How bias in the production of phenotypic variation shapes and is shaped by adaptive evolution

Dissertation

in fulfilment of the requirements for the degree

Doctor rerum naturalium (*Dr. rer. nat.*)

of the Faculty of Mathematics and Natural Sciences

at Kiel University

submitted by

Michael Barnett

Department of Microbial Population Biology

Max Planck Institute for Evolutionary Biology

2022

First examiner:	Prof. Dr. Paul Rainey
Second examiner:	Prof. Dr. Hinrich Schulenburg
Date of oral examination:	17th of November, 2022

Zusammenfassung

Diese Arbeit stellt Evolutionsexperimente an einem bakteriellen Modellorganismus vor, welche zeigen, wie die Prozesse, die phänotypische Variation erzeugen, mit der natürlichen Selektion interagieren. Insbesondere untersuchen die Experimente, auf welche Weise die Systeme, die die Variabilität eines Organismus definieren - die Genotyp-Phänotyp-Abbildung (G-P-Abbildung) und die Mutationsprozesse, die diese Abbildung neu konfigurieren - Wahrscheinlichkeitsverzerrungen bei der Erzeugung von Variation schaffen und wie diese Verzerrungen den Verlauf der adaptiven Evolution beeinflussen, und von ihr beeinflusst werden.

In Kapitel 2 untersuche ich, ob die klassische darwinistische Vorstellung vom "Überleben des Stärkeren" ausreicht, um die Phänotypen zu verstehen, die im Zuge der Anpassung entstehen, oder ob das Ergebnis auch von der Wahrscheinlichkeit bestimmt wird, dass ein bestimmter Phänotyp durch Mutation entsteht. Insbesondere führte ich ein Experiment durch, in dem sich Populationen wiederholt an die gleiche Umgebung anpassen mussten, wobei die sich entwickelnden adaptiven Lösungen durch genetische Manipulationen entfernt wurden. Indem ich das Evolutionsexperiment mit einem so manipulierten Stamm wiederholte, bei dem die zuvor entdeckten Phänotypen also nicht mehr verfügbar waren, war die Evolution gezwungen, während jeder Anpassungsrunde neue lebensfähige, phänotypische Lösungswege zu finden. Durch diesen Prozess fand ich heraus, dass die Phänotypen, die die Evolution anfangs bevorzugte, nicht unbedingt die fittesten erreichbaren Möglichkeiten waren; es gab latent Phänotypen mit höherer Fitness, die die Evolution jedoch nicht nutzte. Die Rekonstruktion der für jeden Phänotyp erforderlichen Mutationspfade, und die Bewertung sowohl der Fitness als auch der Wahrscheinlichkeit jedes Mutationsschritts zeigte, dass die Wahrscheinlichkeit eines Phänotypen, durch Mutationen zu entstehen,eine entscheidende Rolle bei der natürlichen Auslese spielt. Diese Ergebnisse zeigten, dass die adaptive Evolution zuweilen auf dem wahrscheinlichsten Mutationspfad verläuft und zu einem suboptimalen Phänotyp führt: das "Überleben des Wahrscheinlichsten".

Da Wahrscheinlichkeitsverzerrungen bei der Erzeugung von Variationen selbst ein Produkt der Evolution sind, besteht die Möglichkeit, dass sie durch natürliche Selektion geformt werden können, um die Anpassung zu erleichtern - ein umstrittener Gedanke, der als "Evolution der Evolvierbarkeit" bekannt ist. In Kapitel 3 stelle ich ein Evolutionsexperiment vor, bei dem explizit nach Evolvierbarkeit selektiert wurde. Hier wurden bakterielle Stämme herausgefordert, wiederholt einen einzigen fokalen Phänotyp zu aktivieren und dann zu inaktivieren, der im Laufe der Zeit abwechselnd vorteilhaft und nachteilig war. Wurde der Zielphänotyp zu einem bestimmten Zeitpunkt nicht erreicht, starb die betreffende Entwicklungslinie aus und wurde durch eine gegenwärtig bestehende Linie aus der

Population ersetzt (geboren). Dieser Prozess des Aussterbens und der Geburt einzelner Linien ermöglichte es der natürlichen Auslese, auf eine Einheit einzuwirken, die sich über die Generationszeit einzelner Zellen hinaus fortpflanzt und somit über einen Zeitraum, in dem die Variation des evolutionären Potenzials, oder der Evolvierbarkeit, für die Auslese sichtbar werden konnte. Durch eine umfassende Sequenzierung des gesamten Genoms konnte ich jede Mutation identifizieren, die entlang der Entwicklungslinie auftrat, und so aufzeigen, wie sowohl die G-P-Abbildung als auch die Verzerrung der Mutationswahrscheinlichkeit die Evolution beeinflussen und von ihr beeinflusst wurden, und wie dies den Erfolg oder Misserfolg einer bestimmte Linie beeinflusste. Ich fand heraus, dass die erfolgreichsten Linien ihre Evolvierbarkeit durch die Verzerrung der Mutationswahrscheinlichkeiten dahingehend erhöhten, dass ein schneller Wechsel zwischen den phänotypischen Zielzuständen ermöglicht wurde. Darüber hinaus wurden aufgrund der erhöhten Geschwindigkeit, mit der der Zielphänotyp in diesen Linien auftrat, zusätzliche Anpassungsschritte möglich, die die Fitness der Zellen in Bezug auf andere Aspekte der Umwelt optimierten - eine Möglichkeit, die in Entwicklungslinien, die nicht über die Fähigkeit zum schnellen Wechsel verfügten, blockiert wurde. Diese Ergebnisse sind der bisher deutlichste experimentelle Beweis für die Fähigkeit der Evolution, selbstverstärkend zu wirken.

(This abstract was kindly translated into German by Michael Schwarz and Michael Sieber)

Abstract

This thesis presents evolution experiments using a bacterial model organism that examine how the processes that generate phenotypic variation interact with the process of natural selection. More specifically, the experiments investigate in what ways the systems that define the variability of an organism – the genotype-phenotype map (G-P map) and the mutational processes by which this map is re-configured – create biases in the production of variation and how these biases both shape and are shaped by the course of adaptive evolution.

In Chapter 2, I examine whether the classical Darwinian notion of the 'survival of the fittest' is sufficient to understand the phenotypes that arise in the course of adaptation or whether the outcome is also determined by the likelihood that a phenotype is generated by mutation. To this end, I conducted an experiment that repeatedly challenged replicate populations to adapt to the same environment, while continually precluding the evolved adaptive solutions via genetic engineering. By re-playing evolution with an engineered strain in which previously discovered phenotypes were unavailable, evolution was forced to find new viable phenotypic solutions during each round of adaptation. Through this process I found that the phenotypes that evolution initially favoured were not necessarily the fittest possibilities available; latent phenotypes of higher fitness existed yet evolution did not make use of them. Reconstruction of the mutational paths required to reach each phenotype, and evaluation of both fitness and likelihood of each mutational step, revealed that the likelihood of the phenotypes being manifest by mutation played a key role in determining their use by natural selection. These results revealed that adaptive evolution may at times proceed along the most likely mutational path to a sub-optimal phenotype: the 'survival of the likeliest'.

As biases in the production of variation are themselves the product of evolution, the possibility exists that they can be shaped by natural selection to facilitate adaptation – a controversial notion known as the 'evolution of evolvability'. In Chapter 3, I present an evolution experiment that explicitly selected for evolvability. Specifically, bacterial lineages were challenged to repeatedly activate and then inactivate a single focal phenotype that was in turn beneficial then deleterious across time. Failure to reach the target phenotypic state at any one point resulted in extinction (death) of that lineage and replacement (birth) by a contemporary extant lineage from the population. This lineage-level death-birth process allowed natural selection to operate on an entity reproducing beyond the generation time of individual cells and therefore over a timescale where variation in evolutionary potential, or evolvability, was visible to selection. Through extensive whole-genome sequencing I identified each mutation that occurred along the trajectory of the evolving lineages, revealing how both the G-P map and mutational biases had shaped and were shaped by evolution, as well as how this influenced the success or failure of a given lineage. I found that the most successful lineages

increased their evolvability through the establishment of a mutational bias that facilitated rapid switching between the target phenotypic states. Moreover, due to the increased speed with which the target phenotype was generated in these lineages, additional adaptive steps became possible that optimized cell fitness with respect to other aspects of the environment – a possibility that was stalled in lineages not possessing the rapid-switching ability. These results represent the clearest experimental evidence yet for the capacity of evolution to act in a self-facilitating manner.

Table of Contents

Chapter 1

Introduction

Contemporary evolutionary theory has its roots in the Modern Synthesis of the 1930s and 1940s, a result of the union between Mendelian genetics with Darwin's Theory of Natural Selection (Fisher, 1930; Wright, 1931; Haldane, 1932; Huxley, 1942). Codified in this framework is a gene-centric and statistical perspective of evolutionary change, the formal theory of this change being described by population genetics. From the population genetic perspective, evolution is seen as "change in the genetic composition of populations" or "change in allele frequencies"; the causes of evolution are therefore the forces that shift the frequencies of existing variants (Dobzhansky, 1982). This perspective is certainly a necessary one – an essential heuristic for understanding and predicting evolution – but as a definition or explanation of evolution it is unsatisfyingly incomplete.

The limitation of the population genetic approach is that it is concerned exclusively with the fate of variants once they have come to exist and ignores how they came to be. In other words, it deals with the consequence of variation in a population rather than the source of variation within the individual organism (Sober, 1984; Amundson, 2005; Stoltzfus, 2021). Population genetics instead bypasses the complex internal structure of organisms by mapping genotype directly to fitness, treating the intermediary steps as a black box (Amundson, 2001; Nuño de la Rosa & Villegas, 2020). As a result, possible influences on evolution arising through steps in the production of variation – through how mutations tend to alter the genotype, through how such mutations are articulated into changes in phenotype, or through how such changes might affect the potential for subsequent organismal change – are inadequately represented in this framework. Understanding these steps in the production of

variation and their role in the evolutionary process is however essential for a complete understanding of evolution. It is these steps that are the focus of this thesis.

That the formal basis for our understanding of evolution emphasizes the consequence of variation and engages little in its production is undoubtedly due in part to technological limitations at the time of the field's major developments. However, it is also the result of influential historical arguments that dismissed the role of internal organismal properties in orienting the course of evolution and promoted natural selection as the only direction-giving factor (Gould, 2002; Amundson, 2005; Stoltzfus & Cable, 2014). In this introduction, I will summarize the relevant historical arguments, their influence on the study of evolution, and how modern responses are reshaping the field, opening up new opportunities to investigate the production of variation and its fundamental role in the evolutionary process.

1.1 Internal versus external forces of evolution

During the nascent period of evolutionary genetics in the late 19th and early 20th centuries, genotypes could only be inferred through the inheritance patterns of phenotypes, and this very much limited what one could learn about the genetic basis of variation. Such studies could only be conducted with sexual organisms in which, through various crossings, geneticists could shuffle pre-existing variation and reveal its phenotypic consequence in the form of altered morphology. Although the word gene existed, there was of course no idea of a gene or their regulation as we understand it today. Genes were the inherited determinants of altered morphologies – distinguishable only if they were located on different chromosomes that segregated, or if they could be physically uncoupled through recombination (reviewed in Portin, 1993). When applied to simple discrete traits controlled by a single locus, this process led to Mendel's famous discoveries of the phenomena of dominance, heterozygosity and recessiveness in his experiments with pea plants (Mendel, 1866). This was the first encounter with a mechanism for the relationship between genotype and phenotype.

Mendel's clear demonstration of the discrete inheritance of discrete phenotypes was naturally seen by some as evidence that evolution proceeded by similar discrete steps of variation, a notion in line with the existing school of thought known as 'saltationism' and seemingly opposed to Darwinian gradualism (Gould, 2002). In the Origin of Species (1859), Darwin had proposed that precise adaptation is possible only if organisms can come to fit their environments by many fine-grained adjustments, each slightly adaptive and accumulating gradually across the immensity of geological time (Darwin, 1859). Darwinian natural selection shaped the form of species from a near completely malleable substrate, as a potter's hands mould clay, leading Poulton (1908) to characterize non-Darwinian views of the role of variation as "the revolt of the clay against the power of the potter".

Gradualism requires variation to be abundant in amount, slight in extent, and non-directional (Gould, 2002).

The history of evolutionary thought during this early period is complex, not only due to the number of differing positions held and a continuity between them, but also due to the mischaracterizations of positions propagated in some popular narratives (Stoltzfus & Cable, 2014; Amundson, 2005). However, the major contention can be summarized with a metaphor appropriated from Galton via Gould (2002). The metaphor aptly summarizes the two explanatory poles of evolution that characterize the argument, both historically and to some extent today: explanations that invoke internal properties of the organism ('orthogenetic' or 'internalist' explanations), and explanations that invoke external forces, specifically natural selection ('adaptationist' or 'externalist' explanations). As re-imagined by Gould[1], the metaphor compares an evolving species to a ball struck on a billiard table. If the ball is of normal spherical shape, then the direction imposed on its 'evolution' is solely imparted by the strike of the cue, as well as the arrangement of existing balls and the table edge – in other words its direction is imposed externally by the environment. If, however, the ball is of irregular shape – a polyhedron with some limited number of facets – then the path it takes will not only be influenced by the cue and table, but also by the angles and facets specific to the ball itself. Evolution would in this case be limited to the possibilities afforded by adjacent planes of the polyhedron and the omnipotence of external forces (i.e., natural selection) would be diminished (Gould, 2002).

The gradualist perspective as advocated by Darwin is analogous to the condition of the spherical ball (variation is abundant, slight, and non-directional) whereas a pattern of saltational evolution results from the tumbling of a polyhedron. So, if variation was generated in discrete steps as Mendel's work had shown, then evolution thereby progresses in steps (i.e., is saltational) and the direction of each step must, by necessity, be due to internal (orthogenetic) properties of the organism. Darwin's insistence on near continuous variation was thought by some of his supporters to be ill-advised, but Darwin knew the ground he would risk giving up – the role of natural selection as the primary creative force in evolution (Gould, 2002). Here, the historical context needs to be appreciated. Darwin's Theory of Evolution by Natural Selection was only one of many theories of evolution at the time, but the first that completely refuted an inherent and progressive ascent in complexity toward humankind (Gould, 2002). Alternative orthogenetic theories at the time explained this process as being driven by some generally mysterious internal mechanism, one imbued with purpose and often associated with the

[1] Galton's original metaphor referred to pushing a "rough stone" - the stone would "resist" movement from external forces and only with sufficient force would a threshold be crossed and topple the stone to a new equilibrium state (Galton, 1869).

divine. Gradualism allowed the question of where adaptive variation came from (a serious problem to the thinkers of the time) to be effectively subsumed under Natural Selection, and a completely mechanistic and non-teleological account of adaptation could be made (Gould, 2002).

That a major tenet of the Modern Synthesis is the kind of continuous phenotypic variation that is able to account for Darwin's gradualism is primarily owed to the work of Fisher (Fisher, 1930). Fisher was a principal founder of both population genetics and the related field of quantitative genetics and was a dedicated and uncompromising proponent of Darwinian gradualism (Plutynski, 2006). He set out to prove the consistency of Mendelian genetics with continuous variation and to mathematise natural selection. Beginning in 1918 and culminating in his foundational work "The Genetical Theory of Evolution" in 1930, Fisher systematically discredited the proposed alternatives to Darwin's gradualist notion of natural selection, leading to the widespread dismissal of the role of internal organismal properties in influencing the course of evolution. It is worth going through each of these arguments, as they were, and continue to be, profoundly influential in the study of evolutionary genetics, adaptation, and evolution in general.

1.1.1 Fisher's arguments for gradualism

1.1.1.1 The infinitesimal model

The first argument addressed the claimed contradiction between discrete inheritance and continuous variation. Fisher resolved this by what became known as the 'infinitesimal model' (Fisher, 1918). The essential idea is that alleles acting at many different loci and contributing only fractionally to a trait can supply near-continuous variation thanks to the combinatorial possibilities afforded by sexual recombination (Turelli, 2017). This conclusion relies on the idealized (large, panmictic, linkage-equilibrated) populations that Fisher assumed. The abundance of possibilities not only supported the notion of abundant variation required for gradualism but also meant the effect of any single allele could be defined as an average across all possible backgrounds in which it occurred (Hansen, 2006). The averaging of allelic effects by this means meant epistasis (the interaction between loci) was negligible: any epistatic effect could be shuffled out of existence and ultimately contributed only statistical noise to the patterns of phenotypic evolution. Thus, the effect of an allele could be divorced from the genetic background in which it occurred, and this internal complexity of the organism could be ignored. Such a decomposition effectively 'solved' the problem of variation for Darwinian natural selection and justified treating the relationship between genotype and phenotype as a fine-grained one-to-one mapping of minute change in genotype to minute change in phenotype – a spherical ball.

1.1.1.2 The opposing forces argument

Fisher's next argument, also given by Haldane (1932), intended to refute the notion of orthogenesis by representing its hypothesized internal tendency to vary (the shape of the polyhedron) as a mutation rate and pitting this against an opposing force of selection. By casting the internal tendency as a force or pressure, rather than a constraint or a bias, Fisher and Haldane set it up as a competitor to natural selection. If the 'pressures' of mutation and selection are opposed, then the frequency of a mutation in a population will be restricted to an equilibrium where these two forces balance. Even very weak selection would set this equilibrium at a very low level. If mutation were to determine the direction of adaptive evolution, it would need to operate at a strength exceeding that of natural selection and mutation rates would therefore need to be orders of magnitude greater than was empirically observed. Fisher (1930) hence concluded: "the whole group of theories which ascribe to hypothetical physiological mechanisms, controlling the occurrence of mutations, a power of directing the course of evolution, must be set aside". Variation was deemed to be non-directional and shaped exclusively by natural selection.

1.1.1.3 The geometric model

Finally, Fisher (1930) presented an argument intended to demonstrate the illogic of saltational evolution (i.e., evolution via steps of discrete variation). The argument relies on a geometric formalism that conceptualizes the process of adaptation as movement of a point (representing a population) in a multi-dimensional space whose axes represent different phenotypic traits of the organism and whose origin represents an optimal combination of trait values. Environmental change displaces the population from the optimum and the population is then guided back by the action of natural selection through mutations that can occur in any direction (isotropic variation). Fisher used this model to demonstrate that mutations of small effect were more likely to be beneficial and therefore more likely to contribute to natural selection – an increasing vector of mutational effect scaling with a decreasing probability of benefit. At the extreme limit, mutations with infinitesimally small effects have a 50% chance of being beneficial in low (one or two) dimensional spaces. Variation was therefore deemed to be small in extent.

1.1.2 Responses to Fisher

Fisher's arguments epitomize the externalist and adaptationist paradigm – so much so that they effectively act as null hypotheses for the detection of other influences on evolution. Responses to these arguments from more modern sources necessarily reflect more internalist or otherwise alternative perspectives, such as those emphasizing the role of genetic drift. In the following sections

I explore both direct responses to Fisher's arguments, as well as a number of indirect but conceptually related 'responses'.

1.1.2.1 Responses to Fisher: the infinitesimal model argument

Fisher's first argument, in which he introduced the infinitesimal model to resolve the apparent contradiction of Mendelian genetics and continuous variation, has somewhat stood the test of time, although more so as a mathematical tool than an accurate reflection of biological reality as Fisher had intended (Hansen, 2006). Despite its idealized assumptions, the infinitesimal model does enable accurate short-term predictions about a population's response to selection and has proved practically useful in animal and plant breeding (Falconer & Mackay 1996). The fundamental assumption – that of the polygenic nature of quantitative traits – has also been validated in many cases by molecular data, variation in human height for example being statistically affected by so many genes that it has been referred to as effectively 'omnigenic' (Boyle et. al 2017). However, empirical studies of quantitative traits have also demonstrated, beyond any doubt, that the effect size of loci is variable and that variation in many traits is underpinned by a small number of loci of large effect alongside a large number of loci of small effect (Roff, 2007).

The infinitesimal model has also been fruitfully extended into the statistical analysis of covarying traits (Lande, 1979). However, this approach explicitly deviates from Fisher's vision in that it incorporates the idea of internal properties of the organism imposing constraints on natural selection – specifically by having some dimensions of phenotypic variation being more variable than others (Cheverud, 1984). Indeed, a seminal study (Schluter, 1996) using this approach showed that patterns of quantitative genetic variance and covariance have biased the amount and direction of morphological evolution in a number of species over millions of years – adaptation appearing to the follow genetic 'lines of least resistance' or the overall direction in which genetic variance is greatest.

1.1.2.2 Responses to Fisher: the opposing forces argument

The 'opposing forces' argument Fisher used to dismiss orthogenetic variation, although logically true in how it was framed, does not actually discount the possibility of orthogenetic change. For example, the argument as stated allows orthogenetic tendencies to dominate in the case of neutral change. Although coined by Wright in 1932, the significance of genetic drift was not fully appreciated until 1968, when Kimura showed the high incidences of mutations in sequence data could not be accounted for by selection (Kimura, 1968; see also King and Jukes, 1969). Biases in the production of variation introduced by internal properties of the organism therefore dominate the most common kind of molecular evolution.

More significantly, as pointed out by Stoltzfus & Yampolsky (2001), pitting mutation rate and selection as opposing forces is not the correct framing to discount the role of biased variation directing evolution in an adaptive context either. Internal properties need not be a driving force that overcomes the force of natural selection in order to direct evolution; they need only to bias the way that variation is introduced into a population. That is, some forms of variation merely need to arrive earlier in the population than others. Stoltzfus & Yampolsky (2001) provide an illustrative example by considering an organism with a latent capacity to mutate into one of two different phenotypes – one with a higher likelihood of being manifest by mutation but with lesser fitness, and one with a lesser likelihood but higher fitness. Their simple population genetic model simulating this scenario demonstrated that, contrary to the opposing forces argument, bias in the introduction process can indeed affect the outcome of evolution without the need for unrealistic mutation rates or neutrality. Rather, the likeliest type can fix in the population before the fittest type has the chance to establish. This idea will be explored further in **Chapter 2 Disentangling Fitness from Likelihood in Adaptive Evolution** of this thesis.

1.1.2.3 Responses to Fisher: the geometric model argument

Fisher invoked the geometric model to justify mutations of small effect being those most likely to be beneficial and therefore most likely to contribute to evolution. However, Kimura (1983) subsequently pointed out that this argument does not hold if one considers the probability of fixation of these small effect mutations. In realistic population sizes, the probability of escape from genetic drift becomes an important factor, and Kimura showed that in this case it is far more likely for mutations of intermediate effect to fix. Fisher's idealized populations had meant drift was irrelevant.

A further revision of the geometric model was made by Orr, who considered not a single mutational step as Fisher and Kimura did, but the entire random walk toward the optimum. His analysis demonstrated that adaptation would first proceed by a few mutations of large effect, followed by diminishing effect sizes as the population approaches the optimum (Orr, 1998). This indeed appears to recapitulate the ubiquitous pattern of diminishing fitness returns in evolving microbial populations (Couce & Tenaillon, 2015). Interestingly, there is one prominent experiment that does support Fisher's notion of beneficial small effect mutations more commonly contributing to evolution than those of large effect (Burch & Chao 1999), but contrary to Fisher's conclusion this applies only to small populations – beneficial mutations of large effect are rare and therefore less likely to be generated in smaller populations.

Under Fisher's geometric model, the omnipotence of natural selection in shaping evolution required variation to be isotropic – it must be equally possible in all directions. The field of evolutionary

developmental biology ('evo-devo') is founded upon the rejection of this isotropism (Salazar-Ciudad, 2021). Morphological variation in multicellular organisms is very obviously not equally possible in all directions but is instead constrained by the internal (i.e., developmental) properties of the organism. In other words, mutation generates variation at the genetic level, but the process of development determines how this genetic variation is translated into the phenotypic variation on which natural selection acts. By seeking to integrate an understanding of this internal developmental process into the evolutionary process, evo-devo represents a modern instantiation of the orthogenetic ideas Fisher had dismissed.

Ideas relating development to evolution have their own long history, most famously represented by Haeckel's recapitulation theory – the notion that "ontogeny recapitulates phylogeny" (Haeckel, 1866; Gould, 1977). Once the Modern Synthesis perspective had taken hold however, the fields of developmental biology and evolutionary biology were mostly studied independently (Gilbert, 2003; Gould, 1977). Two significant exceptions to this were Goldschmidt and Waddington. For example, Goldschmidt's goal was "to convince evolutionists that evolution is not only a statistical genetical problem but also one of the developmental potentialities of the organism." (Goldschmidt et al., 1951). In the same vein Waddington declared, "Changes in genotypes only have ostensible effects in evolution if they bring with them alterations in the epigenetic processes by which phenotypes come into being; the kinds of change possible in the adult form of an animal are limited to the possible alterations in the epigenetic system by which it is produced." (Waddington, 1953).

Evo-devo did not fully take shape until the 1970s when technologies like recombinant DNA and sequencing enabled the genetics underpinning development to be probed. This led to the synthesis of developmental biology and molecular genetics from which modern evo-devo resulted. Catalysing this synthesis was the earlier discovery of the regulation of the lac operon (Jacob & Monod, 1961). The identification of regulatory genes and elements acting as inducers and repressors of structural genes was the first insight into the molecular logic of the genotype to phenotype connection. A further catalyst was the discovery of the surprising similarity in protein-coding genes between apes and humans, which suggested that changes in gene regulation that control the context – the timing, location, and interactions – of these proteins must be critical to morphological evolution (King & Wilson, 1975; Gilbert, 2003). Indeed, a deep understanding of how changes in gene regulation and regulatory networks shape morphological variation has become the defining feature of evo-devo (Carroll, 2008; Uller et al., 2018).

1.2 Bias in the production of phenotypic variation

Aided by the first glimpses into the actual internal properties of organisms, evo-devo reintroduced the study of the production of variation to the study of evolution. In this section, I will briefly describe some examples of how such internal properties can be understood to generate biases in the production of phenotypic variation (relative to an assumption of isotropic variation) and how these biases can influence the course of evolution.

1.2.1 Mutational bias

The first thing to consider is the generation of genetic mutation. Although a common shorthand is to say that mutations are random, this is true only in the sense that they do not preferentially act to increase fitness in the present environment – evolution is 'blind' (Lenski & Mittler, 1993). The manner by which mutations affect the genome is in fact extremely non-random in that changes are not all equally likely. Different kinds of mutations (nucleotide substitutions, duplications, insertions, deletions, inversions) all occur at different and context-specific rates. These contexts may include the local sequence surrounding a genomic location, DNA secondary structures, sequence homology, along with epigenetic influences such as transcriptional activity, chromatin structure, nucleotide methylation, and of course the properties of the cell's replication and repair machinery (Ségurel et al., 2014; Makova & Hardison, 2015; Schroeder et al., 2018). While such biases are well established to affect neutral evolution, a growing body of evidence from both experimental and natural populations is also uncovering the role of these local and global mutational biases in influencing the path of adaptive evolution (Horton et al., 2021; Payne et al., 2019; Stoltzfus & McCandlish, 2017; Lind et al., 2019). An additional source of mutational bias arises from mobile genetic entities like transposons that live within genomes, which, through their own process of replication, generate specific forms of mutation at potentially high rates (Cooper et al., 2001; Lynch & Walsh, 2007). For example, the 'copy and paste' nature of some of these entities in conjunction with carriage of regulatory elements has been implicated in re-wiring of gene regulatory networks in ways that would otherwise be exceedingly improbable (reviewed in Cowley & Oakley, 2013).

1.2.2 The genotype-phenotype map

Even if mutations were truly random in the strict equiprobable sense, the consequence of mutation on the phenotype would not be. Rather, an organism's existing structure and the rules of construction that connect genotype to phenotype – the genotype-phenotype map (G-P map) – dictate that some phenotypic variants will be more readily manifest by mutation than others (Alberch, 1991; Wagner & Altenberg, 1996).

This is most easily illustrated by considering the genetic code and how this connects codons to amino acids (a simple molecular phenotype). If we examine this code, we see it has a particular asymmetrical structure, with some amino acids being coded by more codons than others. For example, the amino acid arginine is coded by six different codons whereas histidine is coded by two and tryptophan by one. Randomly generating an arginine codon is therefore three times as likely as a histidine, and six times as likely as tryptophan. Indeed, King and Jukes (1969) noted that the frequency of amino acids in actual protein sequences closely correspond to that predicted by randomly generated sequences; amino acids represented by multiple codons are more frequently observed than those represented by only one. King (1971) later suggested that the "probability of fixation of an amino acid is a function of its frequency of arising by mutation, and this will happen more often to amino acids with more codons". The asymmetry of the genetic code makes certain phenotypes more likely to be manifest by mutation than others.

Furthermore, the fact the code is redundant – with a many-to-one mapping to most amino acids – not only means the genotype can change while the phenotype does not, but these silent changes can then open up different adjacent possibilities of amino acids that subsequent mutations can access (Cambray & Mazel, 2008). Take the arginine codons CGG and CGT for instance: the former can access five amino acids (leucine, proline, glycine, glutamine, tryptophan) through single point mutations, while the latter loses access to two of these (glutamine, tryptophan) but gains access to three additional amino acids (histidine, serine, cysteine). Thus, the paths available to evolution are contingent upon the sequence that is being modified, but the sequence space can be explored by moving between neutral (synonymous) states that open up new adjacent phenotypic possibilities.

Similar properties are evident for other simple molecular phenotypes that have a predictable 'development' from genotype to phenotype. One such phenotype is RNA secondary structure, which can be predicted by calculating the minimum free-energy fold of a given sequence and is a good proxy for its ultimate (but far more complicated to predict) 3D conformation. RNAs also have actual catalytic, structural, and regulatory functions, thus making this a biologically relevant phenotype (Schuster et al., 1994). The simplicity of RNA secondary structure allows extensive spaces of possible genotypes to be mapped to their corresponding phenotype and the properties of the resulting G-P map to be examined. This has revealed a number of significant properties that affect how evolution navigates genotype and phenotype space. For example, although genotype space is incredibly vast (a 30 nucleotide long molecule has 4^{30} or 10^{18} possible genotypes), the space of possible phenotypes is far smaller – many genotypes map to few phenotypes (Schuster et al., 1994; Cowperthwaite & Meyers, 2007). Akin to the synonymous mutation of the genetic code although at a much broader scale, this redundancy means that most phenotypes are connected through 'neutral networks' of genotypes

(Cowperthwaite & Meyers, 2007; Wagner, 2011). Sequence space can consequently be explored by a series of single-step mutations that do not affect RNA structure until a mutation becomes possible that results in a new structure of potentially higher fitness. In contrast to the predictions of gradualism, this process gives rise to long periods of phenotypic stasis punctuated by sudden evolutionary change – a pattern also observed for morphological traits in the fossil record and for cellular traits in laboratory evolution experiments (Gould & Eldredge, 1972; Elena et al., 1996).

Another property of the RNA secondary structure G-P map, again reminiscent of the genetic code example, is that different possible phenotypes are distributed asymmetrically across genotype space, with few common phenotypes that can be reached by many different genotypes and a greater number of rarer phenotypes. Indeed, just as the most common amino acids are over-represented within proteins, it has been shown that nature uses only the most common RNA secondary structures – those that are easiest to 'find' through blind search of the space of possibilities (Dingle et al., 2022; Cowperthwaite et al., 2008). Additionally, small movements in genotype space do not necessarily correspond to small changes in phenotype space; a single step mutation can result in a very different RNA structure, meaning 'saltational' events are not uncommon (Sumedha et al., 2007).

Most empirical systems are of course considerably more complex than the RNA secondary structure example discussed above, and the G-P map must be expanded to consider the entire physiology of the organism. Here, a comprehensive mapping from a space of possible genotypes to the space of possible phenotypes is obviously not feasible, even for the simplest of organisms. However, it can be feasible to experimentally examine the local neighbourhood of possible phenotypes available to an organism through mutation accumulation lines or mutagenesis (Besnard et al., 2020; McGuigan & Aw, 2017). It can also be possible to observe the alternative adaptive phenotypes available to simple organisms like bacteria through experiments that iteratively challenge the organism to adapt to a particular environment while continually precluding the discovered solutions via genetic engineering (Lind et al., 2017). Indeed, this is the approach I take in **Chapter 2 Disentangling Fitness from Likelihood in Adaptive Evolution.** Somewhat similar to the examples of accessibility of different phenotypes in the genetic code or RNA secondary structure examples, the concept of mutational 'target size' can be used in these contexts to understand the varying potential for different organismal changes to be generated by mutation (Lind et al., 2015; Besnard et al., 2020). For example, a phenotype that can be generated by breaking the function of a particular gene is associated with a large target size – many possible mutations are able to achieve this outcome. Such a phenotype would be more likely to be introduced early into a population and so gain a corresponding advantage over phenotypes with a smaller target size (Stoltzfus & Yampolsky, 2001). Taking a slightly different perspective, the number of genes underpinning a particular phenotypic trait will in general affect its tendency to vary, also influencing

patterns of evolutionary change (Besnard et al., 2020). Other biasing properties of G-P maps evident when considering the entire physiology of an organism include the correlated patterns of variation that exist between different phenotypic traits or 'pleiotropy' – the result of these traits sharing components of the G-P map such as the same gene or developmental pathway. When multiple different traits of an organism change in response to a single mutation, certain directions of phenotypic evolution can be facilitated and other directions constrained (Schluter, 1996; Uller et al., 2018).

Such articulating rules and structures and their inherent constraints and biases are encountered at all levels of organization throughout the G-P map, from the genetic code to proteins, through gene regulatory networks to the mechanics of morphogenesis (Maynard Smith et al., 1985; Uller et al., 2018; Alberch, 1991). At each level, by determining how genetic variation translates to phenotypic variation, the G-P map defines the range of possible adaptive phenotypes available to the organism and – along with biases in the mutational process – the likelihood of these being manifest by mutation.

1.2.3 G-P maps and mutational bias both shape and are shaped by evolution

What is particularly significant about the G-P map and mutational bias – and what also makes them so difficult to study – is that they are themselves the products of evolution. They therefore both shape and are shaped by evolution (Laland et al., 2015; Wagner & Altenberg, 1996; Hogeweg, 2012). More concretely: by articulating the phenotypic effect of mutations, the G-P map structures the space of possible phenotypes available to evolution, while the map itself is altered by each mutation. Similarly, the genomic replication and repair systems that influence the rates and biases by which the genome is altered (as well as those features that affect this process in a context-specific manner) are encoded within the genome that is being altered. As a result of this causal feedback, the space of possible phenotypes and their likelihood of being manifest by mutation are constantly changing throughout evolution as mutations accumulate.

This self-referential nature of evolution has deep implications. First, it imparts a history that the future evolution of a lineage is then constrained by – evolution is limited to exploring the possibilities adjacent to the historical trajectory of a lineage (Calcott, 2009; Maynard Smith, 1970; Maynard Smith et al., 1985). The result is that the course of evolution is often historically contingent: diverging lineages arrive at configurations of their G-P map (either through selection or neutral evolution) where certain evolutionary paths are closed, or new paths are opened (empirical examples of this process are reviewed in Blount et al., 2018). This may be something as simple as opening up an adjacent possible amino acid through a synonymous nucleotide substitution as in the aforementioned genetic code example, but in principle it can involve a change to the rules of development themselves. This

historical dependency has been studied at varying scales under different guises. For example, the concept of mutational epistasis (whereby the fitness effects of mutations interact in non-additive ways) means fitness effects become dependent on the order in which the mutations occur, and so constrains the paths natural selection will follow through genotype space (Weinreich et al., 2006; de Visser & Krug, 2014). At a much greater scale, the evolutionary history of species or higher-level clades leads to distinct patterns of morphological potential that is part of the study of 'developmental constraints' (Maynard Smith et al., 1985).

Most intriguingly, that the G-P map and mutational bias are themselves products of evolution allows the possibility for natural selection to shape them, such that the likelihood of mutation generating phenotypes that are adaptive is increased, or deleterious mutations decreased. For example, it has been argued that the genetic code itself has been optimized by natural selection to minimize errors arising both from mistranslation and mutation (Freeland & Hurst, 1998). Both the historically contingent nature of evolution and the possibility of natural selection increasing the likelihood of adaptive mutations will be explored in detail in **Chapter 3 The Evolution of Evolvability via Lineage Selection** of this thesis.

1.3 Technology, microbes, and experimental evolution

One of the reasons for ignoring the importance of the production of variation in the evolutionary process has been a technological one. Until very recently, ability to quantify variation in genotype and how it maps to phenotypic variation has been limited. However, we now find ourselves in an era where the precise content of the genome is readily accessible through whole genome sequencing, causative mutations of phenotypic change identifiable in a number of model systems, and where expanded techniques of inquiry allow the effects of mutations to be mechanistically linked to phenotype. Our ability to quantify variation as well as to recognise biases in how it is produced would have been beyond the imaginations of the founders of the Modern Synthesis. It is easy to forget how quickly technology has advanced: the sequencing of over 500 complete genomes at a resolution enabling the identification of a single mutation in each, as reported in **Chapter 3 The Evolution of Evolvability via Lineage Selection**, would have been unthinkable even a decade ago.

In particular, the combination of small genomes with modular architectures, and the ease with which genetic changes can be introduced, has made preliminary characterisation of physiological G-P maps in bacteria possible. Microbes also afford the possibility of watching evolutionary change unfold in real time through experimental evolution. We can see how evolution actually proceeds, rather than attempting to simulate it using models with questionable assumptions or trying to reconstruct an evolutionary history from extant variation. By tracking replicate populations founded by clones, we

can observe how new variants arise by mutation and quantify the repeatability of evolutionary trajectories. Experimental evolution, coupled with knowledge of bacterial G-P maps and their mutational processes, provides a unique basis for experimental investigation of how internal properties of organisms may bias the production of variation and so bias the outcome of evolution. Moreover, through genetic engineering it is possible to modify the bacterium's G-P map and mutational processes, allowing direct testing of their influence.

In this thesis, I describe evolution experiments using the bacterium *Pseudomonas fluorescens* SBW25 that demonstrate the importance of understanding the G-P map and mutational bias in shaping the course of adaptive evolution. Each experiment will be introduced in detail at the beginning of the relevant chapter. Briefly, in **Chapter 2 Disentangling Fitness from Likelihood in Adaptive Evolution** I examine why adaptive evolution arrives at particular phenotypes when other alternative phenotypes are available. Is this solely the result of fitness? Or does bias in the production of variation – specifically differences in the likelihood of the different phenotypes being manifest by mutation – also determine the path of adaptive evolution? In **Chapter 3 The Evolution of Evolvability via Lineage Selection**, I report the results of an experiment that implements a birth-death process at the level of lineages to explicitly select for those with a greater potential to generate adaptive phenotypes (evolvability). Thus, while chapter 2 focusses on how bias in the production of phenotypic variation may shape the course of adaptive evolution, chapter 3 extends to this to consider how bias in the production of variation can itself be shaped by adaptive evolution.

Chapter 2

Disentangling fitness from likelihood in adaptive evolution

"In evolution, selection may decide the winner of a given game but development non-randomly defines the players."

Pere Alberch (1980)

2.1 Introduction

The notion of the 'survival of the fittest' implies that fitness determines the phenotypes that arise in the course of adaptive evolution. Yet this notion does not capture the process by which the phenotypic variation on which natural selection acts is generated. Adaptive phenotypes are ultimately generated by mutation; therefore – assuming the phenotypes observed are not the only possibilities available – a causal understanding of adaptation requires knowing to what degree a phenotype was the 'fittest', and to what degree it was the most likely to be generated by mutation. Differing likelihoods are expected due to biases in the rates and kinds of mutations that occur, as well as how these mutations are then articulated into phenotypic change through the structure of an organism's genotype-phenotype map (G-P map) (Wagner & Altenberg, 1996; Maynard-Smith et al., 1985; Louis, 2016; Uller et al., 2018; Stoltzfus & Yampolsky, 2009). Together, mutational bias and the G-P map define the space of possible phenotypes available to an organism and the likelihood that a given phenotype will be manifest by mutation. Understanding to what extent likelihood contributes to the outcome of adaptive evolution alongside fitness is necessary to understand and predict the course of evolution.

