause stochastic switching between phenotypic states. Despite the wide range of switching strategies found in nature, their evolutionary origins and adaptive significance remain poorly understood. Recently in the Rainey laboratory, a long-term evolution experiment performed with populations of the bacterium *Pseudomonas fluorescens* SBW25 saw the *de novo* evolution of a phenotypic switching strategy. This provided an unprecedented opportunity to gain insight into the evolution and maintenance of switching strategies.

The derived 'switcher' genotype was detected through colony level phenotypic dimorphism. Further microscopic examination revealed the cellular basis of phenotypic switching as the bistable (ON/OFF) expression of a capsule. Transposon mutagenesis demonstrated that the structural basis of the capsule was a colanic acid-like polymer encoded by the Pflu3656-*wzb* locus. Subsequently, whole genome re-sequencing enabled elucidation of the series of mutational events underlying the evolution of capsule bistability: nine mutations were identified in the switcher. Present in both forms of the switcher, the final mutation – a point mutation in a central metabolic pathway – was shown to be the sole mechanistic cause of capsule switching; it 'set the stage' for a series of molecular events directly responsible for bistability.

Two models were proposed to explain capsule switching at the molecular level: the genetic amplification-reduction model, and the epigenetic feedback model. Collective results of biochemical and genetic assays proved consistent with the epigenetic model, whereby a decrease in flux through the pyrimidine biosynthetic pathway activates an already-present feedback loop. Subsequent analysis of a second switcher (evolved independently of and in parallel with the first) revealed a radically different genetic route leading to phenotypically and mechanistically similar capsule switching.

In addition to providing the first empirical insight into the evolutionary bases of switching strategies, the work presented in this thesis demonstrates the power of natural selection – operating on even the simplest of organisms – to forge adaptive solutions to evolutionary challenges; in a single evolutionary step, selection took advantage of inherent intracellular stochasticity to generate an extraordinarily flexible phenotype.

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Paul Rainey for encouragement, ideas and discussions throughout this research. Thanks also to my co-supervisor, Dr. Mathew Goddard from the University of Auckland. I am grateful for the opportunity to present and discuss this research at the Evolution Conference (Christchurch, NZ; 2007) and the Gordon Research Conference in Microbial Population Biology (New Hampshire, USA; 2007). Attendance at these conferences – and indeed the entire project – would not have been possible without the generous financial support of a Tertiary Education Committee Bright Future Top Achiever's Doctoral Scholarship.

I would like to acknowledge the endless insights and observations contributed by past and present members of the Rainey lab. Particular thanks to Dr. Hubertus Beaumont for his prior work on the reverse evolution experiment, and continued interest and collaboration. For assistance with fitness assays, I thank Dr. Christian Kost. Thanks to Frederic Bertels for invaluable bioinformatic assistance and for writing the cell counting programme used throughout this thesis. For support with transposon mutagenesis of $6w^4$, I thank Sylke Nestmann. Thanks to Yunhao Liu for efficient technical assistance and washing all those microcosms!

I am grateful to Professor Ahmed Abdelal, Professor Susan Powers-Lee and Dr. James Thoden for their expert insight into the effects of *carB* mutations, and Professor Mogens Kilstrup for sharing his extensive knowledge of nucleotide biosynthetic pathways. Thanks also to Professor James Sneyd for helpful discussions about mathematical switching models, and to Professor Marti Anderson for statistical expertise. Special thanks to Dr. David Ackerley for kindly providing the pSX over-expression system and assisting with associated cloning and expression protocols.

For providing suggestions for the improvement of this thesis, I would like to thank Professor Paul Rainey, Dr. Hubertus Beaumont, Dr. Christian Kost, Dr. Michael McDonald and, in particular, Frederic Bertels.

For the opportunity to present this research to a wide range of audiences, I thank the Tertiary Education Committee, organisers of the MacDiarmid Young Scientists of the Year Awards. Special thanks to Professor Paul Rainey, Professor Gaven Martin, Dr. Hubertus Beaumont, Frederic Bertels and Elaine Gallie for supporting my entry through to the finals of the 2009 competition.

I thank Professor Don Love and Dr. Brian Murray for the opportunity to tutor and demonstrate various undergraduate student courses.

Finally, I would like to express my gratitude to my family and friends for the endless patience and personal support, without which this thesis would have never been completed. Thanks for always being there Mum, Frederic and Rachele!

Table of Contents

Table of Abbreviations

Abbreviation	Meaning
ACP	Acetylated cellulosic polymer
AP-PCR	Arbitrary primed-polymerase chain reaction
BLAST	Basic local alignment search tool
bp	base pairs
c-di-GMP	Cyclic-dimeric-guanosine monophosphate
CP	Carbamoyl phosphate
CPSase	Carbamoyl phosphate synthetase
DGC	Di-guanylate cyclase
dNTP	dinucleotide triphosphate
DUF	Domain of unknown function
g	gram/gravity
GDP/GTP	Guanosine di-/tri- phosphate
GDP-Fuc	GDP-L-fucose
Gm	Gentamicin
HK	Histidine kinase
KB/kb	King's medium B/kilobase
Km	Kanamycin
LB	Lysogeny broth
M-W-W test	Mann-Whitney-Wilcoxon test
NF	Nitrofurantonin
OD	Optical density
PDE	Phosphodiesterase
REE	Reverse evolution experiment
RR(r)	Response regulator (receiver)
SM	Smooth morph
SSI	Site-specific inversion
SSM	Slipped-strand mispairing
Tc	Tetracycline
TCSTP	Two-component signal transduction pathway
UDP-Gal	UDP-D-galactose
UDP-Gluc	UDP-D-glucose
UDP-GlucA	UDP-D-glucuroinc acid
UMP/UDP/UTP	Uracil mono-/di-/tri- phosphate
WS	Wrinkly spreader
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

Chapter 1:
Introduction

1.1 Phenotypic flexibility is the cornerstone of life

The ability to accommodate environmental change is essential for the survival and evolution of all species. Accordingly, organisms have developed a plethora of mechanisms for responding to challenges posed by the environment. Individual organisms are able to sense environmental cues and respond with appropriate phenotypic modifications, a process known as 'acclimation'. In cases where acclimation results in a sub-optimal phenotype, environmental challenges may be met with adaptation – a population level evolutionary response mediated by natural selection. For adaptation to occur, the following two conditions must be satisfied: (1) population members must display heritable variation in the relevant phenotype, and (2) this variation must result in differential reproductive success. Under these circumstances, the phenotype (and corresponding genotype) best suited to survival spreads through the population over successive generations (reviewed in Stearns & Hoekstra, 2005: p.1-4).

The efficacy of adaptation is dependent upon the generation of heritable variation *via* mutation during DNA replication and repair. As any given mutation has a relatively high probability of causing deleterious effects, rates of mutation are inevitably low - mutation rates vary between species; however, they have been estimated as $\sim 10^{-9}$ and $\sim 10^{-10}$ mutations *per* base pair (bp) *per* generation in eukaryotes and prokaryotes (Drake *et al.*, 1998) - and consequently the generation of genetic variation is a slow process. However, survival in unpredictable, rapidly changing environments depends upon correspondingly rapid phenotypic flexibility. Organisms have responded to this challenge by evolving mechanisms that enable rapid, heritable switching of phenotypes within clonal populations. These mechanisms have been the subject of intensive investigation, and a number of molecular causes have been identified. Some phenotypic switching mechanisms have a genetic basis, achieving switching through selective

elevation of mutation rates in relevant genomic loci (known as 'contingency loci' (Moxon *et al.*, 1994)). In recent years, phenotypic switching mechanisms that do not have a genetic basis have attracted increasing attention. By definition, such mechanisms are epigenetic (Rando & Verstrepen, 2007). Examples of both genetic and epigenetic switching mechanisms are discussed in the following sections.

1.1.1 Phenotypic switching mechanisms with genetic bases: contingency loci

Contingency loci are genomic regions that reversibly mutate between different forms at high frequency ($>10^{-5}$ mutations *per* locus *per* generation), causing heritable alterations in protein expression and phenotype (reviewed in Moxon *et al.*, 1994; Henderson *et al.*, 1999; Hallet, 2001; van der Woude & Bäumler, 2004; Wisniewski-Dyé & Vial, 2008). To date, the vast majority of documented contingency loci occur in genes encoding virulence factors of pathogenic bacteria. Variation in the expression of such virulence factors enables populations to rapidly generate phenotypic dichotomy with respect to the relevant virulence factor (usually cell surface components such as flagella or pili), while simultaneously conserving essential functions encoded by other genes. The majority of contingency loci result in simple ON/OFF switching of protein expression ('phase variation'), but some result in three or more expression states ('antigenic variation'). This distinction between phase and antigenic variation is largely superficial; they share many of the molecular mechanisms underlying mutation of the locus. A number of molecular mechanisms have been identified, the most common of which are site-specific inversion, slipped-strand mispairing and – if the definition of contingency loci is relaxed slightly – differential DNA methylation. Each of these mechanisms is discussed in detail in the following sections.

1.1.1.1 Site-specific inversion

Site-specific inversion (SSI) is a mutational mechanism based on repeated inversion (flipping) of a defined DNA element by specialized, directed recombinases. The invertible element is defined by short flanking inverted repeats (IRs), the sequence of which is recognised by the corresponding recombinase. The expression and activity of

the recombinase are often regulated by external environmental signals, giving an unconventional level of environmental control over the rate at which mutation occurs. Following recognition, the recombinase subsequently mediates a (reversible) homologous recombination event between the two IRs, resulting in the inversion of the DNA segment between the repeats. Phase variation occurs when the inversion event alters the transcription or translation of a structural gene. In one orientation of the element, gene expression occurs ('ON' phase) and in the other, it does not ('OFF' phase). There have been numerous documentations of SSIs mediating phase variation (Zieg *et al*., 1977; Abraham *et al*., 1985; Zhao *et al*., 1997; Krinos *et al*., 2001). Discussed below is one of the best-characterised examples of contingency loci mediated by SSI – that which underlies phase variable expression of *E. coli* type 1 fimbriae.

1.1.1.1.1 Type 1 fimbriae of uropathogenic E. coli

Type 1 fimbriae (also known as pili) are proteinaceous extensions found on the surface of uropathogenic *E. coli*. Encoded by the *fim* locus, type 1 fimbriae are subject to phase-variable expression mediated by inversion of a 314 bp regulatory element (Abraham *et al*., 1985; Freitag *et al*., 1985; reviewed in van der Woude & Bäumler, 2004). Inversion of the regulatory element occurs between two 9 bp inverted repeats that flank the promoter of *fimA*, which encodes the major fimbrial structural subunit. Thus, inversion of the regulatory element results in differential expression of *fimA*; when the promoter is in the correct orientation, transcription proceeds and type 1 fimbriae are expressed ('ON' phase). The alternative orientation results in loss of both *fimA* transcription and fimbrial expression ('OFF' phase) (see Figure 1.1).

The inversion event is mediated by two independently transcribed site-specific recombinases, FimE and FimB. Although FimE and FimB share 48 % amino acid identity (van der Woude & Bäumler, 2004), they have different DNA specificities. FimE acts preferentially when the regulatory sequence is in the 'ON' orientation, switching the locus to the 'OFF' orientation, while FimB mediates inversion in both directions (Figure 1.1). This setup is the basis for environmental regulation of the inversion event - the expression and activity of FimE and FimB can be differentially

modulated in response to environmental signals, including temperature. The optimal temperature for FimB activity is 37-40˚C, while FimE activity decreases as temperature increases towards 37˚C. The net effect of this is to increase the frequency of OFF→ON inversion events at 37˚C (and above) – temperatures that cells are likely to encounter during the fimbriae-dependent host infection process.

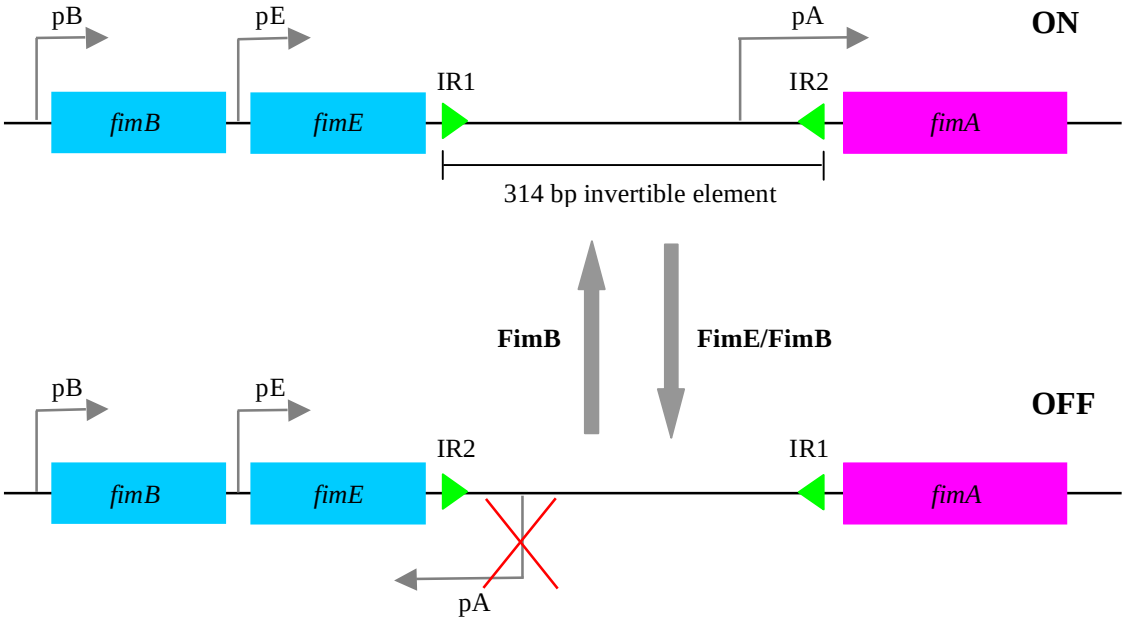


Figure 1.1: SSI-mediated phase variation of type 1 fimbriae of uropathogenic *E. coli*. Phase variation is controlled by the FimE/FimB-mediated inversion of a 314 bp element containing the promoter of *fimA* (p=promoter, pA=*fimA* promoter, IR=inverted repeat). Correct orientation of the *fimA* promoter leads to *fimA* transcription and expression of fimbriae ('ON' phase, top), while the opposite promoter orientation prevents *fimA* expression ('OFF' phase, bottom). Figure adapted from van der Woude & Bäumler (2004).

1.1.1.2 Slipped-strand mispairing

Repetitive DNA has long been recognised as an important feature of eukaryotic genomes. With the advent of whole genome sequencing technology, the prevalence of repetitive elements in prokaryotic genomes is becoming increasingly obvious. These repetitive elements usually take the form of variable number tandem repeats (VNTRs; also known as microsatellites) - homopolymeric tracts of repeated units between one and seven nucleotides in length. VNTRs are subject to reversible expansion and contraction *via* slipped strand mispairing (SSM; reviewed in Levinson & Gutman, 1987; van Belkum *et al*., 1999). SSM occurs as a result of occasional, stochastic misalignment

of mother and daughter DNA strands during DNA replication or repair. The misalignment may occur at the leading or lagging end of the repetitive region, resulting in the respective gain (expansion) or loss (contraction) of repeats in the newly synthesised DNA. In turn, these reversible genetic alterations can lead to changes in protein expression and phenotype. SSM within the coding region of a gene may alter the translational reading frame, leading to radically altered or truncated proteins (Peak *et al*., 1999; Kearns *et al*., 2004; Hendrixson, 2006). Depending on the genetic architecture of a locus, SSM in coding regions may generate phase or antigenic variation in protein expression. Alternatively, SSM can cause phase variable gene transcription; for example SSM may lead to change in the crucial spacing between the -10 and -35 sites for RNA polymerase binding, leading to distinct 'ON' and 'OFF' phases of gene expression (Willems *et al*., 1990). Intriguingly, the transcription of a *Haemophilus influenzae* type III restriction-modification system methyltransferase has been reported to phase-vary *via* SSM within a tetranucleotide repeat tract (De Bolle *et al*., 2000). As the methyltransferase itself functions to control transcription of numerous genes, this results in the coordinated phase variable expression of multiple genes, dubbed the 'phasevarion' (Srikhanta *et al*., 2005). A well-characterized example of SSM-mediated phase variation is described below.

1.1.1.2.1 *Opacity proteins in* Neisseriae

Outer membrane opacity proteins (Opa) proteins are cell surface proteins required by both *Neisseria gonorrhoeae* and *Neisseria meningitidis* for successful invasion of human hosts. *Neisseria* strains may express a range of Opa proteins, each encoded by its own independent operon. *N. gonorrhoeae* contains 11-12 Opa loci, while *N. meningitidis* encodes 3-4. Expression of each Opa locus is under independent (but mechanistically similar) SSM-mediated phase variable control (reviewed in Meyer *et al*., 1990). Collectively, Opa proteins in *Neisseria* are antigenically variable with cells expressing no Opa proteins, a single type of Opa protein, or one of many possible combinations of Opa proteins (reviewed in Virji, 2009). Opa loci consist of a single *opa* gene with a 5' signal peptide. Following export of the Opa protein, the signal peptide is removed from the amino terminus by proteolytic cleavage. Differential expression of the Opa protein is achieved by SSM of a pentameric coding repeat (CR; 5'-CTCTT-3')

within the signal peptide. The reading frame of the main coding segment of the *opa* gene is shifted according to the number of repeats present (Stern *et al.*, 1986; Stern & Meyer, 1987). Only one reading frame results in correct translation of the Opa protein ('ON' phase); the other two reading frames encode severely truncated Opa proteins ('OFF' phase) (Figure 1.2).

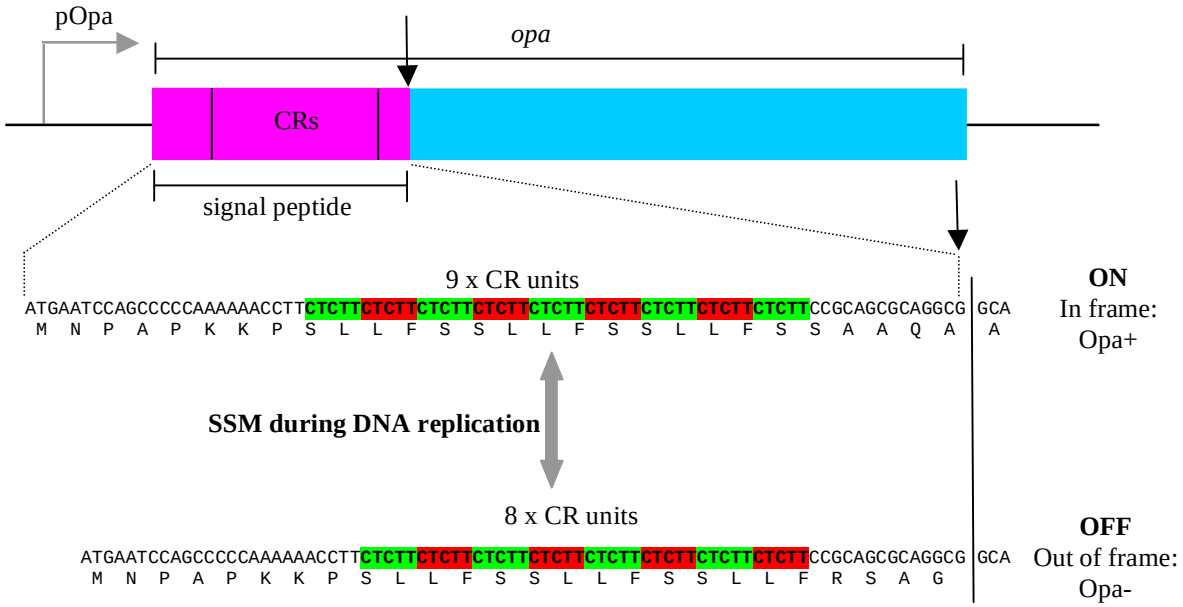


Figure 1.2: Prototypical SSM-mediated phase variation of a *Neisseria* Opa protein. The signal peptide of each Opa gene contains variable numbers of a pentameric coding repeat (CR; 5'-CTCTT-3'). The number of CRs determines whether the downstream Opa coding sequence is transcribed, and thus whether the Opa protein is expressed. The DNA and amino acid sequences for nine (in frame, Opa+; top) and eight (out of frame, Opa-; bottom) CRs are shown in detail. (p=promoter, black arrows indicate site of proteolytic cleavage of signal peptide after translation). Figure adapted from Meyer *et al.* (1990).

1.1.1.3 Differential DNA methylation

DNA methylation is a common mechanism for regulation of gene expression in eukaryotes and prokaryotes alike (reviewed in Mathews *et al.*, 2000: p.927-929; Casadesús & Low, 2006). During the past 20 years it has become apparent that, under the correct conditions, DNA methylation systems can cause phase variable gene expression in bacteria (Blyn *et al.*, 1990; Henderson *et al.*, 1997; Nicholson & Low, 2000; reviewed in Henderson *et al.*, 1999). In such cases, no change in DNA sequence occurs between the different phenotypic forms. Thus, strictly speaking, differential

DNA methylation is not a genetic switching mechanism. However, since methylation acts directly on DNA, it is included here as a pseudo-genetic mechanism. The logic behind this will become clearer when true epigenetic mechanisms - with minimal dependence on DNA - are discussed (see section 1.1.2).

In cases where phase variation results directly from differential methylation patterns, methylation is carried out by deoxyadenosine methylase (Dam), which methylates the adenosine of the short palindromic sequence 5'-GATC-3'. Usually, Dam-mediated methylation occurs on the adenosine of both the forward and reverse strands (full methylation), although in some cases (e.g. during DNA replication) only one strand is methylated (hemi-methylation). The essential feature of Dam methylation-mediated phase variation is competition for DNA binding sites between Dam and transcriptional regulatory proteins on newly-synthesized, hemi-methylated DNA: methylation prevents binding of transcriptional regulatory proteins, and binding of transcriptional regulatory proteins prevents methylation. Thus, there ensues something of a race (albeit unfairly weighted in favour of the previous phase) between Dam and the transcriptional regulator. The type of transcriptional regulator involved determines the 'ON' and 'OFF' phases – binding of a positive regulator will result in the 'ON' phase, while binding of a negative regulator will result in the 'OFF' phase. In more complex cases, both positive and negative regulators contribute to the final phenotypic outcomes. Such a case, the phase variable expression of the *E. coli* pap pilin, is described below.

1.1.1.3.1 Pap pilin of uropathogenic E. coli

In addition to phase-variable type 1 fimbriae (see section 1.1.1.1.1), a number of uropathogenic *E. coli* strains express a phase-variable pyelonephritis-associated pili (pap) pilin. Like type 1 fimbriae, the pap pilin is a cell surface structure that mediates adhesion to host tissues (reviewed in van der Woude *et al*., 1996). Reviewed in Hernday *et al*. (2002) and Casadesús & Low (2006), understanding of the molecular mechanisms responsible for phase variable expression of the pap pilin remains incomplete, and investigations in this area are ongoing. However, research to date has provided considerable insight into the mechanistic bases of pap pilin phase variation.

The pap pilin is encoded by the *pap* gene cluster, which contains nine genes (*papA-I*). Oscillation between the 'ON' and 'OFF' phases is controlled at the transcriptional level by Dam-mediated differential DNA methylation at two sites, GATC-1 and GATC-2. Both sites lie within the 416 bp *pap* regulatory region, and affect two divergently transcribed loci, *papI* and *papBA*. The 'ON' phase occurs when GATC-1 is unmethylated and GATC-2 is methylated, while the 'OFF' phase occurs when GATC-1 is methylated and GATC-2 is unmethylated (Figure 1.3). Switching between the states is controlled by the competitive action of Dam and Lrp, a transcriptional regulatory protein with complex activity. Lrp acts as both a repressor and an activator of the *papBA* promoter, depending on which of two binding sites it is occupying. When bound in the GATC-1 region, it acts as an activator of *papBA* transcription ('ON' phase), and when bound in the GATC-2 region it acts as a repressor ('OFF' phase). PapI, encoded by *papI*, stimulates the 'ON' phase by promoting Lrp binding in the GATC-1 region. Regulation of PapI expression is complex and poorly understood, but *papI* transcription appears to be promoted by PapB. Thus, it appears that the combined action of PapI and PapB perpetuates both the 'ON' and 'OFF' phases by influencing stochastic binding or removal of Lrp during the hemi-methylated state of DNA replication or repair.

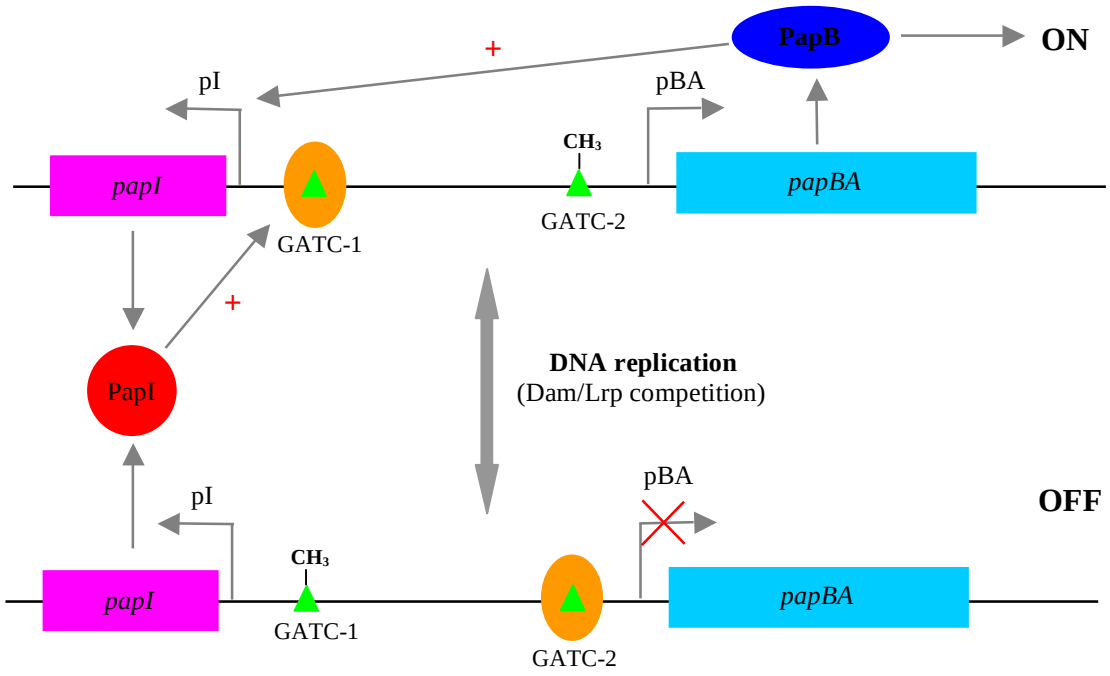


Figure 1.3: Dam methylation-mediated phase variation of the *E. coli* pap pilin. The intergenic region between *papI* and *papBA* contains two methylation sites, GATC-1 and GATC-2 (green triangles), differential methylation of which leads to differential gene transcription. See text for further details. (orange ellipse=Lrp, p=promoter, CH_3=methyl group).

1.1.2 Phenotypic switching with epigenetic bases: bistability

Traditionally, genetically identical microbial individuals (e.g. clonal bacterial populations) grown under identical environmental conditions have been assumed to be phenotypically uniform. This belief was born out of past experimental techniques, which generally measured average, population-level phenotypic responses. Recent advances in scientific technology have increased the resolution of phenotypic assays, enabling the phenotypes of individuals to be observed (Elowitz *et al.*, 2002; Rosenfeld & Rodwell, 2005). Consequently, it is becoming increasingly clear that isogenic individuals are not always, or perhaps even usually, phenotypically identical (reviewed in Dubnau & Losick, 2006; Veening *et al.*, 2008). It has been proposed that much phenotypic variation is owed to 'noise' - stochastic, cell-to-cell fluctuations in gene expression. Many proteins exist in relatively low copy numbers in cells and so small, stochastic changes in protein expression may lead to large downstream effects. In particular, it is easy to imagine that slight fluctuations in the expression of major transcriptional regulators may set off a snowball cascade of effects, culminating in the phenotypic bifurcation of a population (i.e. bistability; reviewed in Kœrn *et al.*, 2005).

An example of noise-generated bistability is provided by genetic competence in the soil bacterium, *Bacillus subtilis.* In stationary phase, some cells in genetically identical *B. subtilis* populations become genetically competent. That is, they gain the ability to take up naked DNA from the environment (Maamar & Dubnau, 2005; reviewed in Dubnau & Losick, 2006). The competent phenotype occurs in only 10 % of cells; 90 % of the (isogenic) population remain non-competent. Genetic and microscopic analyses indicate that competence bistability results from noise in the expression of *comK*, which encodes the master regulator of competence-related gene expression (see Figure 1.4). In exponential phase, ComK is rapidly degraded in all cells, and so competence is not achieved. When stationary phase is reached, a quorum sensing cascade leads to the expression of ComS, a protein that protects ComK from degradation. The working model for competence bistability is that cells reaching a threshold level of ComK spontaneously switch to the competent state (equally, those that drop below the ComK threshold switch back to the non-competent state). The source of the noise and the exact mechanism of switching have yet to be elucidated.

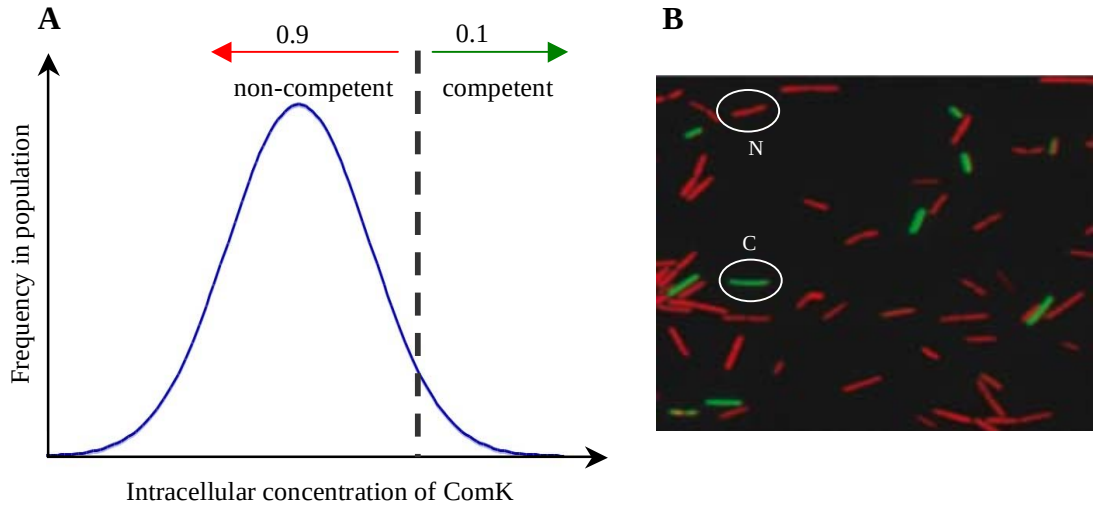


Figure 1.4: Noise in *comK* expression leads to phenotypic bistability in *B. subtilis*. (**A**) Graph showing an (arbitrary) distribution of the intracellular levels of ComK in a *B. subtilis* population. A threshold level of ComK exists (grey dotted line), below which individuals are non-competent, and above which cells switch to competence. (**B**) Fluorescence microscope image of stationary phase *B. subtilis* cells with a ComK-GFP translational fusion. Competent cells (C) appear green (high ComK and correspondingly high GFP). Non-competent cells (N) appear red (low ComK and low GFP, counter-stained with DNA-intercalating agent propidium iodide). Photograph adapted from Dubnau & Losick (2006).

1.1.2.1 Molecular mechanisms of bistability

An emerging concept in biology, population bistability is caused by stochastic differences in gene expression between cells. At the molecular level, the emergence of bistability requires one major component: a self-sustaining feedback system (reviewed in Ferrell Jr., 2002). These systems ensure that expression differences are self-perpetuating, leading to the observed semi-stability of each state. However, the mere presence of feedback is insufficient to cause bistability; the *degree* of feedback is crucial. If feedback is too strong (or weak), the system will exist continually in a single, unistable state. Two main types of feedback mechanisms have been proposed to cause bistability: positive autoregulation and double negative feedback (reviewed in Dubnau & Losick, 2006). Operation of these feedback systems has been tested *in vivo* by combining the key elements in *Saccharomyces cerevisiae* (Becskei *et al.*, 2001) and *E. coli* (Gardner *et al.*, 2000), respectively. Additionally, natural examples of bistability caused by both mechanisms have been described (Ptashne, 1992; Maamar & Dubnau, 2005; Smits *et al.*, 2005; Zordan *et al.*, 2006).

1.1.2.1.1 Positive autoregulation controls Candida albicans *opaque-white switching*

The pathogenic yeast *Candida albicans* produces cells in two distinct phenotypic states – opaque and white (Slutsky *et al*., 1987). It has been demonstrated that these types are heritable, and each can give rise to the other type (Rikkerink *et al*., 1988). Although named for the types of colonies they produce, each state is a composite of several differential phenotypes. Microarray analyses by Tsong *et al*. (2003) showed that these phenotypic differences are due to the differential expression of many genes. More recently, Zordan *et al*. (2006) provided empirical evidence to support a molecular model whereby white-opaque switching results from stochastic fluctuations in the expression of *wor1*, a gene encoding a master transcriptional regulator (Figure 1.5). The authors demonstrated that Wor1 binds to its own promoter, and activates its own transcription. Thus, *wor1* and its protein product constitute a self-sustaining feedback loop; a particular Wor1 threshold exists, above which cells switch spontaneously to the opaque form. Once reached, the threshold level tends to be maintained by self-perpetuating Wor1 expression. In cells where Wor1 stochastically falls below the threshold (e.g. by a decrease in Wor1 on cell division), a spontaneous switch to the white form occurs.

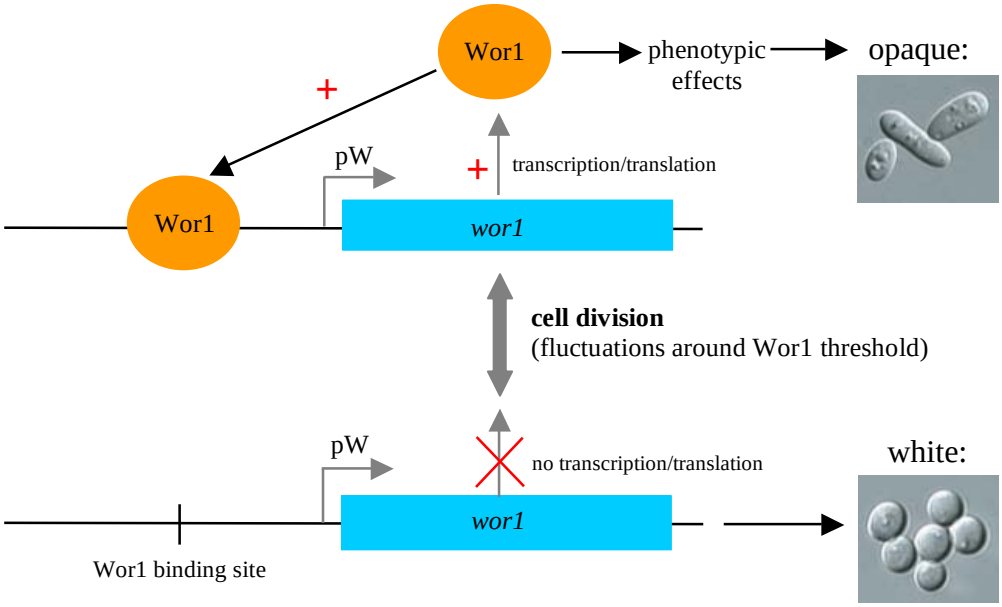


Figure 1.5: Positive autoregulation-mediated white-opaque bistability in *C. albicans*. Wor1, the master regulator of white-opaque bistability forms a self-sustaining feedback loop (indicated by +). In addition to regulating genes involved in the white and opaque phenotypes, Wor1 binds to its own promoter, stimulating transcription. The opaque phenotype (top) results from the accumulation of Wor1 above the required threshold. The white phenotype (below) occurs when the level of Wor1 stochastically drops below this threshold. Microscopic cell images reproduced from Zordan *et al*. (2006). (p=promoter).

1.1.2.1.2 Double negative regulation controls phage lambda lysogeny-lysis switching

Phage lambda is a well-studied bacteriophage that infects *E. coli* by injecting its DNA into the host cytoplasm. Depending on environmental conditions, subsequent events follow one of two mutually exclusive pathways: the lytic or lysogenic pathway. In the lytic pathway, the bacteriophage uses the host cell machinery to replicate (reviewed in Ptashne, 1992). Upon phage-mediated lysis of the host cell, newly synthesized phage particles are released into the environment. The state of lysogeny occurs when the incoming phage DNA integrates itself into the host chromosome. The phage DNA is replicated along with the host DNA, and thus the phage spreads passively through vertical transmission. The 'decision' of which pathway to follow is mediated by cII, a phage-encoded protein that is rapidly degraded by bacterial proteases. Under prosperous environmental conditions (e.g. rich growth medium) the relevant proteases are active and cII levels are low, stimulating the lytic phase. In poor environments (e.g. starvation), protease activity is low, and corresponding high cII levels activate the lysogenic phase. In some lysogenic *E. coli* cells, environmental changes will cause the phage DNA to excise, flipping the cell to the lytic phase. Thus, lambda-infected *E. coli* populations are bistable with respect to the phage lifecycle (Ptashne, 1992).

Phage lambda DNA contains a small number of essential genes, including the structural genes for phage envelope, and two genes encoding transcriptional repressor proteins, Cro and cI. The concerted action of these two regulatory proteins generates a double negative feedback loop that is the basis of the lytic-lysogenic bistable switch. In the passive lysogenic phase, the only lambda protein expressed is cI, which represses transcription of the remaining, divergently transcribed genes. As long as cI is synthesized, the lysogenic state will be maintained (Figure 1.6). Under certain environmental conditions (e.g. in the presence of UV radiation), the cI repressor is inactivated by RecA-mediated cleavage. Once cI repression is alleviated, *cro* and the downstream structural genes are transcribed, triggering the switch to the lytic phase. The repressor encoded by *cro* binds to the *cI* regulatory region, stabilising the lytic state by preventing cI synthesis.

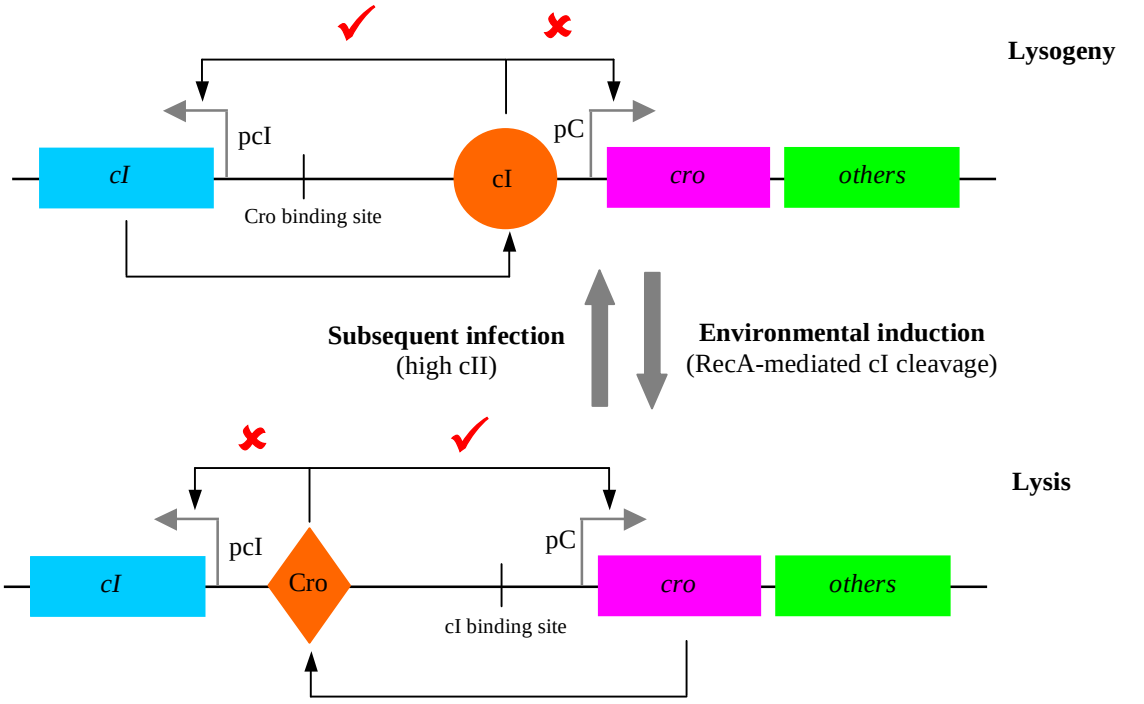


Figure 1.6: Double negative feedback-mediated bistability in the phage lambda lysogeny-lysis switch. The phage lambda genome encodes two mutually repressing repressors – cI and Cro. In the lysogenic state, cI binds to the *cro* promoter and represses Cro biosynthesis (and biosynthesis of lytic structural proteins). Environmental signals (e.g. UV radiation) trigger the RecA-mediated degradation of cI, allowing the production of Cro. Cro binds to the *cI* regulatory region, repressing the synthesis of cI and thereby enabling the continued synthesis of Cro (and other lytic proteins). When lytic phase progeny infect new host cells, each progeny enters either the lytic or lysogenic cycle depending on environmental signals received *via* the phage protein, cII. Figure compiled from Ptashne (1992). (p=promoter).

1.1.3 Adaptive significance and evolutionary origins of phenotypic switching

The prevalence of mechanisms that generate phenotypic diversity among pathogens has led to the assumption that these mechanisms are an evolutionary response to the harsh selection imposed by the host immune system. Recent descriptions of phase variable expression of non-virulence genes in pathogens (Saunders *et al.*, 2000; Seib *et al.*, 2002) and non-pathogenic bacteria (Kearns *et al.*, 2004) indicate that this hypothesis is too simplistic. Additionally, the existence of epigenetic switching mechanisms in a wide range of prokaryotes, yeast and multi-cellular eukaryotes suggests a far-reaching ecological significance (reviewed in Ferrell Jr., 2002; van der Woude, 2006). It seems probable that phenotypic switching mechanisms represent nature's 'bet-hedging' strategies, whereby populations are able to 'spread the risk' phenotypically, increasing

the chance of survival in unpredictable environments (Ferrell Jr., 2002; Veening *et al*., 2008).

In addition to the current adaptive significance of phenotypic switching mechanisms, their evolutionary origins remain elusive (Gould & Lewontin, 1979). It has been proposed that the fluctuating selection imposed by unpredictably changing environments is a major driving force of switch evolution (Slatkin, 1974; Ancel Meyers & Bull, 2002; Kussell & Leibler, 2005; Wolf *et al*., 2005). Several additional factors have been suggested to play a role in switch evolution, including the capacity of a population to respond to environmental changes by mutation and selection ('evolvability'; Wolf *et al*., 2005), frequency of environmental change (Kussell & Leibler, 2005), and the biotic makeup of the environment (Lancaster & Masel, 2009). Despite considerable theoretical interest in the evolution of phenotypic switching mechanisms (Lachmann & Jablonka, 1996; Kussell & Leibler, 2005; Wolf *et al*., 2005; Lancaster & Masel, 2009), supporting empirical evidence is currently lacking.

1.2 The power of bacterial model systems in experimental evolution

Evolution is a (relatively) slow process that requires many generations. As such, early insights into evolutionary processes were largely restricted to those gained from a historical perspective. These limitations were overcome with the realization that microbial systems could be used to observe evolution in real time. Bacteria are comparatively simple organisms, most of which have short mean generation times (~20 minutes) that enable large population sizes to be rapidly attained. Bacteria usually reproduce asexually, providing the opportunity to replicate experiments with isogenic genotypes. Furthermore, bacterial cultures can be frozen at -80˚C almost indefinitely, and subsequently revived from a state of suspended animation. This means that derived strains can be compared to ancestral strains in terms of genotype, phenotype and relative fitness. The tractability of bacterial model systems has made them an attractive option for testing evolutionary theories derived from macroevolutionary observations.

1.2.1 The *Pseudomonas fluorescens* SBW25 experimental system

1.2.1.1 *Pseudomonas fluorescens* SBW25

First isolated in 1989 from the leaf of a sugar beet plant at University Farm, Oxford (Rainey & Bailey, 1996), *Pseudomonas fluorescens* SBW25 is a Gram-negative, rod shaped, aerobic bacterial strain that colonizes the plant rhizosphere (root surfaces) and phyllosphere (leaf surfaces). Similar strains are found in a range of freshwater, marine and terrestrial environments (Rainey, 1999). A physical map (Rainey & Bailey, 1996), full genome sequence and annotation (Silby *et al*., 2009) are available. As described in the sections below, the work of many researchers has transformed *P. fluorescens* SBW25 into an experimental model system used to test evolutionary principles.

1.2.1.2 Diversification in a static microcosm

When cultured on agar plates containing King's medium B (KB), *P. fluorescens* SBW25 produces 'smooth' colonies – symmetrical colonies with a round edge (Figure 1.7). When this smooth type is propagated in a spatially structured environment for 72 hours (~100 generations), it rapidly diversifies into a range of niche-specialist genotypes. Fortuitously, these differences in niche preference are visible at the level of colony morphology, providing a tractable genotypic assay. Using this assay, divergent genotypes are classified into three types: smooth (SM), wrinkly spreader (WS) and fuzzy spreader (FS) morphs, which respectively colonize the broth, air-liquid interface and bottom of the microcosm (Figure 1.7; Rainey & Travisano, 1998). Although little is known about the FS morph, the more abundant WS morph has been studied extensively, providing considerable insight into both the phenotype and the underlying genotype.

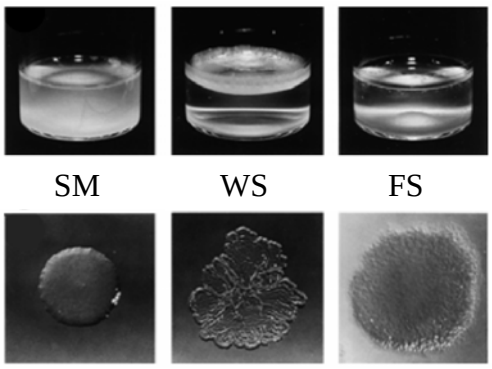


Figure 1.7: Variation in niche specificity and colony morphology in *P. fluorescens* SBW25 morphotypes. Three distinct morphs are recognisable after 72 hours incubation: smooth (SM), wrinkly spreader (WS) and fuzzy spreader (FS). Each morph displays individual niche specificities when cultured in KB microcosms (top) and grows distinctive colonies on KB agar (bottom). Adapted from Rainey & Travisano (1998).

1.2.1.3 The wrinkly spreader phenotype

As illustrated in Figure 1.7, the WS morph is characterized by colonies with a distinct wrinkly-edged morphology on KB agar and colonization of the air-liquid interface of static microcosms. Mutations causing the WS phenotype were identified by a suppressor analysis in which transposon mutants of a single wrinkly spreader phenotype were screened for types defective in both the wrinkled colony morphology and the ability to colonise the air-liquid interface (Kahn, 1998). These types were shown to contain insertions in two principal loci: the *wss* and *wsp* operons (Kahn, 1998; Bantinaki, 2001; Spiers *et al*., 2002). Additional suppressor analyses of WS genotypes - derived from strains in which certain previously recognised loci were removed - identified a further two principal loci: the *aws* and *mws* operons (Gehrig, 2005; McDonald, 2009; McDonald *et al*., 2009). The molecular role of each of these four loci in the WS phenotype is described in the following sections.

1.2.1.4 The structural basis of the wrinkly spreader phenotype

The *wss* (wrinkly spreader structural) operon is a ~15 kilobase (kb) locus containing ten genes (*wssA-J*; Figure 1.8). Together, these genes encode the machinery necessary to synthesize an acetylated cellulosic polymer (ACP). The function of each Wss protein has been predicted on the basis of genetic studies and comparative sequence analyses (Table 1.1; Kahn, 1998; Bohannon, 2002; Spiers *et al*., 2002; Spiers *et al*., 2003). WssB, C and E show homology to three cellulose biosynthetic (Bcs) proteins of *Gluconacetobacter xylinum* (formerly *Acetobacter xylinum*), and are therefore predicted to form the core cellulose biosynthetic component in conjunction with WssD, which shows similarity to a *G. xylinum* cellulase. WssFGHI are predicted to be involved in polymer acetylation (Spiers *et al*., 2003). WssA and WssJ share homology with the *E. coli* YhjQ cell division protein, and also with the cell polarity-determining MinD protein of various species. Consequently, WssA and WssJ are thought to play a role in the localisation of the cellulose synthase complex.

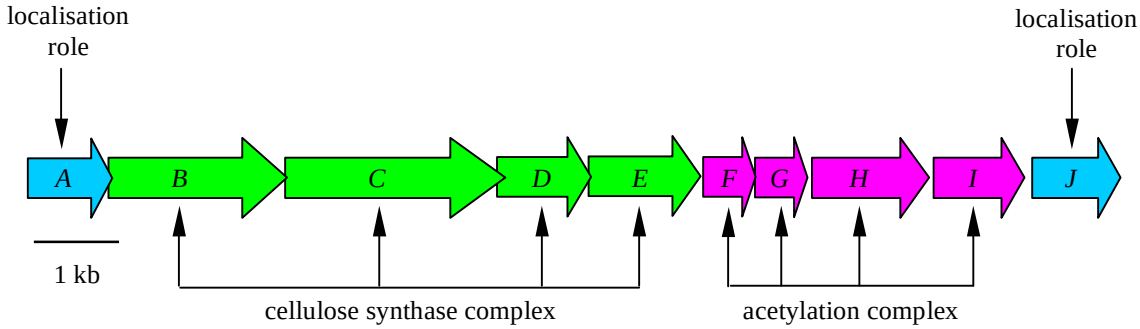


Figure 1.8: Functionally annotated depiction of the *wss* operon. The *wss* operon consists of ten genes: *wssA-J*. These genes encode proteins that act in concert to synthesize ACP. *WssBCDE* (green) encode cellulose synthase subunits, while *wssFGHI* (purple) encode proteins involved in acetylation of cellulose. The products of *wssA* and *wssJ* (blue) are thought to play a role in localisation of the Wss complex.

Gene	Pflu[a]	Size[b]	Predicted function	Pfam domains[c]	*E*-value[c]	Homology[d]
wssA	0300	1035	Sub-cellular localisation	CbiA	4.6×10^{-14}	YhjQ*, MinD‡
wssB	0301	2220	Cellulose biosynthesis	Glycos_transf_2	8.4×10^{-30}	BcsA#, YhjO*
				PilZ	4.8×10^{-22}	
wssC	0302	2268	Cellulose biosynthesis	BcsB	1.8×10^{-271}	BcsB#, YhjN*
wssD	0303	1197	Cellulose biosynthesis	Glyco_hydro_8	6.7×10^{-38}	CMCase*
wssE	0304	3840	Cellulose biosynthesis	BCSC_C	1.0×10^{-171}	BscC#, YhjL*
wssF	0305	666	Acetylation	None	n/a	
wssG	0306	666	Acetylation	None	n/a	AlgFø
wssH	0307	1416	Acetylation	MBOAT	1.6×10^{-53}	AlgIø
wssI	0308	1125	Acetylation	None	n/a	AlgJø
wssJ	0309	1050	Sub-cellular localisation	None	n/a	YhjQ*, MinD‡

Table 1.1: Predicted function, domain characteristics and homology of *wss* genes. [a]Pflu refers to the numeric name of the SBW25 gene. [b]Number of nucleotides. [c]Pfam domain searches (including associated *E*-values) were performed through the Sanger Pfam website. [d]Homology to known proteins: *, ‡, # and ø represent proteins from *E. coli*, *H. pylori*, *G. xylinum* and *P. aeruginosa*, respectively.

Biochemical analyses of WS mats have confirmed that the major structural component is ACP (Spiers *et al.*, 2002). Additional biochemical studies have shown that ACP biosynthesis is the primary phenotypic innovation that enables colonization of the air-

liquid interface (Spiers *et al*., 2003). The proposed molecular model of mat formation is that ACP production and secretion causes individual WS cells to adhere to one another, facilitating colonisation of the air-liquid interface (Spiers *et al*., 2002; Gal *et al*., 2003; Rainey & Rainey, 2003; Spiers *et al*., 2003). It is thought that the ability to colonize the air-liquid interface is adaptive because it provides access to oxygen, a scarce commodity in the static microcosm (Rainey & Rainey, 2003).

1.2.1.5 Genetic routes to the wrinkly spreader phenotype

As the *wss* locus provides the structural basis of WS, deleterious mutations in this region are expected to abolish the WS phenotype. Conversely, mutations and/or environmental stimuli that increase Wss function are expected to generate the WS phenotype. In 1987, structural studies demonstrated that the cellulose biosynthetic enzymes of *G. xylinum* were allosterically activated by the intracellular molecule bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP; Ross *et al*., 1987). Since then, c-di-GMP has been recognised as a widely utilized bacterial second messenger that functions to activate cellulose biosynthetic enzymes (among other enzymes) in a range of bacterial species, including the *P. fluorescens* SBW25 Wss enzymes (Goymer *et al*., 2006; Malone *et al*., 2007; reviewed in Römling *et al*., 2005; Hengge, 2009). C-di-GMP is synthesized from two molecules of GTP by diguanylate cyclases (DGCs), and is broken down to pGpG by phosphodiesterases (PDEs). DGC activity is associated with the GGDEF domain (Paul *et al*., 2004), while PDE activity is associated with the EAL domain (Christen *et al*., 2005; Schmidt *et al*., 2005). The GGDEF and EAL domains are named for the conserved amino acid motifs that are essential components of the enzymatic action sites.

DGC activity is often controlled by two-component signal transduction (TCST), a common method of signal transduction in bacteria (Nixon *et al*., 1986; reviewed in Stock *et al*., 2000). Illustrated in Figure 1.9, the prototypical TCST system consists of a histidine kinase (HK) and a response regulator (RR). An environmental signal is sensed by the HK and passed by a phosphotransfer event to the response regulator (here, the DGC), which generates an appropriate phenotypic response (increased c-di-GMP

levels). Each of the three remaining loci identified in the WS transposon suppressor analyses (*wsp*, *aws*, and *mws*) encodes a DGC. A mutation resulting in stimulation of c-di-GMP production by any of these three DGC is sufficient to cause the production and secretion of ACP (Bantinaki *et al.*, 2007; McDonald *et al.*, 2009).

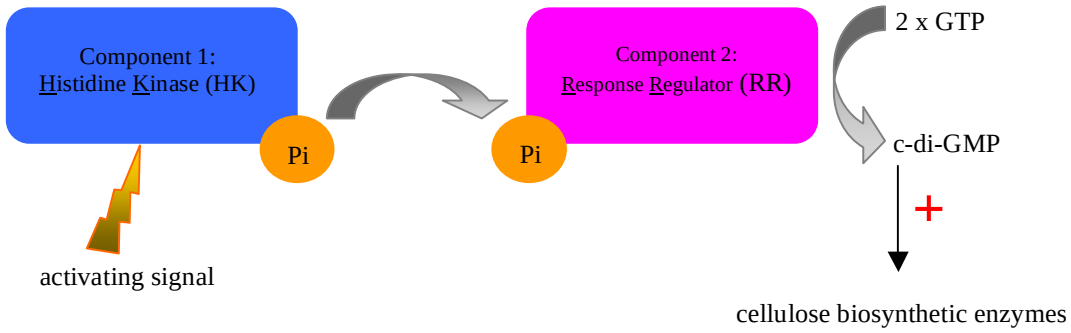


Figure 1.9: Generic activation of cellulose biosynthetic enzymes by traditional two-component signal transduction. An environmental or cellular signal activates the first component of the signal transduction pathway, a histidine kinase (HK), by causing phosphorylation of a conserved histidine residue. A phosphotransfer reaction results in the inorganic phosphate (Pi) being passed to the second component of the cascade, the response regulator (RR). When thus activated (+), the output domain of the RR converts GTP into the bacterial second messenger, c-di-GMP, which in turn activates cellulose biosynthetic (*wss*) enzymes.

1.2.1.5.1 The wsp *locus*

The *wsp* (wrinkly spreader phenotype) locus was first identified in the original WS suppressor analysis (Kahn, 1998) and has since been functionally characterised (Bantinaki, 2001; Bantinaki *et al.*, 2007). Highly conserved in all pseudomonads, the *wsp* operon covers ~8.5 kb, and contains seven genes (*wspA-F* and *wspR*). Together the genes encode a chemosensory pathway with significant homology to the *E. coli* Che pathway (see Table 1.2), a well-studied pathway that allows bacteria to swim directionally with respect to chemical gradients in their environment (Berg, 1975; reviewed in Bren & Eisenbach, 2000).

Protein	Pflu[a]	Size[b]	Predicted function	Pfam domains[c]	*E*-value[c]	Che[d]
WspA	1219	540	Methyl-accepting chemotaxis protein (MCP)	HAMP	1.5×10^{-16}	MCP
				MCPsignal	1.15×10^{-26}	
WspB	1220	170	Scaffold protein	CheW	9.95×10^{-27}	CheW
WspC	1221	419	Methyltransferase	CheR_N	1.15×10^{-11}	CheR
				CheR	2.35×10^{-32}	
WspD	1222	232	Scaffold protein	CheW	6.35×10^{-26}	CheW
WspE	1223	755	Histidine kinase	Hpt	2.55×10^{-12}	CheA/CheY
				HATPase_C	1.35×10^{-22}	
				CheW	3.55×10^{-24}	
				Response_reg	1.15×10^{-30}	
WspF	1224	336	Methylesterase (response regulator)	Response_reg	6.55×10^{-20}	CheB
				CheB_methylest	3.75×10^{-57}	
WspR	1225	333	Diguanylate cyclase (response regulator)	Response_reg	5.85×10^{-19}	CheY
				GGDEF	8.95×10^{-66}	

Table 1.2: Predicted function, domain characteristics and Che homology of Wsp proteins. [a]Pflu refers to the numeric name of the SBW25 gene. [b]Number of amino acids. [c]Pfam domain searches were performed using the Sanger Pfam website. [d]Homology to *E. coli* Che chemosensory pathway.

Based on the operation of the *E. coli* Che pathway, a mechanistic model has been proposed for the Wsp pathway (Figure 1.10; Bantinaki *et al.*, 2007). In this model, WspA, the methyl-accepting chemotaxis protein (MCP), forms a membrane-bound complex with two scaffold proteins (WspB/D) and the histidine kinase, WspE. In the activated (methylated) form, WspA transduces a signal to WspE, which is activated by phosphorylation. In turn, WspE activates two proteins by phosphorylation: WspR (at Asp67; Goymer *et al.*, 2006) and WspF. WspR is a RR containing a GGDEF domain (Table 1.2). When phosphorylated, WspR synthesizes c-di-GMP (Malone, 2005; Malone *et al.*, 2007). Further insight into the regulation of WspR activity is provided by crystal structures of WspR from both *P. aeruginosa* and *Pseudomonas syringae* (De *et al.*, 2008; De *et al.*, 2009). The c-di-GMP produced by WspR activates the cellulose biosynthetic enzymes (Wss), and ACP is synthesized. WspF is a methylesterase; WspE-

mediated phosphorylation of WspF stimulates deactivation of WspA *via* the removal of a methanol group (CH_3OH). WspC, a methyltransferase that constitutively activates WspA, antagonises the methylesterase activity of WspF. Thus, the model predicts that the Wsp pathway continually oscillates between the active and inactive states.

Since the basis of the WS phenotype is overproduction of ACP, any mutation resulting in stimulation of the Wsp pathway could conceivably generate the WS phenotype. The most obvious target for mutations is the negative regulator of the pathway, WspF. According to the above model, any loss-of-function mutation in WspF will result in constitutive activation of the Wsp pathway. Mutations in *wspC*, *wspD* and *wspE* have also been observed. Interestingly, no naturally occurring WS-causing mutations have (yet) been observed in WspR (P.B. Rainey, personal communication).

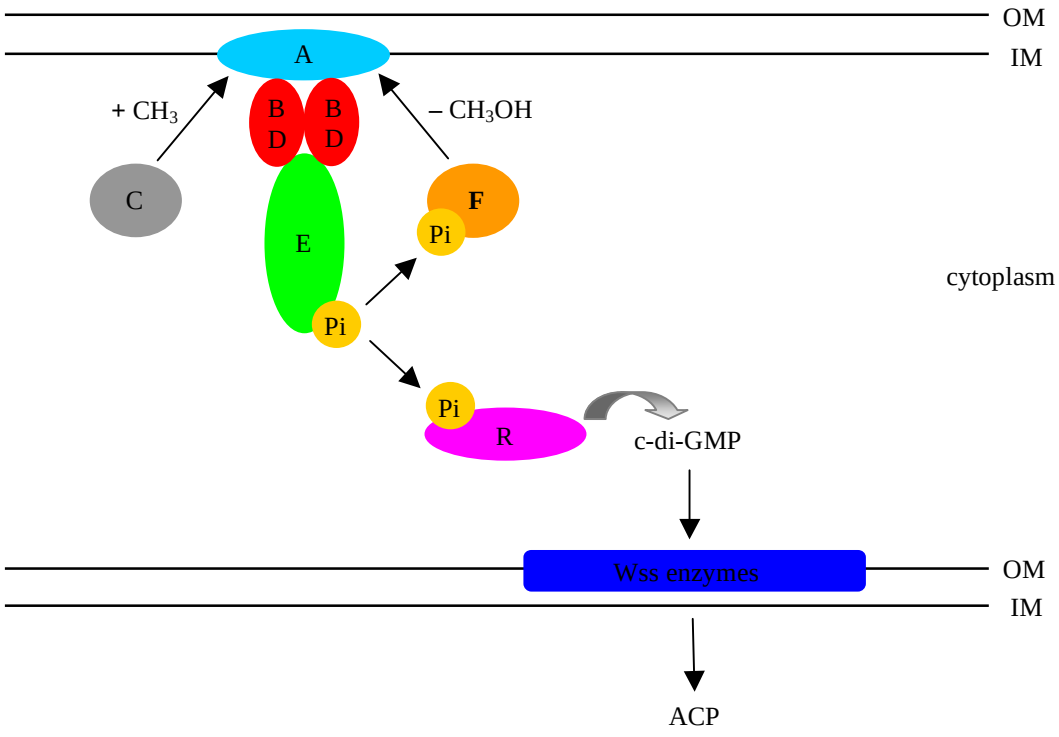


Figure 1.10: Proposed mechanistic model of the Wsp chemosensory cascade. Activation of WspA causes WspE to become phosphorylated. In turn, activated WspE phosphorylates WspR and WspF. Phosphorylated WspR produces c-di-GMP, which activates the production of ACP by the Wss enzymes. Phosphorylated WspF acts to negatively regulate the cascade, removing methanol (CH_3OH) from WspA and thus ending the signal. WspF activity is counteracted by the constitutive action of WspC, which activates WspA by addition of a methyl group (CH_3). Figure adapted from Bantinaki *et al.* (2007). (OM=outer membrane, IM=inner membrane, Pi=inorganic phosphate).

1.2.1.5.2 The aws *locus*

The *aws* (alternate wrinkly spreader) locus is a ~2.3 kb operon containing three genes (*awsX*, *awsR*, and *awsO*). It was implicated in the WS phenotype as a result of a transposon suppressor analysis of a WS genotype derived from an engineered strain (SBW25 with the *wsp* operon deleted; McDonald, 2009; McDonald *et al*., 2009). Recognised domains in each of the three encoded proteins are listed in Table 1.3. The first gene encodes AwsX, a protein conserved among pseudomonads that contains no recognised protein domains. The sole identified motif is a proteolytic cleavage site, indicating that AwsX is a periplasmic protein. Mutational studies have shown that AwsX negatively regulates the product of *awsR*, a DGC. In addition to the recognised HAMP (histidine kinase) and GGDEF domains, AwsR contains two transmembrane helices and an 'I-site' (an R**D motif), indicating that c-di-GMP regulates DGC activity by binding allosterically to AwsR (Christen *et al*., 2006). The final gene, *awsO*, encodes a protein with homology to the outer membrane porin OmpA, and is consequently predicted to form a pore in the outer membrane (Gehrig, 2005).

Protein	Pflu[a]	Size[b]	Predicted function	Pfam domains[c]	*E*-value[c]
AwsX	5211	190	Negative regulator of AwsR	None	n/a
AwsR	5210	420	Diguanylate cyclase (response regulator)	HAMP (histidine kinase)	1.2×10^{-8}
				GGDEF	1.9×10^{-55}
AwsO	5209	163	Outer membrane porin	OmpA	4.2×10^{-26}

Table 1.3: Predicted function and domain characteristics of Aws proteins. [a]Pflu refers to the numeric name of the gene encoding the protein in the SBW25 genome. [b]Number of amino acids. [c]Pfam domain searches (including associated *E*-values) were performed through the Sanger Pfam website.

On the basis of domain characteristics, evolutionary observations and mutational studies (Gehrig, 2005; McDonald *et al*., 2009), the following model has been proposed for Aws function: the outer-membrane porin AwsO binds reversibly to the periplasmic protein, AwsX. AwsX forms a complex with AwsR, a DGC RR associated with the inner membrane. In the inactive state, AwsX is bound to AwsO, preventing physical associations between the two AwsR units (Figure 1.11A). Activation of Aws occurs

when a (currently unidentified) signal results in the dissociation of AwsO from AwsX. The conformational change allows dimerization and consequent activation of AwsR units, and the production of c-di-GMP by the GGDEF domain ensues (Figure 1.11B). C-di-GMP activates the Wss enzymes, and ACP is synthesized. WS-causing mutations have been identified in all three *aws* genes. Presumably, each of these mutations ultimately increases the output of c-di-GMP from AwsR.

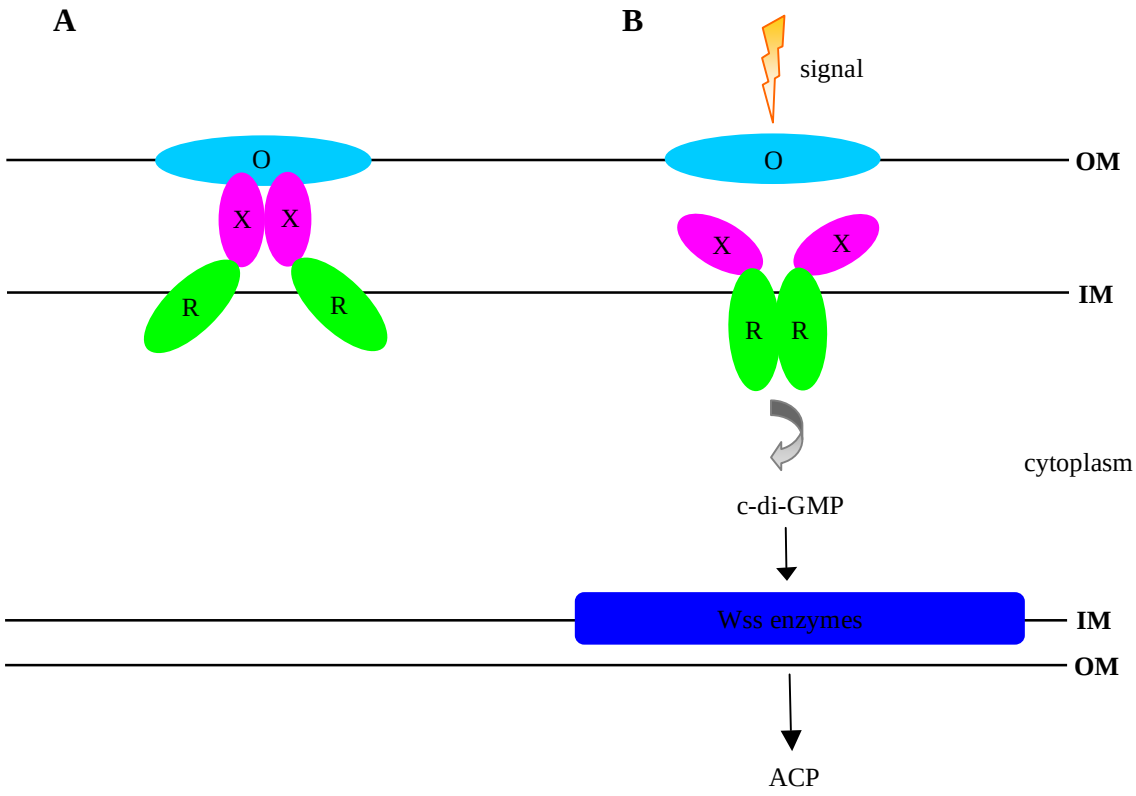


Figure 1.11: Proposed mechanistic model of the Aws pathway. The *aws* operon encodes three proteins – the outer membrane porin AwsO, periplasmic AwsX and the diguanylate cyclase, AwsR. In the inactivated state (**A**), AwsX is bound to AwsO, preventing the dimerization and activation of AwsR. In the activated state (**B**), an environmental signal results in the disbanding of AwsX from AwsO, allowing the dimerization of AwsR. Dimerized AwsR promotes the synthesis of c-di-GMP, which in turn activates the Wss cellulose biosynthetic enzymes, causing the production of ACP (McDonald *et al.*, 2009).

1.2.1.5.3 The mws *locus*

The *mws* (mike's wrinkly spreader) locus was implicated in the WS phenotype during a transposon suppressor analysis of a WS genotype derived from a second engineered strain (SBW25 with both the *wsp* and *aws* operons deleted; McDonald, 2009; McDonald *et al.*, 2009). A ~3.8 kb locus, the *mws* locus consists of a single gene, *mwsR*

(Pflu5329). MwsR is a large protein (1,283 amino acids) with multiple domains. These include three individual PAS (signal transduction) domains, one GGDEF domain and one EAL domain (Table 1.4; see section 1.2.1.5). Additionally, the presence of two transmembrane helices suggests that the protein is localised to the inner membrane.

Domain[a]	*E*-value[a]	Size[b]	Predicted function
PAS_3	1.3×10^{-11}	94	Signal sensory domain
PAS	5.8×10^{-5}	123	Signal sensory domain
PAS	1.3×10^{-10}	123	Signal sensory domain
GGDEF	7.3×10^{-61}	164	Synthesis of c-di-GMP
EAL	3.5×10^{-90}	248	Breakdown of c-di-GMP

Table 1.4: Predicted domain characteristics and associated functions for MwsR (encoded by *mwsR*, Pflu5329). [a]Pfam domain searches were performed using the Sanger Pfam website. [b]Number of amino acid residues.

The co-existence of a GGDEF and EAL domain suggests a role for MwsR in both the synthesis and breakdown of c-di-GMP, and thus in both activating and de-activating the Wss enzymes. Such dual activity has not been previously observed among c-di-GMP regulatory proteins, with all GGDEF/EAL dual proteins observed previously having only PDE activity (Jenal & Malone, 2006). The PDE activity of MwsR was demonstrated by generation of WS upon deletion of the EAL domain (McDonald *et al*., 2009). A role for Mws in the resulting WS phenotype was demonstrated by subsequent deletion of the remainder of *mws*, resulting in the SM phenotype (McDonald *et al*., 2009). Collectively, these experiments show that MwsR possesses both DGC and PDE activity. Thus, the predicted model for MwsR function is that the GGDEF domain increases c-di-GMP levels, while the EAL domain acts to lower c-di-GMP levels (Figure 1.12). Given that mutational causes of WS in *mwsR* increase the net output of c-di-GMP, one might expect to find loss-of-function mutations in the EAL domain or gain-of-function mutations in the GGDEF domain. Indeed, both classes of mutation have been observed (McDonald *et al*., 2009).

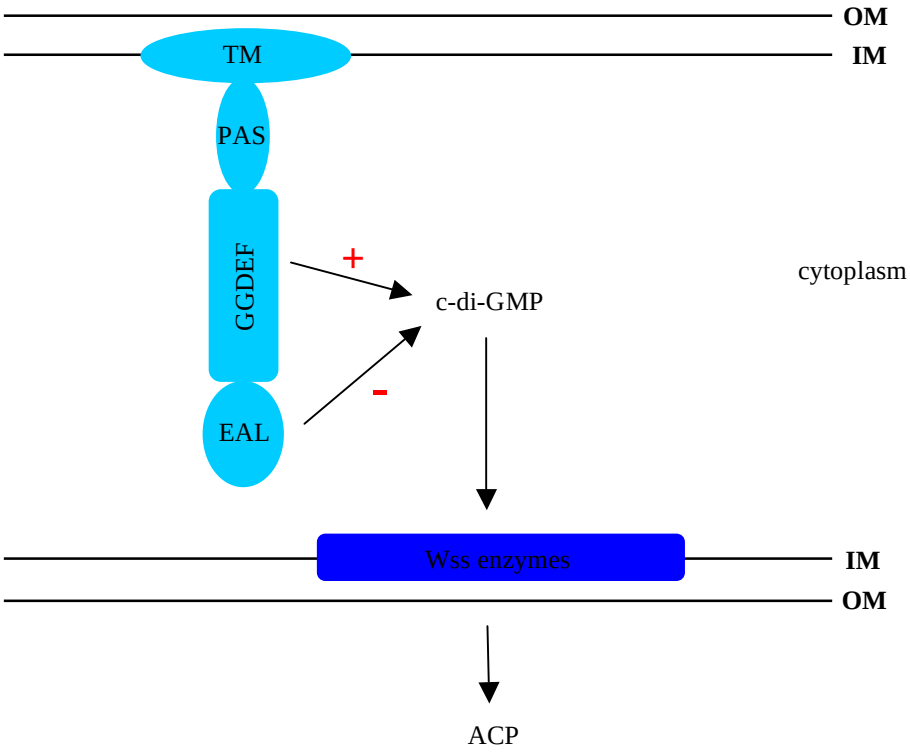


Figure 1.12: Proposed model of MwsR function. MwsR is a large, multi-domain protein. Signals sensed and transduced through the PAS domains control the antagonistic actions of the GGDEF and EAL domains, which synthesize and breakdown c-di-GMP, respectively (McDonald *et al.*, 2009). (OM=outer membrane, IM=inner membrane, TM=transmembrane region)

1.3 Mechanisms of evolution: insights from bacterial model systems

1.3.1 Influences on natural selection

When Darwin and Wallace first proposed the theory of evolution by natural selection (Darwin & Wallace, 1858), it was met with considerable resistance. Since then, the theory has withstood rigorous scientific testing and today natural selection is scientifically accepted as the driving force of evolution; natural selection will operate whenever heritable variation results in differential reproductive success (see section 1.1). Mechanistically, natural selection is entirely dependent on the presence of phenotypic variation within a population, which in turn is fuelled by the generation of random variation (i.e. mutations). Broadly speaking, mutations are the result of unpredictable, random errors during DNA replication and repair (reviewed in Stearns & Hoekstra, 2005: p.102-103). This unpredictability introduces an intrinsic element of

stochasticity into evolution: the chance of any particular mutation occurring within a population. Further elements of chance are introduced by genetic drift – random alteration in genotype frequencies of a population as a result of chance factors (e.g. elimination of a beneficial mutation in an environmental disaster), migration and sex. Thus, the focus of the current debate is the extent to which chance events influence evolutionary outcomes.

1.3.2 The role of evolutionary history: 'replaying life's tape'

Closely intertwined with the role of chance in evolution, historical contingency refers to the concept that evolutionary outcomes are dependent on prior evolutionary events. If at each evolutionary stage there are multiple adaptive outcomes, each outcome may limit subsequent ones, and the mutational path would be of central importance when interpreting the end result (Figure 1.13A). Alternatively, it is possible that one adaptive outcome is so beneficial that all genetic routes will eventually converge on the same phenotypic solution (Figure 1.13B). Note that in both of these views, there is no doubt that natural selection will generate an adaptive form. The uncertainty lies in *which* adapted form. Stephen Gould has expressed this evolutionary dilemma in his book *Wonderful Life*, in which he asks what the world would be like if evolutionary history was repeated, and the 'tape of life' was replayed:

> *"I call this experiment 'replaying life's tape'. You press the rewind button and, making sure you thoroughly erase everything that actually happened, go back to any time and place in the past... Then let the tape run again and see if the repetition looks at all like the original."* (Gould, 1989: p.48).

If each replay of the tape resulted in the evolution of similar life forms, this would support an adaptationist view (Figure 1.13B). On the other hand, the evolution of different life forms on each replay would lend support to the dependence of evolutionary outcomes on historical chance events (Figure 1.13A). Gould himself was a proponent of historical contingency, writing of his gedankenexperiment *"…any replay*

of the tape would lead evolution down a pathway radically different from the road actually taken." (Gould, 1989: p.51). Gould draws support for historical contingency from observations of situations where evolution has engineered anatomically imperfect solutions. His most famous example is that of the panda's bamboo-stripping 'thumb', a dextrous sixth digit evolved by dint of extending a wrist bone (Gould, 1987: p.20-22). A number of scientists take the opposite view to Gould, proposing that genetic architecture of the cell is such that many mutational routes lead to the same evolutionary outcomes (Dawkins, 1986; Schluter, 1996; Conway Morris, 2003; Vermeij, 2006). Support for this view comes from observations of convergent evolution – the phenomenon whereby disparate organisms independently generate similar solutions to the same evolutionary challenges. A famous example of convergent evolution is the eye, similar versions of which have evolved independently in different phyla (Lamb *et al.*, 2007; reviewed in Dawkins, 1986: p.94-95).

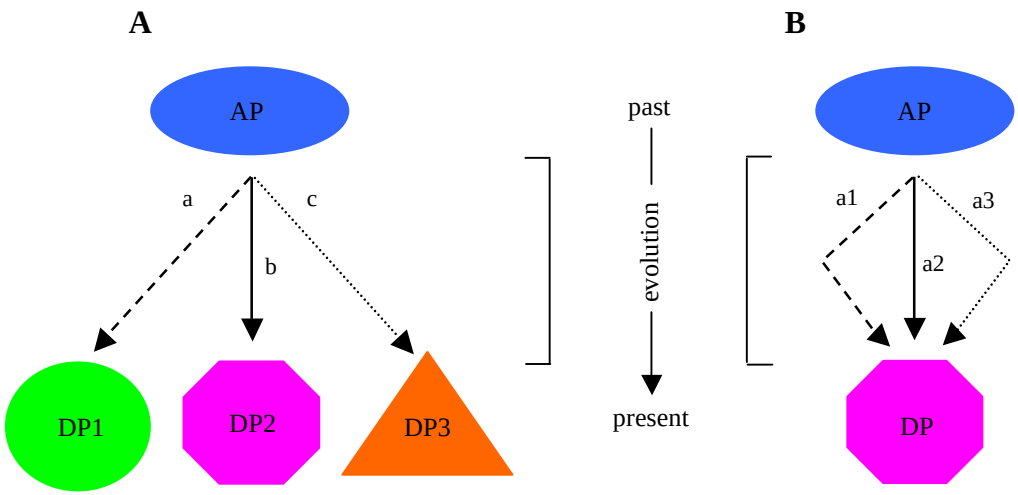


Figure 1.13: Mechanisms of evolution. (**A**) In the historical view, evolution by natural selection is dependent upon historical evolutionary events, and consequently different genetic routes (a, b and c) lead from a common ancestral phenotype (AP) to divergent adaptive descendent phenotypes (DP1, DP2, and DP3). (**B**) In the adaptationist view, evolution by natural selection repeatedly generates the same adaptive descendent phenotype (DP) *via* different genetic routes (a1, a2 and a3). Notably, the adaptationist view requires evolutionary divergence (represented by the first segment of each genetic route) followed phenotypic convergence (represented by the latter segment of each genetic route).

While Gould's 'tape of life' obviously cannot be replayed on a planetary scale, similar experiments have been conducted on a smaller scale. Although some observational experiments have been undertaken in higher organisms (Losos *et al.*, 1998; reviewed in

van der Woude, 2006), the majority of evolutionary replay experiments examine evolutionary trajectories in parallel experimental populations of microbes (Lenski *et al.*, 1991; Finkel & Kolter, 1999). The longest standing of these experiments was begun in 1988 when Professor Richard Lenski founded 12 parallel *E. coli* populations from isogenic ancestors, and allowed them to evolve indefinitely by serial dilution and propagation in glucose-limited minimal medium. Twenty-one years and ~48,000 generations later, the experiment continues, and much insight has been gained by genetic, metabolic and biochemical analysis of the resulting populations (Lenski *et al.*, 1991; Lenski & Travisano, 1994; Travisano *et al.*, 1995; Woods *et al.*, 2006). Perhaps unsurprisingly, the populations have given no clear answer to dilemma of historical contingency. In some respects the populations differ, supporting Gould's historical view. For example, after ~32,000 generations, one line developed the ability to utilize citrate as a carbon source, something which no other line has (yet) shown (Blount *et al.*, 2008). Conversely, a number of similarities across the lines support the adaptationist view. For instance, at 10,000 generations, three of the 12 lines had evolved mutator strains with defects in DNA repair (Sniegowski *et al.*, 1997). In order to reach more definite conclusions about the evolutionary role of chance and history, additional cases of parallel evolution must be examined.

