



Experimental evolution of environment dependent gene regulation

Development of experimental systems

Andrew James Higgins

A thesis submitted in accordance with the requirements of the degree of

Master of Philosophy

Under the supervision of

Dr Max Reuter and Prof Jürg Bähler

University College London

Department of Genetics, Evolution and Environment

October 2018

DECLARATION OF AUTHORSHIP:

I, Andrew James Higgins, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ACKNOWLEDGEMENTS

First and foremost I must express my sincerest gratitude to my supervisor Max Reuter for his indispensable guidance throughout this project. His unerring patience and understanding through a difficult time for myself has been essential in allowing me to make it this far.

I am thankful to my co-supervisor Jürg Bähler for his guidance. I also wish to thank both of my supervisors for the opportunity to work on this exciting project in such a stimulating environment. It was an honour to work with so many brilliant colleagues and I will always be thankful for the good times I spent in excellent company at UCL.

Christos Andreadis assisted me in the laboratory and worked relentlessly on this project, whilst also providing me with crucial guidance on experimental techniques.

Quentin Saintain and Claire Fourcade both contributed significantly to stages of the laboratory work for which I am very grateful.

Samuel Marguerat, Xavi Marsellach and other members of the Bähler lab provided indispensable early guidance on working with fission yeast and specific advice on how to progress my project.

Thanks also go to my family for supporting me through difficult times during this project. Especially to my mother, Janet Higgins, my brother, Peter Higgins, and my father, Colin Higgins, for their support and comments on drafts of this thesis.

Andrew Higgins

October 2018

ABSTRACT

The effects of environment on evolution can be explored by experimentally controlling the environment experienced by a population. Data can be collected continuously on evolutionary change and related to the experimental environment. Further, the controlled conditions allow theoretical predictions to be tested. This thesis reports on the development of two experimental evolution systems that can be used to investigate the effects of environmental change on the evolution of gene regulation. In both systems the fission yeast *Schizosaccharomyces pombe* grows under defined selection pressures in two alternating environments. Conditions in both environments are under tight control, with one selecting positively for expression of a target gene, and the other selecting negatively against expression. Alternating growth between the two environments creates a selection pressure for environment dependent regulation of the gene. This is an example of phenotypic plasticity – an environment dependent phenotypic change. Thus, the two systems can be used to investigate phenotypic plasticity and gene regulation, including testing of related theories.

The first system targets the expression of the *ura4* gene. This gene is necessary for the production of uracil, so an environment lacking uracil selects strongly for expression. The alternate environment contains the compound 5-fluoroorotic acid (FOA) which is metabolised by URA4 into a toxin, thus strongly selecting against expression. The second system targets the expression of an introduced green fluorescent protein (GFP) gene using fluorescence activated cell sorting (FACS). The sorting can alternately select for high and low expressing cells from a population. Environmental conditions between the sorts can be altered to provide a cue for the selection the population will face next, thus allowing evolution of environment dependent expression. Experimental work in developing and testing these systems is presented.

IMPACT STATEMENT

The experimental systems we have developed can be used to test theories and provide novel insight in the fields of phenotypic plasticity and gene regulation. These two phenomena are fundamental aspects of the biology of all life. Despite extensive study their true nature and importance is still debated. The modern synthesis of evolutionary theory has proved to be the most powerful predictive theory in biology to date. Yet, in its current state, it does little to account for the mechanistic link between genotype and phenotype, often treating it as a ‘black box’. Gene regulation is a crucial and complex part of this link and its understanding may contribute to significant theoretical progress.

Evolutionary theory attempts to encompass the causal relationships between environment, genotype and phenotype. Clearly this should include situations where the environment changes. Yet phenotypic plasticity has struggled to find a place in the core of the modern synthesis. This may be due to the aforementioned ‘black box’ treatment of the mechanisms of phenotype production, which are crucial to the manifestation of plasticity. Perfect matching of phenotype to environment would always be favoured by selection, yet this is far from what we find in nature. This alludes to significant constraints on plasticity which must be definitive in the nature of adaptation.

Experiments using our systems can provide insights into these important phenomena by testing theories of gene regulation mechanisms and evolution, along with potentially revealing novel evolutionary phenomena. We have developed two systems that select positively and negatively on expression in environments detectable by our organism of study – *Schizosaccharomyces pombe*. There are various parts of these systems that we have made significant progress in refining, including genetic manipulations allowing deep exploration of our experimental outcomes. Our preliminary experiments indicate that the progress we have made offers novel and unique research avenues, especially in the study of eukaryotic organisms. Various aspects of phenotypic plasticity theory can be tested and explored, such as: the evolution of different strategies to adapt to a changing environment; the effects of different fitness landscapes; the effects of cue reliability; the constraints imposed by relaxed selection and biological limits; and the costs imposed by plasticity.

The fundamental role of phenotypic plasticity and gene regulation in all of biology means that further understanding of these phenomena has the long term potential to contribute to the study and improvement of health and disease; the development and utilisation of species in agriculture; and the understanding of ecological systems allowing the preservation of species and environments.

TABLE OF CONTENTS

Abstract	- 4 -
Impact statement	- 5 -
Table of contents	- 6 -
1 Introduction	- 9 -
1.1 <i>Concepts in phenotypic plasticity and gene regulation</i>	- 9 -
1.2 <i>Experimental evolution systems</i>	- 14 -
1.3 <i>Theories and questions in phenotypic plasticity and gene regulation</i>	- 17 -
2 Background	- 20 -
2.1 <i>Phenotypic plasticity</i>	- 21 -
2.1.1 Environmental change.....	- 22 -
2.1.2 Cue reliability.....	- 23 -
2.1.3 Costs.....	- 23 -
2.1.4 Relaxed selection	- 24 -
2.1.5 Biological limits.....	- 24 -
2.1.6 Plasticity, alternative strategies and bet-hedging	- 25 -
2.1.7 Modelling the evolution of phenotypic plasticity	- 27 -
2.2 <i>Environment dependent gene regulation</i>	- 28 -
2.2.1 Mechanisms	- 29 -
2.2.2 Modelling the evolution of gene regulation.....	- 32 -
2.3 <i>Schizosaccharomyces pombe and key genes</i>	- 33 -
2.3.1 URA4	- 33 -
2.3.2 FUR4.....	- 34 -
2.3.3 URG1, URG2 and URG3.....	- 34 -
2.3.4 GFP	- 34 -
2.3.5 CUF1 and CTR4	- 35 -
2.3.6 FEP1 and STR3.....	- 35 -
3 Aims and objectives	- 37 -
3.1 <i>URA4 system</i>	- 37 -
3.1.1 Environmental conditions	- 37 -
3.1.2 Synthesis of strains.....	- 38 -
3.1.3 Measuring phenotypes and fitness	- 39 -
3.1.4 Preliminary evolution experiment.....	- 39 -

3.2	<i>GFP system</i>	- 39 -
3.2.1	Synthesis of strains	- 39 -
3.2.2	Artificial environmental cues	- 39 -
3.2.3	Cell sorting protocol	- 40 -
4	Methods and results	- 41 -
4.1	<i>URA4 system</i>	- 41 -
4.1.1	Dosing of uracil and FOA - optimal concentrations for selection	- 41 -
4.1.2	Strain construction	- 44 -
4.1.3	Sequencing	- 48 -
4.1.4	The fitness landscape: Relating expression and fitness in two environments using the <i>nmt1</i> repressible promoter	- 48 -
4.1.5	Choosing and transforming a suitable promoter from experimental evolution	- 53 -
4.1.6	Preliminary evolution experiment	- 55 -
4.1.7	Creating a library of randomly mutagenised promoters and analysing their effects on fitness and expression	- 57 -
4.2	<i>GFP system</i>	- 60 -
4.2.1	Strain construction	- 60 -
4.2.2	Growth in a detectable environment	- 60 -
4.2.3	Growth and selection – a preliminary experiment	- 64 -
5	Discussion	- 66 -
5.1	<i>Challenges in constructing experimental evolution systems</i>	- 66 -
5.1.1	Methods for alteration of the promoter region of <i>ura4</i>	- 66 -
5.1.2	URA4 deletion and exploiting nitrogen source restriction	- 67 -
5.1.3	<i>Pfta5-ura4</i> experimental evolution	- 69 -
5.1.4	Introduced mutation in <i>Pfta5</i>	- 69 -
5.1.5	Growth in a detectable environment for GFP expression selection by FACS	- 70 -
5.2	<i>Future work – methodological objectives</i>	- 71 -
5.2.1	Maintaining an evolving population	- 71 -
5.2.2	Constraints on plasticity	- 72 -
5.2.3	Fitness landscapes	- 73 -
5.2.4	Bet hedging and variance – single cell expression	- 74 -
5.2.5	A third experimental evolution system	- 75 -

5.3	<i>Future work – testing theories and models of phenotypic plasticity and gene regulation</i>	- 75 -
5.3.1	Testing theories and models of phenotypic plasticity	- 76 -
5.3.2	Testing theories and models of gene regulation.....	- 84 -
5.4	<i>Conclusion</i>	- 87 -
6	References	- 89 -
7	Appendices	- 95 -
7.1	<i>Growth media (EMM – Edinburgh minimal medium)</i>	- 95 -
7.2	<i>Experimental evolution protocol (preliminary experiment – 4.1.6)</i>	- 95 -
7.3	<i>S. pombe transformation - lithium acetate protocol</i>	- 96 -
7.4	<i>RNA extraction</i>	- 97 -
7.5	<i>qPCR</i>	- 98 -
7.6	<i>Protein extraction and western blot</i>	- 99 -
7.7	<i>Error-prone PCR</i>	- 100 -
7.8	<i>Fluorescence-activated cell sorting (FACS)</i>	- 101 -

1 INTRODUCTION

Environment dependent gene regulation is a phenomenon of fundamental importance to the study of phenotypic plasticity and gene regulation, which themselves are two key concepts in the fields of evolutionary biology and molecular genetics. This thesis describes the development of two experimental evolution systems to study environment dependent gene regulation. We begin this introduction by describing the concepts underpinning phenotypic plasticity and environment dependent gene regulation, and highlighting their importance. These, and related concepts relevant to this project, are expanded upon in the Background chapter (2.1 & 2.2). We continue by introducing experimental evolution systems in the context of how they can encompass the concepts we have discussed. Further descriptions of the components of our systems are found in the Background chapter section 2.3. Our aims and objectives in construction of the systems are detailed further in chapter 3. The experimental construction and testing of these systems is described in the Methods and results chapter 4. The final part of this introduction describes the theoretical approaches in the fields of phenotypic plasticity and gene regulation, which provides the basis for the ultimate aims of this project: to test theories and models in these areas using our systems, and to provide novel insights through our findings. The theoretical approaches are detailed further in the Background chapter (2.1 & 2.2). Within the Discussion chapter (section 5.3) we explore exactly how we could use our systems to test such theories and models.

1.1 CONCEPTS IN PHENOTYPIC PLASTICITY AND GENE REGULATION

The phenotype of an organism is determined by its genotype and the range of environments in which it lives. When a single genotype produces different phenotypes in different environments this is known as phenotypic plasticity. Since changing environments are a ubiquitous feature in life we can expect some form of phenotypic plasticity to be the norm for any organism. Even if a phenotype is robust to environmental changes it is likely that this robustness is an evolved response with an underlying mechanism, as changing external conditions will always have some effects on the organism. It is under the conditions of frequent encounters with consistent environmental challenges that we can expect adaptive robustness or phenotypic plasticity to evolve.

When we encounter obvious examples of phenotypic plasticity we are often seeing a gross phenotypic change in response to a significant environmental change. These adaptive changes are complex and require significant, sustained selection pressure to evolve. For example, plants show a clear adaptive response to changing conditions of available sunlight. If they are shaded by other plants, this affects the amount of sunlight available to them, and they will respond with elongated growth allowing them to reach unshaded areas. Adaptive

phenotypic plasticity requires that information from the environment be transduced into the phenotype, i.e. an environmental cue must be detected, and the detection mechanism must then alter the production of the phenotype. In our example the information is sunlight availability, and this is detected by the wavelength composition of the incident light. As we will see, the information is usually transduced by molecular mechanisms.

Genes are usually defined in molecular biology as the parts of the genome that encode proteins. When the genes are translated into proteins this is known as gene expression. This expression can be altered, known as gene regulation. The regulation of gene expression is a fundamental part of the molecular biology of all life and is likely to be a component of many cases of adaptive phenotypic plasticity. Some definitions of phenotypic plasticity even require gene regulation (i.e. changes in expression) to be involved. A gene regulated by the environment is itself an example of phenotypic plasticity.

A gene produces a specific protein that has limited ability to vary in function across different environments. The genotype of an organism is essentially fixed for its lifetime, however, changes in which genes are expressed allows the genome of an organism to produce different phenotypes. This is how gene regulation can allow plastic phenotypes to differ greatly, despite having the same genome encoding for them. This powerful aspect of gene regulation is not limited to plasticity but is seen in almost all aspects of life from the differentiation of cells and tissues in a multicellular organism to the control of growth and reproduction phases in bacteria.

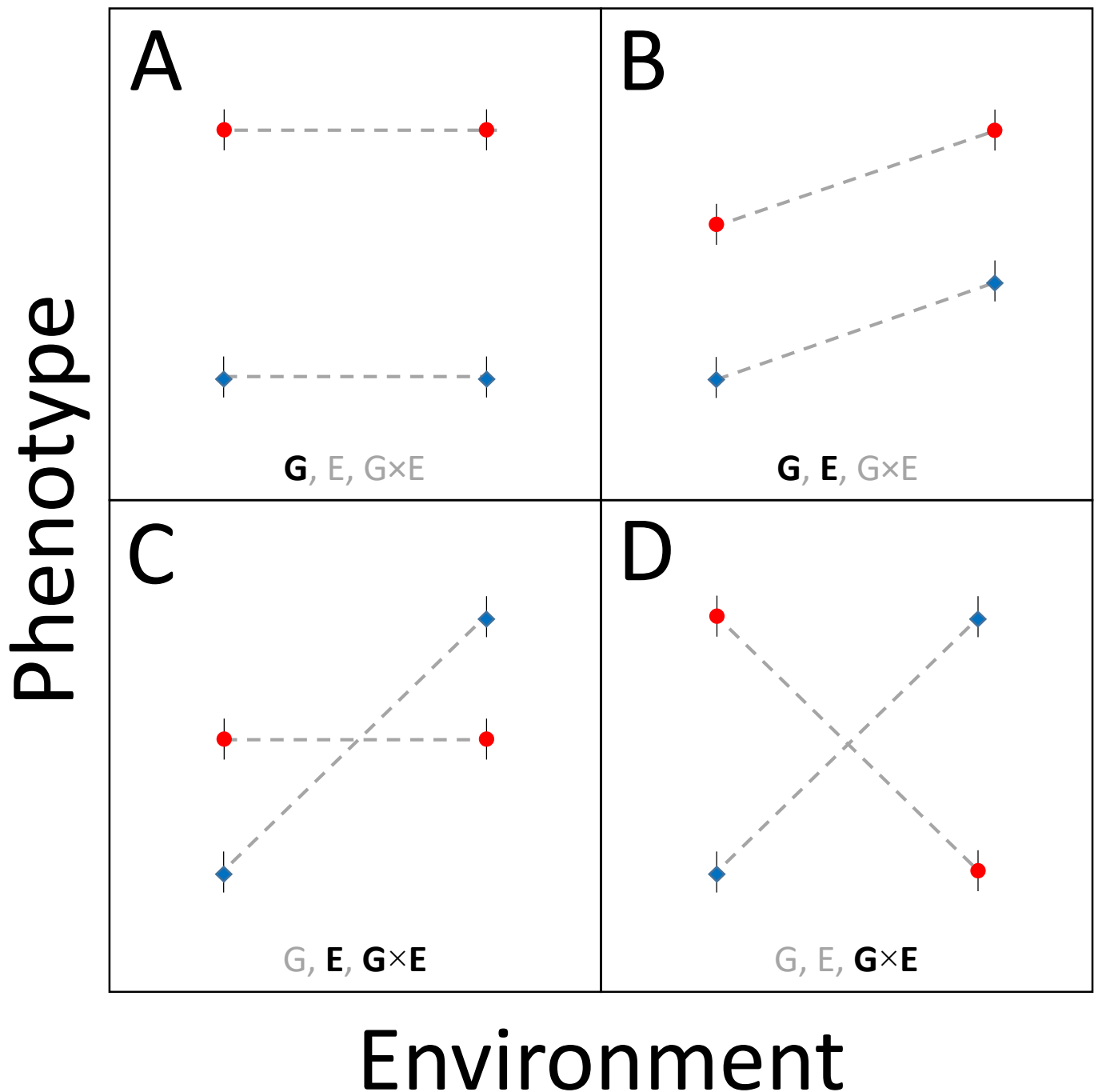
Gene regulation can provide a key step in the transduction of environmental information into phenotypic response. If mechanisms regulating gene expression can evolve to be altered by the environment, then expression can change to produce the required phenotype. Returning to the example of shade response in plants, we see that the XTH proteins of *Arabidopsis thaliana* are upregulated in response to low red-light compositions, the equivalent of shading by neighbouring plants. This upregulation increases cell wall growth leading to elongation of the plant out of the shaded area (Sasidharan et al. 2010).

Now that we have introduced the concept of phenotypic plasticity it will be useful for future discussion to have a stricter definition, though it should be noted that single definitions struggle to capture the scope of this phenomenon. Broadly, phenotypic plasticity can be defined as the expression of different phenotypes in different environments. However, we can be more specific about what constitutes plasticity by specifying the different components of phenotypic variance (V_P) in a population. This variance is made up of genetic and environmental variance (V_G and V_E), and variance due to an interaction between genotype and environment ($V_{G \times E}$), such that $V_P = V_G + V_E + V_{G \times E}$. The components V_E and $V_{G \times E}$ both encapsulate plastic responses and dividing plasticity this way helps us to understand plasticity as an adaptation (**Figure 1**). It

allows us to look specifically at the environmental effects on phenotype for which there is genetic variance ($V_{G \times E}$), and it is in this component where we will understand how plasticity evolves and how it functions as an adaptation (although it is important to say that variance that appears to be solely controlled by the environment, V_E , can still be an evolved adaptation for which there is no detectable remaining genetic variation).

Splitting phenotypic variance as described above leads directly to a fundamental concept in the theory of phenotypic plasticity – the reaction norm. This idea is a cornerstone of plasticity's place within modern theories of genetics and evolution, and it is explicitly or implicitly incorporated into most experiments and models. The reaction norm is simply a quantified phenotypic response to different environments for an individual genotype or a population. The environments can be continuous (and thus also quantitative) or discrete, but the important concept is that the phenotypic response is measured across environments, and thus we have a quantitative relationship between phenotype and environment. A reaction norm can show plastic and non-plastic responses, as exemplified in **Figure 1**. By looking at the reaction norms of individual genotypes we can look both for plasticity and genetic variation for plasticity.

Figure 1. Examples of reaction norms showing different possibilities of genotype and environmental effects on phenotype. The colour of the dots represents a particular genotype or genetically distinct population. **(A)** Genetic variation for phenotype between the red and blue genotypes, but no change in phenotype with environment. **(B)** Genetic variation for phenotype, and a change in phenotype with the environment. However, since both genotypes respond in the same way to the environmental change there is no genotype and environment interaction ($G \times E$). **(C)** The average phenotype of the two genotypes is identical, so there is no overall genetic variation for this. However, there is an overall effect of environment (positive) and there is an interaction between genotype and environment, with the blue phenotype showing a plastic response, unlike red. **(D)** No overall effect of genotype or environment on phenotype, however when looking at individual genotypes there is opposite reactions to the environment, i.e. genotype-environment interaction. (Adapted from DeWitt & Scheiner 2004.)



The genetic and phenotypic variation in plasticity discussed above are predominantly influenced by one factor: evolutionary fitness. This is a quantitative measure of success under selection. The relationship between fitness and changes in phenotype or genotype is known as a fitness landscape. Thus, we have two types of fitness landscape. The phenotype-fitness landscape relates quantitative changes in phenotype to their fitness effects. The genotype-fitness landscape is a quite different concept, as the genotype changes are discrete but can still be connected by mutational distance. This makes the dimensionality of a genotype-fitness landscape very high and it is thus best realised computationally or considered hypothetically. Genotype-fitness landscapes are mentioned in this thesis, but the most important and frequently mentioned fitness landscapes are phenotype-fitness relationships. In plasticity research, changes in the environment are a further addition to the idea of a fitness landscape. Whilst it is possible to consider continuous environmental change as an extra dimension in a fitness landscape this will not be necessary for most of our discussion, where we deal with two discrete environments. In this case we simply discuss the separate fitness landscapes for the two environments, along with the idea of combining them into a ‘fitness set’ which describes how fitness in one environment changes with fitness in the other. The term ‘fitness function’ is sometimes used to describe a specific fitness relationship with phenotype (i.e. a specific instance of a fitness landscape) and is thus interchangeable with the concept of a phenotype-fitness landscape.

Incorporating fitness allows us to conceptualize phenotypic plasticity as an evolved response to a changing environment, but it is not the only possible response. Understanding plasticity means understanding how it evolved, as this will determine what form it will take and when it will occur. A changing environment may select for other strategies over plasticity if the constraints and conditions favour this. Intuitively, plasticity is the optimal strategy, and extensive theoretical work supports this (DeWitt & Scheiner 2004). Alternatives can arise as intermediates or where constraints and conditions favour them. We have already encountered robustness as an alternative to plasticity, but this is not necessarily an alternative strategy, rather just an optimal form of (mechanistic) plasticity where maintaining a fixed (outward) phenotype is favoured. There is a continuum of non-plastic fixed and variable strategies that could arise as bona fide alternatives and this continuum is often divided for conceptual purposes under the terms ‘generalists’, ‘specialists’ and ‘bet-hedgers’. For the two-environment case: generalist strategies are those with a fixed phenotype intermediate to those which have high fitness in each environment; specialist strategies are those which adopt a phenotype which is near-optimal in one environment but sub-optimal in the other; bet-hedgers show phenotypic variance that is under selection, with the distribution of the variance ranging towards the optima of the two environments. Whether plasticity or an alternative strategy arises depends on many factors and constraints, with properties of fitness landscapes being prominent among these. For example, a large overlap between the phenotype-fitness landscapes in the two environments (producing a

convex fitness set) can give a generalist relatively high fitness. In the case of genotype-fitness landscapes there may be regions of low fitness (fitness valleys) between alternative genotypes coding for different strategies, constraining their evolution. In the Background chapter section **2.1.6**, we expand on this topic to include further strategies and combinations along with how conditions and constraints affect their evolution.

1.2 EXPERIMENTAL EVOLUTION SYSTEMS

Experimental evolution systems are artificial or controlled environments where an organism or community of organisms can reproduce for multiple generations, allowing observation of changes in genotype and phenotype. This is useful for testing theories of evolution in a changing environment, as the properties of environmental change can be controlled, and the resultant evolutionary effects observed. Making specific changes to the system and observing the evolutionary outcomes allows us to answer open questions in plasticity, such as when and how it occurs, and what are the prominent constraints on its evolution (Garland & Kelly 2006; Kawecki et al. 2012).

The aims of this project are to create experimental evolution systems which select for environment dependent gene regulation as a form of phenotypic plasticity (Aims and objectives **3**), and to use these systems to test specific theories of how these phenomena evolve. In this report we demonstrate two systems where a pair of tightly controlled environments are used to select for and against expression of a single gene (**Figure 2**; Methods and results **4**). By alternating frequently between growth in these environments we hope to select for regulation of the target gene from a starting point of constitutive expression, which would constitute the evolution of a novel plastic phenotype (**Figure 3**). In the Discussion (Chapter **5**) we explain how experiments using our systems allow testing of a broad range of theories in the fields of phenotypic plasticity and gene regulation.

The organism in which we develop our systems is fission yeast – *Schizosaccharomyces pombe*. The first system targets the *ura4* gene. This gene is necessary for the function of the uracil biosynthesis pathway. Production of uracil is required for survival in an environment containing no uracil, and this environment therefore selects strongly for expression (at the optimal level) of the *ura4* gene. The FOA (5-Fluoroorotic acid) molecule is a uracil analogue which is metabolized by the URA4 protein (orotidine-5-monophosphate decarboxylase). This metabolism creates a harmful product which inhibits growth. Therefore, an environment containing FOA selects strongly against *ura4* expression (when excess uracil is also present, so synthesis is not necessary). By alternating growth of a population between these two environments it is possible that the optimal evolutionary strategy is to evolve regulation of the *ura4* gene. We can artificially place *ura4* under the control of a constitutive promoter with an intermediate level of expression, providing an initial state where there is no regulation and

fitness is similar in both environments. We can then analyse how this strain adapts to the challenge of a changing environment, whether by regulating the *ura4* gene, or possibly by another form of plasticity.

Another complementary system will target a green fluorescent protein (GFP) gene introduced into *S. pombe* under the control of a promoter with constant levels of expression. By exposing a population of cells to repeated rounds of fluorescent activated cell sorting (FACS) and growth, alternating between selection of high and low GFP expression cells, we hope to select for regulation of the GFP gene. Regulation would require the population to be able to detect whether it will be selected for high or low expression, and thus we will provide a signal in the form of growth in alternating concentrations of copper and iron, which are known to cause a transcriptional response. This experiment allows tight control over the evolutionary process our populations go through, allowing detailed testing of models through defined parameters and an ability to alter them. However, it does not reflect the process of regulatory evolution in nature as well as the *ura4* experiment, and therefore we hope they will complement each other in their outcomes.

Figure 2. Hypothetical relationship between fitness and expression in two environments. In environment 1 fitness increases with expression and high expression is selected for, whereas in environment 2 fitness decreases with expression and low expression is selected for. The shape of these fitness functions can vary significantly and our experiment will still be viable, although their shape would have an effect on many aspects of regulatory evolution.

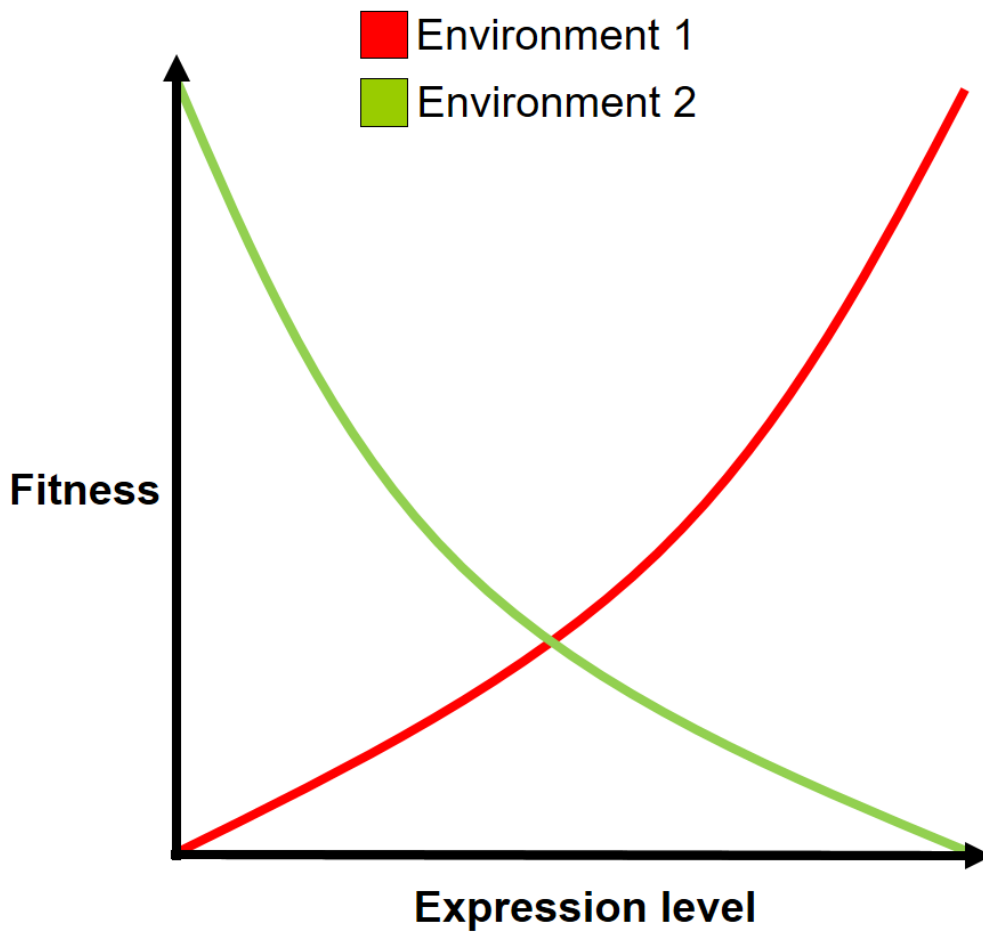
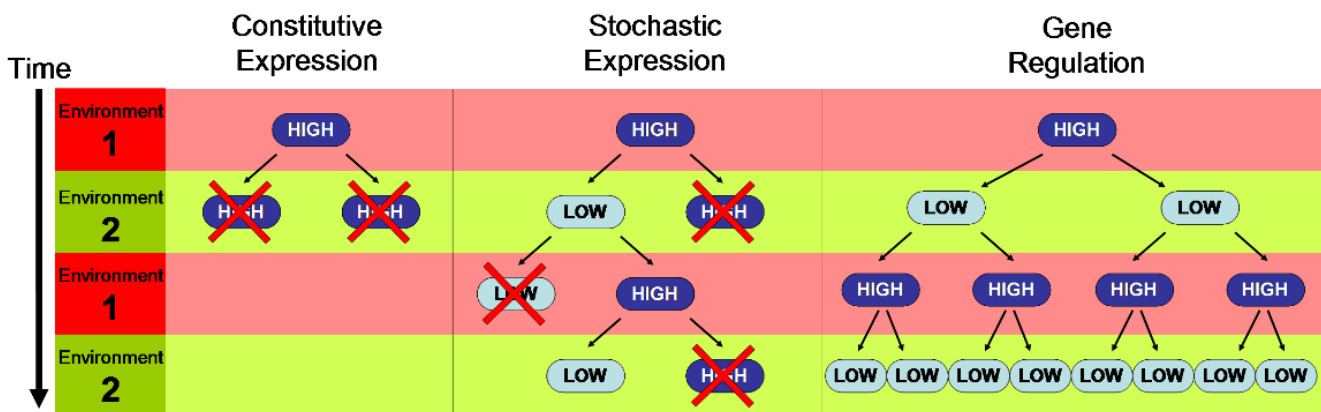


Figure 3. A simplified representation of how different expression patterns will affect fitness during growth in a changing environment. Environment 1 selects for high expression, whereas environment 2 selects for low expression. Cells are labelled with their expression level. Only cells which change their expression can grow well through the changing selection pressures.



This thesis reports on the experimental steps taken to develop the two experimental evolution selection systems described above (Methods and results 4). The use of genetic manipulation to modify the target genes and their regulatory sequences is described. These steps are necessary in the development of the experimental evolution systems, and to assist with their analysis. We also develop further molecular techniques for manipulation of the target genes, allowing us to determine the fitness landscape of our artificial environments via phenotypic and genotypic manipulation. The techniques developed also provide systems for efficient future genetic manipulation. The strains we have developed are analysed in terms of genotype, phenotype and fitness when growing in our selective environments. This has allowed us to refine these environments, understand our experimental systems, and verify that they function as we intended. This report also demonstrates the development of artificially introduced environmental cues, and a preliminary evolution experiment used to develop our experimental evolution system methodology.

1.3 THEORIES AND QUESTIONS IN PHENOTYPIC PLASTICITY AND GENE REGULATION

The ultimate aim of this project is to use the experimental evolution systems we have developed to explore theories of gene regulation and phenotypic plasticity. In this report we describe how multiple tests of specific theories and models could be carried out using our experimental evolution systems (Discussion 5.3). Below we introduce the theories and types of model that exist, and the important questions they tackle, with the implications of specific models being further explored in the Background (2.1 and 2.2).

Many mathematical models of plasticity make use of quantitative measures of the environment-dependent phenotypes of different genotypes, captured in the concept of reaction norms. Some models use a quantitative genetics approach, where genotypes are non-specific instances of genetic variation, and covariation between phenotypes in multiple environments is modelled. Allelic (or gametic) models take a more realistic approach to the genetics of plasticity by following segregation of individual mutants. Optimization models largely ignore the genetic basis of adaptation to simplify the problem and thus focus on more realism elsewhere. Introducing realism often involves introducing mechanistic detail and this is where models of environment dependent gene regulation make their contribution. All the models share similar aims of elucidating when plasticity will evolve, what the process will look like and what form plasticity will take. Thus, our experiments also share the same ultimate goals, and the specific theoretical tests we propose can be summarised in those three questions.