2.1.1 The problem of observing counterfactual outcomes

The challenge is how to disentangle the roles of fitness and likelihood experimentally. To do so requires a study system that allows observation of the possible phenotypes that *could* have evolved but did not. The subset of phenotypes that evolution is observed to arrive at can then be compared to the set of all possible phenotypes that were available. Through this comparison, the rules by which evolution proceeded can then be identified – is the fittest of all possible phenotypes necessarily discovered by evolution? Or are fitter but less likely phenotypes available? Experiments observing such counterfactual scenarios – explicitly investigating the paths not taken by evolution – are not straight-forward. How can the space of possible adaptive phenotypes that could have evolved be examined? Additionally, it must be possible to measure or estimate both fitness and likelihood of the alternative phenotypes so that the relative role of each can be disentangled.

2.1.2 Approaches using microbial experimental evolution

Microbial evolution experiments are uniquely suited to observing the possible phenotypes that could have evolved but did not, allowing experimenters the opportunity to examine counterfactual scenarios in which the roles of fitness and likelihood can be teased apart and tested explicitly. By combining very large population sizes with very short generation times, microbial populations allow us to witness the origin and establishment of novel phenotypes in real time. A single clone can be

divided across multiple replicate experiments to discover the paths initially used by evolution to solve an evolutionary challenge. Most significantly for the study of counterfactual scenarios, the expansive genetic toolkit available to microbiologists, coupled with the modular nature of microbial genomes, allow adaptive solutions that are observed in a first round of evolution to be precluded via genetic engineering. When the experiments are repeated using these engineered strains, the initial paths are unavailable, forcing the use of alternative viable mutational paths that were otherwise invisible to natural selection (McDonald et al., 2009; Lind et al., 2015). Finally, highly sensitive growth and competition assays enable the direct comparison of the fitness of each of the discovered pathways.

2.1.3 The Wrinkly Spreader phenotype

The most successful application of a counterfactual approach to date used microbial evolution experiments to study alternative pathways to a common phenotypic solution: the Wrinkly Spreader (WS) phenotype of *Pseudomonas fluorescens* SBW25 (McDonald et al., 2009; Lind et al., 2015). These experiments demonstrated that the mutational paths commonly used by evolution were not the only ones available: many alternative paths of equivalent fitness could also generate the WS phenotype. However, these alternatives were only ever accessed when the common paths were rendered inaccessible. Since the different paths to reach WS were all of equivalent fitness, the cause of this pattern was able to be explained unequivocally in terms of the likelihood that a given path would be accessed by mutation.

2.1.4 Alternative paths to the Wrinkly Spreader phenotype

P. fluorescens strain SBW25 (henceforth 'SBW25') is an obligately aerobic gram-negative bacterium that, when placed in an unshaken broth-filled microcosm, rapidly proliferates and depletes the dissolved oxygen within the liquid. Cells then compete for position at the air-liquid interface (ALI) where there remains an influx of oxygen. Two distinct adaptive phenotypes reliably and rapidly evolve in response – one of these being WS, named for its distinctive morphology when plated in comparison to the ancestral smooth (SM) morphotype (Rainey & Travisano, 1998). WS colonizes the liquid surface through production of a cellulose exopolysaccharide that acts as a cell-to-cell glue. By binding cells together, and through attachment to the microcosm wall, a mat of cellulose enmeshed with bacteria is established at the ALI.

Generation of the WS phenotype requires mutational activation of a diguanylate cyclase (DGC) – an enzyme that catalyses the production of the secondary messenger molecule cyclic-di-GMP. This molecule then acts as an allosteric activator of the cellulose synthase machinery that produces and exports the cellulose polymer and leads to mat formation. Despite as many as 36 other genes

containing a DGC domain in the genome, the mutations causing WS always activate the same three DGCs (McDonald et al., 2009). These particular DGCs are located in signal transduction pathways subject to post-translational negative regulation, an architecture that imparts a large mutational target for loss-of-function (LoF) mutations, since all that is required to activate the DGC is to 'break' this negative regulation – an outcome readily achieved by a wide spectrum of mutations (McDonald et al., 2009).

Removing the three commonly used DGCs from SBW25 by genetic engineering and replaying evolution from this altered genotype revealed 13 alternative mutational paths involving different DGC-encoding loci that were able to generate WS of equivalent fitness to the common 'negative regulator' paths. That these alternative paths were not initially accessed was accounted for by their relatively smaller mutational target sizes: alternative paths required rarer gain-of-function (GoF) mutations or multiple LoF mutations. Evolution was shown to follow a probabilistic 'path of least resistance' whereby the least specific and fewest mutations to cause a particular phenotype were more likely to be manifest and therefore more likely to be amplified by selection over alternatives.

The frequencies by which the different paths were observed, led Lind et al. (2015) to formulate a probabilistic hierarchy of mutational paths that evolution will tend to proceed by in cases where a phenotype is caused by gene activation. This begins with the most likely paths: LoF to extragenic negative regulators, followed by LoF to intragenic negative regulators[2], then GoF: promoter activating mutations, gene fusion (through interstitial deletions enabling the 'capture' of adjacent promoters), and finally intragenic activating mutations. This order then reiterates but with each mutational class now requiring a double mutation (e.g., a mutation in two separate extragenic negative regulators), subsequently a triple, and so on.

2.1.5 Alternative phenotypes

The previous studies in SBW25 have demonstrated that when phenotypes (and fitness) are equivalent, likelihood can determine the path of genetic evolution. However, the question remains as to whether such features mediating likelihood can lead evolution to entirely different phenotypic (and so presumably different fitness) outcomes. Under what conditions might this occur? The importance of differences in likelihood to the evolution of alternative phenotypes can be examined experimentally in a similar manner to Lind et al. (2015) but instead of removing regulatory pathways to a common phenotype, the structural genes necessary for the phenotype itself can be removed. In fact, a

[2] An extragenic negative regulator is a negative regulatory protein encoded in a separate gene to the element it represses. An intragenic negative regulator refers to a domain of a protein that represses an element of itself.

subsequent study by Lind et al. (2017) did just that. Replaying evolution with the operon encoding the main structural component of WS removed led to the discovery of two additional phenotypes that colonised the ALI. However, although likelihood factors were identified as influencing the paths to these newly discovered phenotypes, both proved to be of lesser fitness than WS. Therefore, although the use of different genetic pathways to WS was explained by likelihood in Lind et al. (2015), the use of WS over alternative phenotypes could be explained in terms of fitness.

Here, I delve deeper into the uncharted phenotypic possibilities available to SBW25 and describe the discovery of novel phenotypes of equivalent or higher fitness than those previously observed. I then show how likelihood played a key role in determining their use by natural selection.

2.2 Results & Discussion

2.2.1 Discovery of additional phenotypes capable of colonizing the ALI

An iterative process was used to discover novel phenotypes capable of colonizing the ALI (Figure 2.1). Discovery of these phenotypes was contingent upon making the previously observed phenotypes inaccessible to natural selection.

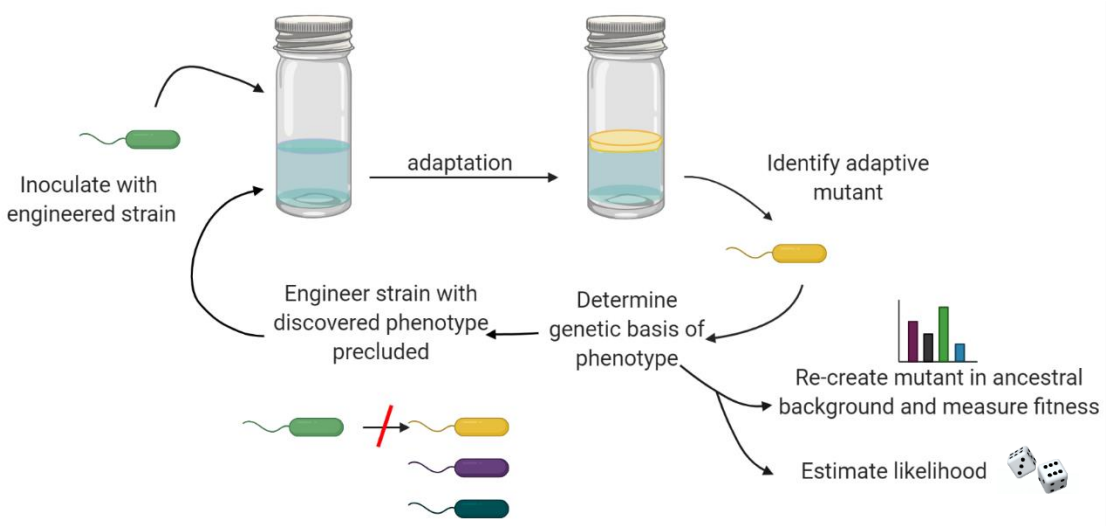

Figure 2.1 Experimental workflow for the discovery of counterfactual evolutionary routes to ALI colonization. Briefly, an engineered genotype unable to produce the previously discovered ALI-adapted phenotypes is inoculated into replicate microcosms. Adaptation to colonize the ALI proceeds and any newly evolved phenotypes are identified by directly sampling of the ALI with an inoculation loop. From here the genetic basis of the mutant is established through whole genome sequencing to identify the causal mutations, and (where possible) suppressor analysis via transposon mutagenesis to establish the genes necessary for the new phenotype. This information is then used to identify the necessary genetic manipulation to preclude the repeated evolution of the newly discovered phenotype. The re-engineered strain is then used to begin the cycle once again. Meanwhile, the causal mutations are reconstructed in the ancestral genotype, fitness measured, and likelihood estimated.

The first step of the discovery process was to preclude those previously discovered phenotypes. Beginning with a SBW25Δ*wss* strain in which the structural basis of the WS phenotype (the cellulose synthase *wss* operon) had previously been deleted by Lind et al. (2017), I first precluded the path to the phenotype known as 'PGA Wrinkly Spreader' (PGA-WS). PGA-WS was one of the two phenotypes discovered by Lind et al. (2017) and, similar to the original cellulose-based WS, was characterized by the over-production of an exopolysaccharide that acted as a cell-cell glue. Over-production of this exopolysaccharide, identified as Poly-β-1,6-*N*-acetyl-D-glucosamine (PGA), also gave the mutants a

distinctive wrinkled colony morphology and furthermore was shown to be caused by the exact same DGC activating mutations as the original cellulose-based WS, indicating an overlapping post-translational regulon for both exopolysaccharide synthases in SBW25. To preclude PGA-WS as a potential adaptive path I therefore deleted its structural basis – the PGA synthase machinery encoded in the *pgaABCD* (*pflu0143-6*) operon. This yielded the genotype SBW25Δ*wss*Δ*pga*, incapable of evolving WS or PGA-WS.

Two known ALI-colonising phenotypes remain accessible from the SBW25Δ*wss*Δ*pga* genotype: one being the Cell-chaining phenotype discovered by Lind et al. (2017) and the other being the Fuzzy Spreader phenotype discovered alongside WS in the original Rainey & Travisano (1998) experiment. Cell-chaining (CC) is caused by a single LoF mutation that disrupts NlpD (Pflu1301), a lipoprotein that localizes to the dividing septum and recruits the amidase AmiC required for complete cell division, thereby leading cells to become linked together in chains and enabling a weak ability to colonise the ALI. Fuzzy Spreader (FS), named for its distinct colony morphology, is caused by a single LoF mutation to a β-glycosyltransferase encoded by *fuzY* (*pflu0478*), leading to modification of the lipopolysaccharide O-antigen structure. The mutated lipopolysaccharide causes cell flocculation and increased adherence to the glass of the microcosm, a process thought to be mediated by an altered charge at the cell surface (Ferguson et al., 2013).

Since both remaining phenotypes (FS and CC) are caused by LoF mutations to structural genes, deleting the responsible gene will not preclude the phenotype but rather create it. Indeed, it is this apparent problem that likely led Lind et al. (2017) to leave intact the FS phenotype in their experiment. As a solution, the LoF targets were instead made polyploid through cloning the target locus and upstream non-coding region (presumably containing the cognate promoter) into a multi-copy plasmid (pSX). In cells carrying this plasmid, the causative LoF mutation becomes recessive, as mutation to the chromosomal copy is compensated by a functional *trans*-acting plasmid copy (Figure 2.2).

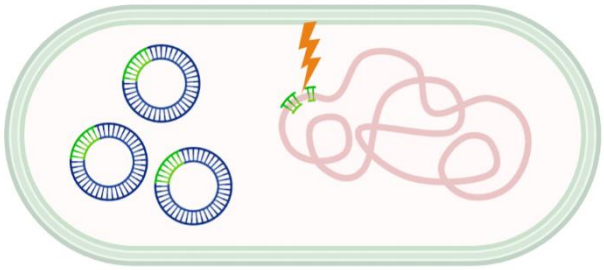

Figure 2.2: A mutational path requiring LoF mutation to a structural gene can be blocked by including additional functional copies of the target gene on plasmids. A gene on the chromosome (green ladder) is the target of a mutation (lightning bolt) that causes a LoF (broken ladder). However additional intact plasmid copies compensate for the LoF, meaning the cell phenotype remains unchanged.

This strategy worked effectively for the FS phenotype, yielding SBW25Δ*wss*Δ*pga* + pSX (*fuzVWXYZ*), a strain unable to evolve either WS, PGA-WS or the FS phenotype. However, the same approach was unable to be used for *nlpD*, as carrying pSX(*nlpD*) induced an abnormal cell-cell attachment phenotype (as observed via microscopy). Fortunately, the low fitness of CC means it remains at very low frequency and hence is unlikely to prevent the evolution of possible alternative phenotypes (Lind et al., 2017). In the following experiments the pathway to the CC phenotype therefore remains intact. To avoid continually referring to increasingly complicated genotypes, I will instead refer to the strains by their inability to generate particular phenotypes e.g., the "(WS, PGA-WS, FS)-deficient" strain. A table of the exact genotypes corresponding to each strain can be found in Chapter 4: Materials and Methods.

The subsequent section (**2.2.2.1**) describes the discovery of three additional ALI-colonizing phenotypes. Genetic details will be reported in this section only to the extent that they help establish the genetic and physical basis of the phenotype such that the phenotypes can be suitably named and the necessary manipulations to preclude their future evolution identified. Detailed analysis of the G-P map, mutational paths, and fitness of each phenotype will then be presented in subsequent sections.

2.2.2.1 The discovery of three additional ALI-colonizing phenotypes

To begin, twenty microcosms were inoculated with ~10^6 cells derived from independent colonies of the (WS, PGA-WS)-deficient strain. Note that FS was here left intact. Microcosms were observed over 14 days during which time, if confluent growth at the ALI was observed, a direct sample was taken with an inoculation loop and streaked to plates. The resulting colonies were then screened for their ability to reproduce the observed mat in a new microcosm. All 20 microcosms containing the (WS, PGA-WS)-deficient strain displayed the characteristic non-confluent ALI-colonization of the FS phenotype, and no novel phenotypes were detected.

In the next step, the same process was conducted for the (WS, PGA-WS, FS)-deficient strain. All microcosms showed negligible ALI colonisation as in the previous genotype with the exception of four microcosms. Examining these microcosms revealed two novel phenotypes.

2.2.2.2 The colanic acid-producing phenotype

Two of the microcosms containing mats appeared to be of the same phenotype; one producing a mat at day 7 and the other at day 11. Direct sampling of the mat with an inoculation loop showed it to be extremely 'gloopy' and led to colonies with a distinct morphology characteristic of cells expressing an exopolysaccharide.

The colony morphology of this new mutant was noted to be very unstable however, with re-streaking leading to many colonies having lost the distinct morphology and appearing as ancestral SM. The tendency of this strain to spontaneously revert to the ancestral SM phenotype meant a suppressor analysis to identify the genes necessary for the phenotype (by screening for loss of the colony phenotype in a population of cells with randomly inserted transposons) was not possible as this would be confounded by false positives. However, sequencing of the two strains revealed mutations implicating the *wcaJ-wzb* operon (*pflu3658–3678*), which is known to encode an exopolysaccharide that has previously been identified as a colanic acid-like polymer (CAP) (Gallie et al., 2015). Deletion of this operon disabled the mat-forming ability of the new phenotype, confirming its essential role as the structural locus. I will therefore refer to this mutant as CAP-Producing (CAPP). The genetic basis and fitness of this mutant is described in detail in section **2.2.4**.

2.2.2.3 The fimbria phenotype

The final two microcosms exhibiting mats appeared distinct from CAPP and were evident at day 11 and 13. Both formed a robust mat but showed no overt difference in colony morphology from the ancestral SM colony. The lack of distinct colony morphology suggested a mechanism of attachment other than constitutive production of an exopolysaccharide and also meant a suppressor analysis at the colony level was not possible. Instead, a suppressor analysis was conducted through screening suppression of the mats themselves when grown in 96-well plates. By this means a total of 3,840 transconjugant mutants were screened for loss of mat formation and a single suppression mutant identified. The transposon insertion locus responsible for suppressing mat formation identified the gene *pflu1610*, predicted to encode a fimbria 'usher' protein.

Fimbria (or pili as they are sometimes referred to) are extracellular appendages in gram-negative bacteria known to be involved in biofilm formation (Vallet et al., 2001). More specifically, a fimbria usher protein is associated with a particular mechanism of fimbria assembly known as the chaperone-usher pathway (Sauer et al., 2004). Chaperone-usher pathways are generally encoded within a single operon that consists of the usher, chaperone, and fimbria subunit proteins (Sauer et al., 2004; Busch & Waksman, 2012). In line with this, the Fim suppression locus and putative usher *pflu1610* in SBW25 has adjacent to it a putative chaperone protein (*pflu1611*) and four putative fimbria (*pflu1612*) or fimbria-like proteins (*pflu1609, pflu1608, pflu1607*). Deletion of the entire locus (Δ*pflu1607-1611*), disabled the mat-forming ability of the new phenotype, confirming its essential role as the structural locus. This phenotype is hereafter referred to as 'Fim' and is further described in section **2.2.5**.

2.2.2.4 The PSL-Wrinkly Spreader phenotype

To preclude future evolution of both newly discovered phenotypes, the CAPP structural locus (*pflu3658–3678)* and the Fim structural locus (*pflu1607-1611*) were deleted to yield a (WS, PWS, FS, CAPP, Fim)-deficient strain.

A total of 60 microcosms inoculated with this new strain resulted in no overt ALI growth being detected over the 14 days of observation, suggesting the SBW25 genome was now exhausted of ALI-colonizing phenotypes – or at least those accessible over the number of generations that can occur within the microcosm over this length of time.

In order to seek additional possible phenotypes requiring rarer mutational paths that were inaccessible over the current experimental conditions, the (WS, PWS, FS, CAPP, Fim)-deficient strain was subjected to a serial transfer experiment. Serial transfer extends the number of generations over which mutations and selection can occur and so removes the restrictions associated with growth in a single microcosm. Twenty replicate microcosms containing this strain were serially transferred at seven-day intervals, with 1% of the population propagated into a fresh microcosm at each transfer. At day 18 and day 20 (equivalent to approximately 20 extra generations of evolution) of this procedure, two microcosms displayed robust mats. The responsible mutants were identified by a subtle but unstable difference in colony morphology from the ancestral SM, meaning a suppressor analysis was again untenable. Sequencing revealed mutations targeting one of the DGCs that activates both WS and PGA-WS, indicating a third exopolysaccharide was being activated by the same cyclic-di-GMP regulon. A structural candidate locus was the predicted exopolysaccharide biosynthetic locus (*pflu2082-pflu2071*), homologous to the PSL-encoding operon (*psl*) in *Pseudomonas aeruginosa* PAO1 in which it encodes one of the primary structural components of *P. aeruginosa* biofilms (Jackson et al., 2004). Deletion of this locus disabled the phenotype, confirming its central role. Due to its genetic and phenotypic connection to WS and PGA-WS this new mutant was designated PSL-WS. The genetic basis and fitness of this mutant is described in section **2.2.4**.

The rarity of PSL-WS suggested the chances of finding additional adaptive ALI-colonising phenotypes beyond this would be difficult and the discovery portion of the experiment was discontinued at this point.

2.2.2.5 Summary of discovery

In sum, three new phenotypes had been discovered: CAPP, Fim and PSL-WS, each represented by two independently evolved mutants that were genome sequenced (Figure 2.3). This brings the total of distinct ALI-colonizing phenotypes available to SBW25 to seven (Figure 2.4)

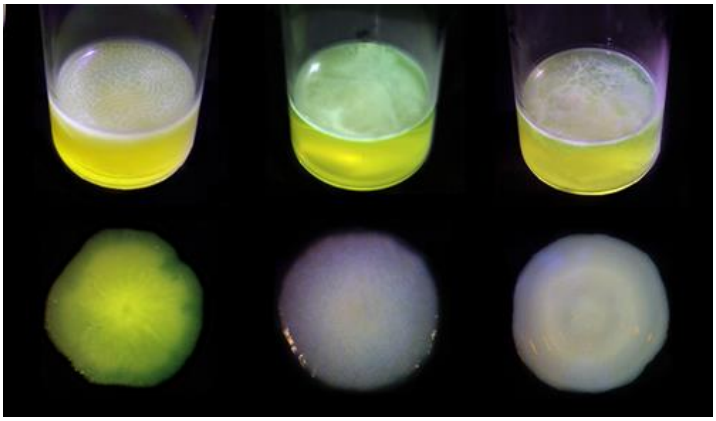

Figure 2.3 Colony and mat morphology of newly discovered phenotypes. From left to right: CAPP, Fim, PSL-WS.

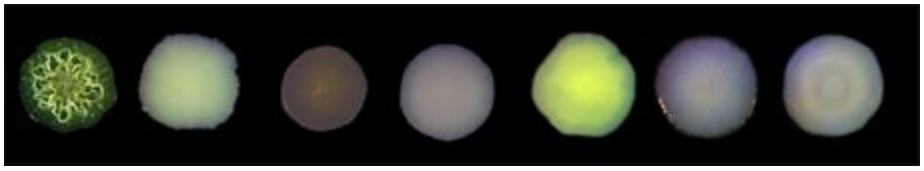

Figure 2.4 Colony morphology of all seven possible phenotypes. From left to right: WS, FS, PGA-WS, CC, CAPP, Fim, PSL-WS.

2.2.3 Reconstruction of mutational paths, fitness assays, and understanding the genetic basis of phenotypes to inform likelihood estimations

In the following sections (**2.2.4 – 2.2.6)** I reconstruct, within the ancestral genotype, the mutational paths required to reach each of the three newly discovered phenotypes. Each mutational step along these paths (all phenotypes were found to require more than one mutation) is then associated with a fitness and a likelihood of occurring.

Measuring both fitness and likelihood can be complicated. In this chapter, I have measured fitness using invasion assays, whereby an initially rare (1:100) mutant population is challenged to invade the ancestral SM type. This is preferable to 1:1 competition between mutants as it avoids the confounds of their physical interaction at the ALI. It should also be noted that fitness is not static within the microcosm environment due to frequency-dependent selection, whereby alternative strategies gain an advantage when rare (Rainey & Travisano, 1998). These dynamics result from the fact that established mats provide a substrate for growth of non-mat forming types (e.g., the ancestral SM), which do not pay the cost of mat-production but nonetheless gain the benefit of access to oxygen at the ALI. Mats can also collapse, providing opportunity for alternative phenotypes to invade. Therefore,

as fitness is expected to change over time, invasion fitness is measured over both 48 hours and 72 hours: the shorter-time frame to measure initial mat establishment, and the longer time-frame to measure mat robustness to invasion from the ancestor or other evolving mutants. Throughout this chapter, fitness will be expressed as the selection coefficient (see Chapter 4 Materials and Methods). Another point to note is that due to the unequal starting ratios in invasion assays, negative selection coefficients are not reliable as the invader is liable to be lost by drift. Therefore, negative selection coefficients resulting from invasion assays are not reported and the minimum selection coefficient reported is zero.

Establishing likelihood presents an even greater challenge than measuring fitness. It first requires estimating a mutational target size: the set of possible mutations able to generate a phenotype (or an intermediate step to a phenotype). For example, in the simplest possible case of a mutational path with an intragenic substitution, the first step to establish a relevant target size is to determine whether the effect of this substitution is equivalent to a clean deletion of the gene; if so, the gene is deemed a gene-wide LoF target and associated with a large target size. If not, then the mutation must be exerting a GoF or domain-specific LoF effect and a more refined analysis is required. This more refined analysis can then be achieved through the collection of additional mutants able to generate the phenotype or through the construction of site-specific mutations to test hypotheses informed by data from orthologs or existing studies in SBW25. Ultimately, estimating target size requires understanding how the causal mutations are altering the structure and function of the bacterium's G-P map. Mutational target size, together with knowledge of the rate of occurrence for each mutation able to 'hit' the target, then informs an estimation of likelihood. Differences in likelihood and fitness between phenotypes will be compared in later sections (**2.2.7** and **2.2.8**).

Note that it was essential that the causative mutations to each phenotype could be reconstructed in the ancestor so as to confirm that they did not rely on the genetically manipulated background from which the phenotypes evolved during the discovery process. This issue had already been encountered with the PGA-WS phenotype discovered by Lind et al. (2017), which depended on the cellulose-deficient Δ*wss* genotype due to an overlapping regulon with WS. If PGA-WS were to evolve from the wildtype SBW25 background it would therefore require a spontaneous mutation equivalent to the artificial deletion of *wss*. This fact makes subsequent comparison of mutational paths to different phenotypes awkward. Consequently, the 'ancestral' genotype in which all subsequently discovered phenotypes will be recreated is SBW25Δ*wss*.

2.2.4 'CAP-producing' phenotype: genetic basis, reconstruction, and fitness

Genome sequencing of the two independently evolved CAPP mutants ($CAPP_A$ & $CAPP_B$) revealed they had been generated through parallel mutational paths, with each mutant containing a large (270,352 & ~298,540 bp) duplication and a substitution in Pflu3677 (P662S & S549G) (table 2.1 and 2.2). Pflu3677 was encoded within the duplicated region and mutations at this locus were only apparent in ~50% of sequencing reads indicating an obvious temporal order in which the duplication necessarily preceded the substitutions in both instances. The two duplications overlapped across ~198,000 bp of the genome, indicating that they targeted a common locus that resided within this region. One immediately relevant candidate here was the 22.7kb *wcaJ-wzb* operon (*pflu3658–3678*) that was shown above (in section **2.2.2.2**) to encode the structural basis of this phenotype. The target of the point mutation, *pflu3677,* is the second gene in this operon. For simplicity I will refer to the structural locus of the CAPP phenotype, the *wcaJ-wzb* operon (*pflu3658–3678*), as the *cap* operon.

Mutant	**Duplication size**	Junction 1		Junction 2	
		Base	**Region**	**Base**	**Region**
$CAPP_A$	270,352 bp	3,905,260	Intergenic *pflu3535 / pflu3536*	4,175,568	*pflu3799*
$CAPP_B$	~298,540 bp	~3,804,538	Intergenic repeat *pflu3448 / pflu3450*	~4,103,078	Intergenic repeat *pflu3722 / pflu3723*

Table 2.1 Details of the duplications found in the CAPP mutants. Duplications can occur by several means, most commonly through RecA-mediated homologous recombination (Reams & Roth, 2015). The junctions of $CAPP_B$ are located at intergenic repeats sharing 700 bp of homology (92.4% identical sites) indicating it was RecA-mediated. The junctions of $CAPP_A$ show no junction homology and either occurred by some other mechanism, or the initial duplication has been modified by further recombination events such that the initial junctions can no longer be identified.

Locus	**Gene name**	**Mutation**	**Effect**
Pflu3677	*wzc*	$CAPP_A$: 1984 G→A	P662S
		$CAPP_B$: 1646 G→A	S549G

Table 2.2 Point mutations found in the two CAPP mutants

Due to their large size, re-creating either duplication in its entirety was not practical. However, it was reasoned that much of the duplicated sequence would be functionally gratuitous and that a targeted

duplication of the relevant region could be constructed, provided the proper function of the duplication could be inferred. One possibility, given that the duplications in both cases extended across the entire *cap* operon, was that they act to increase *cap* expression through a gene dosage effect. Such mutational events are well-known contributors to adaptive evolution in bacteria (Andersson & Hughes, 2009).

An additional possibility was suggested by knowledge of the regulatory architecture of *cap* transcription in SBW25. Substantial work has been conducted regarding the *cap* operon in SBW25 due to its involvement in a stochastic switching phenotype evolved in a previous experiment (Gallie et al., 2009; Gallie et al., 2019; Remigi et al., 2019). In deciphering its molecular basis, a three-gene operon (*pflu3655-3657*) directly adjacent to *cap* (*pflu3658–3678)* was found to encode transcriptional activators of both the *cap* operon and of itself. This configuration establishes a positive-feedback loop whereby the transcriptional activators increase transcription of themselves, and in turn drive increased transcription of the *cap* operon. The large duplications in both the CAPP mutants also encompass this transcriptional activator operon (*pflu3655-3657*). Therefore, as an alternative to the gene-dosage hypothesis, it was hypothesized that duplication of this regulatory region may be inducing the transcriptional positive feedback loop, consequently leading to increased *cap* transcription.

2.2.4.2 Duplication of *pflu3655-3657* is sufficient to generate the first-step to CAPP and recapitulates the effects of the larger evolved duplications

To test whether the duplication of the transcriptional activators alone was sufficient to drive *cap* expression, an artificial duplication of *pflu3655-57* and its upstream intergenic region (presumably containing its cognate promoter) was constructed in the ancestral genotype. This was achieved by cloning said region into a plasmid with transposase activity (pUC18R6K-mini-Tn*7*T-Gm::*pflu3655-57*) that integrated the cloned region as a single copy into the chromosome at a neutral attTn*7* site (Choi & Schweizer, 2006).

The resulting mutant carrying the artificial duplication of *pflu3655-57* exhibited a distinct colony morphology that appeared to be due to over-expression of the colanic acid-like polymer, suggesting the duplication had indeed triggered increased *cap* expression. Furthermore, this mutant exhibited increased fitness (Figure 2.5). Re-creating the known second-step Pflu3677^{P662S} mutation of CAPP$_A$ (established to be the second-step mutation due to its presence in 50% of sequencing reads) in the background of the artificial duplication led to a further increase in fitness, although this was only found to be significant in the 72-hour invasion assay (t_4=3.133, P=0.035, two-tailed t-test), indicating the second-step mutation reinforces the stability of the mat rather than contributing to initial

colonization. That the fitness effect of Pflu3677^{P662S} relied on the preceding duplication is expected given the scenario of the duplication initiating *cap* transcription – *pflu3677* (being the first gene of the *cap* operon) first needs to be transcribed before any mutation can alter the phenotype.

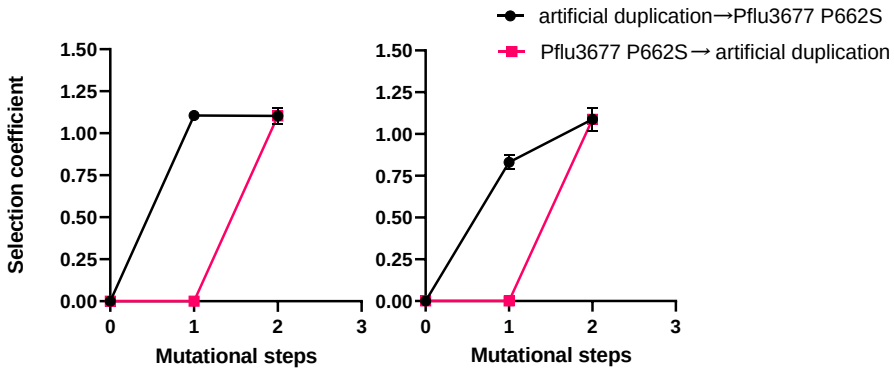

Figure 2.5 The two possible fitness trajectories to reach the CAPP phenotype. In order for fitness to increment at each step, the duplication of *pflu3655-57* must first proceed the substitution in Pflu3677. Selection coefficient calculated from invasion assay (1:100 competition between mutant and ancestral SM) as measured over 48-hrs (left) and 72-hours (right). The selection coefficient of the ancestral SM (i.e., the genotype at 0 mutational steps) was not explicitly measured and is defined as 0 by definition. Error bars represent SEM (n=6 or 3). Details of the invasion assay setup are provided in Chapter 4: Materials and Methods.

Importantly, it was also found that fitness of the reconstructed CAPP$_A$ mutant and the evolved CAPP$_A$ mutant was not statistically different (t_7=2.072, P=0.077, two-tailed t-test). Surprisingly, this suggests that the much larger evolved duplication does not suffer from collateral pleiotropic effects. Taken together, these results indicate that the artificial duplication accurately recapitulated the effect of the evolved duplication and presumably did so by allowing a threshold to be crossed that induced the transcriptional positive feedback loop, leading to increased transcription of the structural *cap* operon. As an aside, colony morphology of mutants with the artificial duplication was stable, suggesting the observed instability of the evolved CAPP colony morphology was due to a high rate of loss of the duplication.

2.2.4.4 Substitutions in Pflu3677 are not equivalent to gene-wide LoF

With the first-step mutation established as duplication of *pflu3655-57,* attention was turned to understanding the functional effect and relevant target size of the second-step mutation in Pflu3677.

Pflu3677 is a predicted tyrosine kinase[3] with studied orthologs existing in *Escherichia coli* and *Streptococcus pneumoniae,* respectively encoding Wzc and CpsD. Wzc and CpsD are tyrosine kinases known to regulate the rate of production and length of capsular polysaccharides (Obadia et al., 2007; Nourikyan et al., 2015; Olivares-Illana et al., 2008) and by this means may alter the physicochemical properties of capsules (Roberts, 1996). They are characterized by an N-terminal transmembrane loop followed by a cytosolic region containing the kinase domain and a C-terminal cluster of tyrosine residues (Grangeasse et al., 2007).

Figure 2.6 Domain architecture of *pflu3677* and locations of the substitutions occurring in the two independent CAPP mutants. Connected grey boxes indicates transmembrane helices. The blue box indicates the conserved cluster of tyrosine residues typical of a tyrosine kinase.

The substitutions observed in both CAPP mutants fall within the cytosolic region and occur near or within the kinase domain, suggesting a somewhat specific mutational target (Figure 2.6). To test whether these mutations were having an effect equivalent to gene-wide LoF, a clean deletion of *pflu3677* was created in the previously reconstructed $CAPP_A$ mutant ($Pflu3677^{P662S}$ + the artificial duplication). If the deletion did not alter the phenotype in any way, then the $Pflu3677^{P662S}$ mutation could be deemed equivalent to a gene-wide LoF. The resulting deletion mutant however exhibited a substantially altered mat and colony phenotype, indicating that the P662S substitution had a GoF effect or domain-specific LoF effect. The fact the mutation exhibits its effect in the presence of the wildtype allele (i.e., it is dominant in a heterozygotic state) led to the conclusion that it is most probably a GoF mutation.

Insight into the physiological and phenotypic effect of the substitutions in Pflu3677 came from microscopy and counter-staining with India ink, which revealed that following the artificial duplication of *pflu3655-57*, all cells appear bound in an impermeable sheath of colanic acid (they are 'capsulated') but if combined with the second-step $Pflu3677^{P662S}$ mutation, cells became non-capsulated (Figure

[3] Tyrosine kinases are one of several kinases existing in bacteria. Kinases catalyse phosphorylation – a reversible protein modification in which a phosphoryl group donated from ATP is covalently bonded to a particular residue, with the energy released during ATP hydrolysis exerting a conformational change of the protein. Together with the reverse reaction (dephosphorylation) catalysed by phosphatases, bacteria use this mechanism as a means of signal transduction and modulation of protein activity.

2.7). Since there was no sign that this was the result of a down-regulation of colanic acid synthesis (at least with respect to colony morphology, double mutants appeared to be producing more rather than less exopolysaccharide), the loss of capsulation suggested the Pflu3677^{P662S} mutation may be altering the properties of the colanic acid polymer in some way, in line with the known function of the Pflu3677 orthologs CpsD and Wzc.

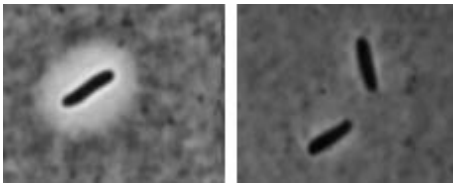

Figure 2.7 India Ink stain of CAPP single and double mutants. The tightly bound exopolysaccharide of the capsule is impermeable to India ink and results in a clear 'halo' around the cell. The first-step mutant artificial duplication of *pflu3655-57* exhibits capsulated cells (left) while the two-step mutant (duplication + P662S) does not (right).

2.2.4.5 CAPP summary

The first step to CAPP requires a duplication of the *cap* transcription factors encoded in *pflu3655-3657*, which is hypothesized to trigger a positive feedback loop of its own expression and increased *cap* transcription. The precise effect of the second step mutation is undetermined but its fitness effect relies on the preceding duplication and was determined to be a GoF mutation to a regulatory tyrosine kinase encoded in second gene of the *cap* operon. As the CAPP mutations successfully generated the CAPP phenotype when constructed in the ancestral genotype, the CAPP phenotype does not rely on the engineered background and could have arisen at an earlier stage (i.e., in the Lind et al. 2017 experiment) by these two mutations alone.

2.2.5. 'Fimbria' phenotype: genetic basis, reconstruction, and fitness

Genome sequencing of the two discovered Fim mutants (Fim_A & Fim_B) also revealed highly parallel mutational paths to this phenotype. Each mutant contained two non-synonymous substitutions: one in Pflu1605 (S558G or A466V) and an identical mutation in Pflu1609 (D94A) (Table 2.3). The Fim_A mutations were reconstructed in the ancestral background and successfully generated the phenotype, indicating Fim did not rely on the engineered background and could (as with the CAPP mutant) have arisen at an earlier stage (i.e., in the Lind et al. 2017 experiment) by these two mutations alone.

Locus	Gene name	Mutation	Effect
pflu1605	*rcsC*	Fim_A: 1672 A→G	S558G
		Fim_B: 1397 C→T	A466V
pflu1609	-	Both: 281 T→G	D94A

Table 2.3 Mutations detected in the two Fim mutants

2.2.5.1 Fitness trajectories to Fim

Pflu1605^{S558G} caused a large increase in fitness (~100%) on its own while Pflu1609^{D94A} had no effect, indicating the mutation to Pflu1605 was necessarily the first step to generate the Fim phenotype (Figure 2.8). The second-step Pflu1609^{D94A} mutation however contributed significant increases in fitness at both 48-hr (t_{10}=4.54, P=0.001, two-tailed t-test) and 72 hrs (t_7=2.93, P= 0.022, two-tailed t-test) when constructed in the background of the Pflu1605 mutation.