1.3.3 Replaying the *P. fluorescens* tape: the reverse-evolution experiment

Designed by Dr. Hubertus Beaumont and Professor Paul Rainey, the reverse evolution experiment (REE) focused on comparative observation of the evolutionary trajectories of experimentally parallel *P. fluorescens* populations (Beaumont *et al.*, 2009). Essentially, the REE involved recording the evolutionary responses of 12 replicate *P. fluorescens* SBW25 populations to 16 alternating rounds of adaptive evolution in static and shaking microcosms. At every round of evolution, a severe bottleneck was imposed on the populations by allowing the survival of only a single genotype from each replicate. This genotype was selected on the basis of phenotypic innovation; it had to form colonies with a different morphology to the genotype used to found that particular round (Figure 1.14A; for a full account of the REE method see section 2.2.9). A result of the alternation between opposing environmental conditions, each consecutive

phenotypic innovation was expected to have a genetic basis. Since founding genotypes were isolated from (and thus adapted to) one environment, they were likely to be maladaptive in the subsequent opposing environment. Thus, alternation between opposing environmental conditions drove the repeated evolution of novel adaptive phenotypes. With the exception of two lines that ceased to respond to selection, the REE culminated in the storage of 16 chronological genotypes for each line. This gave 12 evolutionary series stored in suspended animation, where each genotype differed from its immediate ancestor by a single mutation (Figure 1.14B).

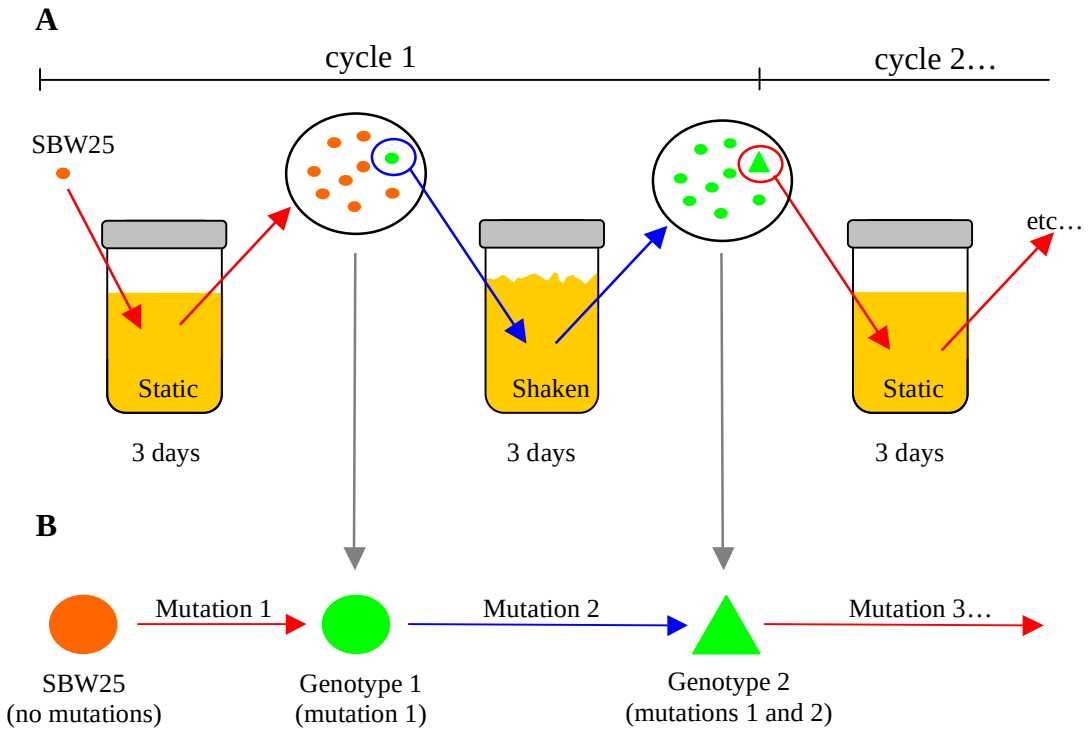


Figure 1.14: The reverse evolution experiment (REE) described in Beaumont *et al.* (2009). (**A**) The REE protocol. *P. fluorescens* SBW25 populations were subjected to repeated bouts of evolution in a cyclic environment. At each change in the environment, a single, phenotypically novel, numerically dominant genotype was selected from which to begin evolution in the subsequent environment. (**B**) The experiment resulted in the storage of an evolutionary series of genotypes, where each genotype (theoretically) contains one additional mutation responsible for its change in phenotype. Here, a single hypothetical evolutionary line is shown; the experiment consisted of 12 parallel lines.

A key difference between the REE and the aforementioned Lenski experiment (see section 1.3.2) lies in the severity of the (artificially applied) selection pressure. The reverse selection regime of the REE continually imposes harsh selection, firstly from

the alternation of environments, and secondly from the extreme, single-genotype bottleneck at the end of every round. A central feature of the REE regime, application of the bottleneck ensured that each of the twelve lines is composed of fixed evolutionary events that could be used as a basis for comparison between lines. The complete complement of mutations in the final genotype of each line may be determined using whole genome re-sequencing techniques, and, since all the previous genotypes are stored in a state of suspended animation, the chronological order of mutations in each line can be determined. Combined with the considerable knowledge of the *P. fluorescens* SBW25 model system (see section 1.2.1), this information will allow in-depth evolutionary analysis at both the genetic and molecular levels. This is an excellent system with which to evaluate the extent to which evolution is contingent upon history, and to untangle the influences of natural selection, chance and history in evolution.

1.3.4 Emergence of the switcher phenotype

Each of the 12 lines in the REE has a unique evolutionary trajectory; every line has a distinct set of phenotypic changes. In two lines (lines one and six), the ninth reversal saw the evolution of an entirely novel type that switched its phenotype rapidly between two different colony states; on an agar plate these genotypes ($1w^4$ and $6w^4$, respectively) each gave rise to two phenotypically distinct colony types (Figure 1.15). These dichotomous genotypes have been dubbed ‘switchers’. The evolution of switchers under the REE regime provides a novel opportunity to gain insight into the molecular basis, adaptive significance and evolutionary origins of a phenotypic switching mechanism (see section 1.1.3). The design of the REE is such that the entire sequence of genetic events leading to the evolution of novel phenotypes can be determined and analysed on both a genotypic and phenotypic level. The primary topic of this thesis is the investigation of the molecular and evolutionary origins of these switcher types.

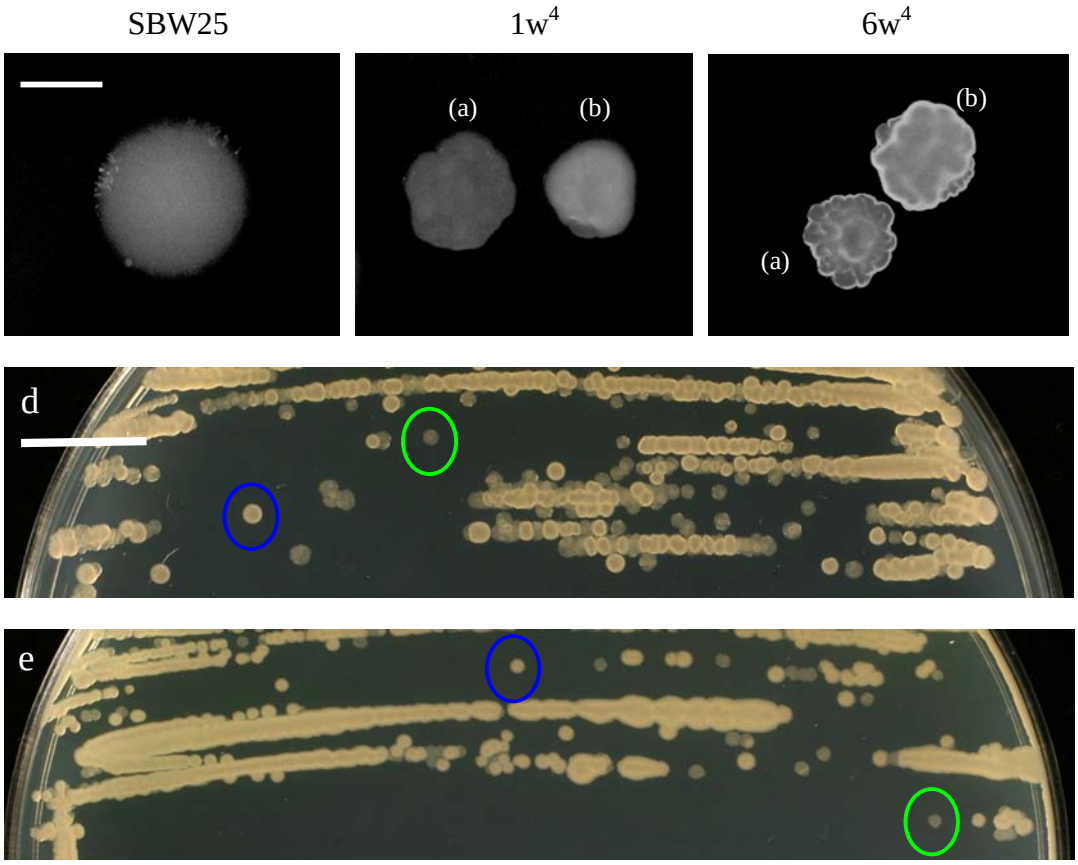


Figure 1.15: The two colony states of the switcher types. Ancestral *P. fluorescens* SBW25 gives rise to smooth colonies on KB agar (left), while the derived strains $1w^4$ and $6w^4$ give rise to two phenotypically distinct types of colonies on King's medium B (KB) agar (middle and right, respectively). Photographs taken under x12.5 dissection microscopy and the contrast/brightness adjusted in iPhoto. Scale bar represents ~3 mm and applies to all three genotypes. Translucent (**d**) and opaque (**e**) $1w^4$ colonies sub-streaked on KB agar (48 hours). Examples of distinct translucent (green) and opaque (blue) are circled for clarity. Scale bar for (d) and (e) represents 10 mm.

1.4 Summary and aims of the current study

The ability of organisms to rapidly generate phenotypic diversity in a population has long been recognised, and a number of underlying mechanisms have been extensively characterised. Due to the prevalence of phenotypic switching mechanisms among pathogenic bacteria, their biological significance has been attributed to evasion of the immune response. More recent observations of phenotypic switching mechanisms in among non-pathogenic species have challenged this assumption, and it is realised that little is definitively understood about the evolutionary origins or adaptive significance

of switching mechanisms. Recently, the evolution of a phenotypic switching mechanism was observed in experimental populations of *P. fluorescens* SBW25. Combined with the extensive knowledge of the *P. fluorescens* model system, the evolution of the switcher under the REE regime presents an unprecedented opportunity to gain insight into both the evolutionary origins and adaptive significance of switching mechanisms. The subject of this thesis is a full characterisation of the phenotypic, evolutionary and molecular bases of the phenotypic switching mechanism. Specifically, this thesis contains:

1. An intensive phenotypic investigation of $1w^4$ (and its evolutionary lineage). This includes a demonstration of heritable, high frequency switching. Additionally, an investigation of the structural bases of bistability using biochemical and genetic techniques (Chapter 3) and a comprehensive transposon mutagenesis screen (Chapter 4) is described.

2. A full analysis of the evolutionary history of $1w^4$. This was achieved using whole genome re-sequencing and conventional sequencing techniques, genomic manipulations and fitness experiments (Chapter 5).

3. Insight into the molecular mechanism underlying bistability. This was achieved using a range of biochemical and genetic analyses (Chapter 6).

4. A study of $6w^4$, a switcher type that evolved in parallel with $1w^4$. This includes a full phenotypic, genotypic, evolutionary and mechanistic characterisation, enabling comparisons to be made between independently evolved switcher types (Chapter 7).

Chapter 2:
Materials and Methods

2.1 Materials

2.1.1 Bacterial strains

Bacterial strains used are listed below in Table 2.1 All strains were stored indefinitely at -80°C in 45 % ($^{v}/_{v}$) glycerol saline.

Strain	Genotype and characteristics	Reference
Pseudomonas fluorescens		
SBW25 (SM)	Ancestral (wild-type) strain isolated from the leaf of a sugar beet plant grown at the University Farm, Wytham, Oxford in 1989; smooth genotype	(Rainey & Bailey, 1996)
LSWS	Niche-specialist mutant evolved from SBW25; wrinkly genotype resulting from a mutation in *wspF* (A901C)	(Spiers *et al*., 2002)
SBW25-Δ*wss*Km	SBW25 with a Km^R gene (*aph3'*) in place of the *wss* operon	(Gehrig, 2005)
SBW25-*lacZ*	SBW25 containing a neutral, chromosomal *lacZ* marker; competitor genotype in fitness experiments	(Zhang & Rainey, 2007)
$1w^0$, $1s^1$, $1w^1$, $1s^2$, $1w^2$, $1s^3$, $1w^3$	Chronology of the first seven genotypes of the reverse evolution experiment (line 1); each consecutive strain contains one additional mutation	H.J.E Beaumont
$1s^4$	Eighth strain of the reverse evolution experiment (line 1); immediate ancestor of the line 1 switcher genotype	H.J.E Beaumont
$1w^4$	Ninth strain of the reverse evolution experiment (line 1); the line 1 switcher genotype	H.J.E Beaumont
$6w^0$, $6s^1$, $6w^1$, $6s^2$, $6w^2$, $6s^3$, $6w^3$	Chronology of the first seven genotypes of the reverse evolution experiment (line 6); each consecutive strain contains one additional mutation	H.J.E Beaumont
$6s^4$	Eighth strain of the reverse evolution experiment (line 6); immediate ancestor of the line 6 switcher genotype	H.J.E Beaumont
$6w^4$	Ninth strain of the reverse evolution experiment (line 6); the line 6 switcher genotype	H.J.E Beaumont

Strain	Genotype and characteristics	Reference
$1w^4$-reD2	Numerically dominant switcher genotype from $1s^4$	H.J.E Beaumont
$1w^4$-reD12	Numerically dominant switcher genotype from $1s^4$	H.J.E Beaumont
$1w^4$-reD1.8	Numerically dominant switcher genotype from $1s^4$	H.J.E Beaumont
$1w^4$-reN1.2	Non-numerically dominant switcher genotype from $1s^4$	H.J.E Beaumont
$1w^4$-reN1.4	Non-numerically dominant switcher genotype from $1s^4$	H.J.E Beaumont
$1w^4$-reN1.5	Non-numerically dominant switcher genotype from $1s^4$	H.J.E Beaumont
$R6s^1$.1-20	20 x strains evolved in parallel from $6w^0$	This study
JG1-183	183 x non-switching strains derived from $1w^4$ by transposon mutagenesis with IS-Ω-Km/hah	This study
JG1ΔCre-JG183ΔCre	Selected transposon mutants (derived from those above) that have undergone Cre recombinase-mediated excision of the transposon	This study
JG6.1-8	7 x strains with reduced capsule biosynthesis derived from $6w^4$ by transposon mutagenesis with IS-Ω-Km/hah	This study
SBW25-*wcaJ-lacZ*	SBW25 strain with a *wcaJ-lacZ* transcriptional fusion	This study
$1s^4$-*wcaJ-lacZ*	$1s^4$ strain with a *lacZ* transcriptional fusion to *wcaJ*	This study
$1w^4$-*wcaJ-lacZ*	$1w^4$ strain with a *lacZ* transcriptional fusion to *wcaJ*	This study
$6s^4$-*wcaJ-lacZ*	$6s^4$ strain with a *lacZ* transcriptional fusion to *wcaJ*	This study
$6w^4$-*wcaJ-lacZ*	$6w^4$ strain with a *lacZ* transcriptional fusion to *wcaJ*	This study
$1s^4$-*carB*mut	$1s^4$ containing the $1w^4$ mutant *carB* allele	This study
$1w^4$-*carB*wt	$1w^4$ containing the wild-type *carB* allele	This study
SBW25-*carB*mut	SBW25 containing the $1w^4$ mutant *carB* allele	This study
$1w^0$-*carB*mut	$1w^0$ containing the $1w^4$ mutant *carB* allele	This study
$6s^4$-*rpoD*mut	$1s^4$ containing the $1w^4$ mutant *rpoD* allele	This study
SBW25-*rpoD*mut	SBW25 containing the $6w^4$ mutant *rpoD* allele	This study
$1w^4$-Δ*wss*	$1w^4$ with the *wss* operon deleted	This study
$1w^4$-Δ*carB*	$1w^4$ with *carB* deleted	This study
$1w^4$-Δ*recA*	$1w^4$ with *recA* deleted	This study
$1w^4$-Δ*arcB*	$1w^4$ with *arcB* (Pflu4892) deleted	This study
SBW25-pSX	SBW25 with pSX	This study
SBW25-pSX-*carB*	SBW25 with pSX containing *carB* for over-expression	This study
SBW25-pSX-*pyrH*	SBW25 with pSX containing *pyrH* for over-expression	This study
SBW25-pSX-*ndk*	SBW25 with pSX containing *ndk* for over-expression	This study
SBW25-pSX-*galU*	SBW25 with pSX containing *galU* for over-expression	This study
$1w^4$-pSX	$1w^4$ with pSX	This study
$1w^4$-pSX-*carB*	$1w^4$ with pSX containing *carB* for over-expression	This study
$1w^4$-pSX-*pyrH*	$1w^4$ with pSX containing *pyrH* for over-expression	This study
$1w^4$-pSX-*ndk*	$1w^4$ with pSX containing *ndk* for over-expression	This study
$1w^4$-pSX-*galU*	$1w^4$ with pSX containing *galU* for over-expression	This study

Strain	Genotype and characteristics	Reference
Re1.4-pSX	1w^4-reN1.4 with pSX	This study
Re1.4-pSX-*carB*	1w^4-reN1.4 with pSX containing *carB* for over-expression	This study
Re1.4-pSX-*pyrH*	1w^4-reN1.4 with pSX containing *carB* for over-expression	This study
6w^4-pSX	6w^4 with pSX	This study
6w^4-pSX-*carB*	6w^4 with pSX containing *carB* for over-expression	This study
6w^4-pSX-*pyrH*	6w^4 with pSX containing *pyrH* for over-expression	This study
6w^4-pSX-*ndk*	6w^4 with pSX containing *ndk* for over-expression	This study
6w^4-pSX-*galU*	6w^4 with pSX containing *galU* for over-expression	This study
Escherichia coli		
DH5α-*λpir*	*supE*44, Δ*lacU*169, *hsdR*17, *recA*1, *endA*1, *gyrA*96, *thi*-1, *relA*1, *λpir*	Invitrogen
TOP10	F', *mcrA*, Δ(*mrr-hsd*RMS-*mcr*BC), Φ80*lacZ*ΔM15, Δ*lac*X74, *deoR*, *recA*1, *ara*D139, Δ*(ara-leu)*7697, *gal*U, *gal*K, *rps*L, Str^R, *endA*1, *nup*G	Invitrogen

Table 2.1: Designations and characteristics of bacterial strains used.

2.1.2 Plasmids and transposons

Plasmids and transposons used in this study are listed below in Table 2.2.

Name	Characteristics	Reference
Plasmid		
pCR8/GW/TOPO	Spe^R, Puc *ori*; 2.8 kb 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 for tri-parental mating	(Figurski & Helinski, 1979)
pCre	A derivative of pUT, carrying the *cre* gene from pRH133, Cm^R, used to excise IS-Ω-Km/hah	(Manoil, 2000)
pSX	A derivative of pUCP22, carrying *lacIQ* and IPTG-inducible *Taq* polymerase promoter, Gm^R, Amp^R; 7 kb *Pseudomonas* expression vector	Ackerley, D.*
Transposon		
IS-Ω-Km/hah	Km^R, ColE1 *ori*, *npt* promoter, *loxP*	(Giddens *et al.*, 2007)

Table 2.2: Designations and characteristics of plasmids and transposons used. *This plasmid was the kind gift of Dr. David Ackerley, Victoria University.

2.1.3 Antibiotics, enzymes and reagents

Unless otherwise stated, all chemicals were obtained from Sigma-Aldrich. Antibiotics were purchased from Melford Laboratories and used at the following concentrations: tetracycline (Tc) 10 or 25 μg ml^{-1} as specified (in 1:1 ethanol:water), kanamycin (Km) 100 μg ml^{-1}, gentamicin (Gm) 10 μg ml^{-1} (liquid cultures) or 20 μg ml^{-1} (plates), spectinomycin 100 μg ml^{-1}, ampicillin 100 μg ml^{-1} and cycloserine 800 μg ml^{-1}. N-[5-Nitro-2-furfurylidene]-1-aminutesohydantoin (NF) was dissolved in dimethyl sulfoxide (DMSO) to a final concentration of 100 μg ml^{-1} and used to inhibit *E. coli* growth. Restriction enzymes (New England Biolabs) were used to carry out digestions at 37°C according to manufacturer's instructions. T4 DNA ligase (New England Biolabs), Taq polymerase, Elongase and PFX Taq (Invitrogen) were used as detailed in the text. The following combinatorial enhancer solution (CES) was added to PCR mixtures where indicated: 2.7 M betaine, 6.7 mM dithiothreitol (DTT), 6.7 % ($^{w}/_{v}$) DMSO, 55 μg ml^{-1} bovine serum albumin (Ralser *et al*., 2006). 5-Bromo-4-chloto-3-indolyl-β-D-galactopyranoside (X-gal, Melford Laboratories) was used at 60 μg ml^{-1} as a chromogenic indicator of β-galactosidase activity.

2.1.4 Media and culture conditions

Unless otherwise specified, all chemicals in this section were obtained from BDH, and all equipment from BioLab. Bacteria were cultured in KB (King *et al*., 1954): 10 g glycerol, 20 g Protease Peptone No.3 (liquid medium, DIFCO) or tryptone (plates, Remel), 1.5 g $K_2PO_4.3H_2O$, 1.5 g $MgSO_4.7H_2O$ L^{-1}, Lysogeny Broth (LB; Bertani, 1951): 10 g NaCl, 10 g tryptone, 5 g yeast extract L^{-1}, or minimal M9 medium (Sambrook *et al*., 1989): 34 g Na_2HPO_4, 15 g KH_2PO_4, 2.5 g NaCl, 5 g 1.8M NH_4Cl, 15 mg 1 M $CaCl_2.6H_2O$, 1 ml 1 M $MgSO_4.7H_2O$, 10 ml 20 % ($^{w}/_{v}$) glucose L^{-1}. Where appropriate, bacteriological agar (H.J. Langdon) was added to a concentration of 1.5 % ($^{w}/_{v}$). Uracil, guanine hydrochloride (dissolved in 1 M NaOH) and L-arginine hydrochloride were added as detailed in the text. Overnight cultures were grown for 16 (or 24 hours where indicated) with shaking at 150 rpm unless specified, either in 25 ml glass microcosms containing 6 ml KB or 30 ml plastic tubes containing 5 ml LB. *P. fluorescens* strains were grown at 28°C while *E. coli* strains were grown at 37°C.

2.1.5 DNA extraction materials

Extraction of genomic DNA required RNA-specific Ribonuclease A (RNase A), proteinase K and lysozyme. Purification of genomic DNA required 10 % ($^{w}/_{v}$) sodium dodecyl sulfate (SDS) and 2 % ($^{w}/_{v}$) cetyl trimethyl ammonium bromide (CTAB, in deionised water: 1.4 M NaCl, 20 mM EDTA, 100 mM Tris HCl; pH 8.0; 0.2 % ($^{v}/_{v}$) 2-mercaptoethanol added immediately prior to use). Further purification was achieved using chloroform washes and filter columns (DNeasy® Blood & Tissue Kit, Qiagen). Final DNA products were re-suspended in TE buffer (in deionised water: 1 M Tris HCl, 0.5 M EDTA; pH 8.0).

2.1.6 Primers

Synthesized by Invitrogen and Integrated DNA Technologies (IDT), primers were re-suspended in deionised water to a concentration of 100 pmol μl^{-1}, and were stored at -20°C. Primers were used at a final working concentration of 10 pmol μl^{-1}.

Name	**Sequence (5'→3')**	**Target**
Identification and ordering of mutations in reverse evolution lines		
MwsR2f	CTCGACCGCTTCAAGCCG	*mwsR* (line 1: 1 & 7*)
Mws8r	GGGCTTCGTTGACTTCGCGC	*mwsR* (line 1: 1 & 7)
Mws6fw	GCCTACGCGCTGACGGTCGC	*mwsR* (line 1: 2 & 8)
MwsR2r	GCCCAGGGAGTCGTTG	*mwsR* (line 1: 2 & 8)
Aws11f	GCTGGTTCAGCTTGATCGAACCC	*awsX* (lines 1 & 6: 3)
Aws16r	ATGGATGCTGCCGATGGTTC	*awsX* (lines 1 & 6: 3)
Aws8f	ATGCCAATGGCAAGTTGCTGG	*awsR* (lines 1 & 6: 4)
Aws18r	CATTGAGTAAGGTGCCAGGGGTGG	*awsR* (lines 1 & 6: 4)
WspF2f	GCCAGTTTCCACGATGACG	*wspF* (line 1: 5, line 6: 1, 2 & 5)
WspF2r	CAGTCACCTGAATTACTGCCTG	*wspF* (line 1: 5, line 6: 1, 2 & 5)
WssAf	CAGAATGAGCCCGACACCAC	*wssA* (line 1: 6)
WssAr	GGCATGTCGAGACCGAGGC	*wssA* (line 1: 6)
CarBf	GTACTTTGAGCCGGTAACACTGG	*carB* (line 1: 9; $1w^4$-reD1.8: 9)
CarBr	GGCTGTCGTTGGACACTTTCAC	*carB* (line 1: 9; $1w^4$-reD1.8: 9)
WssBf	GCCTGACCAACTCGCGGATC	*wssB* (line 6: 6)
WssBr	GGCTGAGGGCTTCGAATTGG	*wssB* (line 6: 6)

Name	**Sequence (5'→3')**	**Target**
2-21f	CGATCGCCTTCCGTTATGG	Pflu1301 (line 6: 7 & 8)
2-21r	AACGACGGGGTAGGAATTGC	Pflu1301 (line 6: 7 & 8)
RpoDf	GCAGATGTTGCAGGAAATGGGTC	*rpoD* (line 6: 9)
RpoDr	CCGTTGCGGTCCAGGTC	*rpoD* (line 6: 9)
PyrHf	AGATCTGATGGGCTCGGAAGAGTTCGG	*pyrH* (1w^4-reN1.4: 9)
PyrHr	AGATCTGTTGGCCTTCCTCGATCAGAGTC	*pyrH* (1w^4-reN1.4: 9)
CarABpf	GCGATGTGGTGGGTGATC	*dapB* (upstream of *carA*)
CarA2f	CGAGTTGCCTTACCTCGTGG	*carA*
CarAf	CAAGTCGCTGTTCGACGGTTC	*carA*
CarINTf	GACCGTTGCCAAGATCATCG	IG *carA/B* (1w^4-reN1.2: 9)
CarB7f	CTGCATCATCGTTTGCTCTATCG	*carB* (1w^4-reN1.5 & -D12: 9)
CarB3f	AGATCTTGGGTGTGACCGAGAAGAACC	*carB* (1w^4-reD1.8: 9)
CarB6f	CAGGACGAAATGCGCGAAC	*carB* (1w^4-reD2: 9)
CarAr	GGTCTTGCACAGGGTGGTTG	*carA*
CarINTr	GAGTTCACCAGGATGACGCG	IG *carA/B*
CarB8r	GCAGCTCTTTGGTCGGAGAC	*carB* (1w^4-reN1.2: 9)
CarB7r	GCGCCGCCAGTAATTTCGTTC	*carB* (1w^4-reN1.5 & -D12: 9)
CarB6r	CGATGCCTTGACCGATACGG	*carB* (1w^4-reD1.8: 9)
CarB4r	AGATCTCGGGTTGACTTCGATGACGTAG	*carB* (1w^4-reD2: 9)
GreAr	CACAGCATTCTGCATACGGC	*greA* (downstream of *carB*)
Allelic replacements		
CarB3f	AGATCTTGGGTGTGACCGAGAAGAACC	*carB*
CarB4r	AGATCTCGGGTTGACTTCGATGACGTAG	*carB*
RpoDf2	GAAGATCTGCAGATGTTGCAGGAAATGGGTC	*rpoD*
RpoDr2	GAAGATCTCCGTTGCGGTCCAGGTC	*rpoD*
WssKO-1b	GAAGATCTGAGTGAGTCGAGCAGATGAC	5' end of *wssA*
WssKO-2b	CAGCATGCGGATCCGTTGACGGACTGATCGAGCGTG CTGAAGG	*wssA*
WssKO-3b	TCCGTCAACGGATCCGCATGCTGGCAATCCCAGTCC GTGGATAAG	*wssJ*
WssKO-4b	GAAGATCTGCAAAGCTCGGTGATATCGTC	3' end of *wssJ*
CarB1	GAAGATCTGTGCGGGCTGTATAACCAG	3' end of Pflu5623
CarB2	CAGCATGCGGATCCGTTGACGGACTTGCAGGACCTA CACGCAG	3' end of *carB*
CarB5	TCCGTCAACGGATCCGCATGCTGCAAATCCGCAGGT CAGTCG	IG *carA/B*
CarB6	GAAGATCTGATGTCGCTGTCCGATTACC	*carA*
RecAKO-1	GAAGATCTCAATTGGCTGCCGAACTGGG	Pflu1188

Name	Sequence (5'→3')	Target
RecAKO-2b	CAGCATGCGGATCCGTTGACGGAGGCGGCCAACACCTGTATAAG	IG Pflu1188/*recA*
RecAKO-3	TCCGTCAACGGATCCGCATGCTGGCTGACACTGACATCTGAAGC	3' end of *recA*
RecAKO-4	GAAGATCTCAAAGCTCCCTTGCCACAGC	IG *recX*/Pflu1191
ArcBKO-1	GAAGATCTCGCGCAACCTGTTCAAGAAC	*arcA*
ArcBKO-2	GGCTACGACGATGCGCATGATGGATCTCCTTCTGTTTAGGTTATCC	5' end of *arcB* (Pflu4892)
ArcBKO-3	ATGCGCATCGTCGTAGCC	5' end of *arcC*
ArcBKO-4	GAAGATCTCAAAGATGCGTTTGGGCCG	*arcC*
Transposon mutagenesis		
Tn*phoA*-II	GTGCAGTAATATCGCCCTGAGCA	IS-Ω-Km/hah
CEKG 2A	GGCCACGCGTCGACTAGTACNNNNNNNNNNAGAG	Non-specific
CEKG 2B	GGCCACGCGTCGACTAGTACNNNNNNNNNNACGCC	Non-specific
CEKG 2C	GGCCACGCGTCGACTAGTACNNNNNNNNNNGATAT	Non-specific
Hah-1	ATCCCCCTGGATGGAAAACGG	IS-Ω-Km/hah
CEKG 4	GGCCACGCGTCGACTAGTAC	5' end of CEKG 2A, B, & C
Expression vectors		
WcaJ-lacZf	GAAGATCTGTATTGCGCCGCGTGATC	5' end of *wcaJ*
WcaJ-lacZr	GAAGATCTGCGCTCAGTAGATATCCTTGG	3' end of *wcaJ*
CarBOE-f	GGAATTCCATATGCCAAAACGTACAGACATAAAAAGC	5' end of *carB*
CarBOE-r	CGGGATCCTCATGCCTTCAATCCTGCGTG	3' end of *carB*
PyrHOE-f	GGAATTCCATATGGCTCAGCAGGGCAGTGGTTATC	5' end of *pyrH*
PyrHOE-r	CGGGATCCTCATTGTTGGCCTTCCTCGATCAG	3' end of *pyrH*
NdkOE-f	GGAATTCCATATGGCTGTTCAACGTACTTTCTCCATCATC	5' end of *ndk*
NdkOE-r	CGGGATCCTTAGCGAGCGGTTACTTCAGTAGC	3' end of *ndk*
GalUOE-f	GGAATTCCATATGATCAAGAAATGCTTGTTCCCAGCAG	5' end of *galU*
GalUOE-r	CGGGATCCTCAGTAAGCCTTGCCAGTCTTGTAG	3' end of *galU*

Table 2.3: Designations, sequences and amplification targets of primers used. *Denotes evolutionary line and chronological mutation that primer confirmed (e.g. line 1, 1 & 7 means reverse evolution line one, mutations one and seven). Incorporated *Bgl*II, *Nde*I & *Bam*HI restriction sites are underlined. IG=Intergenic, MCS=multiple cloning site, N=any base.

2.1.7 Chemically competent cell production materials

During manufacture of *Pseudomonas* chemically competent cells, sterile TG salts of the following recipe were required: 75 mM $CaCl_2$, 6 mM $MgCl_2$ and 15 % ($^{w}/_{v}$) glycerol. Production of chemically competent *E. coli* cells required filter-sterilised TBFI (30 mM KOAc, 100 mM RbCl, 10 mM $CaCl_2$; pH 5.8) and TBFII (10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$, 15 % ($^{w}/_{v}$) glycerol; pH 7.0) buffers.

2.1.8 Microscopy materials

Colony-level microscopy was performed using a Zeiss Stemi 2000-C dissection microscope. In order to visualise the production of cellulose, 0.005 % ($^{w}/_{v}$) Congo Red was added to media. Cell-level microscopy was performed using a Zeiss Axiostar Plus light microscope, coupled with fluorescence lighting (HBO 50/AC) where required. Ten μg ml^{-1} calcofluor (Fluorescent Brightener 28) was added to media for visualisation of cellulose. Where appropriate, 10 % ($^{v}/_{v}$) India ink (Pébéo, NZ) was used to stain cell capsules for visualisation. A Canon PowerShot A640 camera was used to record all microscopic images. Where required, images were adjusted using iPhoto version 7.1.5.

2.2 Methods

2.2.1 Genomic DNA extraction

Genomic DNA was prepared using a combination of CTAB, chloroform and column purification techniques. In order to obtain undamaged DNA, samples were handled gently throughout the procedure. Firstly, 1 ml overnight culture was pelleted by centrifugation (6,000 x g, 10 minutes) and re-suspended in TE buffer (pH 8.0) containing 1 mg ml^{-1} lysozyme. 4 µl 100 mg ml^{-1} RNase A was added and the mixture incubated at 37˚C for 3 minutes. Subsequently, 30 µl 10 % ($^{w}/_{v}$) SDS and 20 µl proteinase K were added and the mixture incubated at 37˚C with gentle mixing every 10 minutes. After 1 hour, 100 µl 5 M NaCl and 80 µl 2 % ($^{w}/_{v}$) CTAB (pre-heated to 60˚C for 1 hour) were added. The resulting mixture was incubated at 65˚C for 10 minutes before addition of 600 µl chloroform and centrifugation (13,300 x g, 5 minutes). The supernatant was transferred to a fresh microcentrifuge tube and washed with a further 600 µl chloroform. Finally, the sample was purified on a filter column according to the manufacturer's instructions. Genomic DNA was re-suspended in a final volume of 50 µl TE buffer (pH 8.0). In cases where the sample was to be used for genome re-sequencing, the quality of the sample was checked using agarose gel electrophoresis (see section 2.2.4) and spectrophotometry (260/280 ratio: 1.8-2.0).

2.2.2 Polymerase chain reaction (PCR)

2.2.2.1 Standard PCR

Polymerase chain reactions were carried out using a CG1-96 Thermal Cycler (Corbett Life Sciences). A standard 25 µl reaction contained: 2.5 µl 10 *x* PCR buffer, 0.5 µl 10 mM dNTP mix, 0.8 µl 50 mM $MgCl_2$, 0.2 µl Taq polymerase, 1 µl of each 10 pmol μl^{-1} primer and 30-100 ng template DNA, made up to 25 µl with deionised water. Following an initial template denaturation step of 3 minutes at 94°C, amplification was performed by 30 cycles of denaturation for 15 seconds at 94°C, annealing for 30 seconds at 55-62°C, and strand extension for 1 minute per kilobase (kb) of target DNA at 72°C. In

cases where PCR products were to be used for cloning into pCR8/GW/TOPO, a final extension step of 10 minutes at 72°C was performed before the sample was cooled to 4°C indefinitely.

In cases of longer amplification targets (2-4 kb), Elongase was used. For each reaction, two separate mixtures were set up as follows: (A) 1 µl 10 mM dNTP mix, 1 µl of each 10 pmol μl^{-1} primer and 100-200 ng template DNA, made up to 20 µl with deionised water, and (B) 4 µl 5 x buffer A, 6 µl 5 x buffer B (giving 1.6 mM Mg^{2+}), 1 µl Elongase, made up to 30 µl with deionised water. These two mixtures were combined to give a final reaction volume of 50 µl. Following an initial template denaturation step of 30 seconds at 94°C, amplification was performed by 35 cycles of denaturation for 30 seconds at 94°C, annealing for 30 seconds at 55-62°C, and strand extension for 1 minute per kb of target DNA at 68°C. In cases where PCR products were to be used for cloning into pCR8/GW/TOPO, 1 µl of standard Taq polymerase was added upon completion of the programme, and a final extension step of 10 minutes at 72°C was performed before the sample was cooled to 4°C indefinitely.

2.2.2.2 Strand overlap extension (SOE-PCR)

SOE-PCR (Ho *et al.*, 1989) was used during the construction of systematic genomic deletions (Chapters 3 and 6). In this process, two overlapping PCR products are amplified and annealed, producing a single PCR product, which is subsequently amplified. DNA on either side of the targeted locus was amplified using the standard PCR techniques described above, with primers that complemented for approximately 25 bp at the site of joining. Equal amounts of the two PCR products were used as templates in an SOE-PCR reaction containing 5 µl 10 x PCR buffer, 1 µl 10 mM dNTP mix, 1.5 µl 50 mM $MgCl_2$, 0.2 µl Taq polymerase, 1 µl of each 10 pmol μl^{-1} primer, 8 µl of template, made up to 50 µl with deionised water. Samples were subjected to the same temperature cycle as described for standard PCR.

2.2.2.3 Arbitrary primed-PCR (AP-PCR)

An AP-PCR technique developed by Manoil (2000) and adapted by Jacobs *et al*. (2003) was employed in Chapters 4 and 7 to amplify transposon-chromosome junctions. This technique involves two successive rounds of PCR; the first PCR contained 2.5 µl 10 x PCR buffer, 1 µl 10 mM dNTP mix, 0.8 µl 50 mM $MgCl_2$, 0.25 µl Taq polymerase, 2 µl of each 10 pmol μl^{-1} primer and 3 µl of the appropriate DNA template, made up to 20 µl with deionised water. Following an initial template denaturation step of 10 minutes at 94°C, amplification was performed by six cycles of 30 seconds at 94°C, 30 seconds at 42°C, and 3 minutes at 72°C, where the annealing temperature was decreased by 1°C for every cycle. Amplification was continued by 25 cycles of 30 seconds at 94°C, 30 seconds at 65°C, and 3 minutes at 72°C. The products of this PCR were diluted by addition of 80 µl of deionised water, and used as templates in the second round of PCR. The second PCR contained: 2.5 µl 10 x PCR buffer, 1 µl 10 mM dNTP mix, 0.8 µl 50 mM $MgCl_2$, 0.25 µl Taq polymerase, 2 µl of each 10 pmol μl^{-1} primer and 2 µl of the appropriate DNA template, made up to 20 µl with deionised water. Following an initial template denaturation step of 10 minutes at 94°C, amplification was performed by 30 cycles of 30 seconds at 94°C, 30 seconds at 65°C, and 3 minutes at 72°C. The sample was cooled to 4°C and prepared for sequencing.

2.2.3 Cloning and transformation techniques

2.2.3.1 Plasmid purification, digestion and ligation

Cloning was carried out under sterile conditions according to standard molecular biology techniques (Sambrook *et al*., 1989). Plasmid DNA was extracted from overnight bacterial cultures using a QIAprep® Spin Miniprep Kit (Qiagen), and stored in 25 µl aliquots at -20°C. Following digestion of the extracted plasmid and desired insert with the appropriate restriction enzyme(s), ligation reactions were carried out by overnight incubation at 4°C with T4 DNA ligase in a final volume of 10 µl, containing 2 µl 5 x ligase buffer, 4.5 µl insert and 3 µl vector. Alternatively, freshly amplified PCR products were cloned directly into pCR8/GW/TOPO using the TOPO® TA cloning protocol. Reactions were set up according to manufacturer's instructions: 4 µl fresh

PCR product, 1 µl salt solution and 1 µl vector were gently mixed and incubated at room temperature (23-25°C) for 10 minutes, before storing at -20˚C.

2.2.3.2 Manufacture and transformation of chemically competent cells

2.2.3.2.1 Manufacture of Pseudomonas *chemically competent cells*

P. fluorescens SBW25, $1w^4$, $1w^4$-reN1.4 and $6w^4$ chemically competent cells were produced from overnight KB cultures using a protocol adapted from Chuanchuen *et al*. (2002). All steps were performed on ice or at 4˚C. 1 ml aliquots of culture were pelleted (13,000 x g, 30 seconds) and re-suspended in 1 ml ice-cold 0.1 M $MgCl_2$. Re-suspensions were pelleted again (13,000 x g, 30 seconds) and re-suspended in 1 ml ice-cold TG salts before incubation on ice for 10 minutes and pelleting (13,000 x g, 30 seconds). Pellets were finally re-suspended in 200 µl ice-cold TG salts. 100 µl aliquots of the newly competent cells were snap frozen and stored at -80˚C until required.

2.2.3.2.2 Manufacture of E. coli *chemically competent cells*

E. coli DH5α-*λpir* chemically competent cells were produced from overnight cultures grown to mid-log phase in 200 ml LB medium. Four 50 ml aliquots of cell culture were pelleted in Falcon tubes (500 x g, 10 minutes). Pellets were re-suspended in 1 ml ice-cold TBFI, and subsequently combined. 10 ml further TBFI was added and the resulting mixture incubated on ice for 60 minutes. Cells were then pelleted (500 x g, 10 minutes) and re-suspended in 4 ml ice-cold TBFII. 50 µl aliquots of the newly competent cells were snap frozen and stored at -80˚C until required.

2.2.3.2.3 Transformation of P. fluorescens *and* E. coli *chemically competent cells*

During transformations, the relevant chemically competent cells were thawed at room temperature (*P. fluorescens*) or on ice (*E. coli*) for 5 minutes before the addition of 4 µl ligation mixture or intact plasmid DNA. Reactions were then incubated on ice for 30 minutes before administration of a heat shock at 45˚C for 4 minutes (*P. fluorescens*) or

42˚C for 30 seconds (*E. coli*). Heat-shocked cells were subsequently incubated on ice for 2 minutes, and recovered in 250 µl LB at 28˚C (*P. fluorescens*) or 37˚C (*E. coli*) for 1 hour. Finally, cells were plated onto appropriate selective medium.

2.2.4 Agarose gel electrophoresis

Agarose gels were run at 100 V until sufficient separation of loading buffer indicator dyes was attained. DNA bands were viewed using a UV Transilluminator and Gel Documentation System. Where appropriate, DNA bands were excised from the gel under UV light using sterile razor blades, and DNA was extracted from the agarose using a QIAquick® Gel Extraction Kit (Qiagen).

2.2.5 DNA sequencing

2.2.5.1 Sanger sequencing

Plasmid DNA and PCR products were prepared for sequencing using a QIAprep® Spin Miniprep Kit and a QIAquick® PCR Purification Kit or QIAquick® Gel Extraction Kit (Qiagen), respectively. Purified DNA samples were sent to Macrogen (Seoul, Korea) for Sanger sequencing and analysis. Sequence traces were analyzed using 4Peaks version 1.7.2 (Mekentosj) and aligned using Geneious version 3.8.5.

2.2.5.2 Whole-genome re-sequencing

Whole-genome re-sequencing was achieved using high-throughput, amplification-based Illumina Solexa technology available through The Allan Wilson Centre Genome Service (AWCGS, Massey University, New Zealand). Many 36 bp sequence reads were generated on a Solexa Genome Analyzer and aligned to the *P. fluorescens* SBW25 genome sequence (Silby *et al*., 2009) using ELAND (Illumina, Inc.) Data that were unable to be aligned in this manner were further analysed using bioinformatic programmes designed in conjunction with and written by Frederic Bertels (Massey

University, New Zealand). Final alignments and analyses were viewed using the Generic Genome Browser (GBrowse) version 1.69 (Stein *et al.*, 2002). Protein domains and associated *E*-values were identified using the Pfam database (Finn *et al.*, 2008).

2.2.6 Conjugation

2.2.6.1 Bi-parental conjugation

Overnight cell cultures of the *E. coli* donor and *P. fluorescens* recipient strains were grown in LB containing appropriate antibiotics. 1 ml of the recipient culture was heat-shocked for 20 minutes at 45˚C before pelleting (13,000 x g, 1 minute). 300 µl of the donor culture was also pelleted (13,000 x g, 1 minute). Both pellets were re-suspended in 200 µl LB, and subsequently mixed thoroughly. Cells were pelleted once more (13,000 x g, 1 minute) prior to re-suspension in 30 µl LB. The resulting concentrated re-suspension was gently dropped and spread onto the centre of a pre-warmed LB agar plate, allowed to dry and incubated for 24 hours at 28°C. After incubation, the inoculum was removed and re-suspended in 1 ml fresh LB. A dilution series was prepared in sterile water and plated onto LB agar plates containing NF (to select against *E. coli*) and Km or Tc (to select for the donor plasmid), as specified. After 48 hours growth at 28˚C, individual colonies were selected and purified by re-streaking, and stored at -80˚C.

2.2.6.2 Tri-parental conjugation

When conjugating from an *E. coli* donor strain not containing the *tra* and *mob* genes necessary for conjugation, a helper plasmid containing these genes was required. An overnight culture of *E. coli* DH5α containing the helper plasmid pRK2013 was grown in LB with Km. 300 µl of cell culture was pelleted and re-suspended in 250 µl LB, and mixed with re-suspended donor and recipient cultures. The method then followed the same protocol as for bi-parental conjugation above.

2.2.7 Enriched two-step allelic exchange

An enriched two-step allelic exchange protocol (Kitten *et al*., 1998) was used during systematic genomic manipulations (Chapters 3, 5, 6 and 7). In order to understand the protocol, it is pivotal to realise that upon entering SBW25 cells, pUIC3 integrates into the host chromosome *via* recombination at a locus homologous to that of its cloned insert. Therefore, the first step of the protocol is conjugation of the desired pUIC3 construct from an *E. coli* donor into the desired *P. fluorescens* recipient (see section 2.2.6.2). Subsequently, 10 µl of a resulting transconjugant *P. fluorescens* overnight culture was used to inoculate 200 ml LB in a 1 L flask (without antibiotic to allow loss of the chromosomal pUIC3 construct). Following 16 hours incubation at 28°C (150 rpm), 400 µl was transferred to 20 ml fresh LB in a 250 ml flask, and incubated for 30 minutes at 28°C (150 rpm). Ten µg ml^{-1} Tc was added prior to a further 2 hours incubation at 28°C (150 rpm). This selected for growth of bacteria that had incorporated the pUIC3 vector and the associated tet^R gene into the host chromosome (i.e. the original transconjugants). Cycloserine, an antibiotic that kills growing *Pseudomonas* cells, was added to the flask, and incubation continued for 4-5 hours. By killing the growing Tc-resistant cells, the addition of cycloserine selected for cells that had undergone a (rare) second round of homologous recombination, leading to loss of the chromosomally located vector (including tet^R and *lacZ*), along with either the original vector insert or the original chromosomal DNA. Cells were then harvested by centrifugation, washed and dilutions spread onto LB agar plates containing X-gal. These were incubated at 28°C for 48 hours before white colonies were picked and purified by re-streaking. Incorporation of the desired DNA fragment was checked by PCR and, in the case of mutation reconstruction, DNA sequencing (see section 2.2.5.1).

2.2.8 Transposon mutagenesis

2.2.8.1 Generation and isolation of transposon mutants

In Chapters 4 and 7, IS-Ω-Km/hah (a transposon which inserts into the *P. fluorescens* genome at random with respect to locus) was used to mutagenize $1w^4$ and $6w^4$ following a method previously used with success in the Rainey laboratory (Giddens *et*

al., 2007). A bi-parental conjugation was set up between the *E. coli* donor strain carrying the IS-Ω-Km/hah transposon on the pSCR001 plasmid and the *Pseudomonas* recipient (section 2.2.6.1). On the appropriate selective medium (LB agar containing NF and Km), $1w^4$ produces sectored colonies. Transconjugant colonies exhibiting a deficiency in colony sectoring (i.e. non-switching) were picked, purified and stored at -80°C. The site of transposon insertion was determined by PCR and sequencing, as described in sections 2.2.2.3 and 2.2.5.1). These sequence data were mapped to the *P. fluorescens* SBW25 genome using Artemis software (Rutherford *et al*., 2000). In cases where annotation of transposon insertion sites was unavailable, gene similarities were determined by BLASTP alignments using Sanger Institute Basic local alignment search tool (BLAST) server.

2.2.8.2 Transposon excision

Excision of genomic IS-Ω-Km/hah inserts may be performed using Cre recombinase-mediated excision. Upon completion, a 189 bp, in-frame fragment of the transposon remains at the site of original insertion, allowing investigation of potential polar effects of the original insertion event. Excision was achieved through bi-parental conjugation between the *E. coli* donor strain containing the pCre plasmid and the non-switching transposon mutant of interest (see section 2.2.6.1). Excision was confirmed by demonstrating sensitivity to Km.

2.2.9 Re-evolution from SBW25, $1s^4$ and $6w^0$ using the REE strategy

The REE strategy described by Beaumont *et al*. (2009) was used during re-evolution of experimental populations from SBW25, $1s^4$ and $6w^0$ (Chapters 5 and 7). Replicate KB microcosms were founded with the appropriate genotype taken from single colonies (purified by re-streaking from -80˚C glycerol stocks). Each replicate population was grown in a static (SBW25 and $1s^4$) or shaken ($6w^0$, 170 rpm) microcosm that was propagated by transfer of a mixed sample (6 µl) to a fresh KB microcosm at 72-hour intervals. Parallel with each transfer, population samples were spread onto KB agar plates to screen for colonies with the most abundant different colony morphology to the

founding genotype (approximately 1,000 colonies screened *per* replicate *per* transfer). Upon detection of a new colony types, a single colony of the numerically dominant type was streaked to single colonies on a KB agar plate to (i) confirm colony morphology heritability, and (ii) impose a single-individual bottleneck. Evolution of a genotype with a different colony morphology marked the end of a round of selection. Cells of new genotypes were taken from control plates, grown in a shaken KB microcosm (16 hours, 170 rpm), and stored at -80 °C. Finally, the *carAB* (for switching genotypes evolved from $1s^4$) or *wspF* loci (for SM types evolved from $6w^0$) of each re-evolved strain was PCR-amplified and sequenced to check for mutations.

2.2.10 Over-expression of nucleotide biosynthetic genes from pSX

Selected genes of the uracil biosynthetic pathway were over-expressed from the pSX plasmid in Chapters 6 and 7. For each gene over-expressed, three independent biological replicates were produced in each of the genetic backgrounds investigated. Firstly, the gene of interest was amplified by PCR and ligated into pCR8/GW/TOPO for sequencing (see sections 2.2.2.1 and 2.2.3.1). Once an error-free clone was obtained, the insert was removed from pCR8/GW/TOPO by simultaneous digestion with *Nde*I and *Bam*HI, ligated into pSX and the resulting construct used to transform chemically competent *E. coli*. *E. coli* clones containing the pSX construct were identified, the construct isolated and used to transform chemically competent *P. fluorescens* cells (see sections 2.2.3.1 and 2.2.3.2). Transformant *P. fluorescens* colonies were purified by streaking, the presence of the desired construct checked by PCR, and stored at -80˚C. The constructed strains were used to perform cell ratio counting assays as described in section 2.2.11.4, with the addition of Gm to the medium.

2.2.11 Biological assays

2.2.11.1 Measurement of growth rates

Overnight cell cultures were produced in KB microcosms (160 rpm), vortexed for 30 seconds, and a 1 ml aliquot of each culture washed in M9 medium. Each aliquot was re-

suspended in M9 to a common cell density (according to the optical density at 600 nm (OD_{600}), which was determined using a $VERSA_{max}$ microplate reader (Molecular Devices)). A 1.5 µl aliquot of each re-suspension was pipetted into 148 µl of appropriate fresh medium, contained within a well of a 96-well microtitre plate (RayLab). The microtitre plate was then placed in a microplate reader (BioTek®), and the OD_{600} of each well was measured at 5-minute intervals for 48 hours (with 5 seconds shaking prior to each read). Replicates of each assay were obtained and mean and standard error values determined.

2.2.11.2 Investigation of niche preference in static microcosms

The ability of *P. fluorescens* strains to form biofilms in KB static microcosms was assessed according to the method described in Rainey & Travisano (1998). Overnight cultures were prepared in KB (180 rpm). 6 µl of each culture was pipetted into a fresh KB microcosm and vortexed for 5 seconds. Following loosening of the caps, vials were incubated without shaking to produce a spatially heterogeneous environment. After 72 hours, each microcosm was examined and photographed prior to vigorous vortexing for 1 minute. 50 and 100 µl aliquots of a 10^6 fold dilution were spread on KB agar plates and incubated at 28°C for 48 hours. Analysis of colony morphologies on the plates involved visual, microscopic and photographic examination.

2.2.11.3 Microscopic analyses

2.2.11.3.1 Light microscopy

Cells of *P. fluorescens* strains subjected to microscopic analysis were grown as detailed in the text; commonly in KB microcosms (16 hours) or on KB agar plates (48 hours). In most cases, India ink was added to cells to allow capsule visualisation. Cells were diluted 1:5 in sterile water, mixed thoroughly with India ink, and allowed to incubate at room temperature for approximately 1 minute. Microscope slides were prepared by dropping 6 µl of the mixture onto the slide and covering with a glass cover slip. Care was taken to avoid the formation of air bubbles, as these cause non-homogeneous cell distributions on the slide. Slides were left at room temperature for 10 minutes to allow

motile bacteria to cease swimming. Then, each slide was placed under the light microscope and viewed under x40 or x63 magnification. Where appropriate, microscopic images were recorded for later analysis.

2.2.11.3.2 Fluorescence microscopy

Analysis of ACP production by *P. fluorescens* strains was done using a combination of calcofluor staining and fluorescence microscopy. Overnight cell cultures of the appropriate strains were prepared, and 5 µl of each culture was dropped onto a KB agar plate containing calcofluor. Drops were allowed to dry at room temperature and plates were subsequently incubated for 24 hours. The inoculum was scraped from the plates and re-suspended thoroughly in 200 µl sterile water. 8 µl drops of the mixture were pipetted onto microscope slides, covered with glass cover slips, and viewed at x40, x63 or x100 magnification under fluorescence lighting.

2.2.11.3.3 Dissection microscopy

Analysis of *P. fluorescens* colony phenotypes was done under a dissection microscope. Samples were prepared as described above for fluorescence microscopy, with the removal of calcofluor from medium and the addition of Congo Red to the KB agar plates where indicated. Samples were viewed under x12.5 magnification, and images were recorded where appropriate.

2.2.11.4 Calculation of relative frequencies of the switcher phenotypes – the capsule counting assay

Quantitative measures of relative frequencies of the switcher phenotypes were obtained through measurement of the percentage of cells expressing capsules in populations of interest, according to the following protocol. For each strain to be assayed, three or five replicate KB microcosms were inoculated from -80˚C glycerol stocks. The lids of the microcosms were loosened, and the microcosms subsequently incubated for 24 hours

(160-180 rpm). Following vortexing for 1 minute, 6 µl aliquots of cells from each pre-culture were used to inoculate fresh KB microcosms. After incubation for a further 24 hours, each of the microcosms was vortexed for 1 minute, and samples were prepared for microscopic analysis as described in section 2.2.11.3.1. Individual cells recorded in the microscopic images were manually assayed for capsule expression, with the aid of a programme written expressly for this purpose by Frederic Bertels (Massey University, NZ). For each replicate, the state of capsule expression (ON or OFF) was recorded for 500 cells; a maximum of 100 cells were recorded *per* image, to avoid bias from cells accumulating in particular areas of the microscope slide. Cell counting is exemplified below in Figure 2.1. Finally, mean and standard error values were determined for each strain, and compared graphically to those of other strains.

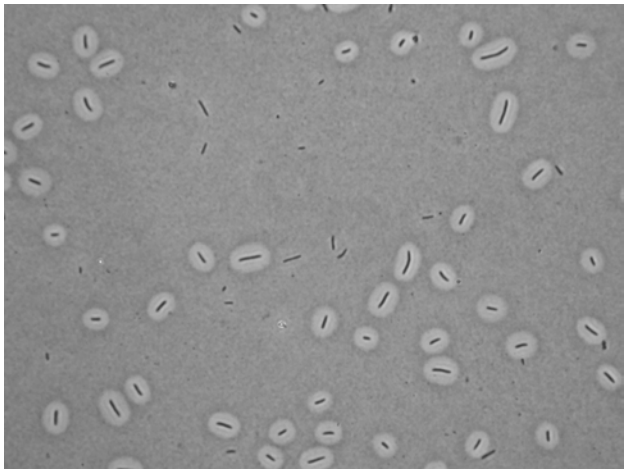


Figure 2.1: A typical image used to count proportions of capsulated and non-capsulated cells in populations of interest. In this assay, the capsulated state of each cell is recorded, starting from the top left and progressing systematically to bottom right, until a maximum of 100 cells have been assayed.

2.2.11.5 Competition experiments

The competition experiments of Chapter 5 were performed with the assistance of Dr. Christian Kost (Massey University, NZ). Strains were purified from glycerol stocks by streaking onto KB agar plates. Following 48 hours incubation, ten single colonies were selected and each one used to inoculate a single KB microcosm. Microcosms were incubated for 18 hours (160 rpm) before vortexing for 30 seconds. Fresh KB

microcosms were inoculated with approximately 5 x 10^6 cells of each competitor, and incubated for 24, 48 or 72 hours in either static or shaking (160 rpm) environments, as indicated. Cell samples taken at 0 hours and 24, 48 or 72 hours were plated or examined directly by light microscopy to determine frequencies of the competitors. Genotypes were distinguished on the basis of morphology or by a neutral *lacZ* marker (Zhang & Rainey, 2007). Relative fitness was calculated as detailed in section 2.2.12.2.

2.2.12 Statistical analyses

2.2.12.1 General statistical tests

Where multiple data points were taken, means or medians, standard deviations, standard errors and/or 95 % confidence intervals were calculated using Microsoft Office Excel 2004 (Microsoft Corporation). R 2003 Version 2.8.1 (The R Foundation for Statistical Computing) was used to apply other statistical tests. Parametric one-sample *t*-tests were used in Chapter 5 to detect deviance of the population mean from a specified value. Parametric and non-parametric (Welch form) two-sample *t*-tests were used to detect significant differences between means. Where normality assumptions for these tests were not satisfied, Mann-Whitney-Wilcoxon (M-W-W) tests were used to detect significant differences between medians.

2.2.12.2 Calculation of differential fitness in competition experiments

Statistical methods devised by Lenski (1991) were used to calculate the relative fitness values during competition experiments in Chapter 5. The density of competitors was calculated at the beginning and end of each competition, and these values were used to calculate Malthusian parameters for both strains (the natural logarithm of the final density of each strain divided by the corresponding initial density). Relative fitness was expressed as the ratio of the two Malthusian parameters.

Chapter 3:
Phenotypic Analysis of 1w^4

3.1 Introduction

Colony morphology is a complex, population level trait that is determined by a range of cellular level phenotypes including cell morphology, cell surface composition and exopolysaccharide biosynthesis (reviewed in van der Woude & Bäumler, 2004). Unsurprisingly, mechanisms that generate diversity in cellular phenotypes often produce corresponding variability in colony morphology. For instance, it has been demonstrated that ON/OFF capsule expression causes opaque and translucent colony forms of *Vibrio parahaemolyticus* (Enos-Berlage & McCarter, 2000), while phase-variable expression of the exopolysaccharide VPS underlies the smooth and rugose colony variants of *Vibrio cholerae* (Yildiz & Schoolnick, 1999). In some cases the relationship between cell phenotype and colony morphology is more complex, with the differential expression of multiple cellular components contributing to colony morphology. This is true for opaque-white colony switching in *C. albicans*, in which cell morphology and differential polymer expression contribute to colony morphology (see section 1.1.2.1.1; Lan *et al*., 2002).

The ability to generate morphologically variable colonies is a defining trait of 1w^4. This strain produces two colony morphs on KB agar: 'translucent' and 'opaque' (see Figure 1.15). The first section of this chapter describes the identification of the cell level phenotypes underlying 1w^4 biphasic colony morphology. This undertaking presented a considerable challenge; a full understanding of the 1w^4 phenotype requires knowledge of the significant evolutionary history separating 1w^4 from SBW25. Thus, each genotype of the evolutionary series was subjected to a range of phenotypic assays that focussed on analysis of polymer production and cell morphology. The second part of the chapter describes the use of a combination of microscopic and genetic techniques to provide an in-depth analysis of relevant phenotypes in 1w^4.

Finally, work in the third section of this chapter demonstrates that 1w^4 phenotypic switching occurs *via* a true switching mechanism. This involved verification that switching between 1w^4 subtypes is a reversible process; all phenotypic forms of 1w^4 must be able to independently generate all other phenotypic forms at a high rate (>10^{-5}). Despite the centrality of reversibility to switching mechanisms, direct demonstrations of reversibility are rare. While it is possible to directly observe reversibility using microscopic techniques, many laboratories do not possess the required technology. Here, a powerful statistical technique is used to demonstrate that individuals of each subtype generate phenotypically mixed populations.

3.2 Aims

1. To investigate the phenotypic history of 1w^4. This will include analysis of colony morphology, cell morphology and polymer biosynthesis in each genotype of the line one evolutionary series.

2. To perform a detailed investigation in 1w^4 of relevant phenotypes identified by the first aim, using a combination of biochemical and genetic techniques.

3. To test the hypothesis that 1w^4 phenotypic switching is reversible.

3.3 Results

3.3.1 Phenotypic analysis of the 1w^4 evolutionary line

In order to obtain insight into the structural basis of translucent-opaque colony switching, the phenotypic history of 1w^4 was investigated. This included analysis of colony morphology, cell morphology, niche preference in static microcosms and ACP biosynthesis in each genotype of the evolutionary line (Figure 3.1). Each of these assays is described in detail below. With some interesting exceptions, the overall trend was that genotypes evolved in the shaken environment (1s^1, 1s^2, 1s^3 and 1s^4) phenotypically resembled SBW25, while those evolved in the static environment (1w^0, 1w^1, 1w^2, 1w^3 and 1w^4) showed WS-like traits.

3.3.1.1 Colony morphology

For each genotype, of a 10^6-fold dilution of an overnight KB culture was produced, and a 25 µl aliquot spread onto KB agar. Following incubation at 28˚C for 48 hours, colony morphology was examined, and comparative photographs were taken using a dissection microscope (Figure 3.1A). Genotypes evolved in shaken microcosms (1s^0, 1s^1, 1s^2, 1s^3 and 1s^4) produced large, smooth-edged colonies, while genotypes evolved in the static environment (1w^0, 1w^1, 1w^2 and 1w^3) produced small, wrinkly-edged colonies. The sole exception to this trend was 1w^4, which simultaneously produced two colony types: WS-like 'translucent' colonies, and 'opaque' colonies that did not resemble any colonies previously described in genotypes derived from SBW25. For further information, see section 3.3.2.1.

3.3.1.2 Cell morphology and capsule production

Cell morphology and capsule production was assayed using India ink, a dye that counter-stains cells. Cell samples from each strain were prepared according to the method outlined in section 2.2.11.3.1, and photographed under a light microscope (Figure 3.1B). The first nine genotypes of the evolutionary line (including SBW25)

produced rod shaped, motile, non-capsulated cells. 1w^4 gave rise to rod shaped cells of two types: motile, non-capsulated cells (cap-), and non-motile, capsulated cells (cap+; white, non-staining region surrounding cells). A flagella and capsule co-staining technique devised by Leifson (1930; 1951) was attempted. Unfortunately the fragility of flagella attachment rendered this technique ineffective; the majority of observed flagella were detached from cells. However, intact flagella were only ever observed on non-capsulated cells. See section 3.3.2.2.1 for further discussion.

3.3.1.3 Niche preference in static microcosms

For each genotype, three replicate static microcosms were produced according to the method outlined in section 2.2.11.2. Microcosms were incubated for only 48 hours, as this allowed sufficient time for a stable mat to form, but insufficient time for new mat-forming types to evolve. Subsequently, each (undisturbed) microcosm was photographed under identical conditions (Figure 3.1C). Overall, genotypes evolved in shaken microcosms grew throughout the broth phase, while genotypes evolved in the static environment formed a mat at the air-liquid interface. However, the mat-forming phenotype was not clear-cut; three strains (1w^2, 1w^3 and 1s^4) colonised both the broth phase and the air-liquid interface. Notably, 1w^4 did not grow in the broth phase, but formed a thick mat at the air-liquid interface.

3.3.1.4 ACP production: Congo red binding

The observed WS-like phenotypes of some genotypes were suggestive of ACP biosynthesis (see section 1.2.1). To investigate this possibility, the ability of each genotype in the series to bind Congo red was tested. Congo red is a dye that binds various extracellular components including ACP, causing the formation of red colonies. Cells from each genotype were grown on KB agar containing Congo red (see section 2.2.11.3.3). Comparable photographs were taken of each genotype under a dissection microscope (Figure 3.1D). With the exception of 1s^4, genotypes evolved in the shaken environment bound little Congo red. Along with genotypes evolved in the static environment, the 1s^4 colony stained a darker red, signifying stronger Congo red binding.

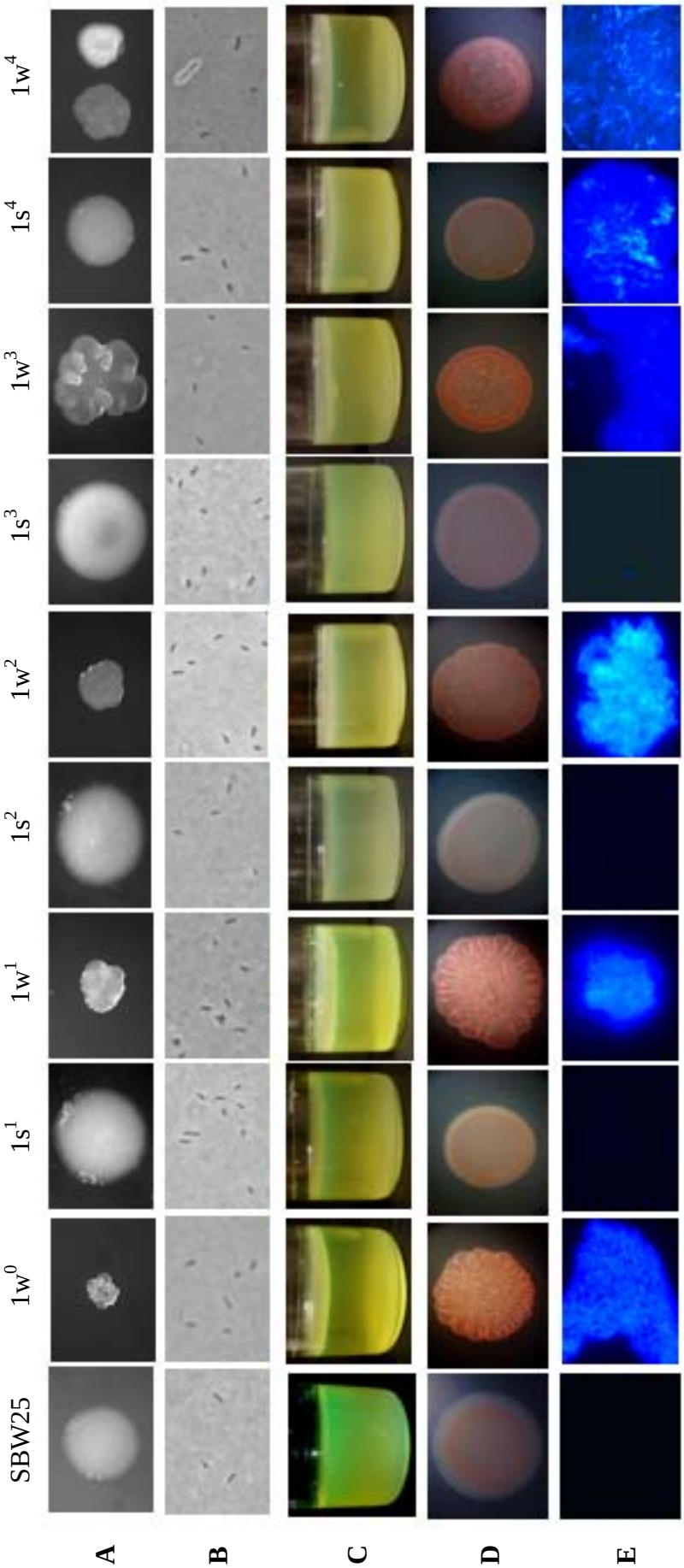


Figure 3.1: Phenotypic analysis of the line 1 evolutionary series. (**A**): Colony morphology on KB agar (48 hours), photographs taken at same magnification. (**B**): Light microscope images of cells counter-stained with India ink (x63). (**C**): Niche preference in a 48-hour static microcosm. (**D**): Dissection microscope images of colony morphology on Congo red agar (48 hours). (**E**): Fluorescence microscope images of ACP-mediated calcofluor binding (x40).

3.3.1.5 ACP production: calcofluor binding

Calcofluor, a dye that fluoresces blue under light of wavelength ~240 nm, binds to ACP with greater specificity than Congo red. Cells of each genotype were grown on KB agar containing calcofluor, as detailed in section 2.2.11.3.2. Samples of each genotype were subsequently viewed and photographed under a fluorescence microscope at x100 magnification (Figure 3.1E). With the exception of 1s^4, genotypes evolved in the shaken environment did not bind calcofluor, while those evolved in the static environment did. This indicated that, of the genotypes in the evolutionary lineage, 1w^0, 1w^1, 1w^2, 1w^3, 1s^4 and 1w^4 synthesized ACP. See section 3.3.2.2.2 for further discussion.

3.3.2 Analysis of 1w^4 colony and cell morphology

In addition to colony morphology, the above phenotypic analyses exposed phenotypic differences between SBW25 and 1w^4 in capsule production and ACP biosynthesis. Capsule production in 1w^4 was dimorphic, while the nature of ACP biosynthesis was unclear. It seems likely that the dimorphic colony phenotype is the result of variable expression of capsules and/or ACP. In order to examine this hypothesis, each of these three phenotypes was investigated in detail.

3.3.2.1 Colony morphology in 1w^4

Translucent/opaque colony switching in 1w^4 was first recognised on KB agar. The sensitivity of this phenotype to medium composition was tested using a range of media. An overnight culture of 1w^4 was produced, washed in Ringer's solution and diluted by a factor of 10^6 before spreading onto agar plates containing KB, LB, M9 or LBS (LB containing 5 % sucrose) medium. Plates were subsequently incubated at 28˚C for 48 hours (72 hours in the case of M9), and visually examined and photographed (Figure 3.2A and B). Complete colony dimorphism was only apparent on KB agar. On LB and M9 agar, colonies were of a single type: a translucent-like base interspersed with opaque-like sectors. Interestingly, growth on M9 was slow, and the colonies small, suggesting a possible metabolic defect in 1w^4. Growth on LBS agar produced extremely

uniform, glutinous colonies. A similar mucoid phenotype was observed when ancestral SBW25 was grown on LBS. The mucoid phenotype is thought to result from the biosynthesis of levan, a homopolysaccharide composed of D-fructofuranosyl residues (P.B. Rainey, personal communication). Presumably, if 1w^4 dimorphism is still present on LBS, it is masked by levan production.

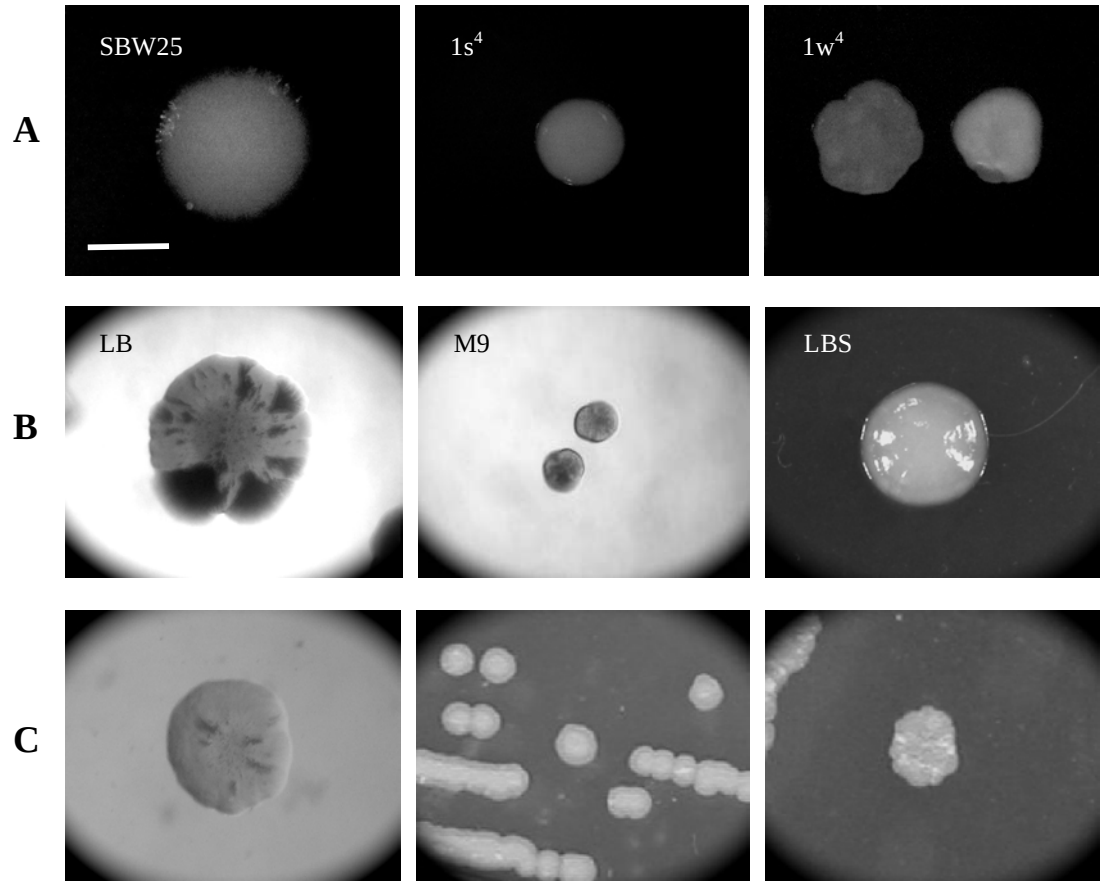


Figure 3.2: The colony phenotype of 1w^4 is dependent on media composition. (**A**): Monomorphic colonies of SBW25 and 1s^4, and dimorphic colonies of 1w^4 grown on standard KB agar (48 hours). (**B**): Morphology of 1w^4 colonies grown on LB (48 hours), M9 (72 hours) and LBS (LB+5 % sucrose; 48 hours) agars. (**C**): 48-hour 1w^4 colonies grown on KB agar made with the following variations in standard tryptone and agar brands: DIFCO Proteose peptone No.3 (left), DIFCO BactoTM Peptone (middle), Sigma tryptone T-9410 and Sigma Bacteriological agar A-5306 (right). Scale bar applies to all nine photographs, and indicates approximately 10 mm. Contrast and/or brightness of some images was adjusted in iPhoto.

Clearly, at least one component of the media significantly influenced the translucent-opaque colony phenotype. This revelation was explored further by alteration of

individual components of KB agar. Given the slow-growing phenotype observed on M9, attention was focussed on tryptone, the main source of amino acids in KB (small differences in preparation of tryptone brands give differences in the final amino acid composition). As above, 1w^4 overnight cultures were produced, washed, diluted and plated on a series of KB agars, each containing a different brand of tryptone and/or agar. Following 48 hours' incubation, examination of colonies revealed that colony phenotype was entirely dependent on the brand of tryptone and, to a lesser extent, agar (Figure 3.2C).

As part of a routine examination, 1w^4 growth was also examined on a range of indicator agars, including nutrient, horse blood (indicator of haemolysis) and MacConkey (indicator of pH) agars. Overnight cultures of SBW25, LSWS and 1w^4 were produced in KB. One ml of each culture was pelleted and re-suspended in 1 ml Ringer's solution. The translucent and opaque fractions of 1w^4 were separated *via* the serendipitously discovered technique of centrifugation, and re-suspended in 850 and 150 µl of Ringer's solution, respectively. Ten µl aliquots of SBW25, LSWS, 1w^4-translucent and 1w^4-opaque were dropped onto each of the agar types. Once dry, the plates were incubated at 28˚C for 48 hours, before visual examination and photographing. An interesting phenotype was observed on MacConkey agar, which contains neutral red, a dye that stains colonies red if the pH drops below 6.8. On this agar, SBW25, LSWS and 1w^4-translucent appeared off-white, indicating a pH above 6.8, while 1w^4-opaque colonies appeared red, indicating a pH below 6.8 (Figure 3.3).

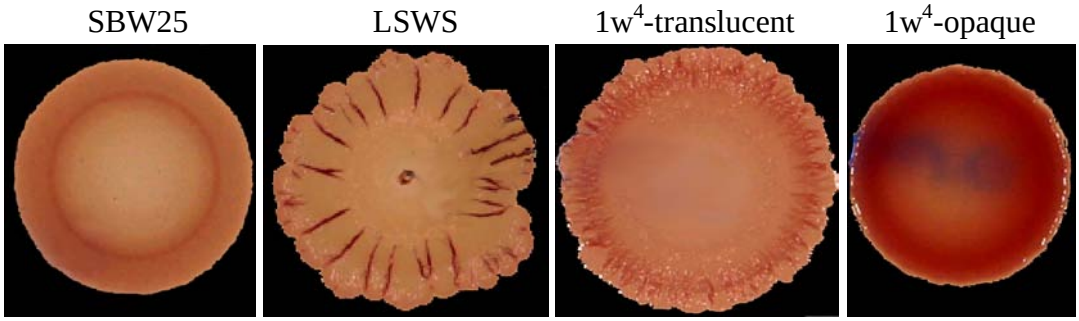


Figure 3.3: SBW25, LSWS and 1w^4 spot colonies on MacConkey agar. After 48 hours' incubation, SBW25, LSWS and the translucent fraction of 1w^4 (1w^4-translucent) did not significantly lower the pH of MacConkey agar. The opaque fraction of 1w^4 (1w^4-opaque) lowered the pH of the agar. Similar adjustments to the contrast and background colour of each photograph were made in iPhoto.

3.3.2.2 Polymer biosynthesis by 1w^4

Previous experience suggested that the capsule and calcofluor binding phenotypes of 1w^4 resulted from the production of two distinct polymers – a previously unidentified capsule polymer, and ACP. To test this, the calcofluor binding ability of 1w^4 fractions enriched for translucent and opaque cells was analysed compared to positive (LSWS) and negative (SBW25-Δ*wss*Km) ACP-producing controls. Overnight cultures of SBW25-Δ*wss*Km, LSWS, 1s^4 and 1w^4 were produced. One ml of the 1w^4 culture was fractionated by centrifugation. Subsequently, SBW25-Δ*wss*Km, LSWS, 1s^4, 1w^4-translucent, 1w^4-opaque and 1w^4 cells were grown on KB agar containing calcofluor, as detailed in section 2.2.11.3.2. Samples from each preparation were subsequently viewed and photographed under a fluorescence microscope (Figure 3.4).