All theories generally agree that plasticity can often be adaptive, but that there can be a variety of constraints on its evolution. We describe how our experiments would allow us to test the predicted conditions in which adaptive plasticity will evolve, and what other strategies

might evolve as alternative or intermediate adaptations (5.3.1.1). We further explain how our experiments could be used to test theories of specific constraints due to: the costs of phenotypes (5.3.1.4); the shape of the fitness landscape (5.3.1.1; 5.3.1.5; 5.3.2.2); the availability of genetic variation (5.3.1.8); relaxed selection (5.3.1.3); hard and soft selection (5.3.1.6) and limited accuracy of environmental cues (5.3.1.5). Which constraints are prominent in nature is an important open question in plasticity research and we discuss how our experiments could make contributions to finding an answer to this. An important factor defining constraints and plasticity will be the mechanisms involved and how they can evolve. We also hope to provide insight here, specifically in the mechanism of gene regulation.

We have already seen that gene regulation is potentially fundamental in phenotypic plasticity and is likely to be important in most known cases; one of the reasons being that it theoretically provides a mechanism for utilising subdivisions of information in the genome. Evolutionary theory implies that every regulated gene must have evolved to its regulated state at some point in history, and thus a way to understand gene regulation is to see how natural selection drives this process of adaptation. Experimental evolution would allow us to analyse this process and its end result, and we describe the specific insights this might provide in relation to molecular mechanisms (5.3.2.1). For example, we suggest how our experimental evolution system could be used to test a theory predicting the effects of pre-existing regulatory DNA sequences on the evolution of regulation (Tirosh et al. 2009). We also explore the predictions of a mathematical and computational model by Proulx & Smiley (2010) which provides a good approximation to our experiments. They use differential equations to model the changes in expression of a single gene evolving in two environments which select for and against expression. Fitness is determined based on the protein concentration and its degradation rate. They extend this model into a population genetics framework to see how a gene might evolve regulation over time. We describe how we can test the predictions they make regarding when regulation evolves, and the changes in the regulatory mechanisms of protein expression and degradation (5.3.2.2). For example, they predict that under convex fitness sets an intermediate constitutive expression level (a generalist strategy) will evolve and that regulation will always evolve from this (provided the necessary mutations arise). They also predict that once regulation has evolved there will be increases in protein degradation rate with concurrent increases in production rate (expression) as an adaptation to respond quickly to environmental change.

Whenever we observe an organism we are seeing one facet of a much more complex entity when we consider its potential phenotypes in different environments. Thus, phenotypic plasticity cannot be ignored when investigating any aspect of biology. Gene regulation is a fundamental mechanism in phenotypic plasticity and beyond this is a key component of cellular biochemistry. However, aspects of phenotypic plasticity and gene regulation beyond their

natural history are not well understood, and there are many open questions relating to constraints and mechanisms. Fundamental principles of evolutionary change in plasticity and gene regulation, and the broader effects of such change on evolution and biology, are just beginning to be elucidated. The models we have discussed are just a few examples of the wide variety of theoretical literature that exists surrounding these topics. Our experimental evolution project hopes to shed light on the validity of these models using a novel technique. By using our systems to gather data on evolution in a changing environment we hope to test models, develop new theories, and reveal novel evolutionary and biochemical processes in the evolution of gene regulation.

2 BACKGROUND

Phenotypic plasticity can be an example of adaptation due to natural selection (Ghalambor et al. 2007), and thus can be investigated through the paradigms of the modern synthesis. We can expect its evolution to be subject to the same potential constraints as any other adaptation, such as its benefits allowing it to overcome genetic drift, and it being an evolutionary stable strategy. However, it also has some unique attributes as an adaptation which determine further potential constraints upon it. As such we do not expect perfect plasticity, where traits are an optimal reflection of every environment the organism encounters. By understanding plasticity evolution and its constraints we can understand when we expect it will occur, how it will evolve and what form it will take.

Environmentally controlled gene regulation is a ubiquitous example and important mechanism of phenotypic plasticity. Gene regulation, whether under environmental control or otherwise, is a fundamental mechanism of cellular biochemistry, and forms a key part of effectively all the processes of life. By understanding the biology of regulation, and including this in evolutionary models, we can begin to understand the constraints on plasticity beyond those proposed by evolutionary theories which generally treat the links between genotype and phenotype as a ‘black box’.

Predictions from theories and models of environmentally controlled gene regulation can differ from those of broader theories and models of evolution in a changing environment, as they incorporate different parameters and constraints. Unsurprisingly, the predictions centre on the effects of the factors which have been included in the model or theory. Broader evolutionary models might emphasise the importance of ubiquitous effects such as fitness landscape shape or relaxed selection in plasticity evolution. Meanwhile, mechanistic models of gene regulation might offer predictions on a completely different scale, focussing on expression changes or the genetic components of regulatory elements. All these effects could be important, and any effects dealt with in broader theories will have a mechanistic basis that will further determine and complicate them. This emphasises the need to create inclusive models, but also to use experimentation to guide the theory towards factors which seem to be prevalent. In most cases the predictions from different models are compatible, though this is not always true. For example, plasticity theory generally predicts that concave fitness sets favour plasticity more than convex sets due to the larger fitness benefits of plasticity over other strategies, whereas the regulatory model of Proulx & Smiley (2010) suggests that convex sets are actually more likely to lead to plasticity (regulation) due to the dynamics of protein production.

The research areas of phenotypic plasticity and gene regulation often have significant divergence in their goals, despite their fundamental connections. Phenotypic plasticity is often investigated within an evolutionary paradigm, centred on how understanding constraints on

plasticity evolution can allow us to predict the dynamics and outcomes of adaptation. Gene regulation is often investigated from a mechanistic perspective, but the field naturally extends to an evolutionary and predictive focus, especially when considering environment dependent gene regulation where the functional perspective is imperative (Hodgins-Davis & Townsend 2009). Following the differences in these research areas, the sections below focus first on the broader field of plasticity (2.1), and then the narrower field of gene regulation (2.2). However, there should be clear and frequent connections between these areas and the questions they raise. Both sections end by looking at mathematical models which attempt to put our understanding of evolution in changing environments onto a rigorous footing (2.1.7; 2.2.2). The final section of this chapter reviews our understanding of the biology of our experimental systems (2.3), which is fundamental to their development, and to the prediction and understanding of experimental outcomes.

2.1 PHENOTYPIC PLASTICITY

Evolutionary models of changing environments which allow the possibility of phenotypic plasticity usually show that it is the optimal phenotype (DeWitt & Scheiner 2004). This provides good support for the theoretical and intuitive concept that a perfectly plastic organism would always be the most successful, as its phenotype would always be the optima for any environment it found itself in (this is how it is defined). In nature this is not what we find, instead there are many species which show little plasticity relative to a perfectly plastic idealised species. The reasons for this appear abundant: a perfectly plastic species would have to code for an infinite number of phenotypes and thus have an infinitely large genome; organisms would need to be able to change their phenotype instantly and continuously in response to new environments; there would be no selection to create or maintain adaptations to so many environments which were never encountered. Whilst extreme, these examples provide manifest demonstration of the biological and evolutionary constraints that could affect plasticity evolution. If plasticity is optimal then adaptation will take phenotypes as far towards plasticity as they can go, and thus research on plasticity has tended to focus on constraints on this. These constraints will determine how far plasticity can go in realistic situations, and thus what form adaptation will take in nature.

Adaptations seem to have been favoured which reduce the need for plasticity by dealing with environmental change in other ways: robustness allows organisms to continue functioning in the same way with limited mechanistic plasticity to buffer the environmental change, removing many of the potential constraints on overt plastic phenotypes; migration allows organisms to place themselves in environments to which they are well adapted; sex is potentially selected for because genes which promote it find themselves in diverse genetic backgrounds, some of which will be better suited to any environmental change which has

occurred. In these adaptations, and in the diversity of species we encounter, we see that changing environments are a huge challenge to life. Plasticity often appears as a relatively minor component of adaptation, despite its theoretical benefits. What are the constraints that stop it being more prevalent and instead lead adaptation towards other solutions to environmental change? Clearly the constraints must be powerful and ubiquitous, defining adaptation everywhere we turn, yet we still do not know which constraints are the most important and what effect they are having.

It is for these reasons that research in phenotypic plasticity is often focussed on constraints on its evolution. There has been significant work on the costs of plasticity, although it is very difficult to reach firm conclusions due to difficulty in finding such costs. These are sometimes distinguished from the concept of ‘limits’ which are constraints that prevent adaptation for reasons other than the adaptation being too costly. This is not always the most useful distinction; limits are diverse and the distinction is not always clear (Auld et al. 2010). Therefore, it can be better to think of all the factors affecting plasticity evolution as constraints. Even when considering alternative strategies to plasticity we are encountering the different possibilities that can occur depending on how plasticity has been constrained. The following sections present key factors in plasticity evolution, but they are not an exhaustive or ordered list in terms of the importance of the constraints they discuss. The final section on modelling the evolution of phenotypic plasticity (2.1.7) summarises the different approaches to modelling plasticity, along with their strengths and weakness. This should provide a context for the predictions provided by these models which appear elsewhere in this chapter and in the Discussion (Chapter 5).

2.1.1 **Environmental change**

The evolution of adaptive plasticity is dependent on repeated exposure to different environments which are detectable by the organism (although without a reliable cue there is the possibility of evolving a strategy of random phenotype switching – 2.1.6). Environmental variation can be both spatial and temporal, and can span a range of frequencies, from occurring repeatedly within the lifetimes of organisms to occurring less than once per generation. In all these situations adaptive plasticity has the potential to occur, but they all have different effects on when it will be favoured and how it will evolve. For example, temporal changes are more likely to favour plasticity than spatial changes, due to their unavoidable effect on the entire population (Moran 1992). The frequency of environmental change can also have a variety of effects; a low frequency could lead towards changes in allele frequencies differentially favoured in each environment, causing shifts in the mean (non-plastic) phenotype over time. Rapid environmental change can be so quick that phenotypic change cannot be fast enough to match the environment before it has changed again, and thus favours a fixed strategy (Levins 1968;

Jablonka & Szathmary 1995). Intermediate frequencies of predictable environmental fluctuations, in contrast, are most favourable for plasticity to occur.

2.1.2 Cue reliability

Adaptive plasticity requires producing the correct phenotype to match the environment. If the environment changes in predictable patterns then this is possible without direct detection, although some information will be needed for synchronisation. In order to respond to the environment the organism requires a mechanism to detect a cue and transduce this information into a phenotypic response. Production of these mechanisms may be a constraint on plasticity evolution (2.1.5), but even when these mechanisms are possible there may be a further constraint due to the cue used for detection not always exactly matching the environment. This is the problem of cue reliability, and it can have significant effects on the evolution of plasticity. If reliability of a cue is below a certain threshold then plasticity may no longer be favoured (DeWitt & Langerhans 2004; Tufto 2000). However, it seems likely that selection would favour adopting a more reliable cue if an environmental mismatch occurs, although this may not always be possible.

2.1.3 Costs

Any phenotype produced by an organism will have costs in terms of energy, resources and time, which could potentially have been spent on other phenotypes with fitness benefits. Therefore, for a phenotype to be favoured by selection its cost-benefit ratio must outweigh that of other possible paths of evolution. Plasticity itself (the mechanisms which alter phenotypes in different environments) can have a cost over and above the cost of the different phenotypes expressed (Callahan et al. 2008). For example, if an organism was to undergo a large morphological change in response to a changed environment, such as the growth of a defensive structure in the presence of a predator, the process of doing this may be costlier than a fixed structure formed through early development. On a different scale we can imagine that the environmental detection mechanisms required for a gene to be expressed in response to environmental change may be costlier than simply expressing the gene constitutively. Such costs of plasticity have often been suggested as a common constraint on its evolution, and some studies do show plastic traits having costs, although these are usually low (DeWitt et al. 1998; van Kleunen & Fischer 2005; van Buskirk & Steiner 2009). This does not necessarily mean that costs are not an important constraint on organisms producing extreme and diverse plasticity, since if high cost is a significant constraint we are implicitly expecting not to find examples of it. Showing that costs exist is also difficult as it requires comparing fitness of individuals with plastic and non-plastic adaptations. Experimental evolution in a changing environment can potentially allow detection of costs, by using a parallel experiment where organisms evolve only in one of the environments. By comparing the phenotypes and fitness of the fixed and changing environment

lines the costs of the production mechanisms can potentially be calculated (discussed further in 5.2.2 and 5.3.1.4).

2.1.4 **Relaxed selection**

Relaxed selection is related to the concept of genetic drift – that selection must have a certain strength to overcome the fixation of random variation in populations, a powerful force in slowing or preventing optimal adaptation of organisms. This can affect a plastic trait, and it can also affect the phenotype produced by plasticity in only some environments. Because this trait is not produced all the time, its selective benefit is not always realised, assuming a low frequency of environmental change such that organisms often or always spend their entire lives with one phenotype. Therefore, purifying selection on the regions of the genome encoding for the environment specific phenotype is weakened. This is similar to the weakening of selection on regions of the genome that are sex-specific in their expression and thus only experience half of the purifying selection strength of a non-sex-specific trait (Snell-Rood et al. 2010). Similarly, an environment-specific trait would also have its purifying selection strength halved if that environment was only experienced by 50% of organisms. We thus expect drift to more easily overcome selection in such traits, and mutational load will cause the phenotypes to be less optimal and fitness to be lower when environments are experienced less, for which there is some empirical evidence (Snell-Rood et al. 2010).

2.1.5 **Biological limits**

There may be situations where biochemistry, physiology or other organismal traits simply do not have the capacity to produce a certain phenotype. This occurrence could be common but is difficult to detect (Auld et al. 2010). We have already discussed examples of such limits when considering the lack of environmental detection mechanisms and the inability of organisms to alter phenotype rapidly enough if the environment is changing with a high frequency. Such biological limits represent fitness peaks on the genotype-fitness landscape. Theoretically, there can be higher fitness peaks coding for plastic phenotypes which are separated from the non-plastic phenotype by fitness valleys. This concept of fitness valleys arises in various ideas of limits of plasticity, and indeed all phenotypes, and is based on the idea that as mutations become more complex their likelihood of being adaptive becomes vanishingly small (i.e. mutations that cross valleys to new peaks are extremely rare), and thus natural selection almost always proceeds by small mutational steps (Darwin 1859; Dawkins 1976, 1982; Dennett 1995). If mutations for plasticity are rare compared to those for other strategies such as generalism then this will also affect the dynamics of evolution, leading to the evolution of other strategies preceding the evolution of plasticity. Crossing such fitness valleys is more likely when they are small, and mutation can jump them. For example, the gene regulation model of Proulx & Smiley (2010) shows how non-regulated gene expression can jump small fitness valleys to regulation.

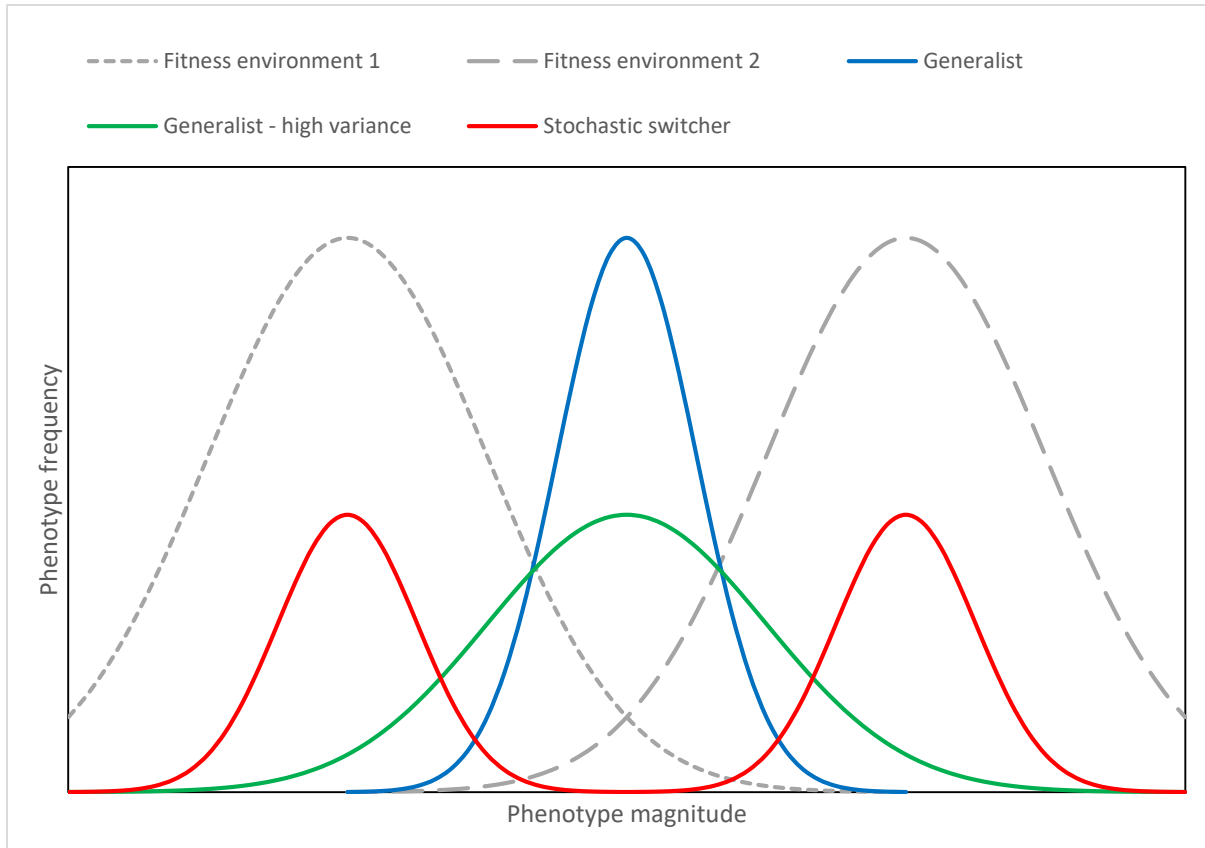
Traversing larger valleys is theorized to be possible due to alterations of the landscape during environmental change and evolution (Steinberg & Ostermeier 2016; Dennett 1995).

2.1.6 **Plasticity, alternative strategies and bet-hedging**

We previously introduced the strategies of generalism and specialism as examples of alternatives to plasticity, and predictions regarding these from the two-environment model of DeWitt & Langerhans 2004 (Introduction **1.1**). Alternative strategies may be favoured when a constraint prevents plasticity evolution, such as its high cost. When fitness sets are more convex than concave generalism may be favoured over specialism due to the relative geometric mean fitness of intermediate and extreme phenotypes. This idea of long term fitness effects over multiple generations helps us understand the fitness benefits of more complex traits such as random phenotypic variance in addition to plasticity, generalism or specialism. Selection for random variance can be considered a form of bet-hedging and is predicted to occur when cues are unreliable or there are biological limits on phenotype production, for example. Further alterations of these strategies can be favoured, for example in a concave fitness set a specialist which bet-hedges asymmetrically only in the environment to which it is not specialised is favoured over a non-bet-hedger and a symmetric bet-hedger (DeWitt & Langerhans 2004).

There is another type of bet-hedging that can be an alternative strategy to plasticity: stochastic switching, which is distinct from the bet-hedging by trait variance we have so far considered. Stochastic switching involves randomly producing distinct phenotypes which are well-adapted to different environments resulting in increased geometric mean fitness; effectively plasticity without making use of environmental cues (**Figure 3**). It is related to bet-hedging by trait variance, in that it could be considered simply as variance with a more complex probability distribution (**Figure 4**) (and it can contain bet-hedging by trait variance within it). However, switching may require distinct mechanisms from variance, and the more distinct the alternate phenotypes are in multiple aspects the less mechanistically and conceptually realistic it becomes to consider them as quantitatively rather than qualitatively different traits. Natural occurrences of bet-hedging by stochastic switching are readily found in microorganisms and often involve gene regulation (Veening et al. 2008). Models of bet-hedging suggest that it is more favourable when environmental change is infrequent (Kussell & Leibler 2005) and when response times are slow (Thattai & van Oudenaarden 2004). When the possibility of a mixed strategy of probabilistic switching between environmental detection (plasticity) and bet-hedging is modelled this is often favoured over any pure strategy, with the probability of bet-hedging increasing with decreasing cue reliability (Arnoldini et al. 2012).

Figure 4. Two different types of bet-hedging strategy phenotype distributions. The fitness functions of phenotype in two environments are shown in grey (the y-axis for these is fitness, not phenotype frequency). The generalist strategy (blue) has a low variance phenotype with a mean between what is optimal in each environment. The higher variance generalist (green) has the same mean phenotype but could hypothetically have a higher geometric mean fitness over many generations experiencing both environments. The other type of bet-hedger is the stochastic switcher (red) which randomly produces two phenotypes which are well adapted to each environment, and therefore has potentially higher geometric mean fitness in changing environments compared to a generalist.



2.1.7 Modelling the evolution of phenotypic plasticity

Phenotypic plasticity researchers have used a wide variety of mathematical and computational approaches in attempts to capture the fundamentals of this phenomena. An ideal framework has not been found, instead different techniques have their own distinct strengths and weaknesses in their scope and realism. I will summarise the logic behind the commonly used approaches, along with their detractions. This should give context to the use of conclusions from specific models when applied to experiments.

The optimality approach encompasses several different mathematical frameworks linked by their aim of finding the dynamics and outcomes of phenotype evolution by following the tendency of a population towards predefined optima for different environments (Leon 1993; Sultan & Spencer 2002; DeWitt & Langerhans 2004). They have the downside of not capturing the genetic underpinning of adaptations but can still be important to delineate when plasticity can be adaptive, without specifically incorporating genetic restrictions. They thus represent an ideal starting point for conceptualizing phenotypic plasticity evolution. Further, their relative simplicity allows productive incorporation of abiotic and biotic factors for added realism (DeWitt & Langerhans 2004; Sih 2004; Stearns & Koella 1986; Krebs & Davies 1996; Savageau 2001), compared with the difficulties of more complex models such as quantitative genetics (Berrigan & Scheiner 2004).

The modelling of phenotypic plasticity using quantitative genetics is a popular approach, for which there are two main mathematical techniques: the character state approach (Via & Lande 1985) and the polynomial function approach (de Jong 1989). Both types of model incorporate statistical measures of genetic variance at the population level to quantify the plasticity that exists as well as the potential of a population to respond to selection for plasticity. The two types of model extend from the same statistics, and do not differ in what they represent from a biological perspective (Van Tienderen & Koelewijn 1994; de Jong 1995; 1999). However, they do offer different advantages depending on the question being asked. Both models share somewhat restrictive assumptions: they assume additive variation over many loci (Crnokrak & Roff 1995), and they assume that genetic variances and covariances remain constant during evolution (Pigliucci 1996).

The character state approach analyses genetic covariance between trait values in pairs of environments. It thus does not explicitly incorporate reaction norms – it makes no direct predictions about what happens in environments outside of those analysed. By using the covariance measure the model quantifies the genetic component of plasticity in a population by incorporating the response of individual genotypes to the environment. Because of its pair-wise approach and lack of an explicit reaction norm this approach is best used for dealing with traits in discrete environments.

The functional approach also makes use of individual trait data to calculate the genetic means, variances and covariances of the trait in multiple environments. These values are then used as the parameters of a polynomial function linking the trait across environments, thus modelling evolutionary properties of the reaction norm. This reaction norm focus, and the incorporation of potentially unlimited environments, makes this approach well suited to continuous environments.

Allelic models look directly at allele frequencies making them more complex than quantitative genetic models, but with less restrictive assumptions and more realistic genetics (Levins 1963; de Jong 1989; Scheiner 1998). By following individual alleles, they are potentially useful for understanding the early evolution of plasticity where only a limited number of loci are involved. Defined loci also allow incorporation of realistic genetic details such as dominance, and biological details such as molecular mechanisms of regulation (Proulx & Smiley 2010). However, most plastic traits will involve many loci potentially limiting this approach (although see Zhivotovsky et al. 1996).

2.2 ENVIRONMENT DEPENDENT GENE REGULATION

The importance of gene regulation as a mechanism of adaptive phenotypic plasticity is well established (Chen & Rajewsky 2007; Connelly et al. 2013; Wray 2007), and this makes it an excellent choice for incorporating into plasticity models and theories to add realism. However, because of its ubiquitous occurrence and fundamental role in cell biology, there is huge scope for studying gene regulation in its own right, from its diversity in nature to theories of its evolution. Before discussing theories and models of regulatory evolution we must first detail its mechanisms, and their possible evolutionary origins.

Gene regulation is ultimately the regulation of the activity of the protein product of the gene (ignoring non-protein coding genes). This can be achieved by altering the concentration of the protein in the area where it is active and altering its activity chemically. The mechanisms that achieve this result are hugely varied. Protein formation is done in steps from transcription of DNA to RNA and translation from RNA to protein. The mechanisms of these processes, or the molecules themselves, can be targeted points of regulation, where the process leading to protein production is stopped or slowed leading eventually to lower protein concentrations.

Control of the mechanisms can depend on other genes which are subject to their own regulation. Thus, a regulatory mutation that affects expression of a gene by affecting regulation of its regulator is known as a *trans* mutation. This is in contrast to a *cis* mutation, which occurs in the DNA proximal to the gene and directly affects regulation, for example in the promoter region upstream of the gene where transcription is initiated and thus controlled. The combination of many genes controlling expression of each other is known as a gene regulatory

network. These can be highly complex and thus allow for diverse systems of regulatory control of genes.

In the case of environmental control of gene regulation, the gene regulatory network will connect to environmental signals. This might involve alteration of a cell surface protein, which then alters another protein inside the cell causing a signal cascade. More simply, some extracellular molecules may be able to enter the cell directly, thus intracellular concentration to some extent reflects extracellular concentration allowing a link to a regulatory element sensitive to this concentration. The signals may connect directly to the protein responsible for a phenotypic response, or they may join the network affecting expression of many regulatory proteins including those involved in the overall phenotypic response. This cellular perspective allows us to see the significant similarities between gene regulation by the extra-organismal environment and by the extracellular environment, which could be composed of many other cells in a multicellular organism. In both cases similar mechanistic and evolutionary possibilities exist for gene regulation and studies of one could be informative for the other.

Evolution of regulation can involve mutations in any of the previously mentioned mechanisms, with changes being made to networks and individual genes, altering how they are regulated by other genes and environmental signals. For a regulatory response to a novel stimuli to evolve it may be necessary to evolve alterations to environmental detection mechanisms or even to evolve entirely novel mechanisms. These are likely to take significantly longer to evolve than the linking of an environmental stimuli which is already detected into a novel regulatory interaction with a gene or network. Novel regulatory network evolution in response to the environment has been found in nature by detecting its signature in phylogenetic studies (Li & Johnson 2010), and environmentally controlled regulation has been evolved by artificial selection (Poelwijk et al. 2011).

We will now look in more detail at some common environmental regulatory mechanisms and how they might evolve. There are other mechanisms which may be involved in environmental regulation, such as relocation and degradation of RNA and protein. There is also the exciting possibility that there are further mechanisms of regulation which have not yet been discovered.

2.2.1 Mechanisms

The mechanisms of regulation can be organised by the stage of protein production at which they occur. Pre-transcription mechanisms include transcription factor binding, chromatin modification and DNA alteration (e.g. methylation). Post transcription mechanisms include mRNA modification (e.g. capping, splicing, poly-A tail addition), translation regulation by RNA interference and other proteins, and mRNA and protein degradation. Many regulatory mechanisms depend to some extent on the DNA directly upstream of a gene, known as the

promoter region. This varies in length in different organisms but is generally in the order of 500bp - 1000bp, usually defined functionally by the region which affects expression (Kristiansson et al. 2009). Other 'enhancer' sites also affect expression via directly interacting with the transcription. These can be up to 1Mbp from the gene but will be physically close to the gene due to folding of the DNA (Pennacchio et al. 2013).

2.2.1.1 Transcription factor binding

The transcriptional machinery of RNA polymerase and its associated proteins (the transcription preinitiation complex) will assemble upstream of a gene prior to transcription. This assembly is affected by various proteins bound to the DNA in this region (the promoter sequence). Some of these binding proteins promote the assembly of this complex (activators), and thus increase expression, whereas some proteins will impede its formation (repressors) and thus decrease expression. The sequences these proteins recognise are known as enhancers and silencers respectively. These proteins, known as transcription factors, bind to specific short DNA sequences of 5-31bp which can be in various positions; often upstream of the gene near where the complex forms in the promoter, but also in introns or much further upstream or downstream in enhancer regions, with DNA folding bringing these factors into positions where they affect the preinitiation complex (Stewart et al. 2012). This is a widespread, principal mechanism of environmentally regulated gene expression (Wray 2007; Rebeiz et al. 2011; Doniger & Fay 2007). The number of binding sites is likely to affect the potential of a promoter to evolve regulation (in the short term) as it affects the number of connections it has to the regulatory network, and thus the number of mutations that can affect regulation (Tirosh et al. 2009). There are theoretical models of how binding sites evolve (Stone & Wray 2001; Mustonen & Lässig 2005), and there are phylogenetic studies elucidating the evolution of binding sites (Chen & Rajewsky 2007).

The nucleotide sequence TATA is a binding site present in many promoters that enhances transcription. Whilst not directly responsible for environmental regulation, it has been found that genes with TATA boxes are more likely to show expression noise, variability over evolutionary time and mutational sensitivity; this has led to the proposal that such genes are more likely to evolve novel regulation, or to evolve it more quickly (Landry et al. 2007; Roelofs et al. 2010; Tirosh et al. 2009).

2.2.1.2 Chromatin modification

Chromatin is the structure in the cell which contains DNA. The DNA strands are wrapped around histone proteins to form nucleosomes, which are then further structured on larger scales. This structuring of DNA has many functions, one of which is regulating gene expression. The chromatin can vary in density, and this affects the ease of access for transcriptional machinery. The density is controlled by enzymes which modify the histone proteins, and by ATP-dependent chromatin remodeling complexes which directly move or remove nucleosomes.

These two processes together constitute chromatin modification. The sequence of DNA also affects how it interacts physically with histone proteins, which provides another mechanism for regulation (Rosin et al. 2012). Chromatin modification has been shown to be involved in environment dependent gene regulation (Dai & Wang 2014; Field et al. 2009) and it has been proposed that the binding pattern of nucleosomes can affect the potential of a promoter to evolve regulation (Tirosh et al. 2009).

2.2.1.3 DNA alteration

Chromatin modification involves alteration of the proteins which package the DNA, but the molecular structure of DNA itself can also be altered in a way which affects regulation. A widespread example is DNA methylation, where a methyl group is added to cytosine or adenine nucleotides. These modifications may affect transcription by directly interfering with the transcriptional machinery, or by recruiting other proteins which affect transcription such as chromatin remodelers. Modifications can be passed on through cell division, and often lead to complete silencing of a gene. This mechanism can be environmentally controlled and thus does represent a form of environment dependent gene regulation (Gilbert 2005; Foust et al. 2016). However, in some species this may not be a relevant gene regulation mechanism, as DNA alteration is not known to occur (for example in *S. pombe* - Capuano et al. 2014).

2.2.1.4 mRNA modification

After a transcript has been produced it can be modified in various ways which will affect how much protein is translated from it, and at what rate. The modifications achieve this by affecting processes such as mRNA degradation, nuclear export, storage and splicing. The alterations are controlled by RNA binding proteins which bind specific sequences on the transcript, thus linking the sequence of the transcript to its regulation. Two common mechanisms of mRNA modification are 'capping', where the five prime end of the transcript is changed to three prime to protect it from degradation, and polyadenylation, where a string of adenine bases are added to the three prime end which also protects the transcript from degradation and can promote translation (Villalba et al. 2011). mRNA modification has been shown to be involved in environmentally controlled regulation, for example in polyadenylation regulation (Liu et al. 2014) and alternative splicing (Cui & Xiong 2015).

2.2.1.5 RNA interference (RNAi)

Various types of RNA molecules affect expression by multiple mechanisms. This includes affecting many of the gene regulation mechanisms we have already mentioned. RNA is involved in the modification of histones, and small RNA strands exist which directly bind and thus regulate mRNA activity, for example by complementary binding, often 'silencing' the mRNA and preventing its translation. Interfering RNAs can also be involved in upregulating genes by binding complementary sequences in the promoter and affecting transcription (Check 2007). RNA interference has been found to be a mechanism of environmental control of

expression, and also has the potential to carry this control across generations (Schott et al. 2014). Interestingly, with regards to plasticity, it also provides a mechanism to respond directly to RNA in the environment, and thus the presence of other organisms (Whangbo & Hunter 2008).

