Evolutionary Innovations and Dynamics in Wagner's Model of Genetic Regulatory Networks

Wang, Yifei

Award date:
2016

Awarding institution:
University of Bath

Evolutionary Innovations and Dynamics in Wagner's Model of Genetic Regulatory Networks

submitted by

Yifei Wang

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Computer Science

December 2015

SUMMARY

The gene regulatory network (GRN) controls the expression of genes providing phenotypic traits in living organisms. In particular, transcriptional regulation is essential to life, as it governs all levels of gene products that enable cell survival and numerous cellular functions. However, there is still poor understanding of how shifts in gene regulation alter the underlying evolutionary dynamics and consequently generate evolutionary innovations.

By employing Wagner's GRN model, this dissertation investigates how the interplay of simple evolutionary forces (mutation and recombination) with natural selection acting on gene regulatory dynamics can generate major evolutionary innovations.

In this dissertation, firstly, I review all currently available research papers using Wagner's GRN model, which is also employed as the computational model used extensively in the remaining chapters. I then describe how Wagner's GRN model and its variants are implemented. Finally, network properties such as stability, robustness and path length in initial populations are investigated.

In the first study, I explore the characteristics of compensatory mutation in the context of genetic networks. Specifically, I find that 1) compensatory mutations are relatively insensitive to the size and connectivity of the network, 2) compensatory mutations are more likely to occur in genes at or adjacent to the site of a previous deleterious mutation and 3) compensatory mutations are more likely to be driven by mutations with a relatively large regulatory impact.

In the second study, I further investigate the evolutionary consequences of the properties of compensatory mutation discovered previously. Specifically, I find that 1) compensatory mutations can occur regardless of patterns of selection, 2) networks with compensatory mutations exhibit proportionately higher robustness when compensatory mutations interact closely with deleterious mutations or have large effects on gene regulation, and 3) regulatory complexity can arise as a consequence of the propensity for co-localised and large-effect compensatory mutations.

In the third study, I provide a mechanistic understanding of how recombination benefits sexual lineages. Specifically, I find that 1) recombination together with selection for developmental stability can drive populations towards the optimum, 2) recombination does not frequently disrupt well-adapted lineages as conventionally expected, and 3) recombination facilitates finding good genetic combinations which are robust to

disruption, although it also rapidly purges weaker configurations.

In the final study, I show that the selection pressure acting on rewiring gene regulation is critical to increasing benefits for sexual lineages whilst mitigating costs of sex and recombination. Specifically, I find that 1) strong selection strength can greatly benefit low-fitness sexual lineages, especially at the early stage, 2) recombination is initially costly, but it can rapidly evolve to compensate for costs of sex and recombination, and 3) sexual lineages with low levels of sex and recombination can outcompete strictly asexual populations under higher selection pressure and lower mutation rates.

The results presented for all of the studies are important for mechanistically understanding evolutionary innovations through altering transcriptional regulatory dynamics. These innovations include 1) facilitating alternative pathway evolution, 2) driving regulatory complexity, 3) benefiting sexual reproduction, and 4) resisting invasion against asexual lineages.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude and thanks to my supervisors, Joanna Bryson and Nicholas Priest, for providing me with a stimulating working environment, teaching me critical and creative thinking, and encouraging me in doing cross-disciplinary research. Your patient guidance and generous support have facilitated my academic growth. Thanks also to my third supervisor, James Davenport, who provided me with a great deal of technical support in HPC. James also greatly helped me with many organisational issues around hosting The 3^{rd} IEEE UK and Ireland Student Branch Congress in September 2013.

Thanks to all of the other people who contributed to my research work presented in this dissertation: Stephen Dorus (Syracuse University), who helped me to conceptualise the research on compensatory mutation and provided useful comments for Chapters 3 and 4; Marios Richards, who did some preliminary modelling work for Chapter 3; Daniel Sloan (Colorado State University), who kindly shared with me his research data for Chapter 4; Daniel Weinreich (Brown University), who provided insightful feedback for Chapter 5, and also generously shared many inspirational ideas for my work on gene regulatory networks; Yinghong Lan (Brown University), who helped design the experiments and discuss the results for Chapters 5 and 6, and also taught me about many evolutionary and ecological concepts and theories as well as helping me analyse data and run simulations.

I would especially like to thank my examiners, James Marshall (University of Sheffield) and Araxi Urrutia (University of Bath), for reading my dissertation. They also provided many insightful conversations and excellent suggestions for further improving this dissertation during the Ph.D. viva. I would like to thank Liz Dexter for her careful proofreading of the dissertation.

I would like to thank Daniel Weinreich again for hosting my research visit to his lab at Brown University from June to August 2014. The research visit greatly facilitated my transition from computer science to biology. Thanks to all members of the Microbial Evolutionary Genetics Laboratory for providing many interesting conversations and useful suggestions during my short stay at Brown University: Yinghong Lan, Scott Wylie, Eugene Raynes, Chintan Modi and Christopher Graves.