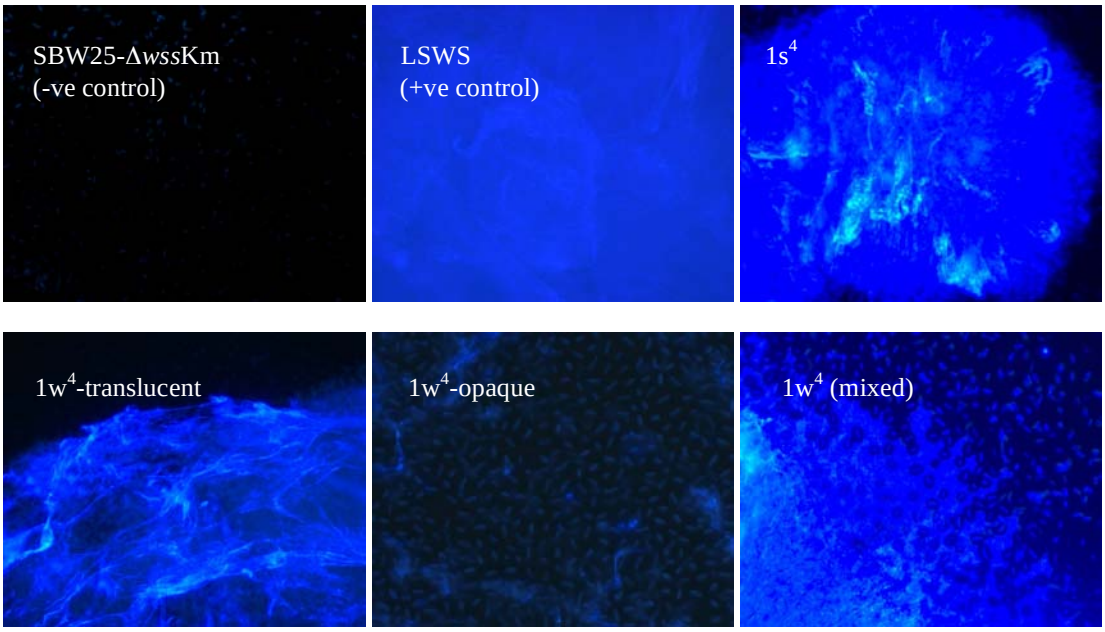


Figure 3.4: Polymer biosynthesis by 1w^4. Fluorescence microscope images showing calcofluor binding by (left to right): SBW25-Δ*wss*Km (no ACP), LSWS (constitutive ACP), 1s^4, 1w^4 –translucent (enriched for non-capsulated cells), 1w^4-opaque (enriched for capsulated cells) and 1w^4 (mixed non-capsulated and capsulated cells). Images taken at magnifications of x40 (top row) or x100 (bottom row).

Observation of samples taken from the translucent fraction revealed ample amounts of calcofluor binding material, while samples from the opaque fraction showed little calcofluor binding. Cells from the opaque fraction appear evenly spaced, as one might

expect if individual cells were separated by a physical, non-staining capsule barrier. Additionally, observations of the mixed fraction showed areas of calcofluor binding consisting almost exclusively of non-capsulated cells. Capsulated cells found at the edges of these areas (presumably these cells became caught up in the sticky ACP polymer) are clearly visible due to the absence of calcofluor binding in the surrounding capsule. Together, these observations demonstrate that 1w^4 synthesizes two distinct polymers: an acidic capsule polymer and ACP.

3.3.2.2.1 Capsule biosynthesis by 1w^4

Like colony phenotype, the capsule phenotype of 1w^4 was found to be sensitive to environmental conditions. A number of environmental factors were found to influence capsule expression, including oxygen availability (shaking speed, tightness of microcosm lids) and incubation time. Notably, capsule biosynthesis was exquisitely sensitive to temperature; temperatures above 28˚C completely repressed capsule production, while temperatures below 28˚C stimulated capsule biosynthesis. Furthermore, growth at 15˚C significantly increased capsule expression in not only 1w^4, but also SBW25 and 1s^4. Media was found to play an important role, with capsule production stimulated in LBS and M9 and reduced in LB, as compared to KB. Conversely, capsule production did not appear to be significantly influenced by osmotic stress induced with increasing levels of NaCl, K_2HPO_4, $MgSO_4$ or $CaCl_2$.

In light of the sensitive nature of capsule production, a specific assay was devised to estimate the relative proportions of capsulated cells in any given population. Described in detail in section 2.2.11.4, the assay uses light microscope photographs of India ink-stained cells from population of interest. From these photographs, the capsule state (+/-) is recorded for 2,500 cells from each population (five replicates, 500 cells *per* replicate), the mean and standard error of the proportion of capsulated cells (between zero (no capsulated cells) and one (all cells capsulated)) in the population is estimated. Subsequently, statistical tests can be used to compare the estimated mean proportions of different populations.

The above assay was used to investigate relative proportions of capsulated cells in the different colony phases of 1w^4. Initially, SBW25, 1s^4 and 1w^4 colonies were grown from glycerol stocks on KB agar. After 48 hours, five individual colonies of each type (SBW25, 1s^4, 1w^4-translucent and 1w^4-opaque) were selected for analysis. In order to maximize potential differences in cell composition, care was taken to select 1w^4 colonies that were either entirely translucent or opaque (rather than sectored). Cells from each of the 20 colonies were assayed. The results of the assay are presented in Table 3.1 and Figure 3.5 (see Appendix A1.1). Importantly, both SBW25 and 1s^4 colonies contained a low proportion of capsulated cells (0.004 for each), demonstrating that these genotypes are also capable of capsule production. Compared to SBW25 and 1s^4 colonies, both translucent and opaque 1w^4 colony variants contained a significantly greater proportion of capsulated cells (P<0.01 and 0.001, respectively). Furthermore, the proportion of capsulated cells in 1w^4-opaque colonies was significantly higher than that found in 1w^4-translucent colonies (P<0.001). Together, these results demonstrate a correlation between capsule production at the cellular level and colony opacity.

Genotype	**Mean ± SE**	**Comparison of population means**		
		Genotypes for comparison	***P*-value**	**95 % CI**
SBW25	0.004 ± 0.00167	-	-	-
1s^4	0.004 ± 0.00126	1s^4 vs. SBW25	1.00	-
1w^4-translucent	0.118 ± 0.0227	1w^4-translucent vs. 1s^4	7.26 x 10^{-3}**	0.0511, 0.177
1w^4-opaque	0.692 ± 0.0575	1w^4-opaque vs. 1s^4	2.79 x 10^{-4}***	0.528, 0.848
		1w^4-opaque vs. –translucent	1.48 x 10^{-5}***	0.431, 0.717

Table 3.1: Relative proportion of capsulated cells in colonies of SBW25, 1s^4, and 1w^4. Mean and standard error (SE) of five replicates are given for each genotype. *P*-values generated from two-sample *t*-tests were used to compare the indicated population means (vs.=versus). Where appropriate, an estimation of the size of the difference is given as a 95 % confidence interval (C.I.). Mean, standard error and *P*-values given to three significant figures.

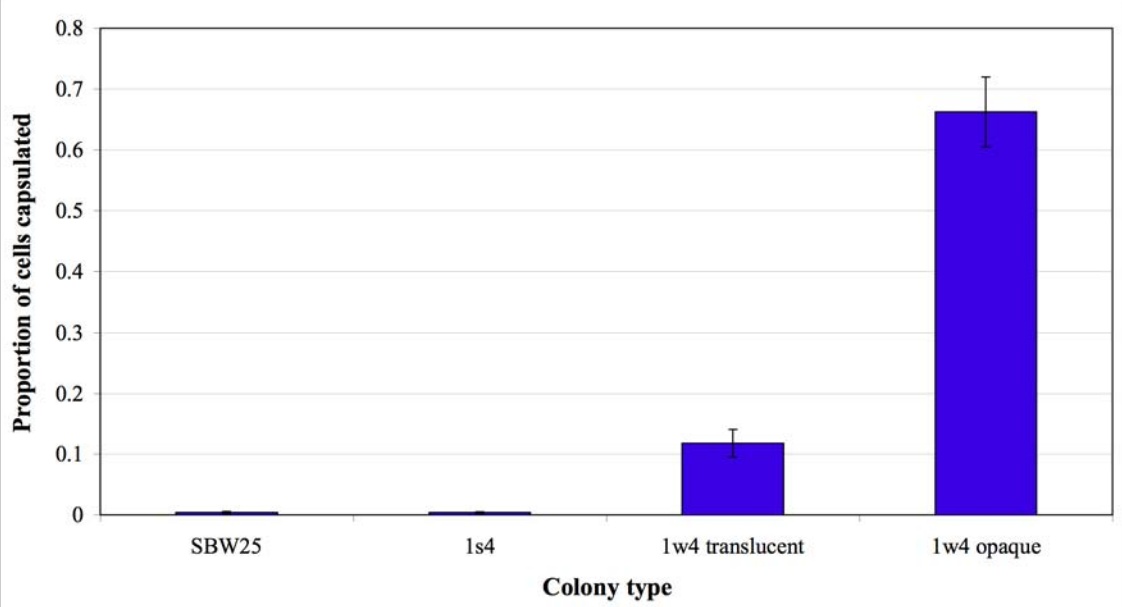


Figure 3.5: Relative proportion of capsulated cells in SBW25, 1s^4, 1w^4-translucent and 1w^4-opaque colonies. Each bar represents the mean of five replicates, and error bars indicate one standard error.

3.3.2.2.2 ACP production in 1w^4

Since ACP is secreted into the extracellular medium, it is not possible to identify which cells in a population produce ACP simply by observing its presence. Thus, at this point it remains unknown whether both capsulated and non-capsulated cells synthesize the ACP observed in 1w^4. Fortunately, the considerable prior knowledge of ACP allowed its involvement in the switcher phenotype to be tested directly; the ACP-biosynthetic *wss* operon was removed from 1w^4, and the phenotypic effects observed. Firstly, SOE-PCR was performed according to the method outlined in section 2.2.2.2. Two separate DNA fragments of 750 bp each were amplified using primers pairs WssKO-1b/-2b and WssKO-3b/-4b (59˚C annealing temperature, 1 minute extension time). These were then combined and used as a template for the second round of PCR that amplified the 1.5 kb deletion fragment using primers WssKO-1b/-4b (59˚C annealing temperature, 1.5 minutes extension time). The deletion fragment was cloned into pCR8/GW/TOPO for sequencing, and a clone harbouring a mutation-free deletion fragment was selected to replace the 1w^4 *wss* locus by enriched two-step allelic exchange (see section 2.2.7). The resulting construct was 1w^4-Δ*wss*, a 1w^4-derived genotype in which the *wss* operon was removed.

Subsequent phenotypic characterization of 1w^4-Δ*wss* included examination of calcofluor binding ability, colony and cell morphology (Figure 3.6). 1w^4-Δ*wss* was no longer capable of binding calcofluor, demonstrating that the calcofluor binding material produced by 1w^4 was indeed *wss*-encoded ACP. Like 1w^4, 1w^4-Δ*wss* produced two types of colonies: an 'opaque' type similar to the opaque type of 1w^4, and a 'smooth' type that resembled smooth colonies of SBW25. At the cell level, 1w^4-Δ*wss* continued to produce both capsulated and non-capsulated cells.

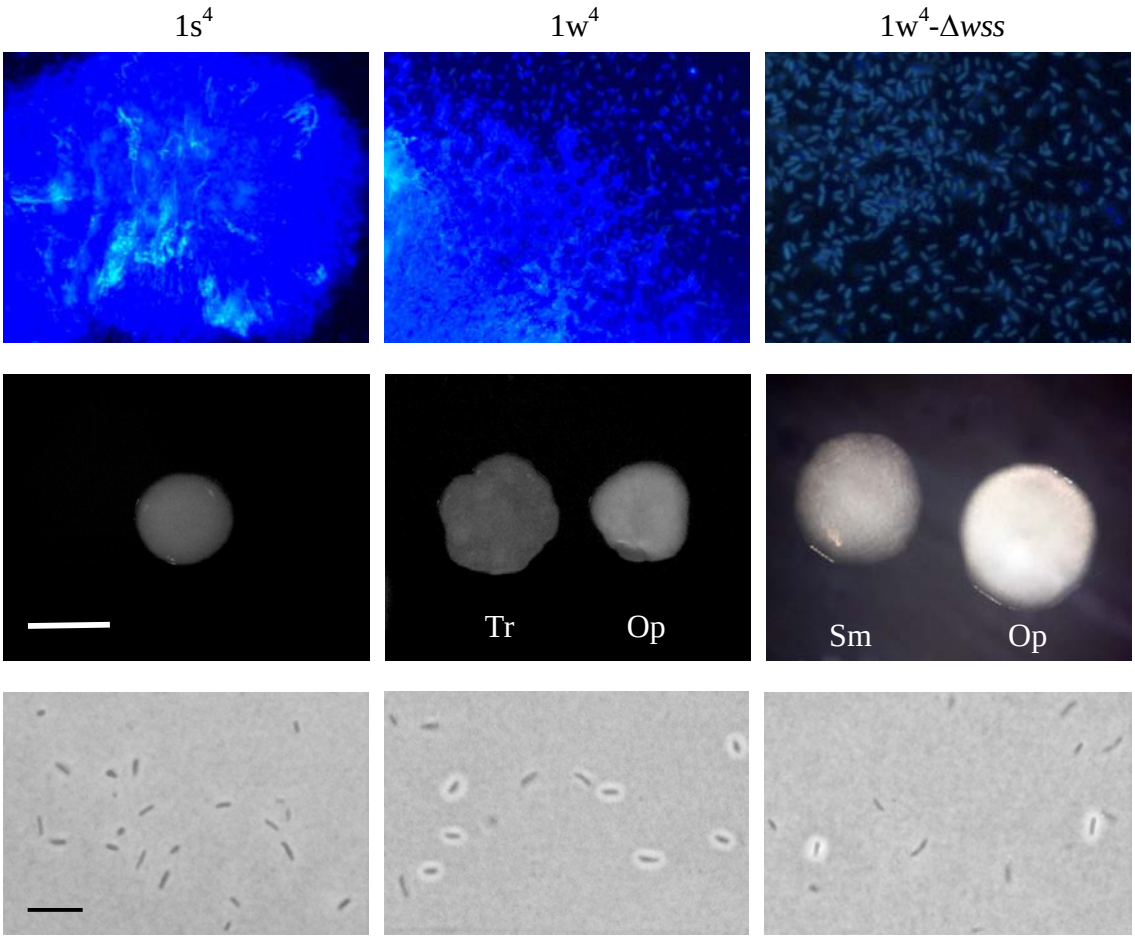


Figure 3.6: Phenotypic characterisation of 1w^4-Δ*wss*, a genotype constructed by removal of the *wss* operon from 1w^4. Deletion of *wss* removes the ability of 1w^4 to bind calcofluor (top row; fluorescence microscope images taken at x40 or x100 magnification), and alters 1w^4 colony morphology (middle row, scale bar indicates approximately 5 mm). Contrastingly, deletion of *wss* does not remove the ability to produce both capsulated and non-capsulated cells (bottom row, scale bar indicates approximately 10 μm).

To assess whether ACP removal had a quantitative effect on capsule expression, a capsule counting assay was performed using five replicate overnight cultures of each of

1s^4, 1w^4 and 1w^4-Δ*wss* (see section 2.2.11.4). Presented in Table 3.2, the results of this assay demonstrate that deletion of the *wss* locus had no quantitative effect on 1w^4 capsule expression (*P*=0.723; Appendix A1.2).

Genotype	Mean ± SE	Comparison of population means		
		Genotypes for comparison	*P*-value	95 % CI
1s^4	0.0016 ± 0.000748	-	-	-
1w^4	0.0844 ± 0.00643	1w^4 vs. 1s^4	1.32 x 10^{-6}***	0.0677, 0.0977
1w^4-Δ*wss*	0.0880 ± 0.00662	1w^4-Δ*wss* vs. 1w^4	0.723	-

Table 3.2: Relative proportion of capsulated cells in cultures of 1s^4, 1w^4 and 1w^4-Δ*wss*. Mean proportion of cells capsulated and standard error (SE) of five replicates are given for each genotype. *P*-values generated from two-sample *t*-tests were used to compare the indicated population means (vs.=versus). Where appropriate, an estimation of the size of the difference is given as a 95 % confidence interval (C.I.). Mean, standard error and *P*-values given to three significant figures.

3.3.3 Statistical demonstration of 1w^4 phenotypic switching

Phenotypic capsule switching may emerge in 1w^4 populations as the result of two mutually exclusive possibilities: (a) uni-directional phenotypic switching (either translucent produces opaque OR opaque produces translucent), or (b) bi-directional phenotype switching (either phase gives rise to either phase). In order to distinguish between these hypotheses, an experiment was undertaken to investigate switching in populations founded by a single cell of either phase. If some populations contained cells of only one type, this would support hypothesis (a). Alternatively, hypothesis (b) would be supported if all populations contained a mixture of cell types.

1w^4 cells were streaked from freezer stocks onto KB agar. After 48 hours growth, a single colony was used to inoculate a KB microcosm. The colony chosen was opaque, as it was known from prior experience that following overnight incubation in a shaking

microcosm, the population would contain approximately equal proportions of non-capsulated and capsulated cells. The inoculated microcosm was incubated for 16 hours (160 rpm) prior to thorough vortexing and microscopic examination to ensure complete separation of cells. Following the method in section 2.2.11.4, the proportion of capsulated cells in the culture was estimated as 0.479 ± 0.0104 (mean ± standard error of five replicate culture samples; see Appendix A1.3).

95 µl of fresh KB was measured into each well of a 96 well plate. A sample of the $1w^4$ overnight culture was diluted by a factor of 10^7 in fresh KB, and 5 µl of this dilution was used to inoculate each well. On the basis that a 16-hour overnight KB culture contains approximately 10^9 cells *per* ml, each well received a theoretical average of 0.5 cells. Assuming a Poisson distribution, it was possible to calculate the probability that a particular well would be inoculated by a particular number of cells (Equation 3.1, where: *P*=probability of *y* cells founding any given well, *y*=number of cells of interest, λ =the expected number of cells *per* well and *e*=base of the natural logarithm). For example, given that the average number of cells founding each population was 0.5, the probability of a well receiving exactly zero cells was 0.606 (Equation 3.2). Furthermore, if there were 96 wells each with a 0.606 chance of receiving zero cells, no bacterial growth was expected in 58 wells (Equation 3.3). The probability of a well being inoculated by increasing numbers of cells and the corresponding expected numbers of wells founded is listed in Table 3.3.

$$P(y) = \frac{\lambda^y \times e^{-\lambda}}{y!} \qquad \textbf{Equation 3.1}$$

$$P(0) = \frac{0.5^0 \times e^{-0.5}}{0!} = 0.606 \quad (3s.f.) \qquad \textbf{Equation 3.2}$$

$$\textit{Expected number of wells} = P(y) \times 96 \qquad \textbf{Equation 3.3}$$

Number of cells	Probability of occurrence[a]	Expected number of wells[b]	
		1 d.p.[c]	Rounded[d]
0	0.606	58.2	58
1	0.303	29.1	29
2	0.0758	7.3	7
3	0.0126	1.2	1
4+	0.00260	0.2	0
Totals	1.00	96	95

Table 3.3: Poisson probabilities of population foundation by increasing numbers of cells, and corresponding expected numbers of each in 96 independent populations. [a]Probability of the specified number of cells founding any given population (to three significant figures). [b]Corresponding expected number of 96 independent populations founded by specified number of cells [c]to 1 d.p. (decimal place; giving a total of 96 wells) and [d] rounded to the nearest number of wells (giving a total of 95 wells only).

After incubation for 24 hours (180 rpm), 40 wells with and 56 wells without bacterial growth were observed. Since this was in good agreement with the expected values of 38 (one or more cells) and 58 (zero cells), the estimation of each well receiving 0.5 cells could reasonably be used in further calculations. Cell samples from the founded populations were stained with India ink and capsule expression examined for each. All 40 populations were composed of a mixture of capsulated and non-capsulated cell types, and produced a mixture of translucent and opaque colonies on KB agar. As the initial inoculum contained approximately equal proportions of capsulated and non-capsulated cells, the probability that each cell type founded at least one of the 29 populations inoculated with a single cell was essentially one (Equations 3.4 and 3.5, where *P(either)*=probability that all populations were founded by a single cell type, and *P(both)*=probability that each cell type founded at least one population). These results demonstrate that 1w^4 capsule expression switches ON and OFF reversibly.

$$P(either) = 2 \times 0.5^{29} = 4.00 \times 10^{-9} \quad (3.s.f.)$$ **Equation 3.4**

$$P(both) = 1 - (2 \times 0.5^{29}) = 0.999999996$$ **Equation 3.5**

3.4 Discussion

3.4.1 Summary of the 1w^4 phenotype

The experiments in this chapter revealed several phenotypic differences between SBW25 and 1w^4 (Table 3.4). In particular, 1w^4 cells were dimorphic, and were shown to differentially synthesize at least two polymers: an unidentified, acidic polymer that forms a thick capsule around the cell, and ACP (see section 1.2.1.4). Correlations between mutually exclusive expression of these polymers and colony morphology were observed; capsule production correlates with colony opacity, while ACP production was revealed as the major phenotypic innovation underlying translucent colony morphology.

Phenotype	**Sub-type**	
	Translucent	**Opaque**
Dimorphism at the colony (population) level		
Edges	Wrinkly	Round
Texture	Rough	Smooth
Colour	Translucent	White
Dimorphism at the cell (individual) level		
Acidic capsule polymer	×	✓
ACP	✓	×
Motility	✓	×

Table 3.4: Phenotypic characteristics of 1w^4 translucent and opaque sub-types at the population and individual level. Colony-level phenotypes are those observed on KB agar at 48 hours. Cell-level phenotypes are those observed in overnight KB cultures. ✓=present, ×=absent.

The degree of 1w^4 dimorphism proved sensitive to a range of environmental stimuli, especially temperature and media composition. It is possible that this was due to alteration of the switch rate in response to specific environmental conditions – a phenomenon that has been observed for both genetic (section 1.1.1) and epigenetic switches (Alby & Bennett, 2009). Alternatively, the observed sensitivity may be attributable to differential fitness of 1w^4 subtypes in individual environments. Either way, sensitivity to environmental factors is not a reflection of the switch mechanism.

3.4.1.1 Insight into the identity of the capsule polymer

Bi-directional switching of capsule expression is an emergent property of 1w^4; high-level capsule production has not previously been observed in *P. fluorescens* SBW25. As such, the structural basis of the capsules is limited to three insights provided from the work in this chapter: (1) the polymer cannot be pelleted by centrifugation at 13,000 rpm, (2) it does not bind calcofluor or stain with India ink, and (3) it appears to be acidic. While similar capsules have not been described in *Pseudomonas*, a number of other enteric bacteria including *Escherichia*, *Salmonella*, *Acetobacter* and *Klebsiella* species are known to produce capsules with these properties (Grant *et al*., 1969). Reviewed by Corbett & Roberts (2008), acidic capsules are diverse structures thought to play a role in protecting cells from environmental stresses such as temperature changes and nutrient availability.

Capsulation has been especially well studied in uropathogenic *E. coli* strains that produce a particularly diverse range of capsule polymers (reviewed in Whitfield, 2006). Briefly, *E. coli* capsule polymers are classified into five groups on the basis of genetic, biochemical and synthetic pathway similarities. Groups 1-4 include polymers tightly associated with the cell surface, and are referred to as K-antigens. The fifth group contains capsules based on a polymer called colanic acid (also known as M-antigen). Like K-antigens, colanic acid forms a discrete capsule, but a large amount of colanic acid is secreted into the surrounding medium. Interestingly, phase-variable expression of capsules has been documented in *E. coli*; differential expression of K1 capsules has been reported to occur *via* the ON/OFF expression of a mobile contingency locus (Vimr & Steenbergen, 2006; King *et al*., 2007).

3.4.1.2 The role of ACP

The work in sections 3.3.1 and 3.3.2.2 illustrated that the production of ACP was an integral part of the phenotypic history of 1w^4. It seems likely that, early in the evolutionary line, the required phenotypic innovation was achieved through a series of mutations that alternately activated and deactivated production of ACP. On the eighth reversal (during which 1s^4, the immediate ancestor of 1w^4, evolved) this trend was

discontinued, and all genotypes thereafter produced ACP. It follows that, as a result of history, ACP was produced by 1w^4; removal of ACP from 1w^4 caused the translucent/opaque colony phenotype to change to smooth/opaque. Together with the differential calcofluor binding of the 1w^4 translucent and opaque fractions, these results indicate that ACP is primarily expressed by non-capsulated cells, and causes the WS-like translucent colony phenotype. One might expect the 1w^4 translucent colonies to resemble those of 1s^4, given that neither express capsules, and both produce ACP. Interestingly, the translucent colony phenotype does not resemble that of 1s^4, suggesting a difference in the quantity of ACP production between 1s^4 and the 1w^4 cap- fraction. Finally, ACP removal does not alter biphasic nature of either colonies or cells, demonstrating that unlike capsule production, ACP expression was not required for either cell or colony level dimorphism.

3.4.2 Estimation of the rate of switching in 1w^4

In order for phenotypic switching to be considered a switching mechanism, a high rate of switching is required. However, measurement of switching rates is wrought with difficulties, and meaningful estimates are rare (Saunders *et al*., 2003). As yet, there are no estimates of switch rates in 1w^4. In order to measure 1w^4 switch rates, three main obstacles that must be overcome. Firstly, the high rate of switching renders population level assays (such as colony morphology) uninformative; colony morphology cannot be used to accurately reflect cell type ratios. Secondly, it is likely that the 1w^4 subtypes have differential fitness, and thus different growth rates, meaning that changes in subtype proportions are the combined result of switch events and subsequent selection. In order to gain insight into switch rates the relative contribution of these factors must be untangled. Finally, it is highly likely that the switch rate is dependent on environmental conditions. Thus, any rate calculated may only be relevant in a very defined set of conditions. Despite these difficulties, it would be useful to have an estimate of switch rates in the environments of the REE. To this end, the above issues are currently being addressed through the use of microscopic technology that will allow observation of individual cells and estimations of relative fitness in specific environments (G. Ferguson, personal communication).

3.4.3 The capsule phenotype of SBW25 and 1s^4

Despite the fact that colony dimorphism is an emergent property of 1w^4, capsulated cells were occasionally observed in ancestral genotypes. Due to the very low frequency at which these capsulated ancestral cells were observed (0.004), it was unclear whether they were the result of standard capsule-inducing mutations, or whether they were produced *via* the same molecular mechanism as those in 1w^4. However, the observed temperature sensitivity of the capsule switching phenotype in both 1w^4 and its ancestors suggested that the switching mechanism is active at a low rate in ancestral forms. Presumably, the switching rate is elevated by one or more mutations in the evolutionary history of 1w^4. Importantly, this elevated transition rate is required for the capsule switching phenotype to be considered a phenotypic switching mechanism, as these mechanisms require a rate of switching of >10^{-5} *per* locus *per* generation (see sections 1.1.1 and 1.1.2.1).

It is important to note that **if** the 1w^4 phenotype is the result of elevation (rather than appearance) of capsule switching, this does not provide any indication of the molecular mechanism underlying switching; it is conceivable that elevation of switch rate could occur as the result of either a genetic or epigenetic mechanism. For example, it is possible that an imperfect repeat in the ancestral forms generates low-level (and therefore non-phase-variable) capsule switching *via* SSM (see section 1.1.1.2). Subsequently, mutation(s) in the 1w^4 evolutionary history may generate a perfect repeat, thus increasing SSM and associated switch rate. Alternatively, an epigenetic feedback loop may exist in the capsule biosynthetic pathway of SBW25 (and derived genotypes). Mutations the 1w^4 evolutionary history may alter the degree of feedback, creating the necessary conditions for observation of capsule bistability (see section 1.1.2.1).

Chapter 4:
Transposon Mutagenesis of 1w^4

4.1 Introduction

Transposon mutagenesis is a powerful tool for analyzing the genetic bases of phenotypes. Using this technique, a library of transposon mutants can be created where each member contains a single transposon insertion that alters the phenotype of interest. Identification of the transposon insertion sites provides insight into the gene(s) required for manifestation of the phenotype. The success of a transposon mutagenesis screen is critically dependent on two factors: (1) the presence of a reliable, easily detectable biological phenotype of interest, and (2) the choice of transposon. The aim of this chapter was to create a transposon mutant library of 1w^4, where each member contained an insertion in a gene required for manifestation of colony bistability. The screening assay made use of the sectored colony phenotype on LB agar (see section 3.3.2.1); large numbers of transposon mutants were screened for loss of colony sectoring.

The transposon selected for the screen was IS-Ω-Km/hah, a highly versatile transposon with several useful attributes (Figure 4.1; Giddens *et al*., 2007):

1. IS-Ω-Km/hah encodes a kanamycin resistance cassette, allowing efficient selection of cells that have integrated the transposon into the genome.

2. Primers specific to the O-end of the transposon have been synthesised (Manoil, 2000), enabling PCR-mediated amplification and sequencing of transposon-chromosome junctions in mutants of interest.

3. Depending on the position and orientation, insertion of IS-Ω-Km/hah can disrupt or activate (through the neomycin phosphotransferase (*nptII*) promoter) gene transcription, allowing concurrent identification of genes that promote the phenotype and genes that repress the phenotype (Jacobs *et al*., 2003).

4. IS-Ω-Km/hah contains two *loxP* sites, enabling pCre-mediated excision of the transposon. This leaves a 189 bp, non-polar insertion. This allows assessment of the involvement of downstream genes (i.e. polar effects) in phenotypic effects.

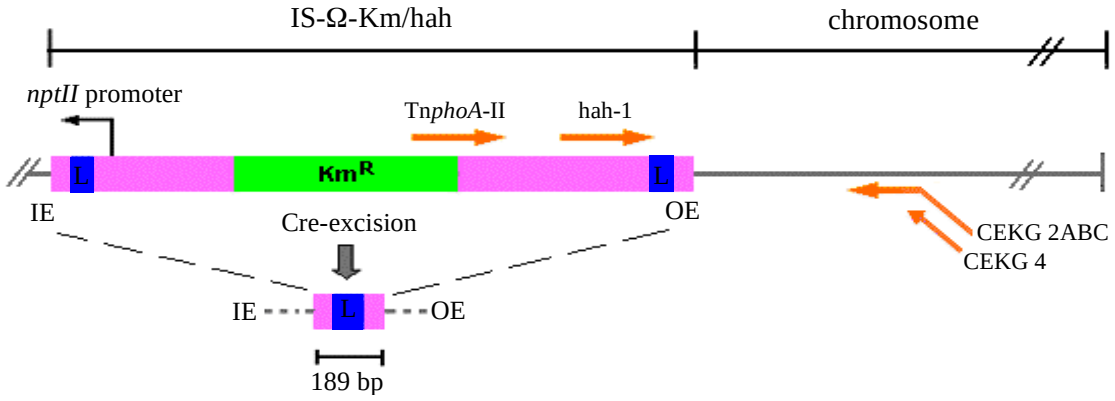


Figure 4.1: Structure and attributes of IS-Ω-Km/hah (Giddens *et al.*, 2007). A combined derivative of IS-*phoA*/hah (Bailey & Manoil, 2002) and pJFF350 (Fellay *et al*., 1989), the IS-Ω-Km/hah transposon has an I-end (IE) and an O-end (OE), which enclose a Km resistance cassette (Km^R) for selection purposes, an *npt* promoter to allow gene activation, and two *loxP* sites (L) to enable Cre recombinase-mediated transposon excision. The O-end (OE) and adjacent chromosomal sequence may be amplified using the indicated primers (orange arrows) in a two-step PCR reaction as described in section 2.2.1.3 (Manoil, 2000). Figure adapted from Gallie (2005).

In this chapter, a reliable, easily detectable phenotype (colony sectoring on LB agar) and a versatile transposon (IS-Ω-Km/hah) were combined to determine the genetic basis of the 1w^4 switching phenotype. The work presented provides considerable insight into the structural bases of the dimorphic capsule phenotype discovered in Chapter 3 and, to a lesser degree, the molecular basis of the switch mechanism.

4.2 Aims

The aim of this chapter is to identify genes required for the dimorphic colony phenotype of 1w^4. This will be achieved using a comprehensive, random mutagenesis screen to isolate and characterize 1w^4 transposon mutants defective in colony switching.

4.3 Results

4.3.1 Transposon mutagenesis of 1w^4

Transposon mutagenesis techniques were employed as outlined in section 2.2.8.1. Transposon mutants were plated on LB+NF+Km medium, upon which 1w^4 forms translucent colonies with distinctive opaque sectors. 68,750 colonies from 41 independent conjugations were screened for loss of sectoring (Beaumont *et al*., 2009). Assuming a Poisson distribution and a genome of 6,000 genes (Silby *et al*., 2009), the probability of not screening an inactivational insertion of any given gene is approximately 1.06 x 10^{-5}.

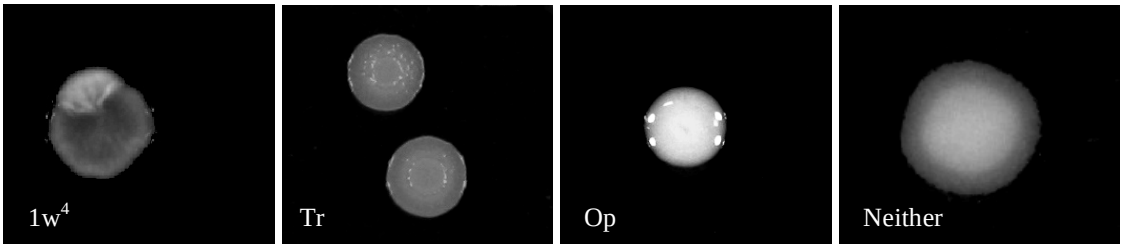


Figure 4.2: The sectoring phenotype of 1w^4 was used to isolate non-sectoring transposon mutant derivatives. Dissection microscope images of (left to right): 1w^4 and derived transposon mutants stuck in the translucent state (Tr), the opaque state (Op), and a state that resembles neither 1w^4 state (Neither). Colonies grown on LB+NF+Km agar for 48 hours. Contrasts and background colours altered in iPhoto.

A total of 183 non-sectoring transposon mutants were obtained (JG1-183). Growth rate, cell morphology and point of insertion were examined for each mutant (see sections 2.2.2.3 and 2.2.5.1). Growth rate was examined by quantifying the growth of each mutant in overnight KB cultures, compared to 1w^4. 55 mutants gave an OD_{600} that was less than 50 % of that of 1w^4 (see Appendix A2.1). It is likely that these mutants exhibit delayed colony-level switching due to slow growth, and so these were discarded from further analyses. The remaining 128 mutants are listed below in Table 4.1. 76 (~60 %) contained insertions in genes directly involved in the biosynthesis of a capsule polymer resembling colanic acid previously described in enteric bacteria. A further 50 either reduce (44, ~34 %) or increase (6, ~5 %) capsule biosynthesis indirectly. The point of insertion in the final two mutants remains unidentified.

Strain	C#[a]	Pflu[b]	Gene[c]	Predicted protein function	Insertion[d]	Cap[e]
Insertions in the colanic acid biosynthetic operon						
JG64	17	3654	-	Two component transcriptional regulator	4047986	-1
JG91	21	3654	-	Two component transcriptional regulator	4048055	-1
JG108	22	3654	-	Two component transcriptional regulator	4048054	-1
JG110	23	3654	-	Two component transcriptional regulator	U	-1
JG133	25	3654	-	Two component transcriptional regulator	4047902	-1
JG149	32	3654	-	Two component transcriptional regulator	4047883	-1
JG157	36	3654	-	Two component transcriptional regulator	4047883	-1
JG183	40	3654	-	Two component transcriptional regulator	4047840	-1
JG54	12	1G	P3655	Initiation of Pflu3655-7 transcription	4048249	-2
JG137	27	IG	P3655	Initiation of Pflu3655-7 transcription	4048176	-2
JG26	6	3656	-	Unknown; hypothetical protein with EAL	4049814	-2
JG39	8	3656	-	Unknown; hypothetical protein with EAL	4049900	-2
JG90	21	3656	-	Unknown; hypothetical protein with EAL	4049383	-2
JG168	38	3656	-	Unknown; hypothetical protein with EAL	4049594	-2
JG5	1	3658	*wcaJ*	UDP-glucose lipid carrier transferase	4050865	-2
JG32	8	3658	*wcaJ*	UDP-glucose lipid carrier transferase	U	-2
JG44	10	3658	*wcaJ*	UDP-glucose lipid carrier transferase	4050904	-2
JG51	11	3659	-	Putative lipoprotein	4054294	-2
JG28	6	3662	*wza*	Capsular polysaccharide translocation	4056427	-2
JG111	23	3662	*wza*	Capsular polysaccharide translocation	U	-2
JG21	3	3663	-	Putative membrane protein	4057344	-2
JG7	11	3666	*wcaI*	Glycosyltransferase	4060642	-2
JG124	24	3669	*waaE*	Heptose kinase/adenyltransferase	4063722	-2
JG53	12	3670	*wcaF*	Acetyltransferase	4064077	-2
JG35	8	3671	-	Glycosyltransferase	4064377	-2
JG77	19	3673	-	Glycosyltransferase	U	-2
JG120	23	3673	-	Glycosyltransferase	4066885	-2
JG150	33	3673	-	Glycosyltransferase	4067727	-2

Strain	C#[a]	Pflu[b]	Gene[c]	Predicted protein function	Insertion[d]	Cap[e]
JG152	33	3674	-	Acetyltransferase	4068319	-2
JG18	3	3675	*wzx*	O-antigen flippase	4069688	-2
JG103*	22	3676	*udg*	UDP-glucose-6-dehyrogenase	4070989	-2
JG1	1	3677	*wzc*	Tyrosine protein kinase	U	-2
JG27	6	3677	*wzc*	Tyrosine protein kinase	4071958	-2
Insertions in genes involved in biosynthesis of colanic acid precursors						
JG62	16	1668	*capD*	UDP-glucose 4-epimerase	1831573	-1
JG31	7	2985	*galU*	UTP–glucose-1-phosphate uridylyltransferase	U	-2
JG49	11	2985	*galU*	UTP–glucose-1-phosphate uridylyltransferase	3252712	-2
JG52	11	2985	*galU*	UTP–glucose-1-phosphate uridylyltransferase	3252734	-2
JG114	23	2985	*galU*	UTP–glucose-1-phosphate uridylyltransferase	3253166	-2
JG130	28	2985	*galU*	UTP–glucose-1-phosphate uridylyltransferase	U	-2
JG103*	22	3676	*udg*	UDP-glucose-6-dehyrogenase	4070989	-2
JG16	3	5243	*pgi*	Glucose-6-phosphate isomerase	U	-2
JG23	4	5243	*pgi*	Glucose-6-phosphate isomerase	U	-2
JG73	18	5243	*pgi*	Glucose-6-phosphate isomerase	5752171	-2
JG75	18	5243	*pgi*	Glucose-6-phosphate isomerase	U	-2
JG81	20	5243	*pgi*	Glucose-6-phosphate isomerase	U	-2
JG123	24	5243	*pgi*	Glucose-6-phosphate isomerase	U	-2
JG153	34	5243	*pgi*	Glucose-6-phosphate isomerase	U	-2
JG154	34	5243	*pgi*	Glucose-6-phosphate isomerase	U	-2
JG160	37	5243	*pgi*	Glucose-6-phosphate isomerase	U	-2
JG163	38	5243	*pgi*	Glucose-6-phosphate isomerase	U	-2
JG175	40	5243	*pgi*	Glucose-6-phosphate isomerase	U	-2
JG17	3	5986	*algC*	Phosphomanno/glucomutase	U	-2
JG55	12	5986	*algC*	Phosphomanno/glucomutase	6542741	-2
Insertions in regulators of colanic acid biosynthesis						
JG74	18	2189	*gacA*	Two component transcriptional regulator	2372777	-2
JG140	28	2189	*gacA*	Two component transcriptional regulator	2372793	-2

Strain	C#[a]	Pflu[b]	Gene[c]	Predicted protein function	Insertion[d]	Cap[e]
JG9	1	3777	*barA*	Two component sensor histidine kinase	4174903	-2
JG10	1	3777	*barA*	Two component sensor histidine kinase	4173951	-2
JG22	3	3777	*barA*	Two component sensor histidine kinase	U	-2
JG40	9	3777	*barA*	Two component sensor histidine kinase	U	-2
JG45	10	3777	*barA*	Two component sensor histidine kinase	U	-2
JG56	14	3777	*barA*	Two component sensor histidine kinase	U	-2
JG67	17	3777	*barA*	Two component sensor histidine kinase	U	-2
JG68	17	3777	*barA*	Two component sensor histidine kinase	U	-2
JG70	17	3777	*barA*	Two component sensor histidine kinase	U	-2
JG106	22	3777	*barA*	Two component sensor histidine kinase	U	-2
JG119	23	3777	*barA*	Two component sensor histidine kinase	4174319	-2
JG138	27	3777	*barA*	Two component sensor histidine kinase	4173233	-2
JG142	28	3777	*barA*	Two component sensor histidine kinase	4173840	-2
JG143	28	3777	*barA*	Two component sensor histidine kinase	U	-2
JG144	29	3777	*barA*	Two component sensor histidine kinase	4174468	-2
JG164	38	3777	*barA*	Two component sensor histidine kinase	4174038	-2
JG166	38	3777	*barA*	Two component sensor histidine kinase	4173899	-2
JG167	38	3777	*barA*	Two component sensor histidine kinase	4173273	-2
JG169	38	3777	*barA*	Two component sensor histidine kinase	4173177	-2
JG173	40	3777	*barA*	Two component sensor histidine kinase	U	-2
JG174	40	3777	*barA*	Two component sensor histidine kinase	4173818	-2
JG177	41	3777	*barA*	Two component sensor histidine kinase	4174881	-2
Insertions in genes that reduce capsule biosynthesis indirectly						
JG48	10	0266	*hslO*	Heat shock response protein 33, chaperonin	U	-2
JG136	27	0266	*hslO*	Heat shock response protein 33, chaperonin	292032	-2
JG139	27	0266	*hslO*	Heat shock response protein 33, chaperonin	U	-2
JG102	22	0335	*ctpA*	Carboxy-terminal protease precursor	U	-1
JG134	25	0478	*wgcD*	Glycosyltransferase	541964	-2
JG161	37	0478	*wgcD*	Glycosyltransferase	U	-2

Strain	C#[a]	Pflu[b]	Gene[c]	Predicted protein function	Insertion[d]	Cap[e]
JG43	10	0816	*prfC*	Peptide chain release factor 3	U	-2
JG135	26	0816	*prfC*	Peptide chain release factor 3	923034	-2
JG57	14	0933	*diaA*	Phosphoheptose isomerase	U	-2
JG165	38	0933	*diaA*	Phosphoheptose isomerase	1036860	-2
JG170	40	0933	*diaA*	Phosphoheptose isomerase	1037323	-2
JG132	25	1304	*dcd*	Deoxycytidine triphosphate deaminase	1441995	-2
JG63	16	1378	*oprM*	Outer membrane efflux protein	U	-2
JG61	16	1553	*fadB*	Fatty oxidation complex, alpha subunit	U	-1
JG145	29	1556	*topA*	DNA topoisomerase I	1706944	-2
JG65	17	1560	*lexA*	Transcriptional repressor of SOS regulon	1709155	-2
JG38	8	IG	*tRNA*	Glu/Gly anticodons TTC/GCC, Pflu1858-9	U	-1
JG46	10	IG	*tRNA*	Glu/Gly anticodons TTC/GCC, Pflu1858-9	2028658	-1
JG109	22	IG	*tRNA*	Glu/Gly anticodons TTC/GCC, Pflu1858-9	U	-1
JG59	15	2720	-	Transmembrane efflux protein	U	-1
JG80	20	4383	*dsbE*	Thiol:disulfide interchange protein	U	-1
JG6	1	4705	*fabG*	3-oxoacyl-[acyl-carrier protein] reductase	U	-2
JG11	1	4713	*rluC*	23S rRNA pseudouridine synthase	5187852	-2
JG41	9	4713	*rluC*	23S rRNA pseudouridine synthase	5187844	-2
JG82	20	4713	*rluC*	23S rRNA pseudouridine synthase	5187277	-2
JG131	25	4713	*rluC*	23S rRNA pseudouridine synthase	5187691	-2
JG141	28	4713	*rluC*	23S rRNA pseudouridine synthase	5187852	-2
JG128	24	4714	*rne*	Ribonuclease E	U	-2
JG147	29	4714	*rne*	Ribonuclease E	5190308	-2
JG2	1	IG	P*algZ*	Alginate biosynthesis transcriptional activator	5217773	-1
JG79	20	4772	-	Conserved hypothetical protein; hydrolase	U	-2
JG47	10	4809	-	Amino transferase	U	-2
JG87	21	4809	-	Amino transferase	U	-2
JG129	24	5055	-	Conserved hypothetical membrane protein	U	-2

Strain	C#[a]	Pflu[b]	Gene[c]	Predicted protein function	Insertion[d]	Cap[e]
JG36	8	5254	*nusA*	Transcription elongation factor	U	-2
JG88	21	5269	*dnaK*	Heat shock response chaperone protein	U	-2
JG4	1	5446	*dedA*	Conserved protein, unknown function	U	-2
JG3	1	5720	*sahA*	S-adenosyl-L-homocysteine hydrolase	U	-2
JG13	2	5720	*sahA*	S-adenosyl-L-homocysteine hydrolase	U	-2
JG72	18	5720	*sahA*	S-adenosyl-L-homocysteine hydrolase	U	-2
JG125	24	5720	*sahA*	S-adenosyl-L-homocysteine hydrolase	U	-2
JG156	36	5720	*sahA*	S-adenosyl-L-homocysteine hydrolase	6271687	-2
JG158	36	5720	*sahA*	S-adenosyl-L-homocysteine hydrolase	6272161	-2
JG181	41	5720	*sahA*	S-adenosyl-L-homocysteine hydrolase	6272159	-2
Insertions in genes that increase capsule biosynthesis indirectly						
JG30	7	3841	-	Exported protein, function unknown	4235937	+1
JG76	19	IG	-	Promoter of *purU* and Pflu4939	5419510	+1
JG113	23	IG	-	Promoter of *purU* and Pflu4939	5419510	+1
JG148	32	4939	-	MvaT-like transcriptional regulator	5419883	+1
JG176	41	5061	*ndk*	Nucleoside diphosphate kinase	5562166	+1
JG178	41	5438	-	Aromatic acid decarboyxylase	5969732	+1
Insertions for which the point of insertion remains unknown						
JG101	22	-	-	-	-	-2
JG171	40	-	-	-	-	-2

Table 4.1: Point of insertion and capsule phenotypes for 128 non-sectoring 1w[4] transposon mutants. [a]Conjugation number during which transconjugant was isolated, [b]Pflu number, the designated number of the gene on the SBW25 chromosome (IG=intergenic), [c]where possible, gene name assigned on basis of BLASTP (dash=previously unnamed gene, P=promoter), [d]precise point of insertion in the 5'→3' direction of the SBW25 chromosome (U=unidentified), [f]capsule phenotype of mutant (-2=no capsules, -1=some capsules but lower proportion than in 1w[4], +1=greater proportion of capsules than 1w[4]). *Marks JG103, which belongs in two categories and so appears in duplicate.

4.3.1.1 Insertions in the colanic acid biosynthetic cluster

The mutagenesis screen identified 33 independent, monomorphic colony mutants with insertions in a ~26 kb cluster of 25 genes: Pflu3654-78 (Table 4.2, Figure 4.3A). This locus resembles the 19-gene *E. coli* K12 colanic acid biosynthetic clusters (Stevenson *et al*., 1996). Although the organization differs, both species have genes for precursor biosynthesis, polymer assembly and export (Figure 4.3B). Capsule expression from the SBW25 locus requires several additional genes, including Pflu3656, Pflu3659, Pflu3660, Pflu3661 and Pflu3663. A regulatory role is predicted for Pflu3656, given the presence of an EAL domain for c-di-GMP breakdown (see section 1.2.1.5). Pflu3659 and Pflu3661 encode putative lipoproteins, while Pflu3660 and Pflu3663 are predicted to encode exported proteins. Transposon insertions among these four genes suggest a possible structural role in colanic acid biosynthesis, indicating that the SBW25 capsule polymer differs from the well-characterized colanic acid of *E. coli* K12.

Protein	**Pflu**[a]	**Ins**[b]	**Size**[c]	**Pfam domains**[d]	***E*-value**[d]	***E. coli* homology**[e]	
						Protein	*%*
Pflu3654	3654	8	439	HAMP HATPase_C	1.8×10^{-9} 2.5×10^{-28}	None	-
Pflu3655	3655	0	243	Trans_reg_C	2.2×10^{-6}	None	-
Pflu3656	3656	4	231	EAL	9.7×10^{-4}	None	-
Pflu3657	3657	0	222	Trans_reg_C	1.2×10^{-11}	None	-
WcaJ	3658	3	471	Bac_transf	8.8×10^{-126}	WcaJ	40
Pflu3659	3659	1	691	DUF940	1.4×10^{-271}	Lipoprotein	42
Pflu3660	3660	0	256	DUF1017	9.3×10^{-23}	Hypothetical	29
Pflu3661	3661	0	217	None	n/a	Lipoprotein	24
Wza	3662	2	370	Poly_export SLBB	8.4×10^{-26} 1.8×10^{-3}	Wza	38
Pflu3663	3663	1	94	None	n/a	None	-
ManC	3664	0	469	NTP_transferase MannoseP_isomer	5.8×10^{-126} 1.2×10^{-128}	ManC/CpsB	60

Protein	**Pflu**[a]	**Ins**[b]	**Size**[c]	**Pfam domains**[d]	***E*-value**[d]	***E. coli* homology**[e]	
						Protein	*%*
WcaH	3665	0	152	NUDIX	8.0×10^{-18}	WcaH/Gmm/ NudD	51
WcaI	3666	1	402	Glycos_transf_1	5.4×10^{-10}	WcaI	61
Fcl	3667	0	325	Epimerase	6.1×10^{-91}	Fcl/WcaG	68
Gmd	3668	0	373	Epimerase	2.6×10^{-107}	Gmd	76
WaaE	3669	1	311	Glycos_transf_2	1.2×10^{-11}	None	-
WcaF	3670	1	192	None	n/a	WcaF	46
Pflu3671	3671	1	381	Glycos_transf_1	1.5×10^{-23}	WfeX	26
Pflu3672	3672	0	440	None	n/a	None	-
Pflu3673	3673	3	410	Glycos_transf_1	5.6×10^{-27}	None	-
Pflu3674	3674	1	187	None	n/a	WcaF	35
Pflu3675	3675	1	433	None	n/a	Wzx	22
Udg	3676	1	445	UDPG_MGDP_dh_N	1.2×10^{-89}	Ugd*	29
				UDPG_MGDP_dh	2.4×10^{-4}		
				UDPG_MGDP_dh_C	3.0×10^{-32}		
Wzc	3677	2	740	Wzz	1.2×10^{-42}	Wzc	42
Wzb	3678	0	146	LMWPc	9.4×10^{-46}	Wzb	49

Table 4.2: Transposon insertions, domain characteristics and similarity to *E. coli* colanic acid biosynthetic proteins. [a]Pflu refers to the numeric name of the SBW25 gene. [b]Ins refers to the number of independent transposon insertions obtained. [c]Number of amino acids. [d]Pfam domain searches were performed using the Sanger Pfam website. [e]Homology to *E. coli* proteins, including % amino acid identity as determined by a BLASTP search (Id.) (*Ugd is not a typing error).

4.3.1.1.1 ACP and capsule expression in colanic acid biosynthetic mutants

The 33 transposon mutants with insertions in 16 of the colanic acid biosynthetic cluster genes produced translucent colonies and, correspondingly, did not express capsules at the cellular level. Accordingly, calcofluor-binding assays (see section 2.2.11.3.2) performed on a subset of the 33 mutants revealed that insertions in many genes of this locus do not prevent the production of ACP (Figure 4.4).

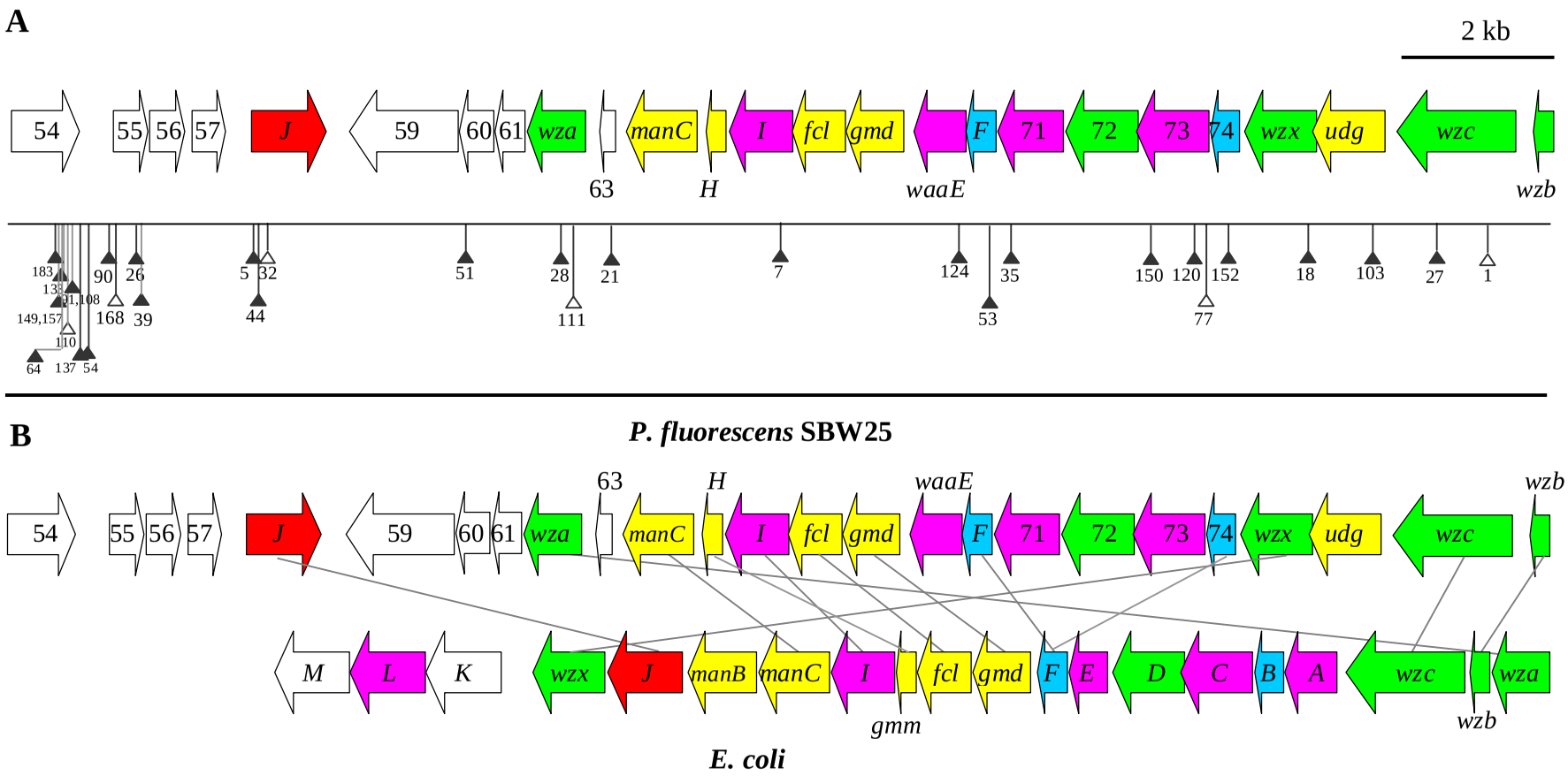


Figure 4.3: The colanic acid biosynthetic gene cluster. (**A**): Transposon insertions in the Pflu3658-78 gene cluster of *P. fluorescens* SBW25 (filled triangles=precise insertion site identified, unfilled triangles=precise insertion site unknown, numbers relate to JG genotype number). (**B**): Comparison of the *P. fluorescens* SBW25 colanic acid biosynthetic locus with that of *E. coli* (see Table 4.2). The SBW25 cluster is predicted to contain five genes involved in biosynthesis of a colanic acid precursor (yellow), *wcaJ* (red), four glycosyl transferases (purple), two acetyl transferases (blue), five genes involved in polymerisation and transport (green) and four genes of unknown function (white).

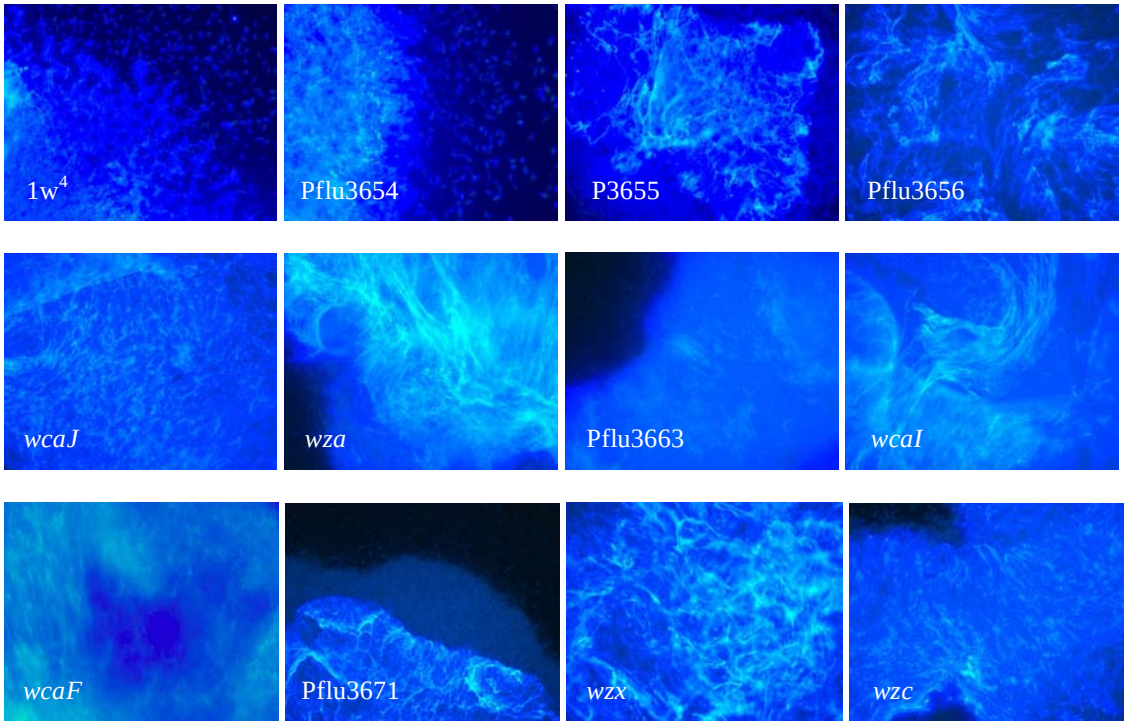


Figure 4.4: Calcofluor binding by non-capsulated 1w^4 transposon mutants with insertions in the colanic acid biosynthetic operon. Images were taken after 16 hours growth on KB+calcofluor agar, at x40 or x100 magnification under a fluorescence microscope (left to right): 1w^4, JG183, JG137, JG26 (top row); JG5, JG28, JG21, JG7 (middle row); JG53, JG35, JG18, JG27 (bottom row).

In order to determine which of the 16 identified genes are essential for capsule expression, representative transposon mutants were selected and used to construct 16 non-polar mutants (see section 2.2.8.2). A capsule counting assay performed on each of the resulting genotypes revealed that each of the 16 genes was essential for capsule production; all strains produced a significantly lower proportion of capsulated cells than 1w^4 (Figure 4.5, Appendix A2.2; M-W-W test P<0.01).

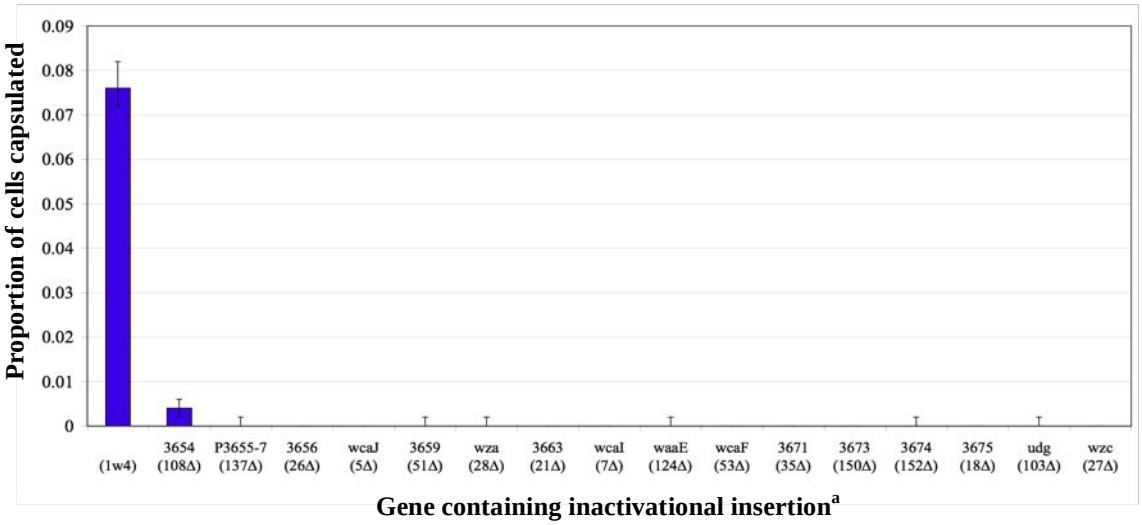


Figure 4.5: Proportion of capsulated cells in genotypes containing 189 bp inactivational insertions in colanic acid biosynthetic genes. Each bar indicates a median of five replicates; error bars indicate range of data. [a]Genotype numbers relating to transposon mutant numbers are given in parenthesis (e.g. 108Δ=JG108-ΔCre); P=promoter.

4.3.1.1.2 Construction of a wcaJ-lacZ *transcriptional fusion*

To determine whether capsule expression in $1w^4$ is controlled at the level of transcription, a *lacZ* transcriptional fusion was constructed in the colanic acid cluster. As the first conserved gene of the cluster, *wcaJ* was selected for the fusion. Firstly, the wild-type *wcaJ* gene was PCR-amplified (see section 2.2.2.1) using the primer pair WcaJ-lacZf/r (58˚C annealing temperature, 1.5 minutes extension time). The resulting PCR fragment was ligated into pCR8/GW/TOPO and used to transform chemically competent *E. coli* (see section 2.2.3.1). The fragment was retrieved from a mutation-free clone by *Bgl*II digestion and ligated into pUIC3 upstream of the *lacZ* gene, creating pUIC3-*wcaJ-lacZ*. This construct was used to transform chemically competent *E. coli* and conjugated into SBW25, $1s^4$ and $1w^4$ (see sections 2.2.3.2.3 and 2.2.6.2). In each background, the construct combined by homologous recombination into the colanic acid gene cluster, creating SBW25-*wcaJ-lacZ*, $1s^4$-*wcaJ-lacZ* and $1w^4$-*wcaJ-lacZ*, respectively.

Each genotype was grown for 72 hours on selective LB media containing X-gal (a chromogenic indicator of *lacZ* expression), and photographed under a dissection

microscope (Figure 4.6). SBW25-*wcaJ-lacZ* and 1s^4-*wcaJ-lacZ* produced uniform, yellowish colonies, while the colonies produced by 1w^4-*wcaJ-lacZ* contained large blue sectors, indicating increased *wcaJ-lacZ* transcription. Microscopic examination revealed blue sections contained a high proportion of capsulated cells. These results are consistent with the hypothesis that 1w^4 capsule expression is the result of increased transcription of genes in the colanic acid cluster.

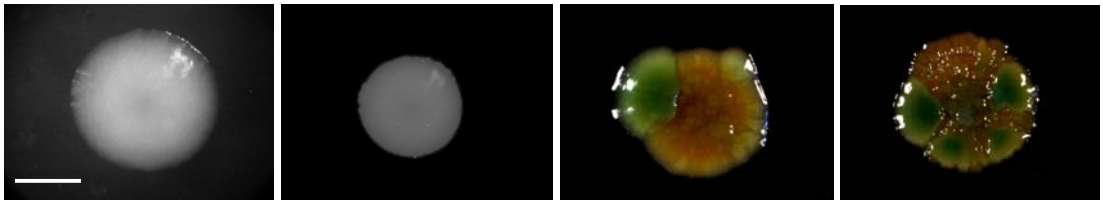


Figure 4.6: The increased expression of colanic acid required for the 1w^4-opaque phenotype is under transcriptional control. 72-hour colonies of SBW25-*wcaJ-lacZ* (left), 1s^4-*wcaJ-lacZ* (left centre) and 1w^4-*wcaJ-lacZ* (right centre and right) on LB+NF+Tc+X-gal agar. Scale bar represents approximately 3 mm. Contrasts and background colours were altered in iPhoto.

4.3.1.2 Insertions in genes involved in the biosynthesis of colanic acid precursors

The biosynthesis of *E. coli* colanic acid requires four nucleotide sugar precursors: UDP-D-glucose (UDP-Gluc), UDP-D-galactose (UDP-Gal), UDP-D-glucuronic acid (UDP-GlucA) and GDP-L-fucose (GDP-Fuc) (Anderson & Rogers, 1963). Eighteen independent non-capsulated mutants were obtained with insertions in genes required for production of these precursors (Table 4.1). These included one insertion in Pflu1668, five insertions in *galU* (Pflu2985), one insertion in *udg* (Pflu3676, see section 4.3.1.1), nine insertions in *pgi* (Pflu5243) and two insertions in *algC* (Pflu5986) (Table 4.3).

Protein	Pflu[a]	Ins[b]	Size[c]	Pfam domains[d]	*E*-value[d]	Precursor[e]
CapD	1668	1	664	Polysacc_synt_2	7.4×10^{-203}	UDP-Gal
GalU	2985	5	279	NTP_transferase	6.0×10^{-18}	All, esp. UDP-Gluc
Udg	3676	1	445	UDPG_MGDP_dh_N	1.2×10^{-89}	UDP-GlucA
				UDPG_MGDP_dh	2.4×10^{-4}	
				UDPG_MGDP_dh_C	3.0×10^{-32}	

Protein	Pflu[a]	Ins[b]	Size[c]	Pfam domains[d]	*E*-value[d]	Precursor[e]
Pgi	5243	9	554	PGI	3.2×10^{-224}	All, esp. GDP-Fuc
AlgC	5986	2	465	PGM_PMM_I	1.3×10^{-45}	All, esp. GDP-Fuc
				PGM_PMM_II	6.9×10^{-56}	
				PGM_PMM_3	1.4×10^{-45}	
				PGM_PMM_IV	1.1×10^{-25}	

Table 4.3: Transposon insertions, domain characteristics and biosynthetic pathways of identified colanic acid precursor biosynthetic proteins. [a]Pflu refers to the numeric name of the SBW25 gene. [b]Ins gives the number of independent insertions obtained. [c]Number of amino acids. [d]Pfam domain searches were performed on the Sanger Pfam website. [e]Precursor affected by insertion (esp.=especially).

4.3.1.2.1 ACP and capsule expression in colanic acid precursor biosynthetic mutants

Each insertion caused a visible reduction in capsule biosynthesis and the production of either translucent colonies (*capD*, *udg* and *algC*) or colonies that resembled neither form of 1w[4] (*galU* and *pgi*). Calcofluor binding assays revealed that insertions in *capD*, *udg*, and *algC* do not prevent ACP production, while those in *galU* and *pgi* disrupted ACP production (Figure 4.7). Insertions in *galU* or *pgi* have previously been shown to destroy the WS phenotype by disrupting ACP production (Gehrig, 2005).

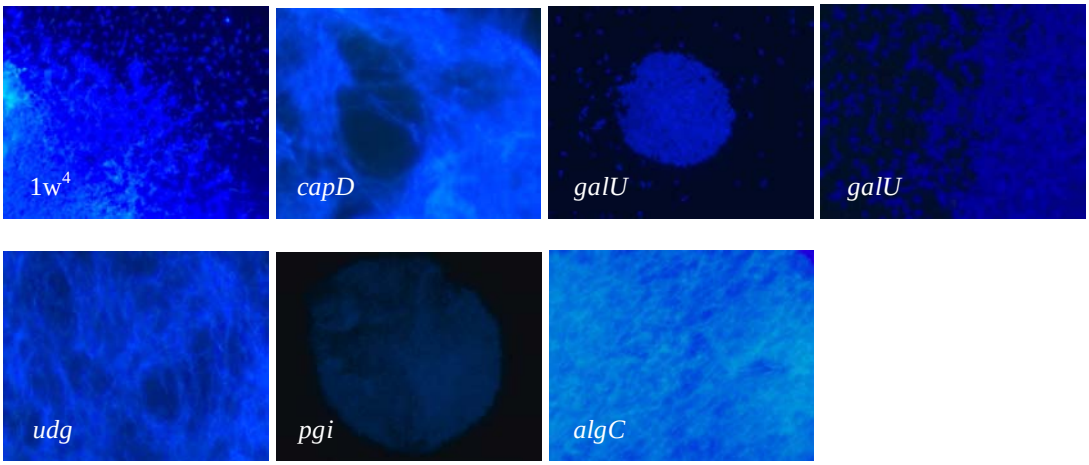


Figure 4.7: Calcofluor binding by non-capsulated 1w[4] transposon mutants with insertions in genes involved in the biosynthesis of colanic acid precursors. All images were taken after 16 hours growth on KB+calcofluor agar, at x40 or x100 magnification under a fluorescence microscope (left to right): 1w[4], JG62, JG49, JG114 (top row); JG103, JG23, JG55 (bottom row).

Representative mutants containing an insertion in each of *galU*, *udg*, *pgi* and *algC* were selected, and a Cre-deletion genotype constructed for each. The four Cre-deletion genotypes were then subjected to a capsule counting assay to determine the quantitative effect of each insertion on capsule expression (Figure 4.8, Appendix A2.2). Disruption of all four genes leads to a significant decrease in capsule expression (P<0.01).

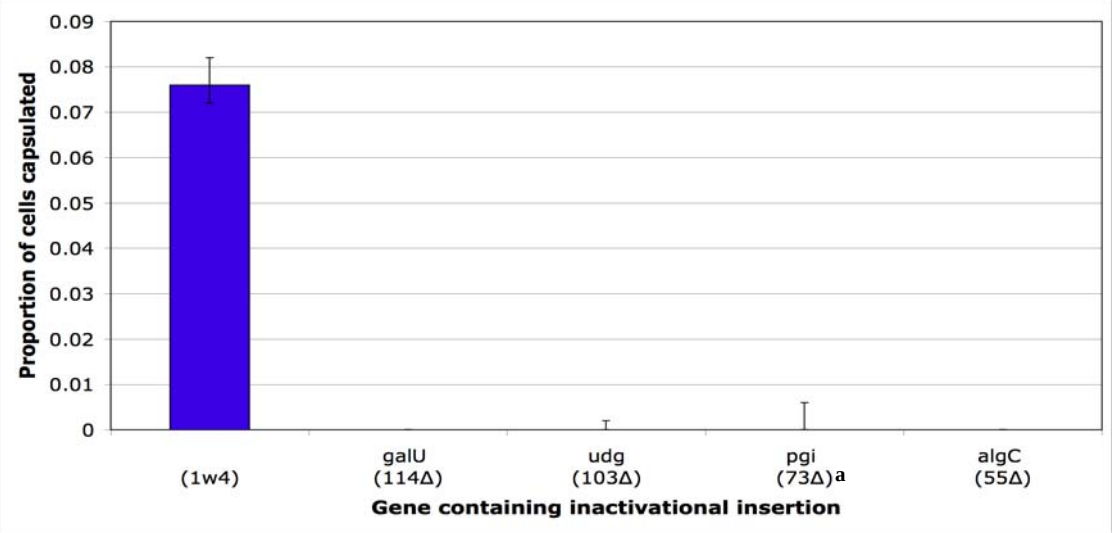


Figure 4.8: Proportion of capsulated cells in genotypes containing 189 bp inactivational insertions in genes encoding precursor biosynthetic genes. Each bar indicates a median of five replicates; error bars indicate range of data. [a]Genotypes relating to transposon mutant numbers are given in parenthesis (e.g. 114Δ=JG114-ΔCre).

4.3.1.2.2 Proposed biosynthetic pathways of colanic acid precursors

Illustrated in Figure 4.9, the biosynthetic pathways of the four colanic acid precursors are interlinked. UDP-Gluc, a common source for the other precursors, is synthesized from glucose-1-phosphate by *galU*-encoded UTP-glucose-1-phosphate uridylyltransferase. UDP-Gluc is converted into UDP-GlucA by the product of *udg* (UDP-Gluc dehydrogenase) and into UDP-Gal by UDP-Gluc 4-epimerase. This enzyme is generally encoded by *galE*, which is absent from SBW25. However, a BLASTP search revealed that Pflu1668 is homologous to CapD, an enzyme that has recently been shown to posses UDP-Gluc 4-epimerase activity in the Gram-negative obligate intracellular parasite, *Rickettsia prowazekii* (Santhanagopalan *et al*., 2006).

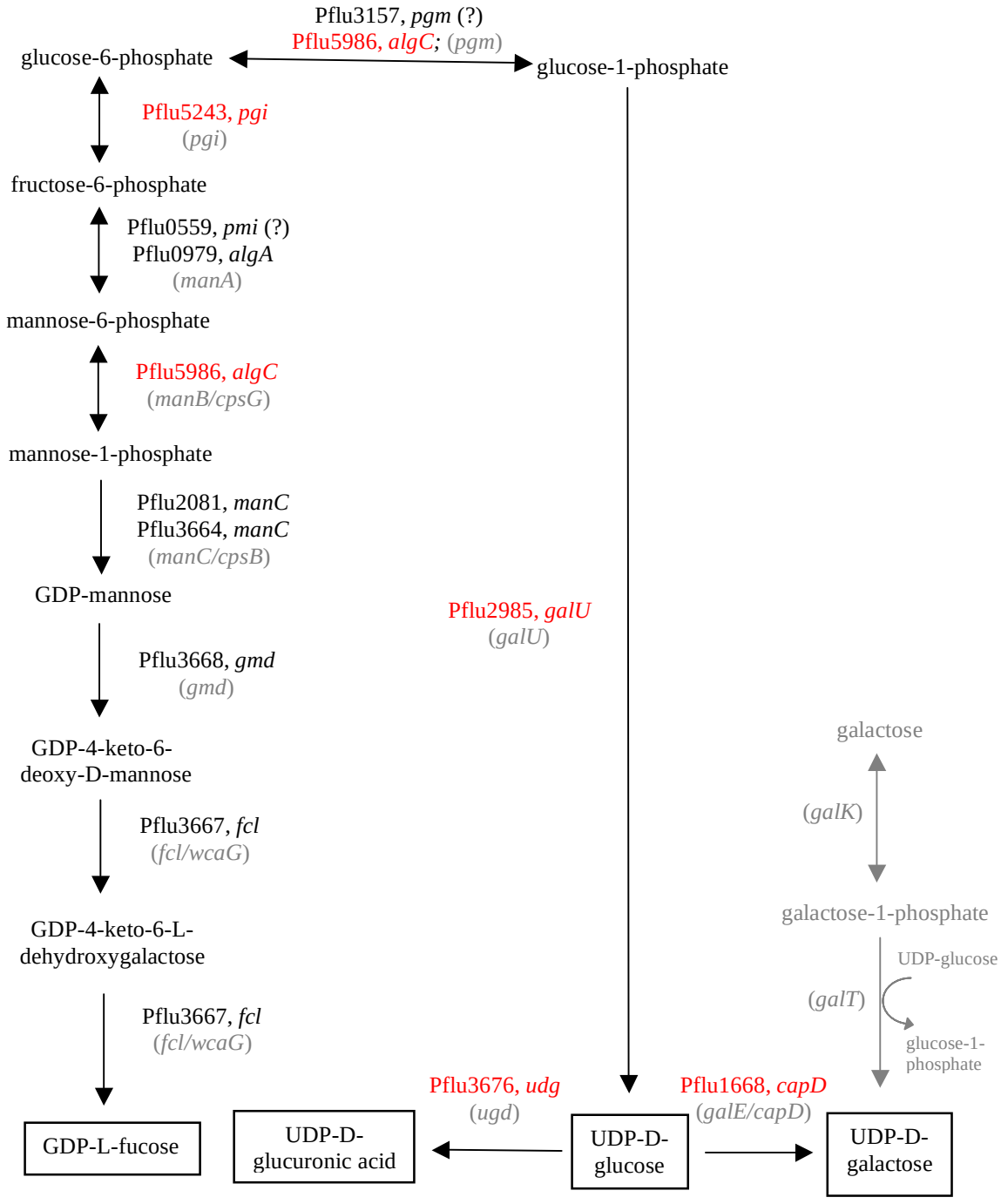


Figure 4.9: Transposon mutagenesis-based biosynthetic pathways of colanic acid precursors in 1w[4]**.** Genes encoding enzymes are shown by Pflu number (and name where appropriate; red=recorded transposon hits. (?) indicates uncertainty of involvement. Corresponding *E. coli* K12 gene names are shown in grey in parenthesis (Markovitz *et al.*, 1967). The *E. coli* UDP-Gal biosynthetic pathway missing from SBW25 is shown in grey to the right.

Finally, UDP-Gluc may be converted into GDP-Fuc by a series of enzymes including glucose-6-phosphate isomerase (Pgi), phosphomanno/glucomutase (AlgC), and a number of genes in the colanic acid cluster (see section 4.3.1.1). The lack of insertions in the committed GDP-Fuc biosynthetic enzymes indicates the existence of an alternative source of GDP-Fuc.

4.3.1.3 Insertions in potential transcriptional regulators of colanic acid genes

The transposon screen identified two genes potentially involved in transcriptional regulation of the colanic acid cluster: *barA* and *gacA*, with 18 and two independent insertions, respectively (Table 4.1). The *barA* gene encodes a trans-membrane protein containing HK and RR domains. The *gacA* gene encodes a smaller, cytoplasmic protein containing RRr and DNA binding domains (see section 1.2.1.5; Table 4.4). A BLASTN search revealed that both of these genes are widely conserved in Gram-negative bacteria, with BarA variably named GacS, LemA and AirS, and GacA also known as UvrY. The *E. coli* BarA/GacA homologs BarA/UvrY encode an atypical TCSTP that controls the transcription of multiple genes in response to environmental stimuli (Rich *et al*., 1994; Pernestig *et al*., 2001; Mondragón *et al*., 2006).

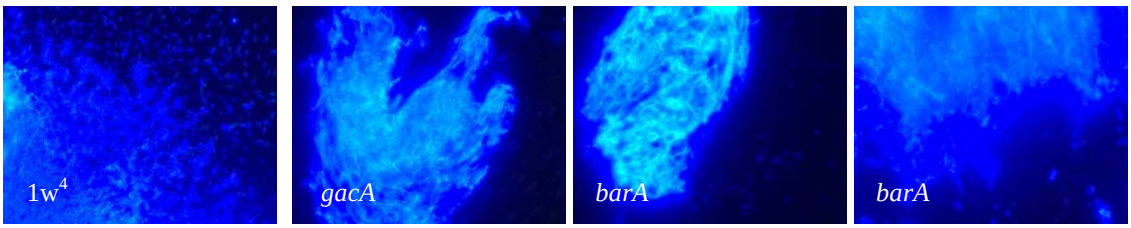
Domain[a]	*E*-value[a]	Size[b]	Predicted function
BarA (encoded by Pflu3777)			
HAMP	2.6×10^{-19}	70	Links signal sensing & histidine kinase domains
HisKA	3.8×10^{-28}	66	HK; accepts activating phosphoryl group
HATPase_c	2.0×10^{-40}	116	Phosphoryl group provision by ATP hydrolysis
Response_reg	4.8×10^{-38}	116	RR; receives phosphoryl signal from HK
Hpt	1.5×10^{-9}	83	Transfers phosphoryl group to an RR
GacA (encoded by Pflu2189)			
Response_reg	2.9×10^{-33}	113	RR; receives phosphoryl signal from HK
GerE	2.6×10^{-22}	58	DNA binding, helix-turn-helix LuxR-like domain

Table 4.4: Predicted domain characteristics and associated functions for BarA and GacA. [a]Pfam domain searches were performed through the Sanger Pfam website. [b]Number of amino acids.

4.3.1.3.1 ACP and capsule expression in transcriptional regulators of colanic acid

Transposon insertions in *gacA* and *barA* reduced capsule biosynthesis and gave rise to distinctive, wrinkly-like colonies. Calcofluor binding assays revealed that disruption of *gacA* or *barA* did not prevent ACP production (Figure 4.10A). Representative strains were selected, and Cre-deletions constructed. Capsule counting assays revealed that both non-polar insertions significantly reduced the proportion of capsulated cells (Figure 4.10B, Appendix A2.2; M-W-W test $P<0.01$).