2.2.2 Modelling the evolution of gene regulation

Many aspects of gene regulation have been the subject of theoretical exploration, with corresponding mathematical and computational models. These range from focussing on the details of molecular mechanisms – such as connecting regulatory sequences to expression using thermodynamics (Sherman & Cohen 2012) – to looking at the bigger picture of how regulatory evolution might progress. Much of this work has looked at the optimisation of expression, regulation and networks, using theoretical approaches from diverse fields (Bedford & Hartl 2009; Kalisky et al. 2007; Savageau 2001). Some models focus on specific mechanisms such as transcription factor binding, looking at how regulatory sites evolve and function, and fruitfully connecting this to actual sequence data (Stone & Wray 2001; Mustonen & Lässig 2005; Moses et al. 2006).

Proulx and Smiley (2010) created a model which captures the process of regulatory evolution in response to a changing environment (paralleling the optimisation models mentioned) but also incorporates details of molecular mechanisms and evolutionary genetics. Since this adds considerable complexity, the model deals with the simple case of one gene regulated in two environments, with a minimised number of parameters and assumptions to maintain generality and simplicity of analysis.

The model considers the rates of transcript production (expression) and protein degradation to give a dynamic value of protein concentration. This then determines fitness depending on how close it is to the optimal level for each environment. They then model evolution based on the environment switching randomly, and mutations occurring for changing the transcript production rate in one or both environments. Mechanisms of production regulation are not specifically modelled, but the model would fit many of those above.

As with other models of plasticity, they predict that regulation (plasticity) will evolve providing that constraints can be overcome. The constraints they find, or introduce, include fitness valleys and costs of expression. They also predict that regulation can evolve directly from an ancestral state in some cases, but that evolution can also first lead to an intermediate expression level. When an intermediate expression level is favoured then regulation will also be favoured and reached when the necessary mutations occur.

This model represents an interesting amalgamation of approaches from different areas of theory and modelling. It incorporates a level of molecular realism by including mechanisms of gene regulation and its optimisation, but there is also a genetic and evolutionary underpinning

to the model. It thus makes interesting and testable predictions, to complement those of less specific plasticity models.

2.3 SCHIZOSACCHAROMYCES POMBE AND KEY GENES

The fission yeast *S. pombe* is a single celled eukaryote distantly related to the budding yeast *Saccharomyces cerevisiae*, each originally isolated from brewing. The cell and molecular biology of both species has been studied in detail. *S. pombe* differs from *S. cerevisiae* in many regards, including usually being haploid and dividing by symmetric fission rather than budding. From extensive laboratory study a variety of molecular genetics tools have been developed for use in *S. pombe* investigations. Experiments show that it has significant phenotypic plasticity and environment dependent gene regulation (Chen et al. 2003; Watt et al. 2008; Beaudoin & Labbé 2006; Pelletier et al. 2003). From the molecular tools and experimental results available we believe there are certain genes in *S. pombe* with potential relevance to achieving our goal of experimental evolution of environment dependent gene regulation, which are detailed below and in **Figure 5**.

2.3.1 URA4

The *ura4* gene of *S. pombe* encodes for the URA4 protein, which is the enzyme orotidine-5-monophosphate decarboxylase. This is the same enzyme produced by the analogue of the gene in *S. cerevisiae* – URA3. It is a catalyst of the decarboxylation of orotidine monophosphate (OMP) to uridine monophosphate (UMP), a necessary step in the biosynthesis of the three pyrimidine nucleotides including uridine triphosphate, the substrate for uracil in RNA transcription. Thus, a functional *ura4* gene is essential for biosynthesis of uracil, and mutants lacking it cannot grow without extracellular uracil.

The compound 5-fluoroorotic acid (FOA) is an analog to the orotate group of orotidine monophosphate, with the only difference being a fluorine atom bonded to the pyrimidine ring at position 5. It is also decarboxylated by URA4 producing 5-fluorouracil, a toxic metabolite thought to cause cell death by nucleotide imbalances and thus incorrect formation of DNA, RNA and proteins (Seiple et al. 2006). Thus, *S. pombe* strains with a functional URA4 gene have their growth inhibited or prevented by FOA depending on the concentration.

Deletion of *ura4* causes lysis at stationary phase (the cessation of exponential growth). This is thought to be due to OMP precursor accumulation (Matsuo et al. 2013), which would otherwise be converted to UMP by URA4. Increased intracellular uracil rescues this phenotype. This may be due to uracil being the signal for downregulation of an earlier step in uracil biosynthesis pathway, thus reducing production of OMP. This could be a mechanism of gene regulation by uracil, although a simpler explanation may be uracil competing with orotate in reactions with URA5, creating UMP and reducing OMP production (Nishino et al. 2015).

2.3.2 **FUR4**

Uracil transport from outside the cell is mainly dependent on the uracil permease FUR4 (de Montigny 1998), although there does seem to be other mechanisms to import uracil, at least under higher concentrations (Nishino et al. 2015). The permease does not appear to be downregulated by increasing extracellular uracil, unlike its analog in *S. cerevisiae* (Séron et al. 1999). However, its activity is upregulated during nitrogen starvation by localisation to the plasma membrane (Nishino et al. 2015), suggesting a possible role in importing nitrogen sources under these conditions. The localization appears to be dependent on an unknown deubiquitination mechanism (Nishino et al. 2015).

2.3.3 **URG1, URG2 and URG3**

These three genes show a significant increase in transcription in response to extracellular uracil (Watt et al. 2008). They are analogs to genes in other species thought to be involved in uracil degradation and utilisation as a nitrogen source (Andersen et al. 2008). This suggests that they are upregulated as a mechanism of achieving an optimal intracellular uracil concentration (as this is not achieved by FUR4 regulation, as appears to be the case in *S. cerevisiae* – Séron et al. 1999), or potentially as a mechanism of efficiently utilising uracil (and other pyrimidines) as a nitrogen source. Further support for this comes from the significant upregulation of the genes during nitrogen starvation (Mata et al. 2002; Kristell et al. 2010).

These three genes are co-localised in a cluster on chromosome 1, and interestingly are also in close proximity to *fur4*. Also, in this region is *toe1*, a transcription factor encoding gene which regulates *urg1*, *urg2* and *urg3* as well as *uck2* which is also in close proximity (Vachon et al. 2013). Thus, *toe1* and *uck2* are both thought to be involved in pyrimidine salvage and could be important in nitrogen starvation and uracil concentration regulation. The co-localisation of genes involved in the same pathway is not the norm (Andersen et al. 2008) and could be due to strong selection on co-regulation or on linkage of co-evolved alleles.

2.3.4 **GFP**

The green fluorescence protein of the jellyfish *Aequorea victoria* has been used as a tool for investigating the expression and localisation of proteins in many species, due its function not depending on the species background it is expressed in (Prasher 1995). GFP can be expressed under the control of different promoters in *S. pombe* and can be attached (tagged) to other proteins to show their localisation and expression without affecting function (Bähler et al. 1998). GFP has a major fluorescence peak under exposure to 395 nm wavelength light, which can be applied to measure expression of populations of cells; with microscopy to examine expression and localisation in individual cells; and using precision lasers to sort cells by expression (Fluorescence-Activated Cell Sorting - FACS).

2.3.5 **CUF1 and CTR4**

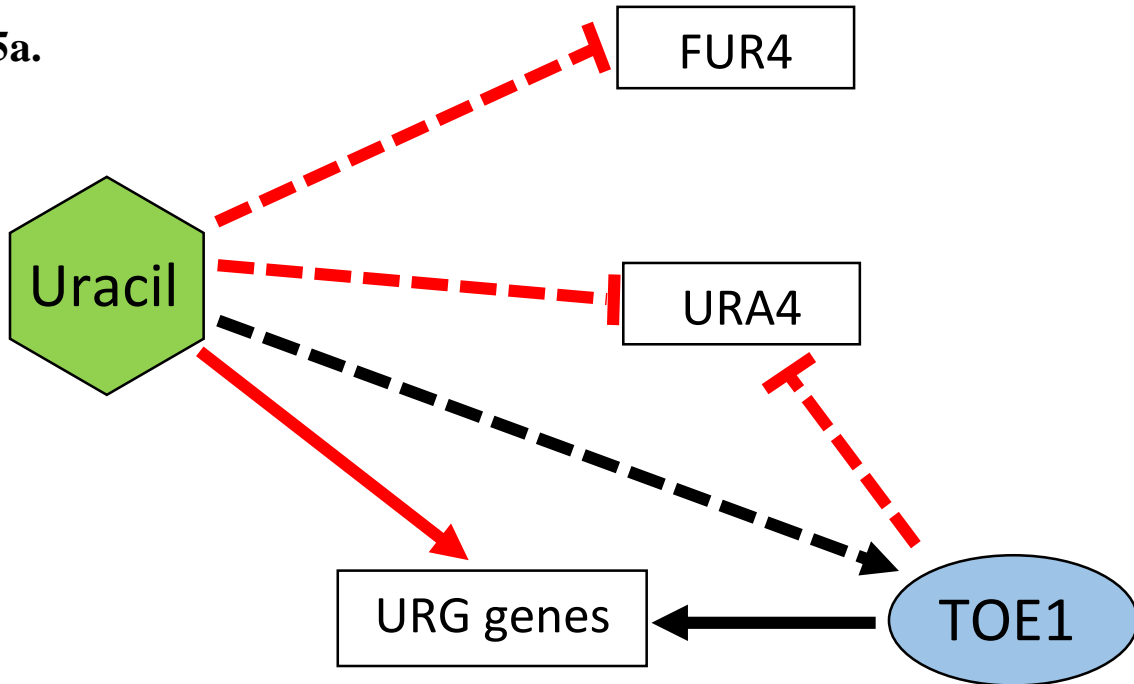
The control of intracellular copper level homeostasis is a clear example of a transcription factor enhancing expression under specific environmental conditions. Under copper starvation the transcription factor CUF1 is localised in the nucleus where it enhances the expression of the copper ion transport protein CTR4, in order to increase copper uptake from the environment (Beaudoin & Labbé 2006). Under conditions of high intracellular copper CUF1 is downregulated by localisation to the cytoplasm. This forms part of a wider cascade of transcriptional responses to copper levels, involving multiple transcription factors (Rustici et al. 2007).

2.3.6 **FEP1 and STR3**

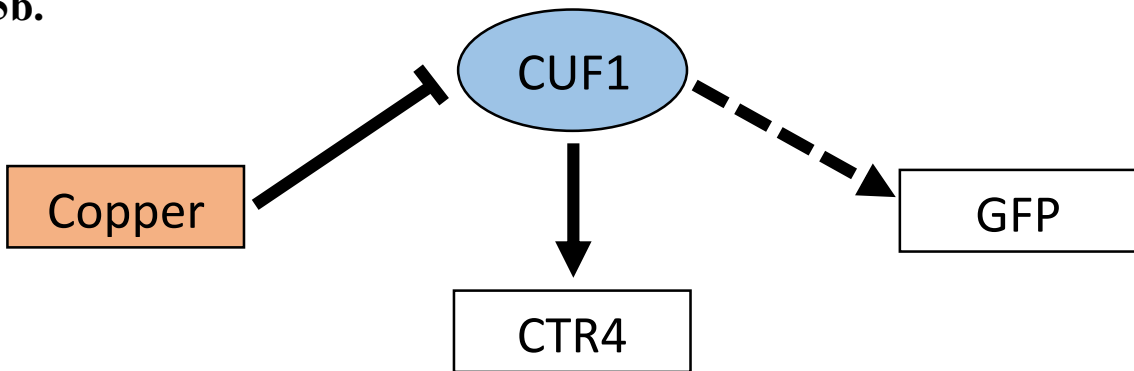
The control of intracellular iron level homeostasis is a clear example of a transcription factor repressing expression under specific environmental conditions. The transcription factor FEP1 is active under high iron conditions. It is a transcriptional repressor which downregulates the expression of iron import genes such as the transporter STR3 (Pelletier et al. 2003; Labbé et al. 2007). FEP1 is inactive under low iron conditions allowing expression of the iron starvation response. As with the copper response, this forms part of a wider cascade of transcriptional responses to iron levels, involving multiple transcription activators and repressors (Mercier et al. 2006; Rustici et al. 2007).

Figure 5. Diagrams showing environmentally controlled regulatory interactions of key genes. Black arrows show direct interactions increasing activity or expression. Red arrows show indirect interactions and have to involve an intermediate factor. Solid lines are known interactions. Dashed connections show regulatory pathways that could potentially evolve to allow optimal expression patterns in response to environmental conditions. Blue ovals are transcription factor genes. Uracil, copper and iron represent environmental concentrations. **5a** shows how the activity of *ura4* could be regulated by environmental uracil concentration. **5b** shows how the expression of GFP could be up-regulated in response to copper starvation. **5c** shows how the expression of GFP could be down-regulated in response to high iron.

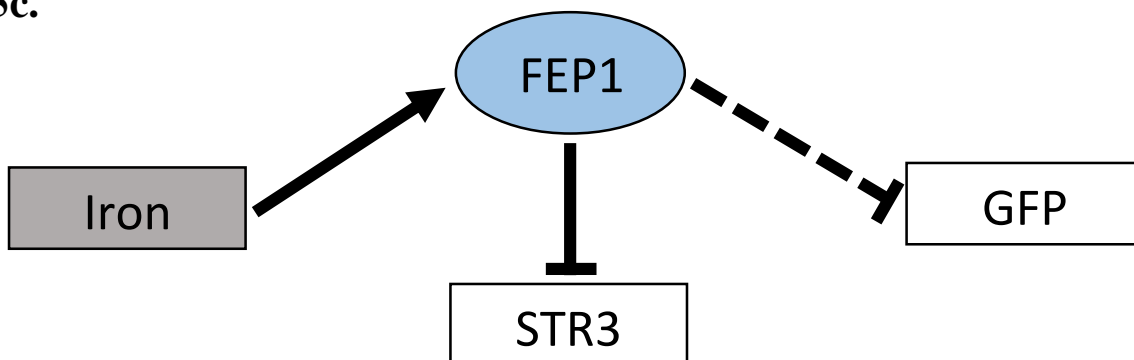
5a.



5b.



5c.



3 AIMS AND OBJECTIVES

The aim of this project is to develop experimental evolution systems that can be used to investigate the evolution of phenotypic plasticity and environment dependent gene regulation. Specifically, by creating a system where two environments select for and against the expression of a target gene, we can select for its regulation by repeated growth alternating between the two environments. Such a system could be used to test theories and models of plasticity and regulation (Introduction 1.3; Discussion 5.3). Our objectives are to take the steps necessary to develop such systems. Below we present these objectives for two systems: a system selecting on the *ura4* gene based on its enzymatic action under different conditions; and for a system selecting on an introduced GFP gene using fluorescence activated cell sorting (FACS).

3.1 *URA4* SYSTEM

3.1.1 Environmental conditions

We aim to grow *S. pombe* in conditions which strongly select for regulation of the *ura4* gene. Thus, our initial objective is to establish the following two environments which select both strongly for and against expression:

Environment ‘no-ura’ (uracil starvation) – Positive selection: The *ura4* gene produces an essential protein in the uracil biosynthesis pathway (orotidine-5-monophosphate decarboxylase). Thus, to produce uracil *S. pombe* requires this gene, and the gene is therefore essential in an environment that contains no uracil. Growth in this environment should select for expression of *ura4*, at an optimal level.

Environment ‘FOA’ – Negative selection: 5-Fluoroorotic acid (FOA) is a uracil analogue which is metabolised by the URA4 protein into a toxin which inhibits growth. Adding FOA to the environment should select strongly against *ura4* expression. Excess uracil must also be present in the environment so that growth can continue without *ura4* expression.

In these environments we also provide the necessary conditions for gene regulation under environmental control, as we know the signal of uracil concentration is already detected by *S. pombe* (2.3; Watt et al. 2008).

Our first objective will be to determine the effects of uracil and FOA on selection using dosing experiments over a range of concentrations. This should be helpful in determining the optimal concentrations to use in experimental evolution, where intermediate growth levels in each environment would be an ideal starting point.

3.1.2 Synthesis of strains

It is key to our experiment to make various genetic modifications to *ura4* and its promoter region. In modifying the protein coding sequence itself, the only modification we will be making is the addition of a GFP tag, so that we can easily track expression of the protein at the cell and population level. All our genetic manipulations will include the addition of a GFP tag. Every modified strain will be sequenced to verify the alterations to its genome are as intended.

3.1.2.1 *Replacement of the native promoter for experimental evolution*

The native *ura4* promoter may be regulated by uracil, as it is part of its synthesis pathway, so we felt it would be sensible to replace its promoter with one which is not related to uracil synthesis. Thus, the purpose of replacing the promoter is to be reasonably certain that *ura4* expression is not affected by our environments at the start of the experiment, and to have expression at an intermediate level, where one environment selects for higher expression, and the other selects for lower. The promoter is not the only part of the genome which could be affecting regulation, but it is likely to be the singular part which can have the largest effects, and by measuring expression levels in both environments we hope to take all reasonable steps to ensure that *ura4* RNA and protein expression will not change between environments in our initial strain.

Our objective will be to choose a number of candidate promoters, insert them in place of the *ura4* native promoter, and determine which is functioning with the desired expression levels for our experiment.

3.1.2.2 *Insertion of a repressible promoter to investigate the fitness landscape*

To test our predictions of how expression changes will affect fitness in our two environments we intend to create a strain with a repressible promoter which can be used to alter expression of *ura4* artificially. As well as verifying the selection pressures in our evolution experiment, this will also allow us to build up fitness landscapes for a range of phenotypes (expression levels) in both environments.

3.1.2.3 *Replacement of the promoter region with a selectable marker*

The sequence of the promoter region of a gene is well established as being a key element in its regulatory control, and thus our objective is to create a protocol which uses a strain inserted with selectable markers in place of the promoter, allowing us to easily alter this sequence. This allows us to see the effects of specific promoter mutations, including mutations that may have evolved during our evolution experiments. It would also allow us to explore the effects of random mutation in this region and introduce genetic variation for selection to act on. Further, we can introduce promoters from a variety of other genes and use these as a starting point for experimental evolution.

3.1.2.4 Creation of a library of mutagenised promoters

By introducing random mutation into the promoter region we will create a library of mutated promoters, allowing us to explore the effects of mutation on the promoter from both a molecular perspective of the direct effects on expression, and an evolutionary perspective of the effects on phenotype regulation and fitness in the two environments. We can also use our mutated promoters as a starting point for evolution experiments. This allows us to see the effects of specific mutations and general variation on the process of evolving regulation, and it also gives us a potential mechanism to speed up evolution if we see that lack of variation is a constraint on our experiments.

3.1.3 Measuring phenotypes and fitness

We intend to investigate the strains we have created, both to verify they are functioning as intended, and to investigate fitness landscapes using altered expression and introduced mutation. We intend to measure fitness by growth rate using a micro-bioreactor and by survivability using serial dilution and colony counting. We will complement this by analysing phenotypes which we assume are the basis of fitness differences, which will also test this assumption. The main phenotypes in this regard are expression levels of the *ura4* mRNA and protein, which can be analysed using quantitative PCR (showing transcription levels) and levels of GFP tags using western blotting or fluorescence (showing both transcription and translation levels).

3.1.4 Preliminary evolution experiment

Our final objective for the URA4 experiment is to carry out a short preliminary evolution experiment where our evolution strain (3.1.2.1) is grown alternately in our two selective environments (3.1.1). We will then collect fitness and phenotypic data from the evolved strains. This will help us establish a growth protocol for future experiments and will provide insight into what we can expect during the early stages of experimental evolution.

3.2 GFP SYSTEM

3.2.1 Synthesis of strains

We intend to select directly on expression of an artificially introduced green fluorescent protein (GFP) gene using fluorescence activated cell sorting (FACS). Thus, the first objective will be to introduce the GFP into a suitable site in the genome, along with *S. pombe* regulatory sequences to produce a constitutive expression level.

3.2.2 Artificial environmental cues

For regulation to evolve we must give the cells environmental information that they are able to detect, which can then be linked into a regulatory pathway for the GFP gene by selection. For the environmental cue we will use the natural response of *S. pombe* to copper and iron and

starvation, where known transcription factors are involved in regulating the transport systems of these metals. When starved of copper the expression activator *cufI* is transcriptionally active (Beaudoin & Labbé 2006). When iron is plentiful (as opposed to iron starvation) the *fepI* expression repressor is active (Pelletier et al. 2003). Therefore, we will select for high expression after growing the cells in an environment low in copper and iron, and for low expression after growth in plentiful copper and iron. This will induce the expression of an activator and a repressor respectively, which should be ideal transcription factors for our GFP gene to evolve high and low expression responses to. There are also many other cellular changes associated with starvation of these metals, including changes in activity of other transcription factors, so there should be a plethora of entry points into these pathways for our GFP gene to evolve (Mercier et al. 2006; Rustici et al. 2007). We will test growth conditions using media which restrict availability of copper and iron and verify the intended effect on expression using qPCR.

3.2.3 Cell sorting protocol

Once the conditions of our two environments and selection have been established we will need to establish a protocol for alternate growth in our two environments, interspersed with selection by cell sorting. We intend for selection to be strong (at least in our initial experiments) so that regulation is strongly favoured and evolves quickly. This means we will only be allowing a low percentage of cells through to the next generation through selection in the cell sorter. However, our population needs to be sustainable through repeated rounds of selection. Thus, we must have a protocol where sufficient growth takes place between selection rounds, and therefore we will determine the population size and growth times needed to establish a workable protocol. The cell sort will produce data on the distribution of expression levels and on the cells which were selected, providing us directly with data on phenotype (expression) and fitness, with fitness being a function of selection based on expression which is entirely under our control via the cell sorting protocol.

Our objective is to conduct two sorts of cells previously grown in the two different environments (starvation and non-starvation media). We will then take the selected cells and verify that they can grow in the alternate environment after selection, thus establishing the basis of our experimental evolution protocol. We will also collect the data from the sort, both to verify the selection parameters are working as intended, and to prepare for analysis of data from future sorts.

4 METHODS AND RESULTS

In this section, methods and results are presented for the development of two experimental systems using the *ura4* and GFP genes as a target for experimental evolution. Both avenues of experimentation have the same goal of using two environments to select positively and negatively on the expression of a single gene in *S. pombe* (see Aims and Objectives 3). Methods are presented in parallel with their results so that the consecutive steps undertaken to develop the required experimental systems can be understood. Important protocols not detailed here can be found in the appendices.

4.1 URA4 SYSTEM

4.1.1 Dosing of uracil and FOA – optimal concentrations for selection

Our initial objective was to understand how our environments would affect fitness, and to choose environments which would create optimal selection pressures. Uracil concentration should have a positive relationship with fitness in a strain with no ability to produce uracil, which partly reflects the expectation that increasing *ura4* expression will also increase fitness, due to increasing production of uracil. FOA concentration should have a negative relationship with fitness in a wild type strain. **Figure 6** summarises how we experimentally verified these relationships.

To find the relationship between uracil concentration and fitness we grew a *ura4* deletion strain at 6 different concentrations of uracil in the ‘biolector’ micro-bioreactor (m2p-labs) which continuously measures optical density to produce growth curves (**Figure 7**). These results show some unexpected anomalies (see Discussion 5.1.2), however the extent of growth at later time points is as expected. We verified this with a further experiment measuring the density of colony forming units after the cultures have entered stationary phase (25 hours growth) in 8 concentrations of uracil (**Figure 6**). This method has the advantage of ignoring dead cells.

The situation of a *ura4* deletion strain may not be relevant to our evolution experiment, as our evolving strain will most likely always be expressing *ura4* to some extent (see Discussion 5.1.2). Therefore, to provide the strongest and clearest selection pressure, we chose a uracil concentration of zero (in which the wild type strain is known to grow well).

ura4 expression under its native promoter is high enough that cells can be very sensitive to FOA, with growth being completely prevented. We therefore grew wild type cells in a range of FOA concentrations below the concentration normally used to completely prevent growth. By doing this we hoped to find a concentration where growth would be slow but still significant. We hoped that at this concentration any lowering of *ura4* expression would translate into immediate fitness benefits (we investigate this below).

We used the biolector to grow cells in six different concentrations of FOA from 0 to 1.2mg/ml for 24 hours. The growth curves produced by this experiment are shown in **Figure 8**. A concentration of 0.4mg/ml significantly reduces growth, so we decided to use this concentration in further preliminary experiments to see if it provided a satisfactory selection level.

Figure 6. Fitness effects of uracil concentration on growth of a *ura4* deletion strain, and of FOA concentration on a wild type strain. In order to ascertain the environmental effects of uracil and FOA on growth we used a *ura4* deletion strain and a wild type strain to measure the growth effects (fitness) of different concentrations of uracil and FOA. Growth (measured by colony forming units at 25 hours) increases with higher uracil concentration (left figure). 25µl was plated from cultures grown at each concentration and used to calculate colony forming units (i.e viable cells) per millilitre of culture. A higher FOA concentration reduces growth (right figure). This figure shows optical density data (i.e. the density of cells and thus growth) from the growth curves in **Figure 8** at the 15 hour time point (the curves being averages of 8 replicates).

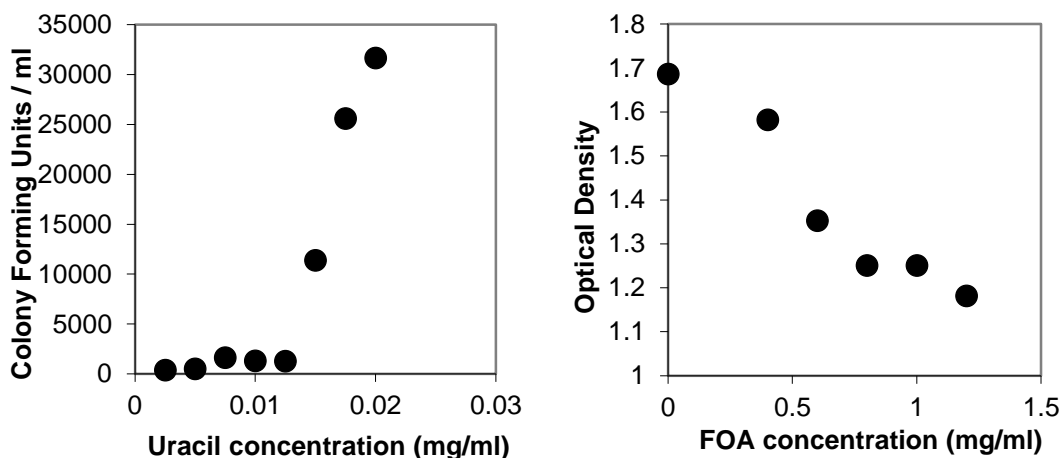


Figure 7. Micro-bioreactor growth curves of a *ura4* deletion strain of *S. pombe* grown in a range of uracil concentrations. Circled in green is a spike caused by cell lysis, and circled in purple is the lag in growth probably caused by uracil metabolism regulation (see **Discussion 5.1.2**). The data for each concentration is an average of 8 replicates. Optical density measurements are taken by the reactor every ten minutes.

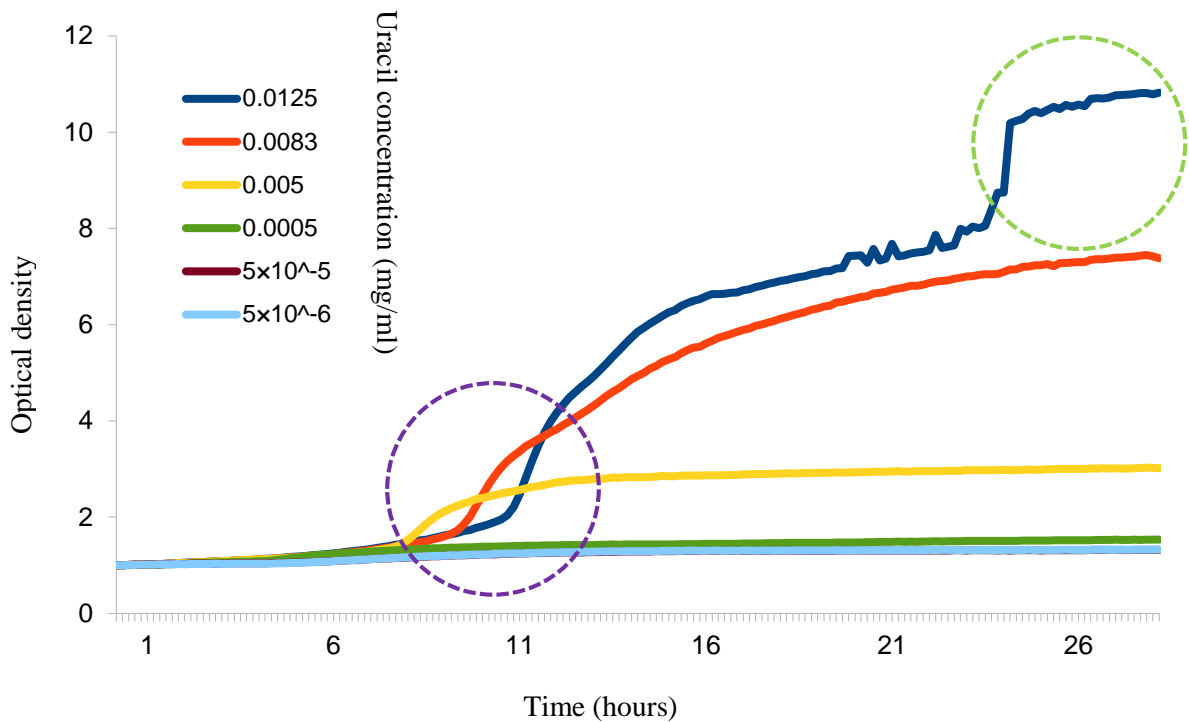
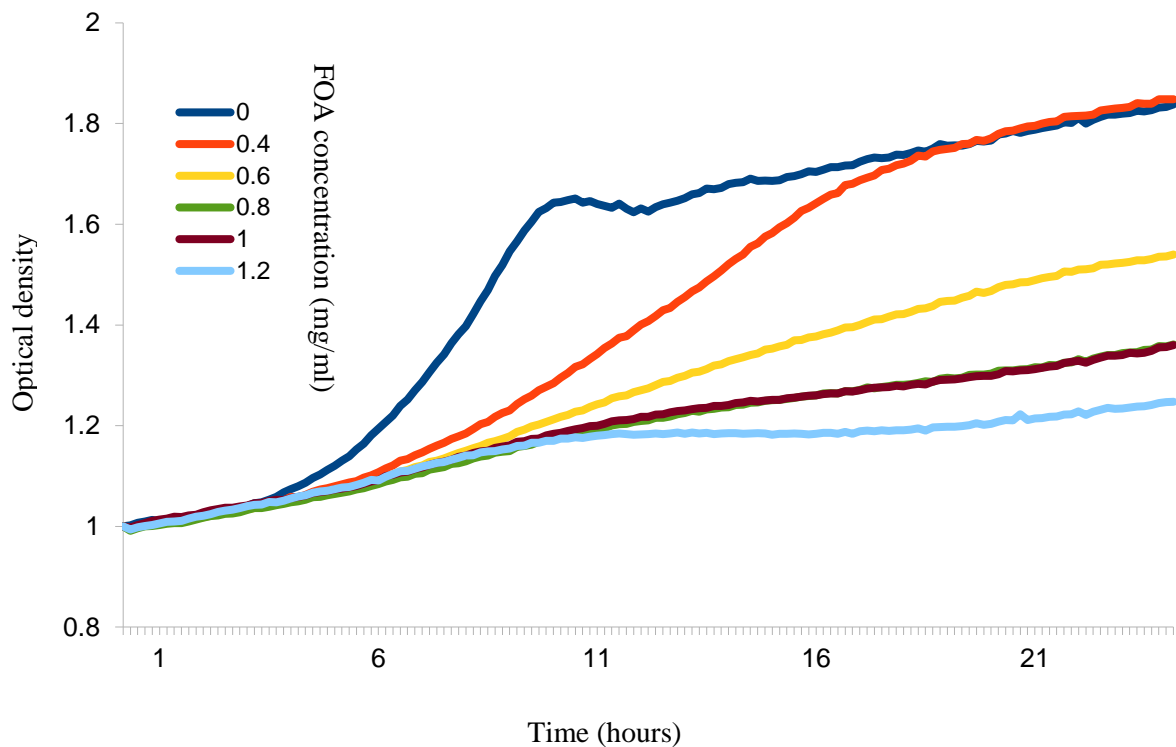


Figure 8. Micro-bioreactor growth curves of a wild type strain of *S. pombe* grown in a range of FOA concentrations. The data for each concentration is an average of 8 replicates. Optical density measurements are taken by the reactor every ten minutes.