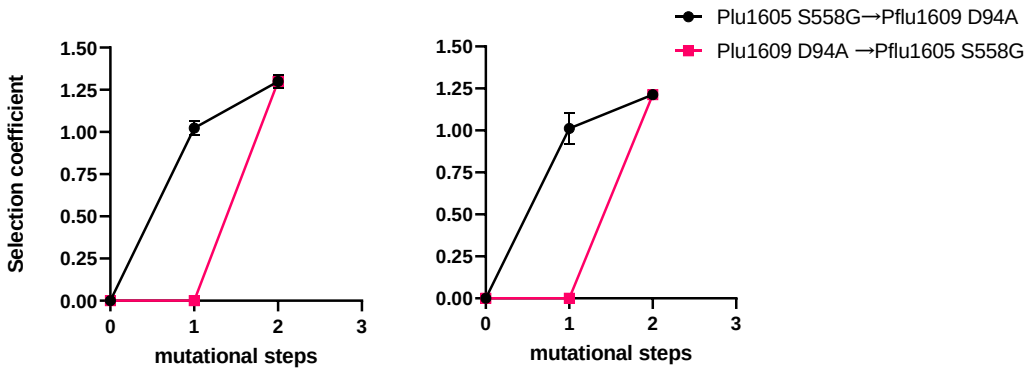

Figure 2.8 Two possible fitness trajectories to Fim_A. In order to increment fitness at each mutational step, the Pflu1605 S558G mutation had to proceed the Pflu1609 D94A mutation. Selection coefficient calculated from invasion assay (1:100 competition between mutant and ancestral SM) as measured over 48-hrs (left) and 72-hours (right). The selection coefficient of the ancestral SM (i.e., the genotype at 0 mutational steps) was not explicitly measured and is defined as 0 by definition. Error bars represent SEM (n=6 or 3).

2.2.5.2. The first-step Pflu1605 mutation is predicted to increase transcription of the adjacent fimbria structural genes

Pflu1605 resides in a four gene cluster (*pflu1603-1606*) homologous to the PvrSR/RcsCB regulatory system in *Pseudomonas aeruginosa* PA14, which has been extensively studied (Mikkelsen et al., 2009; Mikkelsen et al., 2013; Nicastro et al., 2009). PvrSR/RcsCB controls transcription of an adjacent gene cluster encoding a fimbria chaperone-usher pathway, known as *cupD1-4* (Mikkelsen et al., 2013). Similarly, directly adjacent to the PvrSR/RcsCB orthologs (*pflu1603-1606*) in SBW25 is the cluster of genes that form the structural basis of the Fim phenotype (*pflu1612-1607*) and also encode a

chaperone-usher pathway (Figure 2.9). This structural gene cluster is where the target of the second-step mutation for the Fim phenotype, *pflu1609*, resides.

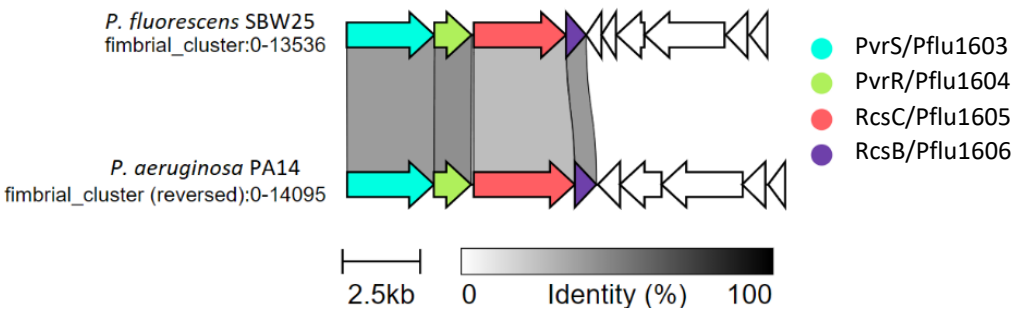

Figure 2.9 Gene cluster comparison of the *pvrSR/rcsCB* and adjacent structural gene region of *P. aeruginosa* PA14 and *pflu1603-1606*. Identity (%) here refers to identical amino acids in the aligned protein sequences. Fimbrial structural genes are shown in white.

Despite the overall homology and synteny of the regulatory regions, the structural regions consist of essentially unrelated genes required for fimbria construction via a chaperone-usher pathway. Such exchangeability of the regulatory PvrSR/RcsCB system with distinct adjacent structural gene clusters is common among *Pseudomonas* (Nicastro et al., 2009). In the case of SBW25, Pflu1603-Pflu1606 is by analogy expected to control the adjacent fimbria gene cluster, and mutations in Pflu1605 are presumably activating transcription of these fimbria genes. This notion is supported by the fitness effects of the two Fim-causing mutations: Pflu1609^{D94A} (encoded in the structural cluster) has no fitness effect in isolation and relies on Pflu1605^{S558G} (encoded in the regulatory cluster) to manifest its effect. Similar to the situation seen with CAPP, this pattern of epistasis supports a scenario whereby increased transcription of the fimbria structural genes is required as the first step mutation before the effect of the Pflu1609^{D94A} can be exerted.

2.2.5.3 Mutation to Pflu1605 is not equivalent to gene-wide LoF

Nicastro et al. (2009) reported a 20-fold increase in adjacent fimbria gene (*cupD1-4*) transcription in a *rcsC* (equivalent to *pflu1605* in SBW25) deletion mutant, indicating its role as an extragenic negative regulator of *cupD* transcription in *P. aeruginosa* PA14. This raised the possibility that the mutations in Pflu1605 lead to increased transcription of the adjacent fimbria structural genes via a similar disruption of negative regulation. However, contrary to this expectation, I found that deleting *pflu1605* from the Fim$_A$ mutant instead abolished the Fim phenotype, indicating the gene was playing an essential role. To better understand the possible effects and relevant target size of the Pflu1605

mutations it is necessary to examine in more detail the structure and function of its ortholog RcsC and the role of this protein in the PvrSR/RcsCB regulatory system.

2.2.5.4 Insights from PvrSR/RcsCB function

PvrSR/RcsCB represents an unusual configuration of two-component regulatory systems (TCRS) functioning in tandem (Nicastro et al., 2009). A TCRS is a basic stimulus-response module in bacteria and in the usual case of a TCRS functioning in isolation, consists of a membrane-bound sensor kinase that detects an environmental signal and a response regulator encoded in a separate gene that receives this signal and mediates the output. Signalling is achieved by a series of phosphate transfer events beginning with detection of the environmental signal, which causes the sensor kinase to auto-phosphorylate – transferring a phosphate from ATP to its conserved histidine residue. The phosphate is then transmitted to a conserved aspartate on the N-terminal receiver domain (Rec) of the response regulator and in turn activates its output domain, which is often a transcription factor. Sensor kinases in TCRSs may also often exhibit phosphatase activity that acts in opposition to the kinase by de-phosphorylating the response regulator (West & Stock, 2001).

In the case of PvrSR/RcsCB, the two sensory components from each TCRS have been integrated to modulate the output of a single response regulator, RcsB (equivalent to Pflu1606), whose output domain is the transcription factor that activates the adjacent *cupD* gene cluster. PvrS (Pflu1603) acts as the sensor kinase of this system, while RcsC (Pflu1605) acts in opposition as the phosphatase. The balance of the opposing activities of the PvrS (Pflu1603) kinase and the RcsC (Pflu1605) phosphatase, hence mediate the level of *cupD* transcription (Mikkelsen et al., 2013). Crucially, although RcsC (Pflu1605) functions as the phosphatase, this protein is also essential for transmitting the signal from the PvrS (Pflu1603) sensor kinase to the response regulator and so is essential for activation of *cupD* transcription. This additional His-Asp phosphotransfer event within RcsC (Pflu1605) is mediated by its internal Rec receiver domain and histidine phosphotransfer domain (Hpt), a configuration known as a phosphorelay (Hoch, 2000).

2.2.5.5. RcsC/Pflu1605 is involved in both positive and negative regulation of the fimbria structural gene cluster

RcsC is therefore both an integral part of the positive regulatory pathway for *cupD* transcription (through its phosphorelay activity) as well as the negative regulatory pathway of *cupD* transcription (through its phosphatase activity). By analogy to this system, Pflu1605 mutations are therefore expected to be affecting the protein's phosphatase activity, which negatively regulates the response

regulator Pflu1606 – the transcriptional activator of the adjacent fimbria structural genes.[4] The locations of both first step Fim_A and Fim_B mutation are compatible with this hypothesis (Figure 2.10).

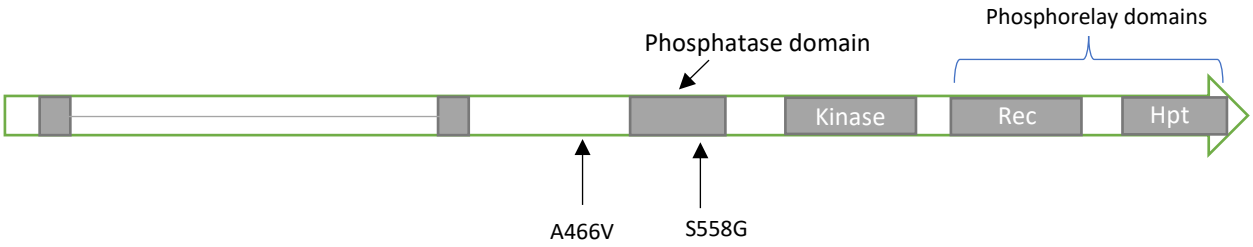

Figure 2.10. Domain architecture of *pflu1605* and locations of the substitutions occurring in independent Fim mutants. Connected grey boxes indicate transmembrane helices. Rec = Receiver domain, Hpt = Histidine phosphotransfer domain. The kinase domain is predicted to be non-functional as in the ortholog RcsC; kinase activity having instead been outsourced to a separate protein, PvrS (equivalent to Pflu1603), as part of the unusual integration of tandem TCRSs (Mikkelsen et al., 2013).

While LoF to the phosphatase domain activates fimbria transcription, phosphorelay activity (encoded downstream in the Rec and Hpt domains) needs to be retained, explaining why deletion of *pflu1605* caused loss of the Fim phenotype. Such 'intragenic pleiotropy' would severely constrain the mutational target size of the first step to Fim by permitting only those mutations that can disable phosphatase activity but do not interfere with the required downstream phosphorelay function and otherwise maintain the overall integrity of the protein.

2.2.5.6. Collecting additional Fim-causing mutations in Pflu1605

To test whether Pflu1605 did indeed represent such a constrained LoF mutational target, an experiment was conducted to collect a range of alternative *pflu1605* mutations capable of generating Fim. To this end, 20 microcosms were inoculated with a (PGA-WS, FS)-deficient strain already containing the neutral second-step mutation $Pflu1609^{D94A}$ and grown for 5 days. Since the second-step mutation to Fim already existed in this strain, only a single mutation (expected to occur in Pflu1605) was required to elicit the full Fim phenotype. By this method, a total of 18 new Fim mutants were isolated and sequenced.

[4] As in its RcsB ortholog, Pflu1606 contains an N-terminal Rec domain and an output domain with a helix-turn-helix (HTH) motif signifying its ability to bind DNA

2.2.5.7. The first-step mutation to Fim requires LoF to a highly constrained extragenic negative regulator

The locations of the additional Pflu1605 mutations are mapped to the gene structure in Figure 2.11. Including those already identified in Fim_A (S558G) & Fim_B (A466V), which were also found among the 18 additional mutations, a total of 11 unique missense mutations were found to generate Fim.

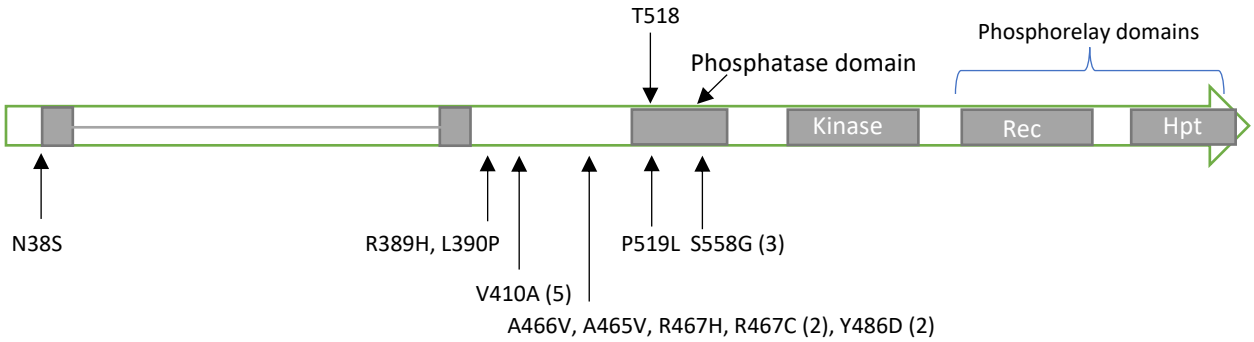

Figure 2.11 Location of 18 additional mutations in Pflu1605. Connected grey boxes indicates transmembrane helices. Rec = Receiver domain, Hpt = Histidine phosphotransfer domain. T518 annotated above the gene indicates the location of a conserved threonine residue necessary for phosphatase activity in *P. aeruginosa* PA14 (Mikkelsen et al., 2013).

With the exception of a single outlier mutation (N38S) found at the N-terminus, all mutations accord to expectations of LoF to the phosphatase activity of Pflu1605. Of particular relevance is P519L: directly adjacent to its equivalent residue in *P. aeruginosa* PA14 is a conserved threonine required for phosphatase activity (annotated in Figure 2.9). Mikkelsen et al. (2013) replaced the equivalent residue in RcsC (T506) with an alanine ($RcsC^{T506A}$), resulting in a substantial increase in fimbria transcription. Taken together I conclude that the first step to Fim requires a mutation to a highly constrained LoF target in a negative regulatory domain.

2.2.5.8 The second-step mutation to Fim requires a GoF mutation to a structural gene

As to the second-step mutation to generate Fim, recall that both of the evolved Fim mutants contained the same D94A substitution in Pflu1609, and that this protein was encoded among the structural genes (*pflu1612-1607*) underpinning the Fim phenotype. That both Fim mutants contained the same D94A mutation suggested a narrow target size or possibly the influence of a mutational hotspot.

Little is known about Pflu1609 and no informative orthologs exist. A single annotated adhesin protein superfamily domain predicted from residues 175 to 301 suggests it functions as an adhesin – a small structural subunit attached to the ends of fimbria that specifies attachment to particular surfaces (Busch & Waksman, 2012). Adhesins have been reported as a target of adaptive mutations in

pathogenic bacteria (Kisiela et al., 2012). The only other predicted feature is a signal peptide at the N-terminus, indicating it is destined toward the SecYEG secretory pathway to be translocated across the inner membrane – in line with the predicted function of an adhesin or fimbria structural subunit (Busch & Waksman, 2012). To test whether the Pflu1609^{D94A} mutation had an effect equivalent to gene-wide LoF, *pflu1609* was deleted from the reconstructed Fim$_A$ mutant background. This resulted in loss of the Fim phenotype, indicating the gene was essential for maintenance of Fim and that the D94A mutation must have a GoF or domain-specific LoF effect. Given its likely role as a structural subunit of fimbria, a domain-specific LoF effect seems unlikely, and I tentatively conclude that the second-step mutation to Fim represents a GoF mutation to a structural gene.

2.2.5.9 Fim Summary

The first step to Fim requires a constrained LoF mutation in Pflu1605 that is predicted to activate transcription of the adjacent fimbria structural genes. The mutational target size of this first mutation is confined to those mutations that disrupt the phosphatase domain of Pflu1605 without interfering with downstream domains involved in phosphorelay, which are necessary for activation of fimbria transcription. The second step requires a GoF mutation to a possible fimbria adhesin (Pflu1609). As these mutations successfully generated the Fim phenotype when constructed in the ancestral genotype, they did not rely on the engineered background and could have arisen at an earlier stage (i.e., in the Lind et al. 2017 experiment) by these two mutations alone.

2.2.6. 'PSL-Wrinkly Spreader' phenotype: genetic basis, reconstruction, and fitness

Genome sequencing of the two PSL-WS (PSL-WS$_A$ & PSL-WS$_B$) mutants again revealed highly parallel mutational paths. Each mutant contained three mutations: an identical Q189H substitution in NlpD, an intergenic point mutation between *nlpD* and *rpoS*, and a mutation of Aws – one of the common DGC pathways known to generate WS and PGA-WS (Table 2.4).

Locus	Gene name	Mutation	Effect
pflu1301 / *pflu1302*	*nlpD* / *rpoS*	PSL-WS$_A$: +36/-71 C→T	-
		PSL-WS$_B$: +87/-20 C→T	-
pflu1301	*nlpD*	Both: 537 C→T	Q189H
pflu5211	*awsX*	PSL-WS$_A$: ΔT229-G261	ΔY77-Q87
pflu5210	*awsR*	PSL-WS$_B$: 196 C→T	F66L

Table 2.4 Mutations found in the PSL-WS mutants

The effect of the Aws mutations is well known – the exact same mutations have previously been found to generate both cellulose WS and PGA-WS phenotypes (McDonald et al., 2009; Lind et al., 2017). This occurs via an increase in cyclic-di-GMP resulting from LoF of the negative regulation imposed by AwsX on AwsR (the DGC in the Aws pathway). Increased cyclic-di-GMP then activates exopolysaccharide synthases – cellulose synthase in the case of wildtype SBW25 and PGA synthase when the cellulose synthase *wss* operon is deleted. It can now be concluded that a third synthase, that producing PSL, is also activated by this same cyclic-di-GMP regulon.

What remains to be understood is the role of the two additional mutations in the PSL-WS mutants and whether the overlapping cyclic-di-GMP regulons of the PGA and PSL synthases will interfere with reconstruction of the PSL-WS phenotype.

2.2.6.2. Two PSL-WS mutations are predicted to target expression of the sigma factor RpoS

Insight into the role of the two other PSL-WS mutations (the Q189H substitution and the intergenic *nlpD/rpoS* point mutation) stemmed from an unusual connection with one of the previously discovered mat-forming phenotypes – the CC (Cell-chaining) phenotype of Lind et al. (2017). The connection relates to the Q189 residue of NlpD. Cell-chaining mutants are consistently found to contain a mutation altering this residue to a premature stop codon (Q189*), causing LoF to the NlpD protein. This specific mutation is found despite many alternative LoF mutations to *nlpD* being able to generate CC of equivalent fitness. The enrichment for mutations at this particular Q189 residue is the result of a potent mutational hotspot hypothesized to be associated with a promoter for the downstream *rpoS* gene that happens to be embedded within *nlpD* at this site (Farr, 2015; Lind et al., 2017). It is thought the initiation of transcription of *rpoS* at this promoter is elevating the local mutation rate (Farr, 2015).

In the case of the $NlpD^{Q189H}$ mutation that occurs in the PSL-WS mutants, rather than targeting NlpD function, this appears to be targeting the embedded *rpoS* promoter itself, with the Q189H substitution being incidental and neutral in regard to NlpD function. Evidence for this comes from a variety of sources. Firstly, reconstructing the Q189H mutation in the ancestral genotype led to normally dividing, un-chained cells, indicating it does not cause LoF to NlpD as in the CC phenotype. Secondly, a study in *P. aeruginosa* PA01 has demonstrated that RpoS transcriptionally activates the PSL synthase-encoding operon, suggesting a similar connection might exist in SBW25 (Irie et al., 2010).[5] Finally, given its

[5] RpoS is a sigma factor: an inter-changeable subunit of RNA polymerase that binds to the core polymerase structure and alters its propensity to attach and initiate transcription at specific promoter sequences (Potvin et al., 2008). Sigma factors are hence global regulators of transcription.

location, the remaining intergenic *nlpD*/*rpoS* mutation in the PSL-WS mutants can conceivably do little else other than also affect *rpoS* expression.

Taken together, both the intergenic mutation upstream of *rpoS* and the mutation of the predicted *rpoS* promoter embedded in *nlpD* would appear to be acting to increase expression of *rpoS* and thereby increasing transcription of the PSL synthase-encoding operon, the structural basis of the PSL-WS mutant. Whether the mutational hotspot identified by Farr (2015) has increased the likelihood of the $NlpD^{Q189H}$ mutation in the present experiment was not examined, but there is no reason to doubt that it would.

2.2.6.1 PSL-WS cannot be re-created in the ancestral background due to an overlapping regulon with PGA

Recall that the previously discovered PGA-WS phenotype was unable to be recreated in the SBW25 wildtype background and relied on the cellulose operon (*wss*) having been deleted (Lind et al., 2017). This is because both cellulose and PGA synthases are activated by the same intracellular signal (increased levels of cyclic-di-GMP). Therefore, recreating the causal mutations that generated the PGA-WS mutants in a wildtype background simply generates a cellulose-based WS.

The situation was found to be somewhat similar for PSL-WS – reconstruction of all three PSL-WS_A mutations in a background with the *pga* operon intact (Δ*wss*) produced a mutant exhibiting a different phenotype than the evolved PSL-WS mutant. Specifically, mutations reconstructed in a (*pga* intact) Δ*wss* background led to hyper-wrinkly colonies and thicker mats that were visibly distinct from either the PSL-WS or PGA-WS phenotypes, suggesting both PSL and PGA synthases were contributing to the hyper-wrinkled phenotype. Although this hyper-wrinkly hybrid is a legitimate phenotypic solution to ALI colonisation and could conceivably have evolved within a microcosm (it may even be of substantial fitness), it will be excluded from further analysis. For the fitness measurements in the following section, the PSL-WS mutations were instead reconstructed in a Δ*wss*Δ*pga* background so as to understand the evolutionary process that occurred *in situ*.

2.2.6.6 Fitness trajectories to PSL-WS

The three mutations required to reach PSL-WS means there are 3! = 6 possible trajectories. Four of the six possible trajectories to the PSL-WS_A mutant were reconstructed (the final two to be created at a later date) and their invasion fitness measured over 48 and 72 hours (Figure 2.12).

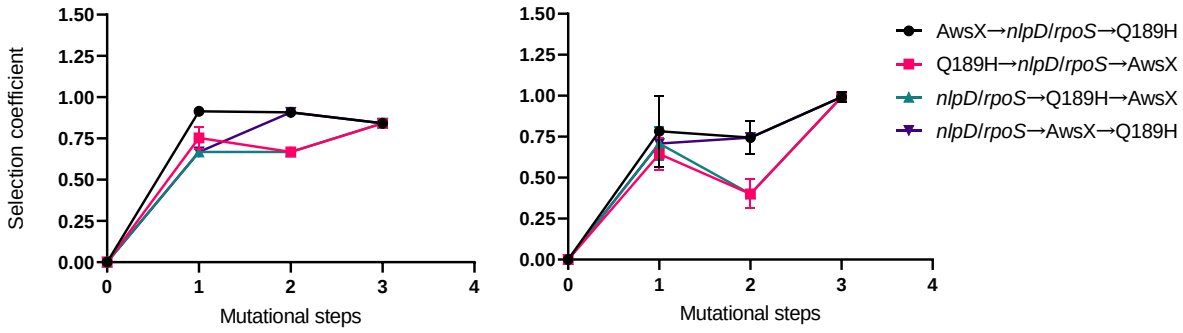

Figure 2.12 Four of the six possible fitness trajectories to reach PSL-WS from a SBW25Δ*wss*Δ*pga* background. Selection coefficient calculated from invasion assay (1:100 competition between mutant and ancestral SM) as measured over 48-hrs (left) and 72-hours (right). The selection coefficient of the ancestral SM (i.e., the genotype at 0 mutational steps) was not explicitly measured and is defined as 0 by definition. Error bars represent SEM (n=3).

No single trajectory as measured over 48 hours or 72 hours increments fitness at each step. That is, all trajectories involve either a neutral or deleterious mutation and selection is therefore expected to favour none of the trajectories. However, since the fitness landscape deforms over time, if both 48 *and* 72-hour fitness measures are considered, a single path where each step results in higher fitness does open up: the intergenic *nlpD/rpoS* point mutation (+36/-71 C→T) → AwsX$^{\Delta Y77\text{-}Q87}$ → NlpD^{Q189H}. Here the first two steps are favoured at 48 hours and the final step favoured at 72 hours. The final two fitness trajectories are required before any firm conclusions can be made regarding the possible mutational path to PSL-WS.

2.1.6.7 PSL-WS Summary

The PSL-WS phenotype required three mutations, although the order in which these might have occurred is not yet clear. One of these mutations involved LoF to the extragenic negative regulator of a DGC that led to increased cyclic-di-GMP levels within the cell and activation of the PSL synthase machinery. Two additional mutations are hypothesized to be affecting expression of the sigma factor RpoS, which was shown to transcriptionally activate the operon encoding PSL-synthase in a different *Pseudomonas* species (Irie et al., 2010). The mutations causing PSL-WS could not be reconstructed in the ancestral genotype due to interference from an overlapping cyclic-di-GMP regulon with the previously discovered PGA-WS phenotype. For the PSL-WS phenotype to be generated from the ancestral genotype, it would therefore require a LoF mutation equivalent to deletion of the PGA synthase-encoding operon.

2.2.7 Comparing fitness

Including those discovered in previous studies, there now exists a collection of six distinct adaptive phenotypes able to evolve from the SBW25Δ*wss* ancestor and colonise the ALI. Moreover, extensive information regarding the mutational paths required to reach each of these phenotypes has been gathered, providing insight into their relative likelihood. It is now possible to ask about the order with which evolution arrived at these different phenotypes and in particular why those phenotypes discovered here (CAPP, Fim, PSL-WS) were not observed to evolve in earlier experiments. To begin, we can compare the fitness of the newly discovered phenotypes compared to those previously discovered (Figure 2.13).

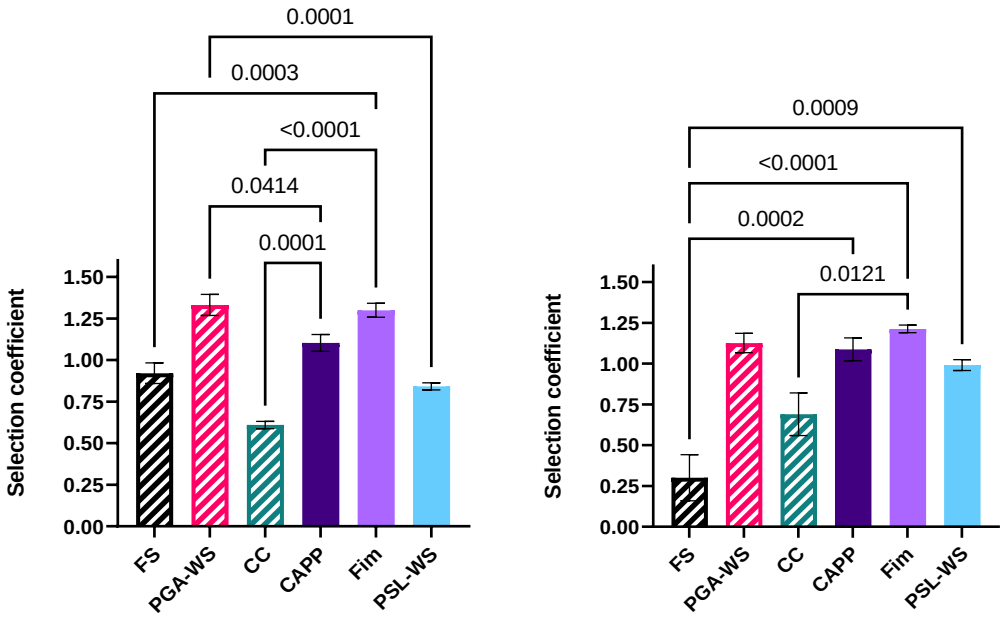

Figure 2.13 Overall fitness measures of alternative ALI-colonizing phenotypes. 48 (left) and 72 hour (right) invasion assays against SBW25Δwss. Only those pairwise comparisons between newly discovered and previously discovered phenotypes (striped bars) that are significant (P<0.05) according to Tukey's multiple comparison test are displayed (ANOVA at 48 hrs $F_{5,24}$=21.46, P<0.0001; ANOVA at 72 hrs $F_{5,17}$=14.77, P<0.0001). The following representative genotypes were used for the previously discovered phenotypes: FuzY^{T443G} for FS; AwsX$^{\Delta Y77\text{-}Q87}$ for PGA-WS and Nlpd^{Q189*} for CC (Ferguson et al., 2013; McDonald et al., 2009; Lind et al., 2017). Error bars represent SEM (n= 6 or 3).

The most important fitness measure to consider here is the shorter 48-hr timeframe (the left bar plot in Figure 2.13) that reflects initial colonisation ability – subsequent fitness is only relevant if a mutant has first managed to establish itself in the population. Here, all three newly discovered phenotypes were found to be significantly fitter than CC and either equally fit (PSL-WS and CAPP) or fitter (Fim)

than FS. One of the newly discovered phenotypes, Fim, was of equivalent fitness PGA-WS, while both CAPP and PSL-WS were significantly less fit.

If Fim, CAPP and PSL-WS are of equivalent or higher fitness in comparison to FS, PGA-WS and CC, then why had they not been observed previously? One possible explanation is that, since the newly discovered phenotypes require two or more mutations, the intermediate steps on the way to realizing the final phenotype might be of very low fitness. To determine whether this is the case, we can examine fitness across each mutational step (Figure 2.14).

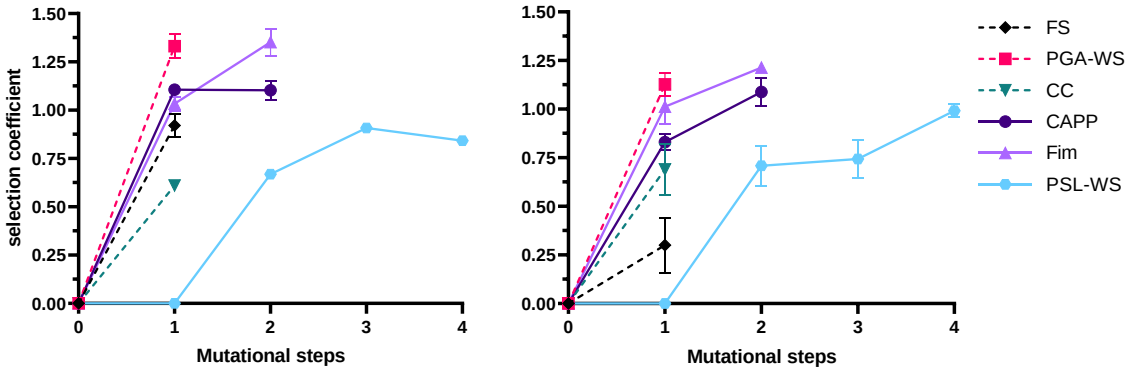

Figure 2.14: Possible fitness trajectories to alternative ALI-colonizing phenotypes. Fitness trajectories of the three previously identified phenotypes (dashed lines) and the three newly discovered phenotypes (undashed lines). Selection coefficient calculated from invasion assays as measured over 48-hrs (left) and 72-hours (right). For phenotypes requiring multiple mutations, only the trajectories that increment fitness at each step are shown. For the trajectory to PSL-WS, I have included four mutations, the first to signify the LoF of *pga* that would be required to avoid generating a PGA/PSL hybrid WS. Error bars represent SEM (n= 6 or 3).

When plotted as in Figure 2.14 the results can now be seen as alternative possible steps on a fitness landscape. From this perspective it is easy to see why the PSL-WS phenotype was not discovered earlier – it would require a neutral first step mutation (equivalent to the deletion of the *pga* operon) before any increase in fitness came about. However, a low fitness of intermediate steps cannot account for the rarity of the CAPP or Fim phenotypes – both have first step mutations that are of equivalent fitness to the FS phenotype (for CAPP t_7=2.00, P=0.08 and for Fim t_{10}=1.33, P=0.21, two-tailed t-tests) and significantly higher fitness than the CC phenotype (for CAPP t_4=16.4, P<0.001 and for Fim t_7=6.24, P<0.001, two-tailed t-tests). Therefore, if interpreted in terms of fitness alone, it would be expected that both the first-step CAPP and Fim mutants should be observed at least as commonly as the FS phenotype, and moreover be significantly more common than the CC phenotype.

At this point it is difficult to say to what extent the first-step CAPP and Fim mutants might exist alongside the previously discovered phenotypes within a microcosm. Of relevance here is the experiment conducted in section **2.2.5.6. Collecting additional Fim-causing mutations in Pflu1605,** which evolved full Fim mutants by the first-step mutation only. Recall that this was achieved through the use of a (PGA-WS, FS)-deficient strain already containing the neutral second-step Pflu1609^{D94A} mutation. After five days growth, 18 out of 20 microcosms were detected to contain full Fim mutants (i.e., those with a mutation in Pflu1605), indicating the Pflu1605 mutation can readily occur within this timeframe in an uncontested population. As both Rainey & Travisano (1998) and Lind et al., (2017) identified adaptive mutants via colony morphology, this first-step Fim mutant (which does not exhibit a unique colony morphology), could have easily escaped detection in these earlier experiments. The one-step CAPP mutant on the other hand does exhibit a distinct colony morphology yet still remained undetected.

While the extent to which the first-step CAPP and Fim mutants compete among the previously discovered phenotypes is uncertain, what can be said is that during the initial discovery experiments conducted here (**section 2.1.2.1**), the fully-realized CAPP and Fim mutants were only ever observed when the FS phenotype was made inaccessible to evolution. This suggested the evolution of FS might be precluding the evolution of the higher-fitness alternatives, CAPP and Fim. Note that since the mutational path to the CC mutant remained intact throughout, this phenotype does not preclude the evolution of any alternatives – an outcome that can be explained by its lower fitness relative to the first step CAPP and Fim mutants.

2.2.7.1 Confirming the precluding effects of FS through an extended evolution experiment

First, it was necessary to confirm that the observation of CAPP and Fim evolving only when FS was inaccessible was not simply due to chance – both CAPP and Fim occurred only rarely among the initial 20 microcosms during the initial discovery experiment (each occurring only twice). It was also important to demonstrate the effect in conditions where mutation supply was not limiting – it could be that the CAPP and Fim phenotypes arise so rarely simply because the limited number of replications within a microcosm does not allow the two mutations necessary for each phenotype to occur.

To test whether FS was indeed precluding the evolution of CAPP and Fim in a condition where mutation supply should not be limiting, a serial transfer experiment was conducted that compared the outcome of adapting populations capable of evolving into Fim, CAPP and FS, (‘FS+’) and another set of populations capable of only evolving into Fim and CAPP (‘FS-’). Serial transfer extends the number of generations over which mutations can accrue and so removes the restrictions on mutation supply associated with growth in a single microcosm.

Twenty microcosms inoculated with a FS+ genotype (Δ*wss*Δ*pga*) and twenty inoculated with a FS- genotype (Δ*wss*Δ*pga* + pSX (*fuzVWXYZ*)) were serially transferred at 7-day intervals, with 1% of the population propagated at each point. The expectation was that microcosms containing the FS- genotype would result in faster and more frequent evolution of the Fim and CAPP phenotypes if the inhibitive effect of FS was not due to chance alone. Indeed, this is what I found (Figure 2.15). In line with the evolution of FS inhibiting the evolution of the fitter alternatives, the rate of evolution of Fim and CAPP was significantly higher in the FS- populations than in FS+ populations (Log rank test, P=0.04; Gehan-Wilcoxon test, P=0.02). Unexpectedly, a PSL-WS type was also found to evolve in one of the FS- microcosms. As the extended evolution experiment was conducted before the fitness assays of PSL-WS had been completed, the high fitness of a AwsX$^{\Delta Y77-Q87}$ mutant had not been revealed – indeed this step reaches equivalent fitness to the first steps to CAPP and Fim. Given the fitness and known likelihood of AwsX$^{\Delta Y77-Q87}$ (the most common cause of classic WS) it may be the case that such single-step PSL-WS mutants would be competing within these populations and potentially inhibiting CAPP and Fim evolution by a similar means as FS. It is expected that precluding evolution of the PSL-WS phenotype and repeating the extended evolution experiment would result in a more overt effect.

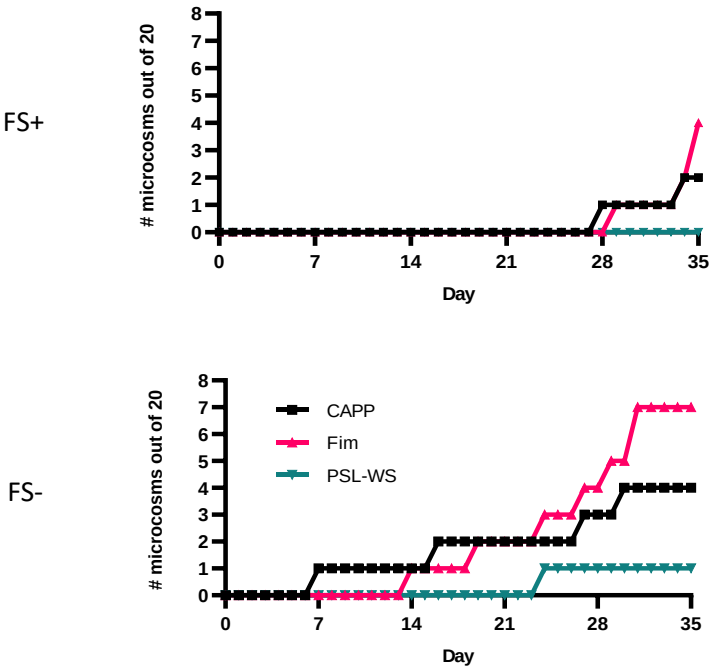

Figure 2.15 FS inhibits the evolution of higher fitness alternative phenotypes over an extended evolution experiment. Results are plotted in terms of the number of microcosms (out of 20) over time that displayed a confluent mat corresponding to CAPP or Fim (or PSL-WS). Top: FS+ populations in which FS was accessible to evolution. Bottom: FS- populations in which FS was inaccessible to evolution. Note that FS was observed after three days in all microcosms in the FS+ populations and no FS- microcosms were found to contain co-existing ALI phenotypes – each contained either Fim or CAPP (or PSL-WS).

It is important to note that some FS+ microcosms do eventually generate CAPP and Fim mutants, indicating that CAPP and Fim are able to evolve from the more limited but persistent pool of ancestral SM (maintained by frequency-dependent selection in the microcosm) and eventually arise, outcompete, and colonise the ALI if given sufficient time. This suggests they are not precluded by some specific and antagonistic fitness interaction with FS. Furthermore, that CAPP can invade a FS+ population, indicates its fitness, although not shown to be statistically significant over the 48-hour fitness assay (as seen earlier in Figure 2.14), is sufficient to invade FS-dominated population.

The results of the extended evolution experiment also suggest there is no subsequent step to improve the ALI-colonising ability of the FS phenotype, at least one that achieves a confluent mat across the ALI, as no microcosms containing only FS displayed anything more than the transient and partial ALI growth characteristic of FS. FS may therefore be a 'dead-end' – a local fitness peak that has no subsequent steps to improve colonisation of the ALI. A similar conclusion can be drawn for the CC mutant which remained intact in both the FS+ and FS- populations – no microcosms producing a confluent mat were found to be the result of a cell-chaining phenotype.

An important and as yet unstated point is that all alternative phenotypes are expected to only arise from the ancestral SM state, e.g., FS does not evolve into CAPP or vice versa. Although not explicitly tested through reconstruction of genotypes with mixed mutational paths to create 'hybrid' phenotypes, such hybrids are expected to result in incompatibilities and costly fitness effects. This was previously demonstrated in the case of the original WS and FS – the hybrid producing weaker mats than a normal WS (Ferguson et al., 2013). In the language of fitness landscapes this means there is reciprocal sign epistasis between the different phenotypes. Reciprocal sign epistasis is necessary to form a rugged fitness landscape – a topology that imposes extreme constraints on the paths natural selection will proceed by (Weinreich et al., 2005). For FS to reach a different peak it would require the sub-population descending in fitness: either returning to the SM ancestral state from which Fim and CAPP again become accessible, or through a presumably low-fitness 'hybrid' Fim/FS or CAPP/FS intermediate.

2.2.8 Comparing likelihood

So far, I have provided evidence that evolution of the FS phenotype can preclude the evolution of higher-fitness alternatives and that this cannot be explained in terms of fitness alone. These results might instead be attributable to differences in the likelihood of each phenotype: if the FS phenotype is more likely to be generated by mutation it would arise earlier in the population and – amplified by selection – gain a corresponding advantage over the equally fit first-step Fim and CAPP mutants.