A



B

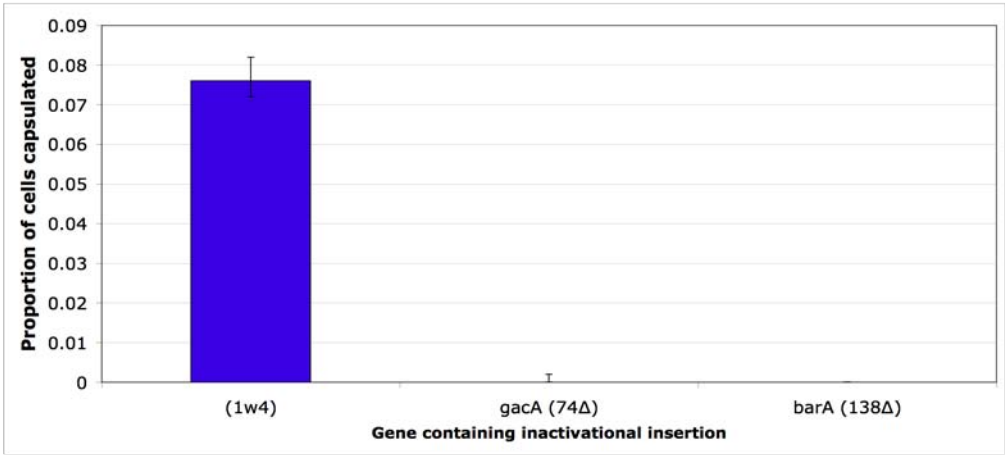


Figure 4.10: Phenotypic investigation of non-capsulated 1w^4 transposon[a] mutants with insertions in potential transcriptional regulators of colanic acid genes. (**A**): Calcofluor binding ability and ACP biosynthesis. All images taken after 16 hours growth on KB+calcofluor agar, at x40 or x100 magnification on a fluorescence microscope (left to right): 1w^4, JG74, JG45 & JG106. (**B**): Proportion of capsulated cells. Bars indicate the median of five replicates; error bars indicate data range. [a]Genotype numbers relating to transposon mutant numbers are given in parenthesis (e.g. 74Δ=JG74-ΔCre).

4.3.1.3.2 The proposed molecular mechanism of the BarA/GacA TCSTP

Present in a wide range of Gram-negative bacteria, the BarA/GacA TCSTP has been shown to control the transcription of genes involved in the synthesis of extracellular

products, enzymes and secondary metabolites (reviewed in Heeb & Haas, 2001). The BarA/GacA TCSTP has not previously been implicated in the regulation of capsular polysaccharides; colanic acid transcription is classically controlled by the Rcs signal transduction system (reviewed in Majdalani & Gottesman, 2005). However, combined with the lack of transposon insertions in SBW25 Rcs-like genes, insertions in *barA/gacA* indicate that colanic acid expression may be under atypical transcriptional control in SBW25. Illustrated in Figure 4.11, the proposed molecular mechanism of colanic acid control involves activation of an unorthodox phosphotransfer reaction from BarA, resulting in DNA binding by GacA.

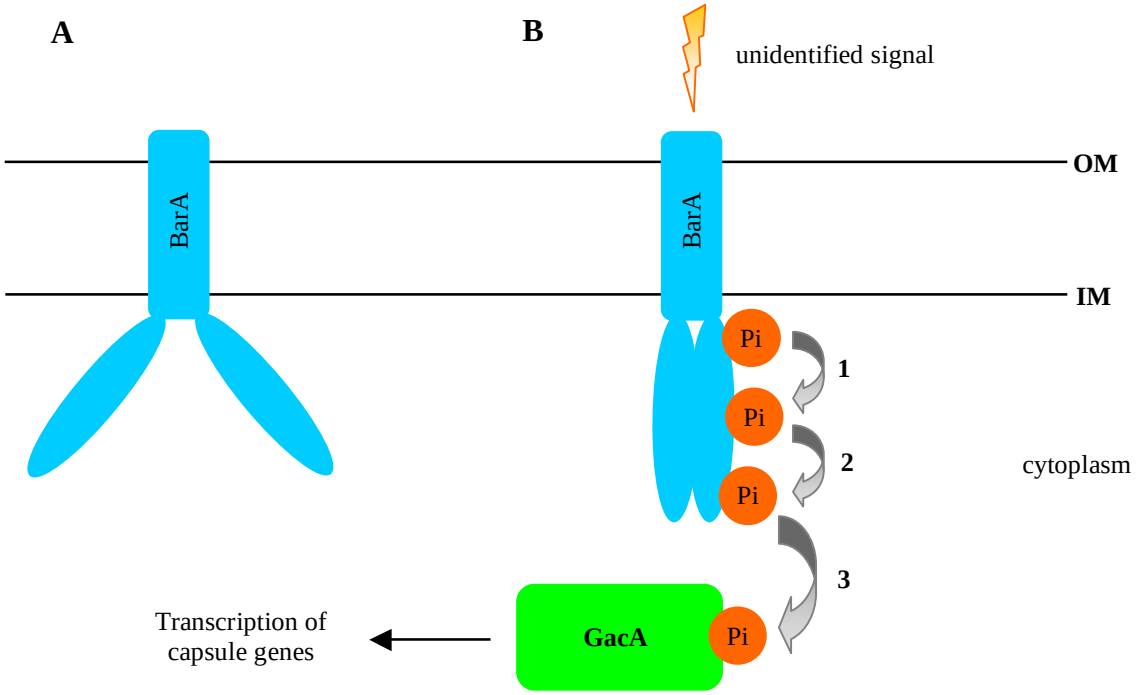


Figure 4.11: Proposed mechanistic model for the activation of BarA/GacA TCSTP in 1w[4]**.** In the inactive state (**A**), BarA and GacA are not phosphorylated, and GacA-activated genes are not expressed. Upon activation by an unidentified environmental signal (**B**), a conformational change takes place in BarA, allowing autophosphorylation to occur at a conserved histidine residue. The phosphoryl group (Pi) is transferred to an aspartate residue in BarA (1), and subsequently to a second, conserved histidine residue (2). The third phosphoryl transfer reaction (3) passes the phosphoryl group to a conserved aspartate residue in the RR receiver domain of GacA. This causes a conformational change in GacA, activating its DNA-binding activity and expression of genes required for capsule expression. Model based on the homologous BarA/UvrY TCSTP of *E. coli* (Pernestig *et al*., 2001) (OM=outer membrane, IM=inner membrane).

4.3.1.4 Insertions in genes that reduce colanic acid biosynthesis indirectly

There are 43 insertions in other genes that reduce capsule biosynthesis, presumably by indirect means. For example, the heat shock chaperone DnaK (JG88) has been shown to control Rcs signalling (Zuber *et al.*, 1995), and Dsb proteins (JG80) have been shown to be involved in post-translational modification of colanic acid proteins (Kadokura *et al.*, 2004). Calcofluor binding assays were performed for selected transposon mutants (Figure 4.12); of those mutants tested, only insertions in the glycosyltransferase-encoding Pflu0478 prevented ACP production. Accordingly, insertions in Pflu0478 were also obtained during a suppressor analysis of WS phenotype (therein named *wgcD*) (Gehrig, 2005). Several insertions of note are discussed in more detail in the following sections.

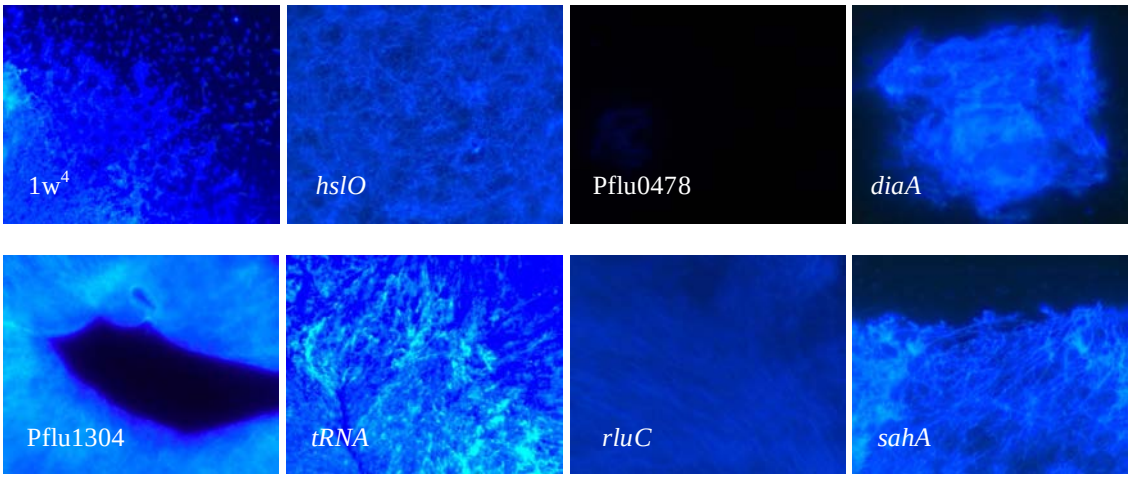


Figure 4.12: Calcofluor binding by 1w^4 transposon mutants with insertions in genes that reduce capsule expression indirectly. Images were taken after 16 hours growth on KB+calcofluor agar, at x40 or x100 magnification under a fluorescence microscope (left to right): 1w^4, JG136, JG161, JG165 (top row), JG132, JG38, JG82, JG72 (bottom row).

4.3.1.4.1 Insertions in rluC *and* rne

In total, seven insertions were obtained in *rluC* and *rne*, two adjacent but divergently transcribed genes; five insertions were obtained in *rluC* (JG11, JG41, JG82, JG131 and JG141), and one insertion was obtained in *rne* (JG147). The precise location of the seventh insertion (JG128) has not been determined, and thus the insertion may occur in either gene. The *rluC* gene encodes pseudouridine synthase, an enzyme that

synthesizes three pseudouridine residues in 23S rRNA (Conrad *et al.*, 1998). While the importance of these residues in capsule expression has not previously been reported, it is possible that they are required for the synthesis of one or more essential proteins. Alternatively, insertions in *rluC* may disrupt the expression of *rne*, which encodes ribonuclease E. Ribonuclease E has been shown to degrade mRNAs involved in the transport of glucose required for colanic acid precursor biosynthesis (Kimata *et al.*, 200; El-Kazzaz *et al.*, 2004). It is conceivable that alterations in glucose transport influence the expression of glucose-requiring polymers such as colanic acid.

4.3.1.4.2 Insertions in a genomic locus encoding tRNAs

Three independent insertions were obtained in the tRNA-encoding locus between Pflu1858 and Pflu1859 (JG38, JG46 and JG109; Figure 4.13). The locus encodes four tRNAs: a Glu (anticodon TTC) /Gly (anticodon GCC) tandem repeat. Additional copies of these tRNAs occur in the genome; two further copies of Glu(TTC) are located between Pflu4205 and Pflu4207, and one further copy of Gly(GCC) is found between *pgsA* and Pflu2192. A non-polar mutation (JG109-ΔCre) was constructed in the first Glu repeat, and a capsule counting assay estimated the median proportion of capsulated cells in JG109-ΔCre to be 0.014 (see Appendix A2.2). This was a significantly lower proportion than that obtained for 1w^4 (M-W-W test P=3.97 x 10^{-3}). Why insertions in this tRNA locus reduce colanic acid expression is unclear; however, since a number of genes required for capsule biosynthesis contain the relevant Glu codon, it is conceivable that a reduction in tRNA availability reduces the level of capsule biosynthetic proteins.

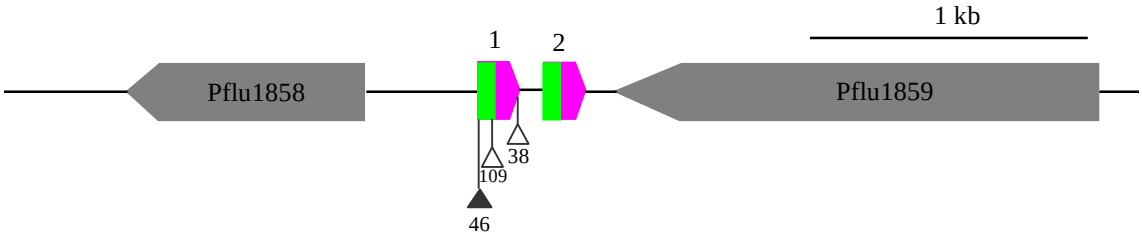


Figure 4.13: Distribution of insertions in the Glu/Gly tRNA locus between Pflu1858 and Pflu1859. The locus contains tandem repeats (1 and 2) of the Glu(TTC)/Gly(GCC) anticodons (green and purple, respectively). Transposon insertions sites are indicated by triangles (filled=exact insertion site, open=approximate insertion site). Numbers correspond to JG genotype number.

4.3.1.4.3 Insertions in sahA

Seven independent insertions were obtained throughout *sahA* (JG3, JG13, JG72, JG125, JG156, JG158 and JG181), which encodes S-adenosyl-L-homocysteine hydrolase. A central enzyme of the methionine biosynthetic pathway, this enzyme interconverts S-adenosyl-L-homocysteine and homocysteine/adenosine. Additional insertions were identified in genes of the methionine biosynthetic pathway, including *hemK*, Pflu5399, Pflu5400 and *hemL* (see Appendix A2.1). However, as these insertions significantly reduced growth rate, they were discarded from the analysis. A non-polar mutation (JG150-ΔCre) was constructed, and a capsule counting assay estimated the median proportion of capsulated cells in JG150-ΔCre to be zero (see Appendix A2.2). This was a significantly lower proportion than that obtained for 1w^4 (M-W-W test P=3.97 x 10^{-3}). It is not obvious why mutations in this gene cause a reduction in capsule (but not ACP) expression. However, given that SahA plays a role in central metabolism, mutations may alter different biochemical pathways giving rise to a range of phenotypic effects. For instance, the methionine biosynthetic pathway is closely linked to nucleotide biosynthesis (through the production of tetrahydrofolic acid, a precursor of purine nucleotides), which may in turn affect the levels of UDP/GDP available for the biosynthesis of colanic acid precursors. This possibility is particularly notable as insertions were obtained in four other loci that influence nucleotide pools: Pflu1304 (JG132), *ndk* (JG176), *nusA* (JG36) and *purU* (JG76, JG113 and JG148) (see sections 4.3.1.5.1, 4.3.1.5.2 and 6.4.3.1).

4.3.1.5 Insertions that increase capsule biosynthesis

Since only a small fraction (six of 127: ~5 %) of transposon mutants resulted in an increase in capsule expression (Table 4.1), each insertion was of considerable interest. Although all six strains produced opaque-like colonies, none of the strains produced entirely capsulated populations; each one continued to produce a significant proportion of non-capsulated cells. Insertions increasing the proportion of capsulated cells were obtained in four genomic loci: Pflu3841 (JG30), *purU*/Pflu4939 (JG76, JG113 and JG148), *ndk* (JG176) and Pflu5438 (JG178). Calcofluor binding assays revealed that –

with the exception of the insertion in Pflu5438 – these insertions greatly reduced ACP production (Figure 4.14). Presumably, this results from a lower proportion of ACP-synthesizing non-capsulated cells (see section 3.4.1).

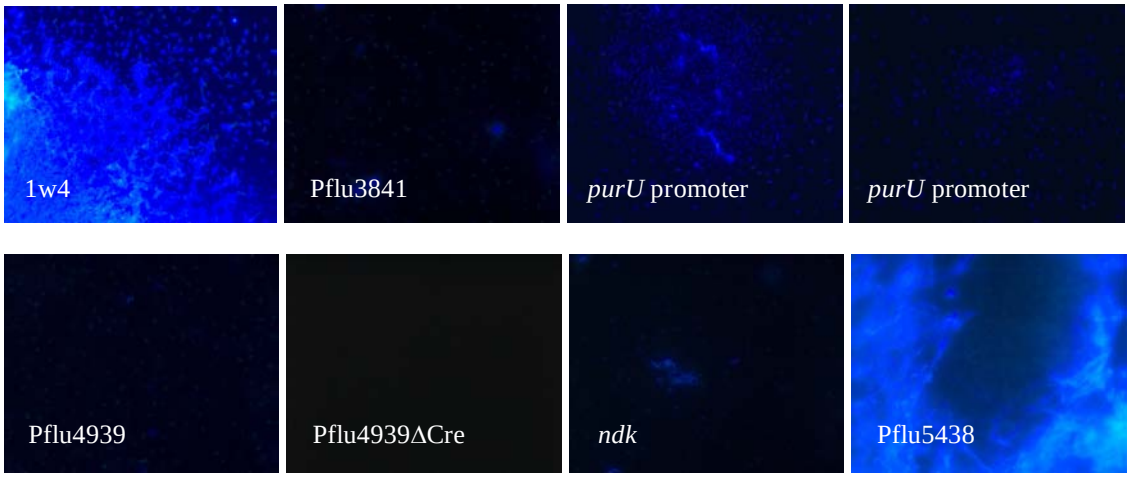


Figure 4.14: Calcofluor binding by 1w^4 transposon mutants with insertions in genes that increase capsule expression. Images were taken after 16 hours growth on KB+calcofluor agar, at x40 or x100 magnification under a fluorescence microscope (left to right): 1w^4, JG30, JG76, JG113 (top row); JG148, JG148-ΔCre, JG176 JG178 (bottom row).

As none of the four loci above presented an obvious connection to capsule biosynthesis, investigative BLASTP and Pfam domain searches were employed (Table 4.5). The product of Pflu3841 is a hypothetical integral membrane protein unique to *P. fluorescens* SBW25 and contains one domain of unknown function (DUF). Pflu5438 encodes an aromatic acid decarboxylase that is conserved across many bacterial species. The nature of the involvement of both Pflu3841 and Pflu5438 in the capsule phenotype remains unknown. The *purU*-Pflu4939 locus encodes both PurU, an enzyme involved in methionine and purine nucleotide biosynthesis, and a predicted transcriptional regulatory protein. Finally, the product of *ndk* is nucleoside diphosphate kinase, a highly conserved protein involved in the biosynthesis of nucleotides. The *purU*-Pflu4939 and *ndk* insertions are discussed in more detail in the following sections.

Protein	Pflu[a]	Ins.[b]	Size[c]	Pfam domains[d]	*E*-value[d]	Homologs[e]
Pflu3841	3841	1	128	DUF2282	1.0×10^{-14}	SBW25
purU	4938	3	282	ACT Formyl_trans_N	1.2×10^{-7} 1.8×10^{-47}	Conserved
Pflu4939	4939		124	None	n/a	*Pseudomonas*
ndk	5061	1	141	NDK	1.3×10^{-75}	Conserved
Pflu5438	5438	1	209	Flavoprotein	6.1×10^{-43}	Conserved

Table 4.5: Predicted function, domain characteristics and homologs of proteins identified as increasing capsules. [a]Pflu refers to the numeric name of the gene encoding the protein in the SBW25 genome. [b]Ins. refers to the number of independent transposon insertions obtained in the gene encoding each protein. [c]Number of amino acids. [d]Pfam domain searches (including associated *E*-values) were performed through the Sanger Pfam website. [e]Conservation level of the protein using a BLASTP homolog search (Conserved=conserved in bacterial genomes outside the *Pseudomonas* genus).

4.3.1.5.1 Insertions in the purU-*Pflu4939 locus*

Three independently obtained transposon mutants contain insertions in the *purU*-Pflu4939 genomic locus. The locus consists of two divergently transcribed genes, *purU* (Pflu4938) and Pflu4939. JG76 and JG113 were found to contain insertions at precisely the same nucleotide of the intergenic region, while JG148 contains an insertion at the 3′ end of Pflu4939 (Figure 4.15). The transcriptional regulator encoded by Pflu4939 belongs to the *Pseudomonas*-specific MvaT family, members of which have been shown to regulate a wide range of genes. These include genes involved in mevalonate metabolism in *Pseudomonas mevalonii* (Rosenthal & Rodwell, 1998), the fimbrial *cup* genes of *P. aeruginosa* (Vallet *et al.*, 2004), and genes encoding the antimicrobial exoproteins of *P. fluorescens* CHA0 (Baehler *et al.*, 2006). Therefore, it is conceivable that Pflu4939 regulates the expression of genes required for capsule biosynthesis.

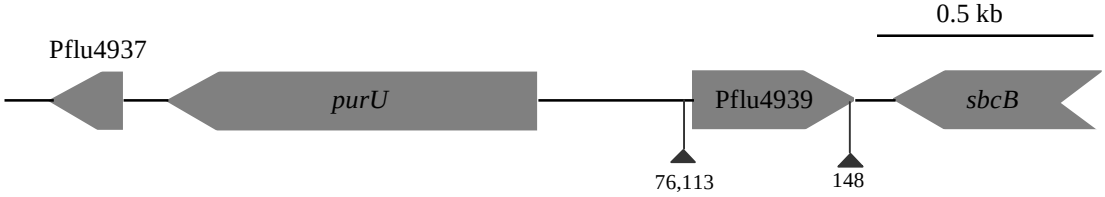


Figure 4.15: Distribution of transposon insertions in the *purU*-Pflu4939 genomic locus. Transposons are indicated as triangles. Numbers refer to JG genotype numbers.

A non-polar mutation was generated in Pflu4939, giving strain JG148-ΔCre, and a capsule counting assay revealed that JG148-ΔCre generated almost exclusively non-capsulated cells (median 0, Appendix A2.2). This phenotype was unexpected, given that JG148 was isolated as a result of increased capsule production. The flipping of the phenotype on Cre-deletion is consistent with the hypothesis that the original capsulated phenotype was the result of increased *purU* expression from the *nptII* promoter of the transposon, and the non-capsulated phenotype was the result of correspondingly low *purU* expression levels. The *purU* gene encodes formyltetrahydrofolate deformylase, an enzyme required in methionine biosynthesis and nucleotide pools (see section 4.3.1.4.3). Interestingly, there are multiple *purU* genes in the SBW25 genome: *purU1* (Pflu2321), *purU2* (Pflu2332), *purU* (Pflu4938) and *purU* (Pflu5647). It is unclear at this stage why the *purU* gene influences capsule expression.

4.3.1.5.2 Insertion in ndk

A single insertion was obtained at the 5′ end of *ndk* (Pflu5061). Ndk is a highly conserved protein that synthesizes all ribo- and deoxyribonucleoside triphosphates. A non-polar mutant was constructed (JG176-ΔCre), and a counting assay estimated the median proportion of capsulated cells in JG176-ΔCre to be 0.544 (Appendix A2.2). This was a significantly greater proportion than that obtained for 1w^4 (M-W-W test P=3.97 x 10^{-3}). Precisely how this phenotype emerges is unknown at present. However, targeted deletion of *E. coli ndk* has been shown to cause nucleotide pool aberrations, which in turn affect RNA, DNA, polysaccharide and signalling molecule biosynthesis (Lu *et al*., 1995; Bernard *et al.*, 2000). Accordingly, a range of phenotypes has been reported for *ndk* mutants, including a mutator phenotype in *E. coli* (Miller *et al*., 2002) and alteration of biosynthesis of the polysaccharide alginate in *P. aeruginosa* (Sundin *et al.*, 1996a; Sundin *et al.*, 1996b). Further insight into the role of the *ndk* gene in the 1w^4 switching phenotype is provided in sections 6.4.3.1, 6.4.3.2.4 and 6.4.4.2.1.

4.4 Discussion

4.4.1 Further insights into the 1w^4 phenotype

Further to the conclusions of Chapter 3, the results in this chapter demonstrate a causal correlation between capsule expression and colony opacity and 1w^4 phenotypic switching; loss of capsule production results in loss of colony dimorphism. Furthermore, the mutagenesis screen identified the genomic locus encoding the structural basis of the capsule phenotype: Pflu3656-*wzb*. While activation of this locus has not previously been described in pseudomonads, homologous loci in *E. coli* and *Salmonella enterica* encode enzymes for the biosynthesis and export of the capsular polysaccharide colanic acid (Stevenson *et al*., 1996; Stevenson *et al*., 2000). Once secreted, the polymer forms an acidic coat around cells, which is consistent with the acidic phenotype of the 1w^4 capsulated fraction on MacConkey agar (see section 3.3.2.1). Finally, the lack of insertions in the *wss* operon corroborates the hypothesis that ACP production is not required for 1w^4 dimorphism (see section 3.3.2.2.2).

4.4.2 A model for biosynthesis and regulation of colanic acid

Collectively, the results of the transposon mutagenesis screen provide significant insight into the biosynthesis of the capsule polymer. The results have enabled the production of a basic model for the synthesis, secretion and regulation of capsule biosynthesis in *P. fluorescens* SBW25 (Figure 4.16).

4.4.2.1 Precursors of colanic acid biosynthesis

The literature suggests that colanic acid is synthesized from four sugar precursors (Anderson & Rogers, 1963). Transposon insertions were isolated in genes that directly affected three of these precursors: UDP-Gluc, UDP-Gal and UDP-GlucA. Notably, the standard UDP-Gal biosynthetic genes (*galT*, *galE* and *galK*) are absent in SBW25, and thus *capD* is predicted to perform the role of UDP-Gal biosynthesis in 1w^4. No insertions were obtained in genes dedicated solely to the biosynthesis of GDP-Fuc.

There are three possible explanations for this: (1) insertions in these genes are lethal, (2) GDP-Fuc is not required for biosynthesis of the capsule polymer, or (3) an alternative source of GDP-Fuc exists. Given the complex nature of central metabolism, the third possibility seems probable. Notably, the insertion in *capD* (JG62) does not abolish capsule production, but merely reduces the proportion of capsulated cells, suggesting that an additional source of UDP-Gal exists.

4.4.2.2 Synthesis and secretion of colanic acid

In order to define the 5' boundary of the colanic acid structural locus, the effect of insertions in Pflu3654-7 must be considered. A total of 14 transposon mutants were isolated in this region: eight in the final ~300 bp of the Pflu3654 (1,320 bp), two in the promoter of Pflu3655-7, and four in Pflu3656. Taken together, these results suggest that each insertion alters Pflu3656 expression, implicating Pflu3656 as the first gene of the colanic acid structural cluster. Like Pflu3656, Pflu3659-61 and Pflu3663 are unique to the SBW25 locus. These genes encode putative lipoproteins and exported proteins that are hereby proposed to perform a structural role in the 1w^4 capsule polymer. In contrast to the amorphous, loosely associated colanic acid capsules of *E. coli* and *S. enterica* (Whitfield, 2006), 1w^4 capsules are morphologically consistent, tightly associated structures. Phenotypically similar *E. coli* capsules have been described, in which colanic acid is anchored to outer membrane lipopolysaccharides (Meredith *et al.*, 2007).

4.4.2.3 Transcriptional regulation of colanic acid biosynthetic genes

Primarily activated in response to osmotic stress, the Rcs system of *E. coli* increases transcription of colanic acid biosynthesis, producing a capsule that functions to protect the cell against desiccation (Majdalani & Gottesman, 2005). In the absence of insertions in likely *P. fluorescens* Rcs homologs, it is proposed that the BarA/GacA system is directly responsible for the transcriptional control of at least some genes of the SBW25 colanic acid cluster. Notably, a large genomic island encompasses the colanic acid gene cluster from *wcaJ-wzb*, indicating the cluster (excluding Pflu3656) was acquired by horizontal transfer from another species (Silby *et al.*, 2009). Under such circumstances, it is possible that the incoming genes incorporated into different, pre-existing SBW25

regulatory systems, and so it is difficult to speculate on the environmental conditions that activate colanic acid production in SBW25. However, previous studies of capsules with similar physical properties indicate that capsule synthesis in *P. fluorescens* is activated in response to temperature and cellular stress (Sledjeski & Gottesman, 1996).

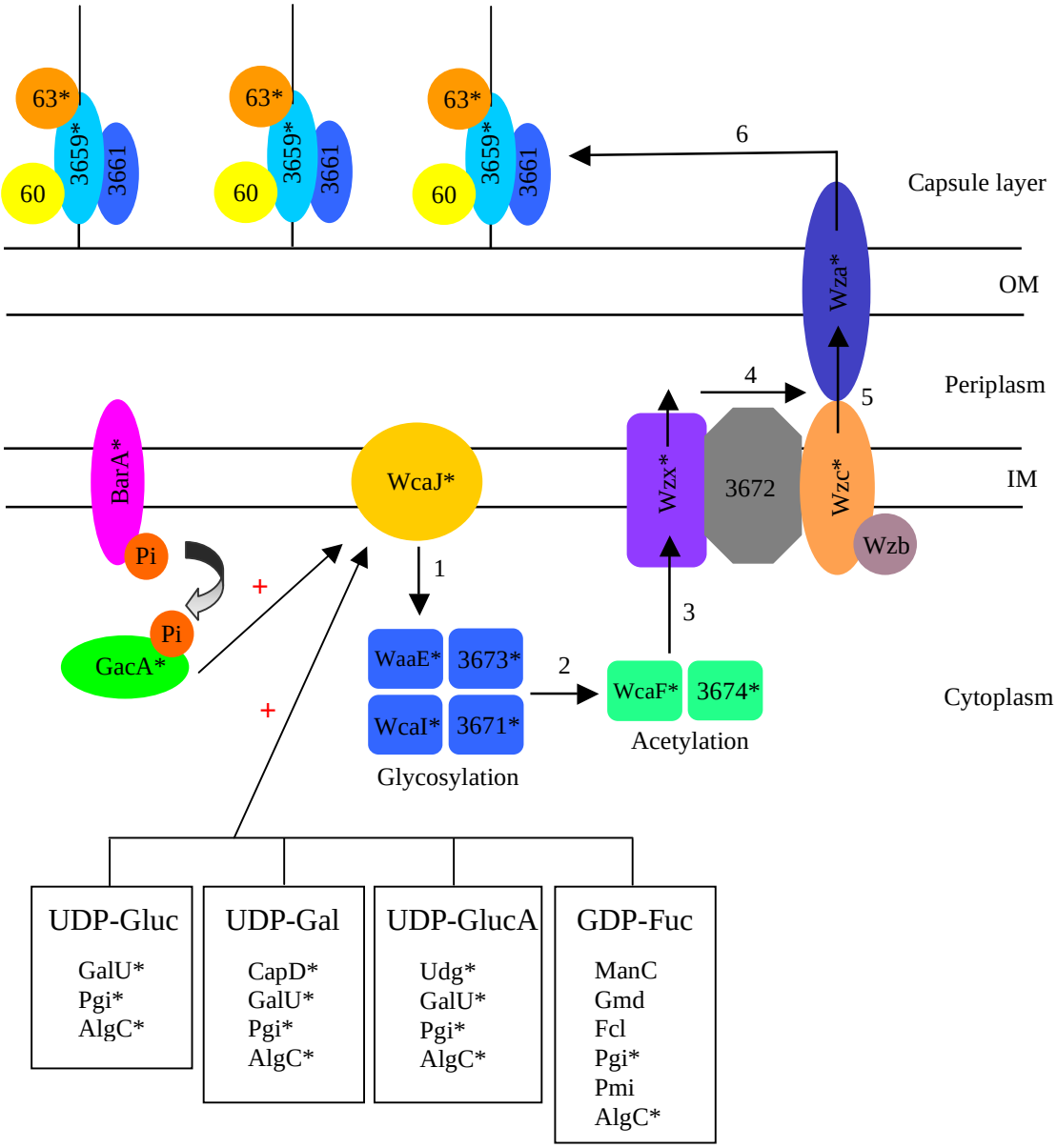


Figure 4.16: A possible mechanism for the biosynthesis of capsules in *P. fluorescens* 1w[4]. Four precursors are synthesized by the genes indicated. These precursors are synthesized into a colanic acid based capsule *via* the detailed route: (1) sequential glycosylation, (2) sequential acetylation, (3) Wzx-mediated transport to the periplasm, (4) Pflu3672-mediated polymerization, (5) modification of phosphoryl groups by Wzc and Wzb, and (6) transport to the extracellular milieu and attachment to a lipoprotein complex by Wza. The transcription of at least some of the polymer biosynthetic genes is thought to be positively regulated by the BarA/GacA TCSTP. (*=transposon insertions reduce/abolish capsule switching). Figure drawn with assistance from Stevenson *et al.* (1996) and Whitfield (2006).

4.4.3 Intracellular nucleotide pools affect the 1w^4 phenotype

An emerging theme of the transposon mutagenesis screen was the importance of nucleotide pools in the 1w^4 capsule phenotype; some insertions affecting nucleotide pools caused a decrease in capsule biosynthesis (those in Pflu1304, *nusA* and *sahA*), while others resulted in an increase in capsule expression (those in *purU* and *ndk*). With the exception of *ndk*, the identified genes were not directly involved in nucleotide biosynthesis. Presumably, this reflects the essential nature of most nucleotide biosynthetic genes. Surprisingly, deletion of *ndk* is not lethal, indicating that nucleoside triphosphates may be synthesized or salvaged from alternate sources (Bernard *et al*., 2000). Traditionally, the role of nucleotides is in RNA and DNA synthesis. Awareness of additional roles for intracellular nucleotides is increasing, with reports of nucleotide pools influencing the biosynthesis of polysaccharides and intracellular signalling molecules (reviewed in Chakrabarty, 1998).

The production of polysaccharides requires polymerisation of nucleotide-associated sugar molecules. Consequently, the synthesis of colanic acid (and ACP) by 1w^4 may be influenced not only by the availability of sugar moieties, but also by the cellular nucleotide pool. Additionally, nucleotide pools may influence the production of c-di-GMP (see section 1.2.1.5), an intracellular signalling molecule known to activate the biosynthesis of various polysaccharides, including ACP in *P. fluorescens* SBW25 and capsules in *V. parahaemolyticus* (Güvener & McCarter, 2003). The possibility that c-di-GMP contributes to the 1w^4 phenotype is supported by two lines of evidence. Firstly, the isolation of 14 non-capsulated transposon mutants predicted to affect expression of Pflu3656, which contains an EAL domain, indicates an inverse relationship between c-di-GMP and capsule expression. Secondly, the observation of multiple alterations in ACP biosynthesis during the evolution of 1w^4 (see Figure 3.1E) indicates a pivotal role for c-di-GMP expression in the phenotypic history of 1w^4. See section 6.4.3.1 for further discussion of the role of nucleotide biosynthesis in 1w^4 phenotypic switching.

4.4.4 Insights into the molecular basis of 1w^4 switching

No transposon insertions were obtained in genes obviously linked to the switch mechanism. For instance, no insertions were obtained in site-specific recombinases or methylases. It is important to note that the absence of candidate mechanistic genes does not necessarily imply an epigenetic switch mechanism; the transposon screen may be of insufficient resolution to uncover some genetic bases (such as small SSM loci). Alternatively, 1w^4 phenotypic switching may operate *via* a previously unrecognised mechanism. Therefore, mechanistic insight provided by the transposon screen is very limited and the nature of 1w^4 phenotypic switching remains unknown.

Chapter 5:
Unraveling the Evolutionary History of 1w^4

5.1 Introduction

The field of evolutionary genetics has been revolutionised by a recent upsurge of powerful technologies that allow sequencing of entire bacterial genomes (reviewed in Hall, 2007). Although subtly different, each of these technologies uses the previously determined DNA sequence of an ancestral reference genome to build the sequence of an evolutionarily derived genome. Thus, the genetic changes that constitute the evolutionary history of genotypes may be elucidated. Such technologies include hybridization-based Roche NimbleGen sequencing, the Roche 454 Life Sciences sequencing-by-synthesis approach and PCR-based sequencing techniques (Illumina Solexa). This chapter describes the use of Solexa technology to elucidate the mutational history of 1w^4.

The Solexa sequencing process consists four main stages: sample preparation, PCR-based amplification, sequencing, and bioinformatic analysis (Figure 5.1). Specifically, pure genomic DNA isolated from the genotype of interest is randomly fragmented, and adapter molecules are ligated onto the both ends of each DNA fragment (A). Thus prepared, single stranded DNA fragments are subsequently bound at random to the inside surface of a flow cell channel, upon which adapter molecules are also bound (B). Free, unlabelled nucleotides and polymerase are added to the flow cell channel, and the first PCR cycle commenced. During this cycle, bound DNA fragments form 'bridges' with nearby adapter molecules, the extension of which results in a double stranded DNA fragment (C). Following denaturation, two single stranded templates are anchored to the flow cell channel (D). As the PCR cycles continue, several million clusters of DNA strands build up in the flow cell channel, each stemming from a single original template (E). Sequencing involves reading a 36 nucleotide DNA sequence from each of these clusters, using the addition of fluorescently labelled nucleotides and laser imaging

(F). Finally, the several million independent reads are aligned to a reference genome to give ~15-40 fold coverage of every base, and bioinformatic analyses are employed to locate differences between the reference and derived genome (G).

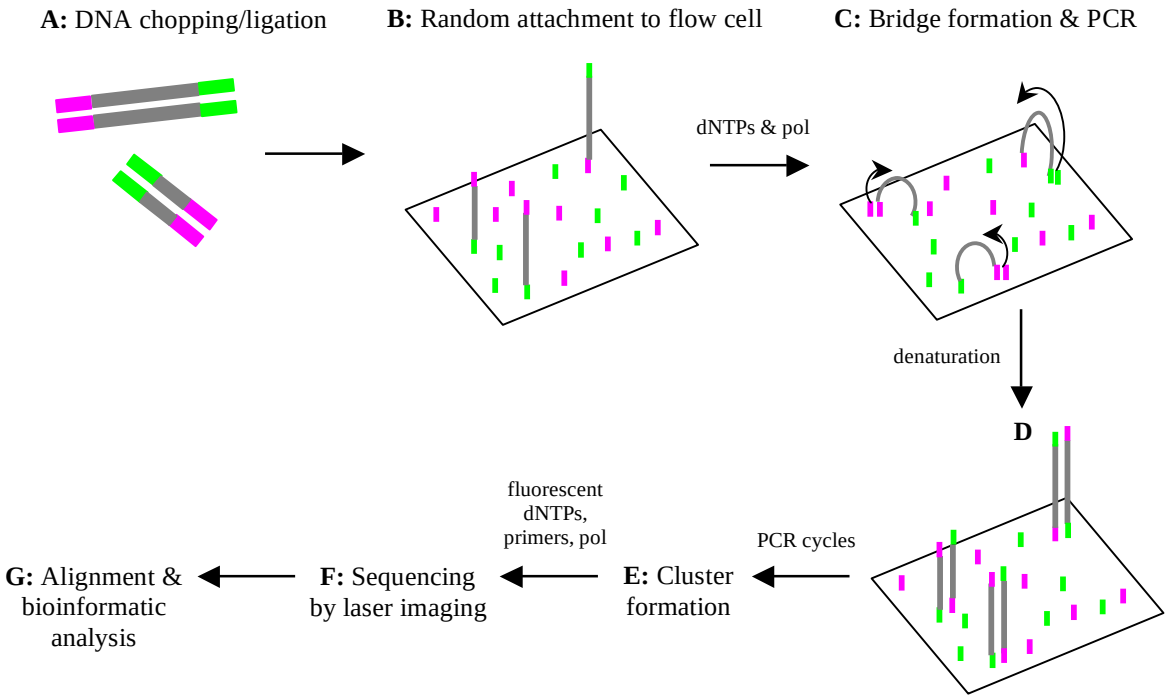


Figure 5.1: A mechanistic description of PCR-based Solexa sequencing technology (Illumina Inc.). See text for details (pol=DNA polymerase).

5.2 Aims

1. To elucidate all genetic discrepancies between the ancestral SBW25 and derived 1w^4. This will be achieved using Solexa genome re-sequencing technology.

2. To determine the chronological order of all mutational events in the history of 1w^4. To this end, PCR and Sanger sequencing techniques will be employed.

3. To identify the mutation(s) required for *operation* of the 1w^4 phenotypic switching mechanism, using genetic manipulation techniques.

4. To investigate which mutation(s) are required for the *evolution* of switching. To this end, evolutionary and competition experiments will be employed.

5.3 Results

5.3.1 Identification of mutations in 1w^4

In order to identify the mutational differences between *P. fluorescens* SBW25 and 1w^4, the genome of 1w^4-reN1.4 (a derivative of 1s^4 containing the first eight mutations of 1w^4, see section 5.3.5.1) was sequenced using the Solexa method. An overnight culture of 1w^4-reN1.4 was produced, and separated into non-capsulated and capsulated fractions by centrifugation (13,000 x g, 10 minutes). Separate genomic DNA extractions were performed on each fraction, as outlined in section 2.2.1. Following quantification, the DNA extracted from the cap- and cap+ fractions was mixed in a 1:1 ratio, ensuring the best chance of discovering any genetic difference between the cap- and cap+ fractions. Subsequently, the DNA was sent for Solexa genome sequencing, and results were analyzed as described in section 2.2.5.2. Nine mutations were identified, eight of which were present in 1w^4: point mutations in each of *wssA*, *wspF*, *awsR* and *mwsR* (x4), and a 33 bp deletion in *awsX* (Table 5.1; see section 6.3.1.2 for information on the ninth 1w^4-reN1.4 mutation). The ninth mutation in the 1w^4 genome, a point mutation in *carB*, was identified during an earlier, partially successful attempt at identification of mutations in 1w^4 (performed in conjunction with Dr. Hubertus Beaumont and Roche NimbleGen). The presence of each of the nine mutations in both the non-capsulated and capsulated fractions of 1w^4 was demonstrated by PCR amplification and Sanger sequencing of each locus, from cap- and cap+ genomic DNA separately (see sections 2.2.2.1 and 2.2.5.1). PCR conditions used for each locus are detailed in Table 5.1.

Gene	Pflu#	Mutation[a]	Primers	Temp[b]	Time[c]	Enzyme	Notes
wssA	0300	164insA	WssAf/WssAr	58	60	Taq	-
wspF	1224	157insG	WspF2f/WspF2r	59	60	Taq	CES[e]
awsR	5210	A1141C	Aws8f/Aws18r	60	60	Taq	-
awsX	5211	Δ229-261	Aws11f/Aws16r	55	60	Platinum PFX	-
carB	5256	C2020T	CarBf/CarBr	60	30	Taq	-
mwsR	5329	G2383A	Mws6fw/MwsR2r	60	30	Taq	-

Gene	Pflu#	Mutation[a]	Primers	Temp[b]	Time[c]	Enzyme	Notes
mwsR	5329	ΔC2553	Mws6fw/MwsR2r	60	30	Taq	-
mwsR	5329	G2778A	MwsR2f/Mws8r	60	30	Taq	-
mwsR	5329	C3094G	MwsR2f/Mws8r	60	30	Taq	-

Table 5.1: Conditions for PCR-amplification of mutations in the 1w^4 evolutionary series. [a]Mutation details provided to distinguish between mutations in the same gene (see Table 5.2 for explanation of nomenclature). [b]Annealing temperature (˚C). [c]Extension time (seconds). [d]Type of polymerase used. [e]Reaction requires addition of CES (section 2.1.3). Mutations listed in order of spatial (not temporal) occurrence in the genome.

5.3.2 Determining the chronology of mutations in line one

Presumably, each of the nine bouts of selection gave rise to one of the nine mutations identified above (see Figure 1.14). Since the 'winning' ancestral genotype was frozen at each round, the order in which the mutations arose could be determined. According to the method outlined in section 2.2.1, genomic DNA was extracted from each of the eight ancestral genotypes (1w^0, 1s^1, 1w^1, 1s^2, 1w^2, 1s^3, 1w^3 and 1s^4). The primer pairs listed in Table 5.1 were used to PCR-amplify the mutated loci in the ancestral genotypes (see section 2.2.2.1). Sanger sequencing (see section 2.2.5.2) revealed the temporal point of occurrence of each mutation. Detailed in Table 5.2 and Figure 5.2, the results revealed that each round of selection saw the acquisition of a single (additional) mutation. Individual mutations are discussed in the following sections.

Strain	Evolution[a]			Gene		Nucleotide change	Amino acid change	Morph[b]
	Cyc	*Rev*	*Days*	*Pflu#*	*Name*			
1s^0	0	0	-	-	-	-	-	SM
1w^0	1	1	3	5329	*mwsR*	G2778A	M926I	WS
1s^1		2	6	5329	*mwsR*	G2383A	E795K	SM
1w^1	2	3	3	5211	*awsX*	Δ229-261	Δ77-87 (ΔYTDDLIKGTTQ)	WS

Strain	Evolution[a]			Gene		Nucleotide change	Amino acid change	Morph[b]
	Cyc	*Rev*	*Days*	*Pflu#*	*Name*			
1s^2		4	6	5210	*awsR*	A1141C	T381S	SM
1w^2	3	5	3	1224	*wspF*	157insG	M52ins(236)[c]	WS
1s^3		6	3	0300	*wssA*	164insA	S54ins(28)[d]	SM
1w^3	4	7	3	5329	*mwsR*	C3094G	R1032G	WS
1s^4		8	12	5329	*mwsR*	ΔC2553	ΔD851(2)[e]	SM/WS
1w^4	5	9	9	5265	*carB*	C2020T	R674C	SW

Table 5.2: Details and chronological order of mutations in the line one evolutionary series. [a]Evolutionary details of each genotype, including the cycle (Cyc) and reversal (Rev) of the REE, and the number of days (Days) taken for the genotype to evolve. [b]Indicates the phenotype of the genotype on the basis of colony morphology on KB agar, ability to form mats in KB microcosms and calcofluor staining (SM=smooth, WS=wrinkly spreader, SW=switcher). [c]M52ins(237) and [d]S54ins(28) indicate frame shifts caused by insertion of a single nucleotide; the number of new residues before a stop codon is reached is in parentheses (in [d]S54ins(28), the residue immediately downstream of S54 (A55), remains unchanged despite the frameshift). [e]ΔD851(3) indicates a frame shift caused by deletion of a single nucleotide; the number of new residues before a stop codon is reached is in parentheses.

5.3.3 Molecular relationships between genotype and phenotype in line one

The first eight mutations in the line one evolutionary series are in loci previously identified as important in the WS phenotype: *wss*, *wsp*, *aws* and *mws* (see section 1.2.1.5). As a result, the morphological effect of the first eight mutations can be predicted with a high degree of accuracy at the molecular level. The final mutation of the series is in *carB*, a highly conserved gene not previously implicated in any phenotype of *P. fluorescens* SBW25. However, some insight into the molecular effects of this mutation was gained through the extensive characterization of this enzyme in other bacterial species (reviewed in Holden *et al.*, 1999). The following sections describe how the molecular effects of each mutation generate the observed phenotype.

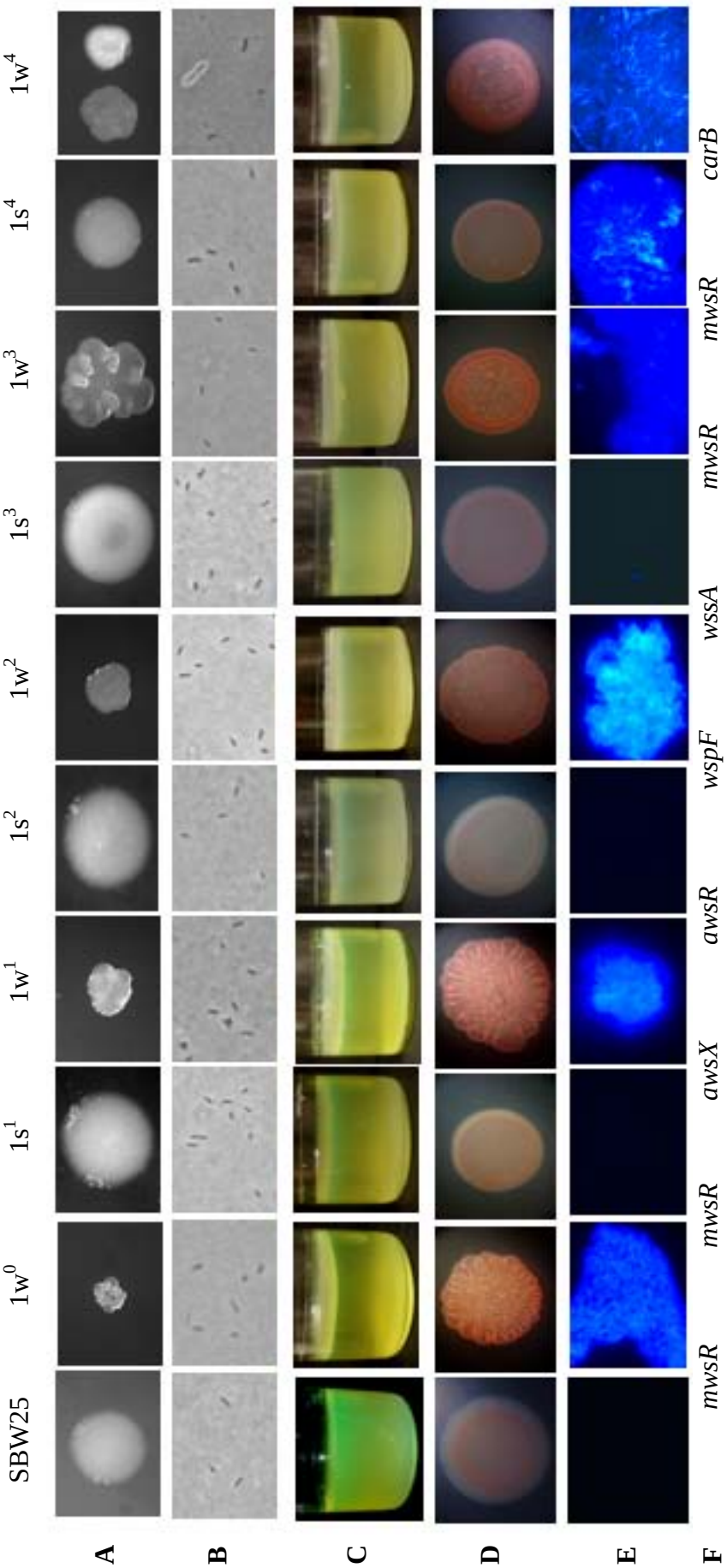


Figure 5.2: Phenotypic and genotypic analysis of the line one evolutionary series. (**A**): Colony morphology on KB agar (48 hours), photographs taken at same magnification. (**B**): Light microscope images of cells counter-stained with India ink (x63). (**C**): Niche preference in a 48 hour static microcosm. (**D**): Dissection microscope images of colony morphology on Congo red agar (48 hours). (**E**): Fluorescence microscope images of ACP-mediated calcofluor binding (x40). (**F**): Gene containing mutation responsible for phenotypic change between each genotype of the series.

5.3.3.1 Cycle one

5.3.3.1.1 Mutation one: mwsR *G2778A*

The first mutation in the evolutionary series caused a change from the smooth colony morphology of SBW25 (1s^0) to the WS-like colony morphology of 1w^0. The mutational origin of this phenotypic change is a transition in *mwsR* (G2778A), which causes a non-synonymous amino acid substitution (M926I) (Table 5.2). Illustrated in Figure 5.3, this substitution occurs in the GGDEF domain, at the amino acid immediately upstream of the GGDEF motif (see Table 1.4). An analysis of the 39 GGDEF domain-containing proteins in *P. fluorescens* SBW25 revealed an isoleucine (I) immediately upstream of the GGDEF motif in six (~16 %), and an alternative aliphatic amino acid (glycine, alanine or valine) occupying this site in a further 12 (~32 %). This indicated that, at least in some cases, IGGDEF is a viable DGC motif. Furthermore, the G2778A *mwsR* mutation was also identified as the mutational cause of a WS-like phenotype in line three of the REE (C. Kost & H.J.E. Beaumont, personal communication). Combined with the WS-like phenotype of 1w^0, this information suggests that G2778A causes a gain-of-function in the *mwsR* GGDEF domain, which in turn is predicted to cause an increase in the production of c-di-GMP and ACP (McDonald *et al*., 2009).

5.3.3.1.2 Mutation two: mwsR *G2383A*

The second genotype of the line one evolutionary series, 1s^1, contains two mutations – the G2778A mutation described above, and a new mutation. Like G2778A, this second mutation is a transition in *mwsR* (G2383A). Illustrated in Figure 5.3, the mutation causes a non-synonymous amino acid substitution in the third PAS domain (E795K; Tables 1.4 & 5.2). As the sole genetic difference between the WS-like 1w^0 and smooth-like 1s^1, this amino acid change is the probable cause of a reduction in *mwsR*-induced ACP biosynthesis (McDonald *et al*., 2009). While the precise mechanism of this reduction remains unknown, it is likely to involve a reduction in signal sensing and/or transmission.

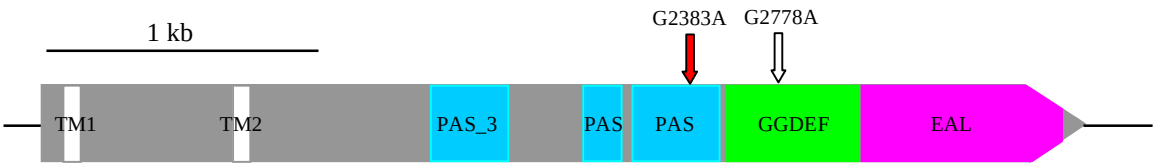


Figure 5.3: Domain structure of *mwsR* and position of mutations one and two. The 1,283 residue MwsR contains two trans-membrane (TM) helices, three PAS domains, a GGDEF domain and an EAL domain. White arrow indicates the approximate position of G2778A (mutation one). Red arrow indicates the approximate position of G2383A (mutation two).

5.3.3.2 Cycle two

5.3.3.2.1 Mutation three: awsX *(Δ229-261)*

The third mutation of the series, a 33 bp deletion in *awsX* (Δ229-261), first appears in the WS-like genotype 1w[1]. The deletion, which occurs between two direct 6 bp repeats (5'-ACCCAG-3'), results in the in-frame deletion of 11 amino acids in the mature AwsX protein (Table 5.2; Figure 5.4). This precise deletion has been previously reported as sufficient to cause a WS phenotype in the SBW25 background (McDonald *et al*., 2009). Presumably, the mutation affects the ability of AwsX to repress the DGC activity of AwsR, and the resulting increase in c-di-GMP and ACP production causes the WS-like phenotype (Gehrig, 2005; McDonald *et al*., 2009).

5.3.3.2.2 Mutation four: awsR *A1141C*

1s[2] contains the fourth mutation of the series – a transversion in *awsR* (A1141C). This results in a non-synonymous amino acid substitution (T381S) in the GGDEF domain of AwsR (Figure 5.4). Given that this mutation is the sole genetic difference between 1w[1] and 1s[2], it is probable that it results in a decrease in AwsR DGC activity induced by mutation three. The corresponding reduction in c-di-GMP and ACP production reverts the phenotype of 1s[2] to a smooth-like state (Gehrig, 2005; McDonald *et al*., 2009).

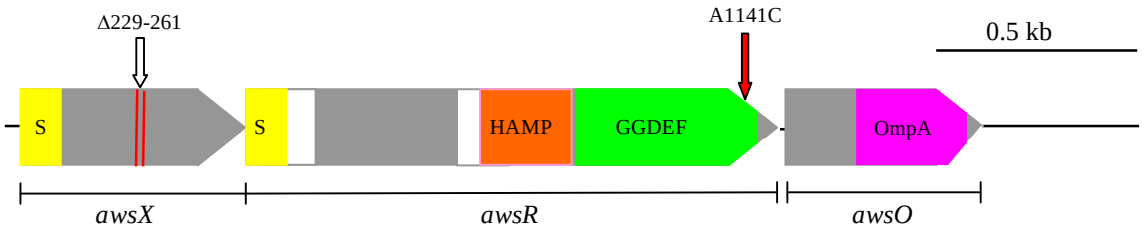


Figure 5.4: Domain structure of the AwsXRO and position of mutations three and four. The 573 bp *awsX* gene contains a proteolytic cleavage site (signal peptide, 'S'). The 1,263 bp *awsR* gene contains a proteolytic cleavage site, two transmembrane helices (white), a HAMP domain and a GGDEF domain. The 492 bp *awsO* gene contains an OmpA domain. White arrow indicates the approximate position of Δ229-261 (mutation three); red lines indicate approximate position of the 6 bp repeats between which the deletion occurred. Red arrow indicates the approximate position of A1141C (mutation four).

5.3.3.3 Cycle three

5.3.3.3.1 Mutation five: wspF *(157insG)*

First occurring in 1w[2], mutation five is a point insertion in *wspF* (157insG). This insertion creates a frame shift in WspF, altering the amino acid sequence from residue 52 in the RR receiver domain, and generating a stop codon after 236 altered residues (Figure 5.5). A considerable number of loss-of-function mutations WspF have been shown to cause the WS phenotype in SBW25 background (Bantinaki *et al.*, 2007). However, this is the first report of an insertion causing a WS-like phenotype. It is highly probable that the frameshift mutation causes a loss-of-function of WspF, which in turn causes constitutive activation of the DGC, WspR (Spiers *et al.*, 2002, Bantinaki *et al.*, 2007). The resulting over-production of c-di-GMP and ACP is presumed to cause the WS-like phenotype of 1w[2].

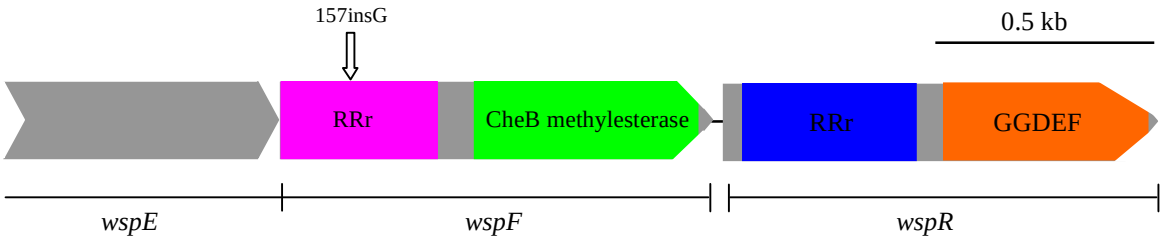


Figure 5.5: Structure of the *wspFR* locus and position of mutation five. The 1,011 bp *wspF* gene contains a response regulator receiver (RRr) domain and a methylesterase domain. The 1,002 bp *wspR* contains an RRr domain and a GGDEF domain. Part of *wspE* (upstream of *wspR*) is shown. White arrow indicates the approximate position of the 157insG (mutation five).

5.3.3.3.2 Mutation six: wssA *(164insA)*

The sixth mutation, a point insertion in *wssA* (164insA), is found in 1s^3 (Table 5.2, Figure 5.6). The insertion causes a frameshift that alters the amino acid sequence of WssA from residue 55, until a stop codon is reached after 28 altered amino acids. While this mutation appears to reduce WspR-mediated ACP production, the repression is not complete (see Figure 5.2E). Presumably, the reduction in ACP production results from loss of WssA function. WssA is thought to play a role in the localisation of the ACP-biosynthetic Wss operon (Kahn, 1998; Spiers *et al*., 2003). If this is the case, then presumably incorrect localisation of the Wss operon indirectly causes a reduction in ACP production in 1s^3. Notably, it is likely that the Wss operon is still functional, and produces the small amounts of calcofluor binding material observed in 1s^3. These results are consistent with the hypothesis that WssA performs a role in the sub-cellular localisation of Wss.

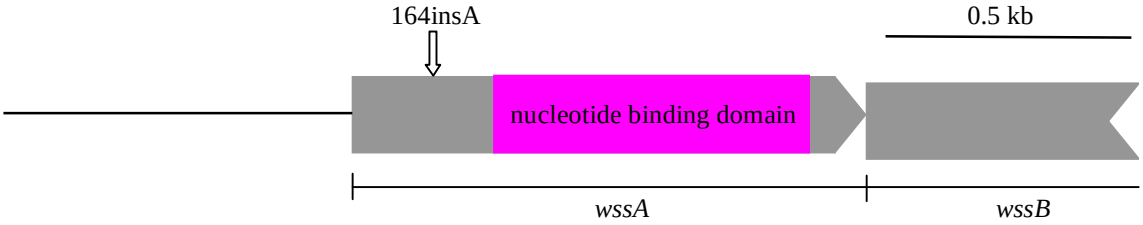


Figure 5.6: Structure of the *wssA* locus and position of mutation six. The 1,035 bp *wssA* gene contains a CbiA nucleotide binding domain, and shows similarity to MinD. The upstream *wss* promoter and downstream *wssB* are shown. White arrow indicates the approximate position of 164insA (mutation six).

5.3.3.4 Cycle four

5.3.3.4.1 Mutation seven: mwsR *(C3094G)*

The seventh mutation is a transversion in *mwsR* (C3094G), giving rise to 1w^3. This mutation causes a non-synonymous amino acid change (R1032G) in the EAL domain of MwsR (Table 5.2, Figure 5.7). Together, the ACP-producing phenotype of 1w^3 and the location of this mutation indicate a loss of PDE activity in the EAL domain, leading to over-production of c-di-GMP by the GGDEF domain.

5.3.3.4.2 Mutation eight: mwsR *(ΔC2553)*

A point deletion in *mwsR* (ΔC2553), the eighth mutation causes a frameshift in the GGDEF domain of MwsR (Table 5.2, Figure 5.7). Given that this frameshift disrupts the GGDEF before the DGC GGDEF motif, one might expect this to cause a complete removal of the ACP (McDonald *et al.*, 2009). However, despite a change to smooth-like colony morphology, 1s^4 still produces a significant amount of ACP, and also colonizes the air-liquid interface (Figure 5.2). This suggests that the removal of c-di-GMP is incomplete; possibly the source of c-di-GMP is WspR, which is still active, (see section 5.3.3.3). However, why WspR-mediated c-di-GMP production should cause ACP synthesis in 1s^4 and not 1s^3 remains unclear.

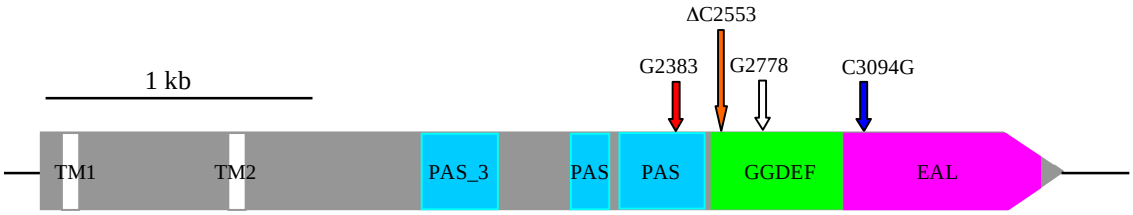


Figure 5.7: Domain structure of *mwsR* and position of mutations one, two, seven and eight. The 1,283 residue MwsR contains two trans-membrane (TM) helices, three PAS domains, a GGDEF domain and an EAL domain. White arrow indicates the approximate position of G2778A (mutation one). Red arrow indicates the approximate position of G2383A (mutation two). Blue and orange arrows indicate approximate positions of C3094G (mutation seven) and ΔC2553 (mutation eight).

5.3.3.5 Cycle five

5.3.3.5.1 Mutation nine: carB *(C2020T)*

Not present in 1w^4-reN1.4, the final mutation in the series is a transition in *carB* (C2020T), which changes monomorphic 1s^4 into dimorphic 1w^4. The *carB* gene (Pflu5265) encodes the long chain of a housekeeping enzyme called carbamoyl phosphate synthetase (CPSase). The small chain of CPSase is encoded by *carA* (Pflu5266), located transcriptionally upstream of *carB*. In SBW25, the *carAB* locus is transcriptionally coupled with *greA* (Pflu5264) and Pflu5263. Although the organization of *carA-carB-greA* is conserved among pseudomonads, the position of Pflu5263 is not (Lu *et al.*, 1997). The *greA* gene encodes a transcription elongation factor, GreA (Borukhov *et al.*, 1992), while Pflu5263 encodes a hypothetical, poorly conserved protein with four predicted transmembrane helices (Table 5.3 and Figure 5.8).

Domain[a]	*E*-value[a]	Residues[b]	Predicted function
CarA (encoded by Pflu5266)			
CPSase_sm_chain	1.4×10^{-91}	3-151	Hydrolysis of glutamate to ammonia
GATase	2.1×10^{-68}	195-373	Transfer of ammonia group from glutamate
CarB (encoded by Pflu5265)			
CPSase_L_chain	7.2×10^{-60}	6-126	Biosynthesis of carboxyphosphate
CPSase_L_D2	1.5×10^{-131}	128-364	ATP binding and hydrolysis
CPSase_L_D3	1.8×10^{-69}	424-547	Oligomerization of subunits
CPSase_L_chain	1.1×10^{-25}	557-671	Biosynthesis of carbamoyl phosphate
CPSase_L_D2	3.3×10^{-17}	673-755	ATP binding and hydrolysis
MGS	4.2×10^{-32}	955-1041	Allosteric regulation by ornithine binding
GreA (encoded by Pflu5264)			
GreA_GreB_N	6.4×10^{-32}	1-74	Transcription elongation factor N-terminus
GreA_GreB	9.7×10^{-32}	80-157	Transcription elongation factor C-terminus
Pflu5263			
None	n/a	n/a	Putative permease protein

Table 5.3: Predicted domain characteristics and associated functions for CarA, CarB, GreA and Pflu5263. [a]Pfam domain searches were performed through the Sanger Pfam website. [b]Amino acid residues incorporated in the indicated domain.

The C2020T mutation causes a non-synonymous amino acid substitution in the second ATP-binding domain of CarB (R674C; Table 5.3, Figure 5.8). A highly conserved, central metabolic enzyme, CPSase synthesizes an intracellular intermediate, carbamoyl phosphate (CP) that is required for the biosynthesis of the amino acid arginine and pyrimidine nucleotides (reviewed in Holden *et al.*, 1999). Interestingly, inactivational insertions in the *carAB* operon have been shown to alter exopolysaccharide production in the Gram-negative extremophile, *Halomonas eurihalina* (Llamas *et al.*, 2003). Precisely how the *carB* mutation results in the 1w^4 bistable capsule phenotype is the subject of Chapter 6.

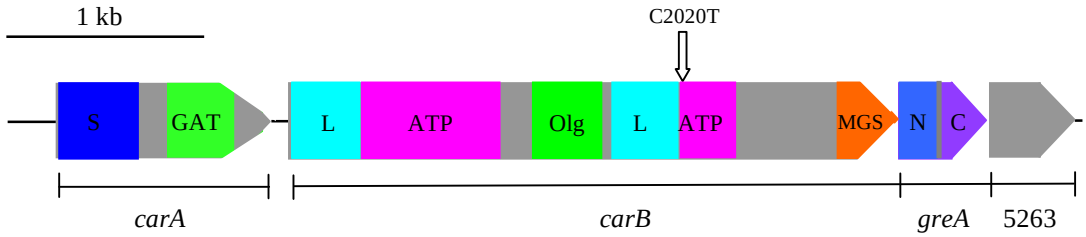


Figure 5.8: Domain structure of the *carA-carB-greA*-Pflu5263 locus, and position of mutation nine. The 1,137 bp *carA* gene contains two domains, the 3,222 bp *carB* gene contains six domains, the 477 bp *greA* gene contains two domains, and the 456 bp Pflu5263 gene contains no recognised domains (see Table 5.3). White arrow indicates the approximate position of C2020T (mutation nine).

5.3.4 Investigating the genetic cause(s) of phenotypic switching

Thus far, the evolutionary history of 1w^4 has been elucidated. The work in this section concentrates on determining which of the nine mutations are required for manifestation of colony and cell dimorphism.

5.3.4.1 Reconstruction of *carB* alleles in various genetic backgrounds

*5.3.4.1.1 Construction of pUIC3-*carB*wt and pUIC3-*carB*mut*

To test the mechanistic role of the *carB* mutation in phenotypic switching, the C2020T and wild type *carB* alleles (named *carB*mut and *carB*wt, respectively) were constructed in various backgrounds. Firstly, separate PCRs were performed using SBW25 (*carB*wt) and 1w^4 (*carB*mut) genomic DNA (see section 2.2.2.1). For each, a DNA fragment of ~1 kb was amplified using the primer pair CarB3f/CarB4r (61˚C annealing temperature, 1 minute extension time). PCR fragments were purified and cloned into pCR8/GW/TOPO for sequencing (see section 2.2.3.1). For both alleles, a clone containing a mutation-free PCR fragment was selected, the inserts retrieved by *Bgl*II digestion, cloned into pUIC3 and the resulting construct used to transform chemically competent *E. coli* cells (see section 2.2.3.2.3). Each set of *E. coli* transformants was screened for incorporation of the construct. The resulting constructs, pUIC3-*carB*wt and pUIC3-*carB*mut, were used to perform *carB* allelic exchanges in various genetic backgrounds, according to the methods outlined in sections 2.2.6.2 and 2.2.7.

5.3.4.1.2 Re-construction of carB*mut in 1s^4 and* carB*wt in 1w^4*

To determine if the *carB* mutation was the sole mutational cause of switching in 1w^4, the *carB*mut allele was constructed in the 1s^4 background, and the *carB*wt allele was constructed in the 1w^4 background. The resulting genotypes, 1s^4-*carB*mut and 1w^4-*carB*wt, were assayed for colony morphology, capsule production and calcofluor binding (Figure 5.9). The phenotype of the re-constructed switcher was identical to that of 1w^4; 1s^4-*carB*mut showed colony and capsule dimorphism, and ACP production. The phenotype of the re-constructed ancestor was identical to that of 1s^4; 1w^4-*carB*wt produced monomorphic colonies and cells, and synthesized ACP. These results demonstrate that the *carB* mutation is necessary and sufficient to cause colony and cell-level dimorphism in the 1s^4 genetic background.

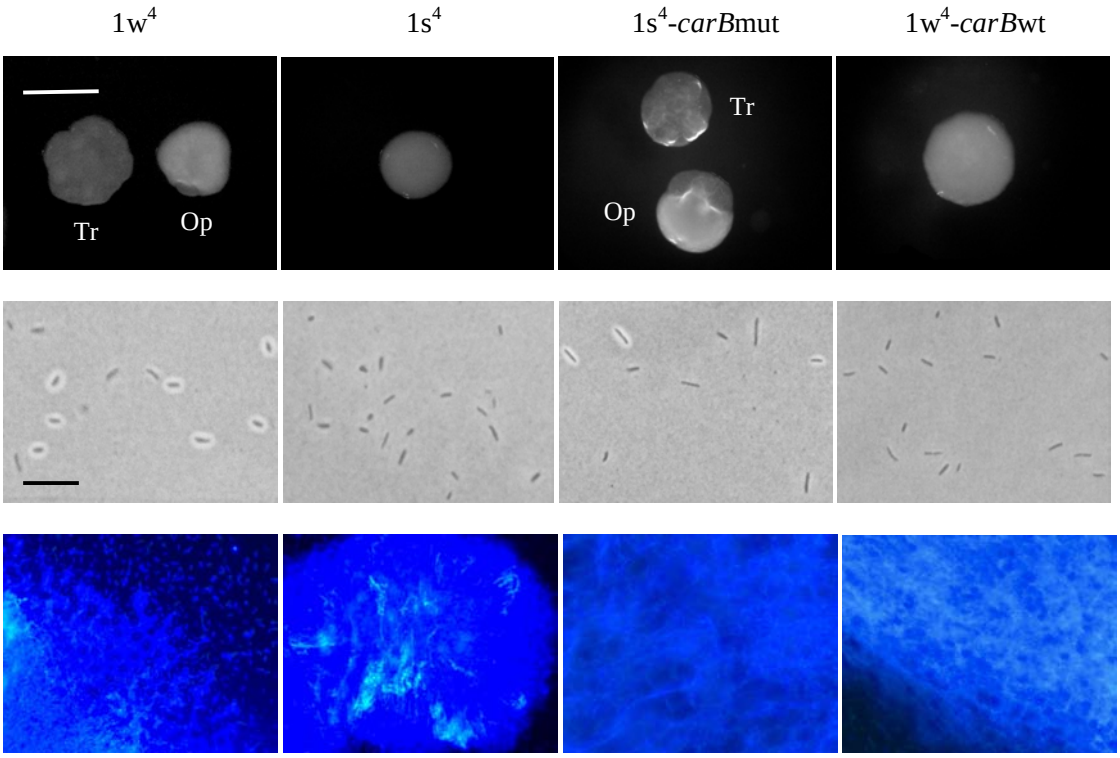


Figure 5.9: Phenotypic characterisation of 1s^4-*carB*mut and 1w^4-*carB*wt. Construction of the *carB* mutation in 1s^4 results in the switcher phenotype, while construction of the wild type *carB* allele in 1w^4 results in a non-switching phenotype: (top: colony morphology on KB agar at 48 hours (scale bar indicates ~3 mm); middle: light microscope images (x40) of India ink counter-stained cells (scale bar indicates ~5 μm); bottom: fluorescence microscope images (x40 or x100) showing calcofluor binding.

5.3.4.1.3 Re-construction of carB*mut in SBW25 and 1w^0*

Next, the dependence of switching on the first eight mutations in the line one series was investigated. The *carB*mut allele was artificially reconstructed in two more distant 1w^4 ancestors: SBW25 (no mutations) and 1w^0 (*mwsR* G2778A mutation), resulting in SBW25-*carB*mut and 1w^0-*carB*mut, respectively. Introduction of the *carB*mut allele into both of these backgrounds resulted in a dimorphic phenotype similar to that of 1w^4 (Figure 5.10). However, the dimorphic colony morphologies of the constructed genotypes were slightly different from those produced by 1w^4; SBW25-*carB*mut produced smooth and opaque colonies, while 1w^0-*carB*mut produced wrinkly and opaque colonies. Unlike 1w^4 and 1w^0-*carB*mut, SBW25-*carB*mut did not bind significant amounts of calcofluor, demonstrating that the *carB* mutation does not cause ACP biosynthesis.

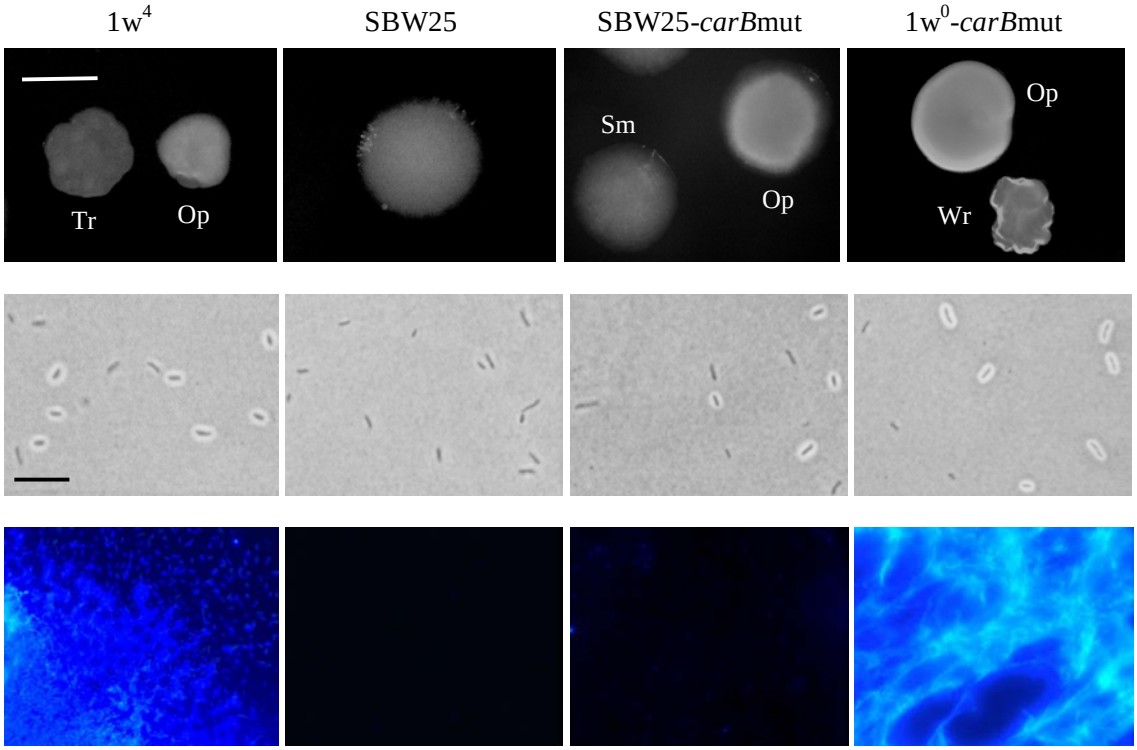


Figure 5.10: Phenotypic characterisation of the artificially constructed genotypes SBW25-*carB*mut and 1w^0-*carB*mut. Construction of the mutant *carB* allele in both genetic background resulted in dimorphism at the colony and cell levels. Top row: colony morphology on KB agar at 48 hours (scale bar indicates ~3 mm). Middle row: light microscope images (x40 magnification) showing India ink counter-staining of cells grown in overnight cultures (scale bar indicates ~5 μm). Bottom row: fluorescence microscope images (x40 or x100 magnification) showing calcofluor binding. See text for details.

5.3.4.1.4 Capsule counting assays for carB *allelic reconstructions*

According to the method outlined in section 2.2.12.4, a capsule counting assay was performed for SBW25, 1s^4, 1w^4, 1s^4-*carB*mut, 1w^4-*carB*wt and SBW25-*carB*mut (Appendix A3.1). Presented below in Figure 5.11 and Table 5.4, the results of this assay provide a number of revelations. Firstly, there is no significant difference between the proportion of capsulated cells in 1w^4 and 1s^4-*carB*mut populations (P=0.278), verifying that the *carB* mutation is the sole cause of the capsulated cell phenotype. This observation was further supported by the fact that no significant difference was found between the proportion of capsulated cells in 1s^4 and 1w^4-*carB*wt populations (P=1.00). Secondly, the proportion of capsulated cells produced in SBW25-*carB*mut populations was significantly greater than in 1s^4 populations (P=7.71 x 10^{-6}), but significantly less than in 1w^4 populations (P=7.84 x 10^{-3}). This suggested that although the *carB* mutation is alone sufficient to cause the capsule switching phenotype, at least some of the additional eight mutations in the 1w^4 background contribute to capsule expression.

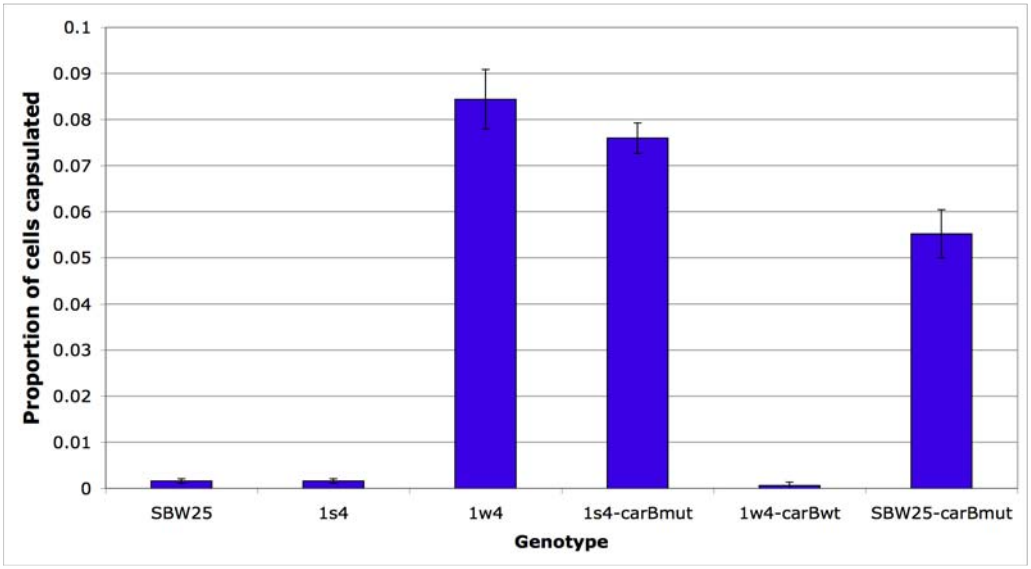


Figure 5.11: Relative proportion of capsulated cells in SBW25, 1s^4, 1w^4, 1s^4-*carB*mut, 1w^4-*carB*wt and SBW25-*carB*mut cultures. Each bar represents the mean of five replicates (error bars indicate one standard error).

Genotype	Mean ± SE	Comparison of population means		
		Vs. genotype 2	*P*-value	95 % C.I.
SBW25	0.0016 ± 0.000748	-	-	-
1s^4	0.0016 ± 0.000748	SBW25	1.00	-
1w^4	0.0844 ± 0.00643	1s^4	1.32 x 10^{-6}***	0.0677, 0.0977
1s^4-*carB*mut	0.0760 ± 0.00323	1w^4	0.278	-
1w^4-*carB*wt	0.00160 ± 0.00748	1w^4	1.32 x 10^{-6}***	0.0677, 0.0977
		1s^4	1.00	-
SBW25-*carB*mut	0.0552 ± 0.00524	1w^4	7.84 x 10^{-3}**	-0.0101, -0.483
		1s^4	7.71 x 10^{-6}***	0.0414, 0.0658

Table 5.4: Relative proportion of capsulated cells in SBW25, 1s^4, 1w^4, 1s^4-*carB*mut, 1w^4-*carB*wt and SBW25-*carB*mut cultures. Mean and standard error (SE) of five replicates are given for each genotype. *P*-values generated from two-sample *t*-tests were used to compare the indicated population means (vs.=versus). Where appropriate, an estimation of the size of the difference is given as a 95 % confidence interval (C.I.). Mean, standard error and *P*-values given to three significant figures.

5.3.5 The first eight mutations play a role in the evolution of switching

The work in section 5.3.4 shows that, mechanistically, the *carB* mutation is the sole cause of the switcher phenotype. Nonetheless, switching took nine rounds of selection to evolve, indicating a possible evolutionary role for the preceding mutations. This possibility was addressed by two evolutionary experiments detailed in the following sections and in Beaumont *et al.* (2009).