4.1.2 Strain construction

In all of our experiments with *S. pombe* we used variations of the reference wild type strain (Leupold's 972 reference strain [JB22] - Leupold 1970). The modifications we made to this strain for all of our *ura4* experiments are detailed here. In many of our transformations we added a GFP tag to the URA4 protein, which serves various purposes including facilitating direct measurement of URA4 protein expression. All transformations were completed using the lithium acetate protocol in Appendix 7.3.

GFP tagging of *ura4*. To measure protein expression of URA4 under the control of its native promoter we tagged the gene with GFP at the C-terminus using the pFA6a-GFP(S65T)-kanMX6 plasmid, as shown in **Figure 9.2** (Bähler et al. 1998).

Relating expression and fitness using the *nmt1* promoter. We used the thiamine repressible *nmt1* promoter to artificially alter expression of *ura4* (See section 4.1.4 for details). This promoter was inserted in place of the *ura4* native promoter using the pFA6a-kanMX6-P81nmt1-GFP plasmid (Bähler et al. 1998). This transformation also included the insertion of a GFP tag onto the N-terminus of URA4 (**Figure 9.3**).

Systems for efficient insertion of multiple promoters. We wanted a system to easily replace the *ura4* promoter, for which we developed two approaches:

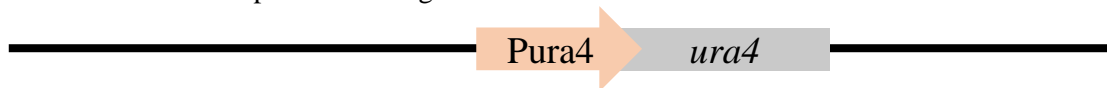
Our first approach used the thymidine kinase (TK) gene as a negatively selectable marker. The *ura4* native promoter was replaced with TK gene and its promoter (Sivakumar et al. 2004 - pFS255 plasmid), along with an N-terminal GFP tag on URA4 (**Figure 9.5**). For this transformation we modified the pFA6a-kanMX6-P81nmt1-GFP plasmid to include the TK gene and its promoter in the place of the *nmt1* promoter (For details see **Figure 10**). The expression of the TK gene makes the strain sensitive to floxuridine (FUdR). This allows us to insert promoters directly in place of the TK gene, with successful transformants being selectable on FUdR media. This method has the potential to rapidly insert many promoters in parallel and was used to insert promoters for experimental evolution (see section 4.1.5).

Our second approach involved first ligating the promoter we wished to insert with a positively selectable marker, and then inserting this construct into the strain. This has advantages and disadvantages when compared to the negatively selectable marker method (see Discussion 5.1.1). To have a *ura4* GFP tag we needed to transform a strain with GFP already in place, so we used the *Pnmt1-ura4* strain described above. Because this strain already contained the kanMX6 positively selectable marker from the *nmt1* promoter insertion we needed to ligate a different marker to the novel promoter we wished to insert. We chose the natMX6 marker, which allows the strain to be selected on nourseothricin 100µg/ml. The method involved PCR amplification of the promoter we wished to insert and the natMX6 marker, with I-SceI restriction sites introduced in the primers. The natMX6 marker (including

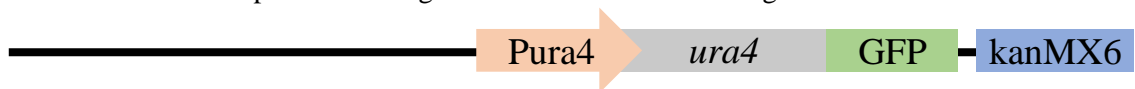
its own promoter) was amplified from the pFA6a-13myc-natMX6 plasmid (Sato et al. 2005). The two PCR products were then digested with the I-SceI restriction enzyme and ligated together. The ligated product was amplified using a further PCR reaction, producing a product ready for transformation (see **Figure 11** for details).

Figure 9. Diagrams of the *ura4* genome region in all the strains used for our *ura4* experiments (see section 4.1.2 for details). Promoters are shown with an arrow. Other boxes show coding regions or entire marker cassettes. Hence not all promoter and terminator regions are shown, for simplicity.

9.1: The native *ura4* promoter and gene.



9.2: The native *ura4* promoter and gene with C-terminus GFP tag.



9.3: The *nmt1* promoter (low expression variant - P81nmt1) and N-terminus GFP tag after transformation into the wild type strain.



9.4: The *nmt1* strain (above) after insertion of a novel promoter using the natMX6 marker.



9.5: The TK-*ura4* strain with the negatively selectable thymidine kinase (TK) marker and GFP tag inserted upstream of the *ura4* gene, ready for insertion of novel promoters.



9.6: The TK-*ura4* strain (above) after insertion of a novel promoter in place of the marker (the incomplete TK allows selection of the successfully transformed strain). The novel promoter is that of *fta5* – this is the strain used in experimental evolution (4.1.5).



Figure 10. Construction of the TK-ura4 strain (section 4.1.2). The intention was to introduce a negatively selectable marker in the place of the native *ura4* promoter, in order to allow efficient future insertion of novel promoters. See figure for further details of the method. Initial plasmids were amplified in an *E. coli* vector.

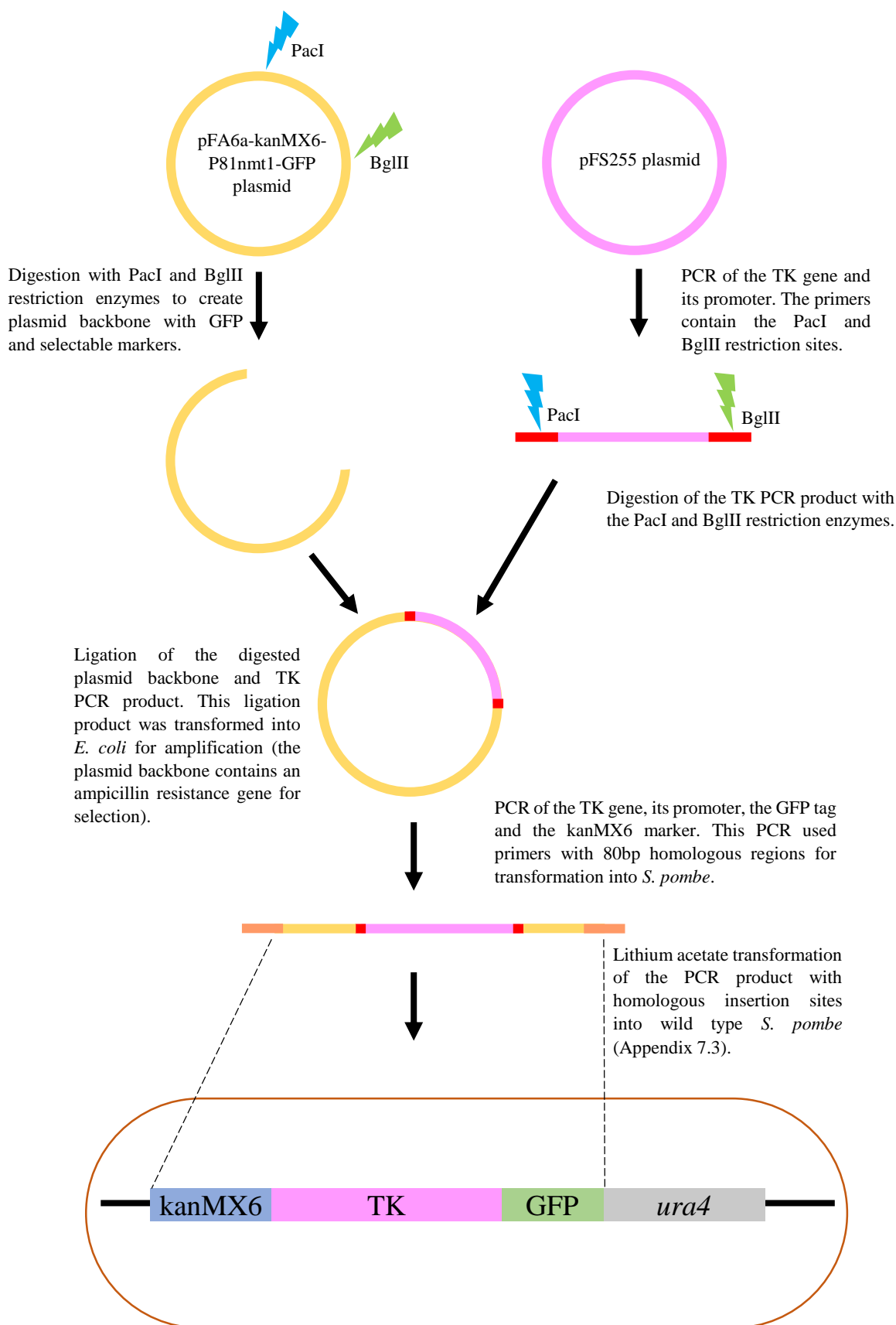
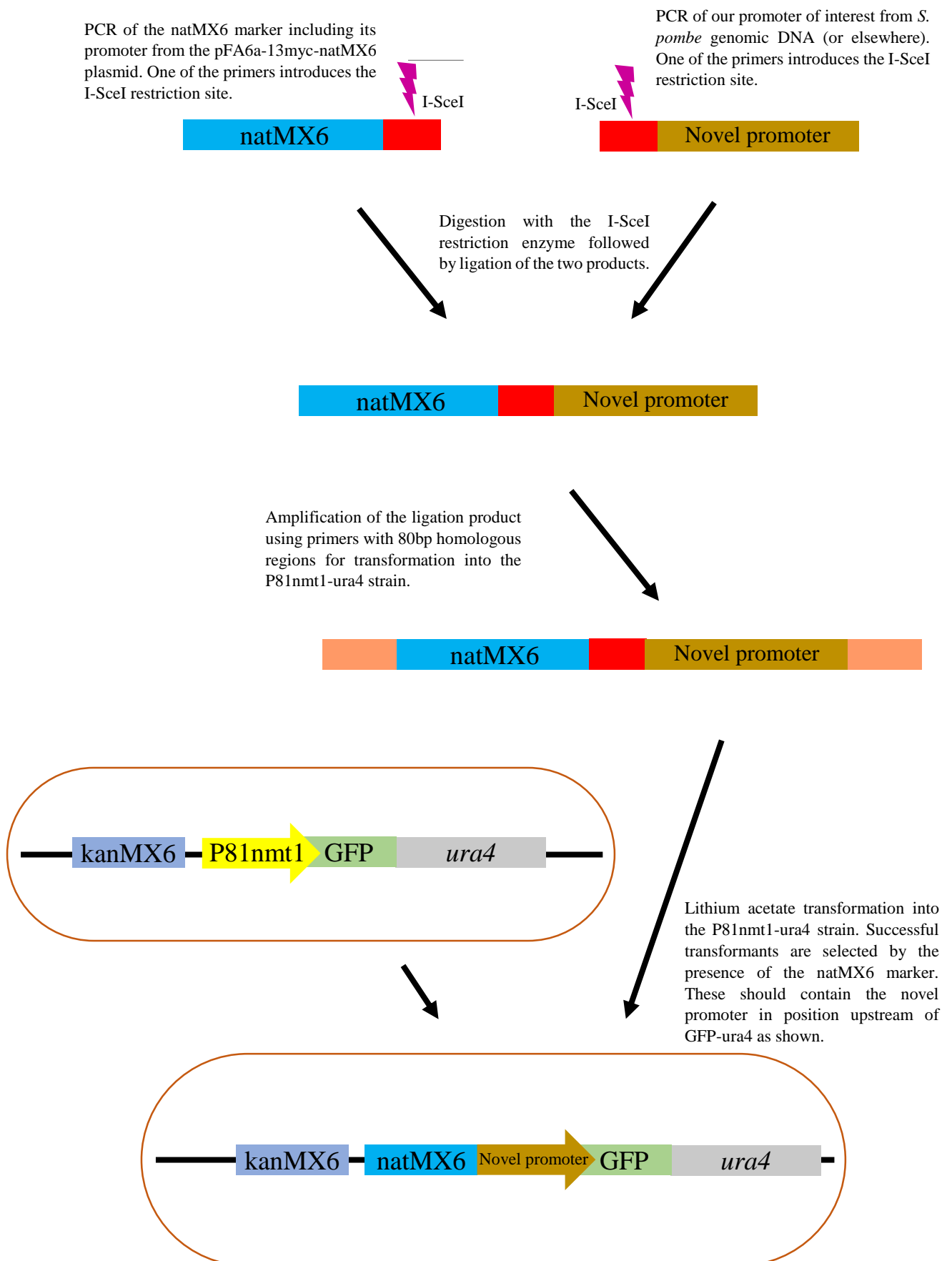


Figure 11. Inserting a novel promoter for *ura4* into the P81nmt1 strain using the natMX6 marker. This is an alternative method for efficiently inserting a promoter for *ura4* whilst having GFP already in place to analyse expression.



4.1.3 Sequencing

Strains from various preliminary experiments were sequenced, either to confirm that transformants matched their hypothetical sequences, or to analyse mutations. We used Sanger sequencing with primers approximately every 300bp. Using the program FinchTV we produced a probable sequence from our chromatograph, and then used ClustalW2 to align the sequence with the hypothetical sequence we intended to produce.

4.1.4 The fitness landscape: Relating expression and fitness in two environments using the *nmt1* repressible promoter

Our experiment relies on fitness and expression having a positive relationship in an environment with no uracil, and the opposite relationship in an environment with FOA (**Figure 2**). We wanted to verify this was the case by controlling the expression of *ura4* in the two environments by putting it under the control of the *nmt1* promoter (section **4.1.2**). This promoter is repressible by growth in different concentrations of thiamine, so we used four concentrations of thiamine to induce different levels of expression (Javerzat et al. 1996). Since thiamine takes time to affect expression, we grew pre-cultures in the thiamine concentrations we intended to use for the experiment, so that expression was consistently at the correct level throughout the experiment. The FOA environment contained FOA at a concentration of 0.4mg/ml which we chose based on the results in section **4.1.1**.

For this experiment we grew 24 hour cultures and then measured growth by serial dilution and colony counting. We controlled for starting conditions by counting cell concentration at the beginning of the experiment and normalising by this, and we ensured that the cells were always in growth phase so that our results were comparable. We used four identical repeat cultures for each environment (**Figure 12**).

The above experiments suggest a fitness landscape which is ideal for our evolution experiment (*cf.* **Figure 2**), where changes in *ura4* expression have contrasting effects on fitness in the two environments. However, the experiments do not allow a direct measurement of expression of the *ura4* RNA or protein. We therefore completed two further experiments to measure expression and relate it directly to fitness, to get a more complete representation of the fitness landscape.

We wanted to confirm that changes in mRNA expression were responsible for the growth changes we were seeing, and that the thiamine concentration was affecting expression as predicted. For this we used quantitative PCR on cells grown in the thiamine concentrations above to determine mRNA levels of *ura4*. mRNA was extracted using an extraction kit (QIAGEN; Appendix **7.4**), and success was confirmed with an agarose gel. We normalised expression by *tub1* using the qPCR protocol in appendix **7.5**. The data is shown in **Figure 13**. The relationship between thiamine concentration and expression is as expected given the results

from the above growth experiment. A *ura4* deletion strain used as a control showed no expression.

As final confirmation of the connection between expression and fitness we wanted to quantify the protein expression using western blot analysis of expression of the URA4 protein. The *nmt1* promoter construct also includes a GFP protein at the N-terminus of URA4 (**Figure 9.3**). We therefore used a GFP antibody to measure protein expression. We could not measure URA4 expression in this way in our wild type strain, so we used a strain with GFP introduced at the C-terminus of URA4 (**Figure 9.2**) with mRNA expression controlled by the native *ura4* promoter. We extracted protein from cells of the above cultures used to analyse mRNA expression (Appendix 7.6). We then used a western blot to analyse GFP protein levels. **Figure 14** shows the results in two separate exposures. Again we see the expected relationship between thiamine concentration and expression, and thus fitness.

Figure 12. Fitness of the P81nmt1-ura4 *S. pombe* strain under two opposing selection pressures. The expression of *ura4* in this strain can be increased by reducing the concentration of thiamine. The environment with a lack of uracil selects for high expression and is shown in red; the highest fitness measured here is close to that of growth under normal conditions, showing that any increase in expression above this will not significantly affect fitness. The environment containing the FOA compound selects for low expression and is shown in green. Fitness relationships with expression are as expected and required for our evolution experiment (**Figure 2.**). Fitness is measured by live cell concentration at 24 hours by serial dilution and counting of colony forming units (CFU). Error bars are standard error from four replicates.

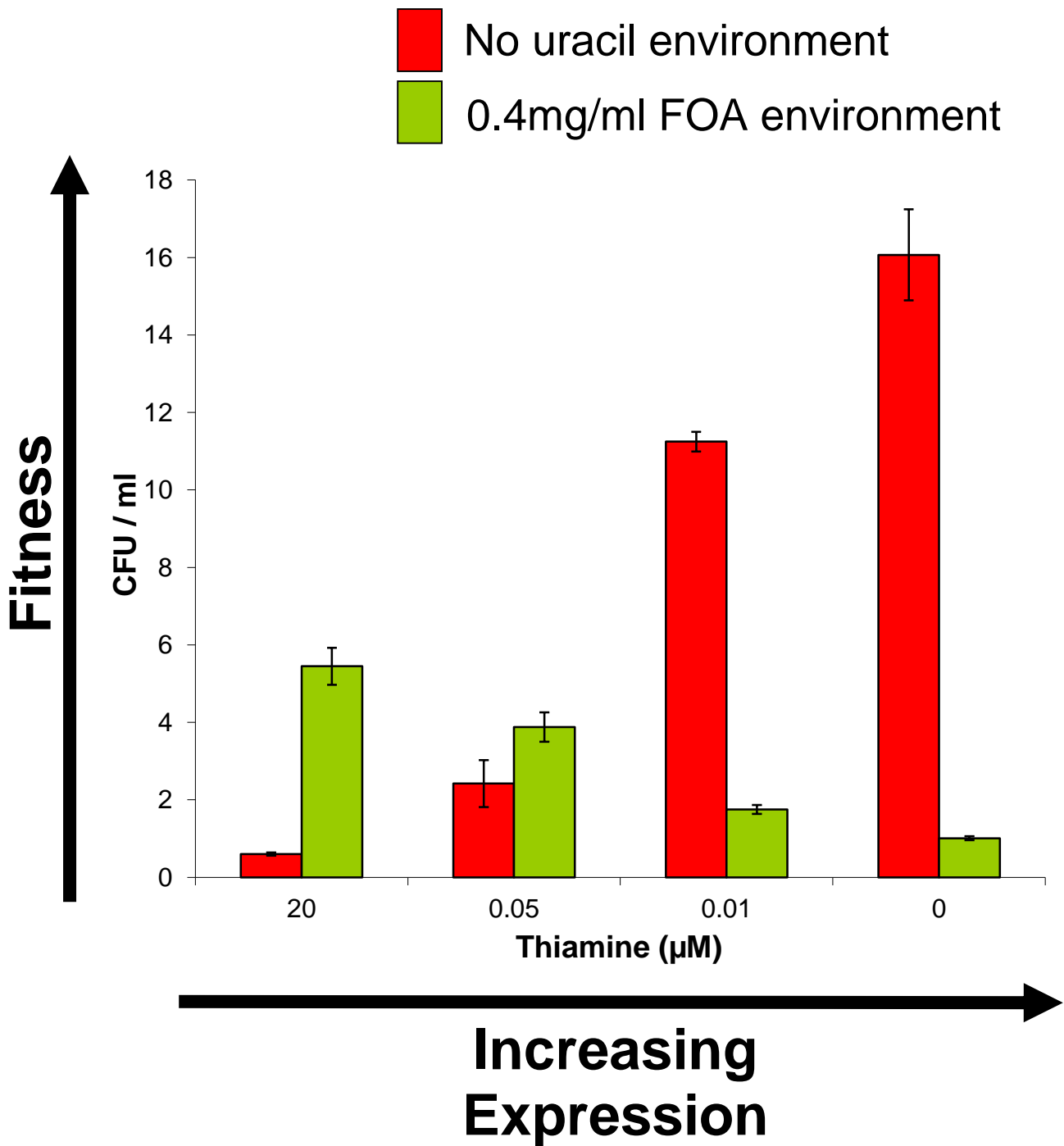


Figure 13. Quantitative PCR of the *ura4* gene under the control of the *nmt1* promoter in a range of thiamine concentrations. Data shown is an average of two replicates. Thiamine is repressing expression as expected. There was a problem with the qPCR reaction for the culture without thiamine (0mg/ml) so this must be repeated. However we know from other experiments (such as the western blot experiment in **Figure 14**) that expression is likely to be high at this concentration, as expected.

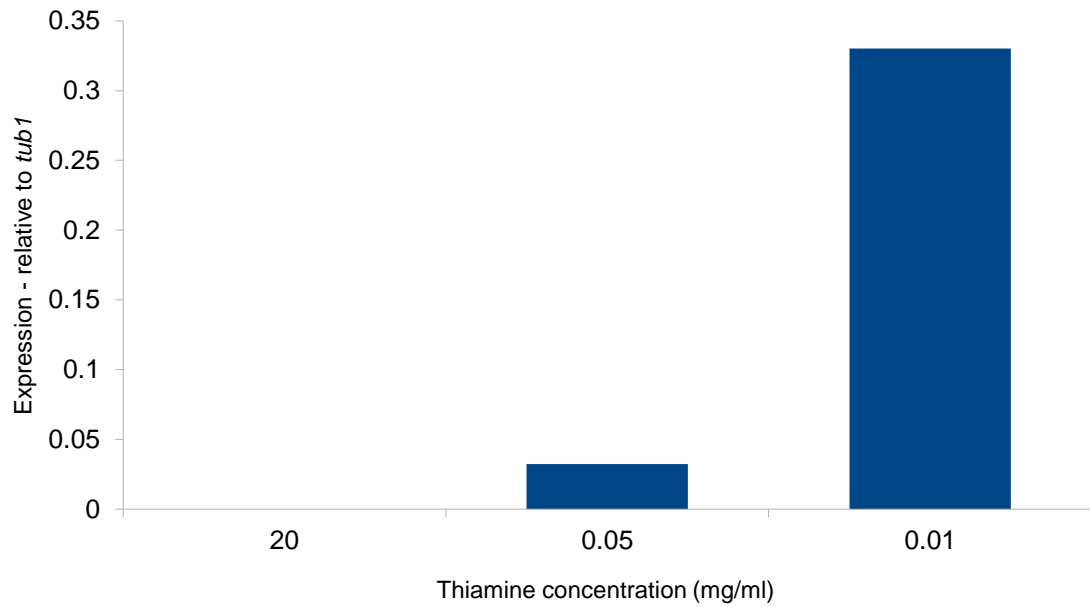
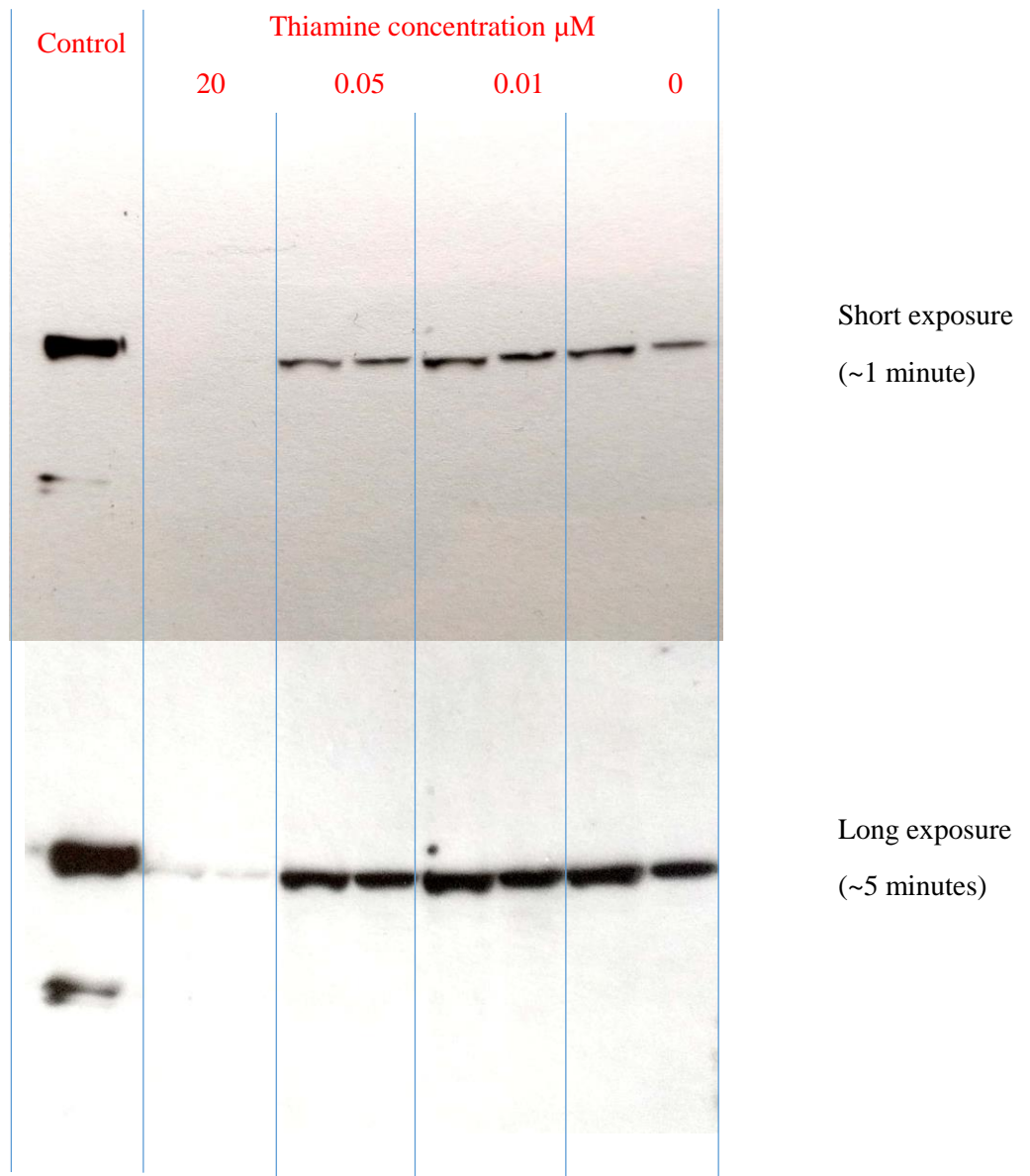


Figure 14. Western blot analysis of URA4 protein expression (using antibodies to the GFP tag) under the control of the *nmt1* promoter in a range of thiamine concentrations. The same gel is shown twice, with different exposure times. Each thiamine concentration has been run twice, with double the amount of protein extract ran on the left. On the far left is a control strain expressing URA4-GFP at a high level under the control of the native *ura4* promoter. In the short exposure we cannot see any expression for 20mg/ml thiamine, and we see lower expression for 0.05mg/ml. In the long exposure we can see small amounts of protein for 20mg/ml thiamine, showing that repression is not complete. These results agree with our expectations and other experiments.



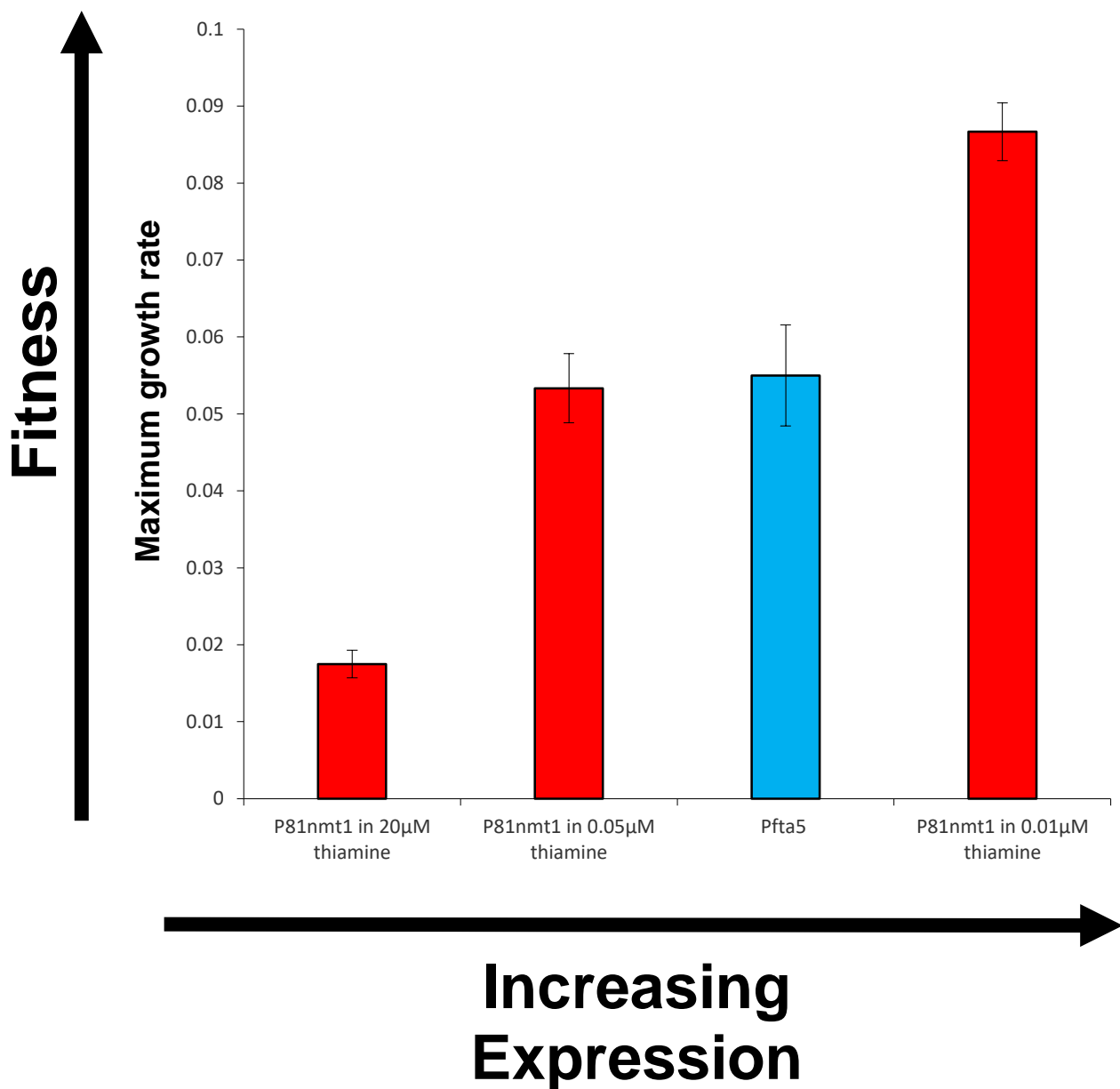
4.1.5 Choosing and transforming a suitable promoter from experimental evolution

For our experiment to be a success we required a promoter that is not regulated in our environments, and has constant (constitutive) expression at an intermediate level. The first step was to find this expression level, which gives intermediate fitness in both environments. Looking at the data from our experiments with the *nmt1* promoter (**Figure 12**) we can see that intermediate expression is significantly lower than that of the native promoter (native promoter data is not shown, but it was found that growth was high without uracil and low with FOA). We used data from a genome wide expression analysis to find genes with significantly lower expression (mRNA count per cell) than *ura4* (Marguerat et al. 2012). We wanted genes which had constitutive levels of expression, so we used genome wide data for expression in various conditions to select genes which did not appear to vary much in expression (Mata et al. 2002; Chen et al. 2003; bahlerlab.info/resources – Gene Expression Viewer). We then reviewed gene ontologies from pombase.org and the literature regarding these genes to be as certain as possible that their expression was constitutive and that they were not likely to be regulated by uracil. The three most promising candidates were *clp1*, *lub1* and *fta5* (also known as *pfl8*).

The promoters of these genes are not accurately defined, so we had to analyse their genomic regions individually to decide on which bases to include. Promoters in *S. pombe* are probably on average less than 829 bases in length and it is likely that most regulatory information will be contained within around 1kbp, although most information will be much closer to the transcription start site (Kristiansson et al. 2009; Lee & Young 2000). For the *fta5* promoter we took 650bp upstream of the coding sequence.

We used the transformation protocol (Appendix 7.3) and the TK-ura4 strain to insert the promoters of the three candidate genes in place to promote the *ura4* gene (**4.1.2; Figures 9.5 & 9.6**). Both the *clp1* and *lub1* promoters had high enough expression to not show any detectable fitness reduction in an environment without uracil. The *fta5* promoter has the lowest expression of the three, and it showed an intermediate level of fitness when compared to the results of the *nmt1* experiment (**Figure 15**). It is the promoter of a cell surface glycoprotein involved in cell-cell adhesion during flocculation (Kwon et al. 2012) which is constitutively expressed and not likely to be related to uracil metabolism. After sequencing we were satisfied that the strain was ready for use in evolution experiments.