The following sections address whether relative likelihood can indeed explain the observed precluding effect imposed by FS. In particular the focus is on the likelihood of the first step mutations necessary to generate CAPP and Fim compared to that required to generate FS. A reminder that likelihood depends on two quantities: the number of possible mutations able to generate a given phenotype or its intermediate steps (the target size), and the rate of occurrence for each mutation that 'hits' the target. Likelihood can be approximated by knowledge from the actual observed spectrum of causal mutations and class-specific mutation rates, combined with an understanding of the logic of how these mutations must alter the structure and function of the bacterium's G-P map to achieve the phenotype.

2.2.8.1 Likelihood of FS

Ferguson et al. (2013) previously characterized the spectrum of mutations able to generate FS. From a collection of 91 independent FS mutants, they reported 59 unique mutations including various substitutions and indels (including those causing frameshifts and premature stop codons) to a single gene, *fuzY,* that encodes an enzyme involved in lipopolysaccharide synthesis. All mutations were shown to generate a FS phenotype with equivalent fitness and a clean deletion of *fuzY* caused the same effect, indicating that this 1,148 bp locus represents a gene-wide LoF target. Some of the larger deletion events recorded in the 59 mutations even extended into neighbouring genes, including *fuzW, X and Z*. This indicates a particularly extensive mutational target size even for a gene-wide LoF target, as maintenance of neighbouring gene function would be expected to impose constraints in most other contexts. Thus, the mutational target is the entire *fuzY* CDS, unconstrained by mutations affecting the upstream and downstream genes, and includes any class of mutation that impairs protein function.

2.2.8.2 Likelihood of Fim

Now consider the likelihood of the first step to Fim that is of equivalent fitness to FS. This requires a domain-specific LoF within a small region of a single gene (*pflu1605*) that disrupts its predicted phosphatase activity but leaves downstream domains involved in phosphorelay unaffected and otherwise maintains the integrity of the protein. The necessity of maintaining C-terminal phosphorelay activity therefore rules out nonsense mutations and frameshifts in the phosphatase domain. Due to the relatively small target size and restricted class of permissible mutations, the likelihood of this first step occurring by spontaneous mutation is clearly far less than FS. Furthermore, to reach full expression of the phenotype and to realize a higher fitness than FS, a highly specific second-step GoF mutation is then required to occur in *pflu1609*. The sub-population of first step Fim mutants would therefore need to reach sufficient numbers in the presence of FS to generate this second GoF mutation.

2.2.8.3 Likelihood of CAPP

As for the first step to the CAPP phenotype, which requires a duplication of the transcriptional regulators encoded in *pflu3655-57*, the likelihood of such a mutation occurring is far more difficult to estimate. Given that any duplication that increases the copy number of the transcriptional regulators could in principle give rise to the first step of CAPP, the mutational target size is potentially very large: a near-infinite number of combinations of junction sites upstream and downstream of *pflu3655-57* would suffice. However, since all but the smallest of these duplications will generate very large numbers of gene dosage imbalances, we expect the vast majority to be associated with pleiotropic fitness costs and be selected against. It may be that the specific duplications observed in $CAPP_A$ and $CAPP_B$ are the only mutations that do not impose large pleiotropic costs, in which case the target size would be extremely small. The mutational target size of duplications is inextricably linked with fitness in a way that other kinds of mutations are not, making it extremely difficult to estimate.

How often the large duplications necessary for the first step of CAPP occur is also frustratingly difficult to estimate. Duplication rates vary drastically between species, between different regions of a given genome, and between mechanisms of duplication (Reams et al., 2010). Very little information about the rate of large duplications is available for SBW25. Long et al. (2018) failed to report any duplication larger than 7bp across 80 mutation accumulation lines of SBW25 (compared to a total of 253 base pair substitutions), suggesting that large scale duplications are extremely rare (or extremely unstable) in SBW25 under neutral evolution. Moreover, of the 522 mutations identified under conditions of adaptive evolution described in Chapter 3 of this thesis, only 4 large duplications were found, suggesting that these types of mutations rarely underpin adaptation in SBW25. Indeed, of all studies on adaptive evolution of SBW25, only one previous study has reported a duplication as the underlying mutation of a selected phenotype (Ayan et al., 2020).

Ultimately, the broad pleiotropic effects and extremely variable rates of large duplications make any sort of locus-specific estimate impossible. Understanding the likelihood of the first step of CAPP would require direct measurement. Since the duplication is expected to result in an increase in transcription of the *cap* operon, there exists the possibility of quantifying the mutation rate to the first step of CAPP (and indeed the first-step of Fim as well) by inserting a selectable marker to the end of *cap* and performing a fluctuation assay; only those cells that had undergone a duplication of the regulatory regions would express the selectable marker and form colonies on restrictive medium. This experiment would allow assessment not only of the rate of duplications resulting in the first step of CAPP, but by sequencing the mutants and identifying junctions, would also provide a broad survey of the different duplications capable of generating CAPP.

Following the duplication, the second-step to achieve full expression of CAPP then requires a GoF mutation within *pflu3677*. As was the case with Fim above, the subpopulation bearing the duplication would need to reach sufficient numbers in the presence of FS to generate the second step GoF mutation.

2.2.8.4 Summary of Likelihood

Although questions remain as to the likelihood of CAPP, relative likelihood is indeed able to explain the observed effect of FS precluding the evolution of Fim. There are different possible means by which this precluding effect could come about. One possibility sees the growth of the early arising FS mutant sub-population diminishing the pool of SM cells from which the first step Fim mutants can arise (i.e., be generated by mutation and escape drift). Another possibility is that the first step Fim mutations do commonly arise within the microcosm, but the sub-populations do not reach sufficient numbers to generate their second-step mutation, and therefore never realize the potential fitness that would allow them to invade an established FS population. It is also possible the negative-frequency dependent selection that occurs in the microcosm is relevant here: the strength of selection for ALI-colonisation may lessen as the first ALI-adapted mutant is amplified, increasing selection for non-mat formers (i.e., the ancestral SM) (Rainey & Travisano, 1998). Negative frequency-dependent selection could therefore conceivably amplify the advantage of an early arising ALI-colonising mutant.

Whatever the exact process, it is the differences in likelihood between phenotypes that are mediating it: only by understanding the interplay of selection and the processes that generate variation can the outcome of evolution within the microcosm be made sense of. This scenario, whereby evolution proceeds via a more likely path and arrives at sub-optimal phenotype, has been referred to as the 'survival of the likeliest' (Leighow et al., 2020).

2.3 Conclusion

That possible phenotypic variation is constrained by the existing structure and generative processes that form an organism is uncontroversial. For instance, epistasis and pleiotropy are well-established features of the G-P map that impose constraints on variation and have a long history: epistasis stems from Sewell Wright (1932) and pleiotropy from Darwin's 'correlation of growth' (Darwin, 1859), later formalized into multivariate quantitative genetics by Lande (1979). Both these properties refer to an underlying structure embodied in the G-P map but are generally identified indirectly through their effect on fitness, and invoked in regard to these fitness effects e.g., pleiotropy can be antagonistic or synergistic, and epistasis positive or negative. They are therefore important and integrated parts of

classical fitness considerations in the study of evolution. What remains far less studied, and a current source of controversy, is how the G-P map, in conjunction with biases in the mutational process, might influence the likelihood of different phenotypes being generated by mutation, and whether or to what degree this also determines the direction of adaptive evolution (Svensson & Berger, 2019; Stoltzfus, 2021).

As a means of isolating the causal role of likelihood, I have here conducted experiments that disentangle the contributions of fitness and likelihood to the phenotypic outcome of adaptation. The first major result – beyond the collective discovery of seven (including WS) distinct phenotypes that SBW25 is able to achieve a single adaptive function – was that the phenotypes evolution tended to favour were not necessarily the fittest. The goal was then to determine the reasons for this. A typical approach is to examine the fitness landscape and the constraints this imposes on the paths natural selection will proceed by (e.g., Weinreich et al., 2006). Indeed, doing so explained many aspects of the results. For example, the rarely observed but fitter phenotypes required two or more mutations that needed to occur in a particular order to increment fitness, whereas less-fit but initially observed types required only a single mutation. However, the results could not be adequately explained in terms of fitness alone. This was resolved by considering the likelihood of mutation generating the different phenotypes – some phenotypes (or the intermediate steps they require) were more or less likely than others. These differing likelihoods were accounted for by examining the bacterium's G-P map and the variable rates of the different classes of mutation by which it could be adaptively re-configured. Variants that were more likely, happened to arrive earlier in the population and so gained a corresponding advantage over rarer, later occurring phenotypes.

The results of this chapter also highlight the obvious but seemingly under-appreciated fact that likelihood can determine the outcome of adaptive evolution in cases of relative neutrality. It is surprising that 4 of the 7 phenotypes (FS, CAPP, Fim & PSL-WS) have first-steps of equivalent fitness. How often are alternative steps to increased fitness relatively neutral? The question is an open one that can be investigated empirically in other organisms using similar methods to those employed here. If somewhat common, it extends the population genetic grounds on which likelihood can be expected to direct adaptive evolution – a situation ordinarily expected to be limited to small populations where mutation supply is limiting such that only the likeliest phenotypes will be generated (Svensson & Berger, 2019; Schenk et al., 2022). In much larger populations where all possible first-step variants co-exist and compete, fitness is expected to always dominate the outcome, particularly so in asexual populations that experience clonal interference (Schenk et al., 2022; Good et al., 2012). Relative neutrality between alternative mutational steps means even in such large populations, likelihood can still be expected to play an orienting role in the outcome of adaptive evolution.

Even if selection does ultimately determine the outcome given a sufficiently large population – and putting aside the fact that such populations are not necessarily achieved in nature – there is still considerable information gained from understanding the underlying features of organisms that impart different likelihoods. In principle, this can allow the bottom-up prediction of evolution (Lind et al., 2019). Such an understanding is also necessary to interpret the patterns of mutation from sequencing data in a wide variety of contexts where the evolutionary and ecological process is most often obscured.

What is needed in these contexts, is mechanistic insight into the causes and magnitude of differences in likelihood that can be expected. Of particular interest is how different kinds of mutations might interact with specific regulatory architectures of the G-P map. For example, the significance of genes encoding negative regulators susceptible to a wide spectrum and number of mutations has been well established by previous work (McDonald et al., 2009; Lind et al., 2015). In a similar vein, and novel to the present study, is the example of the first-step mutation to the CAPP phenotype: a duplication of a transcription factor (*pflu3655-3677*) configured in a positive feedback loop with itself. Given the prevalence of positive feedback loops in regulatory networks, in which they are required for bistable gene expression (Dubnau & Losick, 2006; Zordan et al., 2007), duplication may be a likely and therefore expected means of their constitutive activation. Although perhaps not evident in SBW25, rates of duplication can be particularly high, with rapid gene amplification a well-known mechanism for adaptive evolution to increase expression levels through a gene dosage effect (Andersson & Hughes, 2009). It is plausible that if *pflu3655-3677* were in a different chromosomal context, flanked by large repetitive regions, that CAPP would be observed to evolve more readily.

Furthermore, with respect to the causes and magnitude of differences in likelihood, the results here also call attention to the possibility of extragenetic negative regulators that do not necessarily impart a large target size, as was observed for the first mutational step to reach Fim. Recall that in the hierarchy of likelihood presented by Lind et al. (2015), the most common predicted path to activate a given gene is through LoF of an extragenic negative regulator. In the case of mutation to Pflu1605 (a predicted extragenic negative regulator of the response regulator and transcription factor Pflu1606), the target size is severely limited by the requirement of not breaking the positive regulatory domain encoded within the same gene. This does not contradict the prediction: with respect to Pflu1606 (or the *fim* genes it activates) there may be other less-likely activating mechanisms that remain unobserved. It does however highlight the importance of considering intragenic functional domains and their potentially multi-functional context within a regulatory network.

Chapter 3

The evolution of evolvability via lineage selection

"... the more diversified in structure the descendants from any one species can be rendered, the more places they will be enabled to seize on, and the more their modified progeny will be increased."

Charles Darwin (1859)

3.1 Introduction

Evolution can be viewed as a game where the goal is to keep playing. Taking this perspective brings focus from the usual considerations of evolutionary thought – the properties of individuals and their fitness in the contemporary environment – and shifts it to the properties of lineages that enable their success over long spans of evolutionary time (Thoday, 1953; Eshel, 1973; Nunney, 1999; Palmer & Feldman, 2012; Graves & Weinreich, 2017).

One important determinant of long-term evolutionary success is a lineage's potential to generate adaptive phenotypic variation through genetic change – a property that has been called 'evolvability' (Dawkins, 1989; Kirschner & Gerhart, 1998). This property is required not only for a lineage to keep pace with changing selection pressures, including those imposed by antagonistic co-evolving lineages, but also to outcompete lineages occupying the same niche, and to innovate and diversify into novel niches where competition is absent. In theory, the differential birth and death of lineages that results from variation in their ability to achieve these outcomes can lead to selection favouring more evolvable lineages, analogous to how individual-level selection favours fitter individuals (Conrad, 1990; Dawkins, 1989; Eshel, 1973; Graves & Weinreich, 2017). A more evolvable lineage would in this sense be one associated with a greater likelihood of generating individuals that have adaptive phenotypes. The idea that evolution can act in this self-facilitating manner has been referred to as the 'evolution of evolvability' and despite its intuitive appeal is a controversial notion (Dawkins, 1989; Sniegowski & Murphy, 2006; Lynch 2007; Pigliucci, 2008).

Note that if selection is to act on evolvability in this manner, it must do so as an indirect consequence of the individual-level selection process – evolvability exists only as a potential and so cannot be selected directly. Rather, it is the direct selection of the realized phenotypes of individuals that indirectly favours the systems that enabled those phenotypes to be generated. Indirect selection acting on particular traits of these systems has been referred to as 'second-order' selection, and the traits themselves as 'modifying' traits (Eshel, 1973). Through altering the systems by which phenotypes are generated, a modifying trait affects the variability (potential to vary) of an organism (Wagner & Altenberg, 1996).

3.1.1 Examples of modifying traits

A commonly considered modifying trait is the global mutation rate. By determining the total input of genetic variation, the mutation rate obviously plays a fundamental role in determining evolvability – if this rate is too low, a lineage will be driven extinct through failure to adapt; too high, and the lineage

will be overcome with deleterious variants. Since most mutations with a phenotypic effect are harmful to the organism, there is in general relentless selection for global mutation rates to be minimized to some practical limit (Raynes & Sniegowski, 2014; de Visser, 2002; Lynch et al., 2013). An elevated mutation rate is, however, expected to increase evolvability in situations where adaptation is limited by the supply of variation, such as in rapidly fluctuating environments or in small maladapted populations (de Visser, 2002). Indeed, lineages with increased global mutation rates caused by inactivation of the cell's repair machinery (global 'hypermutators') have been documented as arising in experimental populations of bacteria several times (Elena & Lenski 2003; Pal et al., 2007). Such lineages are found rarely in nature, however – likely because the accumulation of deleterious variants leads to their inevitable extinction (Sniegowski et al., 2000).

Another classic modifying trait is sexual recombination (Maynard Smith, 1978). Through its ability to purge deleterious alleles and bring beneficial alleles together, sexual recombination is widely understood as a modifier that increases evolvability (Barton, 2009; Maynard Smith, 1978). It also appears to be maintained through lineage selection at the species level, with extant species that have reverted to asexuality being found in only the most recently derived phylogenetic branches, suggesting such lineages are destined to become evolutionary dead-ends (Maynard Smith, 1978). Experimental work comparing adaptation of an organism with and without sex has confirmed its advantage is in line with theoretical expectations – sex speeds adaptation and so represents an increase in evolvability (McDonald et al., 2016).

The significance of both global mutation rate and sexual recombination to the evolutionary process have long been considered (Fisher, 1930; Sturtevant, 1937). The potential also exists for a more recently considered kind of modifying trait that is embodied in the architecture of an organism's genotype-phenotype map (G-P map), or that affects mutation rates and spectra in a locus-specific manner (Wagner & Altenberg, 1996; Moxon et al., 1994; Martincorena et al., 2012). Rather than altering the overall rate or mode of genetic change, which promotes variation in all directions, such modifiers have the advantage of promoting or inhibiting particular dimensions of phenotypic variation. As a consequence, there is the possibility for lineages to align patterns of variability with regularities experienced in past selective environments (Watson, 2020). Provided the past resembles the future to some extent, this process can lead to an increase in lineage evolvability – thereby giving the blind process of evolution apparent foresight. While no experiments have demonstrated the de novo evolution of a modifier of this type, the phenomenon has been observed in a broad range of evolutionary simulations (e.g., Crombach & Hogeweg, 2007; Kashtan & Alon, 2005; Parter et al., 2008, Draghi & Wagner, 2008; Zaman et al., 2014; Canino-Koning et al., 2019). A concrete biological example of this kind of modifying trait can be found in bacterial contingency loci (Moxon et al., 1994). These

are hotspots of frameshift mutation resulting from repetitive sequences of nucleotides (either tracts of a single nucleotide or short repeating motifs) that lead to elevated local rates of slipped-strand mispairing (Levinson & Gutman, 1987). Such contingency loci are generally associated with genes encoding surface structures of host-associated bacteria. Since these structures are commonly targeted by the host immune system, high-frequency and reversible frameshift mutations enable the expression of these structures to be rapidly switched on and off – an ability thought to facilitate immune evasion (Moxon et al., 2006).

3.1.2 Objections

Despite the existence of modifying traits in nature that can only be reasonably understood as increasing the evolvability of a lineage (such as sexual reproduction and contingency loci), the degree to which natural selection is thought to optimize for evolvability is highly contentious. While generally accepting its theoretical possibility, those critical of the notion dismiss its relevance based on the narrow conditions it requires and emphasize alternative evolutionary processes that give rise to apparent evolvability as an incidental by-product (Lynch, 2007; Sniegowski & Murphy, 2006; Brookfield, 2001; Dickinson & Seger, 1999).

The primary objection is that selection above the level of the individual is necessarily weak and, consequently, the ability of natural selection to optimize for long-term benefit (i.e., benefits conferred to the lineage) is severely limited. The weakness argument is based on the discrepancy in generation time between the higher-level and that of the individual: the relatively short generation time of individuals allows variation to be rapidly amplified by selection and therefore natural selection most strongly favours the traits of individuals that confer benefits in the immediate environment. This occurs at the expense of long-term benefit: variation that may have future utility is not ‘seen’ in this timeframe and is at risk of becoming extinct before its beneficial fitness effects can be realized. Natural selection is said to be myopic in this sense (Graves & Weinreich, 2017).

The conditions required to overcome this short-sightedness of natural selection are however very simple: competing lineages that vary in evolvability must co-exist for long enough such that their differing evolutionary potential can be manifest into realized phenotypic consequences. If these criteria are satisfied, then variation in evolvability will become visible to natural selection at the lineage level and inevitably lead to the amplification of more evolvable lineages. Experimental insight into the issue is desirable but extremely difficult to come by.

3.1.3 Experimental challenges

Addressing the effect of lineage selection on evolvability is experimentally challenging for a variety of reasons. Foremost is that variation in evolvability is only manifest over many generations and examination of the process therefore requires long periods of evolution and observation. This issue can be largely overcome by experimentally evolving microbes with short-generation times, the approach I take here. However, significant challenges remain. First, such an experiment requires an ability to identify – and track the evolution of – lineages. Second, it must be possible to detect changes in evolvability between lineages and distinguish this from chance events. Third, if changes in evolvability are detected, understanding where this variation in evolvability stemmed from requires extensive knowledge of the organism's genetics. As a consequence, opportunities for experimentally addressing the ability of lineage selection to shape evolvability are rare. In the next sections I describe an existing experiment and how it was modified to overcome each of these issues.

3.1.4 The life cycle experiment

Essential blueprints for examining lineage selection and evolvability can be found in an experiment previously conducted in the Rainey lab and published by Hammerschmidt et al. (2014). As initially conceived, this experiment concerned the evolution of multicellularity and was designed to test a unique hypothesis that makes use of the effective multicellular life-cycle that emerges from the eco-evolutionary dynamics of the bacterium *Pseudomonas fluorescens* SBW25 when grown in a static microcosm.

3.1.5 The emergence of a life-cycle

SBW25 is an obligate aerobe and upon inoculation into a static microcosm begins to proliferate and deplete the dissolved oxygen throughout the liquid. As this occurs, there remains an influx of atmospheric oxygen entering across the air-liquid interface, creating a strong selection pressure for mutant cells able to colonise the liquid surface (Rainey & Travisano, 1998). After 2-3 days a mutant inevitably arises that is characterized by an over-expressed extracellular cellulose polymer that acts as a cell-to-cell glue. As the dividing mutant cells bind to each other, and through adhesion to the microcosm wall, a visible mat is formed at the liquid surface. The new phenotype, termed Wrinkly Spreader (WS) due to its unique colony morphology when plated, rapidly outcompetes the ancestral Smooth (SM) morphotype that occupies the liquid (Rainey & Travisano, 1998).

The WS mat is a product of cooperation – producing the mat is costly to individual cells but allows the collective (group of cells) to colonise an otherwise unavailable niche (Rainey & Rainey, 2003). Due to its cooperative nature, an established WS mat becomes susceptible to invasion by cheaters – mutant

cells reverted to the ancestral SM phenotype that no longer produce the cellulose polymer. Using the mat as a physical substrate for growth, the cheaters gain the benefit of being localized at the air-liquid interface but without paying the cost of cellulose production. They are therefore selectively amplified – their proliferation eventually leading to collapse of the mat and the undoing of the co-operators.

From the perspective of nascent collectives transitioning to a multicellular mode of being, such cheater cells pose a serious problem. A resolution to this problem was suggested after it was observed that if an SM cheater cell was subsequently placed into a fresh microcosm that it would again evolve a WS mat, and again lead to the evolution of cheaters. The observation that the cycle could be repeated led Rainey (2007) to consider the problem of cheaters differently – namely by viewing cheaters instead as propagules of the collective WS mat; 'germ' cells that can disperse and re-establish the mat, or 'soma'. The cheater cells then become an essential means of collective-level reproduction (Figure 3.1).

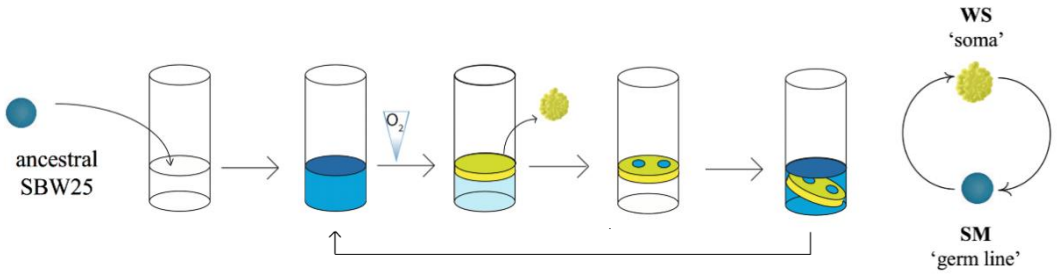

Figure 3.1 The life-cycle. The eco-evolutionary dynamics of SBW25 when grown in a static microcosm (a rich-medium-filled glass vial) naturally gives rise to a repeating cycle of selection that alternately favours two different phenotypes: the cellulose-producing WS that is analogous to a soma (depicted in yellow) and the non-cellulose producing (and motile) SM that is analogous to a germ line (depicted in blue).

Hammerschmidt et al. (2014) performed an experiment to test whether this could be the basis for the evolution of a multicellular entity. To do so they implemented a lineage-level birth-death process that favoured those lineages able to continue the life-cycle, and so through this 'cheat-embracing' strategy, achieve maintenance of the collective (Figure 3.2).

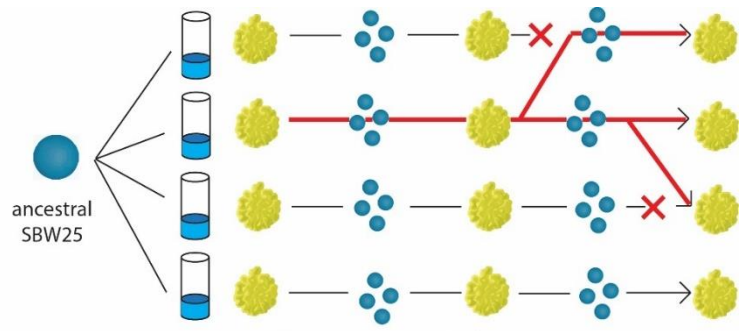

Figure 3.2 The life-cycle experiment imposes a birth-death process at the lineage level. At each stage of the cycle, failure to generate the necessary phenotype results in lineage extinction. Survival of a lineage is rewarded with the chance for reproduction – to fill the vacated microcosm of an extinct lineage. The lineage chosen for reproduction is done so randomly from the set of extant contemporary lineages. Shown here is a population of four lineages traversing five phenotypic transitions.

More specifically, in the Life Cycle Experiment (LCE), single WS colonies independently evolved from ancestral SBW25 were inoculated into replicate microcosms and left for six days. The bottlenecked WS colony was to ensure the WS mat remained free of within-mat conflict. The WS mat was required to remain intact (not collapse) throughout this period, as well as to produce SM cells at the end of the six days. Production of SM cells was determined by plating a sample of each microcosm onto agar and screening for the relevant colony morphology. The SM 'propagules' on each agar plate were then pooled and inoculated into a fresh microcosm and plated after three days to screen for the production of WS types. This process was then iterated down descendant lines of microcosms, establishing lineages that were spatially bound within the microcosms and that varied based on the idiosyncratic mutational paths they took to achieve each phenotypic transition. Failure to reach the target phenotype or failure to maintain the mat, resulted in lineage extinction (death) and replacement (birth) by a random extant lineage from the population. This lineage-level birth-death process effectively allowed for the existence of an entity evolving by natural selection on a timescale greater than that of its constituents, as the necessary 'ingredients' for natural selection to act (variation, reproduction and differential fitness) were manifest at the lineage level (Lewontin, 1970; Godfrey-Smith, 2009). While lineages would at first proceed by a variety of mutational paths to turn WS on and off, the expectation was that a form of regulation might eventually arise through this process. In other words, lineages themselves would adapt.

Indeed, the fittest lineage (i.e., the lineage with the lowest extinction rate) was found to have evolved a form of regulation that enabled it to reliably switch between the two states and therefore had adapted at the lineage level. This was achieved not by means of epigenetic regulation but by an increase in evolvability – the decreased extinction rate was found to be a result of a modifying trait that changed how the two phenotypes were being generated by mutation. Specifically, the fittest lineage carried a loss-of-function mutation to a gene encoding a mismatch repair protein (*mutS*) that caused an elevated global mutation rate. Significantly, this also resulted in a change in the spectrum of mutations such that expansion and contractions of homo-polymeric guanine repeats were particularly common. Such a tract of guanine residues existed in the ancestral SBW25 genome within an enzyme affecting production of the cellulose glue – a periodic expansion and contraction of this tract therefore led to periodic frameshifts that modulated WS expression on and off. Thus, a semi-

stable and reversible genetic switch had been established – a strategy reminiscent of a contingency locus, although dependent on the modifying trait of global hypermutation via inactivation of MutS.

3.1.6 The LCE as a means to examine the evolution of evolvability

The design of the LCE is uniquely suited to examining the evolution of evolvability. Firstly, although the emphasis in the original LCE was on maintenance of the collective, the selective regime in this experiment is also essentially one of alternating periods of selection toward different target phenotypes. This is the exact principle used in numerous evolutionary simulations that have demonstrated the efficacy of lineage selection to increase evolvability (Crombach & Hogeweg, 2007; Parter et al., 2008; Kashtan & Alon, 2005; Kashtan et al., 2007; Canino-Koning et al., 2019). Indeed, the possibility of lineage adaptation outside of an evolvability mechanism is likely to be extremely limited given that selection in the LCE is based on colony morphology – a criterion that favours lineages that have constitutively activated or inactivated the WS phenotype. Lineages may therefore be seen as being 'locked in' to strategies that rely on constructing biases that facilitate on-off mutations of the WS phenotype

The design of the LCE also meets the challenges raised in section **3.1.3 experimental challenges** of tracking lineages, detecting changes in evolvability, and connecting these to changes to properties of the organism. Specifically, the problem of tracking lineages is solved through bottlenecks and the fact that lineages undergo an explicit birth-death process, making it is possible to directly track their rise and fall over the course of the experiment. Changes in evolvability can then be detected by a change in lineage extinction rate. The relatively simple and well-characterised genetics underpinning the WS phenotype (to be addressed in a following section) enable understanding how this variation in evolvability between lineages comes about.

3.1.7 A modified LCE

Motivated by the possibility of discovering alternative switching mechanisms beyond that relying on global hypermutation, as well as an interest in how long lineages could persist even without a defined switching mechanism, I took a subset of the surviving lineages from the LCE and subjected them to extended evolution in a modified version of this experiment (henceforth the modified LCE).

The modified LCE follows the methodology of the original LCE but with two important modifications that aid in tracking lineage evolution and evolvability. First, I implemented a single colony bottleneck on both phases of the cycle so that a single clonal lineage enters each phase (Figure 3.3). This allowed for explicit tracking of the trajectory of a lineage through genotype space, as the exact mutation

causing activation or inactivation of the WS phenotype can be identified through the step-wise sequencing of single clones. Second, I ignored the requirement for the mats to not collapse. This modification was made because extinction due to mat collapse can occur despite a lineage having successfully transitioned to the target SM phenotype. The cause of success or failure at the level of interest is therefore obscured when mat collapse is a criterion for selection.

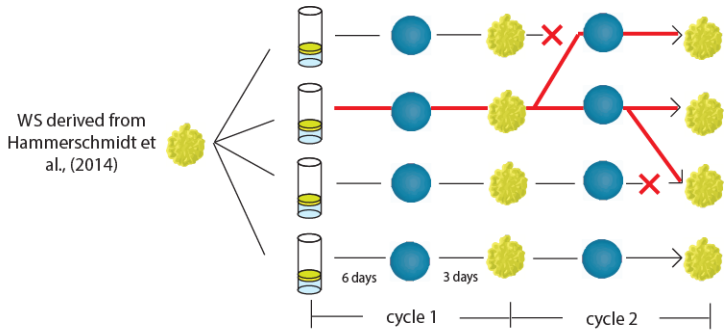

Figure 3.3 The modified LCE. Single colony bottlenecks at both the smooth (blue) and WS (yellow) phase allow direct identification of causal mutations underpinning the transitions between phenotypes

3.1.8 Possible mutational paths to activate and inactivate WS

Before any possible switching mechanism can be arrived at, lineages evolving through the LCE must first avoid extinction by repeatedly turning WS on and off via unique mutations. The extensive knowledge of WS genetics means it is possible to formulate expectations for how this process plays out. For example, consider the first transition from wildtype SM to WS that occurred during the original LCE. Expression of the WS phenotype requires that the cellulose synthase machinery is allosterically activated by the secondary messenger molecule, cyclic-di-GMP. Concentrations of this molecule within the cell are defined by the opposing activities of two enzymes: synthesizing diguanylate cyclases (DGCs) and degrading phosphodiesterases (PDEs). As detailed in the introduction to chapter 2, mutational paths to WS from wildtype SBW25 are exceptionally well-characterized: the first step to WS invariably proceeds via mutations that activate one of three DGCs (known as WspR, AwsR and MwsR). Evolution proceeds via these three paths despite there being many other DGCs also capable of generating WS – a result of the negative regulation these three DGCs are under. Negative regulation imparts a large mutational target size as loss-of-function (LoF) mutation that breaks repression can activate the DGC and generate WS. As a consequence, the three negative regulatory paths are more likely to be accessed by mutation and become visible to selection. Alternative paths to WS that activate other DGCs by different means exist, but these require rarer gain-of-function (GoF) mutations or multiple LoF mutations (Lind et al., 2015). Genome sequencing conducted by Hammerschmidt et al., (2014) on the surviving lineages of the original LCE indicates that the main

targets of mutation in the original LCE were indeed the three negatively regulated DGCs (Figure A.1 in appendix).

Consider now the WS→SM transition following the initial evolution of WS from wildtype SBW25. The activated DGC (Wsp, Aws, or MwsR) now becomes a mutational target for turning WS off via a LoF mutation. However, there is also the possibility of a mutation activating a PDE to counteract the cyclic-di-GMP synthesized by the previously activated DGC. Additionally, there is the possibility for LoF to any one of a number of genes that have been identified as essential for the WS phenotype through extensive suppressor analysis (Gherig, 2005; McDonald et al., 2009). Most prominent here is the wrinkly spreader structural (*wss*) operon that encodes the cellulose synthase machinery, although many other genes involved in glucose metabolism (cellulose is a glucose polymer) and biogenesis of the cell envelope (in which the Wss complex is embedded) are also essential for WS. The different possible mutational trajectories to turn WS off are summarized in Figure 3.4.

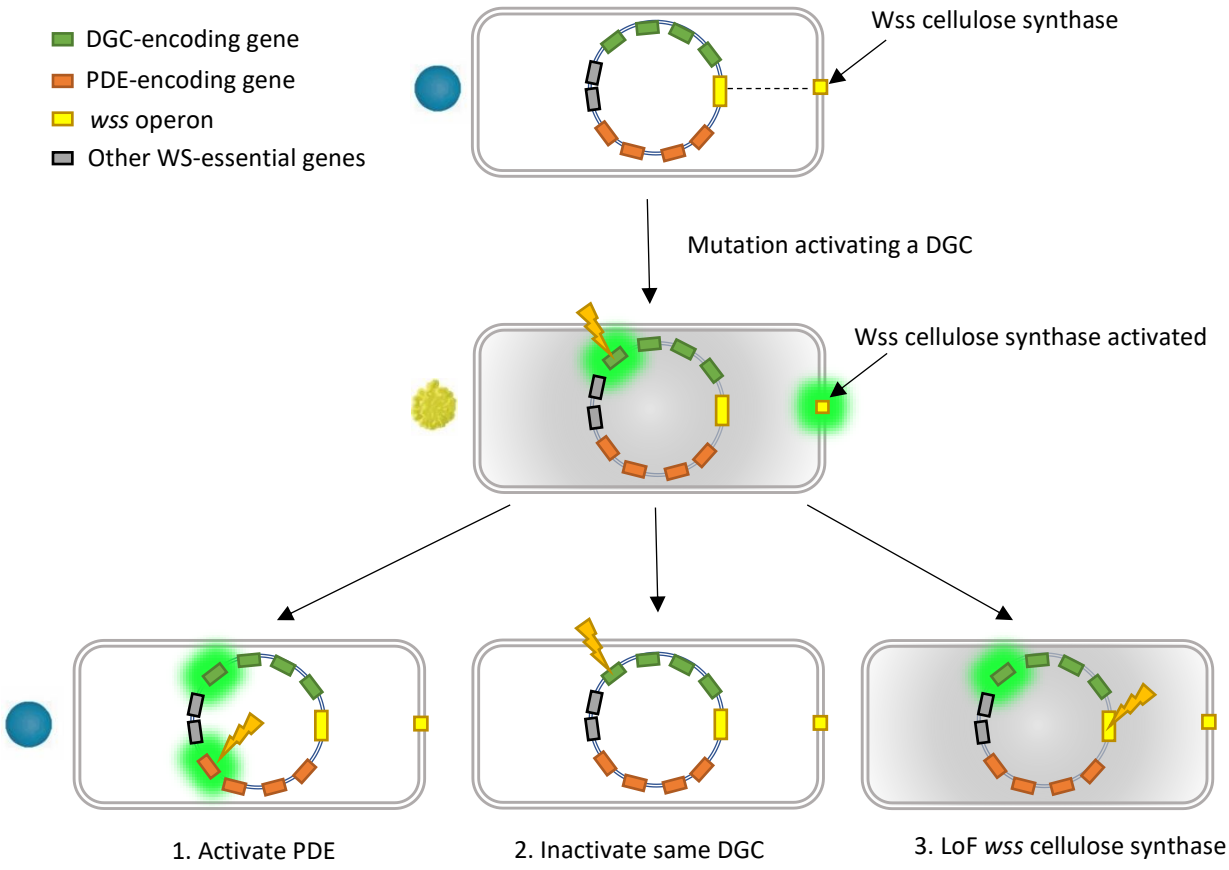

Figure 3.4 Different possible mutational paths to modulate the WS phenotype. The first step to WS invariably proceeds via activation of a DGC, however following this there are multiple different paths to then turn WS off. The green glow indicates that object as being 'active'. The grey gradient permeating the cell indicates

intracellular cyclic-di-GMP levels are high. A fourth possibility to turn WS off (not pictured) is through LoF to one of the other WS-essential genes identified through suppressor analysis.

The specific way that WS is turned off then determines the possible avenues for how WS can be turned back on again – and whether this will be possible at all. For example, LoF to the *wss* cellulose synthase operon necessitates this mutation being repaired or compensated for during the next SM→WS transition to avoid extinction. Any irreparable LoF mutation to *wss*, such as a large deletion, would guarantee lineage extinction – akin to 'locking the door and throwing away the key'. In contrast, the redundancy of genes encoding regulatory DGCs and PDEs that exists in the SBW25 genome provide multiple avenues for repeated on-off transitions and irreparably breaking any one of them will not necessarily lead to extinction. Still, these regulatory paths are finite in number: the probabilistic tendency of evolution to proceed via LoF and the different kinds of mutations this tends to permit (such as large deletions and in general mutations that may be difficult to compensate) means that the initially large total of regulatory genes is expected to be gradually whittled away. Consequently, subsequent phenotypic transitions progressively more improbable as lineages proceed through the experiment (Figure 3.5).

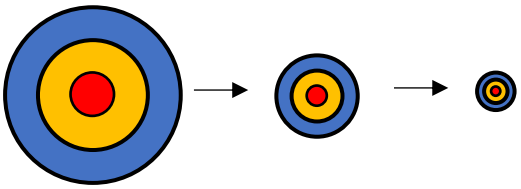

Figure 3.5 WS-generating regulatory target size diminishing. Destructive LoF mutations means the number of regulatory targets able to generate WS, which is at first vast, progressively diminishes – presumably making subsequent transitions more improbable.

The whittling of regulatory WS targets provides a null expectation for the fate of lineages experiencing extended evolution through the modified LCE, and it is useful to frame the challenge to persistence a lineage faces in this way. Long-term lineage survival therefore requires both avoiding sudden genetic dead-ends such as those caused by irreparable LoF to a WS-essential gene, as well as avoiding this slower but inevitable decline in regulatory target size.

Note that since there is always a means of inactivating the WS phenotype (for example, through LoF to *wss*), the target size consideration is less relevant to the WS→SM transition, although here the speed of generating SM is relevant: the earlier SM cells are generated, the more they will be able to spread throughout the mat and the higher their probability of appearing on the screening plate.

3.1.9 Persistence and adaptation

How lineages navigate the different possible mutational paths before arriving at a defined switching mechanism concerns a form of evolvability distinct from the previously discussed notion of evolvability as adaptation. Rather, differential persistence of lineages at this stage depends on the neighbourhood of possibilities surrounding the genotype of a lineage at a given time – a neighbourhood that is a consequence of the idiosyncratic path a lineage had taken to traverse previous transitions and a neighbourhood that will change with each subsequent transition. Any advantage one lineage has over another with respect to this form of evolvability is therefore short-lived. I will refer to this form of evolvability as 'persistence' to distinguish it from evolvability as adaptation, which requires the establishment of a modifying trait that is repeatedly associated with a particular effect (allowing the modifier to be repeatedly amplified by selection and potentially refined). The distinction is crucial, as evolvability in the sense of persistence is uncontroversial and an expected consequence of evolution being restricted to sampling the genotypes adjacent to the historical trajectory of a lineage (Maynard-Smith, 1970).