5.3.5.1 Differential evolution of dimorphic genotypes from SBW25 and 1s^4

The question of whether the first eight mutations play a role in the evolution of switching was addressed in an evolutionary experiment performed by Dr. Hubertus Beaumont. In this experiment, a single bout of REE selection in the static environment was applied to replicate populations of SBW25 (no mutations) and 1s^4 (eight mutations)

(see section 2.2.9), and the frequency of switcher evolution from both genotypes was recorded. Out of 138 replicate SBW25 populations there were zero occurrences of switcher evolution, while switchers 'won' the evolutionary round in three out of 36 replicate 1s^4 populations. Collectively, these results indicate that SBW25 and 1s^4 differed in their capacity to give rise to colony switching (two-tailed Fisher's exact test, P=0.0083; see Beaumont *et al*. (2009)), suggesting that at least some of the early mutations contribute to the evolution of switching. Additionally, 'losing' switcher genotypes (i.e. observed but not numerically dominant) were isolated from a further three 1s^4 populations, while no non-numerically dominant switcher genotypes were observed in SBW25 replicates. Each of the six switcher genotypes evolved from 1s^4 produced translucent and opaque colonies on KB agar, and a mixture of capsulated and non-capsulated cell types (Table 5.5). For further phenotypic and genotypic characterization of these six genotypes, see section 6.3.1.

Genotype	Numerically dominant	Phenotype		Evolution (days)
		Colony	Cell	
1w^4-reD2	Yes	Translucent and opaque colonies	cap-/cap+	9
1w^4-reD12	Yes	Translucent and opaque colonies	cap-/cap+	9
1w^4-reD1.8	Yes	Translucent and opaque colonies	cap-/cap+	9
1w^4-reN1.2	No	Translucent and opaque colonies	cap-/cap+	9
1w^4-reN1.4	No	Translucent and opaque colonies	cap-/cap+	9
1w^4-reN1.5	No	Translucent colonies with opaque sectors	cap-/cap+	9

Table 5.5: Numerical dominance, colony morphology and cell phenotype of the six additional switcher genotypes evolved in a static environment from 1s^4. Genotype names contain information about numerical dominance; D=dominant, N=not dominant. Evolution column gives time required for evolution of genotype in days (three days growth *per* transfer). See text for details.

5.3.5.2 The effect of preceding mutations on biological fitness of the *carB* mutation

An experiment was devised to further investigate the nature of the evolutionary role played by the first eight mutations. The most parsimonious explanation for the results in

section 5.3.5.1 invokes differential relative fitness of the *carB* mutation in the SBW25 and 1s^4 genetic backgrounds; in order to be observed in the REE, an emerging genotype must rise in frequency, and thus requires a fitness advantage over the dominant (starting) genotype. It is possible that 1w^4 (and 1s^4-*carB*mut) has a selective advantage over 1s^4 in the static environment, while SBW25-*carB*mut has no selective advantage over SBW25. In order to test this hypothesis, fitness experiments were performed with the assistance of Dr. Christian Kost (section 2.2.11.5). Four distinct competition experiments were performed: (A) 1s^4 and 1w^4, (B) 1s^4 and 1s^4-*carB*mut, (C) 1w^4 and 1w^4-*carB*wt, and (D) SBW25-*lacZ* and SBW25-*carB*mut (Table 5.6).

Each experiment involved counting the frequency of competitor genotypes in ten replicate microcosms at the beginning and end of a 72 hour, static incubation (Appendix A3.2, Table 5.6). Most of these frequencies were easily counted using distinct colony morphotypes of competitor genotypes on KB agar. The exception was competition D, where insufficiently distinct morphotypes required the use of a neutral *lacZ* marker on KB agar containing X-gal (Zhang & Rainey, 2007). As detailed in section 2.2.12.2, the relative fitness of the competing genotypes was calculated for each replicate. From these data, the mean relative fitness of genotypes in each competition was obtained (Table 5.6; Beaumont *et al.*, 2009). The results showed a selective advantage for 1w^4 and 1s^4-*carB*mut over 1s^4, and a selective advantage for 1w^4 over 1w^4-*carB*wt. No fitness difference was detected between SBW25-*carB*mut and SBW25-*lacZ*.

Competition	**Competitor 1**	**Competitor 2**	**Mean RF ± 95 % C.I.[a]**	***d.f.*[b]**	***P*-value[c]**
A	1w^4	1s^4	1.16 ± 0.0269	9	2.04 x 10^{-4}***
B	1s^4-*carB*mut	1s^4	1.13 ± 0.0262	8	1.25 x 10^{-3}**
C	1w^4	1w^4-*carB*wt	1.18 ± 0.0296	8	4.00 x 10^{-4}***
D	SBW25-*carB*mut	SBW25-*lacZ*	1.05 ± 0.0814	9	0.519

Table 5.6: Relative fitness of indicated genotypes as determined by 72-hour competition experiments in a static environment. [a]Mean and 95 % confidence interval (C.I.) of relative fitness (R.F.; competitor 1/competitor 2) calculated from nine or ten replicates. [b]Degrees of freedom (d.f.) indicate the number of replicates used for calculations (d.f.=number of replicates-1). [c]*P*-values calculated for a two-tailed one sample *t*-test for a significant deviation of relative fitness values from 1.

5.3.5.3 Importance of the biological environment in switcher evolution

During competition D above, it was noted that WS-types evolved from both SBW25-*carB*mut (white WS) and SBW25-*lacZ* (blue WS). This was an intriguing observation as the evolution of WS types, which have a high fitness in the static environment (Rainey & Rainey, 2003), has the potential to influence the relative fitness of the *carB* mutation. To further investigate this possibility, the competition between SBW25-*carB*mut and SBW25-*lacZ* was repeated, and relative fitness data collected from ten replicates at each of 24, 48 and 72 hours (Figure 5.12A, Appendix A3.2). No significant fitness difference was found between SBW25-*carB*mut and SBW25-*lacZ* at 24 and 72 hours (*P*-values=0.585 and 0.519, respectively), while at 48 hours SBW25-*carB*mut was found to have a significant fitness advantage (one-tailed *t*-test $P=3.37 \times 10^{-3}$). The number of white and blue WS types observed at each time point was recorded, and the proportion of WS in each replicate estimated (Figure 5.12B, Appendix A3.3). The SBW25-*carB*mut fitness decrease at 72 hours correlates with an increase in the proportion of WS types. These results suggest that the evolution of WS types from SBW25 causes a reduction in the relative fitness of the *carB* mutation. This reduction is not seen in competitions with 1s^4, in which all conventional mutational routes to WS have been sequentially removed. Thus, it is proposed that the evolutionary role of the earlier mutations was to remove competitor WS types from the biological environment.

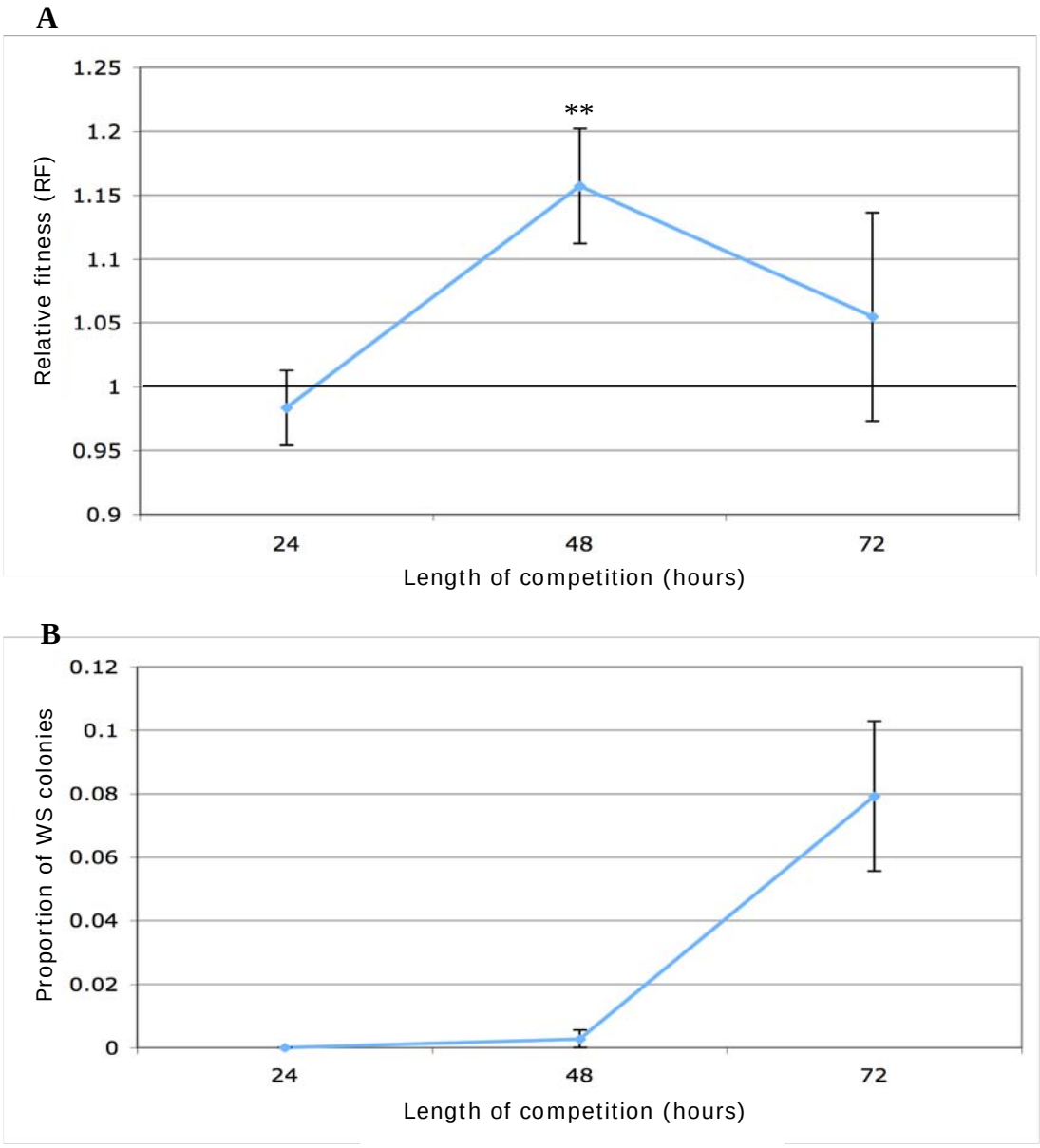


Figure 5.12: Results of 24, 48 and 72-hour competition experiments between SBW25-*carB*mut and SBW25-*lacZ* in a static environment. (**A**) Relative fitness (RF) of competitor genotypes at each time point. An RF of one (solid black line) indicates no difference in fitness; an RF over one (solid black line) indicates a selective advantage for SBW25-*carB*mut (as seen at 48 hours); an RF below one indicates a selective advantage for SBW25-*lacZ*. (**B**) Proportion of observed colonies that are WS at each time point. Both sets of data points are mean values of ten replicates, and all error bars indicate one standard error.

5.4 Discussion

5.4.1 Summary of the 1w^4 genotype

Solexa sequencing revealed that nine mutations occur on the evolutionary path between SBW25 and 1w^4 (Beaumont *et al.*, 2009). Of these, the first eight were in loci previously identified as mutational targets in the evolution of WS, while the final mutation was in *carB*. Genetic analysis revealed that mechanistically, the *carB* mutation is sufficient for manifestation of colony and cell level dimorphism. Subsequent evolutionary studies demonstrated that at least some of the preceding mutations were required for the evolutionary emergence of colony dimorphism under the REE regime. It was proposed that the evolutionary role of these mutations was to remove competing WS genotypes from the REE by sequential inactivation of WS evolutionary targets.

5.4.2 Insights from the mutational series

5.4.2.1 Sub-cellular organisation of Wsp, Aws and Mws

A notable feature of the line one mutational series is that each cycle of evolution results in a pair of mutations in the same genomic locus (with the exception of cycle three). To a certain extent, paired mutations were expected as a result of the opposing environments of the REE; a mutation with a high a fitness in the static environment was expected to have a low fitness in the subsequent shaken environment. However, it was unknown whether the fitness effects of a particular mutation could be effectively reversed only by mutations in the same locus. The high degree of modularity observed suggests that although Wsp, Aws and Mws pathways all activate Wss, each pathway plays a subtly different role in the WS phenotype. Based on preliminary microscopic evidence that WspR localizes to the cell poles (Bohannon, 2002; Gallie, 2005), the currently favoured model is that the sub-cellular localisations of Wsp, Aws and Mws differ. Consistent with this model, inactivation of WspR-mediated ACP biosynthesis by the 1s^3 *wssA* mutation did not prevent subsequent re-activation of MwsR-mediated ACP biosynthesis. Together with the predicted localisation role of WssA (see section

1.2.1.4), these observations imply that the *wssA* mutation does not affect enzymatic activity of Wss, but rather prevents co-localisation of Wsp and Wss.

5.4.2.2 Reconciling genotype and phenotype in 1w^4

The non-modular nature of the cycle three *wspF/wssA* mutations has important implications for the 1w^4 phenotype. The frameshift mutation in *wspF* (1w^2) almost certainly disrupts WspF function, thereby rendering c-di-GMP production by WspR constitutive (see section 1.2.1.5.1). Given that no further *wsp* mutations were observed in the mutational series, it is likely that WspR is constitutively active for the remainder of the evolutionary series. If this is the case, all genotypes from 1s^3 onwards (including 1w^4) are likely to produce elevated levels of c-di-GMP, which may cause a multitude of phenotypic effects, including both capsule and ACP expression (reviewed in Römling *et al*., 2005). Firstly, the transposon mutagenesis screen of Chapter 4 implicated c-di-GMP in capsule expression (see section 4.4.3). Secondly, it is likely that WspR-mediated c-di-GMP production is at least partially responsible for ACP production in 1s^4 and 1w^4. Given the deleterious nature of the final *mwsR* mutation (occurs in 1s^4) with respect to ACP synthesis, the unexpected ACP production in these two genotypes may be explained by WspR activity. The molecular details of why WspR might activate Wss in 1s^4 and 1w^4 and not 1s^3 remain speculative. However, the WspR-mediated activation of Wss *only* after mutation of MwsR indicates the existence of previously unrecognised interactions between the Mws and Wsp signalling systems.

At this stage, the molecular connection between the *carB* mutation and the capsular phenotype remains ambiguous. However, it is worth noting that the biosynthesis of a number of polysaccharides requires the polymerisation of sugar precursors activated by the addition of uridine nucleotides (see section 4.4.3). Included in this group is colanic acid, which requires the polymerization of a four nucleotide-sugar precursors (see section 4.4.2.1). Thus, the pyrimidine nucleotide biosynthetic pathway is a possible link between *carB* and capsule formation. Indeed, inactivational insertions in the *carAB* operon of the extremophile *H. eurihalina* have been shown to alter exopolysaccharide production (Llamas *et al*., 2003).

5.4.3 Insights into the molecular mechanism of phenotypic switching

A pivotal finding of this chapter was that the switch-causing *carB* mutation was present in both the non-capsulated and capsulated forms of 1w^4. Thus, switching between non-capsulated and capsulated forms is not caused by gain and loss of the causal mutation. Rather, the *carB* mutation is indirectly responsible for the switching phenotype, generating a genetic environment in which switching may occur. In itself, this observation does not necessarily indicate an epigenetic switch mechanism, as the *carB* mutation could prime the genome for further genetic change. For instance, if the *carB* mutation reduces (without eliminating) the availability of arginine and/or pyrimidines in the cell, it is conceivable that cyclic amplification and reduction of mutant *carB* gene copies could generate phenotypic switching (Figure 5.13; see section 6.3.3). Amplification and reduction of a mutant *lac* locus in response to selective pressure has previously been documented (Andersson *et al.*, 1998; reviewed in Roth & Andersson, 2004).

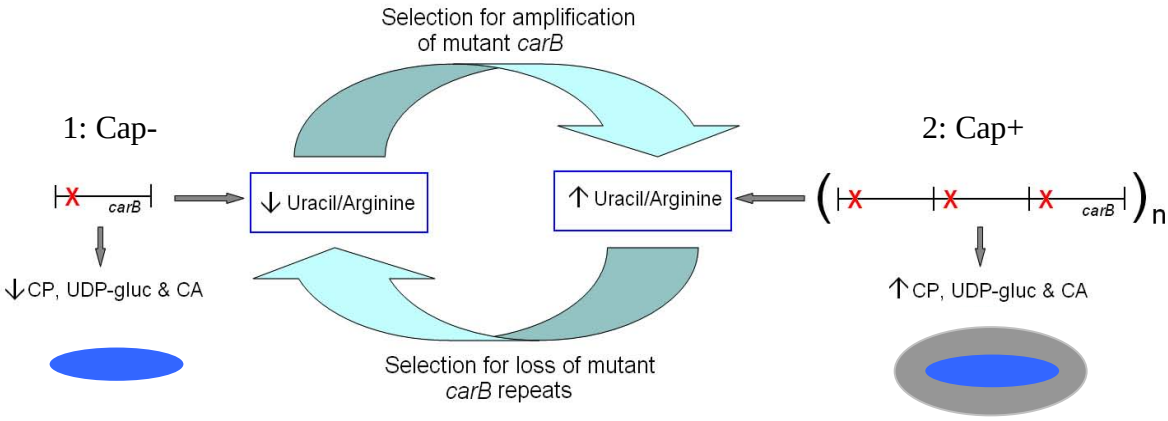


Figure 5.13: The amplification-reduction model of 1w^4 phenotypic capsule switching. In the first phenotypic state (cap-), a single copy of mutant *carB* (mutation represented by the red cross in *carB*) results in a deficiency in uracil and/or arginine. A selective advantage is provided by amplification of mutant *carB* (*via* recombination) through the resulting increase in uracil and/or arginine production. In the second phenotypic state (cap+), the supply of uracil and/or arginine is plentiful, and there is no selective pressure for maintenance of mutant *carB* repeats. Thus, the repeats are lost *via* recombination, returning the cell to the first phenotypic state. Model based on that proposed by Andersson *et al.* (1998). CP=carbamoyl phosphate, CA=colanic acid.

Chapter 6:
The Molecular Mechanism of 1w^4 Phenotypic Switching

6.1 Introduction

The key finding of Chapter 5 was that the *carB* mutation indirectly causes 1w^4 capsule switching; the mutation 'sets the stage' for unidentified molecular events that directly control the switch. In order to deepen understanding of these events, the molecular effects of the *carB* mutation must be investigated. Normally, the *carB* gene functions to synthesize CP, an important intermediate from which both arginine and pyrimidine nucleotides are synthesized (Figure 6.1). As a central metabolic gene, mutation of *carB* may have a multitude of cascading effects in the cell. The work in this chapter deals with the considerable challenge of identifying these effects, and determining which are relevant to 1w^4 phenotypic switching.

6.2 Aims

The aim of this chapter is to investigate the molecular mechanism by which the 1w^4 phenotypic switch operates. Specifically, the aims of this work are:

1. To investigate the phenotype and genotype of the six additional switching types evolved from 1s^4 (see section 5.3.5.1).

2. To analyze the effects of the identified switch-causing mutations on metabolism using comparative analyses, growth assays, genetic and biochemical studies.

3. To directly test the amplification-reduction model of capsule switching (see section 5.4.3).

4. To define the boundaries of the 'switch unit' within the biochemical pathway.

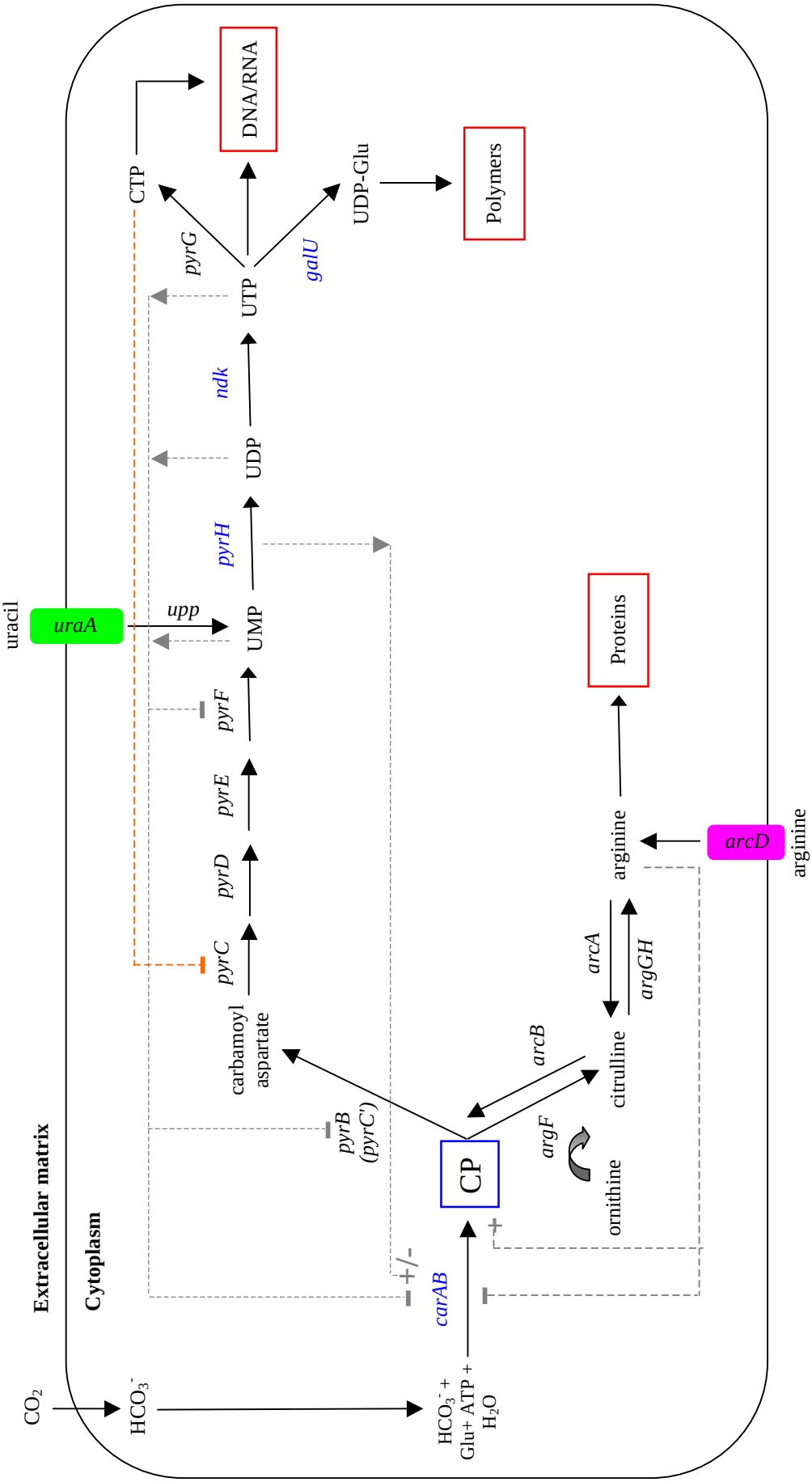


Figure 6.1: Arginine and pyrimidine biosynthetic pathways stemming from carbamoyl phosphate (CP) in *Pseudomonas*. The arginine pathway also operates in reverse. Feedback regulation in each pathway is indicated by dotted lines. Gene names written in blue indicate genes implicated in $1w^4$ capsule switching. Figure drawn with assistance from information in O'Donovan & Neuhard (1970), Llamas *et al.* (2003), Caldara *et al.* (2008) and Turnbough Jr. & Switzer (2008).

6.3 Results

6.3.1 The six independent switchers evolved from 1s^4

In the hope of broadening mechanistic understanding of capsule switching, the six additional switcher genotypes evolved from 1s^4 were investigated. Of these, three were 'winning' switcher genotypes (i.e. the numerically dominant different morph that arose from 1s^4, denoted by 'D'), and three were 'losing' switcher genotypes (i.e. not numerically dominant, denoted by 'N'; see section 5.3.5.1). Characterization of the phenotype and genotype of each switcher genotype is outlined in the sections below.

6.3.1.1 Phenotypic investigation of switcher genotypes

The switcher genotypes were subjected to a similar phenotypic analysis as the line one genotypes; colony morphology, cell morphology, niche preference and calcofluor binding ability were all assayed as outlined in section 3.3.1 (Figure 6.2). At the colony level, the six switchers produced translucent and opaque colonies on repeated sub-streaking. However, there was a noticeable difference in the relative proportions of each type of colony between strains. In particular, 1w^4-reN1.5 gave rise to very few opaque colonies, instead producing mainly wrinkly-like translucent colonies with opaque sectors. Contrastingly, 1w^4-reN1.4 produced a relatively large proportion of opaque colonies. All six genotypes produced both non-capsulated and capsulated cell types, though again in different proportions. Additionally, the capsules produced by 1w^4-reN1.4 appeared significantly larger than those produced by any other strain. Like 1w^4, all switchers preferentially colonized the air-liquid interface of 48-hour static microcosms. With the exception of 1w^4-reN1.4, the switcher strains bound significant amounts of calcofluor on KB agar, signifying ACP production. In summary, while each of the six additional switcher genotypes phenotypically resembles 1w^4, subtle differences exist between strains in the relative proportion of phenotypic forms.

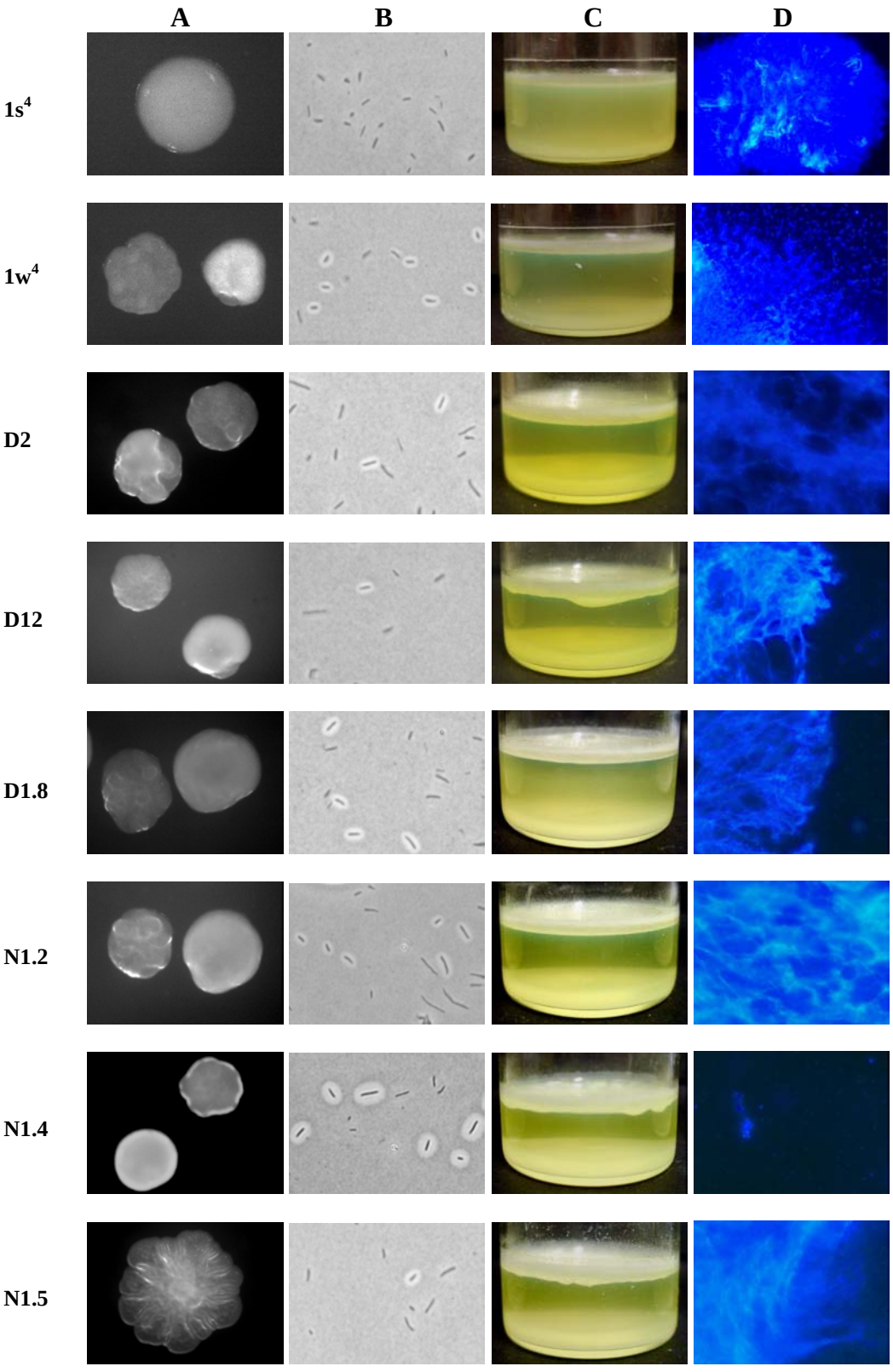


Figure 6.2: Phenotypic analysis of switcher genotypes evolved from 1s^4. (**A**): Colony morphology on KB agar (48 hours); photographs taken at same magnification. (**B**): Light microscope images showing cells from overnight cultures counter-stained with India ink (x40). (**C**): Niche preference in 48-hour static microcosms. (**D**): Fluorescence microscope images of cells after 16 hours growth on KB+calcofluor agar (x40 or x100). Images for each strain are comparable in magnification. Contrast/brightness of some images altered in iPhoto. Names of the re-evolved switcher genotypes are shortened (e.g. D2=1w^4-reD2).

To quantitatively assess differences in capsule expression between switcher genotypes, five replicate overnight cultures of each of 1s^4, 1w^4 and the six additional switcher genotypes were produced, and a counting assay performed (see section 2.2.11.4). Presented in Figure 6.3, the results revealed that although all switcher genotypes produce a significantly greater proportion of capsulated cells than 1s^4 (P<0.05), there are significant quantitative differences in capsule expression between switcher genotypes (Appendix A4.1). Compared to 1w^4, each of 1w^4-reN1.5, 1w^4-reD2, 1w^4-reN1.2 and 1w^4-reD12 produced significantly lower proportions of capsulated cells (P<0.001), and 1w^4-reN1.4 produced a significantly greater proportion of capsulated cells (P=2.53 x 10^{-5}). The proportions of capsulated cells produced by 1w^4 and 1w^4-reD1.8 were not significantly different (P=0.723). Interestingly, the proportion of capsulated cells did not correlate with numerical dominance of the genotype.

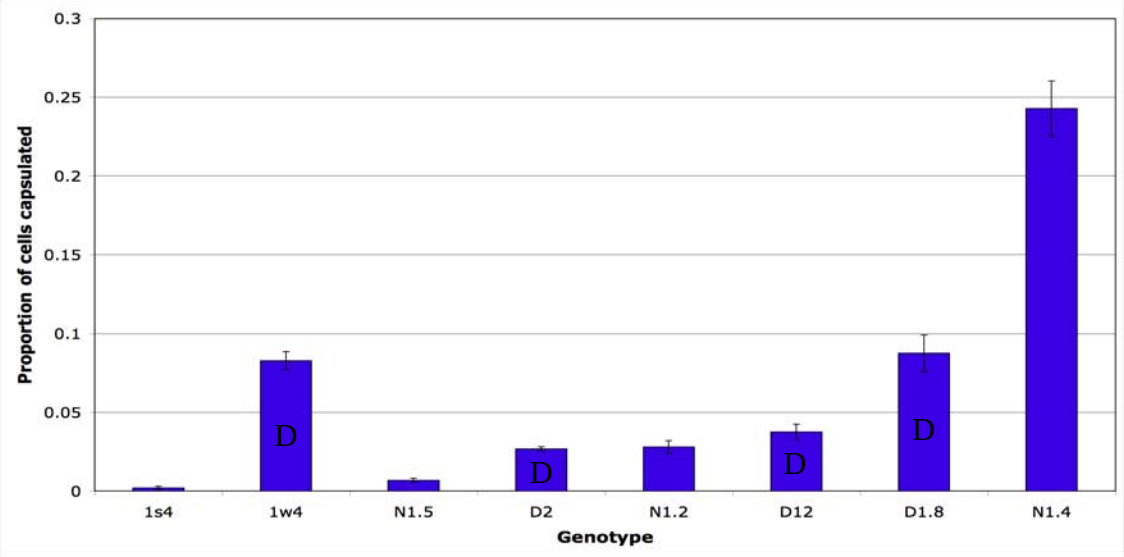


Figure 6.3: Relative proportion of capsulated cells in 1s^4, 1w^4 and the six additional switcher genotypes evolved from 1s^4. Names of the additional switcher genotypes are shortened; '1w^4-re' has been removed from the beginning of each. Numerically dominant switcher genotypes indicated by 'D'. Each bar represents the mean of five replicates, and error bars indicate one standard error.

6.3.1.2 Genotypic investigation of switcher genotypes

Given the similarities between the phenotypes of 1w^4 and the additional switcher genotypes, similarity between the underlying genotypes was investigated. Using the PCR and sequencing techniques outlined in sections 2.2.2.1 and 2.2.5.1, the sequence of

the entire *carAB* locus was determined in each of the six additional switcher genotypes. Seven PCRs were performed on genomic DNA isolated from each of the genotypes, using the conditions outlined in Table 6.1 (Taq polymerase was used for all PCRs).

Primers	**Target**		**Conditions**	
	Gene	**Bases[a]**	**Temperature (˚C)**	**Time (seconds)**
CarABpf/CarAr	*dapB-carA*	5782983-5781651	59	90
CarA2f/CarINTr	*carA-carB*	5781985-5781161	59	60
CarAf/CarB8r	*carA-carB*	5781562-5780694	59	60
CarINTf/CarB7r	*carAB*IG-*carB*	5781091-5780279	59	60
CarB7f/CarB6r	*carB*	5780626-5779581	59	60
CarB3f/CarB4r	*carB*	5779790-5778778	60	60
CarB6f/GreAr	*carB-greA*	5778894-5777870	59	60

Table 6.1: Primer pairs, genomic targets, annealing temperatures and extension times used for PCR-mediated amplification of the *carAB* locus in each switcher genotype. [a]Base numbers refer to SBW25 genome base numbers. IG=intergenic.

Five of the six additional switching genotypes contained a non-synonymous point mutation in *carB* (Table 6.2). Four of these mutations were in ATP binding domains, while the fifth was located between the ATP binding and allosteric domains (Figure 6.4A). Intriguingly, the mutation in 1w^4-reD1.8 was identical to that in 1w^4, a finding consistent with the quantitative similarity in capsule expression between these strains (see section 6.3.1.1). The sixth genotype, 1w^4-reN1.4, did not contain a *carB* mutation. Identified by the Solexa genome re-sequencing performed in section 5.3.1, the switch-causing mutation in this strain is a non-synonymous transition in *pyrH* (Pflu1272; Figure 6.4B). This gene encodes uridylate kinase, an enzyme that operates downstream of CPSase in the pyrimidine biosynthetic pathway (see Figure 6.1). PyrH contains an amino acid phosphorylation domain AA_kinase (E=1.8 x 10^{-64}). Attempts at artificial construction of the *pyrH* mutation in the SBW25 and 1s^4 genetic backgrounds were unsuccessful, probably as a result of severely detrimental *pyrH* disruption.

Genotype	Mutated gene		Nucleotide change	Amino acid change	Primers[a]
	Pflu#	***Name***			
1w^4	5265	*carB*	C2020T	R674C*	CarB3f & CarB6r
1w^4-reD1.8	5265	*carB*	C2020T	R674C*	CarB3f & CarB6r
1w^4-reD2	5265	*carB*	A2477G	N826S	CarB6f & CarB4r
1w^4-reD12	5265	*carB*	G695A	C232Y	CarB7f & CarB7r
1w^4-reN1.2	5265	*carB*	C431T	P144L	CarINTf & CarB8r
1w^4-reN1.4	1272	*pyrH*	C331T	R123C	PyrHf & PyrHr
1w^4-reN1.5	5265	*carB*	C836T	T279I	CarB7f & CarB7r

Table 6.2: Details of nucleotide and corresponding amino acid substitutions causing phenotypic switching in the re-evolved switchers. [a]Forward and reverse sequencing primers with which each mutation was identified (these primers were not always PCR amplification pairs). *The switch-causing mutations in 1w^4 and 1w^4-reD1.8 are identical.

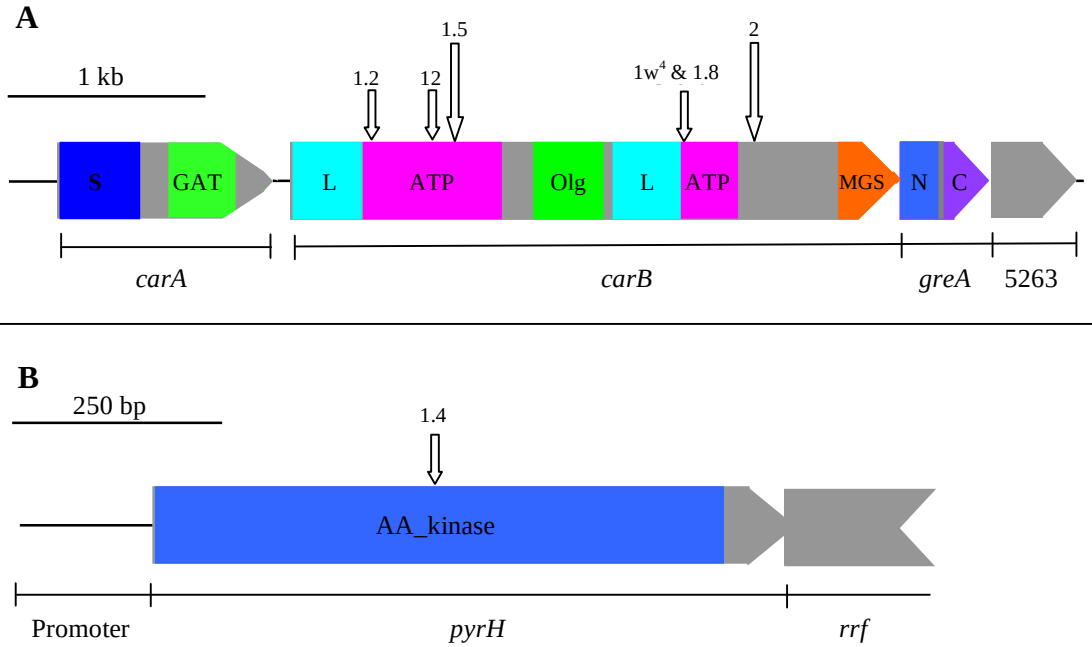


Figure 6.4: Genomic positions of switch-causing mutations. (**A**): Five different *carB* mutations were identified in six different switcher genotypes. (**B**): A single mutation was found in the *pyrH* gene of 1w^4-reN1.4. Genotype names annotated in brief form: '1w^4-reD/N' has been removed from beginning of all names (except 1w^4).

6.3.2 Effects of *carB* mutations

The independent evolution of six mutations in the *carB* gene strongly implicates CPSase and downstream biochemical pathways in phenotypic switching. Since CPSase is a central metabolic enzyme, a mechanism for this is not immediately obvious. In order to gain further insight into the mechanistic operation of the switch, the effects of each mutation were investigated using comparative analyses and growth assays described in the following sections.

6.3.2.1 Insights from comparative structure-function analyses

Fortunately, *E. coli* CPSase and associated biochemical pathways have been extensively analysed, with a range of crystal structures and mutational studies reported (reviewed in Holden *et al*., 1999). These reports allowed investigation of potential effects of the *carB* mutations. Two residues altered by mutation, R674 and N826, are located near the opening of a tunnel through which intermediates are transferred between the active sites of CPSase (Holden *et al*., 1998; Thoden *et al*., 1999; Fan *et al*., 2009). Stapleton *et al*. (1996) reported that mutation of these residues in *E. coli* CPSase (R675A and N827A) resulted in 90 and ~2,000 fold reductions in the maximum velocity (V_{max}) of CPSase activity, respectively. C232 forms part of the CPSase tunnel lining (Holden *et al.*, 1998; Thoden *et al.*, 1999). As such, each of R674C, N826S and C232Y are predicted to create some kind of blockage in the tunnel, slowing the passage of intermediates (J.B. Thoden, personal communication). T279 and P144 do not occur in any specially defined structural region of CPSase; however, considering that the mutations result in substitutions of amino acids with considerably different chemical and physical properties, it is probable that these mutations disrupt the structure and function of CPSase, lowering intracellular levels of CP (S. Powers-Lee, personal communication).

6.3.2.2 Growth assays in minimal media

It was noted earlier that 1w^4 colonies grow slowly on M9 media (section 3.3.2.1). Given that the *carB* mutation is predicted to reduce the function of CPSase, it is possible that

this growth deficiency is due to a reduction in the biosynthesis of arginine and/or pyrimidine nucleotides. In order to test this hypothesis, an experiment was set up to investigate the biochemical requirements of 1w^4. According to the methods outlined in section 2.2.11.2, growth assays were set up for SBW25 and 1w^4 in each of four media: M9, M9 supplemented with 0.6 mM arginine, M9 supplemented with 1 mM uracil, and M9 supplemented with both 0.6 mM arginine and 1 mM uracil. The results are presented in Figure 6.5 and Table 6.3.

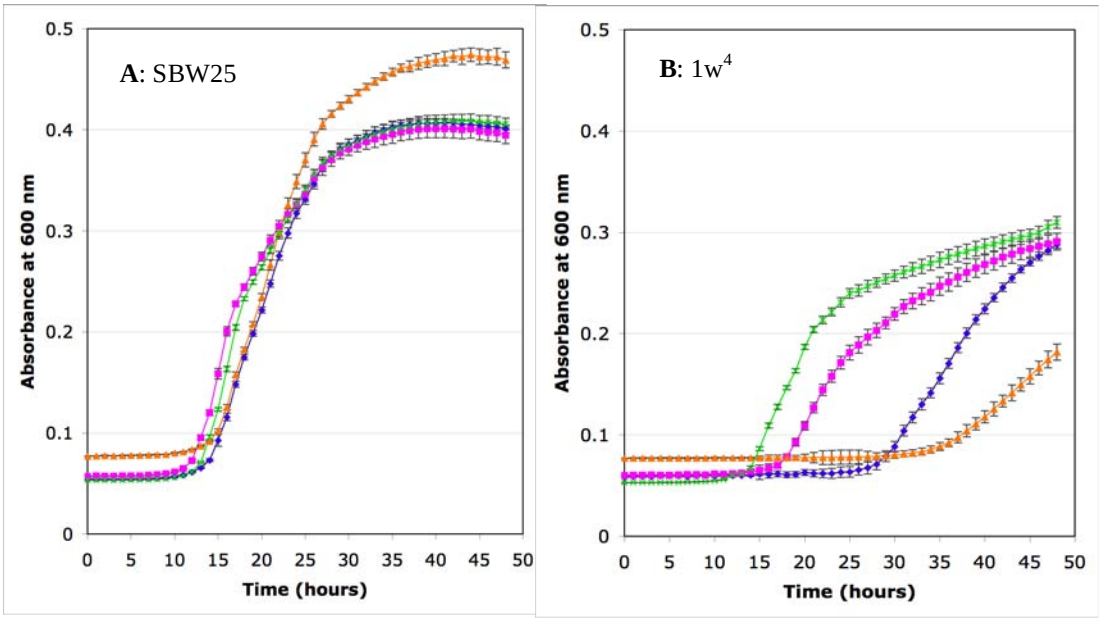


Figure 6.5: Absorbance at 600 nm over 48 hours as a measure of growth rate for SBW25 (A) and 1w^4 (B) in various media. Hourly data values represent the mean of ten replicates, and error bars indicate one standard error. Blue=M9, purple=M9+arginine, orange=M9+uracil, green=M9+arginine+uracil.

The above results demonstrate that, compared with SBW25, 1w^4 growth is reduced in M9, indicating a metabolic deficiency. However, 1w^4 is **not** an auxotroph. This deficiency is partially alleviated by the addition of arginine, or the simultaneous addition of arginine and uracil, while the addition of uracil alone slows 1w^4 growth. These results are a likely reflection of the complex regulation systems governing CPSase biosynthesis and activity; in *E. coli*, the addition of arginine represses transcription and stimulates activity (*via* ornithine binding) of CPSase, while addition of uracil represses both transcription and activity of CPSase (Anderson & Meister, 1966;

Piérard, 1966; Anderson & Marvin, 1968; Piette *et al.*, 1984). Thus, it is probable that the improved growth of 1w^4 on the addition of arginine (or arginine and uracil) results from increased biosynthesis/activity of the mutant CPSase, while reduced growth on the addition of uracil results from a (further) reduction in mutant CPSase activity. Additionally, the inability of large amounts of uracil and arginine to fully restore 1w^4 growth indicates that some of the preceding eight mutations in the 1w^4 background may contribute to the reduced growth of 1w^4 in M9.

Genotype	Medium	Mean Δ OD_{600} ± SE	Relative proportion
SBW25	M9	0.346 ± 0.00397	1
	M9+arginine	0.338 ± 0.00838	1.00
	M9+Uracil	0.392 ± 0.00794	1.13
	M9+arginine+uracil	0.352 ± 0.00581	1.00
1w^4	M9	0.229 ± 0.0188	0.662
	M9+arginine	0.231 ± 0.00481	0.668
	M9+Uracil	0.105 ± 0.00492	0.303
	M9+arginine+uracil	0.257 ± 0.00807	0.743

Table 6.3: Mathematical summary of relative growth rates of SBW25 and 1w^4 in indicated media. Mean change in absorbance over 48 hours (Δ in OD_{600}) ± standard error (SE) was calculated from ten replicates. A relative measure of growth rate was provided by presentation of the mean as a proportion of the mean of SBW25 in M9 (0.346 ± 0.00397); where the mean ± SE encompasses that of SBW25 in M9 the proportion was quoted as 1, otherwise the proportion was calculated as mean/0.346. All values given to three significant figures.

6.3.3 Testing the amplification-reduction model

Given that 1w^4 is not auxotrophic, showing only reduced growth in minimal medium, the amplification-reduction model proposed in section 5.4.3 could conceivably underpin phenotypic switching. In order to test this model, the required *carB* (the amplification target) and *recA* (Pflu1189, required for homologous recombination) genes were

independently removed from 1w^4, and the effects on phenotypic switching observed. Firstly, SOE-PCRs were performed for each gene according to the method outlined in section 2.2.2.2. For the *carB* deletion, two separate DNA fragments of 900 bp and 940 bp were amplified using primers pairs CarB1/2 and CarB5/6 (58˚C annealing temperature, 1 minute extension time). These were then combined and used as a template for the second round of PCR that amplified the 1.8 kb deletion fragment using primers CarB1/6 (59˚C annealing temperature, 2 minutes extension time). For the *recA* deletion, two separate DNA fragments ~550 bp each were amplified using primers pairs RecAKO-1/-2b and RecAKO-3/-4 (60˚C and 59˚C annealing temperatures, 30 seconds extension time). These were then combined and used as a template for the second round of PCR that amplified the ~1.1 kb deletion fragment using primers RecAKO-1/-4 (60˚C annealing temperature, 1 minute extension time). The deletion fragments were cloned into pCR8/GW/TOPO for sequencing, and clones harbouring mutation-free copies of each deletion fragment were selected to independently replace the 1w^4 *carB* and *recA* genes locus by enriched two-step allelic exchange (see sections 2.2.6.2 and 2.2.7). In the case of the *carB* deletion, uracil and arginine were added to all media throughout the exchange protocol. The two resulting strains, 1w^4-Δ*carB* and 1w^4-Δ*recA*, lacked the *carB* and *recA* genes, respectively, in the 1w^4 background.

Phenotypic characterization of 1w^4-Δ*carB* and 1w^4-Δ*recA* included examination of calcofluor binding ability, colony and cell morphology (Figure 6.6). Both 1w^4-Δ*carB* and 1w^4-Δ*recA* produced two types of colonies on KB agar, although 1w^4-Δ*carB* colonies were considerably smaller than those produced by 1w^4, presumably as a result of metabolic deficiency. Both deletion genotypes produced capsulated and non-capsulated cells, and were able to bind calcofluor. A quantitative capsule counting assay (see section 2.2.11.4) demonstrated that neither deletion altered capsule expression levels (P=0.307 and 0.847, respectively; Appendix A4.2). Together, these results demonstrate that 1w^4 phenotypic switching does not result from *recA*-mediated amplification-reduction of *carB*, or any other gene.

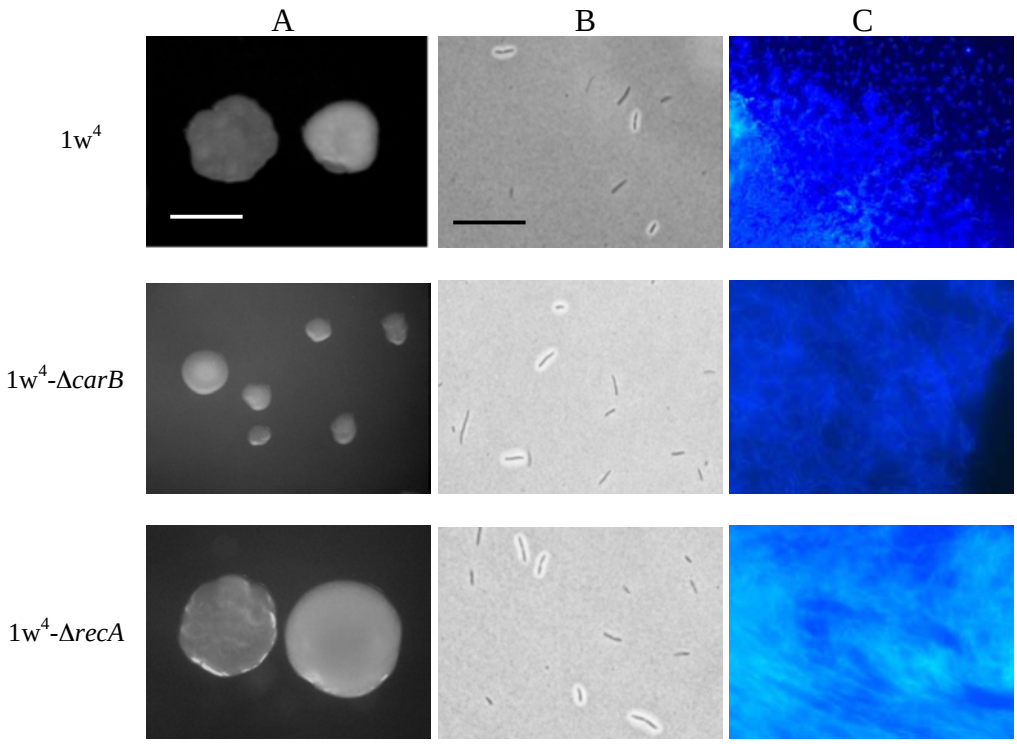


Figure 6.6: Phenotypic characterisation of 1w^4-Δ*carB* and 1w^4-Δ*recA*, 1w^4–derived strains with the *carB* and *recA* genes removed, respectively. Both deletion strains produce two distinct types of colonies on KB agar (48 hours, scale bar represents ~3 mm and applies to all three genotypes; **A**) and cap-/cap+ cells (x40 light microscope images, scale bar represents ~10 μm and applies to all three genotypes; **B**). Additionally, both deletion genotypes retained the ability to bind calcofluor (x40 or x100 fluorescence microscope images, **C**). Images are of comparable magnification between genotypes. Brightness/contrast of some images altered in iPhoto.

6.3.4 The role of arginine and uracil in 1w^4 phenotypic switching

As 1w^4 capsule switching appears to result from a decrease in flux through the CP-arginine and/or CP-UTP pathway(s), it was postulated that addition of the relevant nutrients to the medium would alleviate the metabolic stress and abolish switching. As illustrated in Figure 6.1, Gram-negative bacteria are able to take up arginine and uracil from the extracellular environment, *via* salvage pathways encoded by *arcD* (arginine transport), *uraA* and *upp* (involved in uracil transport and conversion to UMP, respectively). As homologs of these transporters exist in SBW25 (*arcD*: Pflu4890, *uraA*: Pflu0898 and *upp*: Pflu0899, with 46 %, 40 % and 70 % amino acid similarity to their respective *E. coli* counterparts), it was probable that these nutrients could also be salvaged in *P. fluorescens*.

6.3.4.1 Effects of arginine and uracil on 1w^4 phenotypic switching

Colony dimorphism, cell morphology in liquid culture and calcofluor binding by 1w^4 were investigated on four types of medium: KB, KB+arginine, KB+uracil, KB+arginine+uracil (Figure 6.7). In the case of cell morphology, overnight pre-cultures were required in order to acclimatize cells to environmental conditions (see section 2.2.12). Addition of uracil (or arginine and uracil) completely abolished colony dimorphism, and greatly reduced the frequency of capsulated cells. Contrastingly, the addition of arginine alone did not alter colony or cell dimorphism. ACP-production was not noticeably altered by addition of either arginine or uracil.

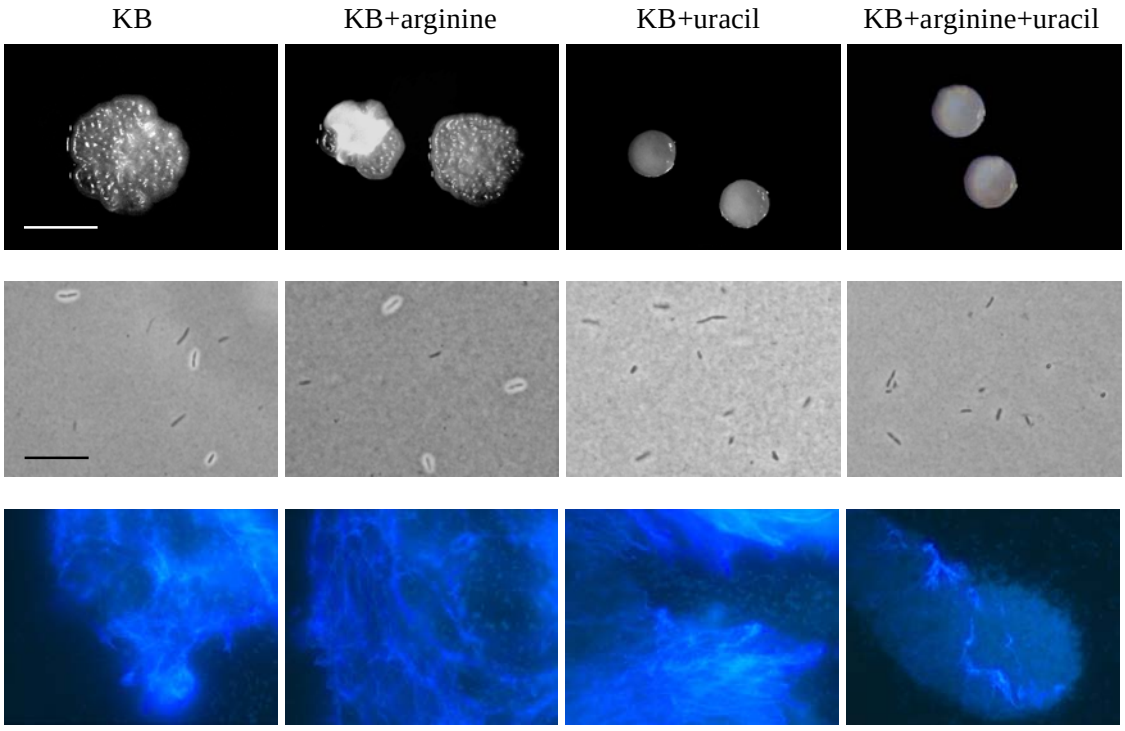


Figure 6.7: Investigation of the effects of arginine and/or uracil addition on the phenotype of 1w^4. Colony morphology at 48 hours (top row, scale bar indicates ~3 mm), cell morphology and capsule production in overnight cultures (middle row; India ink stained cells under x40 light microscope, scale bar indicates ~10 μm), and calcofluor binding (bottom row; x63 fluorescence microscope images). Brightness/contrast of some images altered in iPhoto.

Subsequently, the effects of arginine and/or uracil addition on capsule expression were quantified. A capsule counting assay was performed on 1w^4 cultures containing increasing amounts of arginine and/or uracil (see section 2.2.12.4). Presented in Figure

6.8 and Appendix A4.3, the results demonstrate that the addition of 2 mM arginine had no significant effect on capsule expression (P=0.499), while the addition of uracil or both arginine and uracil significantly reduced capsule expression (P<0.001). Together, these results indicate that a deficiency in the CP-UTP biochemical pathway is responsible for 1w^4 colony dimorphism.

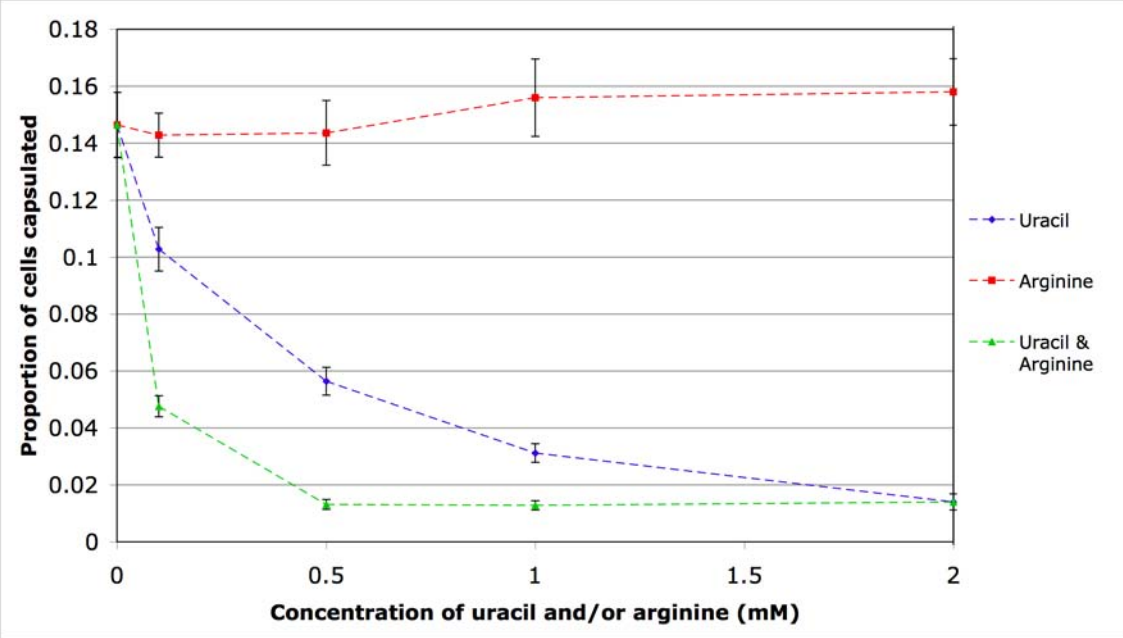


Figure 6.8: Effect of addition of increasing amounts of uracil and/or arginine on capsule expression in 1w^4 populations. The addition of arginine (red) did not alter capsule expression, while the addition of uracil (blue) or uracil and arginine (green) drastically reduced capsule expression. Data points are mean values of five replicates, and error bars indicate one standard error.

6.3.4.2 Effects of uracil on capsule expression in re-evolved switcher genotypes

In order to assess the generality of the conclusions in section 6.4.3.1, the effects of uracil addition on capsule expression in the other switching genotypes evolved from 1s^4 were quantified. A capsule counting assay was performed on 1s^4, 1w^4 and each of the six additional switchers in KB and KB+2mM uracil cultures, with pre-cultures (see section 2.2.12.4). As depicted in Figure 6.9, the addition of uracil significantly reduced capsule production in all genotypes containing a *carB* mutation (P<0.01, see Appendix A4.4). The addition of uracil had no effect on the level of capsule expression in 1s^4 or the *pyrH* switcher, 1w^4-reN1.4 (P=0.771 and 0.778, respectively). This result is not particularly surprising, given that pathway-blocking *pyrH* mutation of 1w^4-reN1.4

occurs downstream of UMP; addition of uracil (which is converted into UMP) would not alleviate flux reduction in this genotype (see Figure 6.1). Collectively, these results lend further support to the hypothesis that switching occurs as a result of a deficiency in the CP-UTP biochemical pathway.

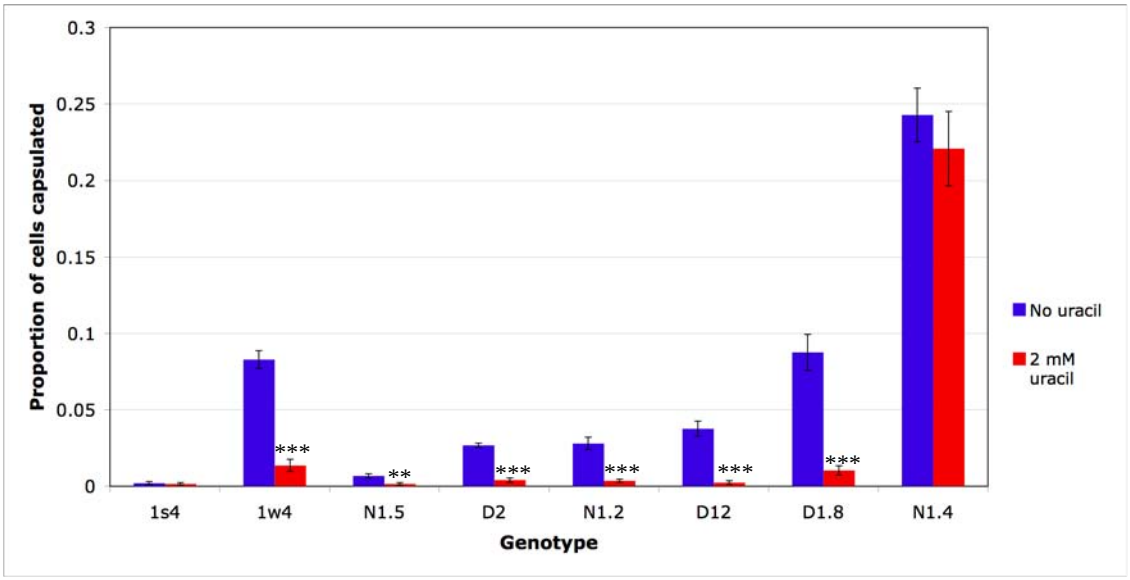


Figure 6.9: Relative proportions of capsulated cells in indicated switcher populations grown in KB and KB+2 mM uracil cultures. Names of the re-evolved switcher genotypes are shortened; '1w^4-re' was removed from the beginning of each. Addition of uracil significantly reduced capsule expression in all genotypes except 1s^4 and 1w^4-reN1.4. Bars represent mean values of five replicates, and error bars indicate one standard error. Stars indicate level of statistical significance (*<0.05, **<0.01, ***<0.001).

6.3.4.2.1 Investigation of an indirect role for arginine in 1w^4 phenotypic switching

The experiments above suggest that arginine is not involved in phenotypic switching. However, as KB medium contains a significant amount of arginine (in tryptone), arginine may play an indirect role in switching; uptake of arginine and conversion to CP *via* the *arc* pathway may contribute to the 1w^4 phenotype (see Figure 6.1). In order to investigate this possibility, an artificial blockage was created in the 1w^4 *arc* pathway by deleting *arcB* (Pflu4892). The *arcB* gene encodes catabolic ornithine carbamoyl transferase (OCT), which converts citrulline into CP and ornithine (anabolic OCT, encoded by *argF* (Pflu1146), catalyzes the forward reaction) (Legrain *et al.*, 1977).

To construct 1w^4-Δ*arcB*, a similar protocol to that described for the *carB* and *recA* deletion strains was used (see section 6.3.3), with the following modifications. For the SOE-PCR, two separate DNA fragments of 570 bp and 490 bp were amplified using primers pairs ArcBKO-1/-2 and ArcBKO-3/-4 (˚C annealing temperature, 30 seconds extension time). These were then combined and used as a template for the second round of PCR that amplified the ~1 kb deletion fragment using primers ArcBKO-1/-4 (58˚C annealing temperature, 1 minute extension time). Phenotypic characterization of 1w^4-Δ*arcB* included examination of calcofluor binding, colony and cell morphology (Figure 6.10). Deletion of *arcB* did not alter 1w^4 colony dimorphism, capsular phenotype or ability to bind calcofluor. A quantitative capsule counting assay (see section 2.2.11.4) demonstrated that 1w^4-Δ*arcB* did not produce a significantly different proportion of capsulated cells to 1w^4 (P=0.723, see Appendix A4.2). Thus, arginine catabolism *via* the *arc* pathway does not appear to play any role in 1w^4 phenotypic switching.

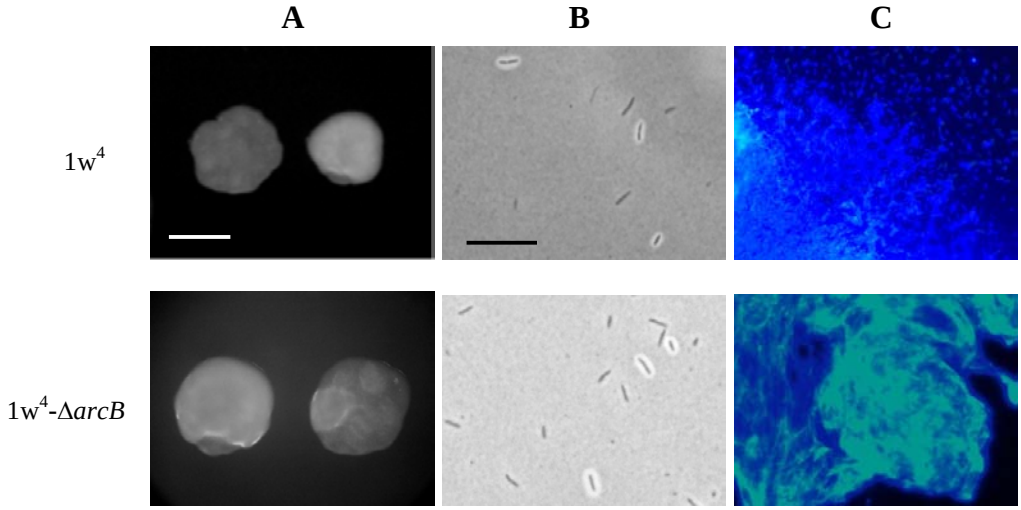


Figure 6.10: Phenotypic characterisation of 1w^4-Δ*arcB*, a genotype constructed by deletion of *arcB* (Pflu4892) from 1w^4. Like 1w^4, 1w^4-Δ*arcB* produced two distinct types of colonies on KB agar (48 hours, scale bar indicates ~3 mm and applies to both genotypes; **A**) and cap-/cap+ cells (x40 light microscope images, scale bar indicates ~10 µm and applies to both genotypes; **B**). Additionally, 1w^4-Δ*arcB* retained the ability to bind calcofluor (x63 fluorescence microscope images, **C**). Images are of comparable magnification between genotypes. Brightness/contrast of some images altered in iPhoto.

6.3.5 Effect of pyrimidine-purine balance on phenotypic switching

From work in the previous sections, it has been elucidated that the cause of capsule switching is a decreased flux through the CP-pyrimidine biochemical pathway. However, rather than low level of pyrimidines *per se*, it is possible that switching results from an imbalance in intracellular pyrimidine and purine pools, which are tightly and co-ordinately regulated under normal cellular conditions (reviewed in O'Donovan & Neuhard, 1970; Chakrabarty, 1998). To investigate this possibility, a capsule counting assay was performed on populations of 1s^4, 1w^4 and 1w^4-reN1.4 in the presence of increasing concentrations of the purine, GTP (see section 2.2.12.4; Figure 6.11; Appendix A4.5). If on/off capsule switching is the result of nucleotide imbalance, the level of capsule expression was expected to increase with the addition of guanine hydrochloride (i.e. upon artificially increasing the imbalance). Indeed, growth in the presence of 1.5 mM guanine hydrochloride significantly increased capsule production in 1w^4 (P=1.5 x 10^{-4}). However, addition of guanine hydrochloride did not alter capsule expression in either 1s^4 or 1w^4-reN1.4 (P>0.1), indicating that capsule switching is not a result of nucleotide imbalance. The increase in 1w^4 is likely to reflect indirect regulatory effects of GTP in the earlier stages of the CP-pyrimidine pathway.

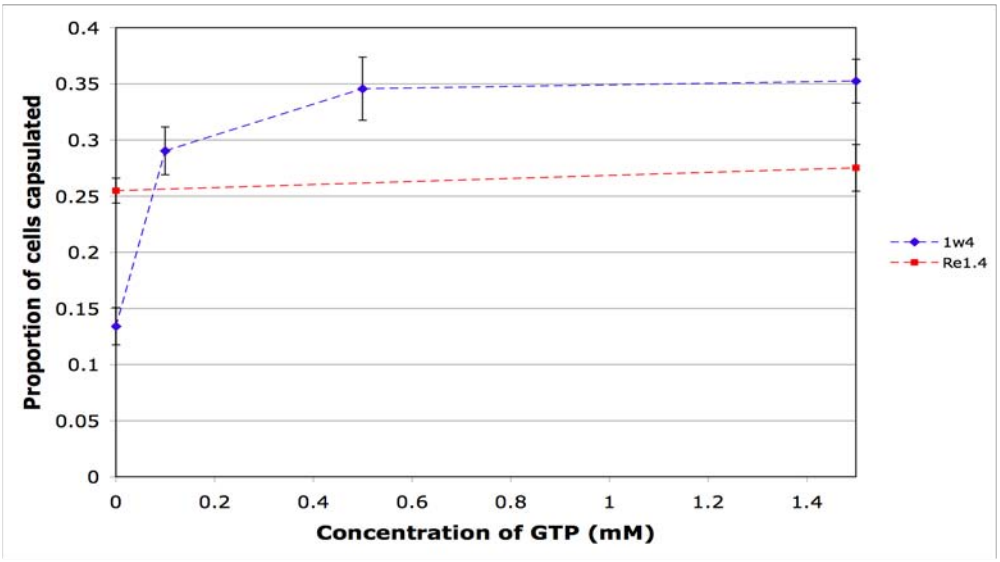


Figure 6.11: Effect of addition of increasing amounts of guanine hydrochloride on capsule expression in 1w^4 and 1w^4-reN1.4 (Re1.4) populations. The addition of guanine hydrochloride increased the proportion of capsulated cells in 1w^4 (blue), but had no effect on 1w^4-reN1.4 (red). Only two data points were obtained for 1w^4-reN1.4: 0 and 1.5 mM guanine hydrochloride. Data points are mean values of five replicates, and error bars indicate one standard error.

6.3.6 Sequential over-expression of UTP biosynthetic genes

The results in previous sections indicate that reduced flux through the pyrimidine biosynthetic pathway activates an epigenetic feedback loop, the operation of which generates phenotypic switching. Extensively studied in a number of species, the UTP biosynthetic pathway is long and complex, with multi-layered and (relatively) poorly understood control mechanisms (see Figure 6.1). Thus, in order to identify potential feedback loops, it was necessary to more closely define the region of the pathway containing the genes and/or intermediates directly responsible for switching (the 'switch unit'). The upstream boundary in the search for the switch unit was indicated by 1w^4-reN1.4 - the ability of a lone mutation in *pyrH* to cause phenotypic switching indicated that reduced flux downstream of *pyrH* was the primary cause of bistability. In an experiment designed to define the second, downstream boundary of the switch unit, genes downstream of *pyrH* were sequentially over-expressed, and the effect on switching observed. The expectation was that if reduction in a particular intermediate contributed to switching, addition of that intermediate (through increased enzyme expression) would alleviate switching. Beyond the downstream boundary of the switch unit, the level of intermediates would have no bearing on switching, and thus over-expression of downstream genes would no longer alleviate switching. To this end, the wild-type *carB* (as a positive control), *pyrH*, *ndk* and *galU* genes were sequentially over-expressed from the pSX plasmid in the 1w^4, SBW25 and 1w^4-reN1.4 backgrounds.

To achieve this, each of the four biosynthetic genes was amplified from SBW25 genomic DNA using primer pairs CarBOE-f/r (Elongase; 56˚C annealing temperature, 3.5 minutes extension time), PyrHOE-f/r (58˚C, 1 minute), NdkOE-f/r (58˚C, 30 seconds) and GalUOE-f/r (57˚C, 1 minute), respectively. Each product was ligated into pCR8/GW/TOPO, and the resulting construct used to transform chemically competent *E. coli*. Clones containing mutation-free fragments were selected, and the fragments retrieved *via* double digestion with *Nde*I and *Bam*HI. Isolated fragments were independently ligated into the pSX vector digested with *Nde*I/*Bam*HI, giving the constructions pSX-*carB*, pSX-*pyrH*, pSX-*ndk* and pSX-*galU*. Along with the empty pSX vector, each construct was used to transform chemically competent SBW25, 1w^4 and 1w^4-reN1.4 cells (see section 2.2.3.1). Transformants were checked for presence of

the insert *via* PCR. This resulted in the following 13 genotypes (three independent biological replicates were produced *per* genotype): SBW25-pSX, SBW25-pSX-*carB*, SBW25-pSX-*pyrH*, SBW25-pSX-*ndk*, SBW25-pSX-*galU*, 1w^4-pSX, 1w^4-pSX-*carB*, 1w^4-pSX-*pyrH*, 1w^4-pSX-*ndk*, 1w^4-pSX-*galU*, Re1.4-pSX, Re1.4-pSX-*carB* and Re1.4-pSX-*pyrH*. Despite several independent attempts, pSX-*ndk* and pSX-*galU* could not be used to transform chemically competent 1w^4-reN1.4, indicating that over-expression of these genes was severely deleterious in the presence of the *pyrH* mutation.

Independent capsule counting assays were performed on each of the three sets of 13 genotypes (see section 2.2.11.4). Except in the cases of control strains (i.e. those not containing pSX or derived construct), 10 μg ml^{-1} Gm was added to the medium. Inducer (IPTG) was not added to the medium, as a preliminary study indicated that phenotypic effects were clearly visible through leaky expression achieved without induction. The mean and standard error of the proportion of cells capsulated in three (non-biological) replicates of each genotype were calculated in each set of 13 genotypes. Following collection of data for all three sets, the means of biological replicates were compared (Appendix A4.6). For every genotype, all biological replicates were comparable – that is, the smallest and largest of the three replicate means and associated standard errors overlapped. Given this, the final 'composite' mean and standard error of the proportion of capsulated cells was calculated using all nine replicates (three non-biological replicates for each of three biological replicates) of each genotype, unless otherwise stated (Figure 6.12, Appendix A4.6). Subsequently, *P*-values for differences between the means of genotypes of interest were calculated using two-sample *t*-tests or, where normality assumptions were not satisfied, M-W-W tests (see section 2.2.12.1).

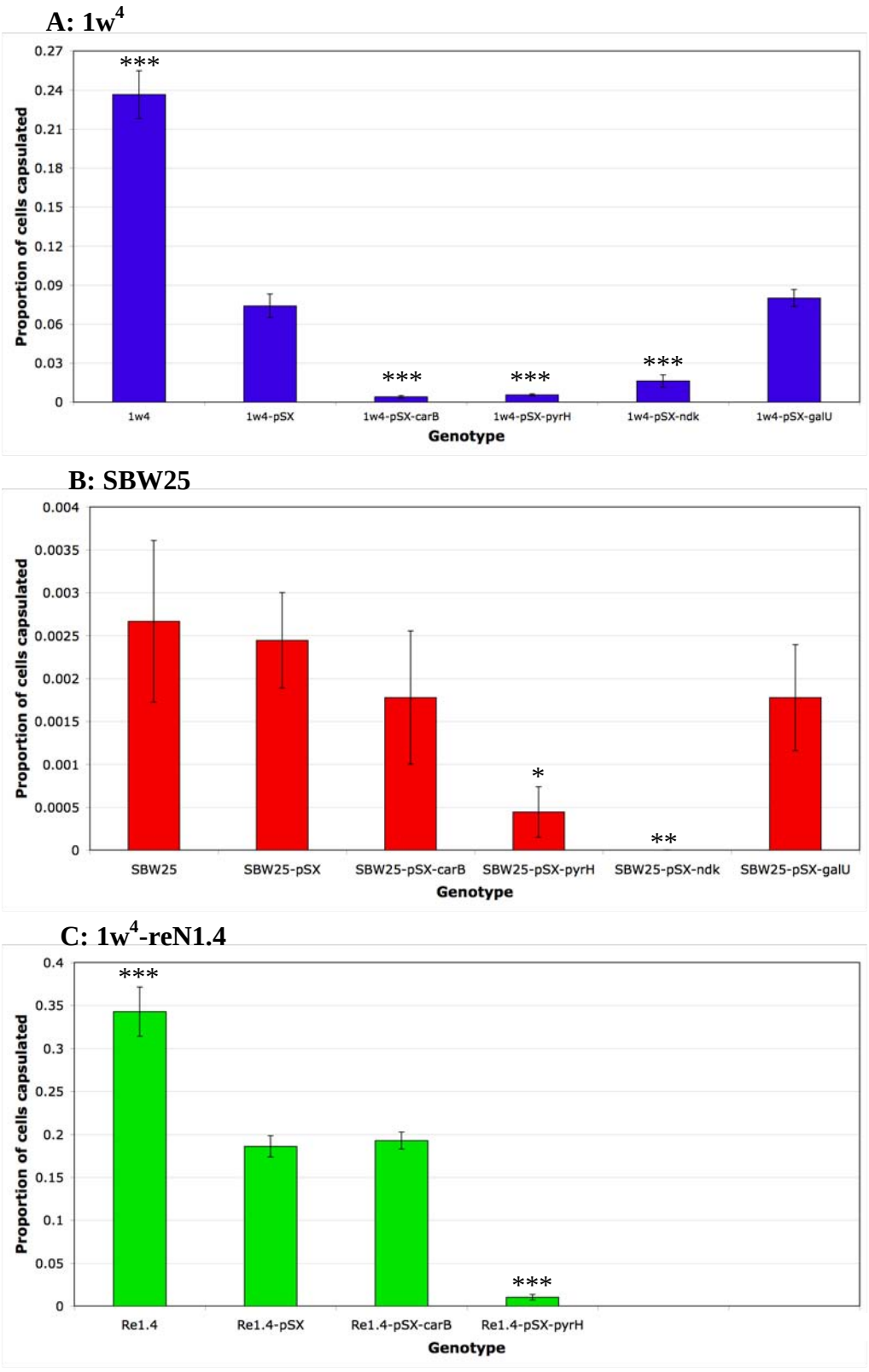


Figure 6.12: Graphs showing the proportion of cells capsulated in 1w^4 (A), SBW25 (B) and 1w^4-reN1.4 (C) populations with the indicated genes over-expressed from pSX. In most cases, bars represent mean values of nine replicates (three non-biological replicates of each of three biological replicate strains; see Appendix A4.6 for details). Error bars indicate one standard error of all nine (biological and non-biological) replicates. Stars indicate a statistically significant difference between the mean and that of the corresponding genotype containing empty pSX (*<0.05, **<0.01, ***<0.001).

The results of this experiment provided significant insight into the molecular mechanisms underlying capsule switching. Firstly, addition of the empty pSX vector significantly lowered the proportion of capsulated cells in both 1w^4 and 1w^4-reN1.4 backgrounds (P<0.001). It is possible that this resulted from Gm addition, as similar effects were noted in the presence of Tc. Alternatively, it is possible that the protocol used to render *P. fluorescens* cells chemically competent (or the subsequent heat shock) caused lasting alterations to the plasma membrane, which in turn alter capsule expression. Given the significant magnitude of the empty pSX effect on capsule switching, this phenomenon should be investigated further. The Gm effect could be investigated by creating a 1w^4-derived genotype with a Gm resistance gene incorporated into the chromosome. If such a strain also demonstrated reduced capsule expression in the presence of Gm (but absence of the pSX vector and transformation protocol, see below), this would support the Gm hypothesis. In order to investigate the heat shock hypothesis, an *ori* locus could be incorporated into the pSX vector to enable conjugation of pSX overexpression constructs from *E. coli* cells, circumventing the need to use transformation protocols directly on *Pseudomonas* genotypes.

Consistent with the hypothesis that the upstream boundary of the switch unit lies beyond *carB*, *carB* over-expression had no effect on capsule expression in SBW25 or 1w^4-reN1.4 (P>0.1), but alleviated switching caused by the *carB* mutation in 1w^4 (P=4 x 10^{-4}). In 1w^4, *pyrH* and *ndk* over-expression both lowered capsule expression below that observed for the control genotype, 1w^4-pSX (P<0.001), while *galU* over-expression did not alter 1w^4 capsule expression (P=0.351). A similar pattern was observed on a smaller scale in SBW25, where *pyrH* and *ndk* over-expression significantly lowered capsule expression (P<0.05), and *galU* over-expression did not (P=0.489). In 1w^4-reN1.4, *pyrH* over-expression significantly reduced capsules (P=0.0168), demonstrating that the *pyrH* mutation causes phenotypic switching in 1w^4-reN1.4 (see section 6.3.1.2). In summary, phenotypic switching in each genetic background is influenced by the over-expression of only two genes: *pyrH* and *ndk.* Collectively, the results of this experiment strongly suggest that the switch unit lies within the *pyrH-ndk-pyrG* segment of the pyrimidine biosynthetic pathway.