I would like to thank the Bath Evolution group for providing me with the learning environment to get to know about a broad research topics in evolution and ecology. Thanks to members of the Bath Evolution group: Laurence Hurst, Jason Wolf, Stephanie Diezmann, Daniel Henk, Xianming Wu and Kangzhen Dong.

Thanks to all members of the Artificial Models of Natural Intelligence (AmonI) research group who have contributed immensely to my personal and professional time at Bath: Daniel Taylor, Bidan Huang, Dominic Mitchell, Rob Wortham, Swen Gaudl, Paul Rauwolf and Andreas Theodorou.

Thanks to the people who gave me many useful feedback and suggestions for my research work: Matthew Brown (Google), Samuel Brown (Georgia Institute of Technology), Jonathan Dawes (University of Bath), Georg Hollander (University of Oxford), Dennis Kostka (University of Pittsburgh), John Mariani (EMBL-EBI), Joseph Marsh (University of Edinburgh), Chris Ponting (University of Oxford), Graham Ritchie (University of Edinburgh), Tim Rogers (University of Bath), Colin Semple (University of Edinburgh) and Kit Yates (University of Bath).

I would like to thank Robert Scheichl (University of Bath), Eike Mueller (University of Bath), Ken Deeley (MathWorks) and all of the course leaders from the Autumn Academy for High Performance Computing at the University of Cambridge for teaching me HPC techniques in C++, Fortran and Matlab to help speed up simulations.

I would like to thank Nihal Sinnadurai (former Chair of IEEE UK & Ireland), Steve Nightingale (former Chair of IEEE UK & Ireland), Ali Hessami (Chair of IEEE UK & Ireland), Mona Ghassemian (former Secretary of IEEE UK & Ireland) and Nick Wright (Treasurer of IEEE UK & Ireland) for being my mentors and colleagues during my term as the Student Activities Chair of IEEE UK & Ireland from 2013 to 2015. Thanks to Andreas Koltes, Sohaib Qamar, Veronika Nesheva, and all IEEE section/region officers and volunteers in the local student branches for providing much support for helping organise numerous student activities across the UK and Ireland.

Thanks to the University of Bath for providing me with the most selective and prestigious *Overseas University Research Studentship* (*ca.* £90,000), which covered both my tuition fees and maintenance expenses. Thanks to the Department of Education of Xinjiang Uygur Autonomous Region, China, for providing me with the prestigious scholarship of *Outstanding Xinjiang Students Studying Abroad* (¥50,000).

I would like to thank all my friends for their support and encouragement. Thanks to my parents and all my family members who supported me in all my pursuits.

Lastly, my special thank goes to my fiancée, Siran Wang, for her faithful love, support, encouragement, patience and understanding. Thanks to her for accompanying me throughout my Ph.D. journey!

Contents

List of Figures

List of Tables

List of Abbreviations

CM	Compensatory Mutation
DM	Deleterious Mutation
Freq.	Frequency
GRN	Gene Regulatory Network
Mut.	Mutation
RS	Relaxed Selection
Rec.	Recombination

List of Symbols[1]

N	Number of genes
$N(0,1)$	The standard normal distribution
M	Population size
W	The genotype, representing all gene-gene interactions
a	A constant determining the rate of change from complete repression to complete activation
c	Network connectivity
$devT$	A constant determining number of iterations in the developmental process
μ	Mutation rate
$s_i(t)$	The expression level of gene i $(i = 1, 2, \ldots, N)$ at time t
$\mathbf{s}(t)$	The phenotype, representing all gene expression levels, $s_1(t), s_2(t), \ldots, s_N(t)$ at time t
$\mathbf{s}_{\mathrm{EQ}}$	All gene expression levels at the equilibrium
$\mathbf{s}_{\mathrm{OPT}}$	All gene expression levels of the optimal (target) phenotype
σ	Selection strength or selection coefficient
τ	A time-constant characteristic for the developmental process under consideration
$w_{i,j}$	The gene i's regulatory impact on the gene j $(i, j = 1, 2, \ldots, N)$

[1]Here, I only list the key symbols that have been extensively used throughout the dissertation.

Chapter 1

Introduction

> I have hitherto sometimes spoken as if the variations so common and multiform in organic beings under domestication, and in a lesser degree in those in a state of nature had been due to chance. This, of course, is a wholly incorrect expression, but it serves to acknowledge plainly our ignorance of the cause of each particular variation.
>
> Charles R. Darwin

1.1 Motivation

Understanding the history of life means uncovering the mechanisms underlying the evolution of innovation on different life scales, ranging from the molecular to the cellular, tissue and organ levels (Wagner, 2011a). One of the most important forms of innovation can be attained through gene regulation, which refers to a process that controls a gene product at a particular time and place (Wagner, 2014). In particular, transcriptional regulation, which is mediated by the binding of proteins to specific DNA sequences or *cis*-regulatory elements, is essential to life, as it governs all levels of gene outcomes that enable cell survival and numerous cellular functions (Karlebach and Shamir, 2008).