3.2 Results & Discussion

Four WS colonies representing surviving lineages from the original LCE were used to seed the modified LCE. The four lineages were referred to by Hammerschmidt et al. (2014) as line 5, line 43, line 54 and line 57. The WS colonies each seeded eight microcosms – establishing 4 independent populations of 8 initially identical (sub-) lineages, that would begin to vary following their first phase of adaptation to the target SM phenotype. Each microcosm contained 6 ml of growth medium as in the original LCE and in the experiment of chapter 2 – the carrying capacity of a single microcosm was therefore approximately 10^{10} cells and the rate of turn-over at stationary phase is unknown. The large population means mutation supply should rarely be limiting for single-step mutations, but multiple-step mutational paths are expected to be restricted.

3.2.1 Fates of the four lineage populations through extended evolution in the modified LCE

The fates of the four lineage populations are displayed below in lineage trees that trace the history of death and birth in each population across time, with the coalescence of the final surviving lineages highlighted (Figure 3.6).

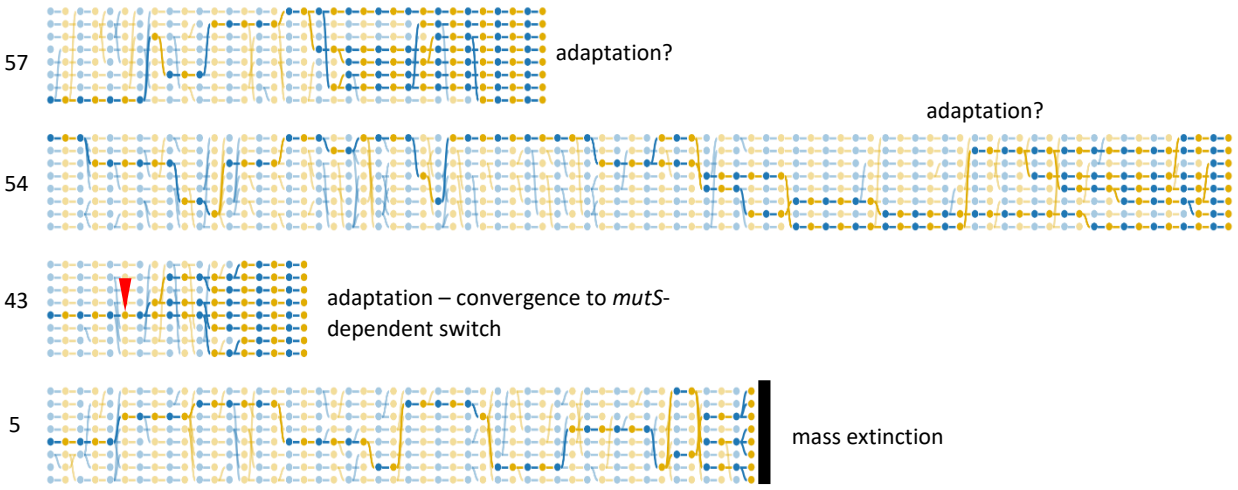

Figure 3.6 Lineage trees depicting the fate of each of the four lineage populations. Each lineage population consisted of 8 microcosms through which a single lineage was passed. The blue nodes on the lineage tree indicate the target phenotype as SM and yellow as WS. Lines connecting nodes indicate a lineage had survived that transition by successfully generating the target phenotype and branch points indicate lineage reproduction events. Lineage reproduction occurred when a lineage failed to generate the target phenotype – allowing a randomly selected contemporary lineage that had successfully transitioned to expand into its vacated microcosm. Potential lineage-level adaptations can be identified through the rapid fixation of a single lineage and subsequent decrease in lineage extinction (indicated by long unbroken horizontal lines of successful phenotypic transitions). This signature of lineage-level adaptation is best observed for the line 43 population,

which was found to converge on a known switching mechanism previously identified by Hammerschmidt et al. (2014) in the original LCE. The red arrow indicates the timepoint at which the triggering *mutS* mutation was found in this population. Both line 54 and 57 populations also show potential signatures of lineage-level adaptation, while line 5 shows no such signal and indeed ended in a mass extinction (as indicated by the black bar).

Briefly, the descendants of line 5 ended in mass extinction after 48 phenotypic transitions in the modified LCE (69 from wildtype SBW25). Mass extinction of the line 5 population was due to the simultaneous failure of all 8 lineages to generate WS cells – a fate repeatable upon two replay experiments. Rather than having arrived at a genetic dead-end (such as irreparable LoF to the *wss* operon), this failure was traced to the fixation of a lineage whose SM cells were able to perform mat formation to some degree and so led to selection of WS cells being lessened. Since this does not relate to evolvability in the sense I am interested in, the details of this process will not be addressed further.

As to the line 43 population, one of the lineages was found to have converged on the same *mutS*-dependent switching mechanism as in the original LCE. This lineage subsequently fixed in the population after 14 transitions (35 from wildtype SBW25). Evolution of the population was halted following the discovery of this mechanism. Work confirming the convergence on the *mutS*-dependent switch will also be omitted here but involved whole-genome sequencing identifying a mutation in *mutS* and Sanger sequencing of the sensitive G-tract demonstrating this locus was indeed modulating WS on and off in a semi-consistent manner in line with the mechanism described by Hammerschmidt et al., (2014).

Analysis of the line 57 population is currently incomplete and will not be addressed further, although the pattern of rapid fixation and low extinction in the coalescence strongly suggests the surviving lineages may have also discovered a switching mechanism. The remainder of this chapter will focus exclusively on the long-evolved line 54 population.

3.2.2 The line 54 population

Evolution of the line 54 population was halted after 80 phenotypic transitions in the modified LCE (101 in total from wildtype SBW25). There were 119 lineage deaths and subsequent births. Figure 3.7 displays a detailed version of the evolutionary dynamics of this population, including the phase-specific extinction rate over time.

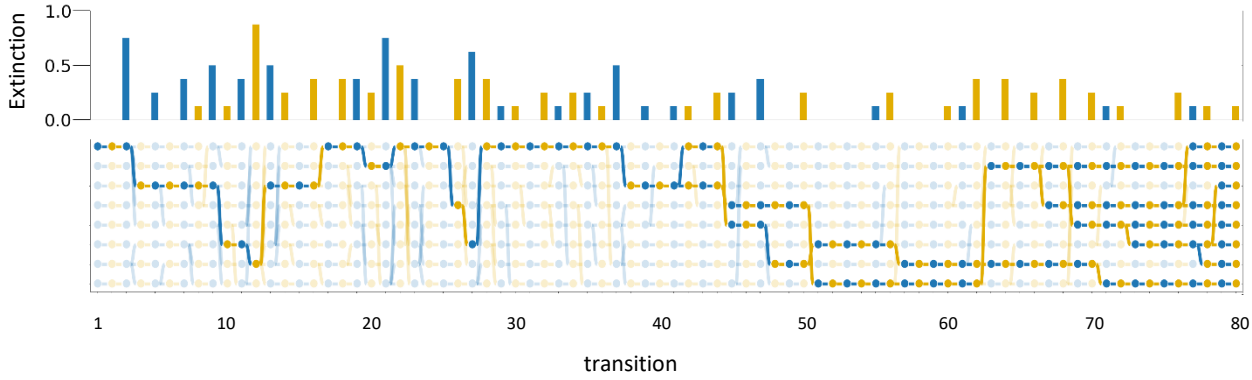

Figure 3.7. Lineage tree and extinction rate of line 54 population. The history of the final surviving lineages is highlighted. Yellow nodes are those points where WS was the target phenotype; blue where SM was the target phenotype. Yellow bars indicate the proportion of extinctions through failure to generate WS; blue bars for failure to generate SM.

The extinction rate appears to decrease as evolution progresses, this being particularly evident during the WS→SM phase (blue bars in top plot of Figure 3.7). Indeed, the low extinction rate has evidently allowed two lineages to coexist for 40 transitions. This is striking given that the longest any two lineages coexisted in the line 5 population (the population that failed to discover a switching mechanism and ended in mass extinction) was only 11 transitions. The low extinction rate suggested the line 54 population may have adapted at the lineage-level.

In order to understand the genetic basis of the remarkable persistence and possible lineage-level adaptation of the line 54 descendants, whole-genome sequencing was conducted on the clones following every one of the 521 phenotypic transitions.

3.2.3 Overview of sequencing results from the line 54 population

Step-wise sequencing of each phenotypic transition that occurred in the line 54 population identified 522 mutation events. The vast majority of phenotypic transitions were associated with a single mutation, allowing the causative mutation of the transition to be unambiguously inferred in most cases. There were 47 transitions associated with two mutations, and only a single transition associated with three mutations. No mutations were identified that could be implicated in alteration of the mutational repair or replication machinery of the cell and the surviving lineages are therefore expected to share the global mutation rate and spectrum of ancestral SBW25.

A number of phenotypic transitions could not be accounted for by mutation and this was due in part to poor sequencing output, although in some cases, no mutations were detected despite accurate sequencing. Here, this was undoubtedly due (at least in part) to the phenotypic plasticity of colonies that was occasionally observed. Specifically, it was at times noted that a single genotype could appear

SM or WS depending on the density of colonies surrounding it (dense arrangements leading to an apparent WS colony and dispersed arrangements giving an apparent SM colony). This plasticity obscured selection for mutationally-caused phenotypic transitions but was inevitably detected shortly after arising. Despite this minor caveat, the current sequence dataset provides a wealth of very precise information that, when mapped to the history of death and birth of the lineage tree, allows the evolutionary dynamics of the lineage population to be understood in great detail.

3.2.4 The spectra and biases of mutations

Of the 522 mutation events identified in the sequencing, 47% (244) were point mutations (i.e., base pair substitutions). Of these, there was a strong bias (155:89) toward transitions (155: purine to purine or pyrimidine to pyrimidine) over transversions (89: purine to pyrimidine or vice versa) compared to the null expectation of twice as many transversions, consistent with a near universal transition:transversion bias across diverse taxa (Stoltzfus & Norris, 2016). Within transitions a two-fold bias of GC→AT mutations was also observed, consistent with a wide range of other bacteria (Hershberg & Petrov, 2010).

The next largest mutational class was deletions, which accounted for 30% (157) of mutations, followed by insertions, accounting for 22% (117) of mutations. Larger-scale structural variants (defined here as mutations encompassing more than a single gene) were rare – only four such duplications were detected, ranging in size from 1547 bp to 2,454,853 bp. Large-scale deletions were similarly rare – only two such deletions were detected (58,704 bp and 58,121 bp). No evidence of mutation mediated by mobile genetic elements was found and indeed none has ever been noted in SBW25 – a finding in contrast to other evolution experiments such as those using *E. coli* (Consuegra et al., 2021).

Remarkably, direct reversions, which are generally expected to be exceedingly rare, accounted for 22% (116 of 522) of all mutations. This extreme tendency was primarily explained by the same 7bp sequence repeatedly cycling between duplication and deletion in *pflu0185*, a previously uncharacterized cyclic-di-GMP regulatory gene, which alone accounted for 76% (88 of 116) of direct reversions. The toggling of a single mutation in this manner suggested the evolution of a genetic switch. The origin and mechanism of this switch will be addressed in a later section, where it will be shown to indeed be responsible for the apparent lineage-level adaptation in the line 54 population. Removing the influence of mutations associated with the switch made deletions just over twice as common as insertions (55:22), a result more consistent with the deletion bias reported in other bacteria (Sung et al., 2016).

3.2.5 Overview of mutational targets and their function

The 522 detected mutations targeted 72 unique genes, although 44 of these genes were only targeted once. The remaining 28 genes, mainly comprising *wss*, PDEs and DGCs (Figure 3.8), that were targeted at least twice therefore accounted for 92% (478 of 522) of mutation events. The vast majority of mutations were intragenic, with only 3.4% (18 of 522) occurring in non-coding regions (excluding the six large structural variants that encompassed both intra- and intergenic regions). Outside of mutations composing the genetic switch there were 31 instances of identical parallel mutations.

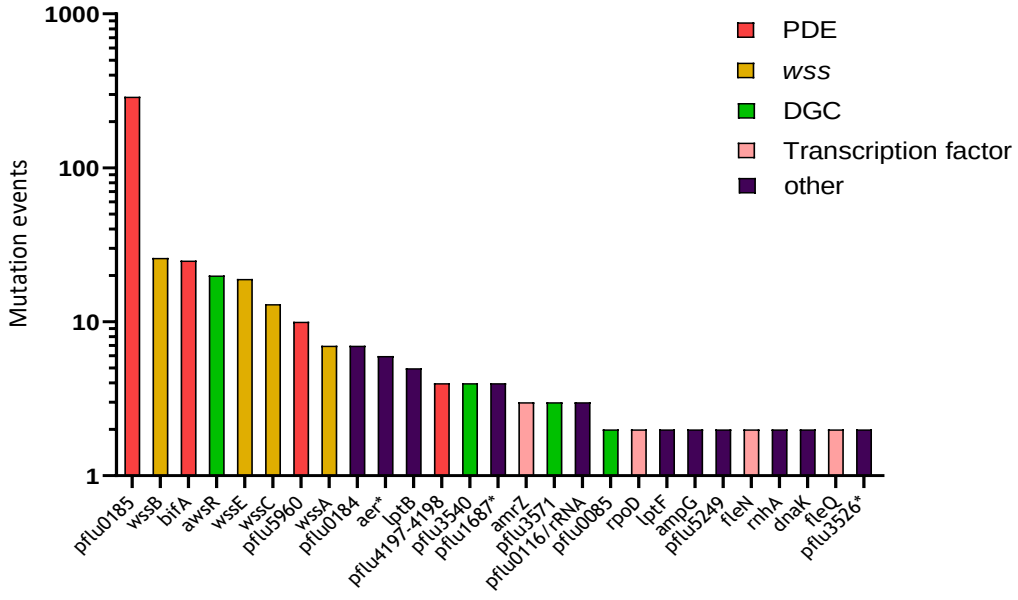

Figure 3.8 Frequency distribution of mutational events for every gene targeted more than once in the line 54 population. An asterisk following the gene names signifies those mutational targets only found in the context of double mutant transitions.

3.2.5.1 Mutational targets and their function: Pflu0185 and other cyclic di-GMP regulators

A single gene, *pflu0185*, alone accounted for 56% (290 of 522) of all mutation events. This extreme enrichment can in part be explained by *pflu0185* harbouring the aforementioned genetic switch, which was alone responsible for 35% (182 of 522) of events. However, *pflu0185* also remained by far the most frequent mutational target even when excluding the genetic switch mutations – a total of 108 other mutation events also targeted this gene. In stark contrast, the next two most targeted genes *wssB* and *bifA*, were targeted only 26 and 25 times, respectively.

Little is known about *pflu0185* – no informative orthologs exist and it has not been encountered in the previous genetic studies of WS. There is nothing overtly unique about Pflu0185 that might explain its

unusual flexibility as a mutational target. It is one of 16 proteins in the SBW25 genome where a DGC domain is fused to a PDE domain, forming a dual-domain DGC-PDE. The majority of proteins containing a PDE in SBW25 are found in this dual-domain configuration, with only five genes containing a PDE domain as a single output. Pflu0185 is predicted ~~to contain two transmembrane helices,~~ indicating it is membrane-embedded, and has an N-terminal PAS sensory domain. Nine of the 15 other dual-domain proteins also contain this same domain architecture.

Since a dual-domain DGC-PDE protein could conceivably function as either a DGC, a PDE, or both, interpreting the effect of a mutation at these loci required resolving this ambiguity. For example, in the context of observing a mutation of a dual-domain protein during a WS→SM transition, is the cause a LoF to DGC activity or GoF to PDE activity? In some cases, this question could be resolved because one of the domains contained a known non-functional active site. For example, BifA in *P. aeruginosa* is known to act exclusively as a PDE and contains the degenerate GGDEF active site sequence GGQEF but has the conserved EAL and DDFGTG motifs necessary for PDE activity (Kuchma et al., 2007). BifA in SBW25 also shares this degenerate active site and the mutations that targeted this locus in the evolution of the line 54 population were consistent with its role as a PDE. Specifically, the very first mutation of *bifA* was responsible for a WS→SM transition – a scenario that can only be interpreted as activating its PDE activity.

As for Pflu0185, it contains canonical active sites for both DGC and PDE activity, however, the pattern of mutation indicated it, like BifA, acts only as a PDE. For example, similar to *bifA*, the first time *pflu0185* was observed to be mutated was during a WS→SM transition. Moreover, unambiguous LoF mutations (e.g., N-terminal frameshifts) to this gene always resulted in WS being turned on.

It is notable that the two most commonly targeted regulatory genes, *pflu0185* and *bifA*, function as PDEs. Indeed, PDEs are more common targets than DGCs even when discounting *pflu0185*. This is surprising, given that there are at least 10 DGCs capable of generating WS outside of the three common DGCs (WspR, AwsR, MwsR), and only four of these were accessed by the line 54 population (Lind et al., 2015). The most frequently targeted DGC in the line 54 population was AwsR. As the WS cell that seeds the line 54 population had previously been shown to be underpinned by AwsR, this was an obvious mutational target for initial phenotypic transitions in this experiment (Barnett, 2016). The two other common DGC targets, WspR and MwsR, had already been activated and subsequently disabled by irreparable deletions during the original LCE, and were therefore not accessible to the line 54 descendants. A likely explanation for the general lack of DGCs is proposed in section **3.2.7 lineage persistence**.

3.2.5.2 Mutational targets and their function: Wss cellulose synthase

Beyond cyclic-di-GMP regulators the next most common target of mutation were genes in the primary structural locus of the WS phenotype, the cellulose-synthase *wss* operon. This large (15.5kb) 10-gene operon accounted for 13% (65) of all mutational events, although only *wssA* (7)*, wssB* (27)*, wssC* (13)*,* and *wssE* (18) were targeted (number of events in brackets). The absence of the last five genes *wssFGHIJ* from this set of targets is expected, as these are involved in cellulose modification via acetylation rather than primary synthesis or export and mutants defective in these genes are still able to form a mat (Spiers et al., 2003).

3.2.5.3 Mutational targets and their function: other components of WS genetic architecture

Beyond the *wss* locus and cyclic-di-GMP regulators, there were very few mutations found in other known components of WS genetic architecture. For example, transcriptional regulators connected to WS (*amrZ*, *fleQ* and *fleN*) together accounted for a total of 8 mutations. Moreover, genes that had been identified as essential for the WS phenotype through previous suppressor analysis (and which therefore represent a potential means of inactivating WS via LoF) were almost entirely absent – no metabolic loci essential for WS were observed and only two mutations were found in loci related to cell envelope biogenesis. These were *fuzW* (*pflu0476*), a gene involved in lipopolysaccharide (LPS) [6] synthesis, and *wswA* (*pflu1661*) the first gene in a cluster of eight known as the wrinkly spreader cell wall biogenesis locus (*wsw*) (Ferguson et al., 2013; McDonald et al., 2009). Surprisingly, the mutations in each of these genes did not inactivate WS as expected but instead were responsible for turning WS on. In both cases, the mutations in *fuzW* and *wswA* were found to be compensating for LoF to one of the cellulose synthase *wss* genes that had underpinned the prior WS→SM transition. This unexpected ability of genes involved in cell envelope biogenesis to compensate for LoF in *wss* will be examined in section **3.2.7.3 Intergenic compensation to Wss loss-of-function**.

As to why many of the loci previously identified by suppressor analysis were not used to inactivate WS via LoF (and then become a potential target for activating mutations of WS in the next phase), this is likely due to the cell-level fitness cost of defects in these genes. For example, one means of inactivating WS is through LoF to *mreB*, which causes the cell to lose its rod-shape and become spherical – as a side-effect its colony also appears SM and it is unable to form a mat. Unsurprisingly, the mutant is also extremely unfit (Yulo et al., 2019). Although an extreme example, it is likely that mutation to other

[6] LPS was introduced in chapter 2 in the context of the Fuzzy Spreader (FS) phenotype, which is caused by LoF to one of the LPS synthesis enzymes (FuzY).

metabolic or cell envelope components also incur costly pleiotropic fitness effects. At minimum, a mutant that interrupts the metabolism of cellulose will continue to incur the energetic and possible pleiotropic costs (such as restricted motility) of cyclic-di-GMP synthesis, while a mutant with an altered cell envelope will incur this and the cost of continued cellulose production. Similar considerations are relevant to those mutants that inactive WS via LoF to the *wss* operon, compared to those that disrupt cyclic-di-GMP regulation directly. Indeed, although a number of mutations targeted *wss,* given the large size of the operon this may contribute to the limited number of *wss* LoF mutations and this may be due to the fitness deficit.

3.2.5.4 Mutational targets and their function: odd loci

Although the majority of mutational targets underpinning the phenotypic transitions had some known connection to the genetic architecture of the WS phenotype, a number of unusual mutational targets were also found to underpin phenotypic transitions. In each case they were the only mutation found and so presumably were responsible for the transition itself. Table 3.1 lists these loci, along with a brief description of the mutational trajectories in which they occurred.

locus	**gene name**	**trajectory**	**function**
pflu3545 & *pflu3544*	*tssL* & *vasK*	PDE inactivated (WS on) → 197bp deletion in *tssL* (WS off) → 13 bp deletion in *vasK* (WS on) → extinct	Type 6 secretion systems
pflu5592	*dnaK*	Pflu0185 PDE inactivated (WS on) → *dnaK* Q497K + A372V (WS off) → *dnaK* R34L (WS on) → *rpoD* T482P (WS off) → pflu0185 PDE inactivated (WS on) →…	Molecular chaperone
pflu2645	*rnhA*	pflu0185 PDE activated (WS off) → *rnhA* V121G (WS on) → *wssB* inactivated (WS off) → extinct	Ribonuclease
		pflu0185 PDE activated (WS off) → *rnhA* G140D (WS on) →pflu0185 PDE activated (WS off) →…	
pflu0883, pflu0884 pflu1056, pflu1057	*lptB, lptH, lptF, lptG*	Varied	Lipopolysaccharide transport system

Table 3.1 description of loci with no previously known connection to WS detected as mediating WS on-off transitions

Of particular note here are mutations in the *lpt* genes, which are components of a complex encoding the lipopolysaccharide transport system. Intriguingly, mutations in *lpt* were found to compensate for LoF mutations of the same *wss* gene as those compensatory mutations in *fuzW* and *wswA* mentioned in the preceding section. As with these two genes, *lpt* also functions in cell envelope biogenesis (see further discussion in section **3.2.7.3 Intergenic compensation to Wss loss-of-function**).

3.2.5.5 Mutational targets and their function: double and triple mutant(s)

In the case of the 47 phenotypic transitions that were associated with two mutations, it was generally the case that one of these was the obvious cause of the transition, while the second mutation targeted a gene that had no known connection to the genetic architecture of WS. It is in this context where many of the genes mutated only once throughout the experiment were observed. Undoubtedly, many of these have neutral fitness effects and are the result of genetic hitchhiking with the primary transition-mediating mutation. This is almost certainly the case for the six mutations that resulted in synonymous substitutions. A table of all loci that occurred exclusively in the context of a double mutant transition is presented below (Table 3.2).

locus	gene name	mutation events	function
pflu5729	*dguC*	1	D-amino acid homeostasis
*pflu1372**		1	hypothetical protein - polysaccharide deacetylase family
pflu4184	*cvpA*	1	colicin production membrane protein
*pflu1593**		1	hypothetical protein
pflu2227		1	putative homocysteine S-methyltransferase
pflu2188-2189	*-gacA*	1	transcription factors
pflu0637	*bfiS*	1	sensor kinase
pflu4943		1	hypothetical membrane protein
pflu1654		1	hypothetical protein; Glycosyl transferase family 2 domain
pflu1877/pflu1878		1	Intergenic -165/-748
*pflu1815**	*gltA*	1	type II citrate synthase
pflu0491		1	outer membrane TolC efflux protein
pflu4413		1	chemotaxis-specific methylesterase
pflu4395	*atuD*	1	putative acyl-CoA dehydrogenase
pflu4551	*aer*	6	aerotaxis receptor Aer
pflu1657	*fnl1*	1	polysaccharide biosynthesis protein
pflu4593	*xdhB*	1	xanthine dehydrogenase
pflu4190	*fimV*	1	Motility regulator (flagella & type IV pili)
*pflu5416**	*lipA*	1	lipoyl synthase
pflu2181	*rocS1*	1	two-component system sensor kinase
pflu1687		4	methyl-accepting chemotaxis protein
pflu1479		1	hydrolase; metallo-β-lactamase
pflu3409		1	methyl-accepting chemotaxis protein
*pflu1728**	*tpm*	1	thiopurine S-methyltransferase
*pflu5739**	*mdcE*	1	malonate decarboxylase subunit gamma

Table 3.2 Table of all loci found exclusively in the context of a double mutant. Synonymous mutation targets are marked with an asterisk. The specific genes discussed in the main text have been highlighted.

Although most of the genes in table 3.2 were mutated only once throughout the experiment, two of these genes, *pflu4551* and *pflu1687,* were targeted multiple times. The function of these genes is also very intriguing – *pflu4551* is an ortholog of the aerotaxis receptor *aer* in *P. aeruginosa*, and *pflu1687* is a predicted methyl-accepting chemotaxis protein. Both genes are therefore involved in

environmental sensing and cell motility. While many of the other secondary mutations could be the result of hitchhiking, the repeated targeting of these two loci indicate that these secondary mutations are themselves adaptive. Moreover, it was noticed that in every case, these mutations occurred during WS→SM transitions mediated by the *pflu0185* genetic switch as the primary mutation, suggesting the switch may be facilitating the secondary adaptive mutations. Furthermore, it was found that the only triple mutant detected in the entire history of the population also occurred during a WS→SM transition and involved the genetic switch alongside a mutation in *pflu4551* (the *aer* ortholog) and an additional mutation in *pflu4190* – an ortholog of the motility regulator *fimV* in *P. aeruginosa* (Semmler et al., 2000; Buensuceso et al., 2016). This implicated both these secondary mutations as adaptive and affecting cell motility. One way of understanding this pattern of secondary mutations is that if the genetic switch has increased the speed that the target phenotype is being generated, then additional adaptive steps become possible that optimize cell fitness with respect to other aspects of the environment, in this case through apparent changes in chemotaxis and motility (see further discussion in section **3.2.9 Secondary adaptive mutations**).

3.2.6 A map of subsequent sections

The remainder of the results will be broken into three main sections. The first of these, section **3.2.7 Lineage persistence**, examines how lineages navigated the existing genetic paths to achieve phenotypic transitions and, more specifically, how the idiosyncratic path a given lineage took influenced its capacity to survive future transitions. This will involve identifying the genetic precursors of extinction, illustrations of the whittling effects of regulatory targets in line with the null expectation described in the introduction, and impressive cases of compensation in response to LoF in one of the structural *wss* genes.

Section **3.2.8 Lineage adaptation** will address the origin and mechanism of the genetic switch in *pflu0185*. I will also speculate as to why *pflu0185* was so frequently targeted by mutation outside of those composing the switch, and whether and how the general flexibility of this gene for on-off transitions might have contributed to the evolution of the genetic switch.

Finally, section **3.2.9 Secondary adaptive mutations** will assess the evidence that lineages possessing the genetic switch experience more rapid evolution within the microcosm, leading to additional (second and third) adaptive steps that optimize aspects of cell fitness unrelated to the primary phenotypic transition.

3.2.7 Lineage persistence

As detailed in the introduction, there exist a variety of different paths by which lineages can modulate the WS phenotype, and the unique path a lineage takes determines the possible avenues for subsequent transitions – leading to differential lineage persistence. The paths used by the line 54 population have already been outlined: lineages traversed a small subset of possible WS-modulating loci, primarily cyclic-di-GMP regulatory genes and the cellulose synthase-encoding *wss* operon, and in particular a single PDE-encoding gene, *pflu0185*.

To get a broad picture of how the different paths were used throughout time, the loci that mediated each transition can be mapped to the lineage tree. Figure 3.8 below depicts the lineage tree colour-coded with the function of the loci that were found to mediate each transition, as well as those transitions that were mediated by the genetic switch in *pflu0185*.

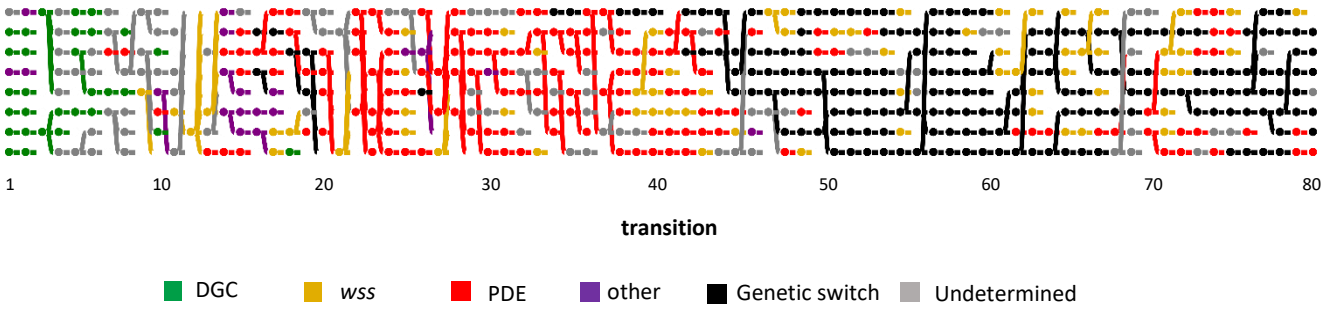

Figure 3.8. Lineage tree colour-coded by the function of mutational targets. Mutations composing the genetic switch in *pflu0185* are shown in black, while all other mutations in *pflu0185* are included in the PDE (red) category.

At first, lineages in the population primarily transition via DGCs – this was expected, at least initially, as cellulose production of the WS cell the seeds this population was underpinned by the DGC AwsR. Despite the initial targeting of DGCs, this quickly gives way to alternative gene targets; PDEs (primarily *pflu0185*) come to dominate. In fact, the last DGC is activated at transition 12 – following this not a single DGC is ever activated again. The change from DGCs to PDEs opens a whole new set of targets, and WS on-off cycles driven previously by DGC on-off mutations, are now driven by PDE off-on cycles: activating mutations in PDEs counteract a previously established source of cyclic-di-GMP resulting in WS→SM transitions, while LoF mutations in PDEs increase intracellular cyclic-di-GMP levels resulting in SM→WS transitions.

The rapid change from targeting DGCs to PDEs is surprising, particularly since lineages had not exhausted all possible DGCs. Many known DGC paths remain intact in the lineages following transition 12 yet are never accessed. One possible explanation is that the source of cyclic-di-GMP in these

lineages is no longer localized to any single gene but is instead distributed across multiple loci – therefore LoF mutations to a single one of these DGC cannot turn WS off, while activation of a PDE can. Alternatively, the source of cyclic-di-GMP could be encoded by a pleiotropic gene whose LoF comes with associated fitness deficit. Whatever the case, the key to this change appears to be a particular WS-activating path involving two mutations: AmrZ + DipA, which fixed in the population at transition 14. Unlike the more common mutational paths that activate WS, this path to activate WS does not lead directly to activation of a specific DGC and instead involves a LoF to the negative regulator of *wss* transcription AmrZ and LoF to the PDE DipA (Lind et al., 2015). While the fixation of this lineage may explain the predominance of PDEs in the line 54 population, it does not explain why mutations occurred almost exclusively in a single PDE (Pflu0185).

The specific mechanisms of lineage persistence, including the whittling of regulatory targets and the arrival of lineages at apparent genetic dead-ends, can be studied at fine resolution by zooming in on specific branches of the lineage tree. Two illustrative examples, offspring trajectories A and B that descend from the same AwsR LoF parent mutation that was responsible for a WS→SM transition are shown in Figure 3.9. To maintain focus on the general characteristics of lineage persistence and the causes of extinction in the main text, detailed step-wise descriptions of these subtrees, illustrating the logic by which all the mutations within the tree were able to be interpreted and assigned function, have been relegated to the legends of Figures 3.10 and 3.11.

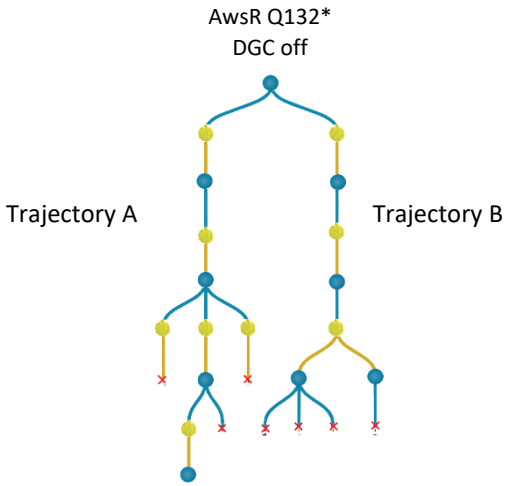

Figure 3.9. Lineage sub-tree extracted two trajectories following a LoF mutation to AwsR. The initial LoF is a premature stop-codon (Q132*) that happened to be randomly selected for a reproduction event at transition 5 of the experiment. The offspring lineages I will refer to as trajectory A and B. A red X denotes a lineage extinction event.

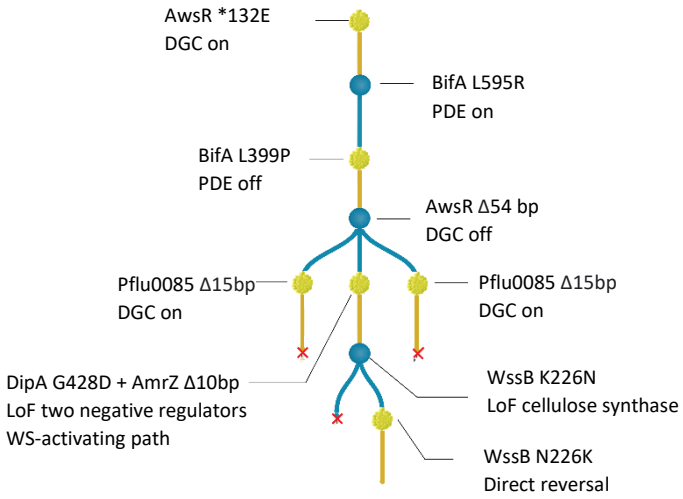

Figure 3.10. Trajectory A: the evolutionary dynamics of a lineage following LoF to the DGC AwsR. This trajectory begins with a reversal of the Q132* premature stop codon in AwsR, although to a different amino acid than the original (*132E). AwsR is therefore active once again: the WS phenotype is generated. The lineage could conceivably then inactivate WS again via LoF to AwsR; however, instead a new gene is mutated: the dual-domain DGC-PDE *bifA* (*pflu4858*). Although this gene encodes both a DGC and PDE, in this context the mutation can only be interpreted as activating PDE activity, which in turn leads to degradation of cyclic-di-GMP produced by AwsR and loss of the WS phenotype. The next phenotypic transition is via a LoF to the now-active BifA, inactivating its PDE activity and restoring WS. Following this, a 54bp deletion in AwsR inactivates WS and irreparably destroys this DGC for future cycles. As it happens the lineage at this point is randomly selected for two reproduction events, and this effectively creates a replicated experiment to examine the consequences of AwsR becoming inaccessible for future evolution in these lineages. The expectation from the work of Lind et al., (2015) is that one of the rarer 'hidden' paths to WS that are only taken when the common DGC paths (WspR, AwsR and MwsR) are inaccessible will now be traversed. In line with this, two of the offspring traverse parallel mutational paths via a 15bp deletion within *pflu0085*, the DGC identified as the most common of the hidden paths to WS, by virtue of its intragenic negative regulatory architecture and associated large target size. The remaining offspring takes a unique path – one of the rarest of the hidden paths uncovered by Lind et al., (2015) involving two extragenic negative regulators: DipA and AmrZ. The two lineages that generated WS via *pflu0085* go extinct through failure to generate the SM phenotype. As there is always a mutational target to turn WS off (via LoF to the *wss* cellulose synthase operon, for example), this cannot be due to lacking an appropriate target. As for the lineage that proceeded via the double DipA + AmrZ mutations, this becomes the ancestor of all surviving lineages of the line 54 population. As described in the main text, it is also the last time a WS activating path is used – subsequent transitions in this lineage are mediated solely by PDEs or structural loci.

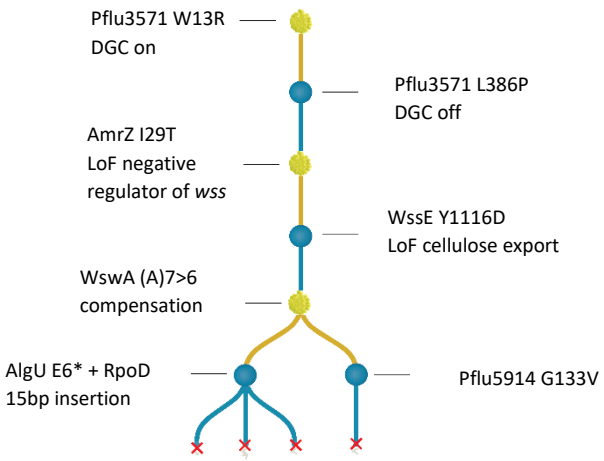

Figure 3.11. Trajectory B: the evolutionary dynamics of a lineage following a LoF to the DGC AwsR. In contrast to trajectory A, the stop codon in AwsR is not reversed and instead one of the rare paths to WS, the DGC Pflu3571, is mutated to generate WS. The next step sees a LoF to the previously activated DGC Pflu3571, hence inactivating WS, which is then activated again via mutation to AmrZ, a negative regulator of *wss* transcription. At the same time, a second 1 bp deletion causing a frameshift also occurs in *pflu4184*, a gene predicted to encode colicin V production protein, a small peptide antibiotic produced by *E.coli* (Zhang et al., 2005). Because there is no obvious functional connection to the WS phenotype and a similar AmrZ mutation (M25A) is found by itself to cause WS in a different lineage in the population, this second mutation is likely the result of genetic hitchhiking. The next step then inactivates WS via LoF to the *wssE,* a gene necessary for cellulose export. The following mutation to re-activate WS does so by apparently compensating for *wssE* mutation via a frameshift in *wswA*, a gene identified in the early suppressor analysis of WS with a predicted role in cell-wall biogenesis (McDonald et al., 2009) . Ordinarily, a LoF mutation to *wswA* is expected to result in WS suppression. Following compensation via the mutation in *wswA*, the lineage reproduces twice. One of the offspring mutates Pflu5914, a predicted membrane protein that has no known connection to the WS phenotype and may therefore be conditional on taking the *wswA* path. The other offspring inactivates WS via an N-terminal premature stop codon in AlgU (E6*), a known suppression locus of WS and therefore expected target of WS→SM transition via LoF (McDonald et al., 2009). A second mutation occurs alongside AlgU, a possible LoF in the house-keeping sigma factor RpoD. The relevance of the RpoD mutation is difficult to interpret given its number of potential effects but, as AlgU is also a sigma factor, it is likely to be functionally relevant and not merely the result of hitchhiking. Following the AlgU + RpoD mutation, the lineage is then randomly selected for two reproduction events and all three identical offspring immediately go extinct through failure to generate WS, a shared fate suggesting the lineage has found itself at a genetic dead-end.

3.2.7.1 The genetic causes of extinction

Lineage persistence means avoiding extinction. As outlined in the introduction, extinction was expected to result from irreparable LoF mutations that rendered certain paths inaccessible for future phenotypic transitions. This could occur both gradually through the whittling away of potential

regulatory targets or through sudden genetic dead-ends that result from irreparable LoF to a gene essential for the WS phenotype. The process of whittling of regulatory paths is difficult to resolve quantitatively due its slight and cumulative nature, although it is possible to observe specific instances within the lineage tree. Of particular note is a sub-tree (Figure 3.12) extracted from late in the evolution of the line 54 population (transition 71 to 78). During this period, the *pflu0185* genetic switch had become by far the most frequented path for WS on-off cycles. Indeed, the trajectory depicted in Figure 3.12 was preceded via two transitions mediated by the switch. However, following these two uses of the switch, the parent of this trajectory transitioned to WS through the irreversible inactivation of Pflu0185 via an 18bp deletion. As a consequence of this whittling step, offspring lineages lose access to the most flexible gene in the genome, as well as the genetic switch it harbours. The offspring persist for some time following this, in part due to a fortuitous lineage reproduction event. However, all descendant lineages succumb to extinction due to a failure to generate WS shortly after. Their fate is sealed by the loss of *pflu0185*, the principal remaining path to WS.