Figure 15. The fitness of the *Pfta5-ura4* strain compared to that of the *P81nmt1-ura4* strain in varying thiamine concentrations in an environment that does not contain uracil. The decreasing thiamine concentration increases the expression of *ura4* under the control of the thiamine repressible *nmt1* promoter, and thus fitness increases with increasing expression of *ura4* as this increases production of uracil (see section 4.1.4 and Figure 12, 4 replicates). The strain with *Pfta5* controlling *ura4* expression has intermediate fitness, indicating intermediate expression, as required for experimental evolution. Fitness is approximated as the maximum growth rate which is calculated from optical density growth curves obtained from the biolector (5 replicates). All error bars show standard error.

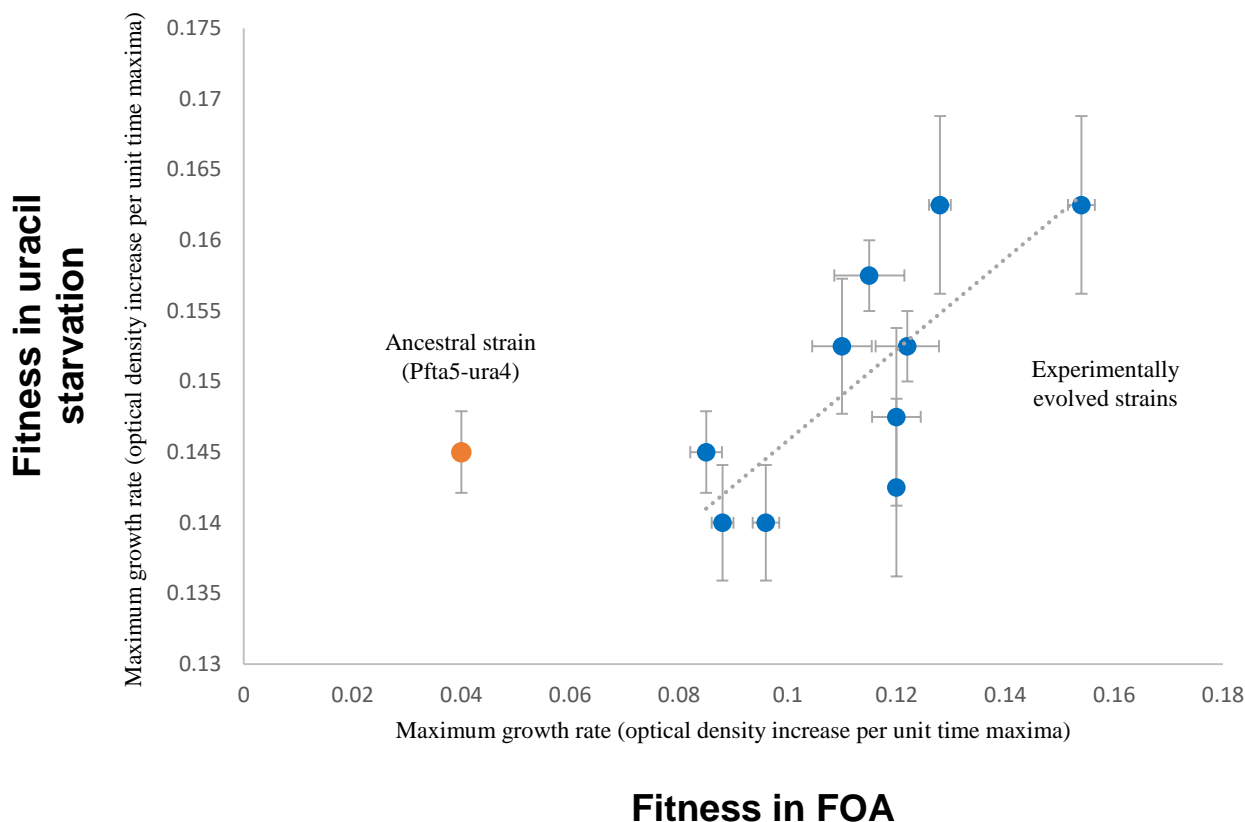


4.1.6 Preliminary evolution experiment

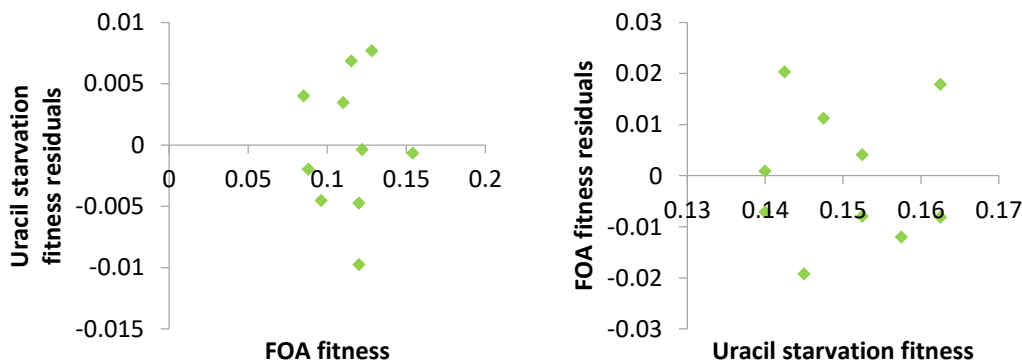
We completed a 10 day preliminary experimental evolution run using our Pfta5-ura4 strain. 10 separate cultures were grown in environments alternating approximately every 24 hours. Thus the lines went through 5 full cycles of a changing environment. Cultures were grown in 60ml culture tubes with 4ml of media for growth. Cells were centrifuged and resuspended in the appropriate media with each environmental change, in order to keep the initial cell concentrations the same (for full methodology see Appendix 7.2).

After 10 days the fitness of the evolved strains was measured in both environments in the biolector by calculating maximum growth rate from the growth curves. There were a variety of changes in fitness relative to our ancestral strain in both environments (**Figure 16**). Fitness increased in all strains in the FOA environment, and in many strains in the uracil deprived environment (Two sample, one-tail t-tests: $p < 0.05$). There is also a correlation in fitness between the two environments in the evolved strains (Linear regression analysis $p < 0.01$).

Figure 16. The fitness of 10 experimentally evolved strains after 10 days alternating growth between uracil starvation and FOA, compared to the fitness of the ancestral strain. The two fitness measures from both environments are plotted against each other, showing the positive correlation in fitness in between environments. A linear regression analysis gives a probability of 0.01 that the regression coefficient is zero, and the residual plots (below) show no clear trends, thus we can be fairly confident that high fitness in one environment predicts high fitness in the other. Fitness is approximated by maximum growth rate calculated from optical density measured during 24 hours growth in the biolector micro-bioreactor. Data is from 4-5 replicates with standard error shown.



Residual plots for linear regression analysis:



4.1.7 **Creating a library of randomly mutagenised promoters and analysing their effects on fitness and expression.**

We assume a high likelihood that evolutionary change in expression and regulation will occur with mutations in the promoter region. In the evolution experiment we allow mutations to occur at their normal rate, which provides the raw material for selection to act on. As a parallel to this we can also introduce random mutations into the promoter artificially, by creating synthetic promoters using error-prone PCR and then inserting them in position to promote *ura4*. We wanted to explore the initial mutations that could occur in our evolution experiment, so we introduced mutated *fta5* promoters, producing a library of strains which were effectively mutated versions of the ancestral Pfta5-ura4 genotype.

To create our library of randomly mutated *fta5* promoters we used an error prone PCR of the promoter from the wild type (Appendix 7.7; McCullum et al. 2010). The conditions we used for this PCR should produce unbiased mutations, and by changing the conditions of the reaction we can change the mutation rate (Wilson & Keefe 2001). We aimed for an average mutation rate of 1 per 300 bases. This ensures a good proportion of promoters with just one or two mutations within the first 300bp, where the majority of mutations with strong effects on regulation are likely to be found. This should give us an idea of fitness effects in the mutation space close to the wild type promoter, and hopefully reveal some beneficial mutations, possibly causing regulation.

We created seven strains with randomly mutated promoters. A single error prone PCR reaction produced the mutated promoters. These were amplified again with homologous primers for insertion into our TK-ura4 strain (4.1.2). Successful transformants should contain a random selection of mutated promoters, and should have effectively the same genotype as the Pfta5-ura4 ancestral strain, apart from the mutations (Figure 9.6). The promoters were sequenced and found to contain a higher rate of mutation than predicted, with most of the promoters having 4 or more mutations in the first 300bp (Figure 17). We measured the fitness of these strains relative to the ancestral *Pfta5-ura4* strain as we did with our experimental evolution strains (Figure 18). There are significant differences in fitness in both environments in some strains (Two sample, one-tail t-tests: $p < 0.05$), but many of the strains do not differ greatly from the wild type. The average fitness of the mutants does not appear to differ from the wild type (One sample, two-tail t-test not significant).

To further understand how the mutations in these strains are affecting fitness we looked at their *ura4* expression. Looking at both mRNA and protein expression (Appendices 7.5; 7.6) we see some variation in expression, with strain 4 showing increased mRNA expression and strain 3 showing increased protein expression, relative to the other mutants (Figures 19 and 20).

Figure 17. Sequence alignment to the wild type *fta5* promoter of 7 mutated promoters inserted to promote *ura4*. Sequences were aligned using blast (bl2seq), and are numbered to correspond with the fitness data in **Figure 18**. The base position starting from the beginning of the promoter is shown (the point where it meets the coding sequence of the gene). SNPs are marked in red.

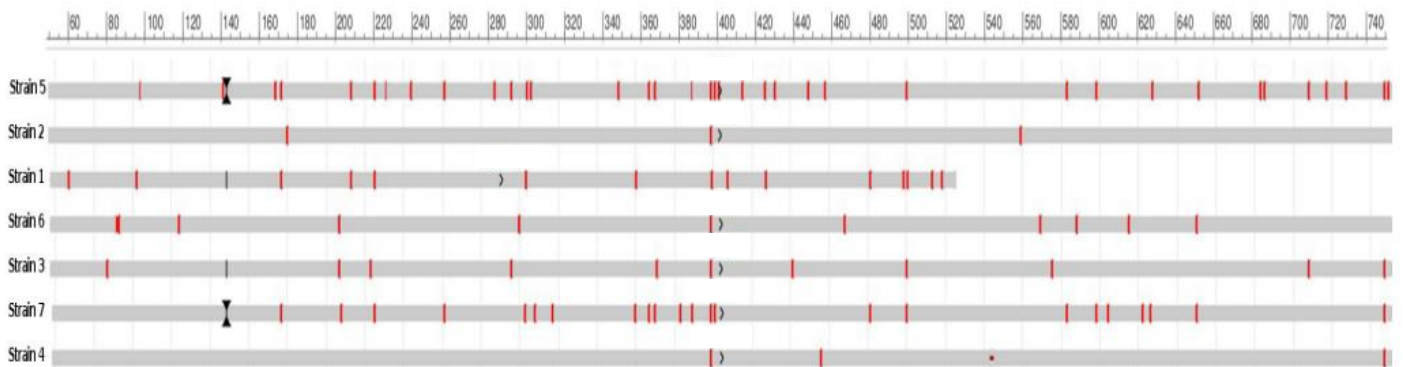


Figure 18. The fitness of seven Pfta5-*ura4* strains with mutated promoters compared to the fitness of the ancestral strain in the two environments of our evolution experiment. The data points are numbered to correspond to the promoter sequence alignments of the strains in **Figure 17**. Fitness is approximated by maximum growth rate calculated from optical density measured during 24 hours growth in the biolector micro-bioreactor. Data is from 4-8 replicates with standard error shown.

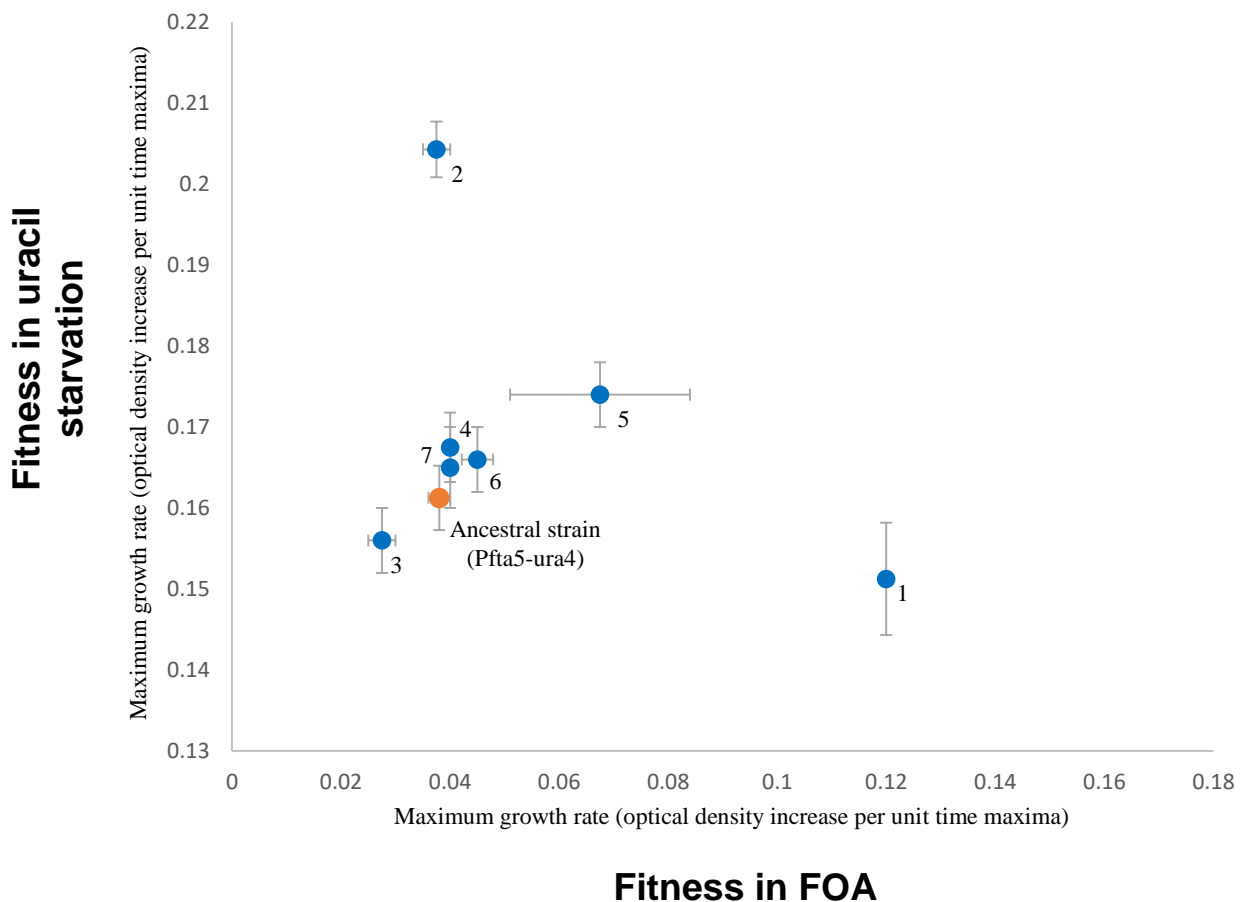


Figure 19. Quantitative PCR of the *ura4* gene under the control of 7 mutated *fta5* promoters, with wild type and *ura4* deletion controls. Expression is measured relative to *tub1*, and each strain has been measured twice (shown in blue and red). Strain 4 has higher expression than the other mutants, but the rest of the variability is not conclusive, and more accurate methods or more replicates are required to say if and how expression has changed in these strains.

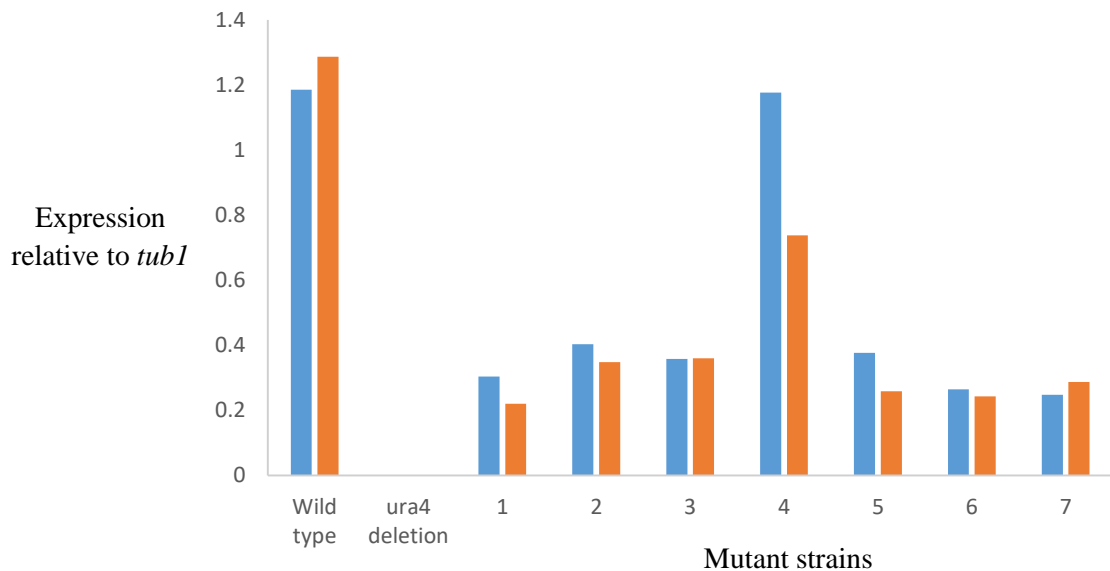
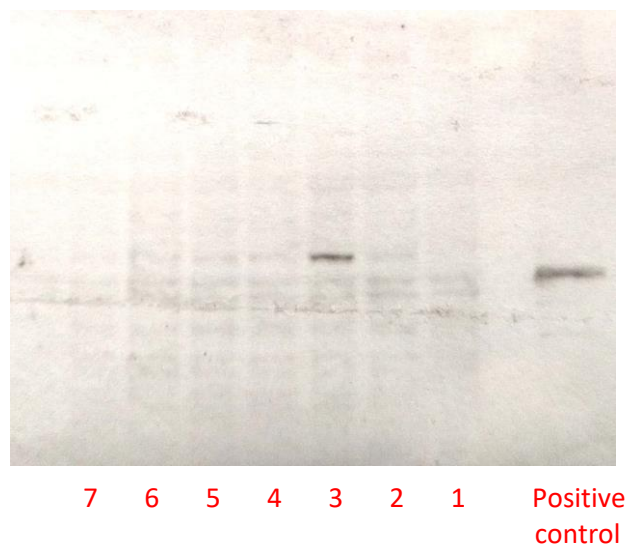


Figure 20. Western blot analysis of URA4 protein expression (using antibodies to the GFP tag) under the control of 7 mutated *fta5* promoters. This method does not allow us to discern small variations in expression, but we do see expression in all of our strains. There is increased expression in strain 3, which does not follow the pattern of mRNA expression, where strain 4 is the only one with increased expression (**Figure 19**).



4.2 GFP SYSTEM

The aim of this experiment is to use fluorescent-activated cell sorting (FACS) to select for regulation of an introduced green fluorescent protein gene (GFP). Using cycles of growth and selection of the highest and lowest expressing cells (alternately) we intend to select for regulation in response to environmental cues (see Aims and Objectives 3.2).

4.2.1 Strain construction

It is important for both cell sorting and evolution that the GFP gene is introduced under the control of a promoter with constitutive, intermediate expression. We decided to use the P41nmt1-GFP construct from the pFA6a-kanMX6-P41nmt1-GFP plasmid (Bähler et al. 1998) as its transformation into *S. pombe* is well established. The *adh1* terminator is part of the construct to ensure proper expression, and the kanMX6 selectable marker appears upstream in the construct to allow selection of successful transformants. The construct was inserted into the *ade6* locus of the genome, replacing the *ade6* gene and promoter (Appendix 7.3). This is a commonly used locus for insertion, which should ensure proper expression. It also allows for selection of strains with the construct inserted at the proper location (following selection for the kanMX6 marker), as colonies with an *ade6* deletion appear pink on low adenine plates (due to accumulation of a red precursor – Moreno et al. 1991). *ade6* deletion does not affect growth if excess adenine is present in the media, which will be the case for our experiment. After constructing our strain we confirmed that it expressed GFP using qPCR (Figure 21; Appendix 7.5) and a western blot (Appendix 7.6).

4.2.2 Growth in a detectable environment

For regulation to evolve we have to give the cells a signal which they are easily able to detect, which can then be linked into a regulatory pathway for the GFP gene by evolution. We decided to use the natural response of *S. pombe* to copper and iron and starvation (Aims and Objectives 3.2.2).

Environment 1 – followed by FACS sorting for low expression: The minimal media we are using (EMM) contains excess copper and iron ions due to the inclusion of CuSO₄ (0.04mg/l) and FeCl₃ (0.2mg/l). Thus these metals are plentiful, and the *cuf1* activator should be inactive, with the *fep1* repressor being active.

Environment 2 – followed by FACS sorting for high expression: An environment of iron and copper starvation is induced by adding chelators of copper and iron ions to the minimal media: 100µM bathocuproine disulphonate (BCS) and 300µM ferrozine respectively. This should produce an increase and decrease of activity in *cuf1* and *fep1* respectively.

To be certain that transcriptional activity was changing between our environments as predicted we analysed the expression of *str3* and *ctr4*, which are known to be transcriptionally regulated by *fep1* and *cuf1* respectively (Pelletier et al. 2003; Beaudoin & Labbé 2006). Five candidate

strains for the evolution experiment were selected from the GFP transformation (section **4.2.1**), and were analysed by qPCR (**Figure 22**; Appendix **7.5**). From these we see that after 6 hours growth in copper and iron starvation elicits the transcriptional response we had intended. However, at 24 hours the *ctr4* response seems to be lost. Strains 1 and 5 show the most promising patterns of expression, and are thus good candidates for our evolution experiment. We also measured expression of *fep1*, *cuf1* and GFP, which were as expected (see **Figure 21** for details).

Figure 21. Relative expression of *cuf1*, *fep1* and GFP (mRNA) after 6 and 24 hours growth in two environments. The strains and conditions are those described in **Figure 22** and expression is relative to *tub1*. Each bar shows an average of two replicates. The expression of the transcription factors *cuf1* and *fep1* does not seem to be altered by the environments, except for some strains after 24 hours of growth. This is as expected, since it is the activity of these transcription factors that is known to change, not necessarily their expression, and this can be seen in the expression of their targets *ctr4* and *str3* (**Figure 22**). However the expression of *fep1* in strains 1, 3 and 5 does change significantly after 24 hours growth, so it would seem that *fep1* itself is downregulated by prolonged starvation, with strains 2 and 4 most likely entering stationary phase and thus displaying expression patterns not typical of normal growth. The downregulation of the *fep1* repressor corresponds with the expression of its target *str3*. GFP is expressed in all strains and does not appear to be regulated, as required for our experiment.

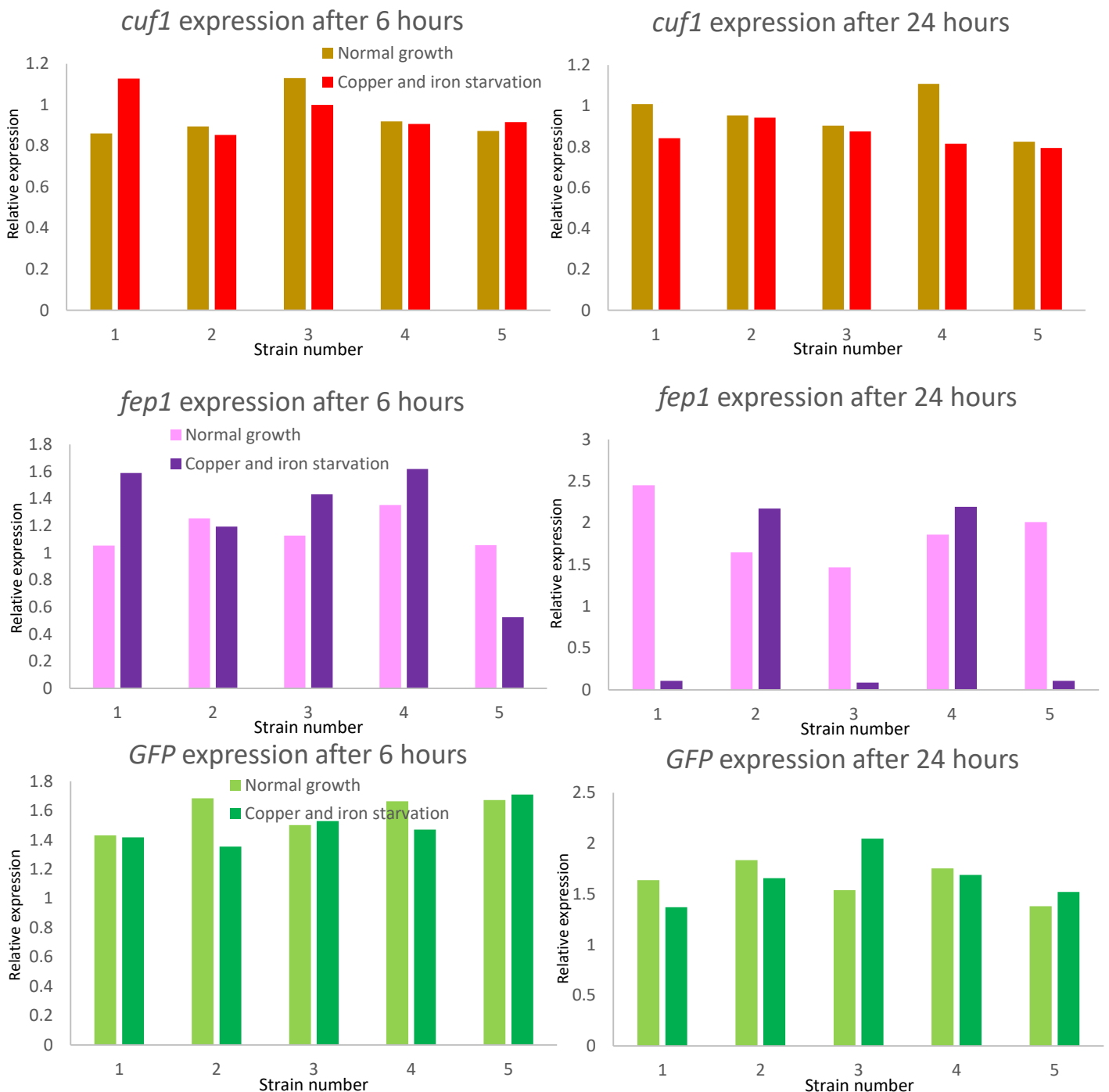
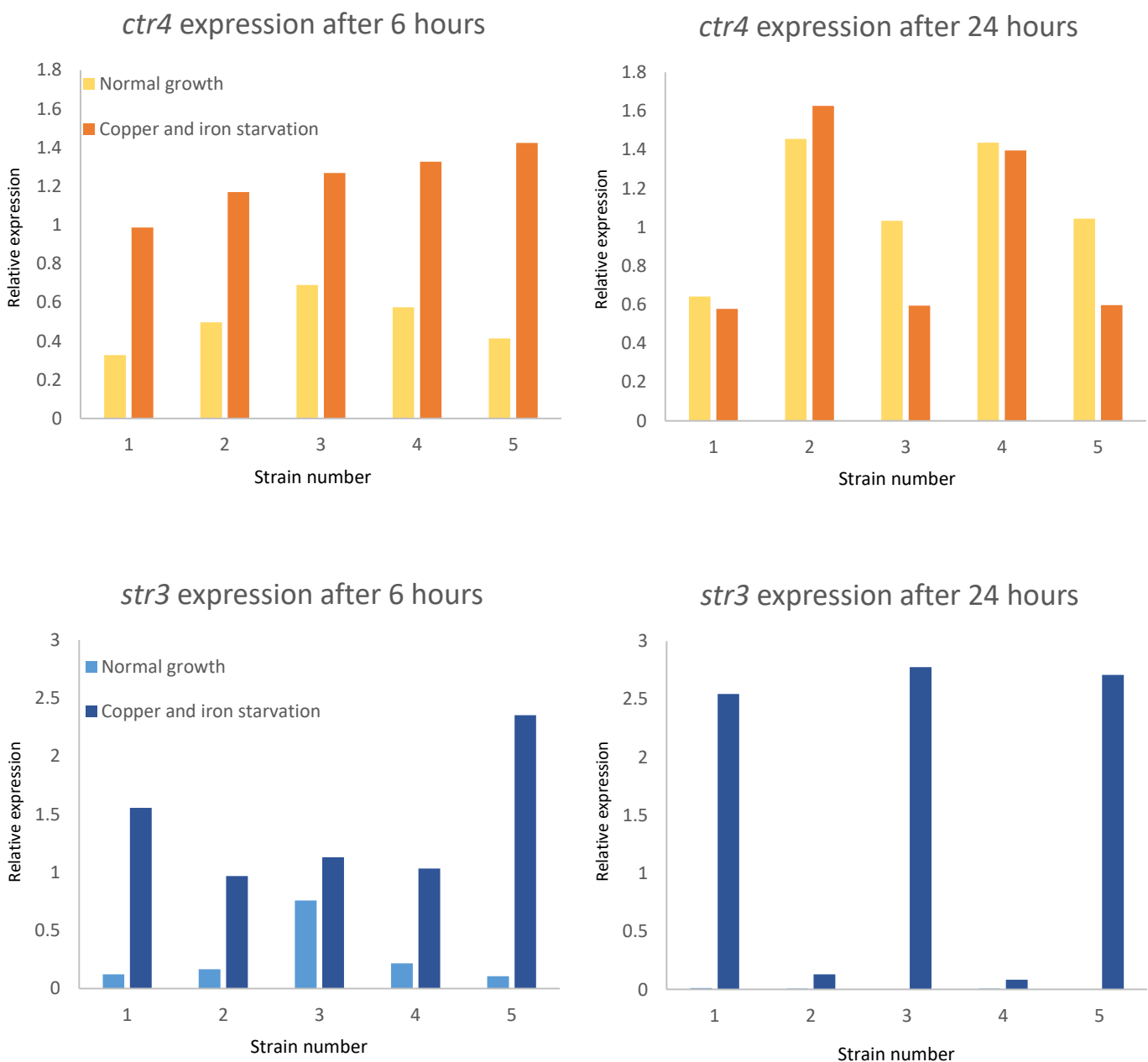


Figure 22. Relative expression of *ctr4* and *str3* after 6 and 24 hours growth in two environments. Each bar shows an average of two replicates. Five strains transformed to contain GFP (Methods and Results 4.2.1) were grown in minimal media (EMM) and adenine ('Normal growth') and in the same environment with added copper and iron chelators BCS and ferrozine ('Copper and iron starvation'). Samples of cells were removed from the cultures at 6 and 24 hours and RNA was extracted for qPCR analysis (Appendix 7.4). Expression is shown relative to *tub1* (a reliable normalisation used in *S. pombe* qPCR). For *ctr4* under starvation conditions we see a pronounced increase in expression after 6 hours when compared to normal growth (as expected for an increase in activity of the activator *cuf1*). However this transcription response seems to have diminished by 24 hours. *str3* in starvation conditions also shows an increase in expression after 6 hours (as expected for a decrease in activity of the repressor *fep1*), and also at 24 hours. However strains 4 and 2 show a much reduced response at this time which is probably due to these cultures reaching stationary phase at an earlier time. Strains 1 and 5 show the best expression pattern for use in our evolution experiment.