However, the evolution of transcriptional regulation is extremely difficult to study experimentally. The main reasons, as summarised in Wagner (2011a), are 1) DNA regions where the regulated transcription can span hundreds of kilobase pairs upstream and downstream of a regulated gene, 2) *cis*-regulatory elements which can function regardless of their orientation and distance from a regulated gene, and 3) the DNA regions surrounding them often evolve rapidly, not only through changes of individual nucleotides but through insertions and deletions of large swathes of DNA. These reasons all present substantial challenges to characterising transcriptional regulation through experiments, due to the limitations of the currently available technologies.

In recent decades, researchers have made tremendous efforts in modelling regulatory networks using computational approaches. Kauffman (1969) introduced basic Boolean networks to study the behaviour of large, randomly constructed nets (Kauffman, 1993). Shmulevich et al. (2002) further developed the probabilistic Boolean networks to include global dynamics and cope with uncertainty. Petri nets, initially proposed by Petri (1962), are used to study large metabolic networks. Friedman et al. (2000) introduced Bayesian networks as a probabilistic framework for discovering interactions between genes based on multiple expression measurements. Differential equations are used to study network dynamics by explicitly modelling the concentration/activity changes of molecules over time (Klipp et al., 2008).

In the mid-1990s, Andreas Wagner proposed a gene regulatory network model where the developmental process was explicitly modelled in the system (Wagner, 1994, 1996). The many-to-one mapping mechanism of genotype to phenotype in Wagner's GRN model enables genes to buffer against and even exploit likely variations in the genome. This mechanism is crucial for evolutionary innovations, because genotypes which control gene-gene interactions can change profoundly without affecting phenotypes, which represent gene activities or expression concentrations (Wagner, 2011a). Wagner's GRN model motivated research on the evolution of genetic networks, and has been successfully employed to study many fundamental evolutionary and ecological questions (Siegal and Bergman, 2002; Bergman and Siegal, 2003; Masel, 2004; Azevedo et al., 2006; MacCarthy and Bergman, 2007a,b; Huerta-Sanchez and Durrett, 2007; Kimbrell and Holt, 2007; Ciliberti et al., 2007a,b; Siegal et al., 2007; Martin and Wagner, 2008; Leclerc, 2008; Borenstein and Krakauer, 2008; Sevim and Rikvold, 2008; Martin and Wagner, 2009; Palmer and Feldman, 2009; Draghi and Wagner, 2009; Fierst, 2010; Lohaus et al., 2010; Wagner, 2011b; Espinosa-Soto and Wagner, 2010; Espinosa-Soto et al., 2011a,b; Fierst, 2011; Rhoné et al., 2011; Le Cunff and Pakdaman, 2012; Pinho et al., 2012; Le Cunff and Pakdaman, 2014; Wang et al., 2014a; Shin and MacCarthy, 2015; Payne and Wagner, 2015; Wang et al., 2015; Wilder and Stanley, 2015; Pinho et al., 2015).

Mutation and recombination are two important sources of genetic variations that can ultimately facilitate evolutionary innovations. Previous studies using Wagner's GRN model have strictly required phenotypic stability (see Chapter 2) for networks, purging individuals with oscillating phenotypic states. However, previous work has not considered the possibility that these 'unviable' networks could restore stability through, for example, compensatory mutations[1]. Therefore, the existing work has largely overlooked innovations that can be driven by those compensated networks.

Many other previous studies have focused on explaining the benefits of recombination from a static viewpoint. However, these studies have not considered how the evolved properties contribute to the maintenance of sex and recombination, especially

[1]Compensatory mutations are mutations that correct a loss of fitness due to earlier mutations.

in the face of invasion by asexual lineages in a competitive regime. Therefore, the existing work has also largely overlooked innovations that can be generated by the underlying evolutionary dynamics via recombination.

This dissertation addresses the research topics overlooked in existing studies by employing Wagner's GRN model. Specifically, I focus on the following research questions:

- What are the characteristics of compensatory mutations when we relax the selection for phenotypic stability?
- How do those networks with compensatory mutations contribute to evolutionary complexity?
- Why can sexual lineages evolve greater benefits than asexual lineages?
- When can sexual lineages resist invasion by asexual lineages in the presence of substantial costs incurred by sex and recombination?

Answering these questions is important for providing a mechanistic understanding of evolutionary innovations through altering regulatory dynamics.

1.2 Dissertation structure

This dissertation mainly presents two related but different research studies. Chapters 3 and 4 are mainly focused on the characteristics of compensatory mutations and their evolutionary consequences. Chapters 5 and 6 are mainly focused on explaining the benefits of sexual reproduction, and how those benefits could recoup the costs of sex and recombination. The detailed structure of the rest of the dissertation is outlined below, along with an overview of each chapter.