6.4 Discussion

6.4.1 Summary of Chapter 6 findings

The work presented in this chapter provides insight into the genetic and molecular-level phenotypic causes of capsule switching. In the first part of the chapter, a second genetic route to capsule switching was identified; in addition to the *carB* mutation, a *pyrH* mutation was found to be capable of causing phenotypic switching. Results from the second part of the chapter provide evidence for the hypothesis that the relevant molecular effect of the *carB* and *pyrH* mutations was to reduce flux through the CP→CTP biochemical pathway. Finally, the over-expression experiments indicate that reduced flux activates an epigenetic feedback loop in the UMP→CTP region of the pyrimidine biosynthetic pathway, which is the direct cause of 1w^4 (and 1w^4-reN1.4) phenotypic switching.

6.4.2 Mutational routes to phenotypic switching: *carB* and *pyrH*

A total of seven capsule switchers were independently evolved from 1s^4. Of these, six were found to contain a mutation in *carB*, while the seventh contained a mutation in *pyrH* (see section 6.3.1). Interestingly, while each mutation caused the capsule switching phenotype, there were differences in the proportion of capsulated cells produced by each genotype. These differences reflect the subtly different molecular effects caused by each mutation. It is possible that quantitative differences in CP→CTP flux reduction result in quantitatively proportional differences in capsule expression; greater flux reductions are predicted to produce greater proportions of capsulated cells.

The observed pattern of six *carB* mutations and one *pyrH* mutation highlights two key points concerning the genetic architecture of the underlying mechanistic pathway. Firstly, there are at least two major mutational routes to phenotypic switching: *carB* and *pyrH*. Secondly, these routes are taken at different frequencies. While this imbalance is likely to be at least partially attributable to a difference in mutational target size - at

3,222 bp, *carB* is ~4.3 times larger than *pyrH* at 744 bp - it may also reflect the organisation of the pyrimidine biosynthetic pathway. Given that the biosynthesis of UTP and CTP is essential for survival, switch-causing mutations must decrease flux through the pathway without abolishing pyrimidine biosynthesis. In other words, only a limited subset of possible deleterious mutations in the pyrimidine biosynthetic genes is able to generate phenotypic switching. Upon examination of the CP→CTP biosynthetic pathway, CPSase (encoded by *carAB*) would appear to be a likely mutational target for flux reduction (see Figure 6.1); it is possible that deleterious mutations in CPSase are partially compensated for by the catabolic arginine pathway (encoded by the *arc* operon), by which extracellular arginine is taken up and broken down into CP independently of CPSase (see Figure 6.1). This hypothesis could be tested directly by analysing the spectrum of mutations among switching genotypes evolved from a 1s^4 derivative in which the catabolic arginine pathway was inoperative. If this salvage pathway indeed acts as 'safety net' for *carB* mutations, one would expect fewer *carB* mutations to arise in such a background.

Three indirect lines of evidence indicate that switch-causing mutations in *carB* are a specific subset of all possible deleterious mutations. Firstly, deletion of *arcB* from 1w^4 did not significantly inhibit the growth of 1w^4 on KB medium (see section 6.3.4.2.1), indicating that the *arc* pathway was not required for CP production in the presence of *carB* mutation C2020T. Secondly, two of the independently evolved switcher genotypes contain the same *carB* mutation (1w^4 and 1w^4-reD1.8). As it is highly unlikely that 1w^4-reD1.8 is a 1w^4 contaminant - isolation of 1w^4-reD1.8 took nine days (and three bottleneck transfers) - this repetition indicates a limited range of possible mutations. Finally, if any loss-of-function mutation in CPSase were sufficient to cause switching, one would expect mutations to occur in *carA*, a gene of 1,137 bp. Despite this considerable target size, no mutations were obtained in *carA*. Deeper insight into the possible mutational routes (and associated molecular effects) leading to phenotypic switching could be obtained by isolating 1s^4 transposon mutants capable of switching. It is likely that such an assay would provide a more complete picture of mutational routes to phenotypic switching, as mutations would not be limited to those that provide a selective advantage under the REE regime.

6.4.3 1w^4 phenotypic switching occurs *via* an epigenetic mechanism

In each genetic route to phenotypic switching, the switch-causing mutation was present in both the non-capsulated and capsulated forms of the relevant genotype (see section 5.3.1 and 6.3.1). This means that switching does not result from rapid gain and loss of the identified mutations. Rather, the *carB* and *pyrH* mutations prime populations for switching. At the beginning of this chapter, it remained uncertain whether the direct mechanistic cause of switching was genetic or epigenetic. In the discussion of Chapter 5, a possible genetic mechanism – the amplification-reduction model based on the work of Andersson *et al*. (1998) – was postulated (see section 5.4.3); a deleterious reduction in CP was proposed to fuel the cyclic gain and loss of mutant *carB* gene repeats, generating respectively the capsulated and non-capsulated switcher forms. The results in section 6.3.3 led to the rejection of this model; phenotypic switching occurs in the absence of the *carB* gene. Furthermore, phenotypic switching in the absence of the *recA* gene (required for homologous recombination leading to gain/loss of repeats) ruled out the possibility of amplification and reduction of *any* gene underlying phenotypic switching (see section 6.3.3).

Despite thorough investigation, there is no evidence of a genetic mechanism underlying 1w^4 phenotypic switching; the transposon mutagenesis screen of Chapter 4 produced no candidate genes involved in a genetic switch, and the powerful genome re-sequencing data of Chapter 5 provided no evidence of a genotypic difference between switcher forms. These results are now further supported by the rejection of the amplification-reduction model. Collectively, this lack of evidence for a genetic switching mechanism is strongly indicative of an epigenetic mechanism underlying 1w^4 phenotypic switching.

6.4.3.1 The epigenetic mechanism of 1w^4 phenotypic switching lies in the pyrimidine biosynthetic pathway

Several lines of evidence indicate that the epigenetic cause of switching is found in the pyrimidine biosynthetic pathway. Firstly, observations of reduced 1w^4 growth in minimal medium (see sections 3.3.2.1 and 6.3.2.2) are suggestive of a non-auxotrophic

metabolic defect. The *carB* mutation implied that the affected pathway was either one (or both) of the pyrimidine or arginine biosynthetic pathways. The first indication that the pyrimidine pathway played a role in switching came from the transposon mutagenesis screen of Chapter 4. In the screen, a transposon mutant was obtained in *ndk*, a gene in later part of the pyrimidine biosynthetic pathway (see section 4.3.1.5.2). Given the central role of *ndk* in pyrimidine biosynthesis, one would logically expect *ndk* mutations to be lethal, or at least highly deleterious. However, it has been shown that loss of *ndk* function can be compensated for by the *adk* (adenylate kinase, Pflu1240; 65 % amino acid identity with *E. coli adk*; Lu & Inouye, 1996). The lack of transposon mutants with insertions in other components of the pathway may be due to an inability to compensate for the severely deleterious or lethal effects of reductions pyrimidine metabolic genes. However, transposon mutants were obtained in several genes indirectly influencing the pyrimidine biosynthetic pathway, including Pflu1304, encoding putative *dcd* (deoxycytidine triphosphate deaminase), which converts dCTP into dUTP (Neuhard & Thomassen, 1971) and *nusA*, encoding transcription elongation factor NusA that has been shown to alter UTP-sensitive transcription attenuation mechanisms during *E. coli pyrBI* transcription (Donahue & Turnbough Jr., 1994). Given the probable role of *nusA* in the regulation of pyrimidine biosynthetic genes, it is possible that *greA* – the transcription elongation factor-encoding gene that is transcriptionally coupled to SBW25 *carAB* – plays a similar role. Transposon mutants that may affect expression of the pyrimidine biosynthetic genes through alterations of the purine nucleotide pool include those in *purU* and *sahA*. No transposon mutants were obtained with insertions in genes obviously linked to the arginine biosynthetic pathway.

Direct evidence for the centrality of the pyrimidine pathway to switching was obtained in section 6.2.1, with the identification of the 1w^4-reN1.4 switch-causing mutation in *pyrH*. This mutation showed unequivocally that changes in the pyrimidine biosynthetic pathway were the cause of phenotypic switching. Additionally, as *pyrH* occurs much later in the pathway than *carB*, the *pyrH* mutation focused attention on the latter part of the pyrimidine biosynthetic pathway. A final piece of evidence underlines the involvement of the pyrimidine pathway - the results of the cross-feeding experiments in section 6.3.4 show that the addition of uracil alleviates phenotypic switching (in *carB* switchers), while the addition of arginine has no effect on switching. In retrospect, this

last observation is not surprising; KB medium contains considerable amounts of arginine (in tryptone), while nucleotides are not present in significant quantities. It is likely that any detected phenotype resulting from a metabolic deficiency originating on KB medium cannot be compensated for by components already present in the medium.

6.4.3.2 Genetics and regulation of the pyrimidine biosynthetic pathway

The expression and operation of the pyrimidine biosynthetic pathway is controlled by a complex network of regulatory systems (see Figure 6.1; reviewed in O'Donovan & Neuhard, 1970; Turnbough Jr. & Switzer, 2008). Extensive study of pyrimidine biosynthesis in a range of bacteria has revealed that while the biochemical pathway is generally conserved, mechanisms of regulation are surprisingly diverse. In particular, there are significant differences in regulation between Gram-positive and Gram-negative bacteria. In *B. subtilis* (Gram-positive), the genes required for UMP biosynthesis lie in a single operon, the expression of which is controlled by the nucleotide-sensitive transcriptional regulator PyrR. In Gram-negative (and some Gram-positive) species, these genes are scattered throughout the genome, and regulatory mechanisms are individually tailored to each. While the expression of some genes in Gram-negative species depends on the PyrR regulator, a number of alternative control mechanisms exist. Rather than utilizing transcription factors, these mechanisms rely on direct sensing of pyrimidine levels by RNA polymerase. They include UTP-sensitive attenuation, re-iterative transcription and CTP-sensitive selection of transcriptional start sites (reviewed in Turnbough Jr. & Switzer, 2008).

As in other Gram-negative species, the pyrimidine biosynthetic genes are widely distributed across the SBW25 genome. Regulation of the *Pseudomonas* pyrimidine biosynthetic genes has been a subject of recent interest and debate (Turnbough Jr. & Switzer, 2008). It appears that the expression of *Pseudomonas* pyrimidine genes is controlled by a mixture of mechanisms, including a mechanism involving an atypical PyrR and mechanisms that affect RNA polymerase directly. As discussed in the following sections, the experiments in this chapter provide novel insight into the regulation of the *P. fluorescens* pyrimidine biosynthetic pathway.

6.4.3.2.1 Regulation of carAB

Encoded by the *carAB* operon (originally named *pyrA*), CPSase converts ammonia or glutamine and bicarbonate into CP, the intermediate from which both pyrimidine nucleotides and arginine are synthesized. Regulation of this operon has been the subject of considerable analysis in both *E. coli* and *S. typhimurium* (Piette *et al.*, 1984; Charlier *et al.*, 1995), and a high level of sequence and organizational conservation suggests similar regulation exists in *Pseudomonas* species (Kwon *et al.*, 1994). Unsurprisingly, the expression of *carAB* is negatively controlled by the end products of the pathways, UTP and arginine. This double regulation is achieved through the differential initiation of transcription from tandem promoters, P1 and P2. Expression from P2 is subject to regulation by arginine; the DNA-binding protein ArgR represses transcription from P2 in the presence of arginine. Expression from P1 is under the control of pyrimidine - and to a lesser extent, purine - nucleotides; transcription from P1 is repressed directly by a large protein complex consisting of integration host factor (IHF), PepA (aminopeptidase A) and PyrH (Kholti *et al.*, 1998). Efficient expression from P1 requires the binding of RutR, a recently discovered pyrimidine sensor (Shimada *et al.*, 2007).

Transcription of *E. coli carAB* has been reported to be under an additional level of control: UTP-sensitive re-iterative transcription (Han & Turnbough Jr., 1998). In this mechanism, elongation of *carAB* transcripts to include the structural gene sequences is dependent upon successful passage of RNA polymerase through a short tract of U residues. Loose pairing between the U tract and the template DNA strand allows for slippage of the transcript during transcription. At high UTP concentrations, repetitive slippage results in the rapid, re-iterative addition of U residues to the 3' end of the transcript. Such transcripts are not extended to contain the structural gene sequences. Low UTP concentrations do not promote reiterative transcription or polymerase slippage, allowing normal transcriptional elongation and translation to occur. For a more detailed mechanistic account of this process, see Turnbough Jr. & Switzer (2008).

In addition to transcriptional control, the activity of *E. coli* CPSase is influenced by intermediates of the pyrimidine and arginine biosynthetic pathways (Anderson &

Marvin, 1968). Firstly, ornithine (the production of which is regulated by arginine) binds to and allosterically activates CPSase. Contrastingly, UMP binding negatively regulates CPSase activity. Thus, the ornithine/UMP balance in a cell is an important factor in determining the rate of CPSase activity.

6.4.3.2.2 Regulation of pyrBC'

In *B subtilis,* the *pyrB* gene is under the control of the nucleotide-sensing regulatory protein PyrR. Mechanistically, PyrR operates by binding to *pyr* mRNA - that is, mRNA transcribed from any *pyr* gene - and promoting the formation of terminator hairpins. The binding of uridine nucleotides to PyrR stimulates mRNA binding, while guanosine nucleotides repress binding. Thus, PyrR acts to balance intracellular pyrimidine and purine levels; *pyr* gene expression is repressed in the presence of uridine nucleotides, and activated in the presence of guanosine nucleotides. Homologs of PyrR exist in *P. aeruginosa*, *P. putida* and *P. fluorescens;* the SBW25 PyrR homolog is encoded by Pflu5757, the gene at the 5' end of the *pyrRBC'* operon (Silby *et al.*, 2009). Interestingly, although a Pfam search shows that the SBW25 protein contains the required PyrR domain (Pribosyltran: residues 1-125, E=3.4 x 10^{-8}), the overall protein sequence differs significantly from that of *B. subtilis* PyrR, including differences at residues thought to be important for mRNA recognition and binding (Turnbough Jr. & Switzer, 2008). In consequence, the mechanism of PyrR action in *Pseudomonas* is presumed to be different to that described for *B. subtilis*.

In *P. fluorescens* SBW25, the *pyrBC'* genes are located downstream of *pyrR*. Notably, *pyrC'* (Pflu5759) differs from *pyrC* (Pflu1154) - although the two are similar in amino acid sequence, *pyrC'* lacks the dihydroorotase activity of *pyrC,* and instead is thought to play a structural role in PyrB function (Brichta *et al.*, 2004). *Pseudomonas* PyrR has been proposed to be involved in the uracil-mediated repression of the *pyrBC'* genes (Turnbough Jr. & Switzer, 2008). Consistent with this hypothesis, addition of guanine to media increased capsule formation in 1w^4, but not 1w^4-reN1.4 (section 6.3.5). The uptake of guanine increases the intracellular purine nucleotide pool, causing further imbalance between purine and pyrimidine nucleotide pools. If the primary effect of this

imbalance was to increase the expression of PyrB (the enzyme that acts immediately downstream of CPSase), one might expect an increase in the turnover of CP (into carbamoyl aspartate). In turn, this could lead to alterations in the expression and/or activity of (mutant) CPSase, and the net effect of increased CP→CTP leading to changes in capsule expression. On the other hand, changes in *pyrB* (and possibly *carAB*) expression would not be expected to alter flux through the pathway in the presence of a downstream blockage caused by a *pyrH* mutation. Thus, guanine salvage has no effect on capsule expression in 1w^4-reN1.4.

6.4.3.2.3 Regulation of pyrC, pyrD, pyrE *and* pyrF

Unlike *pyrB*, there is no evidence supporting a role for PyrR in the transcriptional regulation of any downstream *pyr* genes in *Pseudomonas*. In the absence of any contrary experimental evidence, it is possible (although by no means certain) that these genes are regulated in a similar manner to those in *E. coli* and *S. typhimurium*. In these species, a range of mechanisms involving transcriptional repression in the presence of various pyrimidine nucleotides regulates *pyrC, pyrD, pyrE* and *pyrF* expression. Transcription of the *pyrB* and *pyrE* genes is (independently) regulated by an UTP-sensitive attenuation mechanism controlling transcript elongation from an intrinsic transcription terminator in the gene leader regions. Expression of *pyrC* and *pyrD* is independently controlled by a CTP-sensitive mechanism involving differential selection of transcription start sites. Interestingly, the expression of *pyrF* appears to be repressed by a uridine nucleotide (rather than CTP), but the mechanism of repression remains to be resolved (Schwartz & Neuhard, 1975; Neidhardt *et al.*, 1996: p.587).

6.4.3.2.4 *Regulation of* pyrH, ndk *and* pyrG

Transcriptional regulation of the latter part of the pyrimidine biosynthetic pathway is something of a mystery in Gram-negative bacteria; comparatively little is known regarding the control of *pyrH, ndk* and *pyrG* transcription. In particular, virtually nothing is known of the control of *pyrH* and *pyrG* transcription. However, PyrH directly contributes to the repression of *carAB* transcription in *E. coli* and *S. typhimurium* (see

section 6.4.3.2.1). Genetic studies have shown the expression of *P. aeruginosa ndk* to be regulated by two proteins first recognised for their role in the regulation of alginate biosynthesis: AlgR2 (alternatively known as AlgQ) and AlgH (PA0405) (Schlictman *et al.*, 1995; Bieber Urbauer *et al.*, 2005). Homologs of these exist in *P. fluorescens* SBW25: *algQ* (Pflu5929, BLASTP 62 % amino acid identity with complete *E. coli* AlgR2) and Pflu5755 (BLASTP 73 % amino acid identity with complete *E. coli* AlgH), respectively. Simultaneous inactivation of *algR2* and *algH* completely abolishes both alginate and *P. aeruginosa* Ndk production. Bearing in mind that alginate production requires the polymerization of sugar moieties activated by nucleotide binding (Chakrabarty, 1998), this is suggestive of coordinate regulation of nucleotide biosynthesis and polymerization genes. Therefore, it is possible that proteins involved in the positive regulation of the colanic acid-like polymer of 1w^4 capsules also act as positive regulators of Ndk expression. For example, it is possible that GacA, the DNA-binding RR identified by the transposon mutagenesis screen is required for activation of Ndk transcription. Alternatively, Pflu3655 and/or Pflu3657 (co-ordinately transcribed transcriptional regulators upstream of the colanic acid structural operon) are suitable candidates for such a role (see section 4.3.1).

Ndk is an interesting enzyme with unusually broad specificity – it catalyzes the inter-conversion of a range of nucleoside diphosphates and triphosphates (Stryer, 1988: p.609). The exceptions are adenosine nucleotides; although *ndk* is capable of utilizing and producing ATP, AMP/ADP/ATP inter-conversion is usually carried out by adenylate kinase, encoded by *adk* (Neidhardt *et al.*, 1996: p.587). The action of Ndk is illustrated by Equation 6.1, in which X and Y may represent any of several ribonucleotides or deoxyribonucleotides, including the examples provided by Equations 6.2 and 6.3. A major influential factor in Ndk substrate choice is relative concentrations of the various free nucleotides. The higher the concentration of a particular nucleotide, the more likely it is to be acted upon by Ndk (Ray & Mathews, 1992). Thus, Ndk acts as a kind of buffer for the required relative concentrations of each nucleotide in the intracellular pool. In addition, it has been demonstrated that Ndk possesses ‘enzymic memory’, the phenomenon whereby Ndk molecules under stable environmental conditions are most likely to repeat exactly the previously performed reaction (Katz & Westley, 1980). Enzymic memory is the mechanistic result of briefly maintained, subtly

different enzyme conformations generated by individual reactions performed by a promiscuous enzyme. For further discussion of the role of Ndk in phenotypic switching, see section 6.4.4.2.1).

XDP + YTP ↔ XTP + YDP Equation 6.1

UDP + ATP ↔ UTP + ADP Equation 6.2

GDP + CTP ↔ GTP + CDP Equation 6.3

6.4.4 Epigenetic molecular models for 1w^4 phenotypic switching

Any molecular model proposed to explain switching must satisfactorily account for two aspects of the switching phenotype: (1) the molecular link between the *carB* (or *pyrH*) mutation and capsule formation, and (2) the expression of capsules in some individuals but not others. Each of these aspects is discussed in more detail below, culminating in a model for the molecular basis of phenotypic switching (see Figures 6.13 and 6.14).

6.4.4.1 The molecular link between the pyrimidine biosynthetic pathway and capsule expression

As illustrated in Figure 6.1, the molecular link between switch-causing mutations in the pyrimidine pathway and capsule production is GalU, which catalyzes the reaction UTP + glucose-1-phoshate → UDP-Gluc. UDP-Gluc is a source of all four precursors required for colanic acid-based capsule biosynthesis (see Figure 4.9). Presumably, the colanic acid biosynthetic genes are present and active in switcher cells - possibly as a result of WspR-mediated c-di-GMP production (see section 5.4.2.2) - and convert the precursors into capsule polymer. One might logically expect mutations that reduce UTP biosynthesis to reduce UDP-Gluc and capsule production. Indeed, inactivational insertions in *carAB* of *H. eurihalina* have been shown to eliminate the production of UDP-Gluc-dependent exopolysaccharides (Llamas *et al.*, 2003), and mutations in *P. aeruginosa algR2* and/or *algQ* cause reductions in *ndk* expression and corresponding polysaccharide production (Schlictman *et al.*, 1995).

Paradoxically, the reduction in pyrimidine flux caused by the *carB* and *pyrH* mutations leads to an increase in capsule production, a phenotype that is reversed by the addition of UTP (in the form of uracil, see section 6.3.4.1). In light of the experimental observations in *H. eurihalina* and *P. aeruginosa*, it is possible that the 1w^4 capsulated form does not result from a reduction in UTP *per se*, but rather as the result of excessively high UTP levels generated by mechanisms that overcompensate for the reduction. According to this model, non-capsulated cells contain low to normal levels of UTP, while capsulated cells (of any *P. fluorescens* genotype) contain high levels of UTP; intracellular UTP levels are predicted to fluctuate above (cap+) and below (cap-) a particular level – the 'capsule threshold' that is thought to resemble the 'ComK threshold' discussed in section 1.1.2 – a phenomenon that is magnified when flux through the pyrimidine pathway (and hence UTP) is reduced (Figure 6.13).

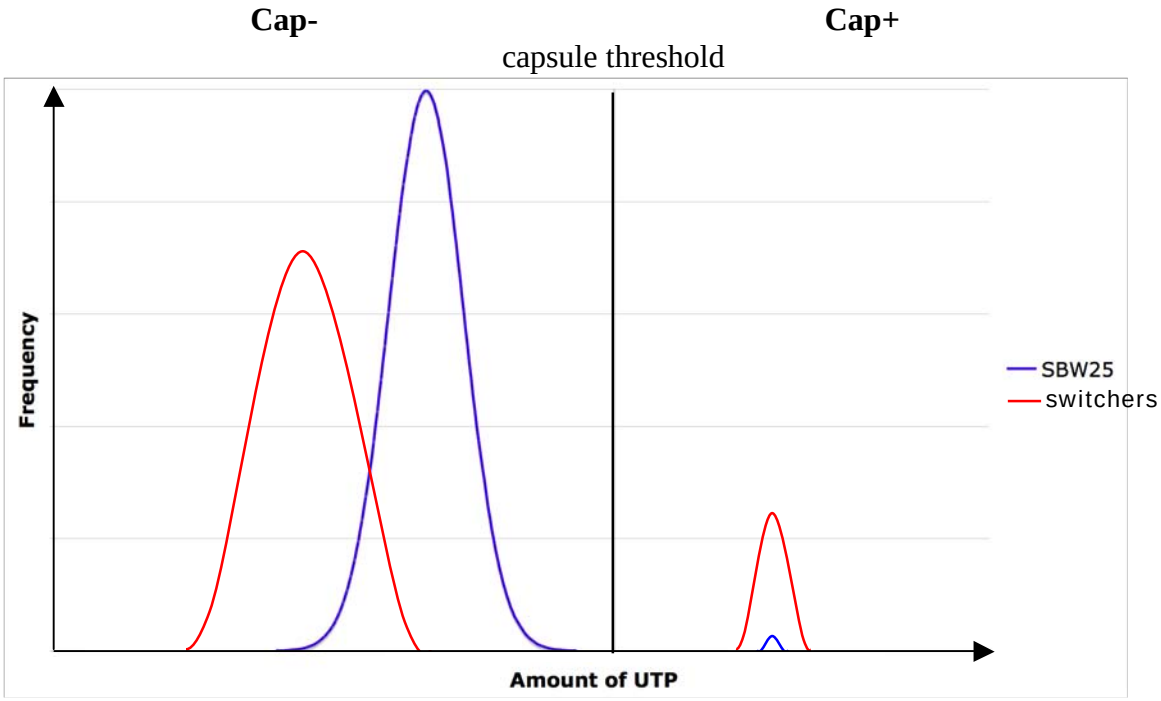


Figure 6.13: Model of the molecular basis of population-level bistability in *P. fluorescens* capsule expression. The intracellular UTP level is proposed to underlie capsule bistability; a capsule threshold exists, UTP levels above which result in capsule expression while lower UTP levels do not. Switch causing mutations alter intracellular UTP levels, causing the relative population distributions (red) to alter from that of wild type populations (blue). See text for further detail.

6.4.4.2 Molecular explanations for bistability: perpetuation of states

Capsule expression by 1w^4 is an all-or-none event; switcher cells either do not express a capsule, or express a full sized capsule. This uncompromising dichotomy implies that intracellular UTP levels do not hover uncertainly about the capsule threshold. Rather, UTP levels are maintained either stably below or above the capsule threshold, indicating the existence of self-perpetuating mechanisms for both phenotypic states (i.e. bistability; see section 1.1.2). The work in this chapter has shown that the molecular basis of bistability is restricted to within the *pyrH-ndk-pyrG* segment of the pyrimidine biosynthetic pathway. Thus, the transcription, translation or activity of one (or more) of these three genes holds the key to bistability.

In section 6.4.3.2, genetic regulation of the pyrimidine pathway was discussed. A main conclusion of this section was that the mechanisms controlling transcription of the genes in this pathway differ significantly between species. Since little experimental work has been previously undertaken in this area with *Pseudomonas* (and virtually none with *P. fluorescens*), insight into potential feedback loops involving *pyrH*, *ndk* and/or *pyrG* in this pathway is limited. It is possible that a positive feedback loop (see section 1.1.2.1) exists in this region; for instance, the expression of any one of *pyrH*, *ndk* or *pyrG* may be positively autoregulated. It is also possible that one component of a double-negative feedback loop (see section 1.1.2.1) exists in this region. Consider a situation where two genes, capsule-promoting gene *a* and capsule-repressing gene *b*, were subject to mutual transcriptional repression. In the presence of A, the product of gene *a*, transcription of *b* would be repressed and the cell would express a capsule. On the other hand, the presence of B, the product of gene *b*, transcription of *a* would be repressed and a capsule would not be expressed. It is conceivable that capsule-promoting gene *a* is present in the UTP biosynthetic pathway (i.e. *pyrH* or *ndk*). However, in the absence of likely candidates for gene *b* from the transposon mutagenesis screen of Chapter 4, this possibility has not been considered further.

Despite the lack of broadly applicable regulatory mechanisms, the pyrimidine biosynthetic pathway itself is almost impeccably conserved among all bacteria

investigated to date (Turnbough Jr. & Switzer, 2008). Thus, it is likely that reported enzyme functions are conserved in *P. fluorescens*, and the possibility that bistability results from post-transcriptional regulation involving PyrH, Ndk and/or PyrG could be investigated to a greater depth. The potential for bistability to originate from the promiscuous activity and enzymic memory of Ndk led to the development of the molecular model described in detail below. In switcher genotypes, the conditions governing net Ndk activity are almost certainly altered, given that the amount of one Ndk substrate (UDP) is severely reduced by the *carB* and *pyrH* mutations.

6.4.4.2.1 A molecular model based on Ndk activity and enzyme expression

A molecular model for capsule switching is illustrated in Figure 6.14. Consider a cell in which UDP levels are lower than those of other diphosphates. In this cell, Ndk will preferentially convert the higher-level diphosphates to triphosphates (an activity perpetuated by the enzymic memory of Ndk, see section 6.4.4.1). Thus, the cell with low UDP levels has correspondingly low UTP levels. Any UTP produced is utilized exclusively for DNA/RNA synthesis, ensuring no UTP-dependent capsules are expressed. Simultaneously, the cell senses that UDP/UTP levels are low, and feedback mechanisms are activated to increase the transcription and translation of the pyrimidine biosynthetic genes (Figure 6.1). As a result, high levels of the (mutant) pyrimidine pathway produce small amounts of UDP, which accumulate in the cell. When a wild-type level of UDP is reached, UDP is able to compete for binding sites on Ndk, and corresponding UTP levels increase. Even though UDP/UTP levels are now satisfactory, the increased quantities of pyrimidine biosynthetic enzymes still exist in the cell, allowing intracellular levels of UDP/UTP to increase to the capsule threshold (see Figure 6.13). If such over-compensation occurs, excess UTP is channelled through GalU into polysaccharide biosynthesis, generating the capsulated form.

The capsulated form is maintained as long as UDP/UTP levels are maintained above the capsule threshold. Importantly, although capsule biosynthesis requires the presence of large amounts of UTP, the process does not alter net levels of uridine nucleotides; UTP channelled into capsule biosynthesis does not form part of the final polymer, but instead

is recycled back into the system as UMP (see Figure 6.1). Thus, uridine nucleotide levels are lowered only by the synthesis of DNA and RNA at cell division. As a result, although a cell may switch during its lifetime, switching is more likely to occur upon replication. Thus, the self-perpetuation of each phenotypic state may be accounted for by a combination of competitive Ndk activity and temporal regulation of pyrimidine enzyme expression.

Aside from the flux-reducing effects of the switch-causing mutations, the above mechanism is consistent with experimental observations. Firstly, over-expression of *ndk* was shown to reduce capsulation in both SBW25 and 1w^4 (see Figure 6.12). In the above model, increasing the availability of Ndk would decrease competition for the enzyme, allowing lower levels of UDP to be immediately converted to UTP and preventing the accumulation of UDP/UTP beyond the capsule threshold. Also consistent with the model is the increase in capsulation observed in JG176, the transposon mutant genotype with an inactivational insertion in the 5' end of *ndk* (see section 4.3.1.5.2). In this genotype, *ndk* levels are likely to be very severely reduced (if not absent), with *adk* acting as a substitute for *ndk*. Thus, competition for *ndk/adk* is likely to be even stronger in this strain, and the level of UDP required for successful competition closer to the capsule threshold. It is conceivable that a greater proportion of individuals would subsequently cross the threshold once UDP pools were restored. The capsule reducing effect of *pyrH* over-expression (see Figure 6.12) is less readily explained; it is possible that increase in PyrH increases the availability of the product, UDP, thus increasing UDP/UTP pools to a more usual level (although how this would be achieved in light of the low UMP substrate levels is unclear). It is also possible that over-expression alters the repressive effect of PyrH on the *carAB* operon (see section 6.4.3.2.4), although the details of such an effect remain unknown.

Finally, the proposed model is able to account for the presence of low-level capsulation in *P. fluorescens* SBW25 and 1s^4. All features of the proposed model are present in ancestral strains; bistability in switcher types is merely activated by reduction of the uridine nucleotide pool. Given that the switch machinery is present in ancestral genotypes, it could potentially be activated at a lower level when the pyrimidine

nucleotide pool stochastically surpasses the 'normal' range. Additionally, activation of the same feedback machinery would increase in ancestral types if pyrimidine nucleotide pools were reduced. Such a reduction could occur as a result of genetic mutation, or as a result of environmental inhibition of the pyrimidine biosynthetic pathway. Interestingly, the activity of *S. typhimurium* CPSase has been shown to be sensitive to temperature: low temperatures result in lower CPSase activity (Han *et al.*, 1990). Thus, activation of the feedback loop at 16˚C in SBW25 (see section 3.3.2.2.1) is consistent with the proposed model.

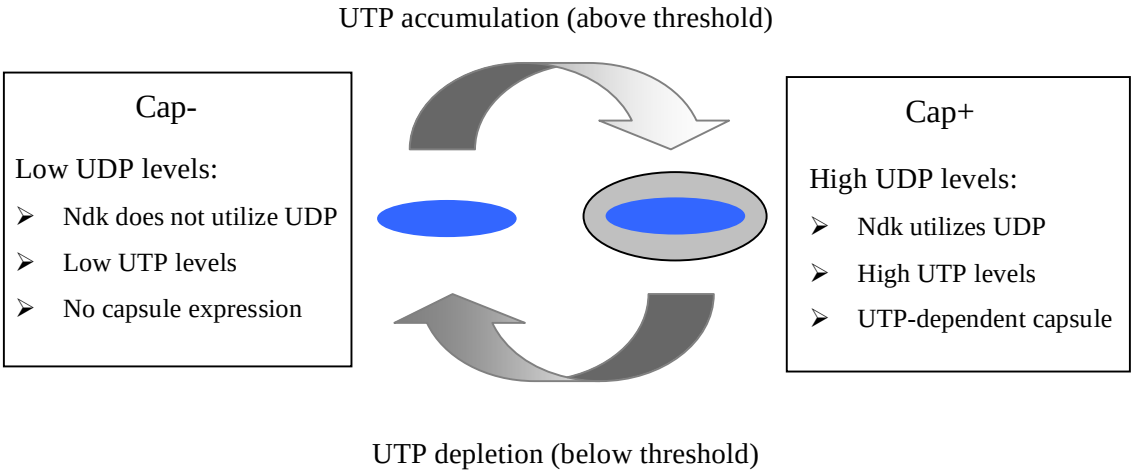


Figure 6.14: Model for the molecular basis of capsule switching. Switch-causing mutations lower intracellular levels of UDP, resulting in poor competition for Ndk, low UTP levels and no capsule expression (left). Upregulation of the biosynthetic enzymes leads to accumulation of UDP, and eventual overcompensation leads to excess UTP being channelled into capsule biosynthesis (right). Downregulation and cell division reduces UTP, and back to the cap- state, where accumulation begins again. See text for further details.

Chapter 7:
Characterization of 6w^4 Phenotypic Switching

7.1 Introduction

As noted in section 1.3.4, translucent-opaque switcher types evolved in two lines of the REE: line one and line six. Thus far, the work in this thesis has concentrated on characterizing one of these types, 1w^4. The focus of this final chapter is to comprehensively characterize the phenotype and genotype of the second switcher, 6w^4. It is hoped that comparisons between 1w^4 and 6w^4 will provide further insight into both the evolutionary and mechanistic bases of translucent-opaque phenotypic switching.

7.2 Aims

1. To phenotypically examine the line six evolutionary series, using the same assays as for the line one evolutionary series (see section 3.3.1).

2. To investigate the structural bases of 6w^4 phenotypic switching, using IS-Ω-Km/hah-based transposon mutagenesis and transcriptional fusion techniques described in Chapter 4.

3. To elucidate the evolutionary history of 6w^4. This will be achieved using the Solexa genome re-sequencing techniques described in Chapter 5.

4. To investigate the molecular mechanisms underlying 6w^4 phenotypic switching, including identification of the mutational cause(s) and investigation of the effect(s) of alterations in pyrimidine and purine nucleotide pools on switching. This will be achieved using assays outlined in Chapter 6.

7.3 Results

7.3.1 Phenotypic analysis of the 6w^4 evolutionary line

Initially, the phenotypic history of 6w^4 was investigated using the same set of assays described for 1w^4 (see section 3.3.1); assays for the analysis of colony morphology, cell morphology, niche preference in static microcosms and ACP biosynthesis were performed in each strain of the evolutionary line (Figure 7.1). Collectively, the results indicate that phenotypic innovation in the early stages of line six was achieved by the sequential ON/OFF switching of ACP biosynthesis, much like in line one. This pattern ceased with the eighth strain (6w^3), from which point alternative phenotypes were observed.

7.3.1.1 Colony morphology

For each strain, a 10^6-fold dilution of an overnight KB culture was produced, and a 25 µl aliquot spread onto KB agar. Following incubation at 28˚C for 48 hours, the morphology of colonies produced was examined, and comparative photographs taken under a dissection microscope (Figure 7.1A). Strains evolved in shaken microcosms (6s^1, 6s^2, 6s^3 and 6s^4) produced large, smooth-edged colonies, while the majority of strains evolved in the static environment produced smaller, wrinkly-edged colonies (6w^0, 6w^1 and 6w^2). 6w^3 was the first exception to this trend, and produced round, translucent colonies that resembled neither the WS nor SM morphs. Like 1w^4, the 6w^4 genotype simultaneously produced distinct translucent and opaque colony types. However, these differed slightly from the corresponding 1w^4 colony types: 6w^4 translucent colonies appeared to have more wrinkly edges, and the opaque colony type was more prevalent in 6w^4 than 1w^4.

7.3.1.2 Cell morphology and capsule production

Cell morphology and capsule production was assayed using India ink and light microscopy. Cell samples from each strain were prepared according to the method

outlined in section 2.2.11.3.1, and photographed under a light microscope (Figure 7.1B). The first seven strains of the evolutionary line (including SBW25) produced rod shaped, motile, non-capsulated cells. The eighth strain, 6w^3, produced cells that were linked together in chains, possibly as the result of incomplete cell division. This phenotype was reversed in the ninth strain (6s^4), cells of which reverted to the rod shaped, motile, non-capsulated type of the first seven strains. As was seen for 1w^4, 6w^4 gave rise to rod shaped cells of two types: motile, non-capsulated cells, and non-motile, capsulated cells. However, the capsules produced by 6w^4 were thicker and more abundant than those seen in 1w^4. It is probable that the greater production of capsular material by 6w^4 was directly responsible for the larger proportion of opaque colony types observed in 6w^4 (see section 7.3.1.1).

7.3.1.3 Niche preference in static microcosms

For each strain, three replicate static microcosms were produced according to the method outlined in section 2.2.11.2, and each microcosm was photographed under identical conditions (Figure 7.1C). Early in the series, strains evolved in the shaken environment (6s^1, 6s^2 and 6s^3) grew throughout the broth phase, while strains evolved in the static environment (6w^0, 6w^1 and 6w^2) formed a mat at the air-liquid interface. This trend was broken by 6w^3, a strain evolved in the static environment. Although 6w^3 formed a mat at the air-liquid interface, the mat was very weak, and of an atypical texture – many thin threads dangled from the mat into the broth phase. Following this, both 6s^4 and 6w^4 grew almost exclusively in the broth phase, although 6w^4 appeared to form a very weak mat in addition.

7.3.1.4 ACP production: Congo red binding

Cells from each strain were grown on KB agar containing Congo red, as outlined in section 2.2.11.3.3. Comparable photographs were taken of each strain under a dissection microscope (Figure 7.1D). With the exception of 6w^3, strains evolved in the static environment bound Congo red, and strains evolved in the shaken environment did not.

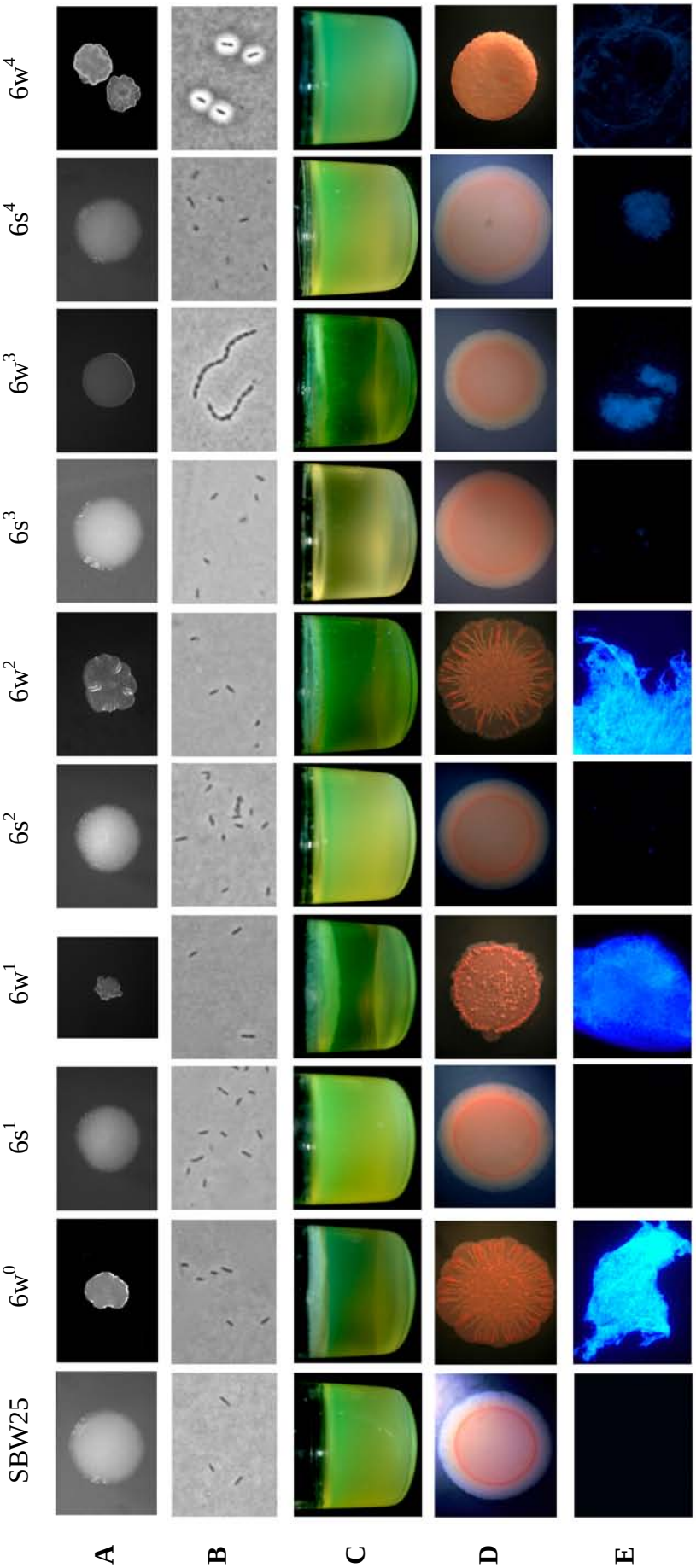


Figure 7.1: Phenotypic analysis of the line 6 evolutionary series. (**A**): Colony morphology on KB agar (48 hours), photographs taken at same magnification. (**B**): Light microscope images of cells counter-stained with India ink (x63). (**C**): Niche preference in a 48 hour static microcosm. (**D**): Dissection microscope images of colony morphology on Congo red agar (48 hours). (**E**): Fluorescence microscope images of ACP-mediated calcofluor binding (x40).

7.3.1.5 ACP production: calcofluor binding

Cells of each strain were grown on KB agar containing calcofluor, as detailed in section 2.2.11.3.2. Samples of each strain were subsequently viewed and photographed under a fluorescence microscope at x100 magnification (Figure 7.1E). The first seven strains showed a fluctuating ability to bind calcofluor; SBW25, 6s^1, 6s^2 and 6s^3 did not bind calcofluor, while 6w^0, 6w^1 and 6w^2 did. From 6w^3 onwards, the strains were not able to bind calcofluor. Notably, the cells of these strains (6w^3, 6s^4 and 6w^4) still bound a basal level of calcofluor, and so occasional blue clumps of cells were visible microscopically. However, these clumps did not produce the extracellular calcofluor binding material seen in earlier strains.

7.3.2 Identification of the 6w^4 capsule polymer

Although the biphasic cellular phenotype of 6w^4 was similar to that of 1w^4, subtle differences were observed in the appearance of capsules. To determine the structural basis of 6w^4 capsules, both transposon mutagenesis and transcriptional fusion techniques were used.

7.3.2.1 Transposon mutagenesis of 6w^4

To gain insight into the structural basis of 6w^4 phenotypic switching, a small-scale transposon mutagenesis was carried out using IS-Ω-Km/hah (see sections 2.2.8.1 and 4.1). A total of ~6,300 colonies from five independent conjugations were screened for loss of phenotypic switching. The method of screening for loss of colony sectoring used in the transposon mutagenesis of 1w^4 was unsuccessful; 6w^4 formed wholly translucent and opaque colonies on the LB+NF+Km selective medium. Therefore, non-switching transformants were selected on the basis of alteration of overall colony morphology, and 53 colonies that were neither translucent nor opaque were selected. Following examination of growth rate and cellular-level phenotypes of each mutant, seven mutants were isolated that grew at a usual rate with visible change in the frequency of capsule

expression. According to the method outlined in section 2.2.2.3, the position of the transposon insertion was identified in these seven capsule-deficient mutants (Table 7.1).

Strain	C#[a]	Pflu[b]	Gene[c]	Predicted protein function	Insertion[d]	Cap[e]
JG6.2	2	0005	-	DNA-binding response regulator	6927	-1
JG6.3	3	0005	-	DNA-binding response regulator	6947	-1
JG6.8	4	0143	*pgaA*	Outer membrane protein	158343	-1
JG6.6	3	0144	*pgaB*	Polysaccharide deacetylase	160795	-1
JG6.1	1	0145	*pgaC*	Glycosyltransferase	162803	-2
JG6.4	3	1211	-	ABC transporter/permease protein	1340996	-1
JG6.7	3	1469	*mucB*	Negative regulator of alginate biosynthesis	1614199	-1

Table 7.1: Insertion sites and capsule phenotypes for seven capsule-deficient transposon mutants of 6w^4. [a]Conjugation number during which transconjugant was isolated, [b]Pflu number, the designated number of the SBW25 gene, [c]where possible, gene name assigned on basis of BLASTP (dash=previously unnamed gene), [d]precise point of insertion in the 5'→3' direction of the SBW25 chromosome, [e]capsule phenotype of mutant (-2=almost no capsules, -1= fewer capsules than 6w^4).

Detailed in Table 7.1, transposon insertion points in the seven mutants implicate four distinct genomic loci in 6w^4 phenotypic switching. While the roles of these loci in relation to switching are not well understood, three of the loci have been previously implicated in colony-level phenotypic change. Firstly, two independent insertions were identified in Pflu0005, an uncharacterized DNA-binding RR. Interestingly, a point mutation in this gene was identified as the cause of a WS-like to SM-like phenotypic change in line two of the REE (C. Kost, personal communication). Secondly, three independent hits were identified in the *pgaABCD* locus (Pflu0143-6). In *E. coli*, this locus is required for the biosynthesis of a polysaccharide adhesin and subsequent for biofilm development (Wang *et al.*, 2004). In SBW25, expression of this locus has been shown to contribute to the formation of weak mats, such as those seen in 6w^4 static microcosms, in the absence of ACP production (Gehrig, 2005). Thus, these mutants suggest that the production of alternative polysaccharides is important in the 6w^4 switching phenotype. Finally, a single insertion was identified in *mucA*, which encodes

a negative regulator of the AlgU sigma (σ) factor. Although the function of this locus is poorly understood, it has been previously identified as an essential component of the Aws-mediated WS phenotype (McDonald *et al.*, 2009).

7.3.2.2 Construction of line six *wcaJ-lacZ* transcriptional fusions

As a result of difficulties screening the non-sectoring 6w^4 transposon mutant colonies for loss of switching, the mutagenesis screen did not reveal the structural basis of the 6w^4 capsule. In order to directly test whether the structural basis of the 6w^4 capsules was the colanic acid-like polymer identified in 1w^4, a *wcaJ-lacZ* transcriptional fusion was created in the 6s^4 and 6w^4 backgrounds. The principles and methods described for the construction of 1w^4-*wcaJ-lacZ* (see section 4.3.1.1.2) were utilized to construct 6s^4-*wcaJ-lacZ* and 6w^4-*wcaJ-lacZ*. These genotypes were subsequently grown for 72 hours on selective LB medium, and photographed under a dissection microscope (Figure 7.2). Under these conditions, 6s^4-*wcaJ-lacZ* produced mainly white colonies, while 6w^4-*wcaJ-lacZ* produced a mixture of white and blue colonies. Microscopic examination showed that blue colonies contained a significantly higher proportion of capsulated cells than white colonies. Together, these results are consistent with the hypothesis that the previously identified colanic acid-like polymer is the structural basis of the 6w^4 capsule. Moreover, given the molecular link between capsule production and the uridine nucleotides (see section 6.4.4.1), this result strongly indicates a role for the pyrimidine biosynthetic pathway in the molecular basis of 6w^4 phenotypic switching.

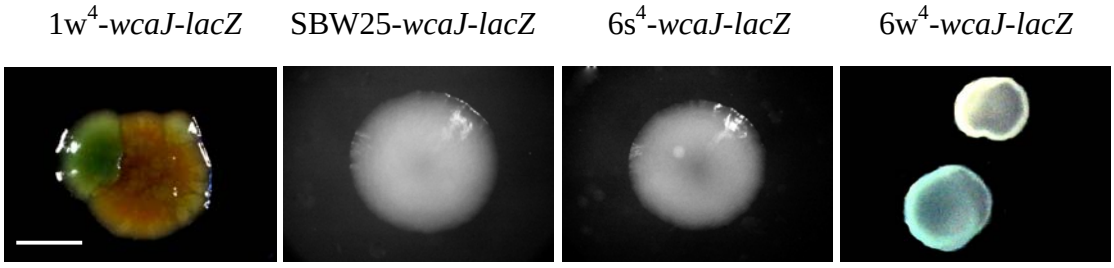


Figure 7.2: Transcription of the colanic acid biosynthetic genes is increased in 6w^4 opaque colonies. Dissection microscope pictures of 72-hour 1w^4-*wcaJ-lacZ*, SBW25-*wcaJ-lacZ*, 6s^4-*wcaJ-lacZ* and 6w^4-*wcaJ-lacZ* colonies on LB+NF+Tc+X-gal agar. The blue colour of the 6w^4 opaque colonies indicates that, as for 1w^4 capsules, colanic acid is the structural basis of 6w^4 capsules. Scale bar represents ~3 mm. Contrasts and brightness of some images were altered in iPhoto.

7.3.3 The evolutionary history of 6w^4

7.3.3.1 Identification of mutations in 6w^4

As described in section 5.3.1, a 1:1 mix of genomic DNA isolated from the translucent and opaque 6w^4 phenotypic forms was used for genome re-sequencing. Seven mutations (out of the expected nine) were easily identified in 6w^4: two substitutions in *nlpD* (Pflu1301), single substitutions in both *awsR* and *rpoD*, a 6 bp deletion in *wssB*, a 15 bp deletion in *wspF* and a 33 bp deletion in *awsX*. The presence of each mutation in both non-capsulated and capsulated 6w^4 fractions was demonstrated by independent PCR amplification and Sanger sequencing from cap- and cap+ genomic DNA (see sections 2.2.2.1 and 2.2.5.1). PCR conditions used for each mutation are detailed in Table 7.2.

Gene	Pflu#	Mutation[a]	Primers	Temp[b]	Time[c]	Enzyme	Notes
wssB	0301	Δ1720–1725	WssAf/WssAr	58	60	Taq	-
wspF	1224	Δ151–165	WspF2f/WspF2r	59	60	Taq	CES[d]
nlpD	1301	C565T	2-21f/2-21-r	60	60	Taq	-
nlpD	1301	A566G	2-21f/2-21r	60	60	Taq	-
awsR	5210	C691T	Aws8f/Aws18r	60	60	Taq	-
awsX	5211	Δ229-261	Aws11f/Aws16r	55	60	Platinum PFX	-
rpoD	5592	T1682C	RpoDf/RpoDr	58	60	Taq	-

Table 7.2: Conditions for PCR-amplification of the line six mutations. [a]See Table 7.3 for explanation of mutation nomenclature. [b]Annealing temperature (˚C). [c]Extension time (seconds). [d]Type of polymerase. [e]Requires CES addition. Mutations listed in order of spatial (not temporal) occurrence in the genome.

7.3.3.2 Determining the chronology of mutations in line six

According to the method outlined in section 2.2.1, genomic DNA was extracted from each ancestral genotype of 6w^4 (6w^0, 6s^1, 6w^1, 6s^2, 6w^2, 6s^3, 6w^3 and 6s^4). The primer pairs listed in Table 7.2 were used to PCR-amplify the mutated loci in ancestral genotypes (see section 2.2.2.1). Sanger sequencing revealed the temporal occurrence of

mutations in the line six evolutionary series (see section 2.2.5.2; Table 7.3). Using this method, no mutations were found to account for phenotypic changes in the first two genotypes of the series ($6w^0$ and $6s^1$). Given the WS-like phenotype of $6w^0$ and the SM-like phenotype of $6s^1$, it was hypothesized that the two missing mutations were in either the *wsp*, *aws* or *mws* operons. Bi-directional Sanger sequencing of these loci in $6w^0$ and $6s^1$ revealed a point deletion in $6w^0$ *wspF*, and wild-type *wspF* in $6s^1$. Thus, the two missing mutations were identified as a mutation in *wspF* followed by an exact back mutation (Table 7.3). Such an event is consistent with the inability to identify the first two mutations in the genome of $6w^4$. Individual mutations and their effects are discussed in the following sections.

Strain	**Evolution[a]**			**Gene**		**Nucleotide change**	**Amino acid change**	**Morph[b]**
	Cyc	***Rev***	***Days***	***Pflu#***	***Name***			
$6s^0$	0	0	-	-	-	-	-	SM
$6w^0$	1	1	3	1224	*wspF*	ΔT475	ΔS159(41)[c]	WS
$6s^1$		2	6	1224	*wspF*	474insT	S158ins(178)[d]	SM
$6w^1$	2	3	3	5211	*awsX*	Δ229-261	Δ77-87 (ΔYTDDLIKGTTQ)	WS
$6s^2$		4	6	5210	*awsR*	C691T	Q231Stop	SM
$6w^2$	3	5	3	1224	*wspF*	Δ151–165	Δ51–55 (ΔLMDLI)	WS
$6s^3$		6	6	0301	*wssB*	Δ1720–1725	Δ574–575 (ΔVA)	SM
$6w^3$	4	7	3	1301	*nlpD*	C565T	Q189Stop	R
$6s^4$		8	3	1301	*nlpD*	A566G	Stop189W	SM
$6w^4$	5	9	3	5592	*rpoD*	T1682C	V561A	SW

Table 7.3: Details and chronological order of mutations in the line six evolutionary series. [a]Evolutionary details including the cycle (Cyc) and reversal (Rev) of the REE, and the number of days taken to evolve. [b]Indicates phenotype on the basis of colony morphology, ability to form mats in static microcosms and ACP production (SM=smooth, WS=wrinkly spreader, R=round, SW=switcher). [c]ΔS159(41) indicates a frame shift caused by the deletion of a single nucleotide; number of new residues prior to a stop codon is in parentheses. [d]S158ins(178) indicates a frame shift by insertion of a single nucleotide; the number of new residues prior to a stop codon is in parentheses (this mutation returns *wspF* to the wild-type state, rendering SBW25 and $6s^1$ isogenic).

7.3.4 Molecular relationships between genotype and phenotype in line six

The first six mutations in the line six evolutionary series are in *wsp*, *aws* and *wss* - loci with previously recognised roles in the WS phenotype (see section 1.2.1.5). The molecular effects of these mutations are well understood. The final three mutations are in *nlpD* and *rpoD*, two genes not previously implicated in any *P. fluorescens* phenotype. Insight into the molecular effects of these mutations draws upon studies in other species. The following sections describe the molecular effects of each mutation in more detail.

7.3.4.1 Cycle one

7.3.4.1.1 Mutation one: wspF *(ΔT475)*

The first mutation in the evolutionary series caused a change from the smooth colony morphology of SBW25 (6s^0) to the WS-like colony morphology of 6w^0. The mutational cause of this change was found to be a point deletion in *wspF* (ΔT475), which causes a frame shift from amino acid residue 159 in the methylesterase domain (Table 7.3 and Figure 7.3). It is highly probable that this mutation results in loss of WspF function, resulting in constitutive activation of the DGC, WspR (Bantinaki *et al.*, 2007). Presumably, the resulting over-production of c-di-GMP and ACP causes the WS-like phenotype of 6w^0.

7.3.4.1.2 Mutation two: wspF *(474insT)*

The second mutation reverts the WS-like phenotype of 6w^0 to the smooth-like phenotype of 6s^1. Surprisingly, the 6s^1 contains the wild-type *wspF* sequence, suggesting that the mutation present in 6w^0 was reverted by precise back mutation to give 6s^1 (Figure 7.3). Such an event would completely restore the functionality of WspF (and the Wsp pathway), reducing WspR activity and abolishing c-di-GMP production and the WS phenotype (Bantinaki *et al.*, 2007).

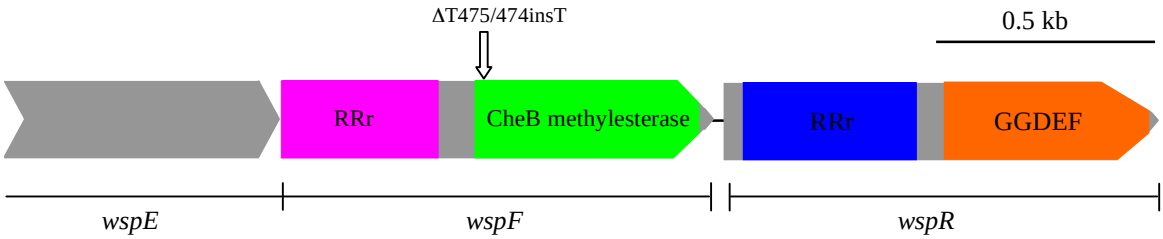


Figure 7.3: Structure of the *wspFR* locus and position of mutations one and two. White arrow indicates the approximate position of both ΔT475 (mutation one) and 474insT (mutation two).

The rarity of exact back mutation events (especially those involving the recovery of lost DNA) begs the question of whether 6s^1 is an SBW25 contaminant carried over from round one. However, four lines of evidence argue strongly against contamination. Firstly, 6s^1 took six days (including a bottleneck transfer) to evolve from 6w^0. If an SBW25 contaminant was present in the round-two inoculum, its selective advantage over 6w^0 in the shaken environment should theoretically drive a rapid rise in frequency, and this would be easily seen as the dominant phenotypic innovation at the first screen (at three days). Additionally, if a genotype is present at such a low frequency as not to be seen in the screen, the chance of it being transferred with the bottleneck population into fresh medium for the subsequent three-day incubation is minimal. Secondly, in the REE regime, colonies were screened for phenotypic innovation and then streaked to single colonies before use in the subsequent round. As colonies are (usually) founded by single cells, this would require two cells (a 6w^0 cell and a smooth cell) to found both the original and the sub-streaked colonies. Thirdly, during the course of this project the 6w^0 glycerol stock (from which round two was inoculated) was used extensively to produce colonies, and no smooth-like colonies were observed on these plates. Finally, as analysis of mutations in the REE lines continues, it is becoming increasingly obvious that *wspF* mutations are not easily reversed; to date, phenotypic innovations arising from WS-generating *wspF* mutations are caused by mutations in *wspF* itself (line six), *wspC* (line 12) or *wss* (line one and six) (C. Kost, personal communication).

Given that 6s^1 is unlikely to be a contaminant, it is possible that a limited set of mutations are able to reverse the *wspF* deletion of 6w^0. This would result in a high frequency of seemingly unlikely reversion events. To test this hypothesis, an evolution

experiment was performed where round two of the line six REE was repeated with 20 replicates. Each replicate was grown and evolved in a shaken microcosm - with transfers and colony screening every three days - until a smooth-type was isolated as detailed in section 2.2.9. Notably, the evolution of smooth types in the majority of replicates took multiple transfers, with two replicates requiring nine transfers (27 days) (Table 7.4). The entire *wspF* gene of the resulting twenty evolved smooth types (named R6s^1.1-20, for re-evolved 6s^1) was PCR-amplified using primer pair WspF2f/WspF2r (59˚C, 1 minute with addition of CES), and the sequence determined. Analysis revealed that all of the twenty smooth types newly evolved from 6w^0 contained the 6w^0 *wspF* deletion (i.e. no exact reversion events occurred), and no other *wspF* mutations (Table 7.4). From these results, it appears that reversion of the *wspF* mutation does not occur with very high frequency, demonstrating that mutations other than a precise back mutation can (and do) generate the SM-like phenotype from 6w^0. Thus, the occurrence of the back mutation in 6s^1 was a (relatively) rare event.

Genotype	**Time required for evolution**		***wspF* mutations**	
	Transfers	**Days**	**ΔT475**	**Other**
Original strains				
SBW25	-	-	×	×
6w^0	1	3	✓	×
6s^1	2	6	×	×
Smooth genotypes newly evolved from 6w^0				
R6s^1.1	3	9	✓	×
R6s^1.2	1	3	✓	×
R6s^1.3	5	15	✓	×
R6s^1.4	3	9	✓	×
R6s^1.5	3	9	✓	×
R6s^1.6	1	3	✓	×
R6s^1.7	1	3	✓	×
R6s^1.8	3	9	✓	×
R6s^1.9	2	6	✓	×
R6s^1.10	1	3	✓	×
R6s^1.11	1	3	✓	×
R6s^1.12	2	6	✓	×
R6s^1.13	9	27	✓	×

Genotype	Time required for evolution		*wspF* mutations	
	Transfers	Days	ΔT475	Other
R6s[1].14	5	15	✓	×
R6s[1].15	6	18	✓	×
R6s[1].16	7	21	✓	×
R6s[1].17	2	6	✓	×
R6s[1].18	7	21	✓	×
R6s[1].19	9	27	✓	×
R6s[1].20	5	15	✓	×

Table 7.4: The occurrence of *wspF* mutations in twenty smooth genotypes evolved from 6w[0] in the shaken environment. With the exception of 6s[1], the 6w[0] *wspF* point deletion (ΔT475) is present in all evolved smooth genotypes, and no other *wspF* mutations were observed.

7.3.4.2 Cycle two

7.3.4.2.1 Mutation three: awsX *(Δ229-261)*

Encountered in 6w[1], the third mutation of the series is the same 33 bp deletion in *awsX* as was encountered in 1w[1] (Δ229-261). For a mechanistic account of this mutation, see section 5.3.3.2.1, and for an illustration see Figure 7.4.

7.3.4.2.2 Mutation four: awsR *(C691T)*

The fourth mutation, a transition in *awsR* (C691T), is found in 6s[2] (Table 7.3). This is a nonsense mutation, generating an ochre stop codon (TAA) at residue 231. In turn, this causes the premature termination of AwsR prior to the GGDEF domain (Figure 7.4). In all likelihood, this results in loss of the AwsR-mediated c-di-GMP production initiated by mutation three (McDonald *et al.*, 2009). The corresponding reduction in ACP production produces the SM-like phenotype of 6s[2].

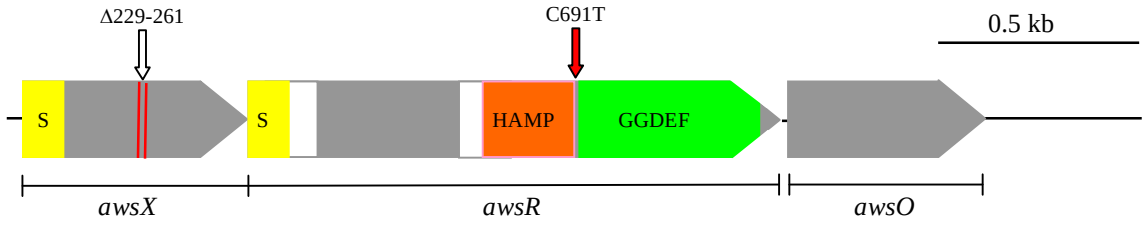


Figure 7.4: Domain structure of the AwsXRO and position of mutations three and four. White arrow indicates the approximate position of Δ229-261 (mutation three); red lines indicate approximate position of the 6 bp repeats between which the deletion occurred. Red arrow indicates the approximate position of C691T (mutation four).

7.3.4.3 Cycle three

7.3.4.3.1 Mutation five: wspF *(Δ151-165)*

First occurring in 6w[2], mutation five is an in-frame, 15 bp deletion in *wspF* (Δ151-165). This results in the deletion of five amino acids (51-55) in the response regulator receiver domain of WspF (Figure 7.5). Given the WS-like phenotype of 6w[2], it is highly probable that the mutation results in ACP production through loss of WspF function, which in turn causes constitutive activation of the DGC, WspR (Bantinaki *et al.*, 2007).

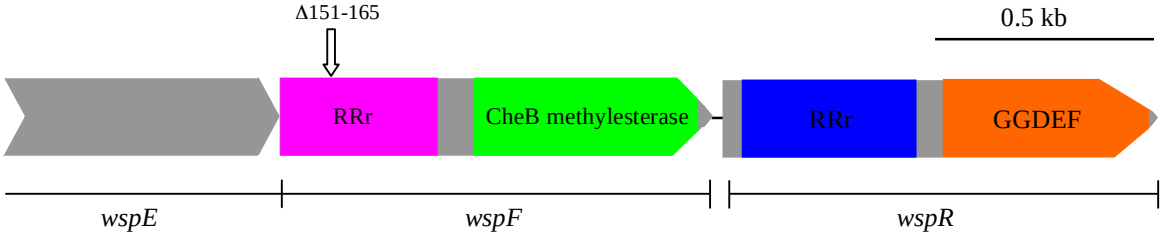


Figure 7.5: Structure of the *wspFR* locus and position of mutation five. White arrow indicates the approximate position of the Δ151-165 (mutation five).

7.3.4.3.2 Mutation six: wssB *(Δ1720-1725)*

The sixth mutation is a 6 bp deletion in *wssB* (Δ1720-1725), giving rise to 6s[3]. This is an in-frame deletion, resulting in the excision of two amino acids (V574 & A575) from the cellulose biosynthetic protein, WssB (Figure 7.6). These two amino acids occur in the PilZ domain immediately downstream of the eighth transmembrane helix, and are highly conserved across species. It is likely that their deletion severely reduces or

abolishes WssB function, and consequently ACP production. Consistent with this hypothesis, ACP production was not observed from this point forward in the line six series, and mutations were no longer obviously linked to ACP production.

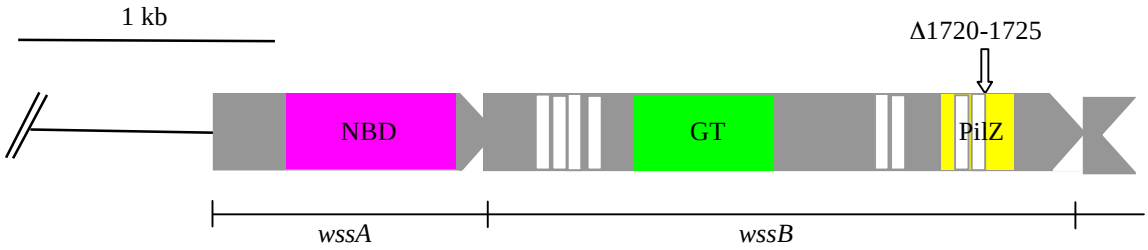


Figure 7.6: Structure of the *wssB* locus and position of mutation six. The 2,220 bp *wssB* contains eight probable transmembrane helices, a glycosyl transferase (GT) domain and a PilZ domain. The upstream *wssA*, part of the promoter and part of downstream *wssC* are shown. White arrow indicates the approximate position of Δ1720-1725 (mutation six).

7.3.4.4 Cycle four

7.3.4.4.1 Mutation seven: nlpD *(C565T)*

6w[3] contains a transition in *nlpD* (C565T), a mutation that may alter the expression of *nlpD* (Pflu1301) and/or *rpoS* (Pflu1302). With respect to NlpD, C565T is a nonsense mutation, introducing a premature amber stop codon (TAG) at amino acid 189 (Figure 7.7). NlpD contains two Pfam domains: LysM (residues 59-102, E=7.6 x 10^{-15}) and Peptidase_M23 (176-271, E=1.5 x 10^{-54}). Conserved among Gram-negative bacteria, NlpD is thought to function in cell wall formation and/or maintenance (Lange & Hengge-Aronis, 1994). It is possible that loss of NlpD function in 6w[3] contributes to cell chain formation through incomplete cell wall biogenesis at cell division.

Downstream of *nlpD* is *rpoS*, a genomic set up that is conserved in many enteric bacteria, including *E. coli* and related species. The *rpoS* gene encodes RpoS, the alternative sigma factor σ^S. Involved in the expression of many stationary phase-induced genes, RpoS contains four Pfam domains: Sigma70_r1_2 (residues 60-96, E=9.9 x 10^{-16}), Sigma70_r2 (99-169, E=1.8 x 10^{-27}), Sigma70_r3 (173-255, E=2.7 x 10^{-28}) and Sigma70_r4 (267-320, E=9.6 x 10^{-20}). Biochemical research has shown that transcription of *E. coli rpoS* during stationary phase occurs from a promoter located

~600 bp into the *nlpD* coding region (Lange & Hengge-Aronis, 1994; Lange *et al.*, 1995). As mutation seven occurs in the same vicinity as this promoter (Figure 7.7), it is possible that the phenotypic effects result from a change in *rpoS* expression.

7.3.4.4.2 Mutation eight: **nlpD** *(A566G)*

The eighth mutation is a transition in *nlpD* (A566G), creating 6s[4]. As with mutation seven, it is unclear whether mutation eight alters the function of NlpD and/or RpoS (Figure 7.7). A566G causes a non-synonymous change in NlpD; it converts the stop codon created by mutation seven into an aromatic tryptophan residue. Notably, this residue was changed from glutamine to STOP (mutation seven), and subsequently from STOP to tryptophan (mutation eight). Restoration of translation by mutation eight may account for the reversal of cell and colony phenotypes in 6s[4]. It is also possible that mutation eight mediates phenotypic change through alteration of *rpoS* transcription.

Interestingly, six further *nlpD* mutations have been identified in lines two and 11 of the REE, all of which affect *nlpD* nucleotide residues 565 or 566 (C. Kost, personal communication). Line two contains four *nlpD* mutations; the first and second are identical to mutations seven (C565T) and eight (A566G) of line six, while the third and fourth affect bases 566 and 565, respectively. Line 11 contains two *nlpD* mutations; the first is identical to mutation seven (C565T) while the second also alters nucleotide 565. Because ordinary loss-of-function mutations may be generated by a wide spectrum of mutations, the independent occurrence of mutations affecting these two residues is indicative of a specific molecular effect. Precise details of this effect remain unknown.

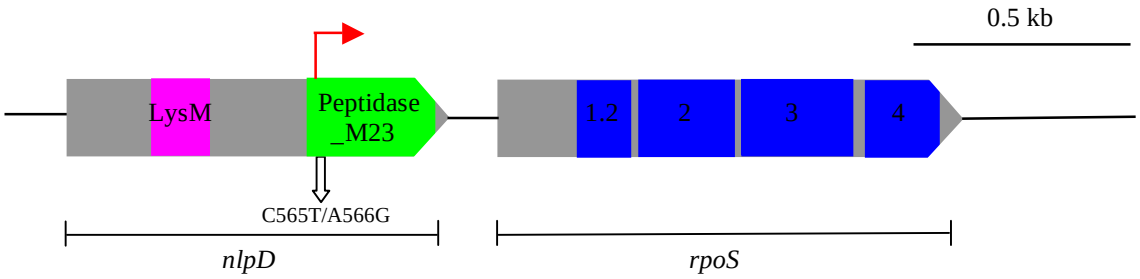


Figure 7.7: Structure of the *nlpD-rpoS* locus and position of mutations seven and eight. The 837 bp *nlpD* contains two domains while 1,005 bp *rpoS* contains four domains (see text for details). A promoter for high-level expression of *rpoS* in stationary phase is located within the coding region of *nlpD* (red arrow). White arrow indicates the approximate position of C565T and A566G (mutations 7 and 8).

7.3.4.5 Cycle five

7.3.4.5.1 Mutation nine: rpoD *(T1682C)*

The final mutation in the line six evolutionary series is a transition in *rpoD* (T1682C), the structural gene encoding the housekeeping sigma factor RpoD. The mutation generates an amino acid substitution (V561A) in a highly conserved domain of RpoD (Figure 7.8). The resulting change in phenotype indicates that this substitution alters the activity of RpoD. Given the large number of genes controlled by RpoD, the potential molecular effects of such an alteration are virtually endless. At this stage, the molecular mechanism by which this substitution causes phenotypic switching in 6w^4 is unclear.

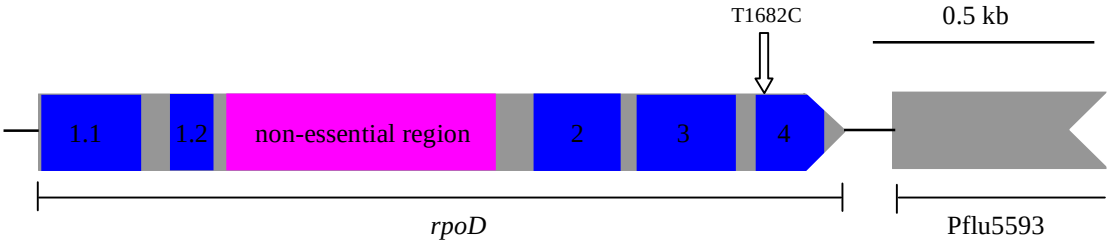


Figure 7.8: Structure of the *rpoD* locus and position of mutation nine. The 1,851 bp *rpoD* (Pflu5562) gene contains six Pfam domains: Sigma70_r1_1 (1.1; residues 3-83, *E*=2 x 10^{-37}), Sigma70_r1_2 (1.2; 97-128, *E*=3.9 x 10^{-16}), Sigma70_ner (non-essential region; residues 139-351, *E*=4.2 x 10^{-133}), Sigma70_r2 (2; 382-452, *E*-=2 x 10^{-28}), Sigma70_r3 (3; 456-538, *E*=1.1 x 10^{-48}), Sigma70_r4 (4; 550-603, *E*=2 x 10^{-25}). The beginning of the downstream Pflu5593 gene is indicated. White arrow indicates the approximate position of T1682C (mutation nine).

7.3.5 Reconstruction of the mutant *rpoD* allele in 6s^4 and SBW25

To test whether the *rpoD* mutation was the cause of phenotypic switching, the *rpoD* mutation was artificially reconstructed in both the 6s^4 and SBW25 backgrounds. To achieve this, the mutant *rpoD* allele was PCR-amplified from 6w^4 genomic DNA, using the primer pair RpoDf2/r2 (annealing temperature 58˚C, extension time 30 seconds). Following the method outlined for *carB* allelic replacements in section 5.3.4.1, the genotypes 6s^4-*rpoD*mut and SBW25-*rpoD*mut were constructed. Subsequently, these genotypes were assayed for colony morphology, calcofluor binding and capsule production (Figure 7.9). 6s^4-*rpoD*mut was phenotypically indistinguishable from 6w^4; 6s^4-*rpoD*mut produced dimorphic colonies and cells, and did not bind calcofluor.

Furthermore, SBW25-*rpoD*mut also produced dimorphic cells and colonies, although the resulting colony morphologies were different from those displayed by 6w^4.

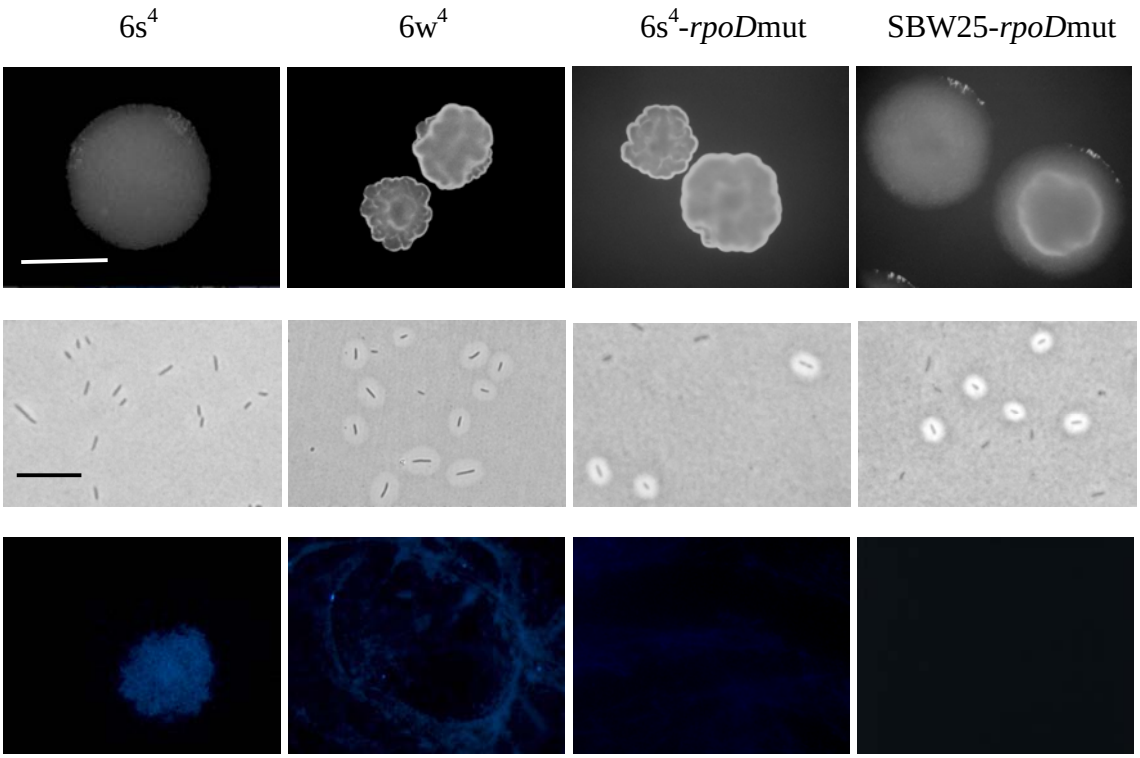


Figure 7.9: Phenotypic characterisation of 6s^4-*rpoD*mut and SBW25-*rpoD*mut. Construction of the mutant *rpoD* allele in both the 6s^4 and SBW25 backgrounds results in the switcher phenotype: (top: colony morphology on KB agar at 48 hours, scale bar indicates ~3 mm; middle: light microscope images (x40) showing India ink counter-staining of cells, scale bar indicates ~5 µm; bottom: fluorescence microscope images (x40 or x100) showing calcofluor binding ability.

According to the method outlined in section 2.2.12.4, a capsule counting assay was performed for SBW25, 6s^4, 6w^4, 6s^4-*rpoD*mut, SBW25-*rpoD*mut and 1w^4 (Appendix A5.1). Presented in Figure 7.10, the results of this assay provide a number of revelations. Firstly, the proportion of capsulated cells in 6w^4 populations is significantly greater than that in 1w^4 populations (P=1.86 x 10^{-4}). Secondly, there is no significant difference between the proportion of capsulated cells in 6w^4 and 6s^4-*rpoD*mut populations (P=0.447), while both genotypes produce populations with a significantly higher proportion of capsulated cells than 6s^4 (P<0.01). This verifies that the *rpoD* mutation causes the capsule switching phenotype in the 6s^4 genetic background. Finally, the proportion of capsulated cells produced by SBW25-*rpoD*mut populations was

significantly greater than in 6s^4 populations (P=2.38 x 10^{-4}), but significantly less than in 6w^4 populations (P=1.55 x 10^{-3}). This suggested that although the *rpoD* mutation is sufficient to cause capsule switching, at least some of the additional eight mutations in the 6w^4 background contribute to capsule expression.

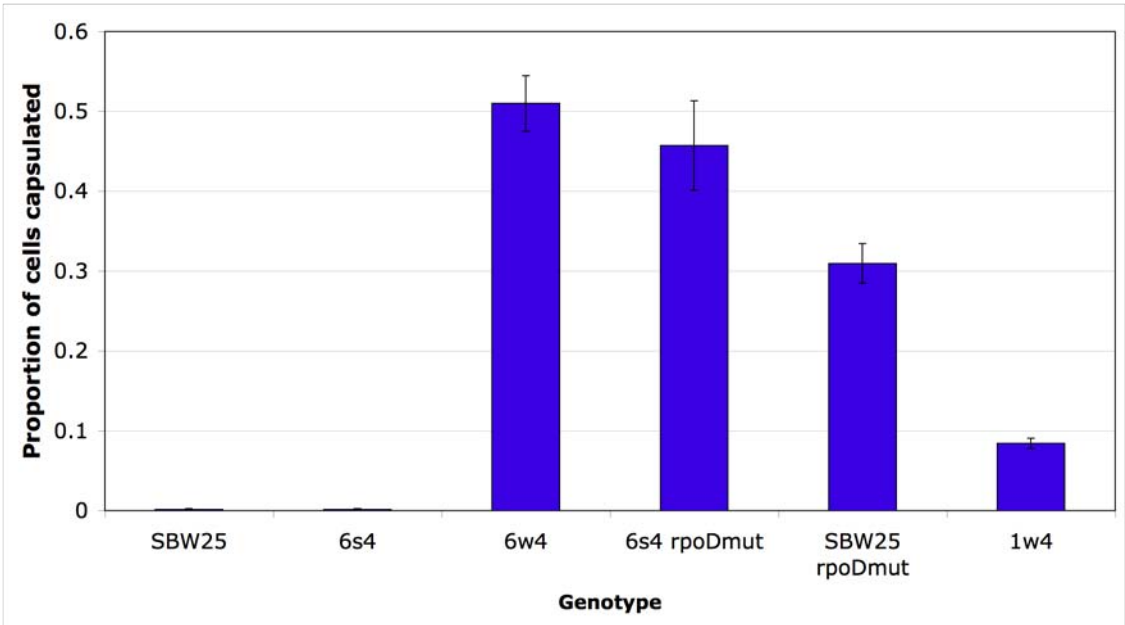


Figure 7.10: Relative proportion of capsulated cells in SBW25, 6s^4, 6w^4, 6s^4-*rpoD*mut, SBW25-*rpoD*mut and 1w^4 populations. Each bar represents the mean of five replicates, and error bars indicate one standard error.