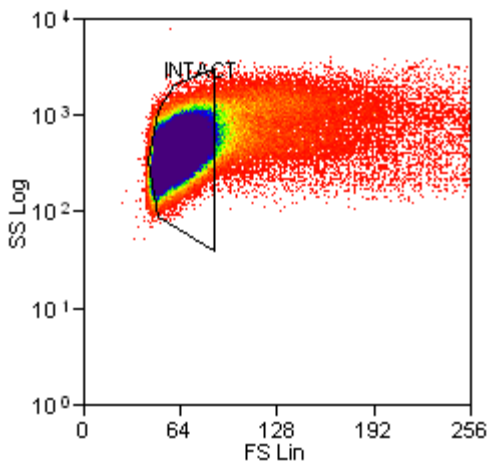


4.2.3 Growth and selection – a preliminary experiment

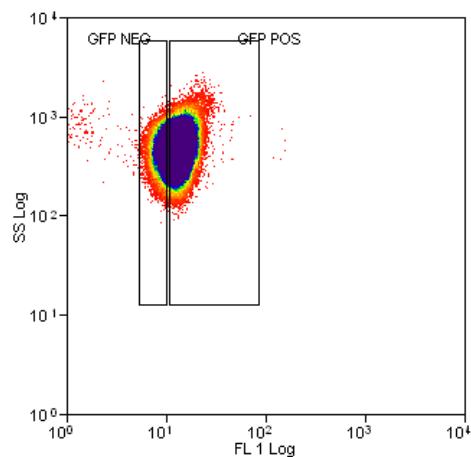
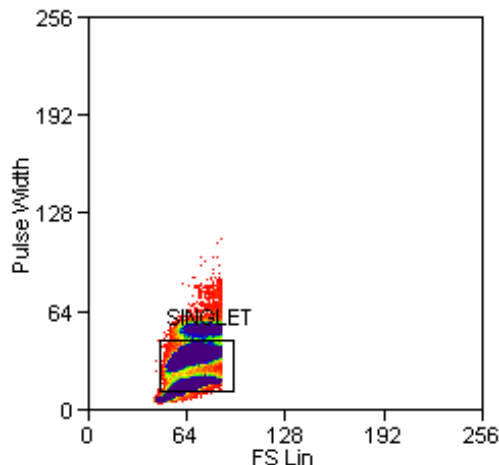
Two FACS sorts were run to assess the viability of our experiment. **Figure 23** shows the fluorescence distributions and sorting of two populations of cells, one under normal conditions and the other starved of copper and iron. The cells were sorted based on an initial sampling run of a smaller number of cells from the populations (data not shown). This gives the approximate distributions for the whole population. Using these distributions cells can be selected at whatever percentile of expression level we wish (see Appendix **7.8** for a description of how selection is achieved by FACS). For this experiment we chose the 20% threshold as our proportion of the population to select. This would give a strong selection coefficient in an evolution experiment, whilst still maintaining a large population size, thus allowing for high mutational input. The input of significant mutation and strong selection should favour rapid evolution. When we sorted for low expression the 20% threshold resulted in collecting over 60,000 cells. If we can select a similar amount of cells consistently then this population should go through approximately 7 cell division cycles in the 21 hours before the next day's sort. This produces a population of 7.68 million cells which will be sufficient for another sort (with a media volume sufficient for the optimal concentration for sorting). The cells were collected into the alternative media to that which they were grown in, and were then incubated and found to grow as expected, thus showing that our population is sustainable through our daily selection.

Figure 23. FACS sorting of the lowest and highest expressers of GFP in populations grown in minimal media and copper & iron starvation respectively. The sort parameters are based on data from a sampling run (not shown). Figures **23a** - **23d** show all of the parameters used to sort a population for low expression. Figures **23e** & **23f** are the equivalent figures for high expression to figures **23c** & **23d**, and show the final distribution of selected cells.

23a ▼ The first parameter selects for intact cells. This uses data on forward scatter (FS) and side scatter (SS) which approximate cell size and density respectively.

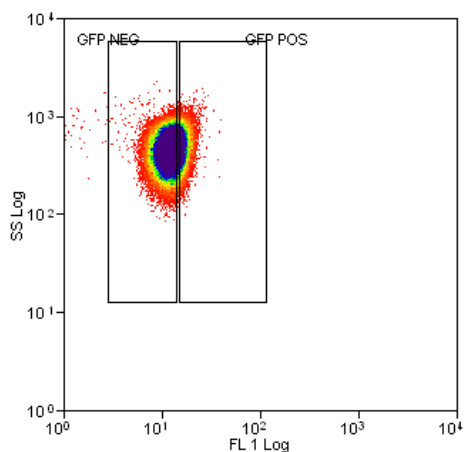
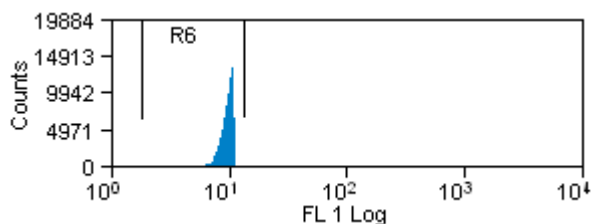


23b ▼ The second parameter selects for single, growth phase cells. This uses data on forward scatter (FS) and pulse width which approximates properties of shape and volume.



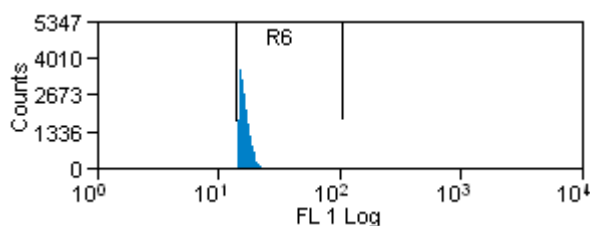
23c ◀ The third parameter selects for the cells with the lowest GFP expression (GFP NEG). In this run we selected 20% of the cells which had already been identified as intact and single. This uses data on GFP fluorescence (FL1) and side scatter (SS).

23d ▼ The lower tail of the distribution of fluorescence (FL1) values at the 20% cut-off. It shows the data from **Figure 23c** as a frequency distribution.



23e ◀ Sorting for the cells with the highest GFP expression (GFP POS). 20% of cells are selected which had already been identified as intact and single (these figures are not shown for this population).

23f ▼ The upper tail of the distribution of fluorescence (FL1) values at the 20% cut-off. It shows the data from **Figure 23e** as a frequency distribution.



5 DISCUSSION

The development of experimental systems we have presented lays the groundwork for our aim of studying phenotypic plasticity and gene regulation through experimental evolution. We have established two systems that select for regulation in two genes using two contrasting approaches. The experiments have largely proceeded as planned, however, there are some potential challenges which we discuss how to overcome (5.1). Future methodological objectives are suggested in section 5.2. These will contribute to our plans for future work which are detailed in section 5.3. Our focus in these plans is on testing hypotheses from theories and models of phenotypic plasticity and gene regulation. Finally, we conclude by discussing the scope of our research and its relation to another study with aspects of our own (5.4).

5.1 CHALLENGES IN CONSTRUCTING EXPERIMENTAL EVOLUTION SYSTEMS

5.1.1 Methods for alteration of the promoter region of *ura4*

One of the aims of experimental evolution using our systems is to explore the role of the promoter region in the regulatory evolution of a gene, since this region is well established as exerting significant control over expression. For this purpose, in our *ura4* experiment, we wanted the ability to evolve strains with different promoters of *ura4*. This allows for control over the initial position of our strain on the fitness landscape, and the introduction of promoters which may have varying potential to evolve regulation (Tirosh et al. 2009). We also wanted the option to introduce mutations into the promoter, and to move promoters between generations of our evolution experiment, to analyse the effect of both artificial and evolved mutations. Thus, we developed two methods for efficient introduction of promoters into position upstream of *ura4*.

The first involved the placement of the negatively selectable TK marker upstream of *ura4* (section 4.1.3), which would be removed by successful insertion of a chosen promoter. This method had a high rate of false positives, most likely caused by mutations in the TK marker causing spontaneous resistance to FUDR. This meant that finding successful transformants required further steps to eliminate the false positives. Molecular methods such as PCR followed by restriction fragment length polymorphism analysis (RFLP) require significant time, and thus are contrary to our goal of developing an efficient insertion system. We had some success selecting transformants by their ability to express *ura4*, as these will grow on media without uracil, whereas the false positives would not. However, this represents a selection step on uracil expression that could have an impact on our further experiments involving the same selection, and thus is not desirable. Our second method has the chosen promoter ‘piggy-backed’ onto the positively selectable NAT marker (section 4.1.3), and then inserted in place of the wild type promoter. This method is more reliable and seems to be the best choice given our problems

with the TK marker method, but it does require an extra ligation step before transformation. Ideally we would like to develop a one-step method for insertion, as the TK marker method was originally hypothesized, but so far this has not been possible.

5.1.2 URA4 deletion and exploiting nitrogen source restriction

When growing a *ura4* deletion strain in varying uracil concentrations we had two unexpected results (**Figure 7**). These phenomena may be somewhat unique to this strain, as our evolving strains should always express *ura4* at some level. However, we will see from our investigation that these results do indicate potential problems for our evolution experiment which we should be aware of.

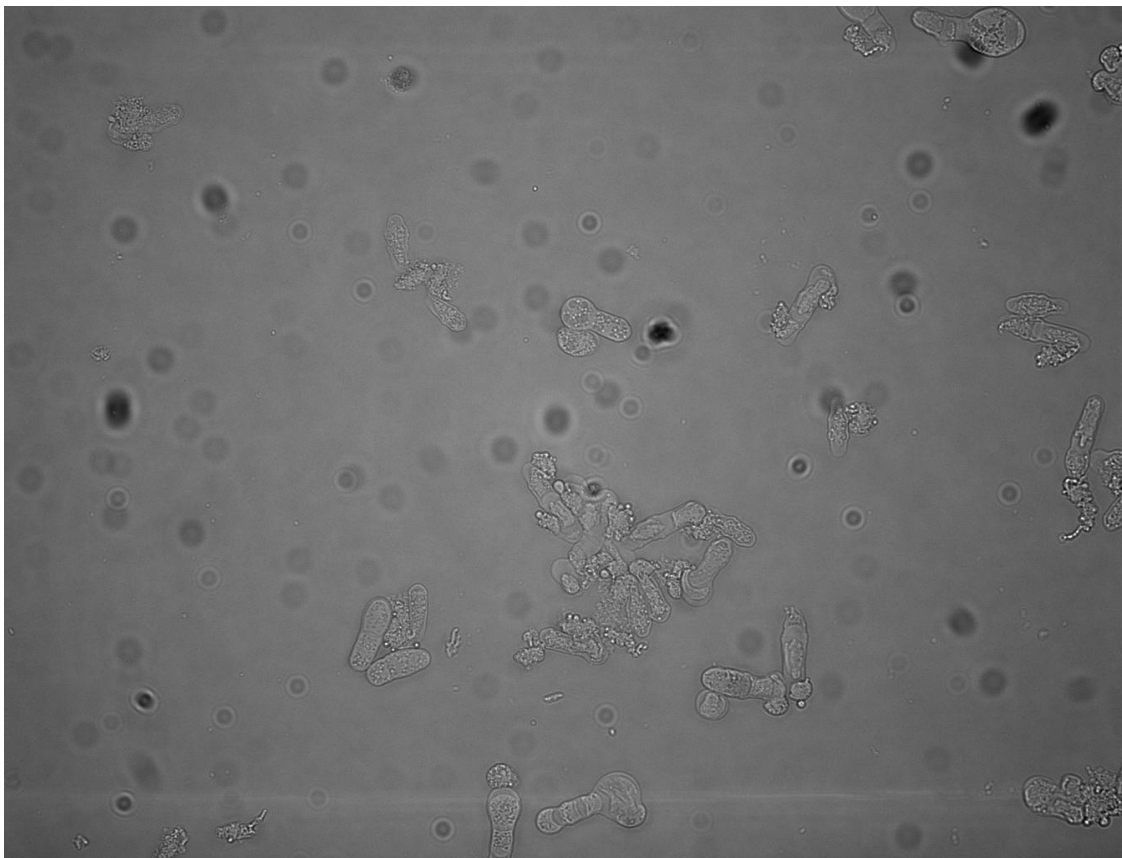
The first phenomenon is a sudden spiking in optical density after growth seems to have ceased (**Figure 7** – the spike is circled in green). This spiking is due to cell lysis, as we discovered by microscopic examination of the cultures (**Figure 24**), and it also occurs at higher uracil concentrations (data not shown). This causes anomalous results in the biolector, as it uses back scattering to measure optical density (separate measurement by transmission did not show such results). Lysis is known to occur at stationary phase in the absence of *ura4* expression, probably due to accumulation of a precursor in the uracil biosynthesis pathway (**2.3.1**; Matsuo et al. 2013). Because *ura4* will probably be expressed at some level in our experiments, and lysis would obviously be strongly selected against, we hope this phenomenon will not affect our experiments, although we should remain aware of it.

The second phenomenon is an initially higher growth rate in lower concentrations of uracil (**Figure 7** – this period of growth is circled in purple). The most likely explanation for this is that the external uracil concentration is affecting a regulation mechanism that normally functions to reduce uracil production or availability when the external concentration is high, so as to maintain a stable intracellular concentration. One candidate for regulation would be the uracil transporter FUR4, which is likely to be solely responsible for uracil uptake, unless external concentrations are very high (Nishino et al. 2015). However, it has been found that uracil concentration does not affect the function of this transporter (Nishino et al. 2015), so it is perhaps unlikely to be causing the effect that we see. Another possibility is regulation of the *urg1*, *urg2* and *urg3* genes which are involved in uracil degradation and utilisation as nitrogen source. These are known to be up-regulated by increased extracellular uracil concentration, presumably to stabilise the intracellular concentration and to utilise uracil as a nitrogen source (**2.3.3**). In our *ura4* deletion strain this would cause low availability of uracil, as none is being produced inside the cell, thus potentially explaining our results.

In our evolution experiment the uracil starvation environment will not contain any uracil, and the FOA environment will contain it in excess. Thus, the regulation suggested due to uracil concentration will only affect fitness in the FOA environment, where the strongest

selection pressure should be caused by FOA. However, the FUR4 transporter is probably solely responsible for uptake of FOA (Jund et al. 1988) – so it is important to consider the possibility of its regulation in our experiments. Specifically, we might expect the transporter to evolve towards down-regulation or even deletion to prevent uptake of FOA, with increased *ura4* expression to compensate. This compensation is in conflict with selection for low expression to prevent metabolism of FOA, and thus would negate our selection for regulation. One possible way to overcome this would be to use uracil as the sole nitrogen source for growth in the FOA environment (another source would have to be used in the ‘no uracil’ environment). This would force the transporter to remain active for growth to occur, making uptake of FOA unavoidable, and thus down regulation of *ura4* the only remaining solution. Although growth of *S. pombe* with uracil as a sole nitrogen source is possible we would need to investigate if the growth rate is high enough for this to be viable in our experiments (see Petersen & Russell 2016 for media; Appendix 7.1).

Figure 24. Light microscope image of *S. pombe* cells with a *ura4* deletion mutation grown under low uracil conditions. The cells are clearly lysed, leading to anomalous optical density readings in the biolector (**Figure 7**; sections 4.1.1 and 5.1.2).



5.1.3 *Pfta5-ura4* experimental evolution

The *Pfta5-ura4* strain seems to be an ideal starting point for experimental evolution (Methods and Results 4.1.5). However, we intend to run further analysis on the strain to confirm its potential. By accurately measuring expression of *ura4* RNA and protein in both environments we can be certain that no regulation is taking place; and by further analysing the fitness of the strain (**Figure 15**) with high accuracy in both environments, we will be sure of its intermediate position on the fitness landscapes. Complementary to this, we would also like to determine more accurate fitness landscapes using variable expression in our *Pnmt1* strains (**Figure 12**) by carrying out more experiments over a larger range of expression values.

A preliminary evolution experiment of 10 days using the *Pfta5-ura4* strain showed increasing fitness, potentially indicating that evolution was occurring with just 5 cycles of environmental change. Statistically significant fitness increases were shown in all strains in the FOA environment, and in some strains in the uracil starvation environment (**Figure 16**). Further investigation is required to understand the mechanisms causing these changes, and whether they involve *ura4* regulation. It is possible that the changes are a result of non-genetic adaptation (i.e. not by natural selection, but plasticity in phenotype), which would explain their rapidity. If this is the case then the changes would not be regulated by each environment, but would simply provide a fitness benefit overall by their effects in one or both environments. When embarking on our full-scale evolution experiment our control lines should help us distinguish such adaptation from the regulatory (plastic) adaptation that we are more interested in.

The evolved strains also showed a positive correlation in fitness between the two environments, and a linear regression analysis showed that high fitness in one environment is likely to predict high fitness in the other. This suggests that a significant amount of the adaptation at this early stage is beneficial in both environments, and thus could indicate the existence of regulatory adaptations. However, they could also be non-genetic adaptations to both environments, genetic adaptations to both environments, or combinations of adaptations to the individual environments. Genetic analysis and our full-scale evolution experiment over a longer time period should help us fully understand the adaptation that occurs.

5.1.4 Introduced mutation in *Pfta5*

Introduction of mutation into the *fta5* promoter region of *ura4* expression led to changes in mRNA and protein expression (**Figures 19** and **20**), and changes in fitness in both environments when compared to the wild type (**Figure 18**). There were more differences than similarities between RNA and protein expression in these strains. For example strain 4 shows relatively high RNA expression, but not protein expression, and has relatively few mutations in its promoter region (**Figure 17**), but does not show significant differences in fitness from the wild type. Our data serves to remind us that RNA and protein expression can be regulated

independently, with only protein level and localisation ultimately affecting uracil and FOA metabolism. Clearly we need to keep an open mind when considering how the genotype and phenotype are related by these mechanisms. However, to understand the details of these relationships we must improve the accuracy of our expression measurements, and make direct comparisons to the wild type promoter.

The changes in fitness of the strains appear to be random, indicating an unbiased effect of mutation on fitness. This is somewhat surprising, as random mutation usually averages a negative effect on fitness. We are looking at a promoter region, where we expect mutations to affect fitness by affecting expression, so it is possible that mutations are causing fairly unbiased changes to expression level. By looking at more mutants we can better understand the average effects of mutation in this promoter, and if we extend this to multiple promoters we may gain insight into the ‘evolvability’ of different promoters, which is thought to vary significantly (Tirosch et al. 2009).

5.1.5 Growth in a detectable environment for GFP expression selection by FACS

For regulation to evolve in response to selection on GFP expression there needed to be differences in transcriptional activity before selection (Methods and Results 4.2). These differences allow the GFP gene to evolve a response by becoming part of one of these regulatory networks. We decided to use the response of *S. pombe* to copper and iron starvation, as this allowed us to alter the activity of the repressor *fep1* and the activator *cuf1*, along with other regulatory elements (Methods and Results 4.2.3). The activity of these transcription factors can be measured by analysing the expression of their targets, *str3* and *ctr4* respectively. Our measurements showed that the transcriptional activity of *cuf1* is restored to normal levels at some point between 6 and 24 hours post addition of the copper ion chelator BCA, which sequesters the copper ions and starves the cells of copper (2.3.5; Figure 23). Further investigation of how the response degrades during this time period will allow us to select a time point for chelator addition that ensures a large transcriptional response in the hours approaching selection. This will be the time when any transcriptional response of GFP will be selected for, and thus is a requirement for our evolution experiment to be a success. However, we must also measure the half-life of GFP and the effect of transcription on protein levels to know whether a transcriptional response at this time will be able to significantly alter GFP levels. The GFP gene is under the control of the *P4Inmt1* promoter, which is repressible by thiamine, allowing us to alter expression and then analyse the effect on GFP levels.

5.2 FUTURE WORK – METHODOLOGICAL OBJECTIVES

5.2.1 Maintaining an evolving population

To maximise the chances of evolutionary change occurring during experimental evolution there are many factors to consider, and these must be weighed against keeping the experiment manageable with the available resources. There is a risk of contamination between replicates when running an evolution experiment with multiple lines growing simultaneously, so using a relatively simple protocol to eliminate this possibility is important for producing independent replicates.

Significant growth in both environments is required, as this is when selection will act on the relative growth of different mutants, and it is these cell divisions which will introduce the mutations that are needed for this selection to occur. Therefore, populations need to be seeded from few enough cells that population growth can occur. However, we need to move enough cells between environments that we maintain a reasonable population size, in order to not bottleneck the population and reduce the mutational input which our experiments require to be successful. Another problem is that if we do not take enough cells, and if growth is slow, then we will be repeatedly diluting the population until it is lost.

In our *ura4* experiment a practical solution could be to take a small fixed volume of cells (to avoid having to repeatedly measure cell concentrations) sufficient for the cultures to reach stationary phase. This keeps the number of transferred cells high, whilst not carrying over any significant volume of media from the previous environment, and allows for substantial growth. This does introduce the complication that our cells are being selected in both exponential growth and stationary phase, but selection should largely occur during growth, so we predict that any selection due to stationary phase will not have a large effect and will not interact with gene regulation phenotypes we are selecting for. We will also use control lines which will be reaching stationary phase, which should help us understand which selection pressures were acting, and when during our experiment they acted. If we wish to use uracil as a sole nitrogen source in the FOA environment (for reasons detailed in 5.1.2) then we will need to test growth rate in this media and adjust our experiments accordingly.

One group of control lines will be grown in a constant environment of uracil starvation (we expect these to increase their expression level of *ura4*) and the other group will be grown in a constant environment of FOA (we expect these to decrease their expression level of *ura4*). These controls may have some significant differences from the switching environment lines, such as spending more time in stationary phase, as they may evolve a high growth rate much faster, due to the simpler challenge presented to them. However, because not much growth occurs in this phase we hope that the controls will still serve their main function – to capture any evolutionary change which is selected for that does relate to the changing environment of

uracil starvation and FOA. A further complication is that FOA takes time to have its full effect on growth. This means that when the cells first enter the FOA environment their growth is not immediately affected, and their growth rate is then slowly reduced over the following hours. After a day the cells are returned to the uracil starvation environment, which gives them time to recover from the toxins they produced when metabolising the FOA. For the control growing in FOA this is not the case, and the effect of the FOA builds up over time leading to a greatly reduced growth rate. These cells may not initially be ideal controls, in that they face quite different conditions from the changing environment lines – they will suffer higher toxin concentrations, they will not be in stationary phase for as long and they will have a smaller population size reducing the effectiveness of mutation and selection to produce evolutionary change. However, we hope that this control will still serve its main purpose, which is to gain an insight into any evolutionary change which can occur in response to FOA which is independent of being in a changing environment of FOA and uracil starvation.

Based on preliminary experiments we have developed a strategy for experimental evolution. We intend to use 60ml glass cell culture tubes with 4ml of media for growth. We will transfer 15µl of culture to the fresh media (new environment) daily. We have calculated that the volume of transferred uracil and FOA is low enough to produce equivalent growth conditions to absolute uracil starvation. When transferring the cells to their new environments we will use long-length pipette tips and sterile technique to avoid cross-contamination. A reasonable aim would be to evolve 40 lines, with 30 in the changing environment, and 5 in each control condition. Samples can be taken from the experiment and frozen for further analysis at any time points we wish.

5.2.2 Constraints on plasticity

A simple way to analyse the constraints on plasticity in our experiment will be to compare those lines which have been selected in changing environments (2-environment) to control lines which have been selected in only one of the two environments (1-environment). This examination is predicated on the fact that all lines will evolve towards optimal phenotypes in the environments they are grown in and that this phenotype is manifested as the expression level or another trait which has identical adaptive function in plastic and non-plastic lines. If these conditions are not met then a similar approach may be possible with an artificially optimised strain, for example with URA4 under inducible control (4.1.4).

At any point in the evolution of our 2-environment lines we can compare their phenotypes and fitness to those of 1-environment lines in the environment they evolved in. We would use 1-environment lines that had undergone extended evolution in their environment, such that their phenotypes were stable. We may find there is no detectable fitness loss or phenotypic difference, implying we have a system with no constraints on the evolution of a plastic phenotype (perfect plasticity), although this would not likely be the case during early

evolution. Any deviations in fitness or phenotype can potentially be matched to theoretical predictions regarding constraints (Section 5.3, especially: 5.3.1.3; 5.3.1.4; 5.3.1.8).

5.2.3 Fitness landscapes

We propose developing a range of fitness sets in our experiments in order to investigate hypotheses related to how different fitness landscapes effect the evolution of plasticity and gene regulation (5.3.1.1; 5.3.1.5; 5.3.2.2).

In our *ura4* experiment we propose creating convex and concave fitness sets by altering the concentrations of uracil and FOA in our environments. With a high concentration of FOA in one environment and little to no uracil in the other we intend to have highly divergent selection, where there is little overlap between fitness functions and thus a concave fitness set. By reducing FOA concentration in one environment, and increasing uracil concentration in other (whilst still being limiting on growth) we should be able to create less divergent selection pressures where there is significant overlap between fitness functions, and the resulting fitness set is convex. Once we have established some control over fitness sets in this manner, we may be able to make smaller alterations to test the effects of a range of fitness sets as they go from concave, through intermediate, to convex.

From our experimental work we can see that developing altered fitness sets in our experiments on the *ura4* gene is likely to be possible, as our data on dosing in both environments shows that we can achieve a graded effect on fitness (4.1.1). However, this does not tell us anything about the fitness functions in these environments of expression (since each experiment looks at only one extreme value for expression), and thus only by further exploration of the relationship between expression and fitness in multiple environments can we begin to see what range of fitness sets might be possible. Promisingly, the fitness functions we have established for our initial environments show some overlap (4.1.4; Figure 12), indicating that varying our environments could lead to significant changes in overlap and thus fitness set shape.

In our GFP experiment the artificial nature of our initial selection parameters leads to fitness functions with no overlap and no gradient, as the fitness of all cells that make it through selection is equal, and these are taken from the extreme upper and lower values of expression, depending on the population. Further to this, the fitness functions change with each round of selection, depending on the expression distribution of cells across the population. Thus, this particular method of experimental evolution has little potential for testing hypotheses related to fitness sets. However, we may be able to perform the experiment using fixed fitness functions where specific expression values are selected for. The fitness of different expression values could be differentiated by running a programme on the cell sorter that chose cells probabilistically rather than deterministically, with cells having an expression value closer to

the chosen fitness maxima having a higher probability of been sorted into the next generation. This would produce fitness functions of expression with smooth gradients, and we could potentially alter functions and thus fitness sets in any way we wished, ideal for testing of the related theories. The downside of this selection regime is that the size of the populations that can be sorted may not be large enough for significant mutational variation to occur, such that a viable number of mutants are selected for by the regime to continue the experiment. This could be overcome by introducing mutation artificially, or by giving small populations more time to grow. Our original method of selecting depending on population expression parameters should gradually select towards stronger regulation, making each step towards further regulation more mutationally accessible. Another possibility would be a hybrid method, where the initial population analysis is used to sort cells based on a set fitness function, but with less severe selection pressure if the population's expression distribution is a long way from the fitness maxima.

If we were successful in developing both methods of selection in this experiment then we potentially have examples of 'soft' and 'hard' selection allowing us to test hypotheses related to the effect of these on plasticity evolution (5.3.1.6; Van Tienderen 1991). Soft selection can be achieved by altering the selective ranges continually to take the same number of individuals with the highest and lowest expression, thus making selection frequency dependent. Hard selection can be achieved by selecting for a specific expression range and not deviating from this for the entire experiment.

5.2.4 **Bet hedging and variance – single cell expression**

Many theories and models in phenotypic plasticity include bet-hedging as a possible adaptation to a variable environment. This can be in the form of trait variance or stochastic switching between multiple traits (2.1.3). To look at these traits we need to be able to assess the phenotypes of individual cells in our experiments. The simplest approach is to use the GFP tags that have been attached to the genes of interest in all of our experiments. If we suspect other genes of being responsible for bet-hedging phenotypes we could tag these with an alternative fluorescent protein with a different wavelength peak so that multiple genes could be analysed in this way simultaneously.

The GFP experiment using FACS sorting already measures individual cell expression so the data required to investigate bet-hedging hypotheses should be immediately available. For the *ura4* experiment we could use the FACS machine, or a more simple flow cytometry machine that measures fluorescence. A simpler but more time consuming alternative is to measure individual cell fluorescence using microscopy.

5.2.5 A third experimental evolution system

We hypothesize another experimental evolution system, very similar to the *ura4* system, targeting the *fcy1* gene in *S. cerevisiae*. Hypothetically, there are two specific environments which should select strongly for and against the expression of *fcy1*. To select positively for expression we would grow the cells in a defined media where cytosine is the sole nitrogen source. Expression of the *fcy1* cytosine deaminase gene will therefore be selected for as it is an essential part of the pathway for utilising cytosine as a nitrogen source. To select negatively against expression we would use an environment with purine as the sole nitrogen source, and we would add 5-fluorocytosine to the media (5FC). 5FC is metabolised by the *fcy1* gene to form the toxin 5-fluorouracil (5FU), so its presence selects against *fcy1* expression. Further, the use of purine as a sole nitrogen source ensures that there can be no selection against transporting 5FC into the cell, as the same transporter gene (*fcy2*) is responsible for uptake of purine, cytosine and 5FC (Paluszynski et al. 2006). This potentially provides similar advantages to the use of uracil as a sole nitrogen source in the *ura4* system (in the FOA environment; see section 5.1.2). This experiment would provide an interesting complement to our experiments in *S. pombe*, but significant work would have to be undertaken to investigate its viability, including initial research to ensure that *fcy1* is not required to utilise purine as a sole nitrogen source.

5.3 FUTURE WORK – TESTING THEORIES AND MODELS OF PHENOTYPIC PLASTICITY AND GENE REGULATION

Following the success of our preliminary experiments we are in a position to begin experimental evolution in order to test specific theories and models, with proper consideration of the challenges discussed earlier in this chapter. In this section we introduce a range of potential experiments to test various specified theories and models of phenotypic plasticity and gene regulation, with rigorous discussion of the hypotheses being tested and the corresponding implications of specific results. Firstly, we summarise the imminently achievable aims, details of which are found in the corresponding following sections:

Our evolution experiments will have multiple independent lines evolving in parallel. This should allow us to immediately analyse any patterns we see between lines in the evolved phenotypes, and in the molecular mechanisms responsible for them. Whether or not we find patterns is itself interesting, and if we find repetition in the strategies and phenotypes evolved, and the dynamics of their evolution, we can analyse these in the light of our hypotheses (5.3.1). We can also look for specific molecular mechanisms (5.3.2.1), whether mutations are occurring in *cis* or *trans*, and how mechanisms change over time. Analysis of expression phenotypes and protein degradation from our strains will also allow us to test specific predictions from the models of Proulx & Smiley (2010. 5.3.2.2). Sequencing the genes under regulation should allow us to investigate relaxed selection in these strains, when compared to the sequences of our

control lines (5.3.1.3). Comparisons with our control lines should also allow us to test further hypotheses related to constraints (5.3.1.4; 5.3.1.8).

5.3.1 Testing theories and models of phenotypic plasticity

Three of the most fundamental questions that can be asked when considering the evolution of a plastic phenotype are when it will be favoured over the other strategies, what form it will assume, and what will be the dynamics of its evolution. Various models of plasticity address these fundamentals using contrasting approaches, prompting variety in experimental testing. Experimental evolution with our systems involves a clonal population (although there is the possibility of introducing sex), with individual mutations affecting plasticity likely giving large fitness benefits leading to quick invasion. We also have a somewhat defined optima of phenotype (4.1.4). However, there is not likely to be large variation for our trait over many loci (unless we introduce such variation – 4.1.7) and there is no reason to assume additive variation or constant covariation. Overall this makes the popular quantitative genetic approach a poor choice for our experiment. More suitable are optimality and allelic models. In the following sections we present specific predictions for our experiments, mostly from models in these areas. The general theme linking all predictions is that perfect plasticity will always be the adaptation with the highest fitness, but deviation from this will be the norm due to the numerous constraints that are likely to occur. By investigating such constraints we hope to gain some insight into the bigger question of which constraints are prominent more generally in plasticity evolution, along with the effects this has on adaptation. Many models assume randomly fluctuating environments, as any regularity in environmental change could favour periodic variation in phenotype that is not fully controlled by the environment. Therefore, we may need to use a regime of randomised environment switching if we wish to properly test these models.

5.3.1.1 Plasticity and alternative strategies

The model of multi-moment reaction norms (DeWitt & Langerhans 2004; DeWitt 2016) provides a general framework for evolution in a changing environment. It allows for strategies of generalism, specialism, bet-hedging and plasticity, as well as combinations of these. The model predicts end-point strategies, but we can make simple qualitative predictions of the possible dynamics of evolution in our experiments from the relative fitness of the strategies.

The assumptions and parameters of this model are well suited to our experiments. It is presented for two-environments with equal frequency. Environmental variation is assumed to be intergenerational (coarse-grained), which will be true of our experiment where there is exponential growth of a population of single cells occurring before switching to another environment. There is assumed to be no soft selection, which is true of our URA4 experiment, as there is an optimal phenotype that does not change for the course of the experiment (the environmental regime remains the same). This assumption is violated for our GFP experiment

if we are continually changing selection parameters (making fitness frequency dependent), but we could perform this experiment with fixed parameters (5.2.3).

The fitness functions of the expression phenotype in the two environments of our *ura4* experiment do not greatly overlap, leading us to predict a concave fitness set (4.1.4). First, we consider strategies with no random phenotypic variance (bet-hedging). The model predicts that with concave fitness sets plasticity has a much higher fitness compared to a generalist (non-plastic intermediate) strategy under a wide range of parameters, unless costs of plasticity are very high, or cues are inaccurate. The generalist strategy always has higher fitness than the specialist strategy (non-plastic extreme). Therefore, we would predict plasticity to be the stable end-point of experimental evolution, and if the population transitions through any other strategies they would be generalist (intermediate expression).