Chapter 2: The Wagner gene regulatory network model

The Wagner gene regulatory network model has been successfully employed as a powerful computational tool to study the evolution of genetic networks, robustness, epistasis, sexual reproduction, plasticity, evolvability, etc. In this chapter, I first review all currently available research papers that fall into the framework of Wagner's GRN model. Then, the detailed implementation of the Wagner model, as well as its variants, is described and discussed. Finally, I investigate network characteristics such as stability, robustness and path length in initial populations. Similar to previous studies, I find that generally small networks with a sparse connectivity have a higher initial stability as well as initial robustness, and also have a shorter path length. These results are important, as they provide a mathematical and simulation foundation for the research work in the following chapters.

Chapter 3: Characteristics of compensatory mutation in gene regulatory networks

It is well-established that gene pathway evolution is constrained by natural selection because it removes maladapted mutations through which novel and beneficial gene combinations may evolve. The evolutionary constraint is expected to be especially prohibitive for genes within gene regulatory networks, as the need for the simultaneous, coordinated expression of many genes seems to make it less likely that additional rounds of mutations could restore function. By using Wagner's GRN model and treating deleterious mutations more like parasites — as reducing rather than eliminating fitness — I am able to explore their dynamics and account for recent empirical findings. In this chapter, I show that the frequency of compensatory mutation is not only relatively high but is also relatively insensitive to the size and connectivity of the network. I find that compensatory mutations are likely to occur in genes at or adjacent to the site of a previous deleterious mutation, in contrast to the more distributed locations of neutral mutations. The results also show that compensation is driven by mutations with a relatively large regulatory impact, whereas neutral mutations are more likely to have a small regulatory impact on networks. These findings show that compensatory mutations may play a more important role in evolution than previously thought, and indicate that gene pathway evolution may be far less constrained than previously considered.

Chapter 4: Compensatory mutation generates regulatory complexity through non-adaptive processes

Although there has been sustained interest in the process of adaptation, recent evidence indicates that major features of genome organisation may evolve without substantial influence from adaptive selection. However, few studies have focused on identifying specific mechanisms that generate biases in the loss of genetic variation, the major way in which non-adaptive processes contribute to evolution. It is difficult to identify non-adaptive processes underlying regulatory evolution with biological experiments because there are no natural systems in which the pattern of gene regulation is completely resolved and in which we can segregate adaptive from non-adaptive processes. It is, however, possible to employ an *in silico* gene regulatory paradigm to identify specific sources of bias in the accumulation of particular configurations of gene regulation without competitive adaptive selection. In this chapter, using Wagner's GRN model, I show that compensatory mutations can occur regardless of patterns of selection. Although they have low robustness, networks with compensatory mutations exhibit proportionately higher robustness when compensatory mutations interact closely with deleterious mutations or have large effects on gene regulation. I show that this location- and size-specific robustness systematically biases which networks are lost

by purifying selection, which, over time, increases the regulatory complexity of the entire population. These findings are important because they provide an explanation of how major features of genome organisation, development and biodiversity can emerge through non-adaptive processes.

Chapter 5: Recombination is constructive in the context of selection for phenotypic stability

Recombination is ubiquitous in multicellular plants, animals and even fungi. Many studies have shown that recombination can generate plenty of genetic innovations, but it is also believed to damage well-adapted lineages, causing debates over how organisms cope with such disruptions. Using Wagner's GRN model, in this chapter, I show that recombination may not be as destructive as expected. Provided only that there is selection for phenotypic stability, recombination can establish and maintain lineages with reliably better phenotypes compared to asexual reproduction. Contrary to expectation, this does not appear to be a simple side-effect of higher levels of variation. A simple model of the underlying dynamics demonstrates a surprisingly high robustness in these lineages against the disruption caused by recombination. Contrary to expectation, lineages subject to recombination are less likely than asexual lineages subject to simple mutation to produce offspring suffering purifying selection for phenotypic stability. These findings indicate the fundamental differences between recombination and high mutation rates, which have important implications for understanding both biological innovation and hierarchically structured models of machine learning.

Chapter 6: Selection pressure benefits low-fitness individuals and mitigates the costs of sex and recombination

The maintenance of sex has long been a mystery to evolutionary biology. Although meiotic recombination helps purge deleterious mutations and has a key role in generating evolutionary innovations, it is not clear that these benefits can recoup the costs of sex and recombination. By employing Wagner's GRN model, in this chapter, I am able to test how selection pressure affects the underlying evolutionary dynamics in sexual lineages. In the first study, I find that, compared with asexual lineages, low-fitness sexual lineages can gain a higher benefit when they are subjected to higher selection pressure, especially at the early stage. This indicates that selection pressure can facilitate fast adaptation for low-fitness individuals via recombination. In the second study, where I include both the recombination cost and the twofold cost (the competitive advantage of asexual lineages relative to sexual lineages) in the system, I show that although recombination is initially costly, it rapidly evolves (through rewiring gene regulation) to compensate for the costs of sex and recombination in even a single bout. I further explore the parameter space and find that sexual lineages with low levels of sex

and recombination can outcompete strictly asexual populations under higher selection pressure and lower mutation rates. These results have important implications for explaining the maintenance of sex and recombination in the context of genetic networks.