7.3.6 Investigation of the molecular mechanism of 6w^4 phenotypic switching

Although the capsule switching phenotype of 6w^4 is similar to that of 1w^4, the causal mutation in 6w^4 is not obviously linked to the pyrimidine biosynthetic pathway. In order to investigate whether the pyrimidine biosynthetic pathway is involved in 6w^4 switching, 6w^4 was subjected to assays involving over-expression of various intermediates and components of the pyrimidine biosynthetic pathway. These assays are described in detail in the following sections.

7.3.6.1 Effect of uracil on 6w^4 phenotypic switching

Firstly, the effect of uracil addition on 6w^4 phenotypic switching was investigated. When 6w^4 was grown on KB agar containing 1 mM uracil, no obvious difference in

colony morphology was observed (likewise, the addition of 0.6 mM arginine had no discernable effect on 6w^4 colony morphology). In order to test if the addition of uracil had a smaller quantitative effect on 6w^4 capsule expression, a capsule counting assay was performed on 6w^4 cultures containing increasing amounts of uracil (as described for 1w^4 in section 6.3.4.1, see Appendix A5.2). Illustrated in Figure 7.11, the results demonstrate that the addition of uracil had no significant effect on capsule expression, (*P* for difference between 0 and 2 mM uracil=0.581). This demonstrates that if 6w^4 phenotypic switching is the result of alterations in pyrimidine biosynthesis, the disturbance occurs downstream of UMP - the entry point of the uracil salvage pathway – as in 1w^{4}reN1.4, the *pyrH* switcher (see Figure 6.1).

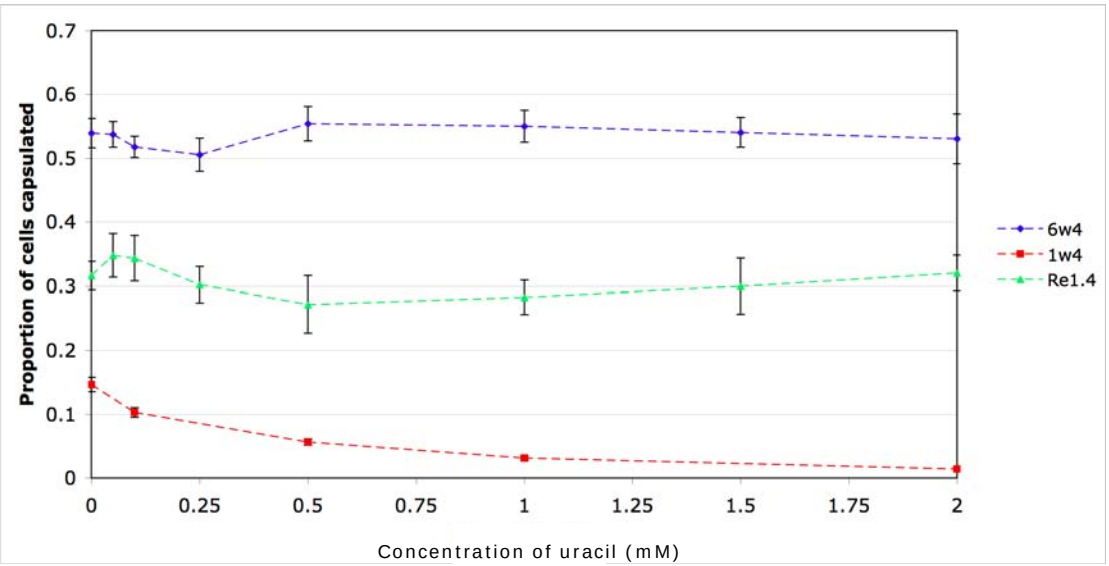


Figure 7.11: Effect of addition of increasing amounts of uracil on capsule expression in 6w^4 (blue), 1w^4 (red) and 1w^4-reN1.4 (Re1.4, green) populations. The addition of uracil did not alter capsule expression in 6w^4 or 1w^4-reN1.4 populations, while 1w^4 capsule expression was significantly reduced by addition of uracil. 1w^4 and 1w^4-reN1.4 data are replicated from Figure 6.8 for completeness. Data points are mean values of five replicates, and error bars indicate one standard error.

7.3.6.2 Effect of guanine hydrochloride on 6w^4 phenotypic switching

A quantitative capsule counting assay was performed on populations of 6s^4 and 6w^4 in the presence of 1.5 mM guanine hydrochloride (see section 2.2.12.4; Appendix A5.3). Presented in Table 7.5, the results show that growth in the presence of 1.5 mM GTP had

no significant effect on 6w^4 capsule expression (*P*=0.410). Bearing in mind that the purine pool plays a role in the regulation of *carAB* expression (see section 6.4.3.2.2), the insensitivity of 6w^4 capsule expression to guanine hydrochloride suggests an alteration somewhere downstream of *carAB*. This observation is consistent with the inability of uracil salvage to alter 6w^4 capsule expression (see section 7.3.6.1).

Genotype	**Mean ± SE in KB**	**Mean ± SE in KB+1.5 mM GTP**	**Comparison of means**
			Two sample *t*-test *P*-value
6s^4	0.00200 ± 0.000632	0.00280 ± 0.00102	0.524
6w^4	0.749 ± 0.0126	0.775 ± 0.0271	0.410

Table 7.5: Relative proportion of capsulated cells in 6w^4 populations grown in KB and KB+1.5 mM guanine hydrochloride. Mean and standard error (SE) of five replicates are given for each strain. *P*-values generated from two-sample *t*-tests were used to compare the indicated population means. Mean, standard error and *P*-values given to three significant figures.

7.3.6.3 Sequential over-expression of UTP biosynthetic genes in 6w^4

The results thus far indicate that the *rpoD* mutation causes a disruption in the pyrimidine biosynthetic pathway somewhere downstream of UMP (i.e. *pyrH*, *ndk* or *pyrG*). In order to directly test this hypothesis, selected pyrimidine biosynthetic genes were over-expressed on the pSX plasmid in the 6w^4 background. The method described in section 6.3.6 was used to generate three biological replicates of the following five genotypes: 6w^4-pSX, 6w^4-pSX-*carB*, 6w^4-pSX-*pyrH*, 6w^4-pSX-*ndk* and 6w^4-pSX-*galU*. Independent capsule counting assays were performed on each set of genotypes; the mean and standard error of the proportion of cells capsulated in three (non-biological) replicates of each genotype were calculated in every set. A final composite mean and standard error of the proportion of capsulated cells was calculated using all nine replicates (three non-biological replicates for each of three biological replicates) of each genotype. Subsequently, *P*-values for differences between the means of genotypes of interest were calculated using two-sample *t*-tests or M-W-W tests (Appendix A5.4).

The results of this assay are presented in Figure 7.12 and Appendix A5.4. Firstly, as previously noted in 1w^4 and 1w^4-reN1.4, addition of the empty pSX vector significantly lowered the proportion of capsulated cells in 6w^4 (P=1.14 x 10^{-11}). Interestingly, *carB* over-expression significantly increased capsule expression in 6w^4 (P=1.10 x 10^{-5}), an effect that indicates a regulatory interaction between CP and downstream pyrimidine genes. Over-expression of *pyrH* and *galU* had no significant effect on 6w^4 capsule expression (P>0.1). This is particularly interesting in the case of *pyrH*, the over-expression of which significantly reduced capsule expression in all other backgrounds examined to date (see section 6.3.6). Finally, the over-expression of *ndk* significantly reduced 6w^4 capsule expression (P=5.01 x 10^{-7}), indicating a direct role for *ndk* in 6w^4 capsule bistability. Overall, the results clearly demonstrate the involvement of the pyrimidine biosynthetic pathway – particularly *ndk* - in 6w^4 phenotypic switching.

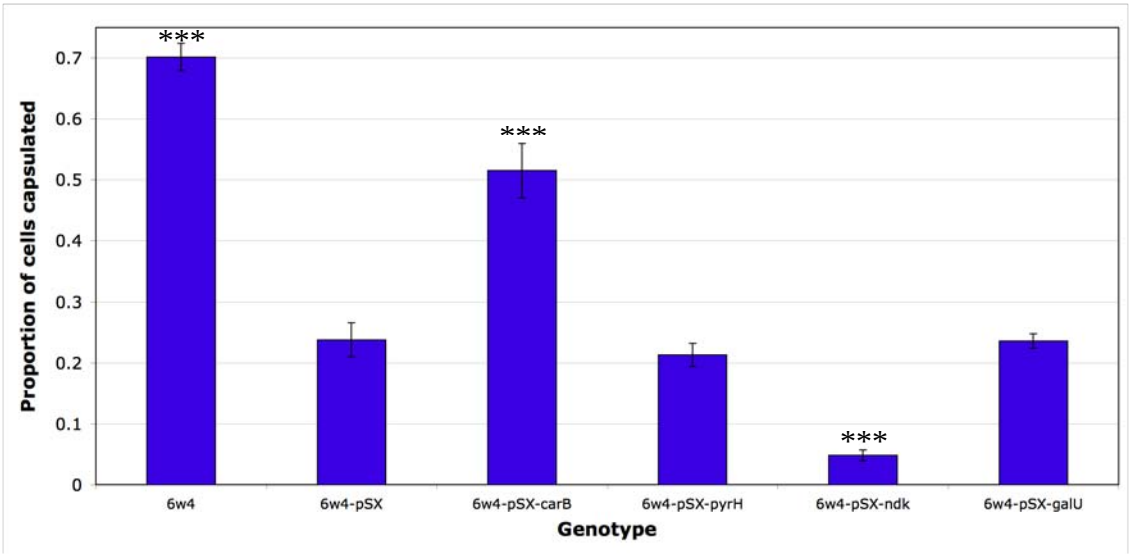


Figure 7.12: Graph showing the proportion of cells capsulated in 6w^4 populations with the indicated genes over-expressed from pSX. Bars represent mean values of nine replicates (three non-biological replicates of each of three biological replicate strains). 6w^4-pSX-*galU* is an exception, with data collected from only six replicates (three non-biological replicates from each of two biological replicates). Error bars indicate one standard error calculated from all nine replicates. Stars indicate a statistically significant difference between the mean and that of the corresponding genotype containing empty pSX (***<0.001).

7.4 Discussion

7.4.1 Summary of the 6w^4 phenotype

Work in the first part of this chapter revealed a number of phenotypic features common to both 6w^4 and 1w^4; both genotypes produce dimorphic colonies and a mixture of motile, non-capsulated and non-motile, capsulated cell types. The main structural component of capsules in both genotypes is a colanic acid-like polymer. However, the phenotypes of the two strains differ on several counts; the proportion of capsulated cells is far greater in 6w^4 than 1w^4. Furthermore, 6w^4 capsules are larger and rounder than their 1w^4 counterparts, indicating that both 6w^4 populations and individuals synthesize greater amounts of capsule polymer than 1w^4. A second point of difference between the two genotypes is that unlike 1w^4, 6w^4 does not synthesize ACP. In fact, the presence of a deletion in the WssB cellulose synthase (see section 7.3.4.3.2) strongly indicates that 6w^4 is incapable of synthesizing ACP. This observation supports the earlier conclusion that although ACP plays an indirect role in the line one switching phenotype, it is not required for capsule switching (see section 3.4.1.2). Finally, the isolation of three independent capsule-deficient mutants with transposon insertions in the *pga* locus suggests the involvement of cell surface adhesins in the 6w^4 phenotype (Wang *et al.*, 2004). The precise role of these adhesins is unclear, but may be related to their ability to form weak mats in a static microcosm (the environment in which 6w^4 arose) in the absence of ACP production (Gehrig, 2005).

7.4.2 Summary of the 6w^4 genotypic history

Elucidation of the evolutionary history of line six revealed a number of interesting similarities and differences to that of line one. Both lines exhibit a high degree of mutational modularity. That is, the mutations acquired during a single cycle of the REE (one round of evolution in static and shaken environments, respectively) tended to occur in the same locus: *wspF/wspF*, *awsX/awsR*, *wspF/wssB*, *nlpD/nlpD*, *rpoD* (line six) and *mwsR/mwsR*, *awsX/awsR*, *wspF/wssA*, *mwsR/mwsR*, *carB* (line one). As noted for line one (see section 5.4.2.1), this modularity is indicative of different molecular roles for

the *wsp*, *aws* and *mws* signal transduction pathways. In both lines, the first deviation from the modularity trend occurs at the fifth and sixth mutations, both of which are loss-of-function mutations in *wspF* followed by *wss* mutations. This is also the point at which the pathway trajectories diverge significantly; the line one *wssA* mutation does not appear to abolish ACP production, and subsequent mutations continue to affect ACP production. Contrastingly, the *wssB* mutation of line six completely abolishes ACP production, and subsequent mutations achieve phenotypic innovation by ACP-independent means. Intriguingly, *mwsR* does not feature in the line six series, probably as a consequence of the *wspF* (and *wssB*) mutation occurring first. Notably, as no *mwsR* mutations are found in line six, *mwsR* mutations do not appear to cause colony-level phenotypic effects other than ACP-based WS morphotypes.

Another point of comparison between the line one and six evolutionary series is the switch-causing mutation itself. Although the respective mutations in line one and six – *carB* in line one and *rpoD* in line six – generate very similar bistable capsule phenotypes, they do not occur in the same locus or obviously mechanistically linked loci. Collectively, these findings provide a fully–characterized example of the realization of similar phenotypic solutions *via* different evolutionary routes. While repeated evolution of the same adaptive phenotype is supportive of an adaptationist view of evolution, the fact that the remaining ten REE lines (~83 %) did not evolve capsule switching is consistent with the historical view of evolution: that evolutionary outcomes are - in many cases – heavily influenced by historical evolutionary events (see section 1.3.2).

7.4.2.1 The three line six *wspF* mutations

A total of three *wspF* mutations were identified in line six. Two of these are the first mutations of the series: a point deletion in *wspF* and an exact back mutation (see section 7.3.4.1). Although by no means impossible, exact reversal events are statistically unlikely. However, as explained in section 7.3.4.1.2, a number of lines of evidence strongly indicate that the chances of 6s^1 being a contaminant are also very low. Thus, whether or not an exact reversal event occurred between 6w^0 and 6s^1 cannot be

conclusively demonstrated. The important point is that 6s^1 is isogenic to SBW25, the original ancestor. This means that phenotypic switching in line six arose after only seven mutations, while line one phenotypic switching arose after nine. Whether all seven (in the case of 6w^4) and nine (in the case of 1w^4) were required for the evolution of switching remains a point of speculation. Given that mechanistically, phenotypic switching requires only a single mutation (in *carB*, *pyrH* or *rpoD*), it seems probable that (some of) the earlier mutations play a role in alteration of the biotic environment, enabling the evolution of bistability under the REE regime. That is, the earlier mutations are thought to eliminate the genetic routes to the advantageous WS phenotype. Once the WS phenotype can no longer be realised, mutations causing other phenotypic effects are able to emerge. Thus, it is postulated that the *wss* mutations in line one and line six played a pivotal role in the evolutionary emergence of phenotypic switching.

The third and final *wspF* mutation in the line six series occurs during the fifth round of selection (see section 7.3.4.3.1). As in line one, the subsequent mutation is in the *wss* operon. No naturally occurring mutations that revert the WS phenotype to SM have ever been observed in *wspR* (P.B. Rainey, personal communication). Given that the WS phenotype in *wspF* mutants is caused by the over-activity of WspR (see section 1.2.1.5.1), this somewhat surprising observation implies that *wspR* mutations are deleterious in a *wspF* mutant. Why this is the case remains to be seen, however it seems plausible that the c-di-GMP produced by WspR is utilised for cellular functions other than ACP production.

7.4.2.2 Mutations in *nlpD*

Following the *wssB* mutation, the genetic history of line six enters previously unknown territory. The two consecutive *nlpD* mutations and their phenotypic effects are very curious; both mutations occur in the same NlpD codon, and independent alterations of this codon in two other REE lines (see section 7.3.4.4) indicate that the mutations are altering transcription from an *rpoS* promoter contained within the *nlpD* coding sequence (Lange *et al.*, 1995). The *rpoS* gene encodes RpoS (σ^{38}), a protein required for the initiation of transcription of genes required in the stationary phase of population growth

(see section 7.4.3.1). Such genes include those involved in the stress response, some metabolic genes and some genes involved in cell morphology (Lange *et al.*, 1995). In *E. coli*, transcription of *bolA* and *ftsQAZ* is dependent upon RpoS (Neidhardt *et al.*, 1996: p.78). The *bolA* gene encodes a morphogenic protein, while the *fts* genes are involved in the temporal regulation of cell division. As homologs of these genes exist in *P. fluorescens* SBW25 (*bolA*: Pflu4312; *ftsQAZ*: Pflu0950-2), it is conceivable that alteration of the expression of these genes is at least partially responsible for the phenotypic changes observed in 6w^3/6s^4 cellular and colony morphology.

7.4.3 Insights into the mechanism of phenotypic switching

The sole mutational cause of phenotypic switching in line six is the *rpoD* mutation. However, it is worth noting that some of the preceding mutations were required in order to attain the quantitative level of capsulation seen in 6w^4 (see Figure 7.10). The role of these mutations in increasing capsulation is unclear, but this trend was also seen in line one, where the *carB* mutation alone resulted in a lower proportion of capsulated cells than was seen in the presence of all nine mutations (see section 5.3.4.1.4). This may be the result of previous mutations causing increased levels of WspR-generated c-di-GMP, which in turn may cause activation of CA enzymes. Additionally, this hypothesis could account for greater proportion of capsulated cells in 6w^4, as no c-di-GMP is used for ACP production. As with the *carB* mutation, a full molecular model of 6w^4 capsule switching requires a both a molecular link between the *rpoD* mutation and capsule production, and an explanation of why capsulation occurs in only a subset of individuals.

7.4.3.1 A molecular link between *rpoD* and capsule expression

RNA polymerase catalyzes transcription (reviewed by Slonczewski & Foster, 2009: p.259-260). Prokaryotic core RNA polymerase consists of five subunits: two α subunits (encoded by *rpoA* - Pflu5502), one β subunit (encoded by *rpoB* - Pflu5534), one β' subunit (encoded by *rpoC* – Pflu5533) and one ω subunit (encoded by *rpoZ* – Pflu5594). When assembled, inactive core RNA polymerase binds non-specifically to

DNA. Transcription initiation requires the addition of a sixth component – the σ subunit. A considerable number of different σ subunits (or σ factors) exist; each one is expressed under particular environmental conditions. Association of a σ factor with core RNA polymerase (then termed the 'holoenzyme') results in reduced affinity for non-specific DNA, and increased affinity for the promoter sequences of environment-specific genes. The holoenzyme binds to these regions and promotes both transcription initiation and elongation. Additionally, many regulatory proteins that inhibit the activity of σ factors exist. These are termed 'anti-σ factors'.

Every bacterial species encodes its own range of σ factors. The *P. fluorescens* SBW25 genome encodes 30 putative σ subunits, including the widely conserved RpoN (σ^{54}, encoded by *rpoN* – Pflu0082), RpoS (σ^{38}, encoded by *rpoS* – Pflu1302), RpoD (σ^{70}, encoded by *rpoD* – Pflu5592) and RpoH (σ^{32}, encoded by *rpoH* – Pflu5777). The primary σ factor in many bacterial species – including *P. fluorescens* - is RpoD. In these species, RpoD is required for transcription of genes required for cell growth, including the genes of the pyrimidine biosynthetic pathway (Mitchell *et al.*, 2003; Shultzaberger *et al.*, 2007; reviewed in Turnbough Jr. & Switzer, 2008). Direct demonstrations of RpoD-initiated transcription of *Pseudomonas* pyrimidine biosynthetic genes include *P. aeruginosa pyrG* (Walsh *et al.*, 1999) and *ndk*, the transcription of which is negatively regulated by the anti-RpoD protein, AlgR2 (see section 6.4.3.2.4; Schlictman *et al.*, 1995). Interestingly, mutations that alter the UTP-sensitivity of *E. coli* RNA polymerase have been identified in the β chain of RNA polymerase (Hammer *et al.*, 1987). These mutations reduce the efficiency of UTP-dependent attenuation mechanisms, leading to constitutive expression of *pyrE* and possibly *pyrBI* (see section 6.4.3.2.3).

7.4.3.2 Molecular effects of the *rpoD* mutation

As detailed in section 7.3.4.5.1, the 6w^4 *rpoD* mutation (T1682C) causes an amino acid change (V561A) in the conserved fourth domain of RpoD. As yet, there is no crystal structure of this section of RpoD (although a structure is available for residues 114-448 of *E. coli* RpoD (Malhotra *et al.*, 1996)). On the basis of DNA binding assays, the fourth domain is thought to be involved in DNA binding at the -35 region of RpoD-

dependent promoters (Dombroski *et al.*, 1992). Presumably, the phenotypic effects of this mutation are caused by alterations in RpoD binding to certain promoter sequences. Notably, the mutation causes a valine to be substituted by an alanine, residues that possess side chains with similar physical and chemical properties. Therefore, it is likely that the substitution does not completely abolish the DNA-binding function of RpoD - an observation consistent with the viability of the mutation. Indeed, amino acid substitutions that alter transcription initiation have been described in the fourth domain of *E. coli* RpoD (Hu & Gross, 1988; Siegele *et al.*, 1988). For these reasons, it is proposed that the T1682C mutation causes a reduction in the affinity of RpoD for associated promoters. This would reduce (but not abolish) the expression of RpoD-dependent genes, including many (or perhaps all) pyrimidine biosynthetic genes. Consistent with this prediction, uracil salvage was unable to influence capsule production by 6w^4 (see section 7.3.6.1); the uptake of exogenous uracil could not compensate for pathway blockages occurring downstream of UMP (i.e. *pyrH*, *ndk* and/or *pyrG*; see Figure 6.1). For similar reasons, guanine salvage did not alter 6w^4 capsule expression (see section 7.3.6.2).

The immediate result of an *rpoD*-dependent reduction in expression of the pyrimidine pathway is likely to be much the same as a *carB* or *pyrH*-dependent reduction: lower levels of uridine nucleotides (see section 6.4.4.1). Ultimately, 6w^4 capsule bistability is proposed to result from a UDP/UTP reduction according to the Ndk-dependent molecular switching model outlined in Figure 6.14. Although the mechanistic details of the model hold for 6w^4, the proportion of cells in the capsulated state is significantly greater than for 1w^4. This may reflect increased severity of the initial reduction in uridine nucleotides – not only is the UDP level reduced in 6w^4, but also the Ndk level. Therefore, competition for Ndk binding is predicted to be more intense in 6w^4 than 1w^4, and the level of UDP required to be competitively successful correspondingly higher. Thus, the competitive UDP level will be nearer the capsule threshold, and cells more likely to stochastically cross the threshold. Consistent with this hypothesis, *ndk* over-expression was able to significantly reduce capsule expression in 6w^4 (see Figure 7.12), presumably as a direct result of reduction for Ndk competition. Contrastingly, *pyrH* over-expression had no significant effect on 6w^4 capsule expression. Although this observation does not negate the Ndk-dependent switching model, it is interesting that

pyrH over-expression was able to compensate for a *carB* (and *pyrH*) mutation, but not an overall reduction in pathway expression. Taken together, these results suggest that *pyrH* over-expression has no effect on UTP biosynthesis when Ndk levels are low.

Chapter 8:
Concluding Discussion

8.1 Overview

8.1.1 Project background

Life exists in continually changing environments. In order to survive, organisms must be able to change in parallel with the environment. A common strategy to cope with unpredictable environments is to rapidly generate phenotypic diversity within a population in the hope that if and when the environment changes, some individuals will already be suited to life in the new environment. Such risk-spreading strategies are known as 'bet-hedging', or alternatively, 'phenotypic switching'. Phenotypic switching strategies are widely distributed in nature, with examples found in organisms ranging from bacteria to humans (Tonegawa, 1983; Danforth, 1999; Balaban *et al.*, 2004; Venable, 2007). The probable long-term fitness advantages of phenotypic switching strategies for infectious pathogens have long been recognised; since the first report of 'diphasic salmonellas' by Andrewes (1922), the number of documented bet-hedging strategies among pathogens has increased dramatically (reviewed in Henderson *et al.*, 1999; Veening *et al.*, 2008). Despite this prevalence, insight into the evolutionary origins and adaptive significance of these strategies remains theoretical. Historically the abundance of switching mechanisms in pathogenic microbes led to the assumption that phenotypic switching was an evolutionary response to the rapidly changing immune environment. However, the discovery of an increasing number of these strategies in phenotypes and species not in contact with the immune response has prompted renewed interest in the evolutionary causes of phenotypic switching strategies (reviewed in van der Woude, 2006; Veening *et al.*, 2008).

At the mechanistic level, phenotypic switching mechanisms have been extensively characterized (reviewed in Moxon *et al.*, 1994; Henderson *et al.*, 1999). The first

examples of switching mechanisms were found to operate on a genetic basis; switching was achieved by turning the expression of phenotypes ON and OFF *via* high rates of reversible mutation in specific genomic loci (contingency loci). More recently, epigenetic switching mechanisms – those that do not involve a change in DNA sequence between phenotypic forms – have been documented in both prokaryotes and eukaryotes (Ferrell Jr., 2002). Examples of epigenetic switches described to date operate *via* the activation of self-perpetuating feedback loops that generate heritable differences in gene transcription. While both genetic and epigenetic switching mechanisms result in the generation of interchangeable phenotypic states, little is known of their evolutionary relationship; it remains unclear whether their evolutionary origins are identical, linked or completely dissimilar (see section 8.2.2).

Extensive prior work has rendered the *P. fluorescens* SBW25 system ideal for evolution experiments; the work of many researchers has contributed to the large bank of molecular and genetic knowledge of the system (Rainey & Bailey, 1996; Rainey & Travisano, 1998; Kassen *et al.*, 2000; Spiers *et al.*, 2002; Rainey & Rainey, 2003; Spiers *et al.*, 2003; Brockhurst *et al.*, 2006; Goymer *et al.*, 2006; Kassen & Bataillon, 2006; Bantinaki *et al.*, 2007; Jasmin & Kassen, 2007; McDonald *et al.*, 2009; Silby *et al.*, 2009). Armed with this knowledge, the REE experiment devised by Dr. Hubertus Beaumont and Professor Paul Rainey was designed to investigate the extent to which evolutionary outcomes are contingent upon historical events. Described in Beaumont *et al*. (2009), the REE regime resembled the immune environment: firstly, *P. fluorescens* populations were subjected to repeated bouts of environmental change, and secondly, at each change a phenotypically different individual was rewarded by selection. Notably, only a single phenotypically different genotype was allowed to survive at each round, a principle that has been dubbed the 'exclusion rule' (see section 8.2.2). Application of the regime to 12 initially clonal populations of *P. fluorescens* SBW25 saw the independent evolution of colony level phenotypic switching on the ninth round of selection in two lines. Observation of the real-time evolution of a switching mechanism provided an unprecedented opportunity to gain insight into the evolutionary origins and adaptive significance of switching mechanisms in general. Thus, the focus of the work in this thesis was to characterize the phenotypic, evolutionary and molecular mechanisms underlying the REE-derived phenotypic switching types.

8.1.2 Review of findings

Phenotypic investigation of $1w^4$ - the line one switcher genotype - was undertaken in Chapters 3 and 4. The work in Chapter 3 showed that the opaque and translucent colony phenotypes that led to the isolation of $1w^4$ were correlated with differential polymer expression at the cellular level; the majority of cells from opaque colonies expressed a previously unidentified, acidic capsule polymer, while most cells from translucent colonies synthesized ACP. It was subsequently demonstrated that although ACP production played a role in the $1w^4$ phenotype, it was not required for phenotypic switching. Additionally, it was established that switching between capsulated and non-capsulated cell types was reversible. That is, a population founded by a single cell of either type gave rise to a mixed population of cells. The transposon mutagenesis screen of Chapter 4 demonstrated a causal link between expression of the capsule polymer and the opaque colony phenotype. Furthermore, results strongly indicated that the structural basis of the capsule is a polymer resembling colanic acid of enteric bacteria. Capsule-reducing transposon insertions identified genes required for the biosynthesis of four nucleotide precursors (UDP-Gluc, UDP-Gal, UDP-GlucA and GDP-Fuc), the locus encoding the polymer assembly and export machinery (Pflu3656-*wzb*) and potential transcriptional regulators of the biosynthetic locus (*barA/gacA*). These findings carry particular significance because the expression of colanic acid-like polymers has not previously been described in pseudomonads.

In Chapter 5, the complete evolutionary history of $1w^4$ was unravelled, revealing a series of nine mutations in the $1w^4$ genome (also reported in Beaumont *et al.*, 2009). Of these, it was shown that only the final mutation – a point mutation in *carB* of the arginine and pyrimidine biosynthetic pathways – was required for the operation of the phenotypic switch. Importantly, this mutation was present in both capsulated and non-capsulated forms of $1w^4$, demonstrating that its causal effect was indirect. That is, although the *carB* mutation causes capsule switching, switching is not achieved through direct gain and loss of the *carB* mutation. Evolutionary experiments revealed that at least some of the first eight mutations in the line one series were required for the evolution of phenotypic switching; while the *carB* mutation could arise in SBW25 (the original ancestor), it was unable to spread through the population in the absence of (at

least some of) the preceding mutations (Beaumont *et al.*, 2009). On the basis of fitness experiments and prior molecular knowledge, it was proposed that the early mutations in the series function to sequentially abolish the three readily realized mutational routes to the WS phenotype (the *wsp*, *aws* and *mws* loci). It was hypothesized that once WS types were unable to arise in a single mutational step, the switch-causing mutation was able to spread through the population, enabling it to be screened and selected in the REE.

The direct molecular bases of switching were addressed in Chapter 6. Firstly, six independent capsule switching genotypes evolved by Dr. Hubertus Beaumont from $1s^4$ (the immediate ancestor of $1w^4$) were analysed. This revealed five additional genetic routes to phenotypic switching: four novel mutations in *carB*, and a mutation in *pyrH*, a gene that acts downstream of *carB* in the pyrimidine biosynthetic pathway. On the basis of studies in other species, a genetic link was proposed between the *carB/pyrH* mutation and capsule expression: UTP, an intermediate of the pyrimidine biosynthetic pathway, is required for the production of the nucleotide-sugar precursors of colanic acid (Chakrabarty, 1998; Llamas *et al.*, 2003). Growth assays in minimal and supplemented media demonstrated that $1w^4$ was not auxotrophic for arginine or uracil. Instead, the collective results and observations of Chapter 6 strongly support the hypothesis that the *carB* and *pyrH* mutations reduce the activity of CPSase and UMP-kinase (the enzymes encoded by *carAB* and *pyrH*, respectively), causing a subsequent reduction in flux through the pyrimidine biosynthetic pathway. In the case of $1w^4$, this flux reduction - and resulting capsule bistability – is readily reversed by addition of uracil to the extracellular environment. Taken together with the genome re-sequencing data (Chapter 5), subsequent rejection of the amplification-reduction model led to the conclusion that $1w^4$ capsule switching was an epigenetic phenomenon; the *carB* (or *pyrH*) mutation simply 'set the stage' for a non-genetic mechanism to generate bistability. In consequence, this work described the characterization of the first epigenetic switch evolved under real-time observation.

The final experiment in Chapter 6 restricted the location of the epigenetic capsule switching mechanism to the *pyrH-ndk-pyrG* segment of the pyrimidine biosynthetic

pathway. Interestingly, the same genes were shown to underlie the low-level capsule switching seen in SBW25. This demonstrated that all cellular machinery and circuitry underlying the switch was already present in SBW25, and was simply activated by flux-reducing mutations in the pyrimidine biosynthetic pathway. Put another way, natural selection simply took advantage of existing molecular noise to give rise to a phenotypic switching, bet-hedging strategy. The chapter concluded with the proposal of a molecular model to explain capsule switching, based on differential levels of intracellular UDP/UTP: low to normal levels are proposed to result in the non-capsulated phenotype, while high UDP/UTP levels are postulated to result in expression of the UTP-dependent capsule. Switching between the two states was proposed to be the combined result of competition for binding sites on promiscuous Ndk, and temporal regulation of the pyrimidine biosynthetic genes (see section 6.4.4). The suggested basis of the $1w^4$ switching mechanism in enzyme activity is unusual; most epigenetic switching mechanisms described to date operate at the level of gene transcription (Ferrell Jr., 2002).

The final results chapter contained an account of the analysis of $6w^4$, the switcher genotype evolved in line six of the REE. This work showed that the phenotype of $6w^4$ was similar to that of $1w^4$; both genotypes produced two phenotypically distinct types of colonies as the result of cellular differences in the expression of capsules consisting of a colanic acid-like polymer. However, $6w^4$ did not synthesize ACP, an observation corroborated by the identification of a loss-of-function mutation in *wssB*. In addition, complete elucidation of the evolutionary history of line six revealed a locus not previously recognised as capable of causing phenotypic innovation – the *nlpD/rpoS* locus. The switch-causing ninth mutation in line six was a point mutation in the primary *P. fluorescens* σ factor, *rpoD*, which functions to initiate the transcription of genes required for cell growth and survival. It was demonstrated that - as in line one - this final mutation was the sole mechanistic cause of capsule switching. Subsequent demonstration that the *rpoD* mutation could be partially compensated for by the over-expression of *ndk* indicated that a reduction in flux through the pyrimidine biosynthetic pathway was (again) the direct cause of capsule switching. It was proposed that the *rpoD* mutation reduces the affinity of the σ factor for one or more promoter regions

within the pyrimidine biosynthetic pathway, reducing expression of the pathway and corresponding UDP/UTP levels. In consequence, it is highly probable that the same circuitry underlies capsule switching in the *carB*, *pyrH* and *rpoD* switchers (and at a lower level in ancestral genotypes), providing an in-depth example of parallel phenotypic evolution by radically different genetic routes.

8.2 Future directions

The work reported in this thesis has culminated in the proposal of a molecular model explaining the events leading to capsule bistability. There are a number of opportunities for further research to contribute to deeper understanding of capsule switching. Although the proposed epigenetic model of capsule switching is able to explain the experimental data collected to date, direct testing of the model has been limited. Therefore, insight into the accuracy of the model would be gained by testing specific predictions made by the model (see section 8.2.1). It would also be of interest to further characterize the colanic acid-like capsule polymer. To this end, analysis of the capsule polymer using mass spectrometry is currently underway in conjunction with Professor Michael Jennings (Griffith University, Australia). On a broader note, there are a number of attractive avenues for further research into the evolutionary bases of phenotypic switching in general. Of particular interest in this area are the identification of environmental factors that promote the evolution of phenotypic switching, and further evolution of the capsule switching phenotype (see section 8.2.2).

8.2.1 Testing the predictions of the epigenetic model of capsule switching

The Ndk model of capsule switching proposed to explain current experimental data makes at least three specific predictions regarding molecular effects that underpin capsule switching: (1) the *carB*, *pyrH* and *rpoD* mutations each cause a reduction in activity of the respective enzymes, (2) non-capsulated cells contain low to normal levels of UDP/UTP while capsulated cells contain higher than usual levels of UDP/UTP, and (3) relatively low UDP levels result in poor competition of UDP for Ndk binding sites,

allowing UDP to accumulate in the cell. In order to better understand the molecular events leading to switching, each of these predictions should be tested empirically.

In order to test the first prediction, the wild type and mutant forms of CPSase (encoded by *carAB*) and UMP-kinase (encoded by *pyrH*) could be purified and the enzymatic properties of each determined. Comparisons between the kinetics of the reactions catalyzed by the wild type and mutant forms of the enzymes would show the effect of each mutation on the rate and amount of product formation (as *per* Fan *et al.*, 2009). Data on the relative enzymatic properties of the various *carB* mutants would also be useful for testing the hypothesis that greater reductions in enzyme activity are correlated with a larger proportion of capsulated cells (see section 6.4.2). Demonstration of reduction in RpoD affinity for DNA upon mutation would require additional biochemical assays, in which the relative affinities of wild type and mutant RpoD for various *carAB, pyr* and *ndk* promoters would be determined.

In order to test the second prediction regarding intracellular levels of uridine nucleotides, a recently developed biochemical assay could be utilized to determine the average amount of UDP and/or UTP levels in any given population (Kochanowski *et al.*, 2006). In this assay, a method of ion-pair reverse phase high performance liquid chromatography (RP-HPLC) is used to separate and quantify the levels of eight free nucleotides (UDP, UTP, AMP, ADP, ATP, GDP, GTP and CTP) and several nucleotide sugars (including UDP-Gluc and UDP-Gal) from cell extracts. Such an assay would be extremely useful in demonstrating the relationship between different nucleotides in different genotypes. For instance, the levels of UDP and UTP could be determined in populations of SBW25 and the separated translucent and opaque forms of the *carB*, *pyrH* and *rpoD* switcher genotypes. Values obtained for SBW25 could then be used as a marker; according to the model, the average value for non-capsulated switcher populations should be below that of SBW25, while the value for the capsulated fractions would be above that of SBW25. Furthermore, the ability of the assay to simultaneously quantify UDP-Gluc and UDP-Gal would allow testing of the hypothesis

that high UTP levels correlate with high UDP-Gluc and UDP-Gal levels in the opaque fractions of the switcher genotypes.

Finally, the prediction that when present in relatively low amounts, UDP will not be converted to UTP by Ndk should be tested using *in vitro* biochemical assays. Assays designed to measure the biochemical activity of mutant Ndk forms using radiolabelled nucleotides could be adapted for this purpose (Sundin *et al.*, 1996a). Ideally, this biochemical assay requires an estimate of both the relative levels of nucleotides and the amount of Ndk present in both cell types of each genotype. The nucleotide estimate could in principle be obtained from the nucleotide quantification assay detail in the paragraph above. An estimate of the amount of Ndk is likely to prove problematic. If this is the case, the enzymatic process could be observed using a range of Ndk concentrations.

8.2.2 Further research into the evolutionary origins of phenotypic switching

As detailed in section 1.1.3, the evolutionary origins of switching phenotypes have long been an area of theoretical speculation (reviewed in Veening *et al.*, 2008). Thus, the environmental conditions promoting the evolution of switching phenotypes in the REE are of considerable significance and interest. The work in Chapter 5 elucidated the genetic events leading to the evolution of switching. However, a number of evolutionary avenues remain to be pursued, and these have formed the basis of a several current projects undertaken by other researchers in the Rainey laboratory. For instance, while it was shown that at least some of the first eight mutations in the line one series were required for the evolutionary emergence of switching (see section 5.3.5), the question of precisely *which* mutations were required remains to be answered. In sections 5.3.5.3 and 7.4.2.1, it was hypothesized that the switcher-promoting effect of the earlier mutations resulted from the removal of genetic routes to the WS phenotype, which was able to out-compete simultaneously arising switcher types. It is hoped that current investigations in this area by Sylke Nestmann will deepen our understanding of the role of the biotic environment in the evolution of switching (Lancaster & Masel, 2009).

In addition to the biotic environment, several factors have been proposed to contribute to the evolution of phenotypic switching, including the fluctuating selection imposed by changing environments, frequency of environmental change and evolvability (see section 1.1.3). Evolution of phenotypic switching in the REE provides an ideal opportunity to assess which of these factors played an active role in the selecting for the evolution of switching. Recently, Dr. Eric Libby has used mathematical models and simulations to investigate these questions (Libby & Rainey, 2010). His results indicate that three features of the regime promoted the evolution of phenotypic switching: (1) the application of a single-colony bottleneck at the beginning of each round of evolution, (2) the 'exclusion rule' (the rule whereby the selected colony had to be phenotypically different to the founder), and (3) the short length of each round. As pointed out by Libby and Rainey (2009), each of these three features is analogous to an aspect of the immune response: the bottleneck is analogous to immune-mediated destruction of the majority of pathogens, the exclusion rule is mirrored by the targeted response of the acquired immune system to the founding phenotype, and the short length of time reflects the rapidity with which the immune response acts. This indicates that these features of the immune response played (and continue to play) a pivotal role in the evolution of phenotypic switching in pathogens. As such, the results carry significant implications for understanding the effect of both the immune response and imposed treatment regimes on the evolution of pathogenic species. Furthermore, these observations highlight the possibility that phenotypic switching may be selected for in *any* environment where these three features are present (e.g. predator/prey interactions, parasite/host symbioses), providing insight into the probable evolutionary origins of phenotypic switching mechanisms found in non-pathogenic situations.

Given the similarities between the REE regime and the environment of the immune response, it has been postulated that imposing further rounds of REE selection on the switcher genotypes may lead to the evolution of a more traditional pathogenic switching mechanism; the central metabolic bistability mechanism may be the evolutionary forerunner of a contingency locus that operates *via* a genetic switching mechanism (G. Ferguson and P.B. Rainey, personal communication). Given the epigenetic nature of the REE switcher types, the evolutionary origins of contingency loci continue to be elusive; whether the evolutionary bases of bistability and contingency loci are in any way related

remains to be seen, and research to this effect is currently being carried out by Dr. Gayle Ferguson.

The thinking behind the 'evolutionary forerunner' hypothesis is as follows: at the population level, switching is a winning strategy in the REE – upon continued evolution of $1w^4$ under the REE regime, the switcher survives each round by quickly and repeatedly generating the phenotypically opposite type (Beaumont *et al.*, 2009). However, switching is caused by mutations that alter central metabolism, and such mutations are almost certainly deleterious to individual cells. Once phenotypic switching exists, this problem could be overcome in a series of mutational steps leading to a genetic switch directly in the capsule expression genes, thus eliminating the need for mutations in central metabolic pathways. For instance, a number of 'interrupted' polymeric tracts exist in the capsule-encoding Pflu3656-*wzb* locus and the *barA/gacA* regulatory loci. It is conceivable that a point mutation could arise in one such area, generating an uninterrupted polymeric tract that could control transcription or translation through an SSM mechanism (see section 1.1.1.2). For example, ~90 bp upstream of *gacA* – in the probable promoter region - lies the sequence 5'-TATATATATG-3', which could by a single G→A transition at the tenth position become a pentameric dinucleotide repeat tract that would, in all likelihood, be subject to SSM. Following such an event, the mutation in central metabolism would no longer be required for phenotypic switching, allowing compensatory mutations to restore usual metabolic function.

8.3 Final comment

Organisms have developed a wide range of strategies for coping with the considerable challenges presented by unpredictable environments; individual level phenotypic acclimation and population level genetic adaptation each play a role in survival. Life in particularly harsh, unstable environments has seen the evolution of phenotypic switching (or bet-hedging) strategies whereby population-level phenotypic diversity is rapidly generated. Diverse populations are more likely to survive an environmental

change, given that some individuals in the population are likely to be 'pre-adapted' to life in the new environment. The work in this thesis describes the first deconstruction of a series of mutational events leading to the evolution of a phenotypic switching strategy, and provides insight into the molecular events underlying switching. Perhaps the most influential outcome of this research is the unexpected ease with which phenotypic switching evolved; the evolution of switching required only nine rounds of mutation and selection, and a single mutation in central metabolism was the sole mechanistic requirement of switching. This finding carries implications for any situation where organisms are subjected to harshly fluctuating environments, including the treatment of infectious pathogens. Furthermore, the rapid and repeatable evolution of phenotypic switching suggests that these strategies may have been among the earliest solutions to the challenge of life in unpredictable environments, perhaps preceding mechanisms of gene regulation.

Reference List

Abraham, J.M., Freitag, C.S., Clements, J.R. & Eisenstein, B.I. (1985). An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proc Natl Acad Sci USA* **82**: 5274-5277.

Alby, K. & Bennett, R. (2009). Stress-induced phenotypic switching in *Candida albicans*. *Mol Biol Cell* **20**: 3178-3191.

Ancel Meyers, L. & Bull, J.J. (2002). Fighting change with change: adaptive variation in an uncertain world. *Trends Ecol Evol* **17**: 551-557.

Anderson, E.S. & Rogers, A.H. (1963). Slime polysaccharides of the *Enterobacteriaceae*. *Nature* **198**: 714-715.

Anderson, P.M. & Marvin, S.V. (1968). Effect of ornithine, IMP and UMP on carbamyl phosphate synthetase from *Escherichia coli*. *Biochem Biophys Res Comm* **32**: 928-934.

Anderson, P.M. & Meister, A. (1966). Control of *Escherichia coli* carbamyl phosphate synthetase by purine and pyrimidine nucleotides. *Biochem* **5**: 3164-3169.

Andersson, D.I., Slechta, E.S. & Roth, J.R. (1998). Evidence that gene amplification underlies adaptive mutability of the bacterial *lac* operon. *Science* **282**: 1133-1135.

Andrewes, F.W. (1922). Studies in group agglutination I. *J Pathol Bacteriol* **25**: 505.

Baehler, E., de Werra, P., Wick, L.Y., Péchy-Tarr, M., Mathys, S., Maurhofer, M. & Keel, C. (2006). Two novel MvaT-like global regulators control exoproduct formation and biocontrol activity in root-associated *Pseudomonas fluorescens* CHA0. *Mol Plant Microbe Interact* **19**: 313-329.

Bailey, J. & Manoil, C. (2002). Genome-wide internal tagging of bacterial exported proteins. *Nat Biotechnol* **20**: 839-842.

Balaban, N.Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. (2004). Bacterial persistence as a phenotypic switch. *Science* **305**: 1622-1625.

Bantinaki, E. (2001). Characterisation of a novel chemosensory pathway underlying adaptive evolution in experimental populations of *Pseudomonas fluorescens* SBW25. *Department of Plant Sciences*, Oxford: University of Oxford.

Bantinaki, E., Kassen, R., Knight, C.G., Robinson, Z., Spiers, A.J. & Rainey, P.B. (2007). Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. *Genetics* **176**: 441-453.

Beaumont, H.J.E., Gallie, J., Kost, C., Ferguson, G.C. & Rainey, P.B. (2009). Experimental evolution of bet hedging. *Nature* **462**: 90-93.

Becskei, A., Bertrand, S. & Serrano, L. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response. *EMBO J* **20**: 2528-2535.

Berg, H.C. (1975). Bacterial behaviour. *Nature* **254**: 389-392.

Bernard, M.A., Ray, N.B., Olcott, M.C., Hendricks, S.P. & Mathews, C.K. (2000). Metabolic functions of microbial nucleoside diphosphate kinases. *J Bioenerg Biomembr* **32**: 259-267.

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

Bieber Urbauer, R.J., Gilmore, J.M., Rosasco, S.E., Hattle, J.M., Cowley, A.B. & Urbauer, J.L. (2005). Cloning, high yield overexpression, purification, and characterization of AlgH, a regulator of alginate biosynthesis in *Pseudomonas aeruginosa*. *Protein Expr Purif* **43**: 57-64.

Blount, Z.D., Borland, C.Z. & Lenski, R.E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc Natl Acad Sci USA* **105**: 7899-7906.

Blyn, L.B., Braaten, B.A. & Low, D.A. (1990). Regulation of *pap* pilin phase variation by a mechanism involving differential Dam methylation states. *EMBO J* **9**: 4045-4054.

Bohannon, J. (2002). The role of the *wss* operon in the adaptive evolution of experimental populations of *Pseudomonas fluorescens* SBW25. *Department of Plant Sciences*, Oxford: University of Oxford.

Borukhov, S., Polyakov, A., Nikiforov, V. & Goldfarb, A. (1992). GreA protein: a transcription elongation factor from *Escherichia coli*. *Proc Natl Acad Sci USA* **89**: 8899-8902.

Bren, A. & Eisenbach, M. (2000). How signals are heard during bacterial chemotaxis: protein-protein interactions in sensory signal propagation. *J Bacteriol* **182**: 6865-6873.

Brichta, D.M., Azad, K.N., Ralli, P. & O'Donovan, G.A. (2004). *Pseudomonas aeruginosa* dihydroorotases: a tale of three *pyrC*s. *Arch Microbiol* **182**: 7-17.

Brockhurst, M.A., Hochberg, M.E., Bell, T. & Buckling, A. (2006). Character displacement promotes cooperation in bacterial biofilms. *Curr Biol* **16**: 2030-2034.

Caldara, M., Dupont, G., Leroy, F., Goldbeter, A. & De Vuyst, L. (2008). Arginine biosynthesis in *Escherichia coli*: experimental perturbation and mathematical modeling. *J Biol Chem* **283**: 6347-6358.

Casadesús, J. & Low, D. (2006). Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev* **70**: 830-856.

Chakrabarty, A.M. (1998). Nucleoside diphosphate kinase: role in bacterial growth, virulence, cell signalling and polysaccharide synthesis. *Mol Microbiol* **28**: 875-882.

Charlier, D., Gigot, D., Huysveld, N., Roovers, M., Piérard, A. & Glansdorff, N. (1995). Pyrimidine regulation of the *Escherichia coli* and *Salmonella typhimurium carAB* operons: CarP and integration host factor (IHF) modulate the methylation status of a GATC site present in the control region. *J Mol Biol* **250**: 383-391.

Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., Meuwly, M. & Jenal, U. (2006). Allosteric control of cyclic di-GMP signaling. *J Biol Chem* **281**: 32015-32024.

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

Chuanchuen, R., Narasaki, C.T. & Schweizer, H.P. (2002). Benchtop and microcentrifuge preparation of *Pseudomonas aeruginosa* competent cells. *Biotechniques* **33**: 760-763.

Conrad, J., Sun, D., Englund, N. & Ofengand, J. (1998). The *rluC* gene of *Escherichia coli* codes for a pseudouridine synthase that is solely responsible for synthesis of pseudouridine at positions 955, 2504, and 2580 in 23 S ribosomal RNA. *J Biol Chem* **273**: 18562-18566.

Conway Morris, S. (2003). *Life's Solution: Inevitable Humans in a Lonely Universe*. Cambridge, Cambridge University Press.

Corbett, D. & Roberts, I.S. (2008). Capsular polysaccharide in *Escherichia coli*. *Adv Appl Microbiol* **65**: 1-26.

Danforth, B.N. (1999). Emergence dynamics and bet hedging in a desert bee, *Perdita portalis*. *Proc R Soc Lond B Biol Sci* **266**: 1985-1994.

Darwin, C. & Wallace, A. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *J Linn Soc Lond* **3**: 45-62.

Dawkins, R. (1986). *The Blind Watchmaker*. London, Penguin Books Ltd.

De, N., Navarro, M.V.A.S., Raghavan, R.V. & Sondermann, H. (2009). Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. *J Mol Biol* **393**: 619-633.

De, N., Pirruccello, M., Krasteva, P.V., Bae, N., Raghavan, R.V. & Sondermann, H. (2008). Phosphorylation-independent regulation of the diguanylate cyclase WspR. *PLoS Biol* **6**: e67.

De Bolle, X., Bayliss, C.D., Field, D., van de Ven, T., Saunders, N.J., Hood, D.W. & Moxon, E.R. (2000). The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. *Mol Microbiol* **35**: 211-222.

Dombroski, A.J., Walter, W.A., Record, M.J., Siegele, D.A. & Gross, C.A. (1992). Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA. *Cell* **70**: 501-512.

Donahue, J.P. & Turnbough Jr., C.L. (1994). Nucleotide-specific transcriptional pausing in the *pyrBI* leader region of *Escherichia coli* K-12. *J Biol Chem* **269**: 18185-18191.

Dubnau, D. & Losick, R. (2006). Bistability in bacteria. *Mol Microbiol* **61**: 564-572.

El-Kazzaz, W., Morita, T., Tagami, H., Inada, T. & Aiba, H. (2004). Metabolic block at the early stages of the glycolytic pathway activates the Rcs phosphorelay system via increased synthesis of dTDP-glucose in *Escherichia coli*. *Mol Microbiol* **51**: 1117-1128.

Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. (2002). Stochastic gene expression in a single cell. *Science* **297**: 1183-1186.

Enos-Berlage, J.L. & McCarter, L.L. (2000). Relation of capsular polysaccharide production and colonial cell organization to colony morphology in *Vibrio parahaemolyticus*. *J Bacteriol* **182**: 5513-5520.

Fan, Y., Lund, L., Shao, Q., Gao, Y.-Q. & Raushel, F.M. (2009). A combined theoretical and experimental study of the ammonia tunnel in carbamoyl phosphate synthetase. *J Am Chem Soc* **131**: 10211-10219.

Fellay, R., Krisch, H.M., Prentki, P. & Frey, J. (1989). Omegon-Km: a transposable element designed for *in vivo* insertional mutagenesis and cloning of genes in Gram-negative bacteria. *Gene* **76**: 215-226.

Ferrell Jr., J.E. (2002). Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* **14**: 140-148.

Figurski, D.H. & Helinski, D.R. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. *Proc Natl Acad Sci USA* **76**: 1648-1652.

Finkel, S.E. & Kolter, R. (1999). Evolution of microbial diversity during prolonged starvation. *Proc Natl Acad Sci USA* **96**: 4023-4027.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.-R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L.L. & Bateman, A. (2008). The Pfam protein families database. *Nucleic Acids Res* **36**: D281-D288.

Freitag, C.S., Abraham, J.M., Clements, J.R. & Eisenstein, B.I. (1985). Genetic analysis of phase variation control of expression of type 1 fimbriae in *Escherichia coli*. *J Bacteriol* **162**: 668-675.

Gal, M., Preston, G.M., Massey, R.C., Spiers, A.J. & Rainey, P.B. (2003). Genes encoding a cellulosic polymer contribute toward the ecological success of *Pseudomonas fluorescens* SBW25 on plant surfaces. *Mol Ecol* **12**: 3109-3121.

Gallie, J. (2005). Characterisation of the spidery spreader phenotype in *Pseudomonas fluorescens* SBW25. *School of Biological Sciences*, Auckland: University of Auckland.

Gardner, T.S., Cantor, C.R. & Collins, J.J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**: 339-342.

Gehrig, S. (2005). Adaptation of *Pseudomonas fluorescens* SBW25 to the air-liquid interface: a study in evolutionary genetics. *Department of Plant Sciences*, Oxford: University of Oxford.

Giddens, S.R., Jackson, R.W., Moon, C.D., Jacobs, M.A., Zhang, X.-X., Gehrig, S.M. & Rainey, P.B. (2007). Mutational activation of niche-specific genes provides insight into regulatory networks and bacterial function in a complex environment. *Proc Natl Acad Sci USA* **104**: 18247-18252.

Gould, S.J. (1987). *The Panda's Thumb*. New York, Penguin Books Ltd.

Gould, S.J. (1989). *Wonderful Life: The Burgess Shale and the Nature of History*. New York, W.W. Norton & Company, Inc.

Gould, S.J. & Lewontin, R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci* **205**: 581-598.

Goymer, P., Kahn, S.G., Malone, J.G., Gehrig, S.M., Spiers, A.J. & Rainey, P.B. (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**: 1-12.

Grant, W.D., Sutherland, I.W. & Wilkinson, J.F. (1969). Exopolysaccharide colanic acid and its occurrence in the *Enterobacteriaceae*. *J Bacteriol* **100**: 1187-1193.

Güvener, Z.T. & McCarter, L.L. (2003). Multiple regulators control capsular polysaccharide production in *Vibrio parahaemolyticus*. *J Bacteriol* **185**: 5431-5441.

Hall, N. (2007). Advanced sequencing technologies and their wider impact in microbiology. *J Exp Biol* **209**: 1518-1525.

Hallet, B. (2001). Playing Dr Jekyll and Mr Hyde: combined mechanisms of phase variation in bacteria. *Curr Opin Microbiol* **4**: 570-581.

Hammer, K., Jensen, K.F., Poulsen, P., Oppenheim, A.B. & Gottesman, M. (1987). Isolation of *Escherichia coli rpoB* mutants resistant to killing by λ *c*II protein and altered in *pyrE* gene attenuation. *J Bacteriol* **169**: 5289-5297.

Han, B.D., Nolan, W.G., Hopkins, H.P., Jones, R.T., Ingraham, J.L. & Abdelal, A.T. (1990). Effect of growth temperature on folding of carbamoylphosphate synthetases of *Salmonella typhimurium* and a cold-sensitive derivative. *J Bacteriol* **172**: 5089-5096.

Han, X. & Turnbough Jr., C.L. (1998). Regulation of *carAB* expression in *Escherichia coli* occurs in part through UTP-sensitive reiterative transcription. *J Bacteriol* **180**: 705-713.

Heeb, S. & Haas, D. (2001). Regulatory roles of the GacS/GacA two-component system in plant-associated and other Gram-negative bacteria. *Mol Plant Microbe Interact* **14**: 1351-1363.

Henderson, I.R., Meehan, M. & Owen, P. (1997). A novel regulatory mechanism for a novel phase-variable outer membrane protein of *Escherichia coli*. *Adv Exp Med Biol* **412**: 349-355.

Henderson, I.R., Owen, P. & Nataro, J.P. (1999). Molecular switches - the ON and OFF of bacterial phase variation. *Mol Microbiol* **33**: 919-932.

Hendrixson, D.R. (2006). A phase-variable mechanism controlling the *Campylobacter jejuni* FlgR response regulator influences commensalism. *Mol Microbiol* **61**: 1646-1659.

Hengge, R. (2009). Principles of c-di-GMP signalling in bacteria. *Nat Rev Microbiol* **7**: 263-273.

Hernday, A., Krabbe, M., Braaten, B. & Low, D. (2002). Self-perpetuating epigenetic pili switches in bacteria. *Proc Natl Acad Sci USA* **99**: 16470-16476.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. & Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**: 51-59.

Holden, H.M., Thoden, J.B. & Raushel, F.M. (1998). Carbamoyl phosphate synthetase: a tunnel runs through it. *Curr Opin Struct Biol* **8**: 679-685.

Holden, H.M., Thoden, J.B. & Raushel, F.M. (1999). Carbamoyl phosphate synthetase: an amazing biochemical odyssey from substrate to product. *Cell Mol Life Sci* **56**: 507-522.

Hu, J.C. & Gross, C.A. (1988). Mutations in *rpoD* that increase expression of genes in the *mal* regulon of *Escherichia coli* K-12. *J Mol Biol* **203**: 15-27.

Jacobs, M.A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C., Levy, R., Chun-Rong, L, Guenthner, D., Bovee, D., Olsen, M.V. & Manoil, C. (2003). Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* **100**: 14339-14344.

Jasmin, J.-N. & Kassen, R. (2007). Evolution of a single niche specialist in variable environments. *Proc R Soc Lond B Biol Sci* **274**: 2761-2767.

Jenal, U. & Malone, J. (2006). Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu Rev Genet* **40**: 385-407.

Kadokura, H., Tian, H., Zander, T., Bardwell, J.C.A. & Beckwith, J. (2004). Snapshots of DsbA in action: detection of proteins in the process of oxidative folding. *Science* **303**: 534-537.

Kahn, S.G. (1998). Molecular characterisation of genes essential for adaptive evolution in *Pseudomonas fluorescens* SBW25. *Department of Plant Sciences*, Oxford: University of Oxford.

Kassen, R. & Bataillon, T. (2006). Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nat Genet* **38**: 484-488.

Kassen, R., Buckling, A., Bell, G. & Rainey, P.B. (2000). Diversity peaks at intermediate productivity in a laboratory microcosm. *Nature* **406**: 508-512.

Katz, M. & Westley, J. (1980). Enzymic memory studies with nucleoside-5'-diphosphate kinase. *Arch Biochem Biophys* **204**: 464-470.

Kearns, C.B., Chu, F., Rudner, R. & Losick, R. (2004). Genes governing swarming in *Bacillus subtilis* and evidence for a phase variation mechanism controlling surface motility. *Mol Microbiol* **52**: 357-369.

Kholti, A.D., Charlier, D., Gigot, D., Huysveld, N., Roovers, M. & Glansdorff, N. (1998). *pyrH*-encoded UMP-kinase directly participates in pyrimidine-specific modulation of promoter activity in *Escherichia coli*. *J Mol Biol* **280**: 571-582.

Kimata, K., Tanaka, Y., Inada, T. & Aiba, H. (2001). Expression of the glucose transporter gene, *ptsG*, is regulated at the mRNA degradation step in response to glycolytic flux in *Escherichia coli*. *EMBO J* **20**: 3587-3595.

King, E.O., Ward, M.K. & Raney, D.E. (1954). Two simple media for the demonstration of pyocyanin and fluorescein. *J Lab Clin Med* **44**: 301-307.

King, M.R., Steenbergen, S.M. & Vimr, E.R. (2007). Going for baroque at the *Escherichia coli* K1 cell surface. *Trends Microbiol* **15**: 196-202.

Kitten, T., Kinscherf, T.G., McEvoy, J.L. & Willis, D.K. (1998). A newly identified regulator is required for virulence in *Pseudomonas syringae*. *Mol Microbiol* **28**: 917-929.

Kochanowski, N., Blanchard, F., Cacan, R., Chirat, F., Guedon, E., Marc, A. & Goergen, J.-L. (2006). Intracellular nucleotide sugar contents of cultured CHO cells determined by a fast, sensitive and high-resolution ion-pair RP-HPLC. *Anal Biochem* **348**: 243-251.

Kœrn, M., Elston, T.C., Blake, W.J. & Collins, J.J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6**: 451-464.

Krinos, C.M., Coyne, M.J., Weinacht, K.G., Tzianabos, A.O., Kasper, D.L. & Comstock, L.E. (2001). Extensive surface diversity of a commensal organism by multiple DNA inversions. *Nature* **414**: 555-558.

Kussell, E. & Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**: 2075-2078.

Kwon, D.-H., Lu, C.-D., Walthall, D.A., Brown, T.M., Houghton, J.E. & Abdelal, A.T. (1994). Structure and regulation of the *carAB* operon in *Pseudomonas aeruginosa* and *Pseudomonas stutzeri*: no untranslated region exists. *J Bacteriol* **176**: 2532-2542.

Lachmann, M. & Jablonka, E. (1996). The inheritance of phenotypes: an adaptation to fluctuating environments. *J Theor Biol* **181**: 1-9.

Lamb, T.D., Collin, S.P. & Pugh Jr., E.N. (2007). Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup. *Nat Rev Neurosci* **8**: 960-976.

Lan, C.-Y., Newport, G., Murillo, L.A., Jones, T., Scherer, S., Davis, R.W. & Agabian, N. (2002). Metabolic specialization associated with phenotypic switching in *Candida albicans*. *Proc Natl Acad Sci USA* **99**: 14907-14912.

Lancaster, A.K. & Masel, J. (2009). The evolution of reversible switches in the presence of irreversible mimics. *Evolution* **63**: 2350-2362.

Lange, R., Fischer, D. & Hengge-Aronis, R. (1995). Identification of transcriptional start sites and the role of ppGpp in the expression of *rpoS*, the structural gene for the σ^S subunit of RNA polymerase in *Escherichia coli*. *J Bacteriol* **177**: 4676-4680.

Lange, R. & Hengge-Aronis, R. (1994). The *nlpD* gene is located in an operon with *rpoS* on the *Escherichia coli* chromosome and encodes a novel lipoprotein with a potential function in cell wall formation. *Mol Microbiol* **13**: 733-743.

Leifson, E. (1930). A method of staining bacterial flagella and capsules together with a study of the origin of flagella. *J Bacteriol* **20**: 203-211.

Leifson, E. (1951). Staining, shape, and arrangement of bacterial flagella. *J Bacteriol* **62**: 377-389.

Legrain, C., Stalon, V., Noullez, J.-P., Mercenier, A., Simon, J.-P., Broman, K. & Wiame, J.-M. (1977). Structure and function of ornithine carbamoyltransferases. *Eur J Biochem* **80**: 401-409.

Lenski, R.E. (1991). Quantifying fitness and gene stability in microorganisms. *Biotechnology* **15**: 173-192.

Lenski, R.E., Rose, M.R., Simpson, S.C. & Tadler, S.C. (1991). Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat* **138**: 1315-1341.

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

Levinson, G. & Gutman, G.A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* **4**: 203-221.

Libby, E. & Rainey, P. (2010). Selecting a "switcher". *In preparation for submission to Proc Natl Acad Sci USA.*

Llamas, I., Suarez, A., Quesada, E., Bejar, V. & del Moral, A. (2003). Identification and characterization of the *carAB* genes responsible for encoding carbamoylphosphate synthetase in *Halomonas eurihalina*. *Extremophiles* **7**: 205-211.

Losos, J.B., Jackman, T.R., Larson, A., de Queiroz, K. & Rodríguez-Schettino, L. (1998). Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279**: 2115-2118.

Lu, Q. & Inouye, M. (1996). Adenylate kinase complements nucleoside diphosphate kinase deficiency in nucleotide metabolism. *Proc Natl Acad Sci USA* **93**: 5720-5725.

Lu, C.-D., Kwon, D.-H. & Abdelal, A.T. (1997). Identification of *greA* encoding a transcriptional elongation factor as a member of the *carA-orf-carB-greA* operon in *Pseudomonas aeruginosa* PAO1. *J Bacteriol* **179**: 3043-3046.

Lu, Q., Zhang, X., Almaula, N., Mathews, C.K. & Inouye, M. (1995). The gene for nucleoside diphosphate kinase functions as a mutator gene in *Escherichia coli*. *J Mol Biol* **254**: 337-341.

Maamar, H. & Dubnau, D. (2005). Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. *Mol Microbiol* **56**: 615-624.

Majdalani, N. & Gottesman, S. (2005). The Rcs phosphorelay: a complex signal transduction system. *Annu Rev Microbiol* **59**: 379-405.

Malhotra, A., Severinova, E. & Darst, S.A. (1996). Crystal structure of a σ^{70} subunit fragment from *E. coli* RNA polymerase. *Cell* **87**: 127-136.

Malone, J. (2005). Analysis of the structure-function relationship of WspR, a GGDEF response regulator. *Department of Plant Sciences*, Oxford: University of Oxford.

Malone, J.G., Williams, R., Christen, M., Jenal, U. & Spiers, A.J. (2007). The structure-function relationship of WspR, a response regulator with a GGDEF output domain. *Microbiology* **153**: 980-994.

Manoil, C. (2000). Tagging exported proteins using *Escherichia coli* alkaline phosphatase gene fusions. *Meth Enzymol* **326**: 35-47.

Markovitz, A., Sydiskis, R.J. & Lieberman, M.M. (1967). Genetic and biochemical studies on mannose-negative mutants that are deficient in phosphomannose isomerase in *Escherichia coli* K-12. *J Bacteriol* **94**: 1492-1496.

Mathews, C.K., van Holde, K.E. & Ahern, K.G. (2000). *Biochemistry*. 3rd. San Francisco, Addison Wesley Longman, Inc.

McDonald, M.J. (2009). The genetics of *Pseudomonas fluorescens* SBW25: adaptation to a spatially structured environment. *New Zealand Institute for Advanced Study*, Auckland: Massey University.

McDonald, M.J., Gehrig, S.M., Meintjes, P.L., Zhang, X.-X. & Rainey, P.B. (2009). Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. IV. Genetic constraints guide evolutionary trajectories in a parallel adaptive radiation. *Genetics* **183**: 1041-1053.

Meredith, T.C., Mamat, U., Kaczynski, Z., Lindner, B., Holst, O. & Woodard, R.W. (2007). Modification of lipopolysaccharide with colanic acid (M-antigen) repeats in *Escherichia coli*. *J Biol Chem* **282**: 7790-7798.

Meyer, T.F., Gibbs, C.P. & Haas, R. (1990). Variation and control of protein expression in *Neisseria*. *Annu Rev Microbiol* **44**: 451-477.

Miller, J.H., Funchain, P., Clendenin, W., Huang, T., Nguyen, A., Wolff, E., Yeung, A., Chiang, J.-H., Garibyan, L., Slupska, M.M. & Yang, H. (2002). *Escherichia coli* strains *(ndk)* lacking nucleoside diphosphate kinase are powerful mutators for base substitutions and frameshifts in mismatch-repair-deficient strains. *Genetics* **162**: 5-13.

Mitchell, J.E., Zheng, D., Busby, S.J.W. & Minchin, S.D. (2003). Identification and analysis of 'extended -10' promoters in *Escherichia coli*. *Nucleic Acids Res* **31**: 4689-4695.

Mondragón, V., Franco, B., Jonas, K., Suzuki, K., Romeo, T., Melefors, Ö. & Georgellis, D. (2006). pH-dependent activation of the BarA-UvrY two-component system in *Escherichia coli*. *J Bacteriol* **188**: 8303-8306.

Moxon, E.R., Rainey, P.B., Nowak, M.A. & Lenski, R.E. (1994). Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol* **4**: 24-33.

Neidhardt, F.C., Curtiss, R.I., Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M. & Umbarger, H.E. (1996). Escherichia coli *and* Salmonella*: cellular and molecular biology*. 2nd edition. Washington, D.C., American Society for Microbiology Press.

Neuhard, J. & Thomassen, E. (1971). Deoxycytidine triphosphate deaminase: identification and function in *Salmonella typhimurium*. *J Bacteriol* **105**: 657-665.

Nicholson, B. & Low, D. (2000). DNA methylation-dependent regulation of *pef* expression in *Salmonella typhimurium*. *Mol Microbiol* **35**: 728-742.

Nixon, T., Ronson, C.W. & Ausbel, F.M. (1986). Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes *ntrB* and *ntrC*. *Proc Natl Acad Sci USA* **83**: 7850-7854.

O'Donovan, G.A. & Neuhard, J. (1970). Pyrimidine metabolism in microorganisms. *Bacteriol Rev* **34**: 278-343.

Paul, R., Weiser, S., Amiot, N.C., Chan, C., Schirmer, T., Giese, B. & Jenal, U. (2004). Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev* **18**: 715-727.

Peak, I.R.A., Jennings, M.P., Hood, D.W. & Moxon, E.R. (1999). Tetranucleotide repeats identify novel virulence determinant homologues in *Neisseria meningitidis*. *Microb Pathog* **26**: 13-23.

Pernestig, A.-K., Melefors, Ö. & Georgellis, D. (2001). Identification of UvrY as the cognate response regulator for the BarA sensor kinase in *Escherichia coli*. *J Biol Chem* **276**: 225-231.

Piérard, A. (1966). Control of the activity of *Escherichia coli* carbamoyl phosphate synthetase by antagonistic allosteric effectors. *Science* **154**: 1572-1573.

Piette, J., Nyunoya, C., Lusty, C.J., Cunin, R., Weyens, G., Crabeel, M., Charlier, D., Glansdorff, N. & Piérard, A. (1984). DNA sequence of the *carA* gene and the control region of *carAB*: tandem promoters, respectively controlled by arginine and the pyrimidines, regulate the synthesis of carbamoyl-phosphate synthetase in *Escherichia coli* K-12. *Proc Natl Acad Sci USA* **81**: 4134-4138.

Ptashne, M. (1992). *A Genetic Switch*. 2nd edition. Boston, Cell Press & Blackwell Scientific Publications.

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

Rainey, P.B. & Bailey, M.J. (1996). Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. *Mol Microbiol* **19**: 521-533.

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

Rainey, P.B. & Travisano, M. (1998). Adaptive radiation in a heterogeneous environment. *Nature* **394**: 69-72.

Ralser, M., Querfurth, R., Warnatz, H.J., Lehrach, H., Yapso, M.L. & Krobitsch, S. (2006). An efficient and economic enhancer mix for PCR. *Biochem Biophys Res Comm* **347**: 747-751.

Rando, O.J. & Verstrepen, K.J. (2007). Timescales of genetic and epigenetic inheritance. *Cell* **128**: 655-668.

Ray, N.B. & Mathews, C.K. (1992). Nucleoside diphosphate kinase: a functional link between intermediary metabolism and nucleic acid synthesis. *Curr Top Cell Regul* **33**: 343-357.

Rich, J.J., Kinscherf, T.G., Kitten, T. & Willis, D.K. (1994). Genetic evidence that the *gacA* gene encodes the cognate response regulator for the *lemA* sensor in *Pseudomonas syringae*. *J Bacteriol* **176**: 7468-7475.

Rikkerink, E.H.A., Magee, B.B. & Magee, P.T. (1988). Opaque-white phenotype transition: a programmed morphological transition in *Candida albicans*. *J Bacteriol* **170**: 895-899.

Römling, U., Gomelsky, M. & Galperin, M.Y. (2005). C-di-GMP: the dawn of a novel bacterial signalling system. *Mol Microbiol* **57**: 629-639.

Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S. & Elowitz, M.B. (2005). Gene regulation at the single-cell level. *Science* **307**: 1962-1965.

Rosenthal, R.S. & Rodwell, V.W. (1998). Purification and characterization of the heteromeric transcriptional activator MvaT of the *Pseudomonas mevalonii mvaAB* operon. *Protein Science* **7**: 178-184.

Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Chana, P., Mayer, R., Braun, S., de Vroom, E., van der Mariel, G.A., van Boom, J.H. & Benziman, M. (1987). Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325**: 279-281.

Roth, J.R. & Andersson, D.I. (2004). Adaptive mutation: how growth under selection stimulates Lac+ reversion by increasing target copy number *J Bacteriol* **186**: 4855-4860.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944-945.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular cloning - a laboratory manual*. 2nd edition, Cold Spring Harbour Laboratory Press.

Santhanagopalan, V., Coker, C. & Radulovic, S. (2006). Characterization of RP 333, a gene encoding CapD of *Rickettsia prowazekii* with UDP-glucose 4-epimerase activity. *Gene* **369**: 119-125.

Saunders, N.J., Jeffries, A.C., Peden, D.W., Hood, H., Tettelin, R., Rappuoli, R. & Moxon, E.R. (2000). Repeat-associated phase variable genes in the complete genome sequence of *Neiserria meningitidis* strain MC58. *Mol Microbiol* **37**: 207-215.

Saunders, N.J., Moxon, E.R. & Gravenor, M.B. (2003). Mutation rates: estimating phase variation rates when fitness differences are present and their impact on population structure. *Microbiology* **149**: 485-495.

Schlictman, D., Kubo, M., Shankar, S. & Chakrabarty, A.M. (1995). Regulation of nucleoside diphosphate kinase and secretable virulence factors in *Pseudomonas aeruginosa*: roles of *algR2* and *algH*. *J Bacteriol* **177**: 2469-2474.

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

Schmidt, A.J., Ryjenkov, D.A. & Gomelsky, M. (2005). The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive domains. *J Bacteriol* **187**: 4774-4781.

Schwartz, M. & Neuhard, J. (1975). Control of expression of the *pyr* genes in *Salmonella typhimurium*: effects of variations in uridine and cytidine nucleotide pools. *J Bacteriol* **121**: 814-822.

Seib, K.L., Peak, I.R.A. & Jennings, M.P. (2002). Phase variable restriction-modification systems in *Moraxella catarrhalis*. *FEMS Immunol Med Microbiol* **32**: 159-165.

Shimada, T., Hirao, K., Kori, A., Yamamoto, K. & Ishihama, A. (2007). RutR is the uracil/thymine-sensing master regulator of a set of genes for synthesis and degradation of pyrimidines. *Mol Microbiol* **66**: 744-757.

Shultzaberger, R.K., Chen, Z., Lewis, K.A. & Schneider, T.D. (2007). Anatomy of *Escherichia coli* sigma-70 promoters. *Nucleic Acids Res* **35**: 771-788.

Siegele, D.A., Hu, J.C. & Gross, C.A. (1988). Mutations in *rpoD*, the gene encoding the σ^{70} subunit of *Escherichia coli* RNA polymerase, that increase expression of the *lac* operon in the absence of CAP-cAMP. *J Mol Biol* **203**: 29-37.

Silby, M.W., Cerdeño-Tárraga, A.M., Vernikos, G.S., Giddens, S.R., Jackson, R.W., Preston, G.M., Zhang, X.-X., Moon, C.D., Gehrig, S.M., Godfrey, S.A.C., Knight, C.G., Malone, J.G., *et al.* (2009). Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome Biol* **10**: R51.

Slatkin, M. (1974). Hedging one's evolutionary bets. *Nature* **250**: 704-705.

Sledjeski, D.D. & Gottesman, S. (1996). Osmotic shock induction of capsule synthesis in *Escherichia coli* K-12. *J Bacteriol* **178**: 1204-1206.

Slonczewski, J.L. & Foster, J.W. (2009). *Microbiology: an Evolving Science*. New York, W.W. Norton & Company.

Slutsky, B., Staebell, M., Anderson, J., Risen, L., Pfaller, M. & Soll, D.R. (1987). "White-opaque transition": a second high-frequency switching system in *Candida albicans J Bacteriol* **169**: 189-197.

Smits, W.K., Eschevins, C.C., Susanna, K.A., Bron, S., Kuipers, O.P. & Hamoen, L.W. (2005). Stripping *Bacillus*: ComK auto-stimulation is responsible for the bistable response in competence development. *Mol Microbiol* **56**: 604-614.

Sniegowski, P.D., Gerrish, P.J. & Lenski, R.E. (1997). Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387**: 703-705.

Spiers, A.J., Bohannon, J., Gehrig, S.M. & Rainey, P.B. (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., Kahn, S.G., Bohannon, J., Travisano, M. & Rainey, P.B. (2002). Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics* **161**: 33-46.

Srikhanta, Y.N., Maguire, T.L., Stacey, K.J., Grimmond, S.M. & Jennings, M.P. (2005). The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *Proc Natl Acad Sci USA* **102**: 5547-5551.

Stapleton, M.A., Javid-Majd, F., Harmon, M.F., Hanks, B.A., Grahmann, J.L., Mullins, L.S. & Raushel, F.M. (1996). Role of conserved residues within the carboxy phosphate domain of carbamoyl phosphate synthetase. *Biochem* **35**: 14352-14361.

Stearns, S.C. & Hoekstra, R.F. (2005). *Evolution*. 2nd edition. Oxford, Oxford University Press.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A. & Lewis, S. (2002). The generic genome browser: a building block for a model organism system database. *Genome Res* **12**: 1599-1610.

Stern, A., Brown, M., Nickel, P. & Meyer, T.F. (1986). Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation. *Cell* **47**: 61-71.

Stern, A. & Meyer, T.F. (1987). Common mechanism controlling phase variation and antigenic variation in pathogenic neisseriae. *Mol Microbiol* **1**: 5-12.

Stevenson, G., Andrianopoulos, K., Hobbs, M. & Reeves, P.R. (1996). Organization of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *J Bacteriol* **178**: 4885-4893.

Stevenson, G., Lan, R. & Reeves, P.R. (2000). The colanic acid cluster of *Salmonella enterica* has a complex history. *FEMS Microbiol Lett* **191**: 11-16.

Stock, A.M., Robinson, V.L. & Goudreau, P.N. (2000). Two-component signal transduction. *Annu Rev Biochem* **69**: 183-215.

Stryer, L. (1988). *Biochemistry*. 3rd edition. New York, W.H. Freeman and Company.

Sundin, G.W., Shankar, S. & Chakrabarty, A.M. (1996a). Mutational analysis of nucleoside diphosphate kinase from *Pseudomonas aeruginosa*: characterization of critical amino acid residues involved in exopolysaccharide alginate synthesis. *J Bacteriol* **178**: 7120-7128.

Sundin, G.W., Shankar, S., Chugani, S.A., Chopade, B.A., Kavanaugh-Black, A. & Chakrabarty, A.M. (1996b). Nucleoside diphosphate kinase from *Pseudomonas aeruginosa*: characterization of the gene and its role in cellular growth and exopolysaccharide alginate synthesis. *Mol Microbiol* **20**: 965-979.

Thoden, J.B., Raushel, F.M., Benning, M.M., Rayment, I. & Holden, H.M. (1999). The structure of carbamoyl phosphate synthetase determined to 2.1 Å resolution. *Acta Crystallogr D* **55**: 8-24.

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* **302**: 575-581.

Travisano, M., Mongold, J.A., Bennett, A.F. & Lenski, R.E. (1995). Experimental tests of the roles of adaptation, chance and history in evolution. *Science* **267**: 87-90.

Tsong, A.E., Miller, M.G., Raisner, R.M. & Johnson, A.D. (2003). Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* **115**: 389-399.

Turnbough Jr., C.L. & Switzer, R.L. (2008). Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. *Microbiol Mol Biol Rev* **72**: 266-300.

Vallet, I., Diggle, S.P., Stacey, R.E., Cámara, M., Ventre, I., Lory, S., Lazdunski, A., Williams, P. & Filloux, A. (2004). Biofilm formation in *Pseudomonas aeruginosa*: fimbrial cup gene clusters are controlled by the transcriptional regulator MvaT. *J Bacteriol* **186**: 2880-2890.

van Belkum, A., van Leeuwen, W., Scherer, S. & Verbrugh, H. (1999). Occurrence and structure-function relationship of pentameric short sequence repeats in microbial genomes. *Res Microbiol* **150**: 617-626.

van der Woude, M.W. (2006). Re-examining the role and random nature of phase variation. *FEMS Microbiol Lett* **254**: 190-197.

van der Woude, M.W. & Bäumler, A.J. (2004). Phase and antigenic variation in bacteria. *Clin Microbiol Rev* **17**: 581-611.

van der Woude, M.W., Braaten, B. & Low, D. (1996). Epigenetic phase variation of the *pap* operon in *Escherichia coli*. *Trends Microbiol* **4**: 5-9.