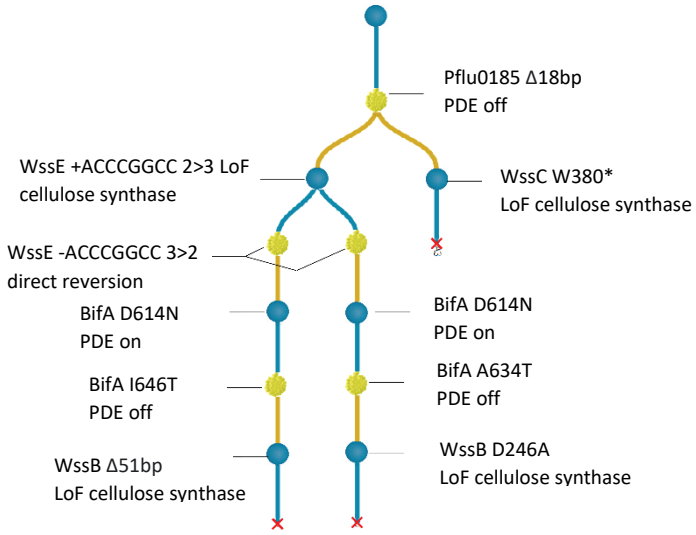

Figure 3.12 Irreparable LoF to the main target of mutation in the experiment (*pflu0185*) leads to lineage extinction.

As to sudden extinctions through irreparable LoF mutations, there was a clear signal of this process occurring at the *wss* locus: 68% (38/56) of extinctions during the SM→WS phase were preceded by parent mutations in the *wss* operon compared to a frequency of 7% (15/226) for lineages that survived the next transition ($P<0.0001$, Fisher's exact test). A lineage was therefore approximately 10 times more likely to go extinct if inheriting a mutation within *wss* compared to any other mutational target.

Beyond *wss*, a small set of structural genes and transcriptional regulators (listed in Table 3.3), when taken as a group, were also found to disproportionately precede extinction: 18% (10/56) of extinctions during SM→WS were preceded by mutations in these loci, compared to a frequency of 5% (12/226) for those lineages that survived the SM→WS transition (P=0.005, Fisher's exact test, two-sided).

Gene	Mutation	Function
pflu5914 (x2)	G113V	Hypothetical membrane protein
algU and *rpoD* (x3)	E6* and 15bp insertion	Sigma factors
pflu4993 ampG	(TGC)4→3	Cell wall recycling
pflu4993 ampG	(TGC)4→0	Cell wall recycling
pflu1476 gcvR	Q30*	Negative regulator of glycine cleavage system transcription
Pflu4418 fleN	L35P	Flagella and exopolysaccharide transcriptional regulator
pflu0883 lptB	G236S	Lipopolysaccharide transport system

Table 3.3 Genes other than *wss* in which mutations were found to precede extinction at the SM→WS transition

Why these particular structural genes and transcriptional regulators listed in Table 3.3 would inhibit or prevent generation of the WS phenotype was not immediately obvious, however a close inspection of the historical trajectory from which they arose reveals that, with the exception of *fleN*, each stem from a mutation in *wss* that was then compensated via a mutation outside of the *wss* operon. For example, retracing the history of both *pflu5914* and the *algU* + *rpoD* double mutant, shows that each evolved from a lineage that compensated for *wssE* disruption via mutation in the scaffold gene *pflu1661* (this trajectory was depicted earlier in Figure 3.11).

Therefore, the majority of extinctions occurring at the SW→WS transition can be linked to *wss* LoF mutations from the immediately preceding WS→SM transition, or that occurred recently in the history of that lineage. As for the alternate WS→SM transition, no functional category or particular mutational target was shown to be associated with extinction during this phase. This was not entirely unexpected given that an inability to generate SM cannot be due to the absence of an appropriate mutational target as in the SM->WS transition – there is always a way to inactivate WS via LoF to one of the WS-essential genes such as those in the *wss* operon. This suggests extinctions that occurred here must have been largely stochastic – driven by lineage 'drift' that could arise both through stochasticity from the dynamics of mutation and individual-level selection within the microcosm, as well as through variation introduced in the colony plating process. Despite there being no association between mutational targets and extinction at the WS→SM transition, as will be described in the later section

3.2.8 Lineage adaptation, there was however a significant association with the specific mutation of the *pflu0185* genetic switch and lineage survival.

3.2.7.2 Repairing LoF mutations in the *wss* operon

Although risky, not all mutations in *wss* are necessarily dead-ends. They can be repaired, through either reversal or compensatory mutation at another site, although the likelihood of either of these possibilities will depend on the manner in which *wss* happened to be inactivated. For example, large deletions are unable to be reversed and are presumably not easily compensated for – too much genetic information has been lost in such an event. On the other hand, some LoF may be quite likely to be reversed or compensated, and so permit lineage survival. Examining the subset of *wss* LoF mutations that were followed by extinction with those that were followed by survival provides a window to observe the biases and constraints that particular inactivating mutations might have with respect to their likelihood of being repaired in the following SM→WS phase. Those LoF *wss* mutations that were able to be repaired and the means by which this occurred are listed in Table 3.4.

***wss* LoF mutation**	**repair mutation**
wssE Y1116D TAC→GAC	*pflu1661 wswA* A7→A6
wssB K226N AAG→AAT	direct reversion N226K
wssE 9bp deletion	*pflu1056 lptF* P306S CCC→TCC
	pflu1057 lptG V331E GTG→GAG
	pflu1056 lptF P306S CCC→TCC
	pflu0884 lptH T74P ACC→CCC
wssB Q271* CAG→TAG	*271S TAG→TCG
wssA 1bp deletion	2bp deletion re-framing
wssB Q627* CAG→TAG	*627W TAG→TGG
wssE G7→G8 +G	direct reversion -G
wssE A779E GCG→GAG	*pflu0476* +T (470/897 nt)
wssB W383* TGG→TGA	direct reversion *383W
	direct reversion *383W
wssC 1 bp deletion	11 bp deletion re-framing
wssE G7>G6 -G	direct reversion +G
wssB -TCT 3→2	direct reversion +TCT
wssB -TCT 3→2	direct reversion +TCT
wssC +TGCTGGTCAACG 1→2	direct reversion - TGCTGGTCAACG
wssE +ACCCGGCC 2→3	direct reversion - ACCCGGCC
	direct reversion - ACCCGGCC

Table 3.4 *wss* LoF mutations and how they were repaired

The first thing to note about this list is that there is not a single example of compensation at any site within *wss* other than the exact residue(s) affected by the inactivating mutation. That is, repair comes about exclusively through two means: (i) direct or near-direct reversions (the same residue substituted to an amino acid other than the original) or (ii) mutations at entirely different genes outside of the *wss* operon (as in *wswA, fuzW* and the *lpt* genes) and never through different sites within *wss* itself. This is in stark contrast to the pattern of compensation observed to occur within some regulatory loci, particularly those flexible PDEs like *pflu0185*, and indicates Wss has a tightly constrained structure-function relationship.

As expected, large deletions do not appear in the list of repaired *wss* mutation. The spectrum of repaired mutations also immediately suggests a biasing role for repetitive sequences, with 6 of the 15 Wss-inactivating mutations being associated with some kind of repeat motif. All of these were repaired by direct reversions at the next phase of the cycle. Indeed, repeat-associated indels were significantly enriched over indels not associated with repeats when comparing the surviving lineages (7 of 9) to those lineages that went extinct (5 of 22) (P=0.012, Fisher's exact test, two-tailed). The bias of repetitive sequences for expansion and contraction is consistent with the well-known elevated mutation rate due to slipped-strand mispairing at repetitive sequences (Levinson & Gutman, 1987).

3.2.7.3 Intergenic compensation to Wss loss-of-function

The most interesting examples of compensation to Wss LoF come about via intergenic mutations. This was a rare occurrence, with only six examples of such events throughout the history of the line 54 population. Interestingly, all occurred in response to mutations of single protein in the Wss complex, WssE (Table 3.5).

***wss* LoF mutation**	**repair mutation**
wssE Y1116D TAC>GAC	*pflu1661 wswA* A7>A6
wssE 9bp deletion	*pflu1056 lptF* P306S
	pflu1057 lptG V331E
	pflu1056 lptF P306S
	pflu0884 lptH T74P
wssE A779E GCG→GAG	*pflu0476 fuzW* +T

Table 3.5 *wss* LoF mutation that were repaired intergenically.

Two of the repair loci, *wswA* and *fuzW*, represent genes identified through early suppressor analysis of WS and as such were expected to be potential targets of WS inactivation and not, as happened to be the case here, a target of WS activation. Both genes are predicted to be involved in aspects of biogenesis of the cell envelope – *fuzW* in LPS synthesis, and *wswA* in cell wall synthesis (Ferguson et

al., 2013; McDonald et al., 2009). Compensatory mutations in these targets each occurred in response to an inactivating substitution in *wssE* and themselves result in frameshifts that presumably cause a LoF, although mutation in *wswA* occurs toward the end of this gene and so may retain some function. The three other examples of intergenic compensation involve three genes (*pflu1056*, *pflu1057* and *pflu0884*) that are each orthologs of components of the lipopolysaccharide transport (*lpt*) system. This system shuttles LPS molecules from the inner-membrane to the outside of the cell, where they form an essential component of the cell's outer-membrane. Therefore, all targets of compensatory mutation to WssE LoF are involved in cell envelope biogenesis.

Compensatory mutations in *lpt* are particularly interesting as none of these genes have previously been associated with the WS phenotype. Moreover, while the lineages that compensated via *fuzW* and *wswA* happened to go extinct either immediately or after one further transition, many *lpt*-traversing lineages persisted and, in a number of cases, traversed phenotypic transitions through mutations in other components of the *lpt* gene complex, as well as other novel mutational targets that appear conditional on this trajectory (Figure 3.13). From an evolvability perspective, the lineages compensating via *lpt* have expanded the target size to WS through the creation of a new epistatic interaction between *wss* and *lpt*.

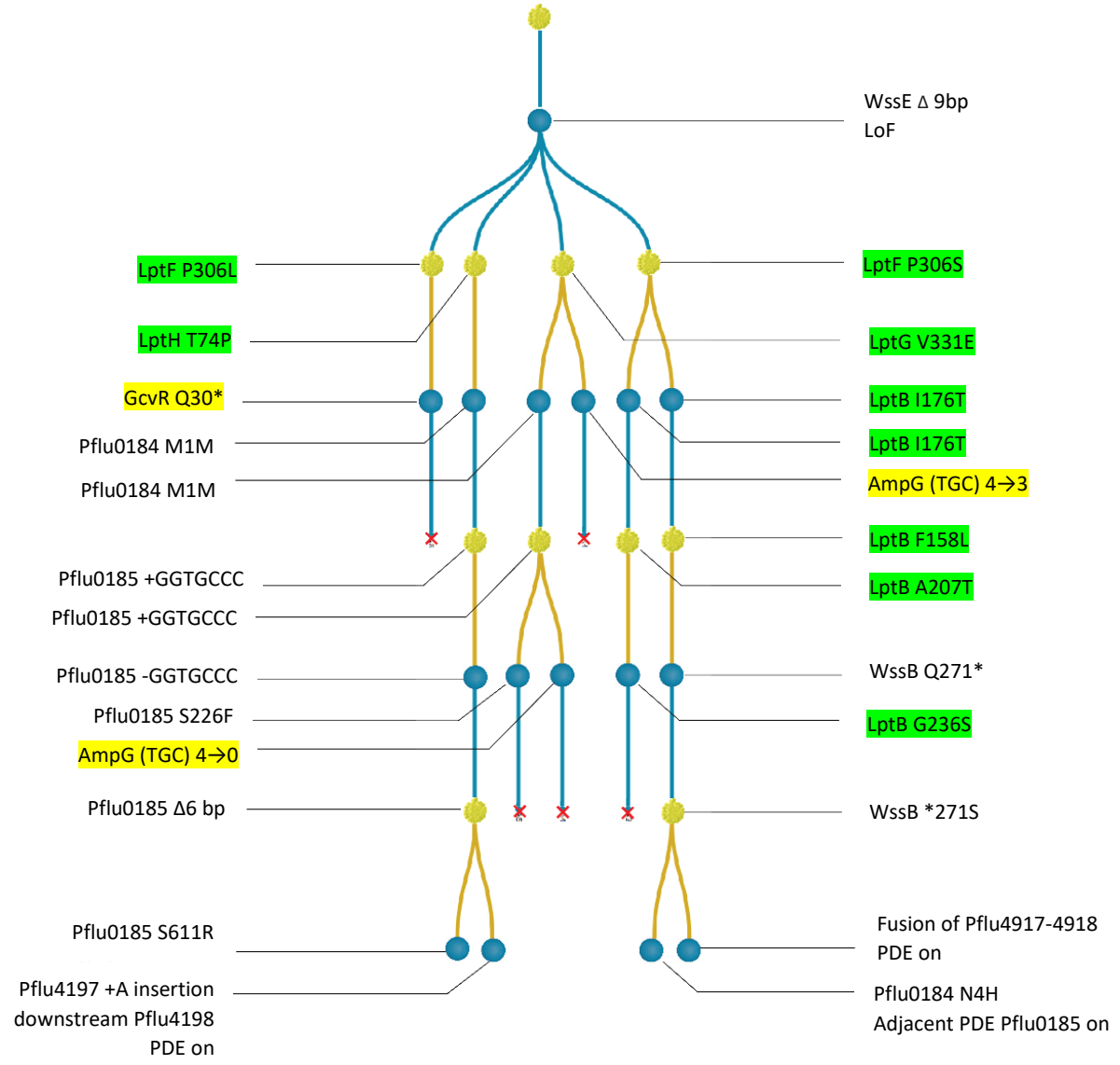

Figure 3.13 Mutations in *lpt* compensate for LoF mutations in *wssE* and open up a novel set of mutational targets to modulate WS. Beginning with the 9 bp deletion in *wssE* the lineage is selected for three reproduction events, leading to the establishment of four replicate lineages. Each re-activates WS via substitutions in *lpt* genes. Some offspring lineages fall extinct, while others eventually return to the cyclic-di-GMP regulators, indicating the potential genetic dead-end of the WssE mutation had been successfully circumvented. *lpt* mutations are highlighted in green. Genes outside of *lpt* but which appear contingent on the *lpt* path are highlighted in yellow. These include *pflu4993*, which is homologous to *ampG*, a permease essential for cell-wall recycling and with a connection to biofilm formation in *E. coli* and *gcvR*, a negative transcriptional regulator of the glycine cleavage system (Mallik et al., 2018). Note that the entire branch depicted here eventually becomes extinct, presumably for reasons unconnected to its traversing of *lpt*, as the lineages which survive the repair return to normal cyclic-di-GMP target, indicating no genetic dead-end was imposed.

Why intergenic compensation to Wss inactivation happened to occur only in response to mutations in this single protein of the Wss cellulose synthase and, further, why compensation comes about exclusively through components involved in cell envelope biogenesis, is more than curious. The first thing to consider in understanding this pattern is the function of WssE.

3.2.7.3.1 WssE function

Although WssE itself has not been studied in *Pseudomonas*, its function can be inferred from a number of well-studied orthologs, which are known by the common designation BcsC (Bacterial cellulose synthase C). Here, BcsC has been identified as a porin – an outer-membrane (OM) protein responsible for channelling nascent cellulose (b-D-glucan chains) synthesized at the inner membrane (IM) by the cellulose synthases (WssB/C) to the outer-cell, where the glucans are formed into cellulose fibres (Acheson et al., 2019). The process is sketched in figure 3.14.

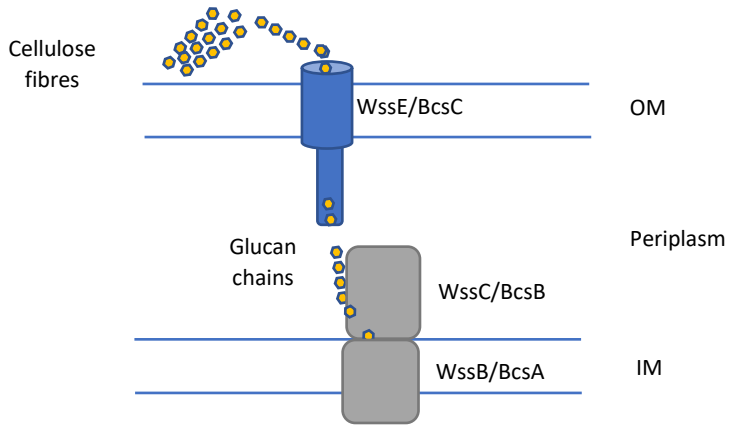

Figure 3.14 Production of cellulose via the Wss cellulose synthase complex. Two synthase proteins, WssB and WssC, embedded in the inner-membrane (IM) catalyse the production of glucose chains (upon allosteric activation via cyclic-di-GMP), which are then passed through the outer-membrane (OM) WssE porin and form cellulose fibres.

The porin BcsC contains an N-terminal domain rich in tetratricopeptide repeats (a common protein interaction module) and a C-terminal β-barrel domain. β-barrels are OM-embedded proteins that form both specific and non-specific channels across the OM through which small molecules are able to pass both in and out of the cell, and it is this domain that identifies BcsC as a porin. Since the OM (unlike the IM) is not energized by ATP (which is absent in the periplasm) any molecule exiting the cell through a porin must move by passive diffusion along concentration gradients or be guided by a periplasmic intermediary that either facilitates diffusion or is energetically coupled to the IM (Braun, 2006). A

particular challenge for these systems is to prevent aggregation, this being especially problematic for amphipathic molecules (containing both hydrophobic and hydrophilic moieties) like cellulose (McNamara et al., 2015). Although the precise details of transport across the periplasm and OM have not been determined for the Bcs complex, it is thought the tetratricopeptide repeat domain of BcsC acts as the intermediary with the IM-embedded cellulose synthase complex and facilitates diffusion of the growing glucan chain through the porin (Acheson et al., 2019).

LoF mutations of WssE in SBW25 are therefore expected to inhibit the transport of cellulose outside of the cell. This presumably leads to the build-up of glucan chains within the periplasm, or potentially interrupts synthase activity if this is somehow coupled to directly to cellulose export. Whether synthesis is coupled to export is apparently unknown, although at least *in vitro* the synthesis of cellulose does not rely on BcsC/WssE (requiring only the synthase components and cyclic-di-GMP) (Omadjela et al., 2013).

As to how the particular LoF mutations in WssE were affecting its function – all three of the compensated WssE mutations occur toward the C terminus, with two residing in the β-barrel domain itself (Figure 3.15). In contrast, WssE mutations that were not compensated for (and so led to extinction) were found to be spread across the gene, with four out of five of these causing frameshifts upstream of the β-barrel domain, indicating they result in complete loss of the channel for cellulose to exit the cell. The different spectra of compensated and uncompensated mutations suggested a complete LoF of the WssE porin is unlikely to be compensated for, although of relevance here is that all mutations that failed to be compensated only had a single chance to repair the mutation i.e., they did not happen to experience a lineage reproduction event following their occurrence. In contrast, two of the three WssE mutations that were compensated experienced occurred in the context of reproduction events.

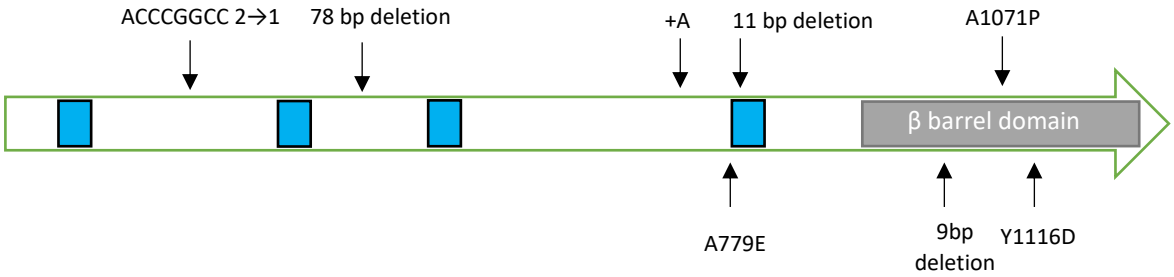

Figure 3.15 Location of WssE mutations that are compensated intergenically. Mutations annotated below the gene are those which were compensated; above, those that were not compensated. Blue boxes indicate tetratricopeptide repeats.

3.2.7.3.2 The 'scaffold-adjustment' hypothesis to explain intergenic WssE compensation

Given that the genes compensating for WssE are all involved in cell envelope biogenesis, and the mutations in WssE that are being compensated maintain the β-barrel domain intact to some degree, it appears that compensation may be occurring via modification of the cell-envelope within which WssE is embedded (figure 3.16). I will refer to this as the scaffold-adjustment hypothesis. The closest example of this phenomenon I could find in the literature was a study by Blanka et al. (2015) who demonstrated that a mutation that causes an alteration of membrane composition can lead to changes in the activity of membrane-embedded sensory proteins, specifically leading to their constitutive activation.

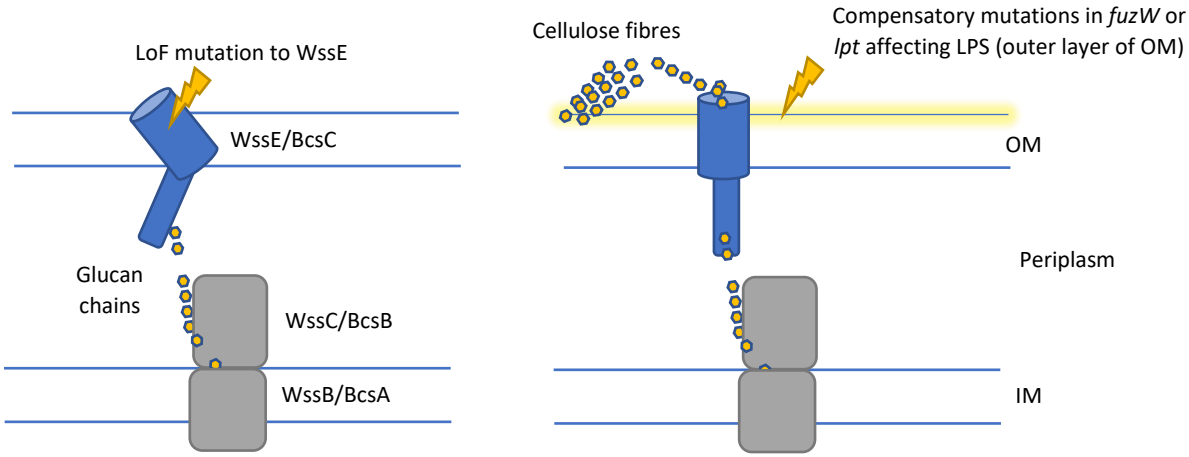

Figure 3.16 The scaffold-adjustment hypothesis for compensation of WssE LoF. A mutation (lightning bolt) causes a LoF to WssE, preventing export of cellulose and mediating a WS→SM transition (left panel). Compensation comes about through mutations that modify some aspect of the cell envelope – pictured here as occurring through modification of the LPS that composes the outer layer of the cell (right panel).

3.2.7.3.3 An alternative hypothesis for WssE compensation via co-option of the Lpt porin

The *lpt* locus is particularly well-characterized, having been closely studied in *E. coli* (reviewed in May et al., 2015) and *P. aeruginosa* (Bollati et al., 2015; Scala et al., 2020). This enables the proposal of an alternative hypothesis for compensation specific to *lpt* mutations and a potential means of testing it.

Like WssE, the Lpt system is responsible for the transport of an amphipathic molecule to the cell-surface. In the case of Lpt, that molecule is the LPS that forms an essential part of the outer-leaflet of the OM. Figure 3.17 below sketches Lpt and Wss complexes for comparison.

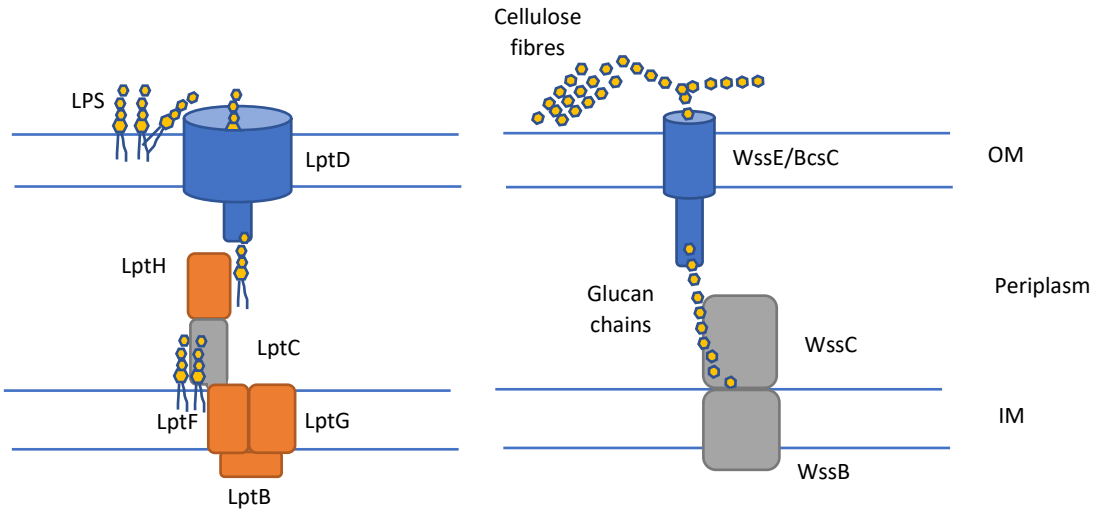

Figure 3.17 Comparison of Lpt and Wss complexes. Orange Lpt proteins indicate those that were targeted by mutation.

The similar functions of Lpt and WssE raises the possibility that the β-barrel porin of the Lpt system, LptD, has been co-opted by Wss to function as the porin for cellulose. As in WssE, the LptD porin contains an N-terminal periplasmic domain that mediates interaction with IM-embedded enzymes and a C-terminal β-barrel domain, in this case encoding a 26-strand β-barrel (Botos et al., 2016) much larger than 16-strand structure of WssE (Acheson et al., 2019). Considering the size of the porin only, LptD could presumably accommodate the nascent cellulose chains that are trapped in the periplasm.

If it were the case that LptD had been co-opted to export cellulose, it would need to continue to secrete LPS as this function is essential to the production of viable gram-negative cells. Compensation would therefore have constructed a new gene complex and an additional function for Lpt, which is now pleiotropic for both LPS and cellulose secretion. A simple means of testing the co-option hypothesis would be to delete the entire WssE β-barrel to determine whether the WssE porin is indeed required for WS expression in these mutants.

3.2.8 Lineage adaptation

Despite the impressive capacity of lineages to repeatedly cycle WS on and off via an array of distinct regulators as well as the ability to compensate and repair LoF mutations to *wss*, there are limits to persistence. Lineages were ultimately destined for extinction unless they could establish a long-term solution to the problem of repeatedly switching between the two phenotypes – long-term survival required lineage adaptation.

The present section examines the evidence for, and the origins and mechanism of, the genetic switch residing in *pflu0185* – a modifying trait that is thought to have been selected at the lineage level through its increased likelihood of generating the target phenotypes. The arrival and expansion of lineages tending to use this switch can be seen in Figure 3.18 below, which depicts both the coalescence of the surviving lineages (upper panel) as well as all transitions that were mediated by the switch (lower panel).

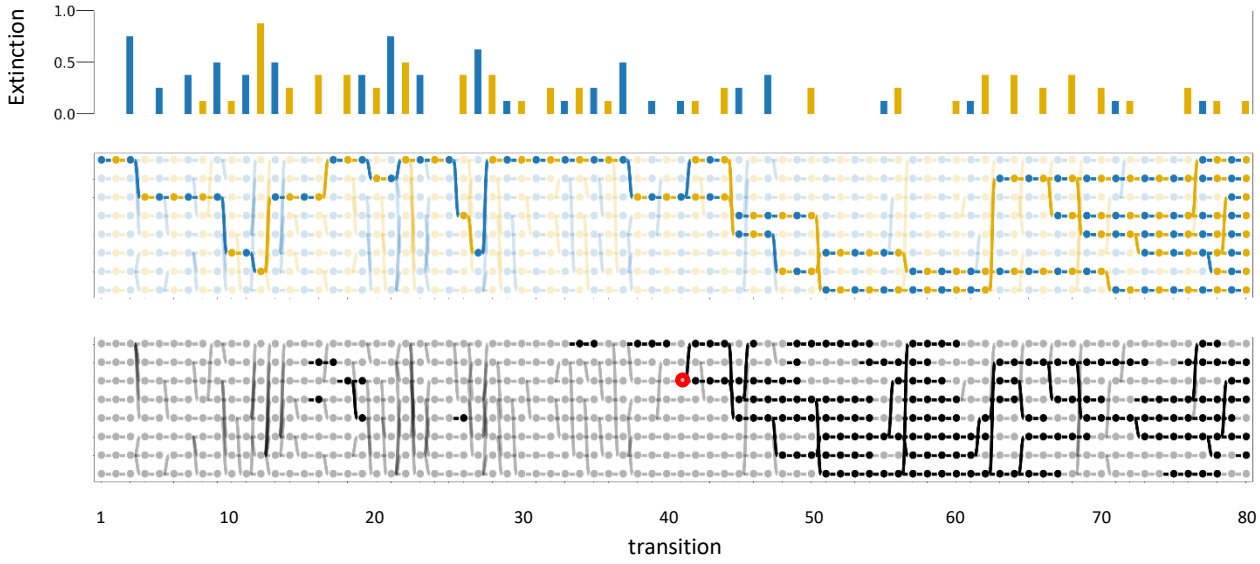

Figure 3.18 Lineage trees comparing coalescence and transitions mediated by the genetic switch. There is a clear association between the pattern of coalescence and use of the switch evident in this figure. Indeed, the point of coalescence of the surviving lineages (indicated by the red circle) coincides precisely with the beginning of the (semi-) stable use of the genetic switch. This point also corresponds with a decrease in the frequency of extinctions during the WS→SM transition (blue bars in top panel). The switch is not perfect however – a number of times a lineage that repeatedly proceeds via it happens to subsequently lose it. Yet, tellingly, those lineages that lose the switch either go extinct or eventually return to the switch. Note that the switch also appears sporadically earlier in the tree, a fact relevant to understanding its origins.

That the switch leads to an increase in lineage evolvability is indicated by a decrease in overall extinction rate from 26% per transition to 9% in those lineages descending from the coalescence point of the final surviving lineages (red-circle in Figure 3.18). Splitting the extinction rate among the two phases reveals that this decrease is mostly accounted for by a reduction in extinctions during the WS→SM phase from 29% to 4%, with the SM→WS phase extinctions showing only a marginal decrease from 21% to 15%. This implies the advantage of the switch is less about the generation of WS and more about increasing the WS→SM mutation rate such that the WS mat is successfully invaded by SM cells (which are then more likely to be present during the colony screening process). Indeed, the mechanistic basis of the switch is able to account for this asymmetry of effect.

3.2.8.1 The mechanism of the genetic switch

Although some variations in its mode of action were observed, the genetic switch primarily consisted of the same two mutations: duplication of a 7 bp sequence 'GGTGCCC' residing at the 5' end of *pflu0185* (the sequence begins at nucleotide 97) followed by deletion of the duplicated copy at the next transition. As the initial 7bp duplication results in a frameshift, this causes the PDE activity of Pflu0185 to be turned off – therefore, cyclic-di-GMP (which has been established by the last WS-activating mutation path; in this case the DipA + AmrZ double mutant) is no longer being degraded and WS is turned on. The subsequent contraction of the GGTGCCC sequence at the next transition corrects the reading frame, thereby restoring PDE activity and turning WS off again. Contraction was by far the most common occurrence following the initial GGTGCCC duplication, occurring in 97% of cases. The 3% of cases that led to further expansion of the GGTGCCC sequence will be explored in section 3.1.4.3. Figure 3.19 displays a hypothetical minimal trajectory for establishment of the switch, along with its common mode of action.

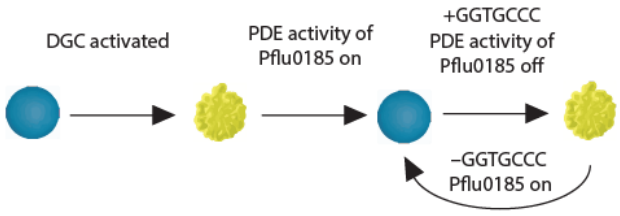

Figure 3.19 Minimal trajectory to establish the pflu0185 genetic switch and its common mode of action

In line with the switch being primarily associated with a decrease in WS→SM extinctions, those lineages inheriting the initial duplication (+GGTGCCC) were found to be significantly more likely to survive the following WS→SM transition than if they had inherited any other cyclic-di-GMP mutation (P=0.0002, Fisher's exact test, two-tailed). In contrast, no such relationship was found between

survival at the SM→WS phase and those lineages that inherited the contraction (-GGTGCCC) of the duplication (P=0.28, Fisher's exact test, two-tailed). This asymmetry of effect can be understood by considering the mechanism of slipped-strand mispairing that likely underpins the WS→SM transition. Slipped-strand mispairing is a well-established cause of elevated rates of expansion and contraction at tandem repeat regions, occurring when the two DNA strands dissociate during synthesis of the repetitive region and are then misaligned upon re-association (Streisinger et al., 1966; Levinson & Gutman, 1987). The tandem repeat formed by the initial GGTGCCC duplication therefore establishes a genomic instability whereby one of the two GGTGCCC units can be readily lost during DNA replication through this mispairing process. The tandem repeat therefore leads to a high rate of reversion of the duplication and a corresponding high rate of WS→SM transitions and decrease in the extinction rate at this phase. In contrast, no obvious mechanism can explain the initial duplication that underpins the alternate SM→WS transition. Nevertheless, the continual regeneration of this tandem repeat is crucial to the functioning of the switch. The tendency of this initial duplication to re-occur will be explored in section **3.2.8.4 Understanding the frequency of the initial GGTGCCC duplication.**

3.2.8.2 An inverted switch

The high rate of reversion of the GGTGCC duplication, and the consequent restoration of the PDE activity of Pflu0185, specifically increases the WS→SM mutation rate. However, this is a consequence of the initial duplication causing a frameshift in *pflu0185* resulting in a loss of PDE activity. If, instead, the initial duplication activated the PDE activity and the reversion resulting in its loss, then we would expect the switch to specifically increase the SM→WS mutation rate. Interestingly, exactly such a mutation occurred. Near the end of the experiment, one of the surviving lineages was found to have inverted the output of the switch (Figure 3.20). This occurred due to a 1 bp deletion in *pflu0185* upstream of the GGTGCCC sequence that was selected initially as an alternative means of inactivating Pflu0185 during a WS→SM transition. As a consequence of this event however, the 7bp duplication of GGTGCCC now corrects the reading frame, turning WS off, and the lineage began to cycle by this means. With the switch inverted, the increased mutation rate associated with the tandem repeat is now acting on the SM→WS transition.

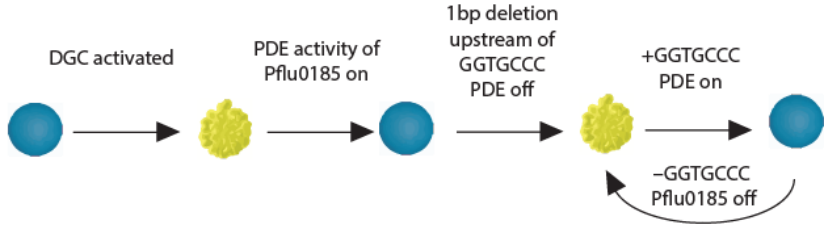

Figure 3.20 Minimal trajectory to establish the inverted pflu0185 genetic switch

3.2.8.3 Rare expansion of the GGTGCCC repeat leads to observable increase in mutation rate

The increased rate of mutation at tandem repeats associated with slipped-strand mispairing should, in addition to the contraction discussed above, also result in expansion of the repeat. Whether expansion or contraction of the repeat occurs depends on whether the daughter or template strand is excluded (looped out) when the DNA is reassociated (Levinson & Gutman, 1987). Repeat expansion is known to cause further increases in the rate of mispairing (Levinson & Gutman, 1987; Lai & Sun, 2003). Indeed, expansions of the GGTGCCC duplication were observed to occur in a small number of lineages. In line with the expectation of an increased mutation following expansion, lineages extending the number of repeats (up to +5 units) exhibited an overt colony-switching behaviour whereby numerous outgrowths of SM morphology appeared at the margins of WS colonies, indicating a further increase in WS→SM mutation rate had occurred in these lineages (Figure 3.21). This colony-switching behaviour was also evident at the +3 stage and indeed, upon waiting an extra day, evident also at the +1 stage (and importantly not at the +0 stage).

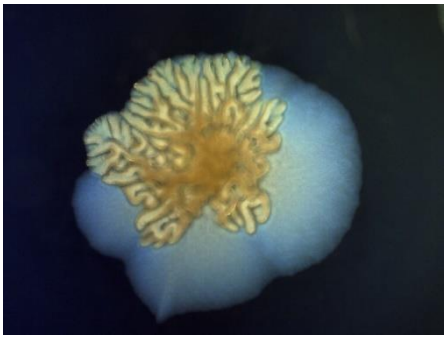

Figure 3.21 Colony-switching behaviour observed in a GGTGCCC +5 genotype after 48 hours of growth

The genotype pictured in Figure 3.21 derived from one of only three lineages that happened to further expand the GGTGCCC sequence rather than contract it at the +1 stage. In each of these three lineages the sequence was first expanded by two additional GGTGCCC units, meaning an addition of 21bp compared to the wildtype sequence. As a multiple of three this corrects the reading-frame and hence turned the PDE back on (and WS off). Following this, in two of three lineages expansion occurred once again – in both cases by an additional two units (+5 in total). That two of the three lineages that expanded the repeat to the +3 stage did so again in the following SM→WS phase indicates that, although expansion was exceedingly rare following the first +GGTGCCC mutation, this tends to become more likely once additional copies exist. Figure 3.22 displays the exact trajectory of one of the lineages that expanded the sequence by 5 additional units.

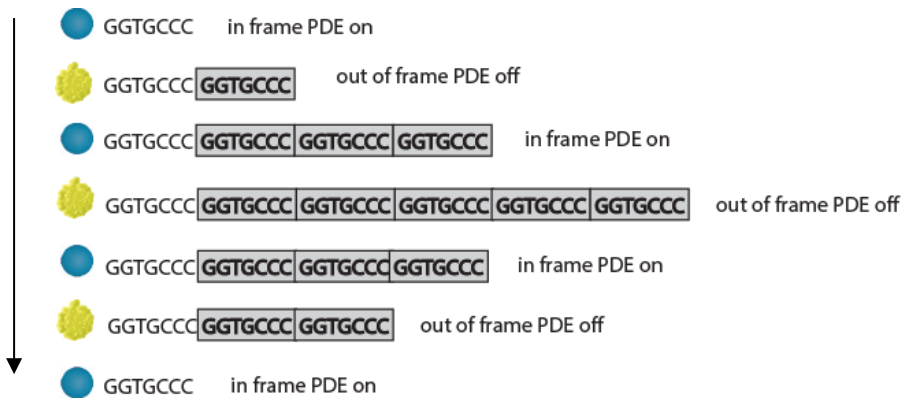

Figure 3.22 One of three rare lineage trajectory with expanded tandem repeats

Following the double expansion, in both lineages the sequence begins to contract, eventually resolving back to wild-type and so losing the associated mutation rate. Although the expanded lineages eventually regressed to a single copy it is plausible that – given the chance for a reproduction event – it could have maintained or increased in length further in one or more of its offspring. For example, consider the lineage at the five +5 unit stage in figure 13 (a state in which the PDE is off). Here a single additional copy of the 7bp repeat could re-align the reading frame, as there would be an additional 42 base pairs (a multiple of 3). A limitation of the small and fixed population size used in the experiment is that it is difficult for any improvements to invade given the condition for reproduction is the extinction of a contemporary lineage (which, due to the stable use of the switch, now rarely go extinct). Shortening the timeframe of both phases or decreasing the population size within a microcosm would increase selection pressure on the lineages and is predicted to drive the evolution of faster switching via extension of the tandem repeats.