Chapter 7: Conclusions

This chapter summarises the main conclusions drawn from Chapters 2–6.

1.3 Contributions

The primary contributions of this dissertation are summarised as follows:

- *Characterises compensatory mutations in the context of genetic networks.* In this dissertation, compensatory mutation is defined as a mutation that can restore a network's phenotypic stability. In the context of genetic networks, I find that 1) compensatory mutations are relatively insensitive to the size and connectivity of the network, 2) compensatory mutations are more likely to occur in genes at or adjacent to the site of a previous deleterious mutation, and 3) compensatory mutations are more likely to be driven by mutations with a relatively large regulatory impact.

- *Identifies how compensatory mutations can drive regulatory complexity through non-adaptive processes.* In this dissertation, Wagner's GRN model is modified to allow periods of relaxed selection, such that 'impaired' networks with oscillating phenotypic states can be rescued by compensatory mutations in subsequent generations. I find that 1) compensatory mutations can occur regardless of patterns of selection, 2) networks with compensatory mutations exhibit proportionately higher robustness when compensatory mutations interact closely with deleterious mutations or have large effects on gene regulation, and 3) regulatory complexity can arise as a consequence of the propensity for co-localised and large-effect compensatory mutations.

- *Provides a mechanistic understanding of how recombination benefits sexual lineages.* In this dissertation, the benefit of recombination is explored under the condition that the selection for the optimum phenotype is largely absent. I find that 1) recombination together with selection for phenotypic stability can drive populations towards the optimum, 2) recombination does not frequently disrupt well-adapted lineages as conventionally expected, and 3) recombination facilitates finding good genetic combinations that are robust to disruption, although it also rapidly purges weaker configurations.

- *Explores how selection pressure recoups the costs of sex and recombination.* In this dissertation, both the recombination cost and the twofold cost have been explicitly included in a regime where sexual lineages compete against asexual lineages. I find that 1) strong selection pressure can greatly benefit low-fitness sexual lineages, especially at the early stage, 2) recombination is initially costly, but can rapidly evolve to compensate for the costs of sex and recombination, and 3) sexual lineages with low levels of sex and recombination can outcompete strictly asexual populations under higher selection pressure and lower mutation rates.

In addition to these contributions, this work also has important implications for the machine learning field, but these are not a major focus of the dissertation. Some preliminary results of solving optimisation problems using the Wagner model have been presented elsewhere in Wang et al. (2014a).

Chapter 2

The Wagner gene regulatory network model

2.1 Introduction

Gene regulatory networks control the expression of genes, thereby providing phenotypic traits in living organisms. They play a central role in cells and govern cell differentiation, metabolism, the cell cycle and signal transduction (Karlebach and Shamir, 2008). Many computational models that aim at capturing the essential structure and dynamics of networks have been developed to uncover the underlying mechanisms of transcriptional networks in nature (Guelzim et al., 2002; Wray et al., 2003; Lynch, 2007b; Tuch et al., 2008; Sorrells and Johnson, 2015; Payne and Wagner, 2015).

One of the most well-established abstract models was proposed and developed by Wagner (1994, 1996). The novel feature in Wagner's GRN model is that it introduces selection for phenotypic stability (for more details, see Section 2.3.6), performed as a separate layer of purifying selection in addition to the selection for a particular or optimal phenotype (Wang et al., 2014a, 2015). Because of this purifying selection imposed on the population, only individuals that can achieve developmental equilibrium (the ability to maintain phenotypic stability, see more details in Sections 2.3.3 and 2.3.6) are able to survive during the evolution. Wagner's central assumption is that an individual's phenotype should be able to buffer against genotypic variations. In other words, the selection for phenotypic stability provides a viable simulation for the known natural phenomenon of an individual's phenotype being expressed as relatively stable whilst its genotype undergoes evolution.

This chapter serves as a mathematical and simulation foundation for the research work presented in Chapters 3–6. Specifically, I first review all currently available research papers that fall into the framework of Wagner's GRN model, classifying them into several application areas in chronological order. Then, implementation details of

Wagner's GRN model and its variants are described and discussed. Finally, I investigate network characteristics such as stability, robustness and path length in initial populations. Similar to previous studies, I find that generally small networks with a sparse connectivity have a higher initial stability. The robustness is also observed to be higher in initially stable networks with a low network connectivity. These results are partly explained by the pattern, shown in this chapter, that small networks with a sparse connectivity generally have a shorter path length and, therefore, are not only able to quickly reach equilibrium phenotypic states but are also more likely to resist perturbations.

2.2 Applications using Wagner's GRN model

In this section, I first present a short, more general overview of gene regulatory network models. Then, all currently available research papers using Wagner's GRN model are thoroughly reviewed. Note that the reviewed papers are grouped by their main research focus. This does not necessarily indicate that they are not relevant to other research topics.