Veening, J.-W., Smits, W.K. & Kuipers, O.P. (2008). Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* **62**: 193-210.

Venable, D.L. (2007). Bet hedging in a guild of desert annuals. *Ecology* **88**: 1086-1090.

Vermeij, G.J. (2006). Historical contingency and the purported uniqueness of evolutionary innovations. *Proc Natl Acad Sci USA* **103**: 1804-1809.

Vimr, E.R. & Steenbergen, S.M. (2006). Mobile contingency locus controlling *Escherichia coli* K1 polysialic acid capsule acetylation. *Mol Microbiol* **60**: 828-837.

Virji, M. (2009). Pathogenic neisseriae: surface modulation, pathogenesis an infection control. *Nat Rev Microbiol* **7**: 274-286.

Walsh, A.G., Burrows, L.L. & Lam, J.S. (1999). Genetic and biochemical characterization of an operon involved in the biosynthesis if 3-deoxy-D-*manno*-octulosonic acid in *Pseudomonas aeruginosa*. *FEMS Microbiol Lett* **173**: 27-33.

Wang, X., Preston III, J.F. & Romeo, T. (2004). The *pgaABCD* locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation. *J Bacteriol* **186**: 2724-2734.

Whitfield, C. (2006). Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu Rev Biochem* **75**: 39-68.

Willems, R., Paul, A., van der Helde, H.G.J., ter Avest, A.R. & Mooi, F.R. (1990). Fimbrial phase variation in *Bordetella pertussis*: a novel mechanism for transcriptional regulation. *EMBO J* **9**: 2803-2809.

Wisniewski-Dyé, F. & Vial, L. (2008). Phase and antigenic variation mediated by genome modifications. *Antonie Leeuwenhoek* **94**: 493-515.

Wolf, D.M., Vazirani, V.V. & Arkin, A.P. (2005). Diversity in times of adversity: probabilistic strategies in microbial survival games. *J Theor Biol* **234**: 227-253.

Woods, R., Schneider, D., Winkworth, C.L., Riley, M.A. & Lenski, R.E. (2006). Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc Natl Acad Sci USA* **103**: 9107-9112.

Yildiz, F.H. & Schoolnick, G.K. (1999). *Vibrio cholerae* O1 El Tor: identification of a gene cluster required for the rugose colony type, exopolysaccharide production, chlorine resistance, and biofilm formation. *Proc Natl Acad Sci USA* **96**: 4028-4033.

Zhang, X.-X. & Rainey, P.B. (2007). Construction and validation of a neutrally-marked strain of *Pseudomonas fluorescens* SBW25. *J Microbiol Meth* **71**: 78-81.

Zhao, H., Li, X., Johnson, D.E., Blomfield, I. & Mobley, H.L.T. (1997). *In vivo* phase variation of MR/P fimbrial gene expression in *Proteus mirabilis* infecting the urinary tract. *Mol Microbiol* **23**: 1009-1019.

Zieg, J., Silverman, M., Hilmen, M. & Simon, M. (1977). Recombinational switch for gene expression. *Science* **196**: 170-172.

Zordan, R.E., Galgoczy, D.J. & Johnson, A.D. (2006). Epigenetic properties of white-opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. *Proc Natl Acad Sci USA* **103**: 12807-12812.

Zuber, M., Hoover, T.A. & Court, D.L. (1995). Analysis of a *Coxiella burnetii* gene product that activates capsule synthesis in *Escherichia coli*: requirement for the heat shock chaperone DnaK and the two-component regulator RcsC. *J Bacteriol* **177**: 4238-4244.

Appendices

Statistical methods used for calculating differences between capsule proportions: the Shapiro-Wilk test for normality was used to determine whether data colleted from each genotype could have come from a population with a normal distribution. In cases where the Shapiro-Wilk $P<0.05$, the normality assumption required for *t*-tests was considered to be violated, and the M-W-W test was performed to investigate differences between population medians. In cases where the Shapiro-Wilk test $P>0.05$, the normality condition was considered satisfied, and *t*-tests were performed on the data to investigate differences between the population means. When data from two or more genotypes was compared, the Levene test was used to test for equal variances of the data collected from each genotype; where $P>0.05$, a parametric two-sample *t*-test was used, while where $P<0.05$, a non-parametric form of the two-sample *t*-test (the Welch test) was used. All values are quoted to three significant figures. Level of statistical significance is represented as follows: *<0.05, **<0.01, ***<0.001.

A1 Appendix material from Chapter 3

A1.1 Raw data & statistical tests for capsulation in colonies

See section 3.3.2.2.1 for an explanation of the data.

Colony type	Raw data	Mean ± SE[a]	Shapiro-Wilk *P*-value
SBW25	0.002, 0.008, 0.002, 0.008, 0.000	0.004 ± 0.00167	0.111
$1s^4$	0.004, 0.004, 0.000, 0.008, 0.004	0.004 ± 0.00126	0.325
$1w^4$-translucent	0.040, 0.106, 0.174, 0.148, 0.122	0.118 ± 0.0227	0.789
$1w^4$-opaque	0.782, 0.646, 0.850, 0.664, 0.518	0.692 ± 0.0575	0.875

Table A1.1.1: Raw data and preliminary statistics on the proportion of cells capsulated in colonies derived from various genotypes. [a]SE=standard error.

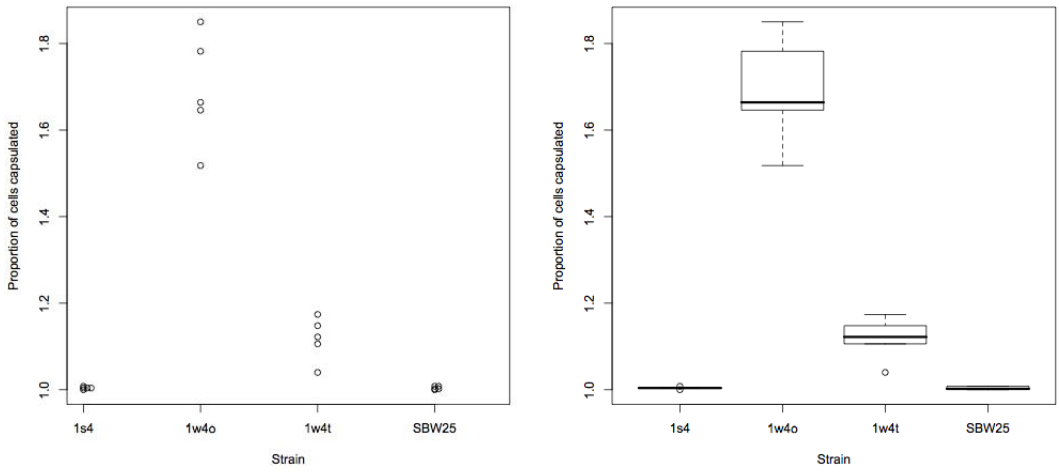


Figure A1.1.1: Comparative dotplots (left) and boxplots (right) of the data in Table A1.1.1. 1w4o=1w^4 opaque colonies, 1w4t=1w^4 translucent colonies.

Comparison[a]		**Levene Test**[b]		***T*-test**[c]		
1	**2**	***F*-stat**	***P*-value**	**Type**	***P*-value**	**95 % C.I.**
SBW25	1s^4	0.514	0.494	Parametric	1.00	-0.00484, 0.00484
1s^4	1w^4-tr.	5.40	0.0487*	Welch	7.26 x 10^{-3}**	-0.0511, -0.177
1s^4	1w^4-op.	6.42	0.0351*	Welch	2.79 x 10^{-4}***	-0.528, -0.848
1w^4-tr.	1w^4-op.	2.23	0.177	Parametric	1.48 x 10^{-5}***	-0.431, -0.717

Table A1.1.2: Comparative statistics on the proportion of cells capsulated in colonies derived from indicated genotypes. [a]Comparison of proportion of cells capsulated in colony types 1 and 2. [b]Levene test statistics: *F*-statistic (*F*-stat) and *P*-value given. [c]The *t*-test *P*-value is for the hypothesis that the true populations means are equal; the 95 % confidence interval (C.I.) estimates the size of the difference.

A1.2 Raw data & statistical tests for capsulation in 1w^4-Δ*wss*

See section 3.3.2.2.2 for an explanation of the data.

Genotype	**Raw data**	**Mean ± SE**[a]	**Shapiro-Wilk *P*-value**
1s^4	0.000, 0.002, 0.002, 0.000, 0.004	0.0016 ± 0.000748	0.314
1w^4	0.062, 0.086, 0.088, 0.102, 0.084	0.0844 ± 0.00643	0.515

Genotype	Raw data	Mean ± SE[a]	Shapiro-Wilk *P*-value
1w^4-Δ*wss*	0.078, 0.080, 0.072, 0.098, 0.112	0.0880 ± 0.00662	0.412

Table A1.2.1: Raw data and preliminary statistics on the proportion of cells capsulated in detailed genotypes. [a]SE=standard error.

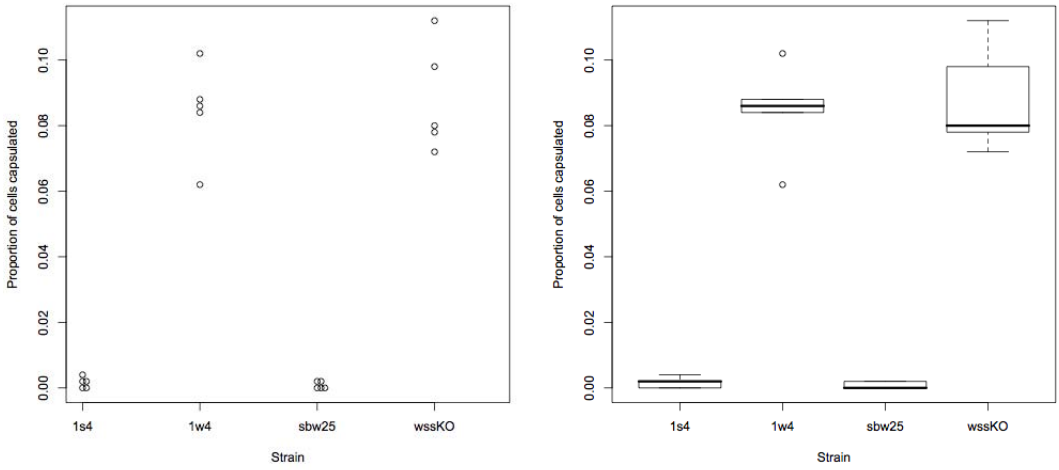


Figure A1.2.1: Comparative dotplots (left) and boxplots (right) of the data in Table A1.2.1. wssKO=1w^4-Δ*wss*.

Comparison[a]		Levene Test[b]		*T*-test[c]		
1	**2**	***F*-stat**	***P*-value**	**Type**	***P*-value**	**95 % C.I.**
1s^4	1w^4	2.52	0.151	Parametric	1.32 x 10^{-6}***	-0.0679, -0.0977
1w^4	1w^4-Δ*wss*	0.178	0.684	Parametric	0.723	-0.0262, 0.0190

Table A1.2.2: Comparative statistics on the proportion of cells capsulated in indicated genotypes. [a]Comparison of proportion of cells capsulated in genotypes 1 and 2. [b]Levene test statistics: *F*-statistic (*F*-stat) and *P*-value given. [c]The *t*-test *P*-value is for the hypothesis that the true populations means are equal; the 95 % confidence interval (C.I.) estimates the size of the difference.

A1.3 Raw data & statistical tests for capsulation in 1w^4 overnight cultures

See section 3.3.3 for an explanation of the data.

Genotype	Raw data	Mean ± SE[a]	Shapiro-Wilk *P*-value
$1w^4$	0.494, 0.502, 0.442, 0.484, 0.474	0.479 ± 0.0104	0.529

Table A1.3.1: Raw data and normality test for the proportion of cells capsulated in the culture used to demonstrate reversible switching between $1w^4$ phenotypic forms. [a]SE=standard error.

A2 Appendix material from Chapter 4

A2.1 Genotype & phenotype of slow-growing $1w^4$ transposon mutants

See section 4.3.1 for an explanation of the data.

Strain	C#[a]	Pflu[b]	Gene[c]	Predicted protein function	Insertion[d]	Cap[e]
Mutants with insertion site determined						
JG83	20	0367	*hutH*	Histidine ammonia-lyase	U	-2
JG112	23	0475	-	Carbamoyl transferase	538025	-2
JG172	40	0475	-	Carbamoyl transferase	537238	-2
JG25	5	0478	-	Glycosyl transferase	U	-1
JG33	8	0478	-	Glycosyl transferase	U	-1
JG42	10	0517	*amiC*	N-acetylmuramoyl-l-alanine amidase	U	-2
JG86	17	0734	-	Outer membrane lipoprotein	U	-1
JG180	41	0739	*hemK*	Glutamine N-methyltransferase	U	-2
JG66	17	0863	*mreB*	Cell shape-determining protein	U	-2
JG105	22	0863	*mreB*	Cell shape-determining protein	U	-1
JG15	2	0864	*mreC*	Cell shape-determining protein	U	-1
JG100	22	0864	*mreC*	Cell shape-determining protein	U	-1
JG179	41	0864	*mreC*	Cell shape-determining protein	U	-2
JG34	8	1548	-	ABC transporter	U	-2
JG8	1	2773	-	Conserved hypothetical protein	U	-2
JG14	2	3801	*ftsK*	Cell division protein	4194738	-1
JG71	18	3801	*ftsK*	Cell division protein	U	-2

Strain	**C#**[a]	**Pflu**[b]	**Gene**[c]	**Predicted protein function**	**Insertion**[d]	**Cap**[e]
JG85	20	3801	*ftsK*	Cell division protein	U	-1
JG93	21	3801	*ftsK*	Cell division protein	U	-1
JG162	37	IG	P*hupB*	Promoter of DNA-binding protein HU-beta	U	-1
JG159	37	4709	-	Conserved hypothetical protein	5184063	-1
JG98	21	4874	*rluB*	Ribosomal large subunit pseudouridine synthase B	U	-2
JG60	15	5399	-	Conserved hypothetical protein	U	-2
JG20	3	5400	-	Thiamine biosynthesis-related protein	U	-2
JG155	36	5400	-	Thiamine biosynthesis-related protein	U	-1
JG24	5	5401	*hemL*	Glutamate-1-semialdehyde aminotransferase	U	-2
JG84	20	IG	*rpoB*	RNA-polymerase beta subunit	6062131	-2
JG58	14	5843	*potG*	Putrescine transport ATP-binding protein	U	-2
JG37	8	5961	*rep*	ATP-dependent DNA helicase	U	-2
JG69	17	6030	-	Membrane-bound sugar binding protein	U	-1
JG104	22	6129	*gidA*	Glucose inhibited division protein A	6712943	-1
Mutants for which insertion site remains unidentified						
JG12	2	-	-	-	-	-2
JG19	3	-	-	-	-	-2
JG29	6	-	-	-	-	-2
JG50	11	-	-	-	-	-2
JG78	19	-	-	-	-	-2
JG89	21	-	-	-	-	-2
JG92	21	-	-	-	-	-2
JG94	21	-	-	-	-	-2
JG95	21	-	-	-	-	-2
JG96	21	-	-	-	-	-1
JG97	21	-	-	-	-	-2
JG99	22	-	-	-	-	-2
JG107	22	-	-	-	-	-2

Strain	C#[a]	Pflu[b]	Gene[c]	Predicted protein function	Insertion[d]	Cap[e]
JG115	23	-	-	-	-	-2
JG116	23	-	-	-	-	-2
JG117	23	-	-	-	-	-2
JG118	23	-	-	-	-	-2
JG121	23	-	-	-	-	-1
JG122	23	-	-	-	-	-1
JG126	24	-	-	-	-	-2
JG127	24	-	-	-	-	-2
JG146	29	-	-	-	-	-2
JG151	33	-	-	-	-	-2
JG182	41	-	-	-	-	-1

Table A2.1.1: Insertion sites and capsule phenotypes for non-sectoring transposon mutants that achieve less than 50 % of 1w^4 OD_{600} in overnight KB culture. [a]Conjugation number during which transconjugant was isolated, [b]Pflu number, the designated number of the gene on the SBW25 chromosome (IG=intergenic), [c]where appropriate, gene name assigned on basis of BLASTP (dash=previously un-named gene, P=promoter), [d]exact point of insertion in the 5'→3' direction of the SBW25 chromosome (U=unidentified), [f]capsule phenotype of mutant (-2=no capsules, -1=some capsules but fewer than 1w^4).

A2.2 Raw data & statistical tests for capsulation in 1w^4 transposon mutants

This data is reported throughout section 4.3.1.

[a]Genotype	[b]Gene	[c]Phenotype	Raw Data	[d]Median
1s^4	-	-2	0.004, 0.000, 0.000, 0.002, 0.002	0.002
1w^4	-	0	0.082, 0.076, 0.074, 0.078, 0.072	0.076
Cre-deletions of insertions in genes involved in biosynthesis of colanic acid precursors				
JG73Δ	*pgi*	-2	0.000, 0.006, 0.000, 0.000, 0.000	0.000
JG55Δ	*algC*	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
JG103Δ	*udg*	-2	0.000, 0.002, 0.000, 0.000 0.000	0.000

[a]Genotype	[b]Gene	[c]Phenotype	Raw Data	[d]Median
JG114Δ	*galU*	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
Cre-deletions of insertions in the colanic acid biosynthetic operon				
JG108Δ	Pflu3654	-1	0.002, 0.002, 0.006, 0.004, 0.006	0.004
JG137Δ	P3655-7	-2	0.000, 0.000, 0.000, 0.000, 0.002	0.000
JG26Δ	Pflu3656	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
JG5Δ	*wcaJ*	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
JG51Δ	Pfu3659	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
JG28Δ	*wza*	-2	0.000, 0.000, 0.002, 0.000, 0.000	0.000
JG21Δ	Pflu3663	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
JG7Δ	*wcaI*	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
JG124Δ	*waaE*	-2	0.000, 0.002, 0.000, 0.000, 0.000	0.000
JG53Δ	*wcaF*	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
JG35Δ	Pflu3671	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
JG150Δ	Pflu3673	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
JG152Δ	Pflu3674	-2	0.000, 0.000, 0.000, 0.000, 0.002	0.000
JG18Δ	Pflu3675	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
JG103Δ	*udg*	-2	0.000, 0.002, 0.000, 0.000, 0.000	0.000
JG27Δ	*wzc*	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
Cre-deletions of insertions in regulators of colanic acid biosynthesis				
JG74Δ	*gacA*	-2	0.000, 0.002, 0.000, 0.000, 0.002	0.000
JG138Δ	*barA*	-2	0.000, 0.000, 0.000, 0.000, 0.000	0.000
Cre-deletions of insertions in other genes				
JG148Δ	IG1858-9 (tRNA)	-1	0.014, 0.008, 0.012, 0.018, 0.016	0.014
JG158Δ	*sahA*	-2	0.000, 0.002, 0.000, 0.000, 0.000	0.000
JG176Δ	*ndk*	+1	0.544, 0.534, 0.554, 0.616, 0.542	0.544

Table 2.2.1: Raw data and median proportion of cells capsulated for $1s^4$, $1w^4$ and selected non-switching, Cre-deleted $1w^4$ IS-Ω-Km/hah genotypes. [a]In case of Cre-deleted transposon mutants, "Cre" has been removed from the end of each name for brevity. *indicates that JG103Δ is duplicated as it

belongs to two categories. [b]Where gene name is unavailable, the Pflu number of the gene is given (P=promoter, IG=intergenic) [c]Phenotypes relate to the production of capsules: 0=same cap+/cap- ratio as $1w^4$, -1=fewer capsules than $1w^4$, -2=fewer capsules than -1, +1=more capsules than $1w^4$. [d]Medians were calculated in Excel (as samples sizes were small and samples non-normally distributed, standard errors could not be meaningfully estimated).

As the raw data samples for many of the above transposon mutants did not satisfy the normality assumption for a *t*-test for a difference between means. Instead, medians were used as central measures, and the M-W-W test was performed for each strain in Table A2.2.1 against $1w^4$. Every transposon mutant (and $1s^4$) except JG176-ΔCre had a significantly lower proportion of capsulated cells than $1w^4$ (*U* statistic=0, one-tailed $P=3.97 \times 10^{-3}$). The M-W-W test for JG176-ΔCre (*ndk*) against $1w^4$ showed that JG176-ΔCre populations contain a significantly greater proportion of capsulated cells (*U*-statistic=0, one-tailed $P=3.97 \times 10^{-3}$).

A3 Appendix material from Chapter 5

A3.1 Raw data & statistical tests for capsulation in *carB* allelic replacements

See section 5.3.4.1.4 for an explanation of the data.

Genotype	Raw data	Mean ± SE[a]	Shapiro-Wilk *P*-value
SBW25	0.002, 0.002, 0.004, 0.000, 0.000	0.0016 ± 0.000748	0.314
$1s^4$	0.000, 0.002, 0.002, 0.000, 0.004	0.0016 ± 0.000748	0.314
$1w^4$	0.062, 0.086, 0.088, 0.102, 0.084	0.0844 ± 0.00643	0.515
$1s^4$-*carB*mut	0.084, 0.076, 0.066, 0.082, 0.072	0.0760 ± 0.00323	0.787
$1w^4$-*carB*wt	0.002, 0.004, 0.000, 0.000, 0.002	0.00160 ± 0.00748	0.314
SBW25-*carB*mut	0.056, 0.060, 0.052, 0.038, 0.070	0.0552 ± 0.00524	0.939

Table A3.1.1: Raw data and preliminary statistics on the proportion of cells capsulated in detailed genotypes. [a]SE=standard error.

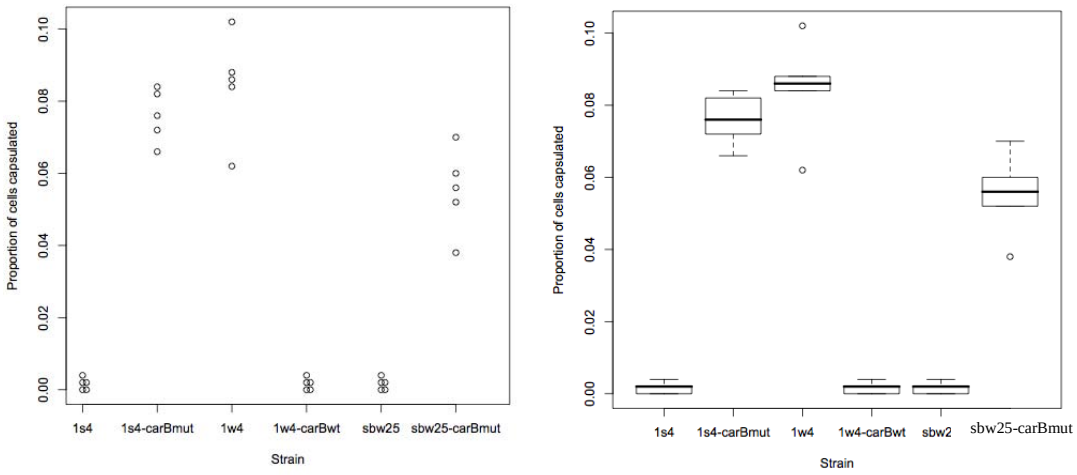


Figure A3.1.1: Comparative dotplots (left) and boxplots (right) of the data in Table A3.1.1.

Comparison[a]		**Levene Test**[b]		***T*-test**[c]		
1	**2**	***F*-stat**	***P*-value**	**Type**	***P*-value**	**95 % C.I.**
SBW25	$1s^4$	0	1	Parametric	1	-0.00244, 0.00244
$1s^4$	$1w^4$	2.52	0.151	Parametric	1.32×10^{-6}***	-0.0679, -0.0977
$1w^4$	$1s^4$-*carB*mut	0.4	0.545	Parametric	0.278	-0.0251, 0.00825
$1s^4$	$1w^4$-*carB*wt	0	1	Parametric	1	-0.00244, 0.00244
$1w^4$	$1w^4$-*carB*wt	2.52	0.151	Parametric	1.32×10^{-6}***	0.0679, 0.0977
$1s^4$	SBW25-*carB*mut	3.91	0.0835	Parametric	7.71×10^{-6}***	-0.0414, -0.0658
$1w^4$	SBW25-*carB*mut	0.0187	0.895	Parametric	7.84×10^{-3}**	0.0101, 0.483

Table A3.1.2: Comparative statistics on the proportion of cells capsulated in colonies derived from various genotypes. [a]Comparison of proportion of cells capsulated in colony types 1 and 2. [b]Levene test statistics: *F*-statistic (*F*-stat) and *P*-value given. [c]The *t*-test *P*-value tests the hypothesis that the true populations means are equal; the 95 % confidence interval (C.I.) estimates the size of the difference.

A3.2 Raw data & statistical tests for relative fitness experiments

For an explanation of the data see section 5.3.5.2. Each of the six tables below contains data from a competition experiments between the two genotypes indicated.

A: $1w^4$ (1) and $1s^4$ (2)

Rep.	Counts at 0 h			Counts at 72 h			Malthusian		Relative Frequency
	D.	$1w^4$	$1s^4$	D.	$1w^4$	$1s^4$	$1w^4$	$1s^4$	
1	25	35	84	25	60	40	7.45	6.17	1.21
2	25	22	55	25	68	52	8.04	6.85	1.17
3	25	41	80	25	40	38	6.88	6.16	1.12
4	25	27	54	25	84	66	8.04	7.11	1.13
5	25	34	88	25	43	119	7.14	7.21	0.991
6	25	28	48	25	100	64	8.18	7.20	1.14
7	25	34	54	25	47	37	7.23	6.53	1.11
8	25	21	62	25	57	27	7.91	6.08	1.30
9	25	45	109	25	51	34	7.03	5.74	1.22
10	25	18	36	50	72	36	7.60	6.21	1.22
Mean RF ± 95 % C.I.									1.16 ± 0.0269

B: $1s^4$-*carB*mut (1) and $1s^4$ (2)

Rep.	Counts at 0 h			Counts at 72 h			Malthusian		Relative Frequency
	D.	1	2	D.	1	2	1	2	
1	25	20	42	50	62	35	7.35	6.03	1.22
2	25	37	70	25	15	28	6.00	5.99	1.00
3	25	41	65	50	55	68	6.51	6.26	1.04
4	25	18	51	50	63	64	7.47	6.44	1.16
5	25	28	58	50	78	46	7.24	5.98	1.21
6	25	25	53	50	126	65	7.83	6.42	1.22
7	25	82	127	*Contaminated with yeast*					n/a
8	25	35	63	50	177	106	7.84	6.73	1.16
9	25	59	64	50	76	60	6.47	6.15	1.05
10	25	26	59	50	112	97	7.68	6.71	1.14
Mean RF ± 95 % C.I.									1.13 ± 0.0262

C: 1w^4 (1) and 1w^4-*carB*wt (2)

Rep.	Counts at 0 h			Counts at 72 h			Malthusian		Relative Frequency
	D.	1	2	D.	1	2	1	2	
1	25	45	93	25	72	50	7.38	6.29	1.17
2	25	27	65	25	30	40	7.01	6.42	1.09
3	25	62	121	25	101	36	7.40	5.70	1.30
4	25	44	58	25	63	28	7.27	6.18	1.18
5	25	41	54	25	51	45	7.13	6.73	1.06
6	25	29	66	25	26	32	6.80	6.18	1.10
7	25	34	56	25	57	33	7.42	6.38	1.16
8	25	21	51	50	83	45	7.59	6.09	1.25
9	25	45	93	50	126	43	7.24	5.44	1.33
10	25	45	55	*Contaminated with yeast*					n/a
Mean RF ± 95 % C.I.									1.18 ± 0.0296

D: SBW25-*carB*mut (1) and SBW25-*lacZ* (2), 24 hour competition

Rep.	Counts at 0 h			Counts at 24 h			Malthusian		Relative Frequency
	D.	1	2	D.	1	2	1	2	
1	25	22	32	50	14	32	5.76	6.21	0.927
2	25	10	40	50	17	43	6.75	6.29	1.07
3	25	26	55	50	15	28	5.66	5.54	1.02
4	50	53	67	50	10	23	5.24	5.84	0.897
5	25	32	44	50	17	40	5.58	6.12	0.912
6	25	36	44	25	14	35	5.96	6.68	0.893
7	25	19	48	50	17	63	6.10	6.49	0.941
8	25	32	68	50	15	44	5.46	5.78	0.944
9	25	17	60	25	6	16	5.87	5.59	1.05
10	25	30	112	50	11	19	5.21	4.44	1.17
Mean RF ± 95 % C.I.									0.983 ± 0.0293

E: SBW25-*carB*mut (1) and SBW25-*lacZ* (2), 48 hour competition

Rep.	Counts at 0 h			Counts at 48 h			Malthusian		Relative Frequency
	D.	1	2	D.	1	2	1	2	
1	25	38	64	50	6	15	4.37	4.76	0.917
2	50	25	106	50	19	28	6.63	5.58	1.19
3	25	29	53	50	29	29	6.21	5.61	1.11
4	25	19	45	50	25	16	6.49	5.18	1.23
5	50	27	42	50	14	24	6.25	6.35	0.985
6	25	33	45	25	34	22	6.94	6.19	1.12
7	25	23	64	50	8	5	5.16	3.67	1.41
8	25	32	69	50	37	41	6.36	5.69	1.12
9	25	11	49	50	17	18	6.65	5.21	1.28
10	50	56	156	50	27	27	6.18	5.15	1.20
Mean RF ± 95 % C.I.									1.16 ± 0.0449

F: SBW25-*carB*mut (1) and SBW25-*lacZ* (2), 72 hour competition

Rep.	Counts at 0 h			Counts at 72 h			Malthusian		Relative Frequency
	D.	1	2	D.	1	2	1	2	
1	25	24	48	50	7	15	4.98	5.05	0.986
2	25	43	81	50	31	11	5.89	4.22	1.40
3	25	36	59	50	1	14	2.63	4.78	0.55
4	25	14	49	50	5	12	5.18	4.81	1.08
5	25	27	43	50	6	5	4.71	4.06	1.16
6	25	29	65	50	13	15	5.41	4.75	1.14
7	25	35	93	50	2	13	3.35	4.25	0.789
8	25	17	59	50	14	9	6.02	4.33	1.39
9	50	78	247	50	4	17	3.94	4.23	0.930
10	25	22	82	50	111	171	7.83	6.95	1.13
Mean RF ± 95 % C.I.									1.05 ± 0.0814

Tables A3.2.1 (A-F): Raw data counts, dilutions (D.), Malthusian parameters and relative fitness values for indicated competitions. Counts were obtained by plating either 25 µl or 50 µl samples (as indicated) of a 10^6-fold dilution for each replicate. All Malthusian parameters, relative fitness (RF) values and confidence intervals (C.I.) quoted to three significant figures.

A3.2.1	Genotype 1	Genotype 2	Mean RF±CI[a]	S.-W.[b]	*T*-test[c]
A	$1w^4$	$1s^4$	1.16 ± 0.0269	0.801	2.04×10^{-4}***
B	$1s^4$-*carB*mut	$1s^4$	1.13 ± 0.0262	0.146	1.25×10^{-3}**
C	$1w^4$	$1w^4$-*carB*wt	1.18 ± 0.0296	0.583	4.00×10^{-4}***
D (24 h)	SBW25-*carBmut*	SBW25-*lacZ*	0.983 ± 0.0293	0.116	0.5854
E (48 h)	SBW25-*carBmut*	SBW25-*lacZ*	1.16 ± 0.0499	0.947	6.74×10^{-3}**
F (72 h)	SBW25-*carBmut*	SBW25-*lacZ*	1.05 ± 0.0814	0.642	0.5187

Table A3.2.2: Statistical tests on raw data for each competition listed in Tables A3.2.1 A-F. [a]Mean relative frequency (RF) ± 95 % confidence interval (CI), [b]*P*-value for Shapiro-Wilk (S.-W.) test for normality, [c]*P*-value for two-tailed one-sample *t*-test that the relative frequency differs from 1.

A3.3 Raw data & statistical tests for WS evolution

For an explanation of the data see section 5.3.5.2.1.

A: 24 hours

Replicate	Evolution of WS types							
	SBW25-*carB*mut		SBW25-*lacZ*		Total		Proportion	
	0 h	24 h	0 h	24 h	0 h	24 h	0 h	24 h
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0
Mean proportion of WS at 24 hours ± SE							0 ± 0	

B: 48 hours

Replicate	Evolution of WS types							
	SBW25-*carB*mut		**SBW25-*lacZ***		**Total**		**Proportion**	
	0 h	48 h	0 h	48 h	0 h	48 h	0 h	48 h
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
9	0	0	0	1	0	1	0	0.0278
10	0	0	0	0	0	0	0	0
Mean proportion of WS at 48 hours ± SE							0.00278 ± 0.00278	

C: 72 hours

Replicate	Evolution of WS types							
	SBW25-*carB*mut		**SBW25-*lacZ***		**Total**		**Proportion**	
	0 h	72 h	0 h	72 h	0 h	72 h	0 h	72 h
1	0	0	0	1	0	1	0	0.0435
2	0	1	0	7	0	8	0	0.160
3	0	1	0	0	0	1	0	0.0625
4	0	0	0	1	0	1	0	0.0556
5	0	0	0	2	0	2	0	0.154
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	3	0	3	0	6	0	0.207

Replicate	Evolution of WS types							
	SBW25-*carB*mut		**SBW25-*lacZ***		**Total**		**Proportion**	
	0 h	72 h	0 h	72 h	0 h	72 h	0 h	72 h
9	0	0	0	0	0	0	0	0
10	0	15	0	20	0	35	0	0.110
Mean proportion of WS at 72 hours ± SE							0.0793 ± 0.0236	

Tables A3.3.1 (A-C): Raw data counts for number and proportions of WS types evolved in the 24, 48 and 72-hour SBW25-*carB*mut vs. SBW25-*lacZ* competitions. Proportions calculated by adding the total number of colonies counted for each replicate (including the SBW25-*carB*mut and SBW25-*lacZ* colony counts listed in the equivalent A3.2.1 tables).

A4 Appendix materials for Chapter 6

A4.1 Statistics for capsulation in switcher genotypes re-evolved from $1s^4$

See section 6.3.1.1 for an explanation of the data.

Genotype	Raw data	Mean ± SE[a]	Shapiro-Wilk *P*-value
$1s^4$	0.002, 0.000, 0.006, 0.000, 0.002	0.00200 ± 0.00098	0.146
$1w^4$	0.084, 0.084, 0.086, 0.098, 0.062	0.0828 ± 0.0052	0.311
$1w^4$–reN1.5	0.004, 0.010, 0.006, 0.010, 0.004	0.00680 ± 0.0012	0.0857
$1w^4$-reD2	0.030, 0.028, 0.022, 0.028, 0.026	0.0268 ± 0.0012	0.493
$1w^4$-reN1.2	0.026, 0.030, 0.014, 0.038, 0.032	0.0280 ± 0.0036	0.709
$1w^4$-reD12	0.056, 0.036, 0.036, 0.026, 0.034	0.0376 ± 0.0044	0.192
$1w^4$-reD1.8	0.088, 0.076, 0.050, 0.112, 0.112	0.0876 ± 0.010	0.461
$1w^4$-reN1.4	0.196, 0.256, 0.228, 0.232, 0.302	0.243 ± 0.016	0.813

Table A4.1.1: Raw data and preliminary statistics on the proportion of cells capsulated in detailed genotypes. [a]SE=standard error.

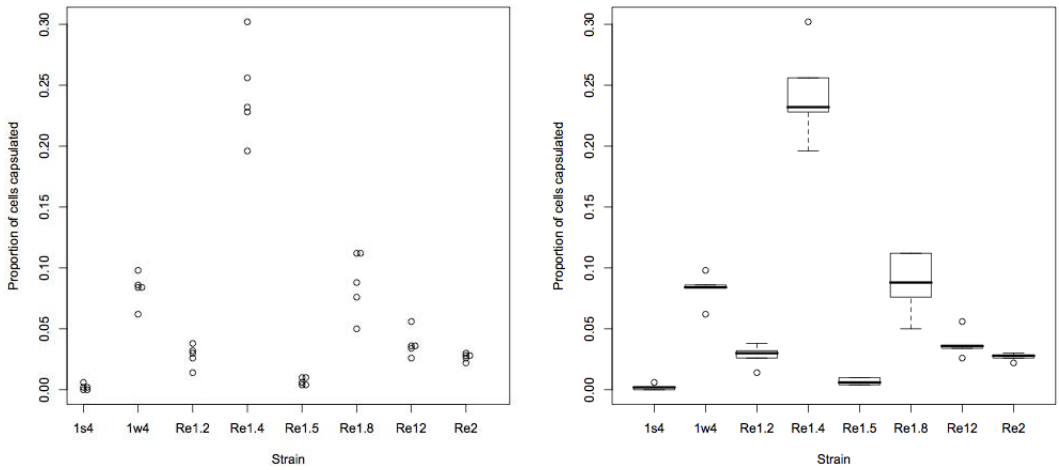


Figure A4.1.1: Comparative dotplots (left) and boxplots (right) of the data in Table A4.1.1. Re-evolved genotype names have been shortened; e.g. Re1.4=1w^4-reN1.4.

Comparison[a]		**Levene Test**[b]		***T*-test**[c]		
1	**2**	***F*-stat**	***P*-value**	**Type**	***P*-value**	**95 % C.I.**
1s^4	1w^4	1.77	0.220	Parametric	7.99 x 10^{-7}***	-0.0671, -0.0945
1s^4	Re1.5	0.571	0.471	Parametric	0.0249*	-0.000779, -0.00882
1w^4	Re1.5	1.33	0.282	Parametric	1.37 x 10^{-6}***	0.0622, 0.0898
1s^4	Re2	0.0909	0.771	Parametric	5.81 x 10^{-7}***	-0.0208, -0.0288
1w^4	Re2	1.50	0.256	Parametric	1.37 x 10^{-5}***	0.0422, 0.0698
1s^4	Re1.2	2.26	0.171	Parametric	2.41 x 10^{-4}***	-0.0164, -0.0356
1w^4	Re1.2	0.0922	0.769	Parametric	5.42 x 10^{-5}***	0.0385, 0.0711
1s^4	Re12	1.48	0.258	Parametric	1.11 x 10^{-4}***	-0.0239, -0.0473
1s^4	Re12	0.0415	0.844	Parametric	3.56 x 10^{-4}***	0.0276, 0.0628
1w^4	Re1.8	7.80	0.0234*	Welch	1.77 x 10^{-3}**	-0.0532, -0.118
1s^4	Re1.8	2.37	0.162	Parametric	0.723	-0.0349, 0.0253
1w^4	Re1.4	3.96	0.0818	Parametric	8.01 x 10^{-7}***	-0.200, -0.281
1s^4	Re1.4	2.05	0.190	Parametric	2.53 x 10^{-5}***	-0.117, -0.203

Table A4.1.2: Comparative statistics on the proportion of cells capsulated in indicated genotypes. [a]Comparison of proportion of cells capsulated in genotypes 1 and 2 (re-evolved genotype names have

been shortened; e.g. Re1.4=1w^4-reN1.4). [b]Levene test statistics: *F*-statistic (*F*-stat) and *P*-value given. [c]The *t*-test *P*-value is for the hypothesis that the true populations means are equal; the 95 % confidence interval (C.I.) estimates the size of the difference.

A4.2 Statistics for capsulation in gene deletion genotypes derived from 1w^4

See sections 6.3.3 and 6.3.4.2.1 for an explanation of the data.

Genotype	Raw data	Mean ± SE[a]	Shapiro-Wilk *P*-value
1s^4	0.000, 0.002, 0.002, 0.000, 0.004	0.0016 ± 0.000748	0.314
1w^4	0.062, 0.086, 0.088, 0.102, 0.084	0.0844 ± 0.00643	0.515
1w^4-Δ*arcB*	0.084, 0.104, 0.068, 0.092, 0.090	0.0876 ± 0.00526	0.858
1w^4-Δ*carB*	0.076, 0.140, 0.102, 0.126, 0.064	0.102 ± 0.0144	0.710
1w^4-Δ*recA*	0.070, 0.068, 0.108, 0.090, 0.096	0.0864 ± 0.00768	0.502

Table A4.2.1: Raw data and preliminary statistics on the proportion of cells capsulated in detailed genotypes. [a]SE=standard error.

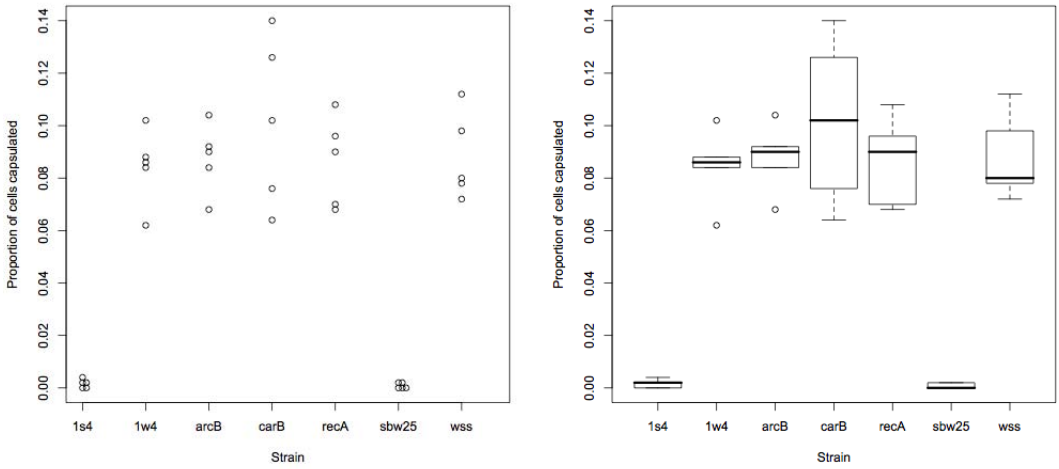


Figure A4.2.1: Comparative dotplots (left) and boxplots (right) of the data in Table A4.2.1. Strain names indicate the deleted gene (e.g. arcB=1w^4-Δ*arcB*).

Comparison[a]		Levene Test[b]		*T*-test[c]		
1	**2**	***F*-stat**	***P*-value**	**Type**	***P*-value**	**95 % C.I.**
$1s^4$	$1w^4$	2.52	0.151	Parametric	1.32×10^{-6}***	-0.0679, -0.0977
$1w^4$	$1w^4$-Δ*arcB*	0	1	Parametric	0.723	-0.0233, 0.0169
$1w^4$	$1w^4$-Δ*carB*	3.79	0.0873	Parametric	0.307	-0.0535, 0.0191
$1w^4$	$1w^4$-Δ*recA*	0.469	0.513	Parametric	0.847	-0.0251, 0.0211

Table A4.2.2: Comparative statistics on the proportion of cells capsulated in various genotypes. [a]Comparison of proportion of cells capsulated in genotypes 1 and 2. [b]Levene test statistics: *F*-statistic (*F*-stat) and *P*-value given. [c]The *t*-test *P*-value is for the hypothesis that the true populations means are equal; the 95 % confidence interval (C.I.) estimates the size of the difference.

A4.3 Statistics for effect of addition of 2 mM uracil and/or arginine to $1w^4$

See section 6.3.4.1 for an explanation of the data.

Medium	Raw data	Mean ± SE[a]	Shapiro-Wilk *P*-value
KB	0.154, 0.166, 0.112, 0.128, 0.172	0.146 ± 0.0114	0.521
KB+arginine	0.128, 0.136, 0.160, 0.176, 0.190	0.158 ± 0.0105	0.705
KB+Uracil	0.012, 0.022, 0.016, 0.010, 0.010	0.0140 ± 0.00228	0.207
KB+arginine+uracil	0.010, 0.010, 0.022, 0.016, 0.012	0.0140 ± 0.00228	0.207

Table A4.3.1: Raw data and preliminary statistics on the proportion of cells capsulated in $1w^4$ populations grown in detailed media. [a]SE=standard error.

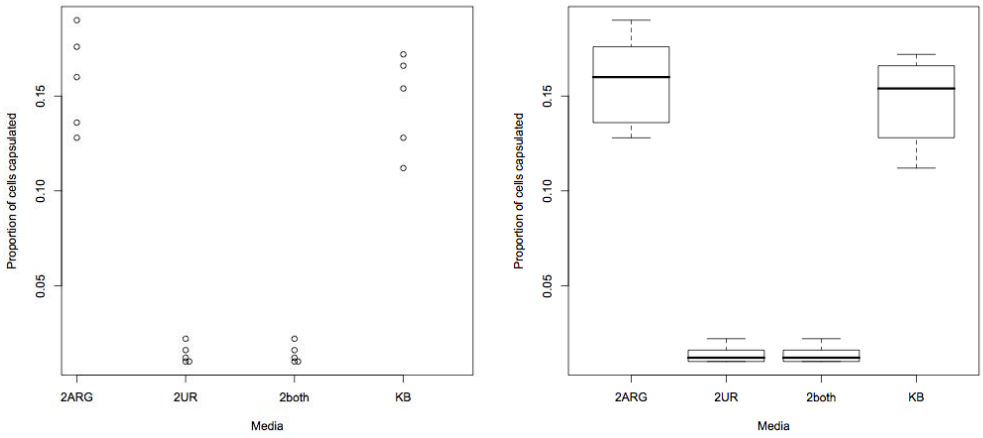


Figure A4.3.1: Comparative dotplots (left) and boxplots (right) of the data in Table A4.3.1. 2ARG=2 mM arginine, 2UR=2 mM uracil, 2both=2 mM arginine and 2 mM uracil.

Comparison to KB[a]	Levene Test[b]		*T*-test[c]		
	***F*-stat**	***P*-value**	**Type**	***P*-value**	**95 % C.I.**
KB+Arg	0.00770	0.932	Parametric	0.499	-0.0261, 0.0493
KB+Ur	4.89	0.0579	Parametric	3.28×10^{-6}***	0.105, 0.159
KB+Arg+Ur	4.89	0.0579	Parametric	3.28×10^{-6}***	0.105, 0.159

Table A4.3.2: Comparative statistics on the proportion of cells capsulated in $1w^4$ populations grown on various media (Arg=2 mM arginine, Ur=2 mM uracil). [a]Comparison of proportion of cells capsulated in KB and indicated media. [b]Levene test statistics: *F*-statistic (*F*-stat) and *P*-value given. [c]The *t*-test *P*-value is for the hypothesis that the true populations means are equal; the 95 % confidence interval (C.I.) estimates the size of the difference.

A4.4 Statistics for effect of uracil addition to re-evolved switcher genotypes

See section 6.3.4.2 for an explanation of the data. For raw data and statistics on each strain in KB medium, see Tables A4.1.1.

Genotype	Raw data	Mean ± SE[a]	Shapiro-Wilk[b]
$1s^4$	0.000, 0.002, 0.004, 0.002, 0.000	0.00160 ± 0.000748	0.314
$1w^4$	0.012, 0.022, 0.016, 0.010, 0.010	0.0140 ± 0.00228	0.207

Genotype	Raw data	Mean ± SE[a]	Shapiro-Wilk[b]
$1w^4$–reN1.5	0.004, 0.002, 0.000, 0.002, 0.000	0.0016 ± 0.000748	0.314
$1w^4$-reD2	0.002, 0.008, 0.006, 0.004, 0.000	0.00400 ± 0.00126	0.967
$1w^4$-reN1.2	0.000, 0.004, 0.006, 0.004, 0.004	0.00360 ± 0.000980	0.135
$1w^4$-reD12	0.000, 0.004, 0.000, 0.006, 0.002	0.00240 ± 0.00117	0.421
$1w^4$-reD1.8	0.008, 0.012, 0.000, 0.018, 0.014	0.0104 ± 0.00306	0.814
$1w^4$-reN1.4	0.142, 0.250, 0.281, 0.206, 0.288	0.233 ± 0.0270	0.446

Table A4.4.1: Raw data and preliminary statistics on the proportion of cells capsulated in detailed genotypes in KB+2 mM uracil. [a]SE=standard error. [b]*P*-value for Shapiro-Wilk normality test.

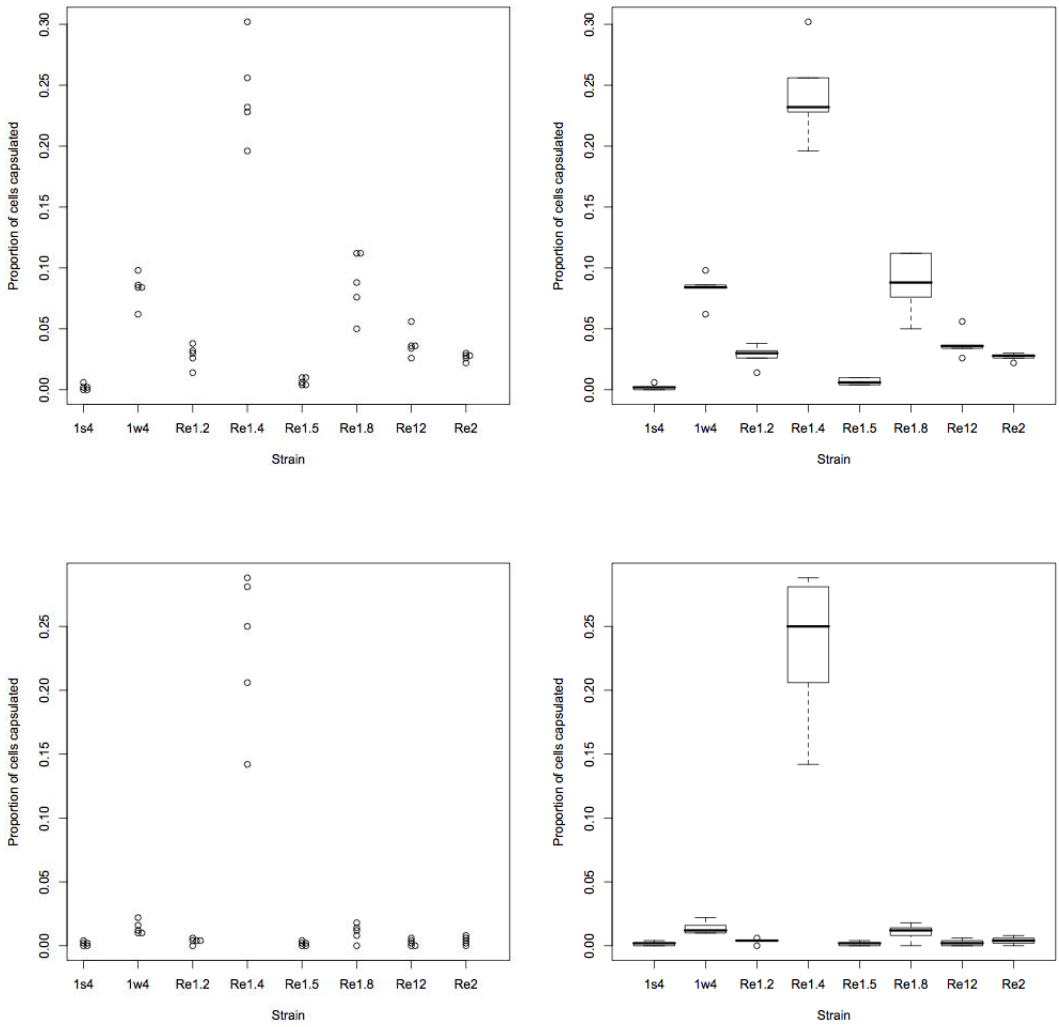


Figure A4.4.1: Comparative dotplots (left) and boxplots (right) of the data in Tables A4.1.1 (KB, top) and A4.3.1 (KB+2 mM uracil, bottom). Re-evolved genotype names have been shortened; e.g. Re1.4=$1w^4$-reN1.4.

Comparison (KB vs. KB+Ur)[a]	Levene Test[b]		*T*-test[c]		
	***F*-stat**	***P*-value**	**Type**	***P*-value**	**95 % C.I.**
$1s^4$	0.2	0.667	Parametric	0.771	-0.00266, 0.00346
$1w^4$	4.89	0.0579	Parametric	3.28×10^{-6}***	0.105, 0.159
$1w^4$–reN1.5	1.8	0.217	Parametric	1.45×10^{-2}*	0.00144, 0.00896
$1w^4$-reD2	0.0909	0.771	Parametric	2.71×10^{-6}***	0.0183, 0.0273
$1w^4$-reN1.2	2.67	0.141	Parametric	3.52×10^{-4}***	0.0149, 0.0339
$1w^4$-reD12	1.26	0.294	Parametric	1.23×10^{-4}***	0.0235, 0.0469
$1w^4$-reD1.8	4.85	0.0589	Parametric	2.13×10^{-4}***	0.0493, 0.0105
$1w^4$-reN1.4	0.642	0.446	Parametric	0.778	-0.0650, 0.0838

Table A4.4.2: Comparative statistics on the proportion of cells capsulated in genotypes grown in KB & KB+2 mM uracil. [a]Comparison of proportion of cells capsulated in genotypes grown in each medium. [b]Levene test statistics: *F*-statistic (*F*-stat) and *P*-value given. [c]*T*-test *P*-value is for the hypothesis that the populations means are equal; 95 % confidence interval (C.I.) estimates the size of the difference.

A4.5 Statistics for effect of GTP addition on capsulation

See section 6.3.5 for an explanation of the data.

Strain	Medium	Raw data	Mean ± SE[a]	Shapiro-Wilk *P*-value
$1s^4$	KB	0.000, 0.004, 0.002, 0.004, 0.000	0.00200 ± 0.000894	0.119
	KB+1.5 mM GTP	0.000, 0.002, 0.002, 0.004, 0.004	0.00240 ± 000748	0.314
$1w^4$	KB	0.134, 0.182, 0.154, 0.222, 0.210	0.180 ± 0.0165	0.735
	KB+0.1 mM GTP	0.242, 0.252, 0.296, 0.362, 0.300	0.290 ± 0.0213	0.531
	KB+0.5 mM GTP	0.298, 0.332, 0.456, 0.320, 0.322	0.346 ± 0.0282	0.0213
	KB+1.5 mM GTP	0.418, 0.348, 0.362, 0.298, 0.336	0.352 ± 0.0195	0.865
Re1.4	KB	0.284, 0.264, 0.258, 0.216, 0.252	0.255 ± 0.0111	0.665
	KB+1.5 mM GTP	0.316, 0.310, 0.206, 0.250, 0.294	0.275 ± 0.0208	0.341

Table A4.5.1: Raw data and preliminary statistics on the proportion of cells capsulated in $1w^4$ populations grown in detailed media. [a]SE=standard error. Re1.4=$1w^4$-reN1.4.

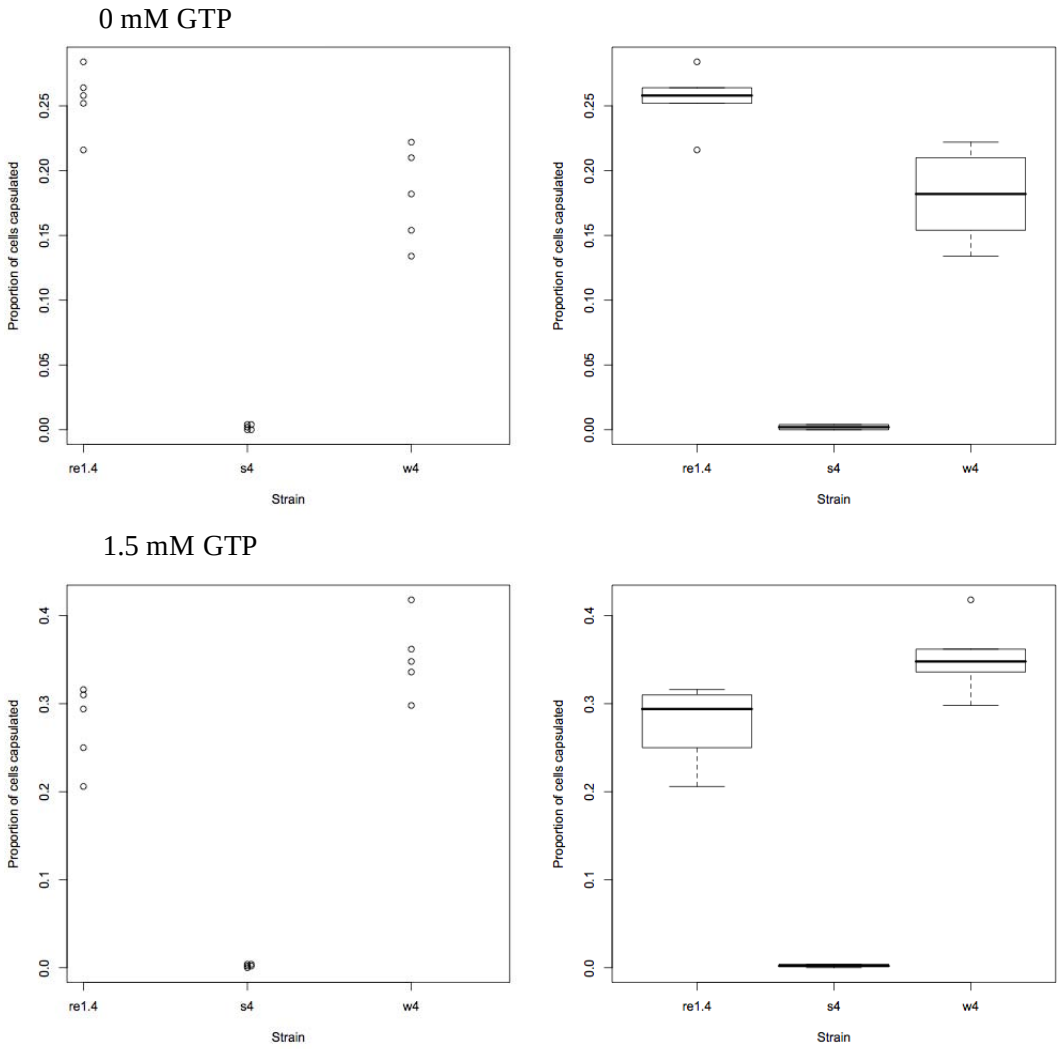


Figure A4.5.1: Comparative dotplots (left; top no GTP, bottom 1.5 mM GTP) and boxplots (right; top no GTP, bottom 1.5 mM GTP) of the data in Table A4.5.1. s^4=1s^4, w^4=1w^4, re1.4=1w^4-reN1.4.

Comparison of 0 & 1.5 mM GuHCl	Levene Test[a]		*T*-test[b]		
	***F*-stat**	***P*-value**	**Type**	***P*-value**	**95 % C.I.**
1s^4	0.4	0.54474	Parametric	0.7404	-0.00309, 0.00229
1w^4	0.00067	0.98004	Parametric	1.496 x 10^{-4}***	-0.231, -0.113
1w^4-reN1.4	1.10354	0.32418	Parametric	0.412	-0.0340, 0.0748

Table A4.5.2: Comparative statistics on the proportion of cells capsulated in indicated genotypes grown on KB and KB+ 1.5 mM guanine hydrochloride (GuHCl). [a]Levene test statistics: *F*-statistic (*F*-stat) and *P*-value given. [b]The *t*-test *P*-value is for the hypothesis that the true populations means are equal; the 95 % confidence interval (C.I.) estimates the size of the difference.

A4.6 Raw data & statistical tests for over-expression experiments

See section 6.3.6 for an explanation of the data.

Genotype	Raw data	Mean ± SE[a]	Overall Mean ± SE[b]	Shapiro-Wilk *P*-value[c]
$1w^4$	1: 0.274, 0.264, 0.214 2: 0.224, 0.254, 0.188 3: 0.252, 0.258, 0.200	0.251 ± 0.0186 0.222 ± 0.0191 0.237 ± 0.0184	0.236 ± 0.0102	0.354
$1w^4$-pSX	1: 0.066, 0.064, 0.050 2: 0.078, 0.088, 0.058 3: 0.048, 0.076, 0.138	0.0600 ± 0.00503 0.0747 ± 0.00882 0.0873 ± 0.0266	0.0740 ± 0.00912	0.0429*
$1w^4$-pSX-*carB*	1: 0.004, 0.004, 0.000 2: 0.004, 0.004, 0.008 3: contaminated	0.00267 ± 0.0013 0.00533 ± 0.0013 n/a	0.00400 ± 0.00103	0.101
$1w^4$-pSX-*pyrH*	1: 0.006, 0.004, 0.004 2: 0.008, 0.006, 0.008 3: 0.004, 0.008, 0.002	0.0047 ± 0.00067 0.0073 ± 0.00067 0.0047 ± 0.00176	0.00556 ± 0.000729	0.198
$1w^4$-pSX-*ndk*	1: 0.048, 0.014, 0.024 2: 0.006, 0.008, 0.016 3: 0.000, 0.008, 0.022	0.0287 ± 0.0101 0.01 ± 0.00306 0.01 ± 0.00643	0.0162 ± 0.00473	0.152
$1w^4$-pSX-*galU*	1: 0.074, 0.070, 0.100 2: contaminated 3: 0.098, 0.094, 0.070	0.0813 ± 0.00940 n/a 0.0873 ± 0.00874	0.0843 ± 0.00590	0.312
SBW25	1: 0.006, 0.002, 0.000 2: 0.002, 0.004, 0.000 3: 0.008, 0.000, 0.002	0.0027 ± 0.00176 0.002 ± 0.00115 0.0033 ± 0.00240	0.00267 ± 0.000943	0.113
SBW25-pSX	1: 0.002, 0.004, 0.002 2: 0.002, 0.000, 0.004 3: 0.000, 0.004, 0.004	0.0027 ± 0.00067 0.002 ± 0.00115 0.0027 ± 0.00133	0.00244 ± 0.000556	0.0254*
SBW25-pSX-*carB*	1: 0.006, 0.000, 0.000 2: 0.000, 0.000, 0.004 3: 0.004, 0.002, 0.000	0.002 ± 0.002 0.00133 ± 0.0013 0.002 ± 0.00115	0.00178 ± 0.000778	0.0107*

SBW25-pSX-*pyrH*	1: 0.000, 0.000, 0.000 2: 0.002, 0.000, 0.000 3: 0.002, 0.000, 0.000	0 ± 0 0.0007 ± 0.00067 0.0007 ± 0.00067	0.000444 ± 0.000294	1.69 x 10^{-5}***
SBW25-pSX-*ndk*	1: 0.000, 0.000, 0.000 2: 0.000, 0.000, 0.000 3: 0.000, 0.000, 0.000	0 ± 0 0 ± 0 0 ± 0	0 ± 0	n/a (all zero)
SBW25-pSX-*galU*	1: 0.002, 0.004, 0.004 2: 0.004, 0.000, 0.000 3: 0.000, 0.002, 0.000	0.0033 ± 0.00067 0.00133 ± 0.0013 0.0007 ± 0.00067	0.00178 ± 0.000619	0.0119*
$1w^4$-reN1.4	1: 0.270, 0.296, 0.334 2: 0.296, 0.278, 0.472 3: 0.324, 0.308, 0.508	0.300 ± 0.0186 0.349 ± 0.0619 0.380 ± 0.0642	0.343 ± 0.287	7.86 x 10^{-3}**
Re1.4-pSX	1: 0.230, 0.160, 0.204 2: 0.210, 0.164, 0.136 3: 0.158, 0.244, 0.168	0.198 ± 0.0204 0.170 ± 0.0216 0.190 ± 0.0272	0.186 ± 0.0123	0.421
Re1.4-pSX-*carB*	1: 0.188, 0.206, 0.152 2: 0.180, 0.162, 0.206 3: 0.208, 0.222, 0.176	0.182 ± 0.0159 0.198 ± 0.0204 0.202 ± 0.0136	0.189 ± 0.00780	0.709
Re1.4-pSX-*pyrH*	1: 0.014, 0.016, 0.022 2: 0.000, 0.008, 0.004 3: 0.006, 0.016, 0.006	0.0173 ± 0.00240 0.004 ± 0.00231 0.0093 ± 0.00333	0.0102 ± 0.00237	0.695

Table A4.6.1: Raw data and preliminary statistics on the proportion of cells capsulated in detailed genotypes. Three biological replicates exist for each genotype (with the exceptions of $1w^4$-pSX-*carB* and $1w^4$-pSX-*galU*, for which there are only two), and three replicate populations were assayed for each. [a]Mean and standard error (SE) for each biological replicate. [b]Where all three means were comparable (all error bars overlapping), an overall mean and SE was calculated using all replicates as independent samples. [c]Shapiro-Wilk test performed using all replicates as independent samples.

Comparison[a]		**Test for difference between population means**		
1	**2**	**Levene *P*[b]**	**Type[c]**	***P*-value[d]**
Comparisons in the $1w^4$ genetic background				
$1w^4$	pSX	n/a	M-W-W	4.11 x 10^{-5}***

Comparison[a]		Test for difference between population means		
1	**2**	**Levene *P*[b]**	**Type[c]**	***P*-value[d]**
pSX	pSX-*carB*	n/a	M-W-W	4.00×10^{-4}***
pSX	pSX-*pyrH*	n/a	M-W-W	4.11×10^{-5}***
pSX	pSX-*ndk*	n/a	M-W-W	4.11×10^{-5}***
pSX	pSX-*galU*	n/a	M-W-W	0.351
Comparison in the SBW25 genetic background				
SBW25	pSX	n/a	M-W-W	0.863
pSX	pSX-*carB*	n/a	M-W-W	0.436
pSX	pSX-*pyrH*	n/a	M-W-W	0.0188*
pSX	pSX-*ndk*	n/a	M-W-W	0.00399**
pSX	pSX-*galU*	n/a	M-W-W	0.489
Comparison in the 1w^4-reN1.4 genetic background				
reN1.4	pSX	n/a	M-W-W	4.11×10^{-5}***
pSX	pSX-*carB*	0.282	Parametric two-sample *t*-test	0.846
pSX	pSX-*pyrH*	0.0168*	Welch two-sample *t*-test	3.32×10^{-7}***

Table A4.6.2: Comparative statistics on the proportion of cells capsulated in colonies derived from various genotypes. [a]Comparison of proportion of cells capsulated in genotypes 1 and 2 (1w^4-reN1.4 shortened to reN1.4). [b]Levene test for equal variances performed where Shapiro-Wilk $P>0.05$ for both genotypes (see Table A4.6.1). [c]Type of test determined by Shapiro-Wilk *P*-values (where either genotype's $P<0.05$=M-W-W; both $P>0.05$=two-sample *t*-test) and Levene *P*-values ($P>0.05$=parametric *t*-test; $P<0.05$=Welch *t*-test). [d]*P*-value for indicated test for difference in population means.

A5 Appendix materials for Chapter 7

A5.1 Raw data & statistical tests for capsulation in *rpoD* allelic replacements

See section 7.3.5 for an explanation of the data.

Genotype	Raw data	Mean ± SE[a]	Shapiro-Wilk[b]
SBW25	0.002, 0.002, 0.004, 0.000, 0.000	0.00160 ± 0.000748	0.314
$6s^4$	0.000, 0.000, 0.002, 0.002, 0.004	0.00160 ± 0.000748	0.314
$6w^4$	0.620, 0.454, 0.498, 0.426, 0.552	0.510 ± 0.0348	0.823
$6s^4$-*rpoD*mut	0.562, 0.426, 0.340, 0.344, 0.614	0.457 ± 0.0562	0.280
SBW25-*rpoD*mut	0.362, 0.364, 0.310, 0.276, 0.236	0.310 ± 0.0247	0.493
$1w^4$	0.062, 0.086, 0.088, 0.102, 0.084	0.0844 ± 0.00643	0.515

Table A5.1.1: Raw data and preliminary statistics on the proportion of cells capsulated in detailed genotypes. [a]SE=standard error. [b]*P*-value for Shapiro-Wilk normality test given.

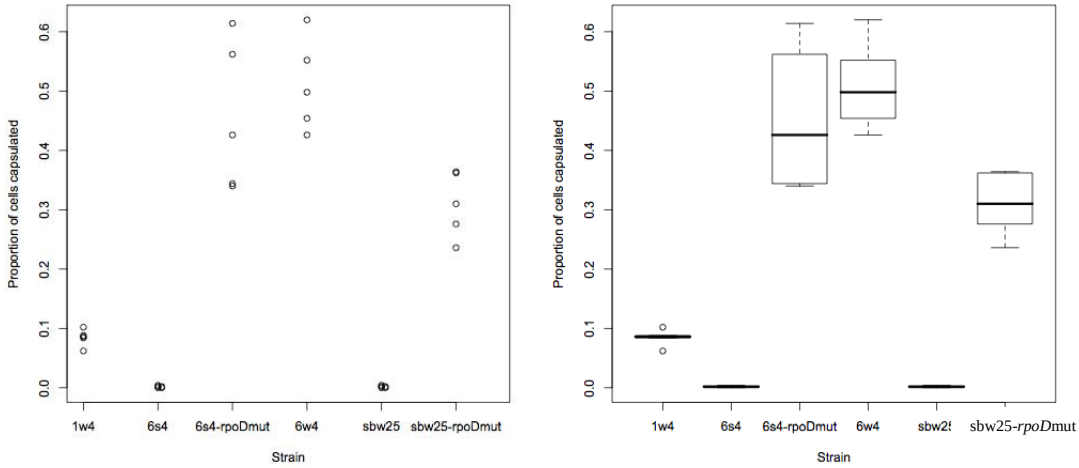


Figure A5.1.1: Comparative dotplots (left) and boxplots (right) of the data in Table A5.1.1.

Comparison[a]		Levene Test[b]		*T*-test[c]		
1	**2**	***F*-stat**	***P*-value**	**Type**	***P*-value**	**95 % C.I.**
SBW25	$6s^4$	0	1	Parametric	1	-0.00244, 0.00244
$6s^4$	$6w^4$	8.31	0.0204*	Welch	1.27×10^{-4}***	-0.412, -0.605
$6s^4$	$6s^4$-*rpoD*mut	9.66	0.0145*	Welch	1.25×10^{-3}**	-0.300, -0.612
$6w^4$	$6s^4$-*rpoD*mut	1.17	0.311	Parametric	0.447	-0.0995, 0.205
$6s^4$	SBW25-*rpoD*mut	11.2	0.0102	Welch	2.38×10^{-4}***	-0.239, -0.377
$6w^4$	SBW25-	0.444	0.524	Parametric	1.55×10^{-3}**	0.102, 0.299

Comparison[a]		Levene Test[b]		*T*-test[c]		
1	**2**	***F*-stat**	***P*-value**	**Type**	***P*-value**	**95 % C.I.**
6w^4	*rpoD*mut 1w^4	5.91	0.0411*	Welch	1.86 x 10^{-4}***	0.330, 0.521

Table A5.1.2: Comparative statistics on the proportion of cells capsulated in various genotypes. [a]Comparison of proportion of cells capsulated in genotypes 1 and 2. [b]Levene test statistics: *F*-statistic (*F*-stat) and *P*-value given. [c]The *t*-test *P*-value tests the hypothesis that the true populations means are equal; the 95 % confidence interval (C.I.) estimates the size of the difference.

A5.2 Raw data & statistical tests for effect of addition of 2 mM uracil to 6w^4

See section 7.3.6.1 for an explanation of the data.

Medium	Raw data	Mean ± SE[a]	Shapiro-Wilk *P*-value
KB	0.540, 0.606, 0.516, 0.490, 0.602	0.551 ± 0.0231	0.386
KB+uracil (2 mM)	0.392, 0.578, 0.624, 0.534, 0.524	0.530 ± 0.0389	0.562

Table A5.2.1: Raw data and preliminary statistics on the proportion of cells capsulated in 6w^4 populations grown in detailed media. [a]SE=standard error.

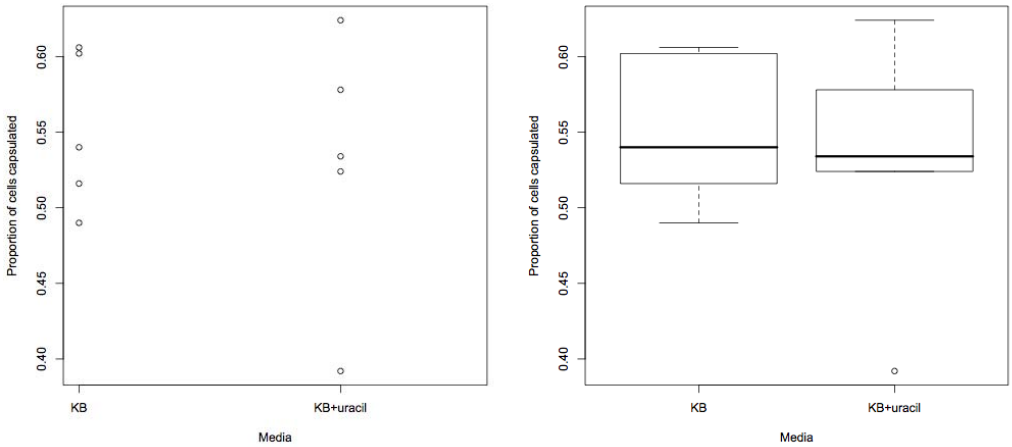


Figure A5.2.1: Comparative dotplots (left) and boxplots (right) of the data in Table A5.2.1.

Comparison to KB	Levene Test[a]		*T*-test[b]		
	***F*-stat**	***P*-value**	**Type**	***P*-value**	**95 % C.I.**
KB+uracil (2 mM)	0.331	0.581	Parametric	0.664	-0.0839, 0.125

Table A5.2.2: Comparative statistics on the proportion of cells capsulated in 6w^4 populations grown on KB+2 mM uracil. [a]Levene test statistics: *F*-statistic (*F*-stat) and *P*-value given. [b]The *t*-test *P*-value is for the hypothesis that the true populations means are equal; the 95 % confidence interval (C.I.) estimates the size of the difference.

A5.3 Statistics for effect of addition of guanine hydrochloride to 6s^4 and 6w^4

See section 7.3.6.2 for an explanation of the data.

Medium	Raw data	Mean ± SE[a]	Shapiro-Wilk
6s^4			
KB	0.004, 0.002, 0.002, 0.002, 0.000	0.002 ± 0.000632	0.325
KB+GuHCl (1.5 mM)	0.000, 0.006, 0.002, 0.002, 0.004	0.0028 ± 0.00102	0.814
6w^4			
KB	0.738, 0.758, 0.734, 0.794, 0.722	0.749 ± 0.0126	0.437
KB+GuHCl (1.5 mM)	0.720, 0.818, 0.778, 0.850, 0.710	0.775 ± 0.0271	0.559

Table A5.3.1: Raw data and preliminary statistics on the proportion of cells capsulated in 6s^4 and 6w^4 populations grown in detailed media (GuHCl=guanine hydrochloride). [a]SE=standard error.

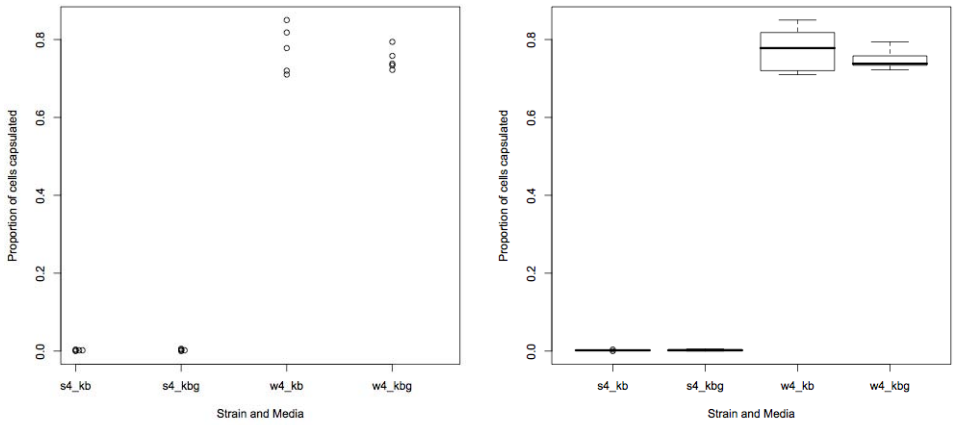


Figure A5.3.1: Comparative dotplots (left) and boxplots (right) of the data in Table A5.3.1. s4=6s^4, w4=6w^4, kb=KB medium, kbg=KB medium containing 1.5 mM guanine hydrochloride.

Comparison of KB and KBG	Levene Test[a]		*T*-test[b]		
	***F*-stat**	***P*-value**	**Type**	***P*-value**	**95 % C.I.**
$6s^4$	0.8	0.397	Parametric	0.524	-0.00197, 0.00357
$6w^4$	2.98	0.122	Parametric	0.410	-0.0430, 0.0950

Table A5.3.2: Comparative statistics on the proportion of cells capsulated in $6s^4$ and $6w^4$ populations grown on KB and KB+1.5 mM guanine hydrochloride (KBG) [a]Levene test statistics: *F*-statistic (*F*-stat) and *P*-value given. [b]The *t*-test *P*-value is for the hypothesis that the true populations means are equal; the 95 % confidence interval (C.I.) estimates the size of the difference.

A5.4 Raw data & statistical tests for line six over-expression experiments

See section 7.3.6.3 for an explanation of the data.

Genotype	Raw data	Mean ± SE[a]	Overall Mean±SE[b]	Shap.-Wilk[c]
	1: 0.682, 0.586, 0.754	0.674 ± 0.0487		
$6w^4$	2: 0.786, 0.726, 0.666	0.726 ± 0.0346	0.702 ± 0.0224	0.619
	3: 0.622, 0.732, 0.760	0.705 ± 0.0421		
	1: 0.276, 0.342, 0.124	0.247 ± 0.0645		
$6w^4$-pSX	2: 0.192, 0.244, 0.188	0.208 ± 0.0180	0.238 ± 0.0279	0.841
	3: 0.180, 0.210, 0.386	0.259 ± 0.0643		
	1: 0.366, 0.660,0.616	0.547 ± 0.0916		
$6w^4$-pSX-*carB*	2: 0.568, 0.732, 0.354	0.551 ± 0.109	0.515 ± 0.0447	0.502
	3: 0.428, 0.462, 0.448	0.446 ± 0.00987		
	1: 0.300, 0.208, 0.184	0.231 ± 0.0354		
$6w^4$-pSX-*pyrH*	2: 0.182, 0.212, 0.158	0.184 ± 0.0156	0.213 ± 0.0190	0.0210*
	3: 0.186, 0.172, 0.318	0.225 ± 0.0465		
	1: 0.022, 0.096, 0.062	0.06 ± 0.0214		
$6w^4$-pSX-*ndk*	2: 0.024, 0.048, 0.056	0.0423 ± 0.00961	0.0482 ± 0.00892	0.898
	3: 0.072, 0.042, 0.012	0.042 ± 0.0173		
	1: 0.216, 0.244, 0.280	0.247 ± 0.0185		
$6w^4$-pSX-*galU*	2: 0.230, 0.196, 0.250	0.225 ± 0.0158	0.236 ± 0.0119	0.993
	3: contaminated	n/a		

Table A5.4.1: Raw data and preliminary statistics on the proportion of cells capsulated in detailed genotypes. Each genotype has three biological replicates (with the exception of 6w^4-pSX-*galU*, for which there are only two), and three replicate populations were assayed *per* biological replicate. [a]Mean and standard error (SE) for each biological replicate. [b]Where all three means were comparable (all error bars overlapping), an overall mean and SE was calculated using all biological and non-biological replicates as independent samples. [c]The Shapiro-Wilk test was performed using all replicates as independent samples.

Comparison[a]		**Test for difference between population means**		
1	**2**	**Levene *P*[b]**	**Type[c]**	***P*-value[d]**
6w^4	pSX	0.759	Parametric two-sample *t*-test	1.14×10^{-11}***
pSX	pSX-*carB*	0.0631	Parametric two-sample *t*-test	1.10×10^{-5}***
pSX	pSX-*pyrH*	n/a	M-W-W	0.248
pSX	pSX-*ndk*	0.103	Parametric two-sample *t*-test	5.09×10^{-7}***
pSX	pSX-*galU*	0.202	Parametric two-sample *t*-test	0.696

Table A5.4.2: Comparative statistics on the proportion of cells capsulated in cultures derived from various 6w^4-based genotypes. [a]Comparison of proportion of cells capsulated in genotypes 1 and 2. [b]Levene test for equal variances performed where Shapiro-Wilk $P>0.05$ for both genotypes (see Table A4.6.1). [c]Type of test determined by Shapiro-Wilk *P*-values (where either genotype's $P<0.05$=M-W-W test; both $P>0.05$=two-sample *t*-test) and Levene test *P*-values ($P>0.05$=parametric *t*-test; $P<0.05$=Welch *t*-test). [d]*P*-value for indicated test for difference in population means.