When the possibility of bet-hedging (phenotypic variance) is introduced into each strategy we can expand our predictions. To test predictions regarding variance we would have to analyse the phenotype of individual cells, for example by tagging URA4 with GFP and thus measuring single cell expression (5.2.4). A specialist strategy may be able to replace a generalist strategy under a concave fitness set, on the condition that the specialist has a high phenotypic variance (significant bet-hedging) compared to the generalist. If we are starting our experiment from an intermediate expression level, we may see initial evolution of a specialist, followed by the evolution of plasticity, indicating that a specialist strategy is more mutationally accessible. The model also predicts that any generalist will have some-level of bet-hedging under a concave fitness set, therefore we expect some change in the variance of our phenotype (expression) if evolution is progressing between generalist strategies. This is likely to be an increase in variance assuming variance is initially low. However, the model doesn't account for the possibility of a bet-hedger which switches randomly between specialist phenotypes, and which could potentially have much higher fitness (see Background 2.1.6 and 5.3.1.2 below). We should also be aware that noise in the phenotype may not be adaptive in the manner presented during early evolution.

The model also incorporates bet-hedging by plastic strategies (trait variance in each environment). From the predictions of the model we expect any plastic strategies that evolve in our experiment to incorporate a level of bet-hedging depending on the limits on perfect plasticity, due to our fitness set being concave. We do not expect bet-hedging due to cue inaccuracy as our inaccuracy should generally be low, however the effects of cue reliability can be further explored with our GFP experiment (see 5.3.1.5). A cost of plasticity never favours bet-hedging. Therefore, assuming accurate cues, if we find bet-hedging in a plastic strategist the model predicts that we should be able to find (by other means) a limit on plasticity. If we find no bet-hedging in plastic strategies the model predicts that we should find no other limits on plasticity (for concave fitness sets). However, we do expect to find bet-hedging dynamics

during evolution towards optimal plasticity, even if the final strategy has no bet-hedging. The variance of this bet-hedging is expected to show skewness towards the optimal value.

Because we can potentially alter the fitness set shape in both our URA4 and GFP experiments (5.2.3) we can test the prediction that limits will only lead to bet-hedging under concave fitness sets. If under a concave set cues are 100% accurate (see 5.3.1.5 for when this is not the case), and we thus find bet-hedging which we conclude to be due to limits, then this bet-hedging should not evolve under convex fitness sets. We could test this both by evolving lines independently from the same starting point under the different sets, or we could evolve first in one set, and then move to the other to see if the predicted change occurred. It should also be noted that over the full range of limits the advantage of a plastic strategy over a generalist decreases significantly under convex fitness sets, and thus plasticity may not evolve or be lost, removing our ability to test the above predictions regarding bet-hedging and plasticity. Another important caveat is that this model incorporates limits on developmental range, rather than any limits due to genetic factors (since it does not include these) and thus we must be cautious when using its conclusions in conjunction with predictions based on limits due to phenomena such as relaxed selection and drift.

Finally, the model also incorporates the possibility of all strategies having asymmetric bet hedging (trait variance can vary between the two environments). This expands on above predictions without voiding them. We would expect any specialists that evolve to have much more extreme variance in the environment they are not specialised in, compared to the one they are specialised in. Generalists are also predicted to be asymmetric in their bet-hedging due to our fitness set not being symmetrical (the fitness functions likely have different shapes in our two environments – 4.1.4). Finally, if the plastic strategists in our experiment are bet-hedging due to limits on plasticity, then we expect this also to be asymmetrical both due to the asymmetrical fitness set, and due to asymmetry between the limits on plasticity in each environment (because expression cannot go below complete silencing in the FOA environment).

We can further test this model by altering the fitness functions creating a convex fitness set. This can be potentially be done in both our URA4 and GFP experiments (5.2.3) and, as mentioned above, we can test the alternative predictions by evolving the same starting strain in each, or moving from one fitness set to the other. The model predicts that plastic strategies generally have less of an advantage over generalist strategies under convex fitness sets, and thus under these we would expect plasticity to evolve slower and less often compared to under a concave set. The model also predicts that under intermediate and convex fitness sets bet-hedging specialists will never out-compete pure generalists, unlike under concave fitness sets, thus we would never expect these to evolve under these fitness sets. Under concave and intermediate fitness sets we would expect a degree of bet-hedging in generalist strategies,

however evolving under a convex fitness set should never favour the evolution of bet-hedging in a generalist. If we find significant costs of plasticity (5.3.1.4) the model predicts that generalists would out-compete plastic strategies under convex fitness sets, unlike under concave fitness sets. We can potentially alter the fitness sets by degrees to provide a more rigorous test for the model (5.2.3).

5.3.1.2 *Bet-hedging by stochastic switching*

Theory suggests that under strongly divergent selection between two environments a stochastic switching strategy may out-compete a generalist strategy by providing higher geometric mean fitness (2.1.6). This type of bet-hedger contrasts with the high-variance bet-hedger discussed above (Figure 4; 2.1.6). We may expect bet-hedgers to evolve, possibly replacing generalists, before being replaced by even higher fitness plastic strategies. This presumes that evolving the connection of phenotype regulation to environmental detection requires less probable mutations than evolution of the optimal phenotypes and a random switching mechanism. If we find this type of bet-hedger only evolves in small proportion of our lines before plasticity evolves we can conclude it is not likely to be more mutationally accessible than plasticity. We can also measure potential fitness benefits of bet-hedging over a generalist by controlling expression using repressible promoters (4.1.4). Models also predict that some individuals will bet-hedge whilst others are plastic, with the proportion depending on cue reliability (5.3.1.5; Arnoldini et al. 2012). If we wish to select for bet-hedging without plasticity this is possible in our GFP experiment, by giving no environmental cue for the upcoming selection.

5.3.1.3 *Relaxed selection*

This section will focus specifically on the *ura4* experiment, although a similar analysis could be completed with the GFP experiment. We predict that the lines in our experiments will evolve towards environment-specific gene expression - the expression of the URA4 gene in the 'no uracil' environment only. This extreme specificity towards expression in a single environment that the organisms only experience approximately half of the time leads to relaxed selection on the gene (2.1.4; Snell-Rood 2010). Therefore, making use of our control lines (5.2.2) we predict that selection should be relaxed to half of its strength in the 2-environment lines (on the assumption that they evolve environment-specific expression), whereas in 1-environment 'no-ura' selection should be at its strongest (URA4 is always required). In 1-environment 'FOA' we would expect selection against maintaining functional URA4 (thus this should not be considered an example of relaxed selection). We have two ways to test these predictions.

First we can look at mutation accumulation among non-synonymous sites (and potentially use synonymous site mutation to ensure mutation rate is similar). We expect this to be highest for 1-environment FOA lines, followed by our 2-environment lines, with 1-environment no-ura lines having the lowest rate.

Secondly we can analyse protein function of URA4 by transforming the protein coding sequence from each of the lines into an identical background strain, either expressed from a plasmid or the genome. Functionality can then be indirectly tested by growth rate in the no-ura and FOA environments. We predict the URA4 genes from the strains evolved in 1-environment no-ura will have the highest growth rate, followed by the sequences from the 2-environment lines, followed by the genes from the 1-environment FOA lines. We expect the opposite relationship for growth rate in FOA. Further direct tests of protein function could be carried out by isolation and testing using molecular methods. Of course the difference in selection strength may not be enough to lead to any detectable relaxation of selection on such a small genomic region, but if there is a difference in the gene it should follow the patterns described.

There may also be relaxed selection elsewhere in the genome, for example, if an upstream regulator is down or up regulated in one of our environments such that the protein it codes for (e.g. a transcription factor) is no longer functioning in one environment, then we would expect relaxed selection. Once we have investigated the regulatory basis for plasticity in our lines (5.3.2) further analyses of these potential sites of relaxed selection will be possible using the methods above.

5.3.1.4 Costs

Costs of plasticity are defined as fitness losses of plastic individuals compared to those producing the same phenotype without plasticity (in an environment to which both are adapted - 2.1.3). Using our control lines (5.2.2): if the phenotypes produced in the 2-environment and 1-environment lines (in the environment they evolved in) are identical, but there is a reduction in fitness of the 2-environment plastic strain then we can conclude there is cost of producing the plastic phenotype. We must be careful to distinguish these from simple costs of each phenotype in our experiments (Murren et al. 2015). However, we assume negligible relative phenotypic cost, as the fitness gain of producing the correct phenotype, and the fitness loss of producing the incorrect phenotype, are very high. To put this another way, due to our experimental setup, the phenotype that gives high fitness in one environment should give very low fitness in the other. If the fitness cost of plasticity itself were comparable to the fitness benefit in each environment for producing the optimal phenotype, then the cost of phenotype may be relevant. We assume this is not the case, based on experiments artificially introducing beneficial phenotypes (4.1.4), however, further experiments of this type could be explored, should we suspect such high plasticity cost relative to phenotypic benefit.

5.3.1.5 Cue reliability

It is predicted that the more unreliable a cue, the less favourable plasticity becomes as a strategy, and the more favourable fixed strategies become (2.1.2). In our GFP experiment we can alter the reliability of cues and see the effects on the evolution of gene expression (this may also be possible with our *ura4* experiment, although much more difficult to control as selection is not

direct and there will be potential reliable cues intrinsic in the selection pressure). We can create selection regimes where the co-occurrence of a certain cue and a certain selection pressure can be varied for different evolving populations. When the cue and pressure always occur at the same time we have 100% accuracy. We can then vary the co-occurrence probability down to 50% to give a completely random occurrence of the cue (no relation to the environment). As reliability increases we predict that plasticity will be more likely to evolve as an end result and will evolve more rapidly (2.1.2).

Models of phenotypic plasticity make more specific predictions relating to cue reliability. The optimality model of DeWitt & Langerhans (2004) predicts plasticity will only outcompete generalists when cues are above 75% accuracy, regardless of fitness set shape and environmental change frequency. We can test this by altering fitness sets (5.2.3) and environmental change frequency along with cue reliability, but we should be mindful that this prediction does not extend to when variance in phenotype (bet-hedging) is introduced into the model.

When bet-hedging is introduced further predictions follow on the form that plasticity will take. From the model we would expect that no bet-hedging would evolve when cues are 100% accurate regardless of fitness set shape (although bet-hedging may evolve anyway due to biological limits – 2.1.5; 5.3.1.8). When cues are around 95% accurate or less we expect bet-hedging to be favoured if the fitness set is concave. We would expect the degree of bet-hedging (the variance in expression) and its fitness benefit to increase as cues became less accurate still. If we altered the fitness set to be intermediate or convex the model predicts that bet-hedging has very little benefit and we may expect it to only evolve slowly or not at all, unless accuracy is very low. We can also extend this experimentation to altering both fitness sets and cue accuracy by intervals to build a multidimensional view of the effects of these factors on plasticity and bet-hedging, and see how this fits with model predictions.

This multidimensional experimentation can be extended to altering environmental frequency as this is predicted by other models to interact with cue accuracy in determining when plasticity is favoured over fixed strategies (Levins 1968; Leon 1993; Proulx & Smiley 2010). The effect of cue accuracy in this context relates to the interaction of environmental frequency and the lag in phenotype production.

Cue reliability can also be altered asymmetrically such that it is more reliable for one environment than the other. The DeWitt & Langerhans model predicts that when this is the case any bet-hedging (variance) in phenotype should be increased in the environment with the less reliable cue.

Further predictions of the effects of cue reliability on bet-hedging come from models which consider stochastically switching bet-hedgers (rather than those with variance around

one phenotype – **2.1.6**). The model of Arnoldini 2012 predicts a population of individuals which stochastically become bet-hedging or plastic (environmentally determined) individuals, with an increasing proportion of bet-hedgers with decreasing cue reliability. Thus, from this model, the prediction is that our experiments will produce strains where some cells bet-hedge and some are plastic, and the proportions will change with reliability.

5.3.1.6 *Hard and soft selection*

The quantitative genetics model of Van Tienderen (1991) makes specific predictions regarding the dynamics of evolution in two environments under hard and soft selection. Under soft selection the phenotype is predicted to gradually change as the frequency of one environment increases relative to the other (from 0 to 1), and to show significant plasticity for duration of evolution. Under hard selection plasticity is minimal and the phenotype changes rapidly over a short period relative to the environmental transition. This prediction is potentially testable with our GFP experiment. Environmental frequency can be altered as evolution progresses. Hard selection can be implemented by selecting only those cells with expression in a specific range, i.e. selecting for a specific phenotype (in the context of the model – a specific phenotype will have high fitness. See section **5.2.3**). Soft selection is implemented by a regime where a certain percentage of the highest and lowest expressing cells are selected (here fitness is related to the phenotype distribution of the population, not an absolute value). We can evolve under these selection regimes and analyse the changes in fitness and phenotype to see if they show a gradual and abrupt change with significant and little plasticity for the soft and hard selection lines as predicted by the model, although it may be necessary to introduce variation for expression artificially (**4.1.7**) at the beginning of the experiment to allow both regimes to act in good approximation to the model.

5.3.1.7 *Adaptation to discrete and continuous environments*

Both the URA4 and GFP experiments have the potential to investigate the idea of adaptation to discrete environments pre-adapting populations to other environments in a range continuous with the discrete environments they evolved in. The URA4 experiment can potentially be carried out in ranges of uracil and FOA concentrations which both alter fitness (**4.1.1**), and the GFP experiment can have a range of specified expression windows selected. This allows us to evolve strains in two discrete environments, and then test their fitness both in ‘intermediate’ environments within the range defined by the two discrete environments, and ‘extreme’ environments outside the range of these. The fitness gain can be measured over a background ancestral strain, and compared to the fitness gain in the discrete environments in which evolution took place.

5.3.1.8 *Limits*

If our 2-environment lines have evolved to a stable phenotype but do not reach the fitness levels of the 1-environment lines there are various possibilities for the cause of the constraint on

plasticity (methodology summarised in section **5.2.2**). If the phenotypes are not identical (for example, URA4 expression - with both its mean and variance being taken into account) then there is potentially a limit on plasticity. This could be a mechanistic limit (biological limit – section **2.1.5**) on the regulatory system that is changing the expression between environments – such as the regulatory control being unable to stabilise expression to the same level as the fine-tuned expression of the unregulated (i.e. constitutively expressed) gene in the 1-environment lines.

Plasticity could be limited by environmental detection. This could be a biological limit such as signalling delay, which could lead to reduced fitness and the same apparent phenotype, because the optimal phenotype was not reached immediately. We must be mindful of this in our measurements of phenotype and fitness; the limit being potentially detectable by only measuring fitness and phenotype once the phenotype had stabilised, or by measuring the dynamics of phenotype production and fitness. Any other inability to detect the environment would also be a biological limit and we could potentially overcome this by providing a more easily transduced environmental signal.

The other limit posed by environmental detection is cue reliability and uncertainty. In our URA4 experiment this should not be an important factor, as we carefully control the environment and thus the availability of any cue. In our GFP experiment there is potential for experimental alteration of cues to investigate the influence of cue reliability on plasticity evolution (**5.3.1.5**).

Limits may due to mutational availability (something that would theoretically be overcome if we gave our experiment enough time), this could be investigated by comparing to evolution of strains with an artificially raised mutation rate (**4.1.7**). Mutation towards a plastic phenotype may also be constrained due to fitness valleys (**2.1.5**; Steinberg & Ostermeier 2016; Proulx & Smiley 2010). If this was the case we might expect different lines to ‘jump’ these valleys at different times, with relatively sudden changes in phenotype of the population, and concurrent complex mutations, possibly of a similar nature between lines.

Finally the limit may be due to genetic drift overcoming selection, which will depend on the strength of selection, the mutation rate and the population size. These can potentially be controlled, measured or estimated in our experiments, and thus the impact of drift understood. Related to the concept of drift is that of relaxed selection, where a phenotype which is only beneficial in one environment will have weakened purifying selection (**5.3.1.3**).

Here we have briefly introduced the possible limits and referred to further experimental avenues to explore if these limits are in action. If none of these experiments reveal limits then we are left to conclude, by process of elimination, that there is an unknown biological limit, which could be overcome to some extent over a longer period of selection.

5.3.2 Testing theories and models of gene regulation

5.3.2.1 Mechanisms in gene regulation

Gene regulation is a complex process with many mechanisms (2.2.1) and is thought to be an important mechanism in phenotypic plasticity (2.2), although the genomic basis of environment specific phenotypes is a debated area (Grishkevich & Yanai 2013). With our experiments we have the potential to investigate how the molecular genetic mechanisms of environmental regulation evolve from their nascence. Although a variety of mechanisms are fairly well understood their evolution is not. If we find patterns in our experiments in the timing and form of arising regulatory mechanisms this is indicative of the dynamics of regulatory evolution in general. This could be verified if signatures of these patterns could be found in natural populations and other species. A great diversity in regulatory evolution between our lines would also be of interest, indicating the possibility of a general diversity and lack of predictability in natural regulatory evolution. We may also discover unknown regulatory mechanisms, since our approach to revealing them is different to more widely used methods.

Current theories of the molecular basis of regulatory evolution also make some more specific predictions of what we might find. Much experimental and theoretical work focuses on the evolution of the promoter region (2.2.1) and we would expect to see mutations here that affect regulation. We would predict a high likelihood of mutations being fixed which increase binding of transcription factors whose regulation is linked to our environmental changes (Grishkevich & Yanai 2013). This would represent a mechanism of linking gene regulation with environmental detection, a strong prediction of what will occur in our experiments. Evolving a response to a particular transcription factor implies a significant expression change in only one environment, with little change in the other, and thus we would predict mutations of this sort to only alter fitness in one environment, with further alterations to base expression levels and regulation then leading to optimal regulation. More detailed theories predict that promoters with features such as an abundance of TF binding sites, nucleosome free regions and a TATA box may be more likely to evolve regulation (Tirosh et al. 2009). We could test this by evolving genes with a variety of promoters (3.1.2.3). To detect the evolution of transcription factor binding we can compare the mutations and surrounding bases to known transcription factor binding sequences, and we can detect binding using chromatin immunoprecipitation (ChIP). Mutations in the promoter region may also be involved in rearrangement of histones, also testable with ChIP.

So far we have discussed *cis* changes, but there is also the possibility of *trans* effects in regulation, and some data even suggest that these may be more important factors in regulatory evolution (Grishkevich & Yanai 2013). These would be expected as mutations in genes or promoter regions that form part of a regulatory network that is (or becomes) connected with our target gene (*ura4* or GFP). Dynamically, we would predict the initial occurrence of *cis*

mutations tying *ura4* and GFP into pre-existing regulatory networks, followed by *trans* regulation changes within these networks. Specifically, we would expect evolutionary change in networks already connected to the environmental changes in our experiment, such as networks involving the *urg1*, *urg2* and *urg3* genes in our URA4 experiment (2.3.3) and the copper and iron response networks in our GFP experiment (2.3.5; 2.3.6). The continued evolution of regulatory networks once plasticity has evolved in order to reduce costs has also been proposed (Murren et al. 2015) so we may expect to see such changes concurrent with a reduction in plasticity costs (5.3.1.4).

Any mutations that we find have spread through the population can be transformed into an ancestral strain to test whether they are adaptive (i.e. whether they increase fitness). Transforming them in isolation may also aid in elucidating their mechanism. However, we must be aware of the possibility of epistatic effects when considering both fitness and mechanistic effects.

5.3.2.2 *Testing a model of gene regulation evolution*

Proulx & Smiley (2010) model the evolution of regulation of a single gene in two randomly occurring environments, looking primarily at the early evolution of transcription rates, making the analysis very applicable to our experiments. They assume no frequency dependence in selection, which should be minimal as our cells will mainly be in their exponential growth phase. By keeping the cells constantly in growth phase, and using the hard selection regime in our GFP experiment (5.2.3), we can be even more certain this assumption is not violated. They analyse evolution from the starting point of expression being constitutive and optimal in one environment, which we could potentially achieve using strains that have undergone selection in one environment. The only assumption which we violate is that the environment switches on a timescale below a single lifespan, however we can assume that our clonal population of fissioning cells has a perpetual biochemistry, and that fitness is a function of expression level (4.1.4), and thus the proxy for total fitness used in the model is still valid.

The model predicts that regulation will usually evolve when it is the global optima, but may move between different strategies of constitutive expression first, paralleling the evolution of specialism and generalism in models of phenotypic plasticity (5.3.1.1). The model predicts that intermediate constitutive expression (ICE) can evolve when a constitutive change towards an intermediate value gives larger fitness benefits in one environment than losses in the other. Thus, ICE is favoured over expression specialised to one environment when we set up our experiments with convex fitness functions (5.2.3). The model predicts that whenever ICE is favoured regulation will be favoured to evolve from any constitutive value. Under conditions where ICE is not favoured regulation can still evolve from a constitutive expression level specialised to one environment. Thus, we may expect to see constitutive expression if our populations evolve towards intermediate or specialised expression levels, but only in the case

of evolution towards intermediate expression levels should we invariably expect the evolution of regulation. In the case of a specialised expression level being favoured (i.e. under concave fitness functions), regulation may evolve, or it may not arise because of a fitness valley, or because of it not being the global optima.

Fitness valleys occur only under specific fitness function shapes where expression changes near endpoints have relatively low effects on fitness. To assess this phenomenon's potential in our *ura4* experiments we would have to ascertain exact fitness functions by artificially altering expression (5.2.3; 4.1.4). In our GFP experiments we have full control of fitness functions, and could alter them to explore functions where fitness valleys could occur. If we find a valley we could use competition with strains that had evolved regulation to show that it is a global optima. The occurrence and duration of constitutive and regulatory strategies will be heavily influenced by the relative mutational availability for the different strategies (2.1.5).

When plasticity does not evolve in the model it is due to the constraint imposed by mismatching of phenotype and environment. This mismatch is caused by a lag time in expression change when switching environments, due to the time taken to produce and degrade the protein. Thus it is intrinsically connected with the rate of environmental switching. If the rate of switching is too high then regulation can never produce a phenotype that matches the environment, and a constitutive level of expression is favoured, specialised to one environment. Therefore, a higher protein degradation rate lets the phenotype change faster, and thus regulation is more likely to evolve under higher rates of switching. The model therefore predicts that if we conducted evolution experiments with different rates of switching we should find a rate above which regulation stops evolving. This result is intuitive when one considers the degree of lag that must occur when changing phenotype in any example of plasticity, and indeed there are models of plasticity which make similar predictions (Moran 1992; Levins 1968; Jablonka & Szathmary 1995).

However, in gene regulation the lag time can be altered by changing the rate of protein degradation, and this possibility is modelled by Proulx & Smiley. They find that small increases in degradation rates will be favoured, and that because these then alter the balance of transcription and degradation, the rate of transcription is increased to compensate. Thus, when looking at our experiments we should expect to find increases in degradation and transcription rates at any point during regulatory evolution. However, this cycle of increase should cease at a certain point depending on the cost of expression, and this prediction is born out in the model when a cost of expression is introduced. The introduction of a cost of expression can also prevent regulation from evolving, but only when costs are large relative to the benefits of expression, which is not likely to be the case in our experiments.

This model provides us with the slightly counter-intuitive prediction that we are more likely to evolve regulation under a convex fitness set, where intermediate constitutive expression is favoured, rather than under the highly divergent selection pressures of a concave fitness set, where theory generally suggests that plasticity provides the largest benefits (DeWitt & Langerhans 2004). However, regulation can still evolve in both cases when protein turnover is high relative to the rate of environmental change, and in our experiments we may find the constraints treated in this model not as relevant as others.

5.4 CONCLUSION

Phenotypic plasticity and environment dependent gene regulation are closely related areas of study, and this is reflected in the similarities of their theoretical exploration. Experiments with our systems are an attempt to answer questions in both areas by observing evolution of a regulated phenotype. The questions in these areas can be summarised as follows: What phenotypes will evolve under changing environments? What are the dynamics of this evolution? What factors will restrict the evolution of these phenotypes? What mechanisms will underlie the phenotypes? This final question concerns a significant factor in answering the others. We hope to complete experiments that will complement theory in providing answers to these questions.

Poelwijk et al. (2011) investigated similar questions in a study of *Escherichia coli* with many parallels to our own. They constructed an operon under the regulation of the *LacI* repressor for which expression was strongly selected for in one environment, and strongly selected against in another environment. They artificially regulated the operon to build a fitness set for the two environments, and then used introduction of mutation by error prone PCR and rounds of selection to enrich for regulatory phenotypes. The regulatory phenotypes they found matched the predictions of optimality theory, given their fitness functions. However, they did discover possible restrictions and dead-ends in evolving regulation, and these were only explained by dissecting the molecular mechanisms behind the phenotype changes. They found significant variety in the effects of different mutations leading to regulatory phenotypes.

The findings of Poelwijk et al. reinforce our approach of working in parallel at both the phenotypic and mechanistic level in analysing our experiments and addressing hypotheses. It is encouraging to see regulation evolve by diverse mechanisms in line with theoretical predictions, with interesting restrictions related to molecular mechanisms. We hope that our future experiments can complement this by reaching broad conclusions that can be understood in terms of general theories of plasticity, as well as specific conclusions related to constraints and molecular mechanisms. Our experiments involve strong selective pressure on the regulation of a single gene, rather than an operon, which could potentially mutate to separate the regulation of its component genes. This allows us to complete a long-term study of the evolution of

plasticity and gene regulation, where complex and dynamic adaptation has the opportunity to occur. Only by allowing these processes to fully play out can we properly relate experimental results to evolution in nature and the theories which attempt to explain it.

We have developed two tightly controlled and complementary systems which show great potential in carrying out the evolution experiments described. We have discussed how, through these experiments, our systems could be used to progress theories and knowledge of gene regulation and phenotypic plasticity. The fundamental importance of these two processes makes understanding of them crucial to the progression of the life sciences. Specifically, there are two unanswered questions in current biology that directly relate to these two processes, and that have sweeping implications for our understanding of life on Earth: Can we find a paradigm for understanding how genotype determines phenotype? Can we explain the complexity and diversity of life? Progress towards answering the first question will surely involve a greater understanding of gene regulation, which is as complex as it is ubiquitous in the determination of phenotype from genotype. The second question is inextricably linked with the unknown constraints that must exist on adaptive phenotypic plasticity. The complexity and diversity of life are no doubt dependent on the varied environmental niches in which elaborate adaptation can allow an organism to thrive. With plasticity, many environments could be exploited by one species, but this is far from what we see. Clearly the constraints on plasticity determine the breadth of a niche that a species can fill. Furthermore, constraints on adaptation will increase ecological complexity creating yet more vacant niches. However, current understanding of which constraints are most influential on plasticity evolution, and the specific effects they have, is decidedly lacking. Insights provided by experimentation with our systems may yield conclusions that directly contribute towards answering these questions of foremost importance.

6 REFERENCES

- Andersen, G., Bjornberg, O., Polakova, S., Pynyaha, Y., Rasmussen, A., Moller, K., Hofer, A., Moritz, T., Sandrini, M. P. B., Merico, A. M., Compagno, C., Akerlund, H. E., Gojkovic, Z. and Piskur, J. (2008). 'A second pathway to degrade pyrimidine nucleic acid precursors in eukaryotes'. *Journal of Molecular Biology*, 380 (4), 656-666.
- Arnoldini, M., Mostowy, R., Bonhoeffer, S. and Ackermann, M. (2012). 'Evolution of Stress Response in the Face of Unreliable Environmental Signals'. *Plos Computational Biology*, 8 (8), 10.
- Auld, J. R., Agrawal, A. A. and Relyea, R. A. (2010). 'Re-evaluating the costs and limits of adaptive phenotypic plasticity'. *Proceedings of the Royal Society B-Biological Sciences*, 277 (1681), 503-511.
- Bähler, J., Wu, J. Q., Longtine, M. S., Shah, N. G., McKenzie, A., Steever, A. B., Wach, A., Philippsen, P. and Pringle, J. R. (1998). 'Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*'. *Yeast*, 14 (10), 943-951.
- Beaudoin, J. and Labbe, S. (2006). 'Copper induces cytoplasmic retention of fission yeast transcription factor Cuf1'. *Eukaryotic Cell*, 5 (2), 277-292.
- Beaumont, H. J. E., Gallie, J., Kost, C., Ferguson, G. C. and Rainey, P. B. (2009). 'Experimental evolution of bet hedging'. *Nature*, 462 (7269), 90-97.
- Bedford, T. and Hartl, D. L. (2009). 'Optimization of gene expression by natural selection'. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (4), 1133-1138.
- Berrigan, D. and Scheiner, S. M. (2004). 'Modelling the evolution of phenotypic plasticity', *Phenotypic plasticity: functional and conceptual approaches* (pp. 82-97). Oxford University Press.
- Callahan, H. S., Maughan, H. and Steiner, U. K. (2008). 'Phenotypic Plasticity, Costs of Phenotypes, and Costs of Plasticity Toward an Integrative View'. *Year in Evolutionary Biology 2008*, 1133, 44-66.
- Capuano, F., Mulleder, M., Kok, R., Blom, H. J. and Ralser, M. (2014). 'Cytosine DNA Methylation Is Found in *Drosophila melanogaster* but Absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and Other Yeast Species'. *Analytical Chemistry*, 86 (8), 3697-3702.
- Check, E. (2007). 'RNA interference - Hitting the on switch'. *Nature*, 448 (7156), 855-858.
- Chen, D. R., Toone, W. M., Mata, J., Lyne, R., Burns, G., Kivinen, K., Brazma, A., Jones, N. and Bähler, J. (2003). 'Global transcriptional responses of fission yeast to environmental stress'. *Molecular Biology of the Cell*, 14 (1), 214-229.
- Chen, K. and Rajewsky, N. (2007). 'The evolution of gene regulation by transcription factors and microRNAs'. *Nature Reviews Genetics*, 8 (2), 93-103.
- Connelly, C. F., Skelly, D. A., Dunham, M. J. and Akey, J. M. (2013). 'Population Genomics and Transcriptional Consequences of Regulatory Motif Variation in Globally Diverse *Saccharomyces cerevisiae* Strains'. *Molecular Biology and Evolution*, 30 (7), 1605-1613.
- Crnokrak, P. and Roff, D. A. (1995). 'Dominance variance – associations with selection and fitness'. *Heredity*, 75, 530-540.
- Cui, P. and Xiong, L. M. (2015). 'Environmental Stress and Pre-mRNA Splicing'. *Molecular Plant*, 8 (9), 1302-1303.
- Dai, H. and Wang, Z. (2014). 'Histone modification patterns and their responses to environment'. *Current environmental health reports*, 1 (1), 11-21.
- Darwin, C. (1859). *On the origin of species*. John Murray.
- Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.
- Dawkins, R. (1982). *The extended phenotype: The gene as the unit of selection*. Oxford University Press.
- De Jong, G. (1989). 'Phenotypically plastic characters in isolated populations', *Evolutionary biology of transient unstable populations* (pp. 3-18): Springer.