2.2.1 General gene regulatory network models

Generally, two types of network model have been developed for quantitatively or qualitatively analysing the evolutionary dynamics of genetic networks (Ciliberti et al., 2007b; Fierst and Phillips, 2015). The first type of model focuses on modelling a specific network or genetic pathway to quantitatively understand, for example, the segment polarity network in *Drosophila* (von Dassow et al., 2000), the oscillatory network in *Escherichia coli.* or the cell-cycle network in yeast (Li et al., 2004). These models typically use differential equations and require the precise measurement of concentrations or activities of gene products modelled through biochemical parameters, for example, binding affinities of transcription factors, dissociation constants of receptors and ligands or rate constants of enzyme kinetics (Ciliberti et al., 2007b).

However, the quantitative information on the parameters used in these models is largely unknown, even for some well-studied experimental systems, due to the limitations of current biochemical techniques (Wagner, 1996; Fierst and Phillips, 2015). Specifically, for many biological networks, we do not have a comprehensive understanding of each circuit in a network and its interactions. Even if such quantitative information is available, it has been difficult to precisely estimate or measure the exact strengths of gene-gene interactions. Therefore, due to the lack of quantitative information in studying genetic networks, the second type of model has been used more broadly to discover general principles that emerge from the dynamics of genetic networks (Ciliberti et al., 2007b; Fierst and Phillips, 2015). A recent review of such models

can be found in Spirov and Holloway (2013). These models typically use general and abstract representations, and therefore do not require measurements or estimates of biochemical information in nature systems. One of the most successful computational gene regulatory network models was proposed and developed by Wagner (1994, 1996).

Wagner's GRN model, initially proposed to describe the evolutionary mechanism of gene duplication (Wagner, 1994), has been employed as a popular *in silico* modelling approach to study epistasis, non-linear interactions between alleles at different loci and complex genetic interactions for a broad range of fundamental research questions in evolution, ecology and systems biology; for example, gene duplication (Wagner, 1994), genetic assimilation and robustness (Wagner, 1996; Siegal and Bergman, 2002; Masel, 2004; Huerta-Sanchez and Durrett, 2007; Kimbrell and Holt, 2007; Ciliberti et al., 2007a,b; Martin and Wagner, 2008; Leclerc, 2008; Espinosa-Soto et al., 2011b; Le Cunff and Pakdaman, 2012; Shin and MacCarthy, 2015; Payne and Wagner, 2015), recombination and sexual reproduction (Azevedo et al., 2006; MacCarthy and Bergman, 2007a; Martin and Wagner, 2009; Lohaus et al., 2010; Wagner, 2011b; Le Cunff and Pakdaman, 2014; Wang et al., 2015), phenotypic plasticity (Bergman and Siegal, 2003; Borenstein and Krakauer, 2008; Fierst, 2011; Espinosa-Soto et al., 2011a; Pinho et al., 2015), evolvability (Draghi and Wagner, 2009; Fierst, 2010; Wang et al., 2014a; Wilder and Stanley, 2015), network topology (Siegal et al., 2007), subfunctionalisation (MacCarthy and Bergman, 2007b), incompatibility (Palmer and Feldman, 2009), modularity (Espinosa-Soto and Wagner, 2010), selection strength (Rhoné et al., 2011) and phenotypic stability (Sevim and Rikvold, 2008; Pinho et al., 2012).

2.2.2 Genetic assimilation and robustness

In a classic experiment by Waddington (1953), a phenotype of crossveinless wings appeared when *Drosophila* pupae of a wild Edinburgh strain were exposed to a temperature shock after puparium formation. Waddington then selected those offspring with crossveinless wings and further observed that the crossveinless phenotype continued to appear even when the temperature shock was no longer applied. He referred to this process as genetic assimilation, whereby environmentally induced phenotypic variations become constitutively produced even if the environmental signal is absent (Waddington, 1953, 1959). Waddington further envisioned a metaphor for the biological development in which cells, represented by balls, roll downhill through a high-dimensional epigenetic landscape, and described the concept of canalisation[1] (also termed robustness) as the deepening of valleys (pathways) down the slope, making the developmental outcome less sensitive to perturbations (Waddington, 1953, 1959; Bhattacharya et al., 2011; Pujadas and Feinberg, 2012). After Waddington, a large number of studies focused on

[1]Canalisation measures the ability of a population to produce the same phenotype regardless of the variability of its environment or genotype.

uncovering the underlying mechanisms by which canalisation can be achieved. However, it is still unclear how canalisation affects the distribution of molecular or genetic variations at different levels of genetic hierarchies or regulatory genes (Gibson and Wagner, 2000; Felix and Barkoulas, 2015).