3.2.8.4 Understanding the frequency of the initial GGTGCCC duplication

While the increased WS→SM mutation rate can be adequately explained by slipped-strand mispairing, the mechanism that generates the initial GGTGCCC duplication (which allowed the switch to function in its most common mode of action) is unknown. First, the GGTGCCC sequence exists in the ancestral SBW25 genome; the establishment of the switch is therefore not contingent upon the creation of this sequence by mutation. Second, it is not the case that the repeated occurrence of the duplication is the result of it being the only remaining path to WS, as the +GTGGCCC mutation simply causes LoF to *pflu0185* that any number of mutations are also able to achieve. Indeed, alternative LoF mutations in *pflu0185* were still occasionally observed to occur once the switch had become established in the population. These include the 18bp deletion in *pflu0185* depicted in Figure 3.12 that led to the

lineage's eventual extinction, and the 1bp deletion upstream from the GGTGCCC sequence that led to the inverted switch (Figure 3.20). It is therefore certain that something is elevating the initial +GGTGCCC mutation rate relative to possible alternative LoF mutations. This is not to say the whittling process has not been relevant, however. In fact, that the +GGTGCCC mutations occurred sporadically earlier on in the lineage tree (see Figure 3.18) yet did not lead to the proliferation of those lineages, suggests whittling played a crucial role in the eventual ascendance of the +GGTGCCC mutation over alternative LoF mutations. In this case, whittling of other mutational paths (both within *pflu0185* and those at other genes) that were more or equally likely to occur progressed until the +GGTGCCC mutation remained the most likely remaining path to inactivate *pflu0185*. This identifies the GGTGCCC sequence as a mutational 'warm' spot of unknown mechanism.

There is the additional possibility that something about the mutational history of the lineage that eventually comes to dominate via the switch has led to an increase in the +GGTGCCC mutation rate. Examining this possibility requires confronting the unusual flexibility of *pflu0185* as a mutational target in general – recall that this gene was not only targeted by the genetic switch mutations but was additionally the most common target of unique mutations throughout the experiment, accounting for 34% (92 of 273) of all unique mutations. Indeed, the lineage that led to the semi-stable use of the switch carried with it 20 mutations that targeted *pflu0185* before the switch became its dominant means of transitioning. It is possible that one or more of these may have had some incidental effect on the mutation rate of the +GGTGCCC mutation, or indeed on the mutation rate of *pflu0185* in general.

Three overlapping questions must be addressed: (1) What is the mechanism causing the elevated duplication rate of the GGTGCCC sequence that explains its repeated occurrence? (2) Do mutations occurring in the history of the lineage that gave rise to the switch increase this rate? (3) Might mutations occurring in the history of the lineage increase the rate of mutations at *pflu0185* in general? Answers to these questions remain elusive at this stage, but it is worthwhile to speculate on some potentially relevant factors.

3.2.8.5 Local sequence context and DNA secondary structure

Local sequence context is known to influence mutation rates by a variety of means (Schroeder et al., 2016). Note that the closest two nucleotide changes in the final surviving lineages occur far from the from the GGTGCCC sequence itself (43 bp upstream and 437bp downstream), indicating the mutational history has not altered the mutation rate via changes in local sequence context. Nevertheless, the wildtype sequence may already harbour elements that increase the likelihood of GGTGCCC duplication. For example, DNA secondary structures known as stem-loops (or cruciforms in

double-stranded configuration) can stall the replication machinery, potentially leading to genomic instability and mutational hotspots (Voineagu et al., 2008). Indeed, a stable hairpin is predicted in the ssDNA at the position of the GGTGCCC sequence (Figure 3.23) and thus the GGTGCC sequence itself might promote mutation.

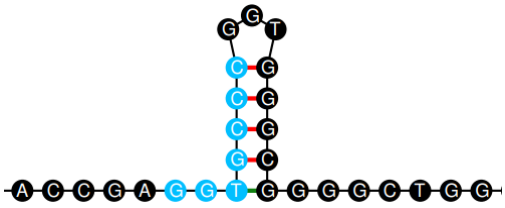

Figure 3.23 Local context of the duplicated sequence that establishes a tandem repeat. The sequence surrounding the GGTGCCC was analysed for the presence of stem-loop structures using mFold (Zuker, 2003). This program folds a segment of ssDNA and predicts the most thermodynamically stable secondary structure. As a crude attempt to identify secondary structure in the sequence surrounding the GGTGCC sequence, I ran this program with default settings on the first 200 bp of *pflu0185*. The folding temperature was repeatedly ramped up until only the most stable structures remained. One of the three such stable stem-loop structures that was still present at a folding temperature of 90°C was partially formed by the GGTGCCC sequence (shown as blue circles).

3.2.8.6 The possibility of transcription-associated mutagenesis affecting *pflu0185* mutability

Mutations outside of the local sequence context may increase the rate of either the GGTGCCC duplication or even mutations at *pflu0185* in general. For instance, any mutation that alters the expression level of a sequence could change the mutation rate as the process of transcription is known to be mutagenic (reviewed in Jinks-Robertson & Bhagwat, 2014). Transcription-associated mutagenesis is thought to occur through a number of mechanisms including separation of the two DNA strands exposing the non-transcribed strand, which is then more vulnerable to mutation. It may also occur through interference between TF binding and the replication and repair machinery of the cell (Kaiser et al., 2016), or through replication-transcription collisions (Lang et al., 2017).

Previous mutations that happened to increase transcription factor binding to the native promoters of *pflu0185*, or alternatively through the creation of novel promoter sequences, could have made *pflu0185* a mutational hotspot. As a rough means of detecting novel promoters, a recently released *Pseudomonas*-specific promoter prediction software developed by Coppens & Lavigne (2020) was used to compare the ancestral and evolved sequences. Intriguingly, three additional promoters were predicted to occur in the evolved sequence (Figure 3.24). These new promoters, alongside mutations upstream that presumably act on the native *pflu0185* promoter, could provide a basis for transcriptional mutagenesis of *pflu0185* itself.

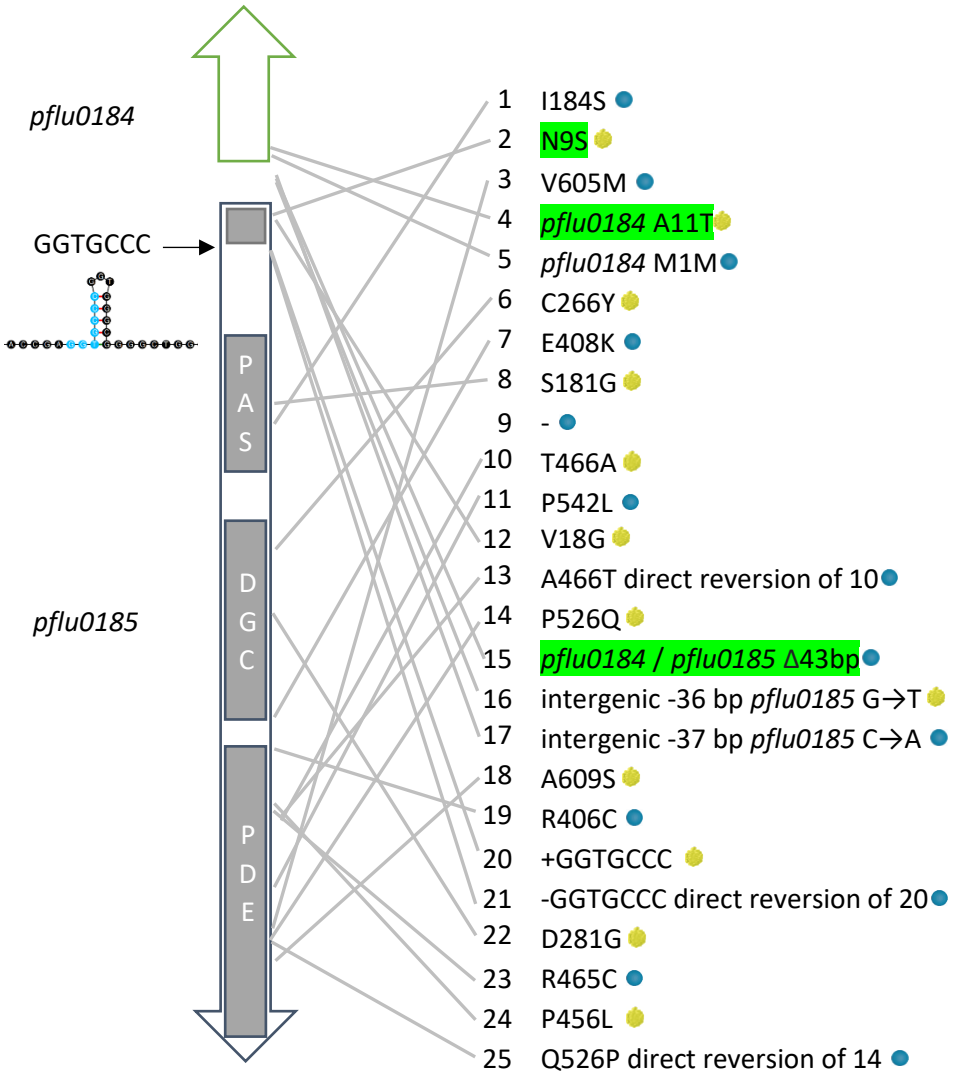

Figure 3.24 Position and order of all *pflu0185* mutations in the lineage prior to the stable use of the genetic switch. Mutations are numbered according to the order in which they occurred. Green highlighting indicates mutations that created novel predicted promoters.

3.2.9 Secondary adaptive mutations

The present section addresses the observation that in some of the phenotypic transitions associated with two (or in one case, three) mutations, genes that had no obvious connection to the transition itself were targeted repeatedly within a lineage and in parallel between lineages. The observed parallelism suggests that these additional mutations are adaptive with respect to some other aspect of cell fitness in the microcosm environment. Moreover, all of the examples occurred during the WS→SM phase and the primary (i.e., transition-mediating) mutation was in every case a toggling of the genetic switch in *pflu0185*. It is therefore hypothesized that the increased mutation rate associated with the switch at the WS→SM phase has facilitated these 'secondary' adaptive mutations, which would otherwise be stalled by the waiting time of the 'primary' transition-mediating mutation. That is, the benefit of the lineage-level adaptation appears to have fed back into evolution within the microcosm and literally 'sped up' evolution – allowing an additional (or in one case, two additional) adaptive steps to occur over the timeframe in which only the transition-mediating mutation would ordinarily occur. In support of this hypothesis, a second mutation (of any kind) was over twice as likely to occur during a WS→SM transition when associated with the genetic switch (20%; 18/88) than with any other transition-mediating mutation during this phase (9%; 13/144), in line with an additional effect of selection over a background of neutral hitchhiking mutations (P=0.0098, Fisher's exact test, two-tailed).

3.2.9.1 Secondary adaptive mutations target motility and chemotaxis regulators

The first genes identified as being targeted with secondary adaptive mutations were *pflu4551* and *pflu1687*. Both encode receptors that feed into the Chemotaxis (Che) pathway – a regulatory complex that integrates signals from multiple receptors within the cell to control flagellar motor proteins, allowing the cell to orient and swim in response to chemical gradients detected in the environment. Typically, the receptors sensing these gradients are membrane-bound Methyl-accepting Chemotaxis Proteins (MCPs). This kind of receptor is remarkably abundant in the SBW25 genome, which has 56 genes annotated as MCPs, all of which are uncharacterized with the exception of WspA.[7]

Pflu1687 is predicted to be a typical MCP but has no characterized orthologs and the signal to which it responds is therefore unknown. Pflu4551, on the other hand, is an ortholog of the aerotaxis receptor Aer in *P. aeruginosa*, which activates Che in response to oxygen gradients (Hong et al., 2004). Unlike

[7] Although annotated as an MCP, WspA does not feed into the Che pathway. Rather, the entire Wsp pathway is paralogous to the Che pathway, with WspA being its single dedicated MCP (Bantinaki et al., 2007).

a typical MCP, this does not occur by direct sensing of a ligand. Rather, Aer indirectly detects changes in external oxygen concentration via monitoring the flux of the electron transport chain in a process sometimes referred to as 'energy taxis' (Edwards et al., 2006).

Beyond Aer and Pflu1687, an additional mutation of potential adaptive significance was found in Pflu4190, an ortholog of the motility regulator FimV in *P. aeruginosa*. FimV has been shown to have a variety of functions in different forms of motility including the biogenesis of type IV pili (the appendage that mediates twitching motility) and in anchoring the flagellar repressor FlgZ at the cell pole (Semmler et al., 2000; Buensuceso et al., 2016; Bense et al., 2019). The mutation in Pflu4190 (henceforth FimV) occurred in conjunction with a mutation in Pflu4551 (henceforth Aer) and a toggling of the genetic switch in the only triple mutant transition detected in the entire line 54 population. Its predicted role as a motility regulator and its appearance alongside the Aer mutation implicates FimV as adaptive and further supports the idea that additional adaptive steps are being enabled by the genetic switch.

3.2.9.2 Additional secondary adaptive mutation candidates

The evidence for secondary adaptive mutations observed in chemotaxis and motility regulators during the WS→SM phase prompted a closer look at some of the other double mutant WS→SM transitions. This revealed four additional loci of potential adaptive significance involved in chemotaxis and flagella regulation, and two genes involved in polysaccharide biosynthesis. Table 3.6 lists all secondary adaptive mutation candidates and Figure 3.25 displays their occurrence in context of the lineage tree.

Primary mutation (*pflu0185*)	**Secondary mutation(s)**	**Function**
(GGTGCCC) 6→4	*aer* Δ3bp	Aerotaxis receptor
(GGTGCCC) 3→1	*aer* S260P	
(GGTGCCC) 2→1	*aer* Δ9bp	
(GGTGCCC) 2→1	*aer* Δ21bp	
	fimV (*pflu4190*) Q640*	Motility regulator (flagella and type IV pili)
(GGTGCCC) 2→1	*aer* E517K	Aerotaxis receptor
(GGTGCCC) 2→1	*aer* G270D	
(GGTGCCC) 2→1	*pflu1687* L27P	MCP
(GGTGCCC) 2→1	*pflu1687* M154I	
(GGTGCCC) 2→1	*pflu1687* S6N	
(GGTGCCC) 2→1	*pflu1687* V215I	
(GGTGCCC) 2→1	*fleQ* R371C	Flagella transcription regulator
(GGTGCCC) 2→1	*fleN* L42Q	Flagella transcription regulator
(GGTGCCC) 2→1	*cheB* T92P	Chemotaxis-specific methylesterase
18bp deletion	*pflu3409* G428D	MCP
(GGTGCCC) 2→1	*pflu1654* nt 260 (T)7→8	polysaccharide biosynthesis; glycosyltransferase
(GGTGCCC) 2→1	*fnl1* (*pflu1657*) S173A	polysaccharide biosynthesis; UDP-glucose 4-epimerase

Table 3.6 Secondary adaptive mutation candidates

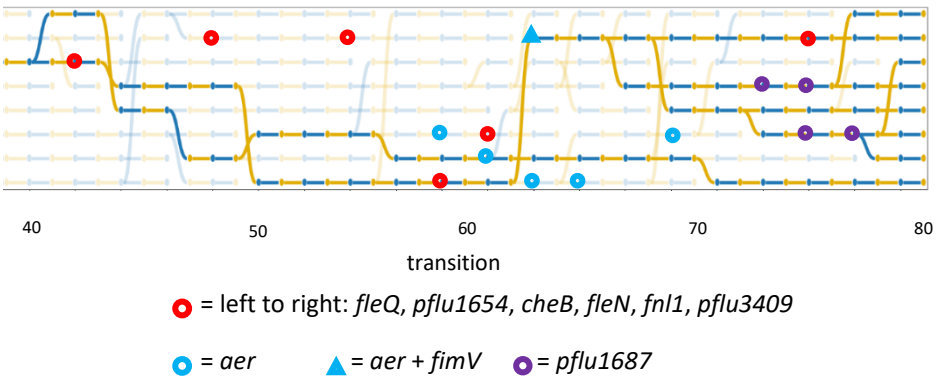

Figure 3.25 Occurrence of secondary adaptive mutation candidates in the line 54 lineage tree. The first half of tree has been omitted as no secondary adaptive candidates were found during this period.

The additional candidates include the flagella transcriptional regulators, FleQ and FleN, another uncharacterized MCP (Pflu3409), and an ortholog of the methylesterase CheB (Pflu4414). The latter is a key player in the Che pathway, where it alters the methylation state of MCPs by removing methyl groups that are constitutively added by the methyltransferase, CheR. The resulting methylation pattern of the MCP determines its sensitivity and is essential for the process of 'sensory adaptation' (Parkinson et al., 2015).

The other two candidates involved in polysaccharide biosynthesis appear to reside within the same operon and are a UDP-glucose 4-epimersase known as Fnl1 (Pflu1654) and a poorly characterized 'hypothetical protein' (Pflu1657) that contains two predicted glycosyltransferase domains. UDP-glucose 4-epimersases catalyse the conversion of UDP-galactose to UDP-glucose (the immediate precursor of cellulose) and mutations in Fnl1 have been found associated with a mucoid colony phenotypes in SBW25 (Harrison et al., 2017). Glycosyltransferases transfer sugar moieties from an activated donor molecule to other sugar molecules (for example from UDP-glucose to cellulose). Both the context in which the Fnl1 and Pflu1657 mutations were found (alongside the genetic switch during a WS→SM transition), their related functions in polysaccharide synthesis, and that they appear to be encoded within the same operon suggested these were secondary adaptive mutations and not merely the result of genetic hitchhiking.

3.2.9.3 Adaptive significance of secondary mutations

If not contributing directly to the WS→SM phenotypic transition then what is the adaptive significance of these mutations? Answering this question requires considering the two possible contexts in which they might have evolved: before the transition (while the cells were WS) or after the transition (while

the cells were SM). Although it may be possible that mutations in motility and chemotaxis genes somehow benefit WS cells during surface colonization, it is most likely the case that they occur following the WS→SM transition in the context of an SM cell growing within the WS mat. The primary reason for this is the inverse regulation of flagella and exopolysaccharide production in *Pseudomonas*: the high levels of cyclic-di-GMP that generate WS are also expected to repress flagella transcription (Baraquet & Harwood, 2013). Clearly, if flagella genes are repressed then mutations affecting their post-transcriptional regulation can have little effect. Exceptions are the mutations in the two polysaccharide-related genes (*fnl1* and *pflu1657*) and the transcriptional regulators FleQ and FleN, which are involved in both flagella and exopolysaccharide regulation. It may therefore be the case that one or more of these mutations was selected in the context of a WS cell.

3.2.9.4 Phenotypic effects of secondary adaptive mutations

To examine the possible phenotypic effects of secondary adaptive mutations I will focus exclusively on the Aer, FimV, and Pflu1687 mutants as these are unambiguously adaptive and, given their predicted functions, likely evolved in the context of a SM cell living within the mat. They should therefore display some identifiable difference in motility and chemotaxis behaviour. Moreover, the phenotypic effects of knockout mutants in Aer and FimV have been studied in other *Pseudomonas* species, providing further insight into their possible phenotypic effects.

3.2.9.4.1 Phenotypic effects of secondary adaptive mutations: Aer & FimV

Aer knockout mutants have been studied in both *P. aeruginosa* and *P. putida* where they were shown to impair aerotaxis (Hong et al., 2004; Nichols & Harwood, 2000). Given the importance of oxygen as the limiting factor of growth in the microcosm, a decrease in aerotactic behaviour would seem maladaptive. Indeed, the mutations that occurred in Aer do not appear to be causing a gene-wide LoF effect. Evidence for this comes from observation of a lineage that incurred a 9bp deletion within *aer* yet in in two offspring lineages later incurred the substitutions G270D or E517K (Figure 3.26). The later occurring mutations would not have arisen if the 9bp deletion had already rendered Aer non-functional. Notably, the two other deletions that occurred in Aer are also in-frame.

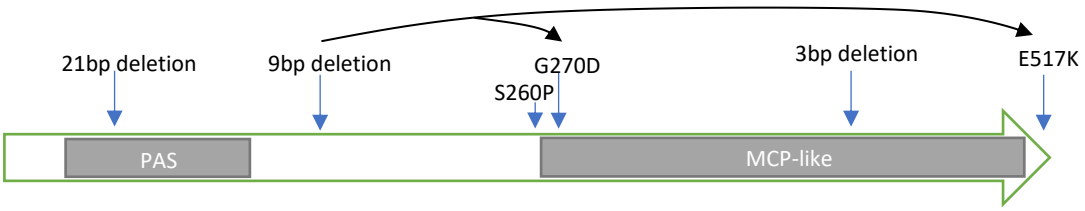

Figure 3.26 Domain structure of Aer and mutations incurred during lineage evolution. Arrows connecting mutations indicate that these mutations occurred within the same lineage. Unconnected mutations occurred in distinct lineages.

The mutation in FimV (that occurred alongside a mutation in Aer in the only triple mutant in the line 54 population) caused a premature stop-codon (Q640*), suggesting it results in either partial or potentially complete LoF (FimV is 861 amino acids long). In *P. aeruginosa fimV* knockouts are known to completely inhibit type IV pilus-mediated twitching motility[8] (Semmler et al., 2000; Wehbi et al., 2011). Other work in *P. aeruginosa* has shown that FimV is also involved in regulation of the flagella motor proteins through interaction with the flagellar motor repressor, FlgZ, which requires FimV for localization at the cell pole in order to enact its repression (Bense et al., 2019). In *P. aeruginosa*, FimV therefore effectively acts as a negative regulator of flagella-mediated locomotion. Assuming a conserved function in SBW25, the Q640* mutation to FimV is therefore expected to decrease twitching motility but increase swimming and swarming motility.

As a preliminary step to detect the phenotypic effects of these mutations on motility, I conducted swarming motility assays on the stored SM cells from before and after the secondary adaptive mutations had occurred. Swarming motility is a flagella-mediated means of surface locomotion and can be readily visualized on semi-solid agar (Kearns, 2010). The secondary adaptive mutations substantially altered swarming behaviour, in terms of both swarming pattern and distance, thereby confirming a phenotypic effect on motility (Figure 3.27).

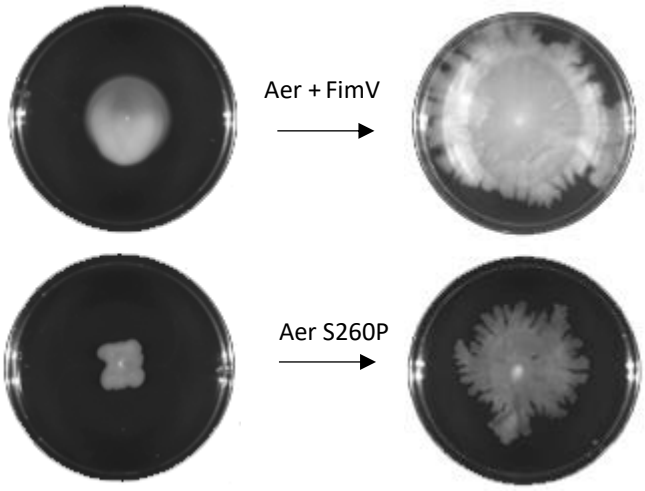

Figure 3.27 Secondary adaptive mutations in Aer and FimV alter swarming motility. Effect on swarming motility for the Aer + FimV mutant and one of the Aer mutants (S260P) is displayed. An intervening WS transition separates the SM cells that are pictured here but as this was mediated by a duplication-deletion cycle of the

[8] During twitching motility, pili extend from the cell and attach to a surface; the pilus is then retracted, allowing the bacterium to move in grappling hook-like manner (Wehbi et al., 2011).

genetic switch in *pflu0185*, has left no genetic trace – the genotypes are therefore identical but for the secondary mutations. Motility plates were imaged 30 hours after inoculation.

3.2.9.4.2 Phenotypic effects of secondary adaptive mutations: Pflu1687

Pflu1687 has not been studied in SBW25 nor does it have any well characterised orthologs. It consists of the common domain architecture of MCPs, including two N-terminal trans-membrane helices that presumably flank an uncharacterized ligand-specific binding domain, followed by a cytoplasmic HAMP domain linked to the MCP domain that interacts with the sensor kinase of the Che pathway, CheA. Figure 3.28 illustrates this domain architecture and the locations of Pflu1687 mutations.

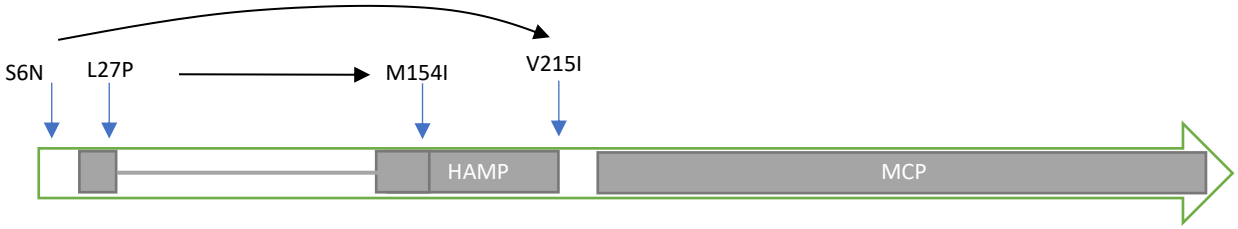

Figure 3.28 Domain structure of Pflu1687 and mutations incurred during lineage evolution. Arrows connecting mutations indicate that these mutations occurred within the same lineage. Connected boxes indicate predicted transmembrane helices.

Swarming motility assays were conducted on the SM cells that followed the S6N→V215I trajectory (Figure 3.29). A clear change in swarming behaviour is observed, this being particularly evident for the increase in swarming distance following the initial S6N mutation. Note that the lineages that incurred mutations in Pflu1687 were descendants of the Aer + FimV mutant (see Figure 3.25 for how these events map to the lineage tree).

Figure 3.29 Secondary adaptive mutations in Pflu1687 alter swarming motility. The starting genotype is equivalent to the Aer + FimV double mutant depicted in Figure 3.27. Since all intervening phenotypic transitions were mediated by duplication-deletions cycles of genetic switch, the genotypes pictured are identical but for the secondary mutations. Motility plates were here imaged 26 hours after inoculation.

Although further work is required to better understand the phenotypic effects of the secondary adaptive mutations, including those candidates outside of Aer, FimV and Pflu1687, it can be confidently concluded that the mutations to the three loci analyzed here are indeed adaptive, affect cell motility, and occur following the WS→SM transition in the context of the mat environment.

3.3 Conclusion

That natural selection favours increasing evolvability is doubted by some (Lynch, 2007; Sniegowski & Murphy, 2006) while considered to be essentially self-evident by others (Kirschner & Gerhart, 1998). The primary objection of critics is the supposed weakness of selection acting above the level of the individual, which means natural selection favours the short-term interests of individuals at the expense of potential long-term benefits (Sniegowski & Murphy, 2006; Lynch, 2007; Brookfield, 2001). Although this objection is legitimate to a degree, the conditions for lineage selection to overcome this myopia are simple – merely requiring that competing lineages with distinct evolvabilities co-exist for long enough such that their differing evolutionary potentials can be manifest. What has been lacking is direct experimental support of the process.

To this end I conducted an evolution experiment that explicitly selected for evolvable lineages. Specifically, small populations consisting of eight lineages were challenged to repeatedly activate and then inactivate a focal phenotype (WS) that was in turn beneficial and then deleterious across time. Failure to reach the target phenotypic state at any one point resulted in extinction (death) of that lineage and replacement (birth) by a contemporary extant lineage from the population. This lineage-level birth-death process allowed natural selection to operate on an entity reproducing beyond the generation time of individual cells and therefore over a timescale where variation in evolvability was visible to selection.

The evolution of one of the populations, referred to as line 54, was subsequently analysed in detail. The evolution of this population can be understood through three related processes, each one necessary for the next: (i) ‘lineage persistence’, whereby the idiosyncratic mutational paths lineages took determined their capacity to survive future transitions; (ii) ‘lineage adaptation’, whereby certain lineages eventually constructed a genetic switch that increased the mutation rate from WS→SM and were thereby selected at the lineage level, and (iii) ‘secondary adaptive mutations’ made possible by

the genetic switch. As the main focus of the modified LCE was on evolvability in the adaptive sense, I will here address only the last two processes.

3.3.1 The birth of a genetic switch

Despite the impressive capacity of lineages to repeatedly cycle WS on and off via a myriad of distinct mutational paths, ultimately lineages were destined for extinction unless they established a reversible and heritable mutational bias. The solution eventually discovered by the line 54 population was local hypermutation consisting of a reversible duplication that toggled on and off the PDE activity of Pflu0185. This mechanism of evolvability is widespread in nature where it has come to be known as a contingency locus (Moxon et al., 2006). While most frequently observed in pathogens, where they provide a means of evading host immunity, rare examples outside of host-associated bacteria are also known. A particularly apt example is the marine bacterium *Pseudoalteromonas atlantica*, which modulates exopolysaccharide production on and off via repeated insertion and excision of a mobile genetic element, which is thought to mediate its alternating life-cycle of surface-dwelling and planktonic phase in the open ocean (Perkins-Balding et al., 1999).

Despite localized hyper-mutability or contingency loci as a mechanism of evolvability not itself being novel, observing the *de novo* evolution of this mechanism certainly is. The results have therefore provided unprecedented insight into the evolution of this phenomenon by making explicit the selective conditions and sequence of events required for its origin and selection. More broadly, the evolution of the genetic switch in *pflu0185* has provided a concrete example of the potency of lineage selection to shape evolvability in an adaptive manner and therefore a confirmation of the controversial evolvability-as-adaptation hypothesis (Sniegowski & Murphy, 2006).

Much of the controversy surrounding evolvability as an adaptation stems from the idea that the short-term interest of the individual is generally opposed to the long-term interest of the lineage (Brookfield, 2001; Dickinson & Seger, 1999; Lynch, 2007). The key to understanding the observed potency of lineage selection both with respect to the evolution of the genetic switch and evolvability beyond this is to see that the incentives of individuals and lineages can be aligned. This is particularly clear in the modified LCE; the colonies selected from the screening plate following each transition likely represented the fittest genotype within the population at the time. Lineage selection simply exploited specific genetic instantiations of equivalent phenotypes by selecting those with a greater capacity for future change.

3.3.2 Secondary adaptive mutations

Due to the increased speed with which the SM phenotype is generated in lineages possessing the genetic switch, additional adaptive steps became possible that optimized cell fitness with respect to other aspects of the environment – a possibility that was not observed in lineages that lacked this rapid-switching ability. Evolution was effectively sped up, just as a global hypermutator is expected to do but without the cost of 'off-target' deleterious mutations and the threat of Muller's ratchet (Sniegowski et al., 1997). This second advantage of contingency loci has not been previously appreciated; the rapid switching ability of contingency loci provides a 'head-start' to adapt to other aspects of the environment and thereby feeds back into the further success of the lineage.

Overall, the experiment described in this chapter serves to demystify how selection operating above the level of individuals can adaptively shape evolvability. Considering what was able to take place in a relatively short period of laboratory evolution in a particularly small population of eight lineages, it is easy to imagine the extent to which similar processes have operated across evolutionary time. An important future task will be to look at the uncharacterised line 57 population that showed a signature of lineage-level adaptation with a currently unknown basis. This will reveal whether the pattern observed for the line 54 population is a general or idiosyncratic outcome of evolution in the modified LCE. As it stands, the results I have presented here represent the clearest experimental evidence yet for the capacity of evolution to act in a truly self-facilitating manner.

3.4 Appendices

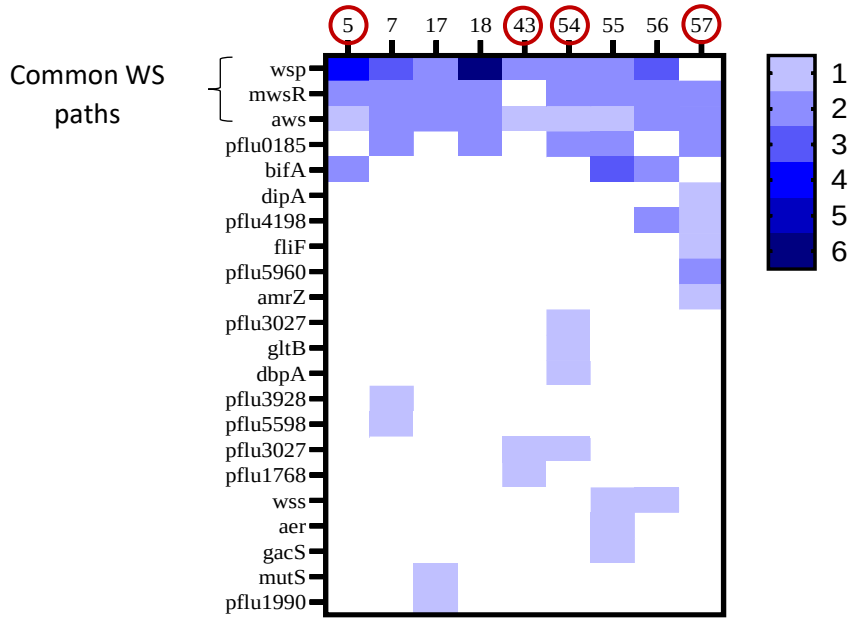

Figure A.1 Heatmap of mutations found in surviving lineages of the original LCE. Red circles highlight the seeding genotypes used in the modified LCE. Note that in the hypermutating line 17 only the subset of the mutations with functional relevance to WS are shown. These sequence data were collected by Hammerschmidt et al. (2014).

Chapter 4

General Materials and Methods

4.1 Bacterial strains and culture conditions

P. fluorescens strains were grown in King's B (KB) medium (King et al., 1954) at 28°C and *E. coli* strains grown in Lysogeny Broth (LB) (Bertani, 1951) at 37°C. For solid media, 1.5% agar was added. Semi-solid KB media with 0.3% agar was used for the swarming motility assays in chapter 3. All overnight cultures were grown from colonies inoculated in 25ml glass vials containing 6 ml of media and shaken at 200 rpm for approximately 16 hours with loosened lids. For evolution of ALI-colonising types and fitness assays, the same 25ml glass vials containing 6 ml of KB ('microcosms') were incubated statically with loosened lids. Bacterial strains were stored indefinitely at -80°C in 33% glycerol saline. The primary bacterial strains used in these experiments are listed in Table 4.1.

Strain	**Genotype/characteristics**	**Reference**
Pseudomonas fluorescens		
Wildtype SBW25	Wildtype strain isolated from the leaf of a sugar beet plant	Rainey & Bailey (1996)
Line 5, line 43, line 54, line 57	WS clones derived from the original LCE	Hammerschmidt et al. (2014)
WS-deficient strain	SBW25Δ*wss*	Lind et al. (2017)
(WS, PGA-WS)-deficient strain	SBW25Δ*wss*Δ*pga*	This work
(WS, PGA-WS, FS)-deficient strain	SBW25Δ*wss*Δ*pga* + pSX (*fuzVWXYZ*)	This work
(WS, PGA-WS, FS, Fim, CAPP)-deficient strain	SBW25Δ*wss*Δ*pga*Δ*fim*Δ*cap* + pSX (*fuzVWXYZ*)	This work
Reconstructed Fim_A mutants	Fim_A causing mutations *pflu1605* A1672G (S558G) and *pflu1609* T281G (D94A) constructed in	This work

	SBW25Δ*wss* both individually and combined	
Reconstructed $CAPP_A$ mutants	$CAPP_A$ causing mutations *pflu3677* G1984A (P662S) and the artificial duplication attTn7::*pflu3655-57* constructed in SBW25Δ*wss* both individually and combined	This work
Reconstructed PSL-WS_A mutants	PSL-WS_A causing mutations *nlpD/rpoS* +36/-71 C→T, *nlpD* C537T (Q189H), and *awsX* ΔT229-G261 (ΔY77-Q87) constructed in SBW25Δ*wss*Δ*pga* both individually and combined	This work
Escherichia coli		
One Shot PIR1	For use with oriR6K plasmids	Invitrogen
Top 10		Invitrogen

Table 4.1 The primary strains used in this thesis

Antibiotics (filter-sterilized and dissolved in ddH20 unless otherwise stated), added to media for selection and plasmid-maintenance as required, were used at the following concentrations: kanamycin (Kan) 100 µg ml^{-1}, ampicillin (Amp) 100 µg ml^{-1}, cycloserine 900 µg ml^{-1}, tetracycline (Tet) 12.5 µg ml^{-1} dissolved in 50% ethanol, nitrofurantoin (NF) 100 µg ml^{-1} dissolved in dimethyl sulfoxide (DMSO). Plasmids used in this thesis are listed in Table 4.2.

Name	**Genotype/characteristics**	**Reference**
pUC18 R6K mini-Tn7T-Gm	R6K, *bla*, Tn7, Gm^R	Choi & Schweizer (2006)
pMRE-Tn7-145 (scarlet)	plasmid pMRE-pSC101ori-oriT-Tn7-AraC-Pbad-tnsABCD-PntpII-mScarlet-I, Gm^R	Schlechter et al. (2018)
pMRE-Tn7-152 (GFP)	plasmid pMRE-pSC101ori-oriT-Tn7-AraC-Pbad-tnsABCD-PntpII-sGFP2, Kan^R	Schlechter et al. (2018)
pCM639 IS-Ω-Km/hah	ColE1, nptII promoter, LoxP, Kan^R	Giddens et al. 2007
pRK2013	Helper plasmid for tri-parental mating; *tra*, *mob,* Kan^R	Figurski and Helinski, 1979
pSX	rep, *bla*, pTac, Gm^R	Owens & Ackerley, 2011
pUX-BF13	Helper plasmid for tri-parental mating; R6K ori, transposition genes, Amp^R	Bao et al. (1991)
pSX (*fuzVWXYZ*)	Contains *fuzVWXYZ* and 5' UTR with native pSX Tac-promoter removed, Gm^R	This work
pSX (*nlpD*)	Contains *nlpD* and 5' UTR with native pSX Tac-promoter removed, Gm^R	This work
pUIC3-mini-GFP	pBR322, *tetA*, *msfGFP*, oriT	Rainey lab, unpublished

Table 4.2 Plasmids used in this thesis

4.3 DNA extraction and purification for sequencing and molecular cloning

Genomic DNA for whole-genome sequencing or for use as PCR templates during cloning was extracted from overnight cultures using the DNeasy DNA Blood & Tissue Kit (Qiagen) according to the manufacturer's spin column protocol. All plasmid DNA was extracted from overnight cultures using the QIAprep Spin Miniprep Kit (Qiagen)**.** For plasmid PCR templates used in cloning, extracted plasmids were linearized with an appropriate restriction enzyme and following the PCR the product was digested with DpnI to remove the remaining plasmid template. All PCR products used in cloning were purified with the QIAquick PCR purification kit (Qiagen). Enzymatic purification with EXOSAP was performed on all PCR products intended for Sanger sequencing to hydrolyse unconsumed dNTPs and remaining primers.