- De Jong, G. (1995). 'Phenotypic plasticity as a product of selection in a variable environment'. *American Naturalist*, 145 (4), 493-512.
- De Jong, G. (1999). 'Unpredictable selection in a structured population leads to local genetic differentiation in evolved reaction norms'. *Journal of Evolutionary Biology*, 12 (5), 839-851.
- De Montigny, J., Straub, M. L., Wagner, R., Bach, M. L. and Chevallier, M. R. (1998). 'The uracil permease of *Schizosaccharomyces pombe*: A representative of a family of 10 transmembrane helix transporter proteins of yeasts'. *Yeast*, 14 (11), 1051-1059.
- Dennett, D. (1995). *Darwin's dangerous idea*. Simon & Schuster.
- DeWitt, T. J. (2016). 'Expanding the phenotypic plasticity paradigm to broader views of trait space and ecological function'. *Current Zoology*, 62 (5), 463-473.
- DeWitt, T. J. and Langerhans, R. B. (2004). 'Integrated solutions to environmental heterogeneity: theory of multimoment reaction norms', *Phenotypic plasticity: functional and conceptual approaches* (pp. 98-111). Oxford University Press.
- DeWitt, T. J. and Scheiner, S. M. (2004). *Phenotypic plasticity: functional and conceptual approaches*: Oxford University Press.
- DeWitt, T. J., Sih, A. and Wilson, D. S. (1998). 'Costs and limits of phenotypic plasticity'. *Trends in Ecology & Evolution*, 13 (2), 77-81.
- Doniger, S. W. and Fay, J. C. (2007). 'Frequent gain and loss of functional transcription factor binding sites'. *Plos Computational Biology*, 3 (5), 932-942.
- Ernande, B. and Dieckmann, U. (2004). 'The evolution of phenotypic plasticity in spatially structured environments: implications of intraspecific competition, plasticity costs and environmental characteristics'. *Journal of Evolutionary Biology*, 17 (3), 613-628.
- Fantes, P. and Nurse, P. (1977). 'Control of cell-size at division in fission yeast by a growth-modulated size control over nuclear division'. *Experimental Cell Research*, 107 (2), 377-386.
- Field, Y., Fondufe-Mittendorf, Y., Moore, I. K., Mieczkowski, P., Kaplan, N., Lubling, Y., Lieb, J. D., Widom, J. and Segal, E. (2009). 'Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization'. *Nature Genetics*, 41 (4), 438-445.
- Foust, C. M., Preite, V., Schrey, A. W., Alvarez, M., Robertson, M. H., Verhoeven, K. J. F. and Richards, C. L. (2016). 'Genetic and epigenetic differences associated with environmental gradients in replicate populations of two salt marsh perennials'. *Molecular Ecology*, 25 (8), 1639-1652.
- Garland, T. and Kelly, S. A. (2006). 'Phenotypic plasticity and experimental evolution'. *Journal of Experimental Biology*, 209 (12), 2344-2361.
- Ghalambor, C. K., McKay, J. K., Carroll, S. P. and Reznick, D. N. (2007). 'Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments'. *Functional Ecology*, 21 (3), 394-407.
- Gilbert, S. F. (2005). 'Mechanisms for the environmental regulation of gene expression: Ecological aspects of animal development'. *Journal of Biosciences*, 30 (1), 65-74.
- Graham, J. K., Smith, M. L. and Simons, A. M. (2014). 'Experimental evolution of bet hedging under manipulated environmental uncertainty in *Neurospora crassa*'. *Proceedings of the Royal Society B-Biological Sciences*, 281 (1787), 6.
- Grishkevich, V. and Yanai, I. (2013). 'The genomic determinants of genotype x environment interactions in gene expression'. *Trends in Genetics*, 29 (8), 479-487.
- Hodgins-Davis, A. and Townsend, J. P. (2009). 'Evolving gene expression: from G to E to G x E'. *Trends in Ecology & Evolution*, 24 (12), 649-658.
- Jablonka, E. and Szathmari, E. (1995). 'The evolution of information-storage and heredity'. *Trends in Ecology & Evolution*, 10 (5), 206-211.
- Javerzat, J. P., Cranston, G. and Allshire, R. C. (1996). 'Fission yeast genes which disrupt mitotic chromosome segregation when overexpressed'. *Nucleic Acids Research*, 24 (23), 4676-4683.
- Jund, R., Weber, E. and Chevallier, M. R. (1988). 'Primary structure of the uracil transport protein of *Saccharomyces cerevisiae*'. *European Journal of Biochemistry*, 171 (1-2), 417-424.

- Kalisky, T., Dekel, E. and Alon, U. (2007). 'Cost-benefit theory and optimal design of gene regulation functions'. *Physical Biology*, 4 (4), 229-245.
- Kawecki, T. J., Lenski, R. E., Ebert, D., Hollis, B., Olivieri, I. and Whitlock, M. C. (2012). 'Experimental evolution'. *Trends in Ecology & Evolution*, 27 (10), 547-560.
- Krebs, J. R. and Davies, N. B. (1996). *Behavioural ecology: an evolutionary approach*: Blackwell, Oxford.
- Kristell, C., Westholm, J. O., Olsson, I., Ronne, H., Komorowski, J. and Bjerling, P. (2010). 'Nitrogen depletion in the fission yeast *Schizosaccharomyces pombe* causes nucleosome loss in both promoters and coding regions of activated genes'. *Genome Research*, 20 (3), 361-371.
- Kristiansson, E., Thorsen, M., Tamas, M. J. and Nerman, O. (2009). 'Evolutionary Forces Act on Promoter Length: Identification of Enriched Cis-Regulatory Elements'. *Molecular Biology and Evolution*, 26 (6), 1299-1307.
- Kussell, E. and Leibler, S. (2005). 'Phenotypic diversity, population growth, and information in fluctuating environments'. *Science*, 309 (5743), 2075-2078.
- Kwon, E. J. G., Laderoute, A., Chatfield-Reed, K., Vachon, L., Karagiannis, J. and Chua, G. (2012). 'Deciphering the Transcriptional-Regulatory Network of Flocculation in *Schizosaccharomyces pombe*'. *Plos Genetics*, 8 (12), 23.
- Labbe, S., Pelletier, B. and Mercier, A. (2007). 'Iron homeostasis in the fission yeast *Schizosaccharomyces pombe*'. *Biomaterials*, 20 (3-4), 523-537.
- Landry, C. R., Lemos, B., Rifkin, S. A., Dickinson, W. J. and Hartl, D. L. (2007). 'Genetic properties influencing the evolvability of gene expression'. *Science*, 317 (5834), 118-121.
- Lee, T. I. and Young, R. A. (2000). 'Transcription of eukaryotic protein-coding genes'. *Annual Review of Genetics*, 34, 77-137.
- Levins, R. (1963). 'Theory of fitness in a heterogeneous environment .1. Developmental flexibility and niche selection'. *American Naturalist*, 97 (893), 75.
- Levins, R. (1968). *Evolution in changing environments: some theoretical explorations*: Princeton University Press.
- León, J. A. (1993). 'Plasticity in fluctuating environments', *Adaptation in stochastic environments* (pp. 105-121): Springer.
- Li, H. and Johnson, A. D. (2010). 'Evolution of Transcription Networks - Lessons from Yeasts'. *Current Biology*, 20 (17), R746-R753.
- Liu, M., Xu, R. Q., Merrill, C., Hong, L. W., Von Lanken, C., Hunt, A. G. and Li, Q. S. Q. (2014). 'Integration of Developmental and Environmental Signals via a Polyadenylation Factor in Arabidopsis'. *Plos One*, 9 (12), 23.
- Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R. and Bähler, J. (2012). 'Quantitative Analysis of Fission Yeast Transcriptomes and Proteomes in Proliferating and Quiescent Cells'. *Cell*, 151 (3), 671-683.
- Mata, J., Lyne, R., Burns, G. and Bahler, J. (2002). 'The transcriptional program of meiosis and sporulation in fission yeast'. *Nature Genetics*, 32 (1), 143-147.
- Matsuo, Y., Nishino, K., Mizuno, K., Akihiro, T., Toda, T., Kaino, T. and Kawamukai, M. (2013). 'Polypeptide Induces Dramatic Cell Lysis in *ura4* Deletion Mutants of Fission Yeast'. *Plos One*, 8 (3), 13.
- McCullum, E. O., Williams, B. A. R., Zhang, J. L. and Chaput, J. C. (2010). 'Random Mutagenesis by Error-Prone PCR'. In J. Braman (Ed.), *In Vitro Mutagenesis Protocols, Third Edition* (Vol. 634, pp. 103-109). Totowa: Humana Press Inc.
- Mercier, A., Pelletier, B. and Labbe, S. (2006). 'A transcription factor cascade involving Fep1 and the CCAAT-binding factor Php4 regulates gene expression in response to iron deficiency in the fission yeast *Schizosaccharomyces pombe*'. *Eukaryotic Cell*, 5 (11), 1866-1881.
- Moran, N. A. (1992). 'The evolutionary maintenance of alternative phenotypes'. *American Naturalist*, 139 (5), 971-989.
- Moreno, S., Klar, A. and Nurse, P. (1991). 'Molecular genetic-analysis of fission yeast *Schizosaccharomyces pombe*'. *Methods in Enzymology*, 194, 795-823.

- Moses, A. M., Pollard, D. A., Nix, D. A., Iyer, V. N., Li, X. Y., Biggin, M. D. and Eisen, M. B. (2006). 'Large-scale turnover of functional transcription factor binding sites in *Drosophila*'. *Plos Computational Biology*, 2 (10), 1219-1231.
- Murren, C. J., Auld, J. R., Callahan, H., Ghahambor, C. K., Handelsman, C. A., Heskell, M. A., Kingsolver, J. G., Maclean, H. J., Masel, J., Maughan, H., Pfennig, D. W., Relyea, R. A., Seiter, S., Snell-Rood, E., Steiner, U. K. and Schlichting, C. D. (2015). 'Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity'. *Heredity*, 115 (4), 293-301.
- Mustonen, V. and Lassig, M. (2005). 'Evolutionary population genetics of promoters: Predicting binding sites and functional phylogenies'. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (44), 15936-15941.
- Nishino, K., Kushima, M., Matsuo, Y. and Kawamukai, M. (2015). 'Cell Lysis in *S. pombe* *ura4* Mutants Is Suppressed by Loss of Functional Pub1, Which Regulates the Uracil Transporter Fur4'. *Plos One*, 10 (11), 17.
- Nurse, P. (1975). 'Genetic control of cell-size at cell-division in yeast'. *Nature*, 256 (5518), 547-551.
- Paluszynski, J. P., Klassen, R., Rohe, M. and Meinhardt, F. (2006). 'Various cytosine/adenine permease homologues are involved in the toxicity of 5-fluorocytosine in *Saccharomyces cerevisiae*'. *Yeast*, 23 (9), 707-715.
- Pelletier, B., Beaudoin, J., Philpott, C. C. and Labbe, S. (2003). 'Fep1 represses expression of the fission yeast *Schizosaccharomyces pombe* siderophore-iron transport system'. *Nucleic Acids Research*, 31 (15), 4332-4344.
- Petersen, J. and Russell, P. (2016). 'Growth and the environment of *Schizosaccharomyces pombe*'. *Cold Spring Harbor Protocols*, 2016 (3).
- Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. and Bejerano, G. (2013). 'Enhancers: five essential questions'. *Nature Reviews Genetics*, 14 (4), 288-295.
- Pigliucci, M. (1996). 'Modelling phenotypic plasticity .2. Do genetic correlations matter?'. *Heredity*, 77, 453-460.
- Poelwijk, F. J., de Vos, M. G. J. and Tans, S. J. (2011). 'Tradeoffs and Optimality in the Evolution of Gene Regulation'. *Cell*, 146 (3), 462-470.
- Prasher, D. C. (1995). 'Using GFP to see the light'. *Trends in Genetics*, 11 (8), 320-323.
- Proulx, S. R. and Smiley, M. W. (2010). 'The Evolutionary Origins of Gene Regulation'. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, 314B (4), 327-340.
- Rebeiz, M., Jikomes, N., Kassner, V. A. and Carroll, S. B. (2011). 'Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences'. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (25), 10036-10043.
- Roelofs, D., Morgan, J. and Sturzenbaum, S. (2010). 'The significance of genome-wide transcriptional regulation in the evolution of stress tolerance'. *Evolutionary Ecology*, 24 (3), 527-539.
- Rosin, D., Hornung, G., Tirosh, I., Gispán, A. and Barkai, N. (2012). 'Promoter Nucleosome Organization Shapes the Evolution of Gene Expression'. *Plos Genetics*, 8 (3), 10.
- Rustici, G., van Bakel, H., Lackner, D. H., Holstege, F., Wijmenga, C., Bahler, J. and Brazma, A. (2007). 'Global transcriptional responses of fission and budding yeast to changes in copper and iron levels: a comparative study'. *Genome Biology*, 8 (5), 57.
- Sasidharan, R., Chinnappa, C. C., Staal, M., Elzenga, J. T. M., Yokoyama, R., Nishitani, K., Voesenek, L. and Pierik, R. (2010). 'Light Quality-Mediated Petiole Elongation in *Arabidopsis* during Shade Avoidance Involves Cell Wall Modification by Xyloglucan Endotransglucosylase/Hydrolases'. *Plant Physiology*, 154 (2), 978-990.
- Sato, M., Dhut, S. and Toda, T. (2005). 'New drug-resistant cassettes for gene disruption and epitope tagging in *Schizosaccharomyces pombe*'. *Yeast*, 22 (7), 583-591.
- Savageau, M. A. (2001). 'Design principles for elementary gene circuits: Elements, methods, and examples'. *Chaos*, 11 (1), 142-159.
- Scheiner, S. M. and Berrigan, D. (1998). 'The genetics of phenotypic plasticity. VIII. The cost of plasticity in *Daphnia pulex*'. *Evolution*, 52 (2), 368-378.

- Schott, D., Yanai, I. and Hunter, C. P. (2014). 'Natural RNA interference directs a heritable response to the environment'. *Scientific Reports*, 4, 10.
- Seiple, L., Jaruga, P., Dizdaroglu, M. and Stivers, J. T. (2006). 'Linking uracil base excision repair and 5-fluorouracil toxicity in yeast'. *Nucleic Acids Research*, 34 (1), 140-151.
- Seron, K., Blondel, M. O., Haguenaer-Tsapis, R. and Volland, C. (1999). 'Uracil-induced down-regulation of the yeast uracil permease'. *Journal of Bacteriology*, 181 (6), 1793-1800.
- Sherman, M. S. and Cohen, B. A. (2012). 'Thermodynamic State Ensemble Models of cis-Regulation'. *Plos Computational Biology*, 8 (3), 10.
- Sih, A. (2004). 'A behavioral ecological view of phenotypic plasticity.'. In D. a. Scheiner (Ed.), *Phenotypic plasticity: functional and conceptual approaches*. (pp. 112 - 125): Oxford Univ. Press.
- Sivakumar, S., Porter-Goff, M., Patel, P. K., Benoit, K. and Rhind, N. (2004). 'In vivo labeling of fission yeast DNA with thymidine and thymidine analogs'. *Methods*, 33 (3), 213-219.
- Snell-Rood, E. C., Van Dyken, J. D., Cruickshank, T., Wade, M. J. and Moczek, A. P. (2010). 'Toward a population genetic framework of developmental evolution: the costs, limits, and consequences of phenotypic plasticity'. *Bioessays*, 32 (1), 71-81.
- Starrfelt, J. and Kokko, H. (2012). 'Bet-hedging - a triple trade-off between means, variances and correlations'. *Biological Reviews*, 87 (3), 742-755.
- Stearns, S. C. and Koella, J. C. (1986). 'The evolution of phenotypic plasticity in life-history traits - predictions of reaction norms for age and size at maturity'. *Evolution*, 40 (5), 893-913.
- Steinberg, B. and Ostermeier, M. (2016). 'Environmental changes bridge evolutionary valleys'. *Science Advances*, 2 (1), 9.
- Stewart, A. J., Plotkin, J. B. and Hannenhalli, S. (2012). 'Why Transcription Factor Binding Sites Are Ten Nucleotides Long'. *Genetics*, 192 (3), 973.
- Stone, J. R. and Wray, G. A. (2001). 'Rapid evolution of cis-regulatory sequences via local point mutations'. *Molecular Biology and Evolution*, 18 (9), 1764-1770.
- Sultan, S. E. and Spencer, H. G. (2002). 'Metapopulation structure favors plasticity over local adaptation'. *American Naturalist*, 160 (2), 271-283.
- Svanback, R., Pineda-Krch, M. and Doebeli, M. (2009). 'Fluctuating Population Dynamics Promotes the Evolution of Phenotypic Plasticity'. *American Naturalist*, 174 (2), 176-189.
- Thattai, M. and van Oudenaarden, A. (2004). 'Stochastic gene expression in fluctuating environments'. *Genetics*, 167 (1), 523-530.
- Tirosh, I., Barkai, N. and Verstrepen, K. J. (2009). 'Promoter architecture and the evolvability of gene expression'. *Journal of biology*, 8 (11), 95.
- Tufto, J. (2000). 'The evolution of plasticity and nonplastic spatial and temporal adaptations in the presence of imperfect environmental cues'. *American Naturalist*, 156 (2), 121-130.
- Vachon, L., Wood, J., Kwon, E. J. G., Laderoute, A., Chatfield-Reed, K., Karagiannis, J. and Chua, G. (2013). 'Functional Characterization of Fission Yeast Transcription Factors by Overexpression Analysis'. *Genetics*, 194 (4), 873.
- Van Buskirk, J. and Steiner, U. K. (2009). 'The fitness costs of developmental canalization and plasticity'. *Journal of Evolutionary Biology*, 22 (4), 852-860.
- Van Kleunen, M. and Fischer, M. (2005). 'Constraints on the evolution of adaptive phenotypic plasticity in plants'. *New Phytologist*, 166 (1), 49-60.
- Van Tienderen, P. H. (1991). 'Evolution of generalists and specialists in spatially heterogeneous environments'. *Evolution*, 45 (6), 1317-1331.
- Van Tienderen, P. H. and Koelewijn, H. P. (1994). 'Selection on reaction norms, genetic correlations and constraints'. *Genetical Research*, 64 (2), 115-125.
- Veening, J. W., Smits, W. K. and Kuipers, O. P. (2008). 'Bistability, Epigenetics, and Bet-Hedging in Bacteria', *Annual Review of Microbiology* 62, 193-210.
- Via, S., Gomulkiewicz, R., Dejong, G., Scheiner, S. M., Schlichting, C. D. and Vantienderen, P. H. (1995). 'Adaptive phenotypic plasticity - consensus and controversy'. *Trends in Ecology & Evolution*, 10 (5), 212-217.

- Via, S. and Lande, R. (1985). 'Genotype-environment interaction and the evolution of phenotypic plasticity'. *Evolution*, 39 (3), 505-522.
- Villalba, A., Coll, O. and Gebauer, F. (2011). 'Cytoplasmic polyadenylation and translational control'. *Current Opinion in Genetics & Development*, 21 (4), 452-457.
- Watt, S., Mata, J., Lopez-Maury, L., Marguerat, S., Burns, G. and Bähler, J. (2008). 'urg1: A Uracil-Regulatable Promoter System for Fission Yeast with Short Induction and Repression Times'. *Plos One*, 3 (1), 8.
- Whangbo, J. S. and Hunter, C. P. (2008). 'Environmental RNA interference'. *Trends in Genetics*, 24 (6), 297-305.
- Wilson, D. S. and Keefe, A. D. (2001). 'Random mutagenesis by PCR'. *Current protocols in molecular biology*. 8, 8.3.
- Wray, G. A. (2007). 'The evolutionary significance of cis-regulatory mutations'. *Nature Reviews Genetics*, 8 (3), 206-216.
- Zhivotovsky, L. A., Feldman, M. W. and Bergman, A. (1996). 'On the evolution of phenotypic plasticity in a spatially heterogeneous environment'. *Evolution*, 50 (2), 547-558.

7 APPENDICES

7.1 GROWTH MEDIA (*EMM – EDINBURGH MINIMAL MEDIUM*)

Based on the media from Nurse 1975 (Solid media is made by adding 2% Difco Bacto Agar):

EMM: 3 g/l Potassium hydrogen phthalate; 2.2 g/l Na₂HPO₄; 5 g/l NH₄Cl; 20 g/l Glucose; 20 ml/l **Salts**; 1 ml/l **Vitamins**; 0.1 ml/l **Minerals**.

Salts (50x): 52.5 g/l MgCl₂·6H₂O; 0.735 g/l CaCl₂·2H₂O; 50 g/l KCl; 2 g/l Na₂SO₄.

Vitamins (1000x): 1 g/l pantothenic acid; 10 g/l nicotinic acid; 10 g/l inositol; 10 mg/l biotin.

Minerals (10,000x): 5 g/l boric acid; 4 g/l MnSO₄; 4 g/l ZnSO₄·7H₂O; 2 g/l FeCl₂·6H₂O; 0.4 g/l molybdc acid; 1 g/l KI; 0.4 g/l CuSO₄·5H₂O; 10 g/l citric acid.

For a media with uracil as a sole nitrogen source (**5.1.2**) the composition is the same apart from the replacement of ammonium chloride (NH₄Cl) with uracil 10 mM (Fantes & Nurse 1977; Petersen & Russell 2016).

7.2 EXPERIMENTAL EVOLUTION PROTOCOL (*PRELIMINARY EXPERIMENT – 4.1.6*)

All cell transfers should be completed with strict sterile technique to prevent cross-contamination.

- 1) Grow 24hr pre-culture. Centrifuge at 650rpm for 3 min and remove supernatant.
- 2) In ten 60ml capped glass tubes add 10ml of cells suspended in EMM at optical density (OD) 0.1 in EMM (For example add 1ml OD 1 cells to 9ml media). These are 10 independent evolution lines.
- 3) Incubate for 24hrs at 32°C and 170rpm (no-ura environment).
- 4) Centrifuge each line at 650rpm for 3 min and remove supernatant.
- 5) Resuspend each line in 10ml EMM with 0.05mg/ml uracil and 0.4mg/ml 5-FOA at OD 0.1.
- 6) Incubate for 24hrs at 32°C and 170rpm (FOA environment).
- 7) Repeat steps 2-6 five times (5 full environmental cycles).

7.3 *S. POMBE TRANSFORMATION - LITHIUM ACETATE PROTOCOL*

DNA to insert is amplified with primers with 80bp homologous sequences at each end for the insertion site.

Solutions required:

LiAc-TE: 1ml LiAc at 1M (or 10x); 1ml TE (10x); Water 8ml.

LiAc-TE-PEG: 1ml LiAc at 1M (or 10x); 1ml TE (10x); PEG 50% 8ml.

- 1) Pre-culture in 50ml liquid YES media until optical density (OD_{600}) reaches 0.5.
- 2) Transfer to 50ml tube and centrifuge for 5min at 4000rpm.
- 3) Pour off supernatant. Wash with 50ml sterile water.
- 4) Resuspend in 1ml water, transfer to a 1ml tube, and centrifuge for 4min at 2500rpm. Wash with 1ml of LiAc-TE.
- 5) Resuspend in 250 μ l LiAc-TE.
- 6) Separate into 100 μ l aliquots (add one negative control which won't have DNA added) and add to each tube: 2 μ l of carrier DNA (e.g. boiled herring sperm cell DNA) and up to 10 μ l of DNA to insert.
- 7) Incubate for 10min at room temperature.
- 8) Resuspend in 250 μ l LiAc-TE-PEG (freshly mixed). Mix gently.
- 9) Incubate for 60min at 30°C. Pre-warm DMSO at 42°C.
- 10) Add 47 μ l of pre-warmed DMSO and mix gently.
- 11) Heat shock at 42°C for 5min. Cool the tube for 2min on ice.
- 12) Centrifuge for 4min at 2500rpm. Remove supernatant. Wash with 1ml sterile water.
- 13) Centrifuge for 4min at 2500rpm. Resuspend in 500 μ l of water and spread on YES plates. Spread 250 μ l/plate with medium beads. Allow to dry then incubate at 32°C.
- 14) After 18 hours duplicate plates onto selection media (can plate straight away onto selective media in previous step but may reduce efficiency).

7.4 RNA EXTRACTION

Solution required: TES – 10mM Tris pH 7.5; 10mM EDTA pH8; 0.5% SDS. Use DEPC water and store at room temperature.

- 1) Harvest cells from culture – approximately 25ml at optical density (OD₆₀₀) 0.2 or equivalent. Centrifuge for 2min at 2000rpm and discard supernatant. Snap freeze pellet on liquid nitrogen or dry ice and ethanol. Alternatively, filter cells and snap freeze the filter disc. Store cells at -70°C.
- 2) Thaw cells on ice. Add 1ml pre-chilled DEPC water, resuspend cells, and transfer to 2ml Eppendorf tubes. Spin 10sec at 5000rpm and remove supernatant.
- 3) Add 750µl of TES (adjust if total cells are over OD₆₀₀ 5). Resuspend cells with pipette and immediately add 750µl acidic phenol-chloroform (shake and ensure well mixed), vortex and incubate at 65°C. Perform this step separately for each sample.
- 4) Incubate at 65°C for 1hr and vortex for 10sec every 10min.
- 5) Place samples on ice for 1min (can increase to 5-10min), vortex for 20sec, and centrifuge for 15min at 20,000rcf at 4°C.
- 6) Pre-spin 2ml ‘phase-lock’ tubes for 10sec. Add 700µl of acidic phenol-chloroform.
- 7) Take 700µl of the water phase from step 5 and add to the phase-lock tubes from step 6. Thoroughly mix by inverting (do not vortex) and centrifuge 5min at 20,000rcf at 4°C.
- 8) Pre-spin 2ml phase-lock tubes for 10sec. Add 700µl of chloroform:isoamyl alcohol (24:1).
- 9) Take 700µl of the water phase from step 7 and add to the phase-lock tubes from step 8. Thoroughly mix by inverting (do not vortex) and centrifuge 5min at 20,000rcf at 4°C.
- 10) Prepare 2ml Eppendorf tubes with 1.5ml of 100% EtOH (-20°C) and 50µl of 3M NaAc pH 5.2.
- 11) Transfer 500µl of water phase from step 9 to the tubes from step 10 and vortex 10sec. Samples can be precipitated at -20°C overnight (or at -70°C for 30min).
- 12) Centrifuge for 10min at 20,000rcf at room temperature. Discard supernatant. Add 500µl 70% EtOH (4°C and made with DEPC water). Do not vortex. Spin for 1min (with tube in same orientation). Aspirate most of the supernatant. Spin for another 5sec then remove the rest of the liquid by pipette. Air dry 5min at room temperature.
- 13) Add 100µl of DEPC water and incubate for 10min at room temperature. Dissolve pellet first by pipetting up and down approximately 30 times until no particles remain, then gently vortex for 10sec. Can run a sample of this on a gel to check the RNA has not degraded
- 14) This protocol produces approximately 200µg of RNA. 100µg of this was purified for qPCR using RNeasy mini spin columns (Qiagen) following kit protocol and then eluted

twice with 30µl RNase-free water and kept on ice. The remaining volume can be stored in a 1:3 ratio with 100% EtOH and stored at -70°C as a backup.

- 15) Run 2µl of purified RNA on a 1% agarose gel. All apparatus was cleaned with 'RNase-Zap' and new buffer used. The loading buffer must be RNase-free made with DEPC water. This should show two clear ribosomal bands.
- 16) Use 'Nanodrop' with 2µl of purified RNA to determine concentration, and adjust the concentration of the purified sample to 2µg/µl for use in qPCR.

7.5 qPCR

DNase treatment (using reagents from Ambion TURBO DNA free kit #1907):

- 1) Mix 10µg RNA; 5µl 10x buffer; 0.5µl DNaseI; 50µl water.
- 2) Incubate at 37°C for 25min.
- 3) Add 10µl of inactivation reagent.
- 4) Incubate at room temperature for 2 minutes and mix occasionally.
- 5) Spin at 10,000xg for 2mins and transfer supernatant (contains RNA) to a fresh tube.

Reverse transcription (using reagents from Superscript II Reverse Transcriptase kit #18064-022):

- 6) Mix 2µl random primers (100ng/µl stock); 10µg RNA (use speed vac to reduce the 50µl above to 9µl or less); 1µl dNTP mix (10mM each); up to 12µl water.
- 7) Heat mixture to 65°C for 7 minutes and quick chill on ice for 2 minutes. Spin down the tube and add 4µl 5X First-Strand buffer; 2µl 0.1M DTT; 1µl RNaseOUT (40U/µl).
- 8) Mix the tube gently and incubate at 25°C for 2min.
- 9) Add 1µl (200 units) of Superscript II Reverse Transcriptase and mix by pipetting gently.
- 10) Incubate at 25°C for 10min.
- 11) Incubate at 42°C for 50min.
- 12) Inactivate the reaction by heating at 70°C for 15min.
- 13) Add 100µl water to have a total of 120µl cDNA. This can be stored if necessary at -20°C.

qPCR (using Fast SYBR Green Master Mix Kit – Applied Biosystems #4385612. qPCR machine – 7900HT Fast):

- 14) Mix the qPCR reaction: 10µl SYBR Green Master Mix (2X); 1µl forward primer (200ng/µl stock); 1µl reverse primer (200ng/µl stock); 8µl cDNA template from step 13. 2 technical replicates are recommended for each sample. Can prepare master mix with primers and add this directly to plate wells containing 8µl cDNA to increase speed for multiples samples using the same primers.
- 15) Transfer the 20µl sample to its assigned well in the plate.

- 16) Standard samples must also be added for quantification. These can either use wild type genomic DNA or a pool of all cDNA samples. Six 1/10th serial dilutions were used to cover a range of start-end points. The dissociation curves for the cDNA samples should fall within the range of curves produced by serial dilution. Reactions can be added to the plate using primers for a gene with high constitutive expression (*tub1* in our case) to normalise expression. One control is also included with no DNA.
- 17) The reaction is then run as per the 7900HT Fast instructions (20 sec enzyme activation, 1sec denaturing, 20sec annealing and extension). The sample results are then analysed against the standard curves.

7.6 PROTEIN EXTRACTION AND WESTERN BLOT

- 1) Harvest cells from culture: 25µl at optical density (OD₆₀₀) 0.2 into Eppendorf tubes.
- 2) Centrifuge at 2500rpm for 5min.
- 3) Resuspend in 300µl lysis buffer.
- 4) Add two protease inhibitors: (i) Complete EDTA-Free (25X). (ii) PMSF (100X – use at 1mM).
- 5) Place 400µl chilled glass beads into screw-cap Eppendorfs, and add cells from step 4 to this.
- 6) Centrifuge for 1min at 6.5m/s (perhaps should be vortexing or shaking, need to check this).
- 7) Ice for 5min.
- 8) Repeat steps 6 and 7.
- 9) Invert the tube and tap liquid away from base. Make a hole in the base with a red hot needle.
- 10) Place this into another open Eppendorf for liquid to flow into. Place the tubes together into a larger Falcon tube.
- 11) Centrifuge this at 4°C for 2min at 2500rpm.
- 12) Can transfer the supernatant to a new tube for a cleaner sample if desired, or to collect insoluble protein from pellet. Centrifuge again at 4°C for 2min at 4000-5000rpm.
- 13) Transfer the supernatant to a new tube. Add to this: 100µl loading buffer (4X); 4µl DTT (100X).
- 14) Boil at 80°C for 10 minutes.
- 15) Load 40µl into western blot wells. Run at 80V for 10-15 minutes then run at 100-120V until the ladder indicates it has ran the distance of the gel.

7.7 ERROR-PRONE PCR

Protocol from McCullum et al. 2010.

Master mixture is prepared as follows:

Reagents	Stock concentration	Volume	Final concentration
Forward primer	100 μ M	15 μ L	1 μ M
Reverse primer	100 μ M	15 μ L	1 μ M
dCTP and dTTP	20 mM	75 μ L each	1 mM
dATP and dGTP	20 mM	15 μ L each	0.2 mM
PCR buffer with Mg^{2+}	10x	150 μ L	1x
$MgCl_2$	1M	8 μ L	~5.5 mM
Nanopure water		1098 μ L	
Final volume		1491	

- 1) Dispense master mixture between 16 PCR tubes with 96 μ L in tube 1 and 88 μ L in tubes 2-16.
- 2) Add 2 μ L DNA template to tube 1.
- 3) Place tube 1 in thermocycler and start the following program:

Step	Temperature	Duration
1 Denaturation	94°C	1 min
2 Annealing	60°C	1 min
3 Extension	72°C	3 min
4 Product storage	4°C	End

Repeat steps 1-3 for 64 cycles.

- 4) Once the program has reached the annealing temperature add 1 μ L of freshly prepared $MnCl_2$ solution and 1 μ L Taq DNA polymerase to the PCR reaction tube.
- 5) Complete 4 cycles of amplification.
- 6) Remove tube 1 and place on ice.
- 7) Transfer 10 μ L of the reaction from tube 1 to tube 2.

- 8) Place tube 2 in the thermocycler and start the program.
- 9) Once the program has reached the annealing temperature add 1 μL of freshly prepared MnCl_2 solution and 1 μL Taq DNA polymerase to the PCR reaction tube.
- 10) Complete 4 cycles of amplification.
- 11) Remove tube 1 and place on ice.
- 12) Repeat steps 7-11 for all tubes up to 16 creating 16 mutagenic libraries with increasing mutation rates.

Libraries can be purified and quantified using gel electrophoresis.

7.8 FLUORESCENCE-ACTIVATED CELL SORTING (FACS)

The first step in FACS is to determine distribution of fluorescence levels among cells in the population. This is done by passing a sample of the population of cells through the cell sorter and measuring their fluorescence, without sorting them for collection. We can choose a threshold value from the sample distribution which should approximately cut-off a specific proportion of the population, for example the 20% most (or least) fluorescent cells. The population is then sorted as follows, until the desired number of cells within the threshold population have been acquired: Cells are passed individually through the sorter using controlled vibration in a narrow stream of media to break the flow into droplets which are likely to contain only one cell each. These then pass through a laser detector which measures size and fluorescence. A charge is then be placed on the droplet, either negative or positive depending on the sorting parameters, and the droplets are then passed through an electromagnetic field which sorts them into two containers based on their charge. Thus we have two populations, one which meets our parameters and one which does not. In our sorts there is selection based on size parameters and fluorescence which allowed us to select intact, singular cells with the fluorescence threshold we needed to obtain our desired percentage of the population (See section **4.2.3** and **Figure 23** for an example of a FACS run).