Wagner (1996) first employed his GRN model, which explicitly incorporates self-development along with the evolutionary process, to investigate canalisation in the context of genetic networks, and reported that the probability of mutations that cause changes in gene expression patterns can be substantially reduced. He referred to this phenomenon as epigenetic stability; that is, the system of epigenetic interactions may compensate or buffer some of the changes that occur as mutations on its lowest levels. Wagner also observed this increased epigenetic stability independently in experiments with variations in network architecture or other model parameters.

Siegal and Bergman (2002) developed Wagner's GRN model, and further showed that selection for phenotypic stability is sufficient for canalisation. Specifically, Siegal and Bergman designed evolutionary scenarios where they measured the phenotypic distance of evolved populations in the face of mutation perturbations under different selection pressures for the optimal phenotype. They reported that networks can evolve greater insensitivity to mutation even without the directional selection for this property; that is, the selection for the optimal phenotype is largely absent. They concluded that genetic canalisation, phenotypic insensitivity to mutation, is an emergent property of complex gene networks.

Masel (2004) introduced external noise at an individual's developmental stage, and further reported that selection for phenotypic stability is also sufficient for genetic assimilation. Specifically, the modelled noise served as an environmental perturbation similarly to the temperature shock described in Waddington (1953)'s experiment, and could consequently affect the phenotype-genotype mapping. Masel then measured the phenotypic diversity in the presence of noise to access the evolution of genetic assimilation. In addition to the phenomenon observed by Siegal and Bergman (2002), Masel concluded that the results supported the utility of Waddington's canalisation as an explanation for genetic assimilation.

Huerta-Sanchez and Durrett (2007) re-examined the previous work of Wagner (1996) and Siegal and Bergman (2002) and proposed a mathematical framework to investigate a simplified version of Wagner's GRN model in more detail. Huerta-Sanchez and Durrett showed that the qualitative observation that systems evolve to be robust is itself a robust conclusion, given that the population size is sufficiently large. They further explained that robust systems by definition of the model are insensitive to mutation and hence have a large amount of viable offspring. Therefore, the evolution of robustness is simply selection for greater reproduction success.

Ciliberti et al. (2007b) studied how robustness varies in networks with different architectures. They showed that robustness to mutations and noise are positively correlated. Here, the noise was modelled as perturbations to initial gene expression patterns, which is different from noise introduced at the developmental stage, as in Masel (2004). Moreover, Ciliberti et al. showed that highly robust networks can be reached from networks with lower robustness through gradual and neutral evolution in one large metagraph[2] of network architectures. In a similar study (Ciliberti et al., 2007a), the same authors further concluded that the robustness emerging from the connected metagraph can simulate long-term innovation in gene expression patterns.

Kimbrell and Holt (2007) studied canalisation in source-sink evolution. Here, the sink was modelled as a low-quality habitat where populations cannot persist without recurrent immigration from a source population, whereas the source was modelled as a high-quality habitat. They showed that the probability of adaptation to the novel habitat decreases when canalisation increases. However, by introducing noise to initial gene expression patterns, as in Ciliberti et al. (2007b), Kimbrell and Holt found that noise can facilitate adaptation to novel habitats.

Martin and Wagner (2008) investigated how multifunctionality affects a network's robustness to mutations and noise. The multifunctionality was modelled as different pairs of initial and equilibrium gene expression patterns. They showed that the number of network architectures decreases dramatically as a result of the increased additional functions that they are required to carry out. Given that the relationship between the robustness of one function and that of other functions to mutations and noise is largely absent, Martin and Wagner concluded that robustness trade-offs of multiple stable phenotypes generally do not arise in such systems.

Leclerc (2008) argued that the common measurement for robustness used previously in Wagner (1996) and Siegal and Bergman (2002) may not be appropriate, as the measurement inadvertently discounts the costs of network complexity. By taking the costs of complexity into account, Leclerc showed that a higher robustness could be observed in sparsely connected networks (low network connectivity) with parsimonious architectures[3] Moreover, the author showed that selection will favour sparse networks if the network architecture is free to evolve.

Espinosa-Soto et al. (2011b) introduced non-genetic perturbations and studied the relationship between a phenotype's mutational robustness and a population's potential to generate novel phenotypic variations. Here, non-genetic perturbations referred to both perturbations from environmental factors such as temperature, diet or biotic iterations, as modelled in Masel (2004), and perturbations from an organism's inter-

[2]Metagraph refers to a graph of graphs, which is a special kind of mutational graph where two genotypes (nodes) are connected in a mutational graph if one genotype can be obtained from the other through a single mutation.

[3]Parsimonious networks refers to sparsely connected and not unnecessarily complex networks.

nal factors such as activity changes in initial gene expression patterns, as modelled in Ciliberti et al. (2007b) and Kimbrell and Holt (2007). Espinosa-Soto et al. found that phenotypic robustness facilitates variability in response to non-genetic perturbations, but not in response to mutations.

Le Cunff and Pakdaman (2012) reviewed previous work using Wagner's GRN model, and derived new observations of emergent properties with respect to robustness in the system. They showed that selection for a specific (target) phenotype also benefits individuals by increasing the probability of stabilising alternative phenotypes revealed under stress. Le Cunff and Pakdaman further showed that a generalised canalisation in the system can drive a population towards robustness in the presence of perturbations, for example, gene deletion, loss of interactions and mutations in regulation activities.