4.4 Sequencing and detection of mutations

Genomic DNA was sequenced using Illumina NextSeq or MiSeq sequencing to a depth of ~50x coverage. Library preparation and sequencing were conducted by the Max Planck Institute for Evolutionary Biology in Plön, Germany using the tagmentation protocol describe by Picelli et al. (2014). The resulting reads were assembled against the reference *P. fluorescens* SBW25 genome (GenBank accession:AM181176.4; Silby et al., 2009) using breseq with default parameters (Deatherage & Barrick, 2014). Mutations were identified by examining breseq's HTML output and in the case of novel predicted junctions, manually inspected using Geneious Prime. Sanger sequencing was also conducted by the Max Planck Institute for Evolutionary Biology using the BigDye Terminator v3.1 cycle sequencing kit followed by the BigDye XTerminator purification kit (ThermoFisher) according to manufacturer's instructions.

4.5 Strain construction

All molecular cloning used isothermal Gibson Assembly reactions with the NEBuilder HiFi DNA Assembly master mix (New England BioLabs) according to the manufacturer's protocol. For construction of point mutations and deletions, regions of length 400-800 bp upstream and downstream of the target site were amplified by PCR (with primers containing the necessary nucleotide change if constructing a point mutation) and subsequently assembled into a pUIC3-mini-GFP vector. For construction of the artificial duplication used to generate the CAPP mutant, the relevant genes and their 5' UTR were amplified and subsequently assembled into pUC18-R6K-mini-Tn7T-Gm. For construction of the plasmids used to block the two ALI-colonising phenotypes caused by loss-of-function to structural genes (FS and CC), the relevant structural genes and their 5' UTR were amplified and assembled into the multi-copy plasmid pSX. All PCRs were performed using master mixes of the high-fidelity polymerases Phusion or Q5. Primer overhangs of the appropriate

complementarity for assembly were generated using the online NEBuilder tool (nebuilder.neb.com). Primer sequences are available on request.

Following the assembly reaction, plasmids were transformed into chemically competent *E. coli* Top 10 cells or, for pUC18-R6K-mini-Tn7T-Gm, into *E. coli* One Shot PIR1 cells (which contain the necessary *pir* genes for the plasmid's R6K origin of replication). Both the pSX and pUC18-R6K-mini-Tn7T-Gm plasmids were then extracted and transformed into SBW25 via electroporation as previously described (Udall et al., 2015), but with pUC18-R6K-mini-Tn7T-Gm requiring the addition of the helper plasmid pUX-BF13 (which contains the necessary transposase genes for integration into the chromosome). For the point mutation and deletion constructs assembled in pUIC3-mini-GFP, integration into SBW25 was achieved by two-step allelic exchange (Rainey, 1999; Bantinaki et al., 2007). Briefly, the first step of this process involved a tri-parental conjugation with *E. coli* carrying the helper plasmid pRK2013 (which contains the *tra* and *mob* genes necessary for conjugation). As pUIC3-mini-GFP is unable to replicate on its own within SBW25, successful transconjugant cells had integrated the pUIC3-mini-GFP plasmid into a single strand of the genomic DNA via homologous recombination. The next step of allelic exchange required a second rarer homologous recombination between the integrated construct and the remaining wild-type allele. Cells having undergone this second homologous recombination event were selected for via growth in the presence of the bacteriostatic antibiotics tetracycline (to which the plasmid confers resistance) and cycloserine (which kills only growing cells via disruption of cell wall biogenesis). By this means, cells containing the plasmid are counter-selected and those having lost the plasmid are enriched – these being either wildtype or those cells that had undergone the second homologous recombination step and so possessing the desired mutant genotype. Colonies without the plasmid were distinguished by their lack of GFP signal on a blue light transilluminator and successful mutants confirmed by Sanger sequencing.

Strains expressing fluorescent proteins for use in fitness assays were constructed via standard conjugation of the recipient SBW25 cells and the *E. coli* donor containing the pMRE-Tn7 plasmid but with media containing 0.1% w/v arabinose to induce the promoter of the Tn7 transposase genes (Schlechter et al., 2018).

4.6 Transposon mutagenesis

Transposon mutagenesis involves the random insertion of a transposon throughout the genome of recipient cells, usually disrupting the gene at the site of insertion. The resulting mutant colonies can then be screened for suppression of the focal phenotype, and the location of the transposon within the genome determined by Sanger sequencing. This process was used to identify the structural basis

of the Fim phenotype according to the previously described method (Giddens et al., 2007). Briefly, this involved a tri-parental conjugation between the recipient Fim mutant and an *E. coli* donor containing the plasmid pCM639 with the IS-Ω-kan/hah transposon along with *E. coli* containing the helper plasmid pRK2013. Since it was not possible to screen for loss of the Fim phenotype by colony morphology, the resulting transconjugants were screened for suppression of mat-forming ability. Random transconjugant colonies were each inoculated into wells of a 96-well plate containing 200 µl of KB media and grown overnight. Suppression mutants were identified on a backlit panel and loss of the mat phenotype was confirmed through observing their growth in a full-sized microcosm. The site of insertion was determined by amplifying the transposon-chromosome junction through an arbitrarily primed PCR (Manoil, 2002) and Sanger sequencing of the resulting product.

4.7 Invasion fitness assays

Fitness was assessed using an invasion fitness assay that competed the ALI-colonising mutant strains against a strain exhibiting the ancestral SM phenotype from a starting ratio of 1:100 (invading mutant population to incumbent SM population). To prevent the ancestral SM strain from itself evolving into ALI-colonizing phenotypes (specifically, the highly likely and high fitness PGA-WS phenotype), the relevant ancestral genotype (SBW25Δ*wss*) was modified such that the three most likely paths to DGC activation were removed (SBW25Δ*wss*Δ*wsp*Δ*aws*Δ*mwsR*). This modification was found to result in far less variance between replicates, particularly for those low fitness phenotypes such as CC. Both the invaders and incumbent ancestral SM strains were fluorescently marked: the invading strains with mScarlet-I and ancestral SM with sGFP2 (Schlechter et al., 2018). To generate the starting populations for the invasion assay, strains were first were grown from colonies for 24 hours at 28°C in a shaking incubator. Each invading strain was then separately mixed 1:100 (10µl to 1000µl) with the SM strain. This mix was then diluted in filtered PBS (0.22µM filter) and the number of cells and initial ratio of scarlet to GFP obtained using flow cytometry by counting a total of 50,000 fluorescent events with the MACSQuant VYB. 6 µl of the 1:100 mix was also used to initiate competition in static microcosms, which ran for either 48 or 72 hours. Following competitive growth, each microcosm was vortexed vigorously to break up cells and again diluted in filtered PBS with the cell counts and final ratio determined by counting 50,000 fluorescent events using the flow cytometer. For each strain, fitness assays were conducted in triplicate at two separate occasions with the exception of a few strains for which the second triplicate is yet to be completed (note that these few strains do not affect any conclusions derived from the fitness assays). Selection coefficients were calculated using the regression model s=ln[R(final)/R(initial)]/t, as previously described (Dykhuizen, 1990), where t is the number of generations and R is the ratio of the invading mutant to the incumbent SM population. The number of generations was determined by ln(final population/initial population)/ln(2).

4.8 The modified LCE

The modified LCE was conducted according to the 'cheat-embracing' regime of the original LCE (Hammerschmidt et al., 2014) with two alterations: mat collapse was ignored as a criterion for lineage survival and a second bottleneck was implemented at the SM stage. To begin, microcosms were inoculated with a single WS colony from selected surviving lineages (line 5, 43, 54 and 57) of the original LCE (Hammerschmidt et al., 2014), establishing four independent sets (lineage 'populations') of eight replicate microcosms per line. The WS cells were then grown for six days, this constituting the WS→SM phase. Following this period, each microcosm was vortexed and 1.5×10^{-7} ml (50 µl of a 3.35×10^{5}-fold dilution) spread on an agar plate. After 48 hours growth each plate was then examined for the presence of SM colonies and a single SM colony used to inoculate a fresh microcosm, initiating the SM→WS phase. SM cells were then grown for three days after which time 2.5×10^{-8} ml (200 µl of an 8.0×10^{6}-fold dilution) of each microcosm was plated. After 48 hours these plates were inspected for WS colonies, and a single colony taken and used to initiate another WS→SM phase. Lineages failing to produce the target SM or WS colony type were deemed extinct, and a random number generator was used to determine an extant lineage that would replace it – a lineage 'reproduction' event. Overnight cultures of each selected SM and WS colony were also made for the creation of frozen stocks.

4.9 Software used

The lineage trees in chapter 3 were generated using Colgen, a custom software designed specifically for visualising results of the LCE (Doulcier, 2019). Signal peptides were predicted using SignalP 5.0 (Almagro Armenteros et al., 2019) and transmembrane helices predicted using TMHMM 2.0 (Krogh et al., 2001). The NCBI Conserved Domain Database (Marchler-Bauer et al., 2015) was used to identify conserved protein domains and the Pseudomonas Genome Database (Winsor et al., 2016) used to identify gene orthologs and synteny. Clinker was used to generate the gene cluster homology and synteny comparison of Figure 2.8 (Gilchrist & Chooi, 2021).

Bibliography

Acheson JF, Derewenda ZS, Zimmer J. Architecture of the Cellulose Synthase Outer Membrane Channel and Its Association with the Periplasmic TPR Domain. Structure. 2019;27: 1855–1861.e3.

Alberch P. Ontogenesis and Morphological Diversification. Am Zool. 1980;20: 653–667.

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37: 420–423.

Amundson R. The changing role of the embryo in evolutionary thought: roots of evo-devo. Cambridge University Press, 2005

Amundson R. Adaptation and development. Cambridge: Cambridge University Press, 2001.

Andersson DI, Hughes D. Gene amplification and adaptive evolution in bacteria. Annu Rev Genet. 2009;43: 167–195.

Ayan GB, Park HJ, Gallie J. The birth of a bacterial tRNA gene by large-scale, tandem duplication events. Elife. 2020;9. doi:10.7554/eLife.57947

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

Bao Y, Lies DP, Fu H, Roberts GP. An improved Tn7-based system for the single-copy insertion of cloned genes into chromosomes of gram-negative bacteria. Gene. 1991;109: 167–168.

Baraquet C, Harwood CS. Cyclic diguanosine monophosphate represses bacterial flagella synthesis by interacting with the Walker A motif of the enhancer-binding protein FleQ. Proc Natl Acad Sci U S A. 2013;110: 18478–18483.

Barton NH. Why sex and recombination? Cold Spring Harb Symp Quant Biol. 2009;74: 187–195.

Bense S, Bruchmann S, Steffen A, Stradal TEB, Häussler S, Düvel J. Spatiotemporal control of FlgZ activity impacts Pseudomonas aeruginosa flagellar motility. Mol Microbiol. 2019;111: 1544–1557.

Bertani G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. J Bacteriol. 1951;62: 293–300.

Besnard F, Picao-Osorio J, Dubois C, Félix M-A. A broad mutational target explains a fast rate of phenotypic evolution. Elife. 2020;9. doi:10.7554/eLife.54928

Blount ZD, Lenski RE, Losos JB. Contingency and determinism in evolution: Replaying life's tape. Science. 2018;362. doi:10.1126/science.aam5979

Bollati M, Villa R, Gourlay LJ, Benedet M, Dehò G, Polissi A, et al. Crystal structure of LptH, the periplasmic component of the lipopolysaccharide transport machinery from Pseudomonas aeruginosa. FEBS J. 2015;282: 1980–1997.

Botos I, Majdalani N, Mayclin SJ, McCarthy JG, Lundquist K, Wojtowicz D, et al. Structural and Functional Characterization of the LPS Transporter LptDE from Gram-Negative Pathogens. Structure. 2016;24: 965–976.

Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017;169: 1177–1186.

Braun V. Energy transfer between biological membranes. ACS Chem Biol. 2006;1: 352–354.

Brookfield JF. Evolution: the evolvability enigma. Curr Biol. 2001;11: R106–8.

Buensuceso RNC, Nguyen Y, Zhang K, Daniel-Ivad M, Sugiman-Marangos SN, Fleetwood AD, et al. The Conserved Tetratricopeptide Repeat-Containing C-Terminal Domain of Pseudomonas aeruginosa FimV Is Required for Its Cyclic AMP-Dependent and -Independent Functions. J Bacteriol. 2016;198: 2263–2274.

Burch CL, Chao L. Evolution by small steps and rugged landscapes in the RNA virus phi6. Genetics. 1999;151: 921–927.

Busch A, Waksman G. Chaperone-usher pathways: diversity and pilus assembly mechanism. Philos Trans R Soc Lond B Biol Sci. 2012;367: 1112–1122.

Calcott B. Lineage Explanations: Explaining How Biological Mechanisms Change. Br J Philos Sci. 2009;60: 51–78.

Cambray G, Mazel D. Synonymous genes explore different evolutionary landscapes. PLoS Genet. 2008;4: e1000256.

Canino-Koning R, Wiser MJ, Ofria C. Fluctuating environments select for short-term phenotypic variation leading to long-term exploration. PLoS Comput Biol. 2019;15: e1006445.

Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008;134: 25–36.

Cheverud JM. Quantitative genetics and developmental constraints on evolution by selection. J Theor Biol. 1984;110: 155–171.

Choi K-H, Schweizer HP. mini-Tn7 insertion in bacteria with single attTn7 sites: example Pseudomonas aeruginosa. Nat Protoc. 2006;1: 153–161.

Conrad M. The geometry of evolution. Biosystems. 1990;24: 61–81.

Consuegra J, Gaffé J, Lenski RE, Hindré T, Barrick JE, Tenaillon O, et al. Insertion-sequence-mediated mutations both promote and constrain evolvability during a long-term experiment with bacteria. Nat Commun. 2021;12: 980.

Cooper VS, Schneider D, Blot M, Lenski RE. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of Escherichia coli B. J Bacteriol. 2001;183: 2834–2841.

Coppens L, Lavigne R. SAPPHIRE: a neural network based classifier for σ70 promoter prediction in Pseudomonas. BMC Bioinformatics. 2020;21: 415.

Couce A, Tenaillon OA. The rule of declining adaptability in microbial evolution experiments. Front Genet. 2015;6: 99.

Cowley M, Oakey RJ. Transposable elements re-wire and fine-tune the transcriptome. PLoS Genet. 2013;9: e1003234.

Cowperthwaite MC, Economo EP, Harcombe WR, Miller EL, Meyers LA. The ascent of the abundant: how mutational networks constrain evolution. PLoS Comput Biol. 2008;4: e1000110.

Cowperthwaite MC, Meyers LA. How Mutational Networks Shape Evolution: Lessons from RNA Models. Annu Rev Ecol Evol Syst. 2007;38: 203–230.

Crombach A, Hogeweg P. Chromosome rearrangements and the evolution of genome structuring and adaptability. Mol Biol Evol. 2007;24: 1130–1139.

Dawkins, R. in Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems (ed. Langton, C. G.) 201–220. Addison-Wesley, Reading, MA, 1989.

de Visser JAGM, Krug J. Empirical fitness landscapes and the predictability of evolution. Nat Rev Genet. 2014;15: 480–490.

de Visser JAGM. The fate of microbial mutators. Microbiology. 2002;148: 1247–1252.

Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol. 2014;1151: 165–188.

Dickinson WJ, Seger J. Cause and effect in evolution. Nature. 1999. p. 30.

Dingle K, Ghaddar F, Šulc P, Louis AA. Phenotype Bias Determines How Natural RNA Structures Occupy the Morphospace of All Possible Shapes. Mol Biol Evol. 2022;39.

Dobzhansky T. Genetics and the Origin of Species. Columbia university press, 1982.

Draghi J, Wagner GP. Evolution of evolvability in a developmental model. Evolution. 2008;62: 301–315.

Dubnau D, Losick R. Bistability in bacteria. Mol Microbiol. 2006;61: 564–572.

Dykhuizen DE. Experimental Studies of Natural Selection in Bacteria. Annu Rev Ecol Syst. 1990;21: 373–398.

Edwards JC, Johnson MS, Taylor BL. Differentiation between electron transport sensing and proton motive force sensing by the Aer and Tsr receptors for aerotaxis. Mol Microbiol. 2006;62: 823–837.

Elena SF, Cooper VS, Lenski RE. Punctuated evolution caused by selection of rare beneficial mutations. Science. 1996;272: 1802–1804.

Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet. 2003;4: 457–469.

Falconer DS, Mackay TFC. Introduction to Quantitative Genetics. 4th Edition, Addison Wesley Longman, Harlow, 1996.

Figurski DH, Helinski DR. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. Proc Natl Acad Sci U S A. 1979;76: 1648–1652.

Fisher RA. The Genetical Theory of Natural Selection. Oxford University Press: Oxford. 1930

Fisher RA. XV.—The correlation between relatives on the supposition of Mendelian inheritance. Earth and Environmental Science Transactions of the Royal Society of Edinburgh. 1919;52(2):399-433.

Freeland SJ, Hurst LD. The genetic code is one in a million. J Mol Evol. 1998;47: 238–248.

Gallie J, Bertels F, Remigi P, Ferguson GC, Nestmann S, Rainey PB. Repeated Phenotypic Evolution by Different Genetic Routes in Pseudomonas fluorescens SBW25. Mol Biol Evol. 2019;36: 1071–1085.

Galton F. Natural inheritance. Macmillan and Company, 1889.

Gilbert SF. The morphogenesis of evolutionary developmental biology. Int J Dev Biol. 2003;47: 467–477.

Gilchrist CLM, Chooi Y-H. clinker & clustermap.js: Automatic generation of gene cluster comparison figures. bioRxiv 2020.11.08.370650. doi:10.1101/2020.11.08.370650

Goldschmidt R. The podoptera effect in Drosophila melanogaster. University of California Press, 1951.

Good BH, Rouzine IM, Balick DJ, Hallatschek O, Desai MM. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. Proc Natl Acad Sci U S A. 2012;109: 4950–4955.

Gould SJ, Eldredge N. Punctuated equilibria: an alternative to phyletic gradualism. Models in paleobiology. 1972;1972:82-115.

Gould SJ. Ontogeny and Phylogeny. Cambridge: Harvard University Press, 1977.

Grangeasse C, Cozzone AJ, Deutscher J, Mijakovic I. Tyrosine phosphorylation: an emerging regulatory device of bacterial physiology. Trends Biochem Sci. 2007;32: 86–94.

Haeckel E. Generelle Morphologie der Organismen. Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von C. Darwin reformirte Descendenz-Theorie, etc. 1866.

Haldane JB. The Causes of Evolution. Princeton University Press, 1932

Hammerschmidt K, Rose CJ, Kerr B, Rainey PB. Life cycles, fitness decoupling and the evolution of multicellularity. Nature. 2014;515: 75–79.

Harrison E, Hall JPJ, Paterson S, Spiers AJ, Brockhurst MA. Conflicting selection alters the trajectory of molecular evolution in a tripartite bacteria-plasmid-phage interaction. Mol Ecol. 2017;26: 2757–2764.

Hershberg R, Petrov DA. Evidence that mutation is universally biased towards AT in bacteria. PLoS Genet. 2010;6: e1001115.

Hoch JA. Two-component and phosphorelay signal transduction. Curr Opin Microbiol. 2000;3: 165–170.

Hogeweg P. Toward a theory of multilevel evolution: long-term information integration shapes the mutational landscape and enhances evolvability. Adv Exp Med Biol. 2012;751: 195–224.

Hong CS, Shitashiro M, Kuroda A, Ikeda T, Takiguchi N, Ohtake H, et al. Chemotaxis proteins and transducers for aerotaxis in Pseudomonas aeruginosa. FEMS Microbiol Lett. 2004;231: 247–252.

Horton JS, Flanagan LM, Jackson RW, Priest NK, Taylor TB. A mutational hotspot that determines highly repeatable evolution can be built and broken by silent genetic changes. Nat Commun. 2021;12: 6092.

Huxley J. Evolution. The modern synthesis. Evolution. The Modern Synthesis.. 1942.

Irie Y, Starkey M, Edwards AN, Wozniak DJ, Romeo T, Parsek MR. Pseudomonas aeruginosa biofilm matrix polysaccharide Psl is regulated transcriptionally by RpoS and post-transcriptionally by RsmA. Mol Microbiol. 2010;78: 158–172.

Jackson KD, Starkey M, Kremer S, Parsek MR, Wozniak DJ. Identification of psl, a locus encoding a potential exopolysaccharide that is essential for Pseudomonas aeruginosa PAO1 biofilm formation. J Bacteriol. 2004;186: 4466–4475.

Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol. 1961;3: 318–356.

Jinks-Robertson S, Bhagwat AS. Transcription-associated mutagenesis. Annu Rev Genet. 2014;48: 341–359.

Kaiser VB, Taylor MS, Semple CA. Mutational Biases Drive Elevated Rates of Substitution at Regulatory Sites across Cancer Types. PLoS Genet. 2016;12: e1006207.

Kashtan N, Alon U. Spontaneous evolution of modularity and network motifs. Proc Natl Acad Sci U S A. 2005;102: 13773–13778.

Kashtan N, Noor E, Alon U. Varying environments can speed up evolution. Proc Natl Acad Sci U S A. 2007;104: 13711–13716.

Kearns DB. A field guide to bacterial swarming motility. Nat Rev Microbiol. 2010;8: 634–644.

Kimura M. Evolutionary rate at the molecular level. Nature. 1968;217: 624–626.

Kimura M. On the evolutionary adjustment of spontaneous mutation rates. Genet Res. 1967;9: 23–34.

Kimura M. The neutral theory of molecular evolution. Cambridge University Press; 1983.

King EO, Ward MK, Raney DE. Two simple media for the demonstration of pyocyanin and fluorescin. J Lab Clin Med. 1954;44: 301–307.

King JL, Jukes TH. Non-Darwinian evolution. Science. 1969;164: 788–798.

King JL. The influence of the genetic code on protein evolution. Biochemical evolution and the origin of life. North-Holland, Viers. 1971:3-13.

King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. Science. 1975;188: 107–116.

Kirschner M, Gerhart J. Evolvability. Proc Natl Acad Sci U S A. 1998;95: 8420–8427.

Kisiela DI, Chattopadhyay S, Libby SJ, Karlinsey JE, Fang FC, Tchesnokova V, et al. Evolution of Salmonella enterica virulence via point mutations in the fimbrial adhesin. PLoS Pathog. 2012;8: e1002733.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305: 567–580.

Kuchma SL, Brothers KM, Merritt JH, Liberati NT, Ausubel FM, O'Toole GA. BifA, a cyclic-Di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by Pseudomonas aeruginosa PA14. J Bacteriol. 2007;189: 8165–8178.

Lai Y, Sun F. The relationship between microsatellite slippage mutation rate and the number of repeat units. Mol Biol Evol. 2003;20: 2123–2131.

Laland KN, Uller T, Feldman MW, Sterelny K, Müller GB, Moczek A, et al. The extended evolutionary synthesis: its structure, assumptions, and predictions. Proc Biol Sci. 2015;282: 20151019.

Lande R. Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain: Body Size Allometry. Evolution. 1979;33: 402–416.

Lang KS, Hall AN, Merrikh CN, Ragheb M, Tabakh H, Pollock AJ, et al. Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell. 2017;170: 787–799.e18.

Leighow, S. M., Liu, C., Inam, H., Zhao, B., & Pritchard, J. R. (2020). Multi-scale predictions of drug resistance epidemiology identify design principles for rational drug design. Cell reports, 30(12), 3951-3963.

Lenski RE, Mittler JE. The directed mutation controversy and neo-Darwinism. Science. 1993;259: 188–194.

Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol. 1987;4: 203–221.

Long H, Miller SF, Williams E, Lynch M. Specificity of the DNA Mismatch Repair System (MMR) and Mutagenesis Bias in Bacteria. Mol Biol Evol. 2018;35: 2414–2421.

Louis AA. Contingency, convergence and hyper-astronomical numbers in biological evolution. Stud Hist Philos Biol Biomed Sci. 2016;58: 107–116.

Lynch M, Walsh B. The origins of genome architecture. Sunderland, MA: Sinauer Associates; 2007

Makova KD, Hardison RC. The effects of chromatin organization on variation in mutation rates in the genome. Nat Rev Genet. 2015;16: 213–223.

Martincorena I, Seshasayee ASN, Luscombe NM. Evidence of non-random mutation rates suggests an evolutionary risk management strategy. Nature. 2012;485: 95–98.

May JM, Sherman DJ, Simpson BW, Ruiz N, Kahne D. Lipopolysaccharide transport to the cell surface: periplasmic transport and assembly into the outer membrane. Philos Trans R Soc Lond B Biol Sci. 2015

Maynard Smith J, Burian R, Kauffman S, Alberch P, Campbell J, Goodwin B, Lande R, Raup D, Wolpert L. Developmental constraints and evolution: a perspective from the Mountain Lake conference on development and evolution. The Quarterly Review of Biology. 1985 Sep 1;60(3):265-87.

Maynard Smith J. Natural selection and the concept of a protein space. Nature. 1970;225: 563–564.

McDonald MJ, Rice DP, Desai MM. Sex speeds adaptation by altering the dynamics of molecular evolution. Nature. 2016;531: 233–236.

McGuigan K, Aw E. How does mutation affect the distribution of phenotypes? Evolution. 2017;71: 2445–2456.

McNamara JT, Morgan JLW, Zimmer J. A molecular description of cellulose biosynthesis. Annu Rev Biochem. 2015;84: 895–921.

Mikkelsen H, Hui K, Barraud N, Filloux A. The pathogenicity island encoded PvrSR/RcsCB regulatory network controls biofilm formation and dispersal in Pseudomonas aeruginosa PA14. Mol Microbiol. 2013;89: 450–463.

Moxon R, Bayliss C, Hood D. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. Annu Rev Genet. 2006;40: 307–333.

Moxon R, Bayliss C, Hood D. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. Annu Rev Genet. 2006;40: 307–333.

Nicastro GG, Boechat AL, Abe CM, Kaihami GH, Baldini RL. Pseudomonas aeruginosa PA14 cupD transcription is activated by the RcsB response regulator, but repressed by its putative cognate sensor RcsC. FEMS Microbiol Lett. 2009;301: 115–123.

Nichols NN, Harwood CS. An aerotaxis transducer gene from Pseudomonas putida. FEMS Microbiol Lett. 2000;182: 177–183.

Nourikyan J, Kjos M, Mercy C, Cluzel C, Morlot C, Noirot-Gros M-F, et al. Autophosphorylation of the Bacterial Tyrosine-Kinase CpsD Connects Capsule Synthesis with the Cell Cycle in Streptococcus pneumoniae. PLoS Genet. 2015;11: e1005518.

Nuño de la Rosa L, Villegas C. Chances and propensities in Evo-devo. Br J Philos Sci. 2020. doi:10.1093/bjps/axz048

Obadia B, Lacour S, Doublet P, Baubichon-Cortay H, Cozzone AJ, Grangeasse C. Influence of tyrosine-kinase Wzc activity on colanic acid production in Escherichia coli K12 cells. J Mol Biol. 2007;367: 42–53.

Olivares-Illana V, Meyer P, Bechet E, Gueguen-Chaignon V, Soulat D, Lazereg-Riquier S, et al. Structural basis for the regulation mechanism of the tyrosine kinase CapB from Staphylococcus aureus. PLoS Biol. 2008;6: e143.

Omadjela O, Narahari A, Strumillo J, Mélida H, Mazur O, Bulone V, et al. BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis. Proc Natl Acad Sci U S A. 2013;110: 17856–17861.

Orr HA. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution. 1998;52: 935–949.

Owen JG, Ackerley DF. Characterization of pyoverdine and achromobactin in Pseudomonas syringae pv. phaseolicola 1448a. BMC Microbiol. 2011;11: 218.

Pal C, Maciá MD, Oliver A, Schachar I, Buckling A. Coevolution with viruses drives the evolution of bacterial mutation rates. Nature. 2007;450: 1079–1081.

Parkinson JS, Hazelbauer GL, Falke JJ. Signaling and sensory adaptation in Escherichia coli chemoreceptors: 2015 update. Trends Microbiol. 2015;23: 257–266.

Parter M, Kashtan N, Alon U. Facilitated variation: how evolution learns from past environments to generalize to new environments. PLoS Comput Biol. 2008;4: e1000206.

Pavlicev M, Cheverud JM, Wagner GP. Evolution of adaptive phenotypic variation patterns by direct selection for evolvability. Proc Biol Sci. 2011;278: 1903–1912.

Payne JL, Menardo F, Trauner A, Borrell S, Gygli SM, Loiseau C, et al. Transition bias influences the evolution of antibiotic resistance in Mycobacterium tuberculosis. PLoS Biol. 2019;17: e3000265.

Perkins-Balding D, Duval-Valentin G, Glasgow AC. Excision of IS492 requires flanking target sequences and results in circle formation in Pseudoalteromonas atlantica. J Bacteriol. 1999;181: 4937–4948.

Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 2014;24: 2033–2040.

Pigliucci M. Is evolvability evolvable? Nat Rev Genet. 2008;9: 75–82.

Plutynski A. What was Fisher's fundamental theorem of natural selection and what was it for? Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences. 2006;37: 59–82.

Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. Empirical fitness landscapes reveal accessible evolutionary paths. Nature. 2007;445: 383–386.

Portin P. The concept of the gene: short history and present status. Q Rev Biol. 1993;68: 173–223.

Potvin E, Sanschagrin F, Levesque RC. Sigma factors in Pseudomonas aeruginosa. FEMS Microbiol Rev. 2008;32: 38–55.

Poulton EB. Essays on evolution 1889-1907. Claredon Press; 1908.

Rainey PB, Bailey MJ. Physical and genetic map of the Pseudomonas fluorescens SBW25 chromosome. Mol Microbiol. 1996;19: 521–533.

Raynes Y, Sniegowski PD. Experimental evolution and the dynamics of genomic mutation rate modifiers. Heredity. 2014;113: 375–380.

Reams AB, Kofoid E, Savageau M, Roth JR. Duplication frequency in a population of Salmonella enterica rapidly approaches steady state with or without recombination. Genetics. 2010;184: 1077–1094.

Reams AB, Roth JR. Mechanisms of gene duplication and amplification. Cold Spring Harb Perspect Biol. 2015;7: a016592.

Remigi P, Ferguson GC, McConnell E, De Monte S, Rogers DW, Rainey PB. Ribosome Provisioning Activates a Bistable Switch Coupled to Fast Exit from Stationary Phase. Mol Biol Evol. 2019;36: 1056–1070.

Roberts IS. The biochemistry and genetics of capsular polysaccharide production in bacteria. Annu Rev Microbiol. 1996;50: 285–315.

Roff DA. A centennial celebration for quantitative genetics. Evolution. 2007;61: 1017–1032.

Sackman AM, McGee LW, Morrison AJ, Pierce J, Anisman J, Hamilton H, et al. Mutation-Driven Parallel Evolution during Viral Adaptation. Mol Biol Evol. 2017;34: 3243–3253.

Salazar-Ciudad I. Why call it developmental bias when it is just development? Biol Direct. 2021;16: 3.

Sauer FG, Remaut H, Hultgren SJ, Waksman G. Fiber assembly by the chaperone-usher pathway. Biochim Biophys Acta. 2004;1694: 259–267.

Scala R, Di Matteo A, Coluccia A, Lo Sciuto A, Federici L, Travaglini-Allocatelli C, et al. Mutational analysis of the essential lipopolysaccharide-transport protein LptH of Pseudomonas aeruginosa to uncover critical oligomerization sites. Sci Rep. 2020;10: 11276.

Schenk MF, Zwart MP, Hwang S, Ruelens P, Severing E, Krug J, et al. Population size mediates the contribution of high-rate and large-benefit mutations to parallel evolution. Nat Ecol Evol. 2022;6: 439–447.

Schlechter RO, Jun H, Bernach M, Oso S, Boyd E, Muñoz-Lintz DA, et al. Chromatic Bacteria - A Broad Host-Range Plasmid and Chromosomal Insertion Toolbox for Fluorescent Protein Expression in Bacteria. Front Microbiol. 2018;9: 3052.

Schluter D. Adaptive radiation along genetic lines of least resistance. Evolution. 1996;50: 1766–1774.

Schroeder JW, Hirst WG, Szewczyk GA, Simmons LA. The Effect of Local Sequence Context on Mutational Bias of Genes Encoded on the Leading and Lagging Strands. Curr Biol. 2016;26: 692–697.

Schroeder JW, Yeesin P, Simmons LA, Wang JD. Sources of spontaneous mutagenesis in bacteria. Crit Rev Biochem Mol Biol. 2018;53: 29–48.

Schuster P, Fontana W, Stadler PF, Hofacker IL. From sequences to shapes and back: a case study in RNA secondary structures. Proc Biol Sci. 1994;255: 279–284.

Ségurel L, Wyman MJ, Przeworski M. Determinants of mutation rate variation in the human germline. Annu Rev Genomics Hum Genet. 2014;15: 47–70.

Semmler ABT, Whitchurch CB, Leech AJ, Mattick JS. Identification of a novel gene, fimV, involved in twitching motility in Pseudomonas aeruginosa. Microbiology. 2000;146 (Pt 6): 1321–1332.

Silby MW, Cerdeño-Tárraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM, et al. Genomic and genetic analyses of diversity and plant interactions of Pseudomonas fluorescens. Genome Biol. 2009;10: R51.

Sniegowski PD, Gerrish PJ, Johnson T, Shaver A. The evolution of mutation rates: separating causes from consequences. Bioessays. 2000;22: 1057–1066.

Sniegowski PD, Gerrish PJ, Lenski RE. Evolution of high mutation rates in experimental populations of E. coli. Nature. 1997;387: 703–705.

Sober E. The Nature of Selection: Evolutionary Theory in Philosophical Focus. Cambridge, Mass.: MIT Press, 1984.

Spiers AJ, Bohannon J, Gehrig SM, Rainey PB. Biofilm formation at the air-liquid interface by the Pseudomonas fluorescens SBW25 wrinkly spreader requires an acetylated form of cellulose. Mol Microbiol. 2003;50: 15–27.

Stoltzfus A, Cable K. Mendelian-mutationism: the forgotten evolutionary synthesis. J Hist Biol. 2014;47: 501–546.

Stoltzfus A, Norris RW. On the Causes of Evolutionary Transition:Transversion Bias. Mol Biol Evol. 2016;33: 595–602.

Stoltzfus A, Yampolsky LY. Climbing mount probable: mutation as a cause of nonrandomness in evolution. J Hered. 2009;100: 637–647.

Stoltzfus A. Mutation, Randomness, and Evolution. Oxford University Press, 2021.

Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, et al. Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. Cold Spring Harb Symp Quant Biol. 1966;31: 77–84.

Sturtevant AH. Essays on Evolution. I. On the Effects of Selection on Mutation Rate. Q Rev Biol. 1937;12: 464–467.

Sumedha, Martin OC, Wagner A. New structural variation in evolutionary searches of RNA neutral networks. Biosystems. 2007;90: 475–485.

Sung W, Ackerman MS, Dillon MM, Platt TG, Fuqua C, Cooper VS, et al. Evolution of the Insertion-Deletion Mutation Rate Across the Tree of Life. G3. 2016;6: 2583–2591.

Turelli M. Commentary: Fisher's infinitesimal model: A story for the ages. Theor Popul Biol. 2017;118: 46–49.

Udall YC, Deeni Y, Hapca SM, Raikes D, Spiers AJ. The evolution of biofilm-forming Wrinkly Spreaders in static microcosms and drip-fed columns selects for subtle differences in wrinkleality and fitness. FEMS Microbiol Ecol. 2015;91.

Uller T, Moczek AP, Watson RA, Brakefield PM, Laland KN. Developmental Bias and Evolution: A Regulatory Network Perspective. Genetics. 2018;209: 949–966.

Voineagu I, Narayanan V, Lobachev KS, Mirkin SM. Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. Proc Natl Acad Sci U S A. 2008;105: 9936–9941.

Waddington CH. Epigenetics and evolution. In R. Brown and J. F. Danielli (eds.), Evolution. (SEB Symposium VII), Cambridge University Press, Cambridge; 1953. pp. 186–199

Wagner A. The Origins of Evolutionary Innovations: A Theory of Transformative Change in Living Systems. OUP Oxford; 2011.

Wagner GP, Zhang J. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat Rev Genet. 2011;12: 204–213.

Watson, R.A. Evolvabiity. In Evolutionary Developmental Biology: A Reference Guide. 2020

Wehbi H, Portillo E, Harvey H, Shimkoff AE, Scheurwater EM, Howell PL, et al. The peptidoglycan-binding protein FimV promotes assembly of the Pseudomonas aeruginosa type IV pilus secretin. J Bacteriol. 2011;193: 540–550.

Weinreich DM, Delaney NF, DePristo MA, Hartl DL. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science. 2006.

Weinreich DM, Watson RA, Chao L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. Evolution. 2005;59: 1165–1174.

West AH, Stock AM. Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem Sci. 2001;26: 369–376.

Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FSL. Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. Nucleic Acids Res. 2016;44: D646–53.

Wright S. Evolution in Mendelian Populations. Genetics. 1931;16: 97–159.

Yampolsky LY, Stoltzfus A. Bias in the introduction of variation as an orienting factor in evolution. Evol Dev. 2001;3: 73–83.

Zaman L, Meyer JR, Devangam S, Bryson DM, Lenski RE, Ofria C. Coevolution drives the emergence of complex traits and promotes evolvability. PLoS Biol. 2014;12: e1002023.

Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD. Interlocking transcriptional feedback loops control white-opaque switching in Candida albicans. PLoS Biol. 2007;5: e256.

Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31: 3406–3415.

Acknowledgments

Being allowed to study evolution every day is wonderful. This privilege I owe entirely to my supervisor Professor Paul Rainey. A million thanks for your support over the years and for doing such interesting science – and for allowing me to try and do some of my own.

While studying evolution and doing science is great and all, writing a thesis is not so much. For getting me through this process I must thank the incredible Dave, Ellen, Joanna, and Bilal. My debt to you all is incalculable but I will pay you back as best I can. Each of you, by being my friend :), also made my life in Plön a happy one. Here I also give special thanks to my amigos and fellow Wrinkly Spreaders Loukas and Andy. Others who have contributed to this journey in various ways that I am grateful for include my officemates Norma, Malavika and Wiola, lab managers Elisa and Daniel, my bungalow flatmate Alan, the remaining Wrinkly Spreaders Juan and Maria, and assorted others: Devika, Demetris, Nico, Karem, Roxy, and Bram.

I am grateful to those who provided administrative and technical assistance throughout this process. In particular I owe many thanks to the wonderful Britta, who has helped me in innumerable ways. Thank you to Alina Waldmann for heroically conducting the hundreds (!) of genomic DNA extractions for the sequencing in Chapter 3, to Guilhem Doulcier for creating the lineage visualisation software and tech support, to Ellen for her *immense* help in conducting the fitness assays of chapter 2, and to Elisa for constructing two of the gene deletions in the same chapter. Your contributions were essential.

To my dear and wonderful family – Mum, Dad, Simon, and Milly the cat – thank you for everything. There is no way I would have been able to do this without your support and guidance.

Last but not least, I thank my darling Rebekka – I have no idea why you put up with me, but I am certainly glad you do.

Declaration

I hereby declare that:

i. Apart from my supervisor's guidance, the content and design of this thesis is all my own work;
ii. This thesis has not been submitted either partially or wholly as part of a doctoral degree to another examining body and no material has been published or submitted for publication;
iii. The preparation of this thesis has been subjected to the Rules of Good Scientific Practice of the German Research Foundation;
iv. No academic degree has ever been withdrawn prior to this thesis.

_____

Michael Barnett