Shin and MacCarthy (2015) investigated how robustness and sensitivity become distributed in a host-parasite model of antagonistic co-evolution. Here, parasites were modelled on species such as cuckoos where mimicry of the host phenotype confers a higher fitness to the parasite but a lower fitness to the host. They found that sensitivity sites[4] are broadly distributed throughout the network and continually relocate. Shin and MacCarthy referred to this phenomenon as 'Whack-A-Mole', inspired by a popular fun park game.

2.2.3 Recombination and sexual reproduction

Recombination is ubiquitous in multicellular plants, animals and even fungi. However, it is still unclear how evolutionary dynamics such as sexual reproduction contribute to the stability of inheritance. All sexual systems exhibit recombination — the reshuffling of parental genetic information which generates novel, heritable gene combinations (Eshel and Feldman, 1970; Feldman et al., 1996; Otto and Feldman, 1997; West et al., 1999). However, sexual reproduction is also considered to be very costly, since it may damage well-adapted lineages and produces fewer offspring. Consequently, why is sexual reproduction maintained? For decades, researchers have made tremendous efforts and proposed numerous possible theories to explain and uncover the mystery of sex and recombination (Eshel and Feldman, 1970; Hurst and Peck, 1996; West et al., 1999; Otto and Lenormand, 2002; Meirmans and Strand, 2010; Wagner, 2011b). Two classic, although still controversial, benefits of sex and recombination are 1) purging deleterious mutations more efficiently, and 2) creating novel gene combinations (Kondrashov, 1993; Otto and Feldman, 1997; Otto and Gerstein, 2006; Kouyos et al., 2007; Barton, 2009; Martin and Wagner, 2009). However, although many observed phenomena, such as improving robustness and facilitating evolutionary adaptation, can be attributed to sexual reproduction, the underlying evolutionary mechanism is still poorly understood

[4]Sensitivity sites refer to the location where mutations are more likely to undermine the network stability.

(Wagner, 2011b).

Azevedo et al. (2006) first employed Wagner's GRN model to study the maintenance of sexual reproduction in the context of gene networks. They showed that sexual populations can evolve a higher robustness than asexual populations. Moreover, they further observed that synergistic (negative) epistasis[5] can evolve from sexual populations as a by-product of selection for phenotypic stability imposed in the system, whereas antagonistic (positive) epistasis[6] evolves from asexual populations, which supports the deterministic mutation hypothesis[7] for explaining the maintenance of sexual reproduction. Azevedo et al. concluded that sexual reproduction evolves genetic properties that favour its own maintenance.

MacCarthy and Bergman (2007a) pointed out that the study conducted by Azevedo et al. (2006) may not have explicitly examined whether sexual populations can outcompete asexual populations under the condition of synergistic epistasis. Specifically, they studied conditions whereby asexual reproduction could nonetheless be favoured by allowing the spontaneous emergence of epistasis in its evolution and introducing a modifier locus that explicitly alters the recombination rate. They found that the fixation time of the asexual mode only has a significant correlation with the level of antagonistic epistasis, but not that of synergistic epistasis. MacCarthy and Bergman highlighted that the deterministic mutation hypothesis may not be a plausible explanation for the maintenance of sexual reproduction.

Martin and Wagner (2009) focused on effects of recombination in the context of genetic networks. They showed that recombination has much weaker effects than point mutations. Moreover, they demonstrated that recombination reduces genetic load[8] and also dramatically increases genetic diversity. Finally, they observed that the effect of recombination can create particular regulatory complexes that are able to mitigate recombination effects that are deleterious to regulatory circuits. Martin and Wagner concluded that the effects of recombination may lead to many benefits, for example, increased genetic diversity and reduced genetic load, which are able to compensate for the disadvantages caused by sexual reproduction.

Lohaus et al. (2010) complemented the results presented in Azevedo et al. (2006) and MacCarthy and Bergman (2007a) by studying the long-term benefits of sexual reproduction. Similar to the previous studies by Azevedo et al. and MacCarthy and Bergman, Lohaus et al. observed that sexual populations can evolve a higher robust-

[5]Synergistic epistasis refers to a situation when the effect on the fitness of two mutations is more radical than would be expected from the effects of the two single mutations.

[6]Antagonistic epistasis refers to a situation when the effect on the fitness of two mutations is smaller than would be expected from the effects of the two single mutations.

[7]This deterministic mutation hypothesis, proposed by Kondrashov (1988), assumes that the majority of deleterious mutations are only slightly deleterious, and affect the individual such that the introduction of each additional mutation has an increasingly large effect on the fitness of the organism.

[8]Genetic load is the reduction in the mean fitness of a population relative to a population composed entirely of individuals having optimal genotypes (Whitlock and Davis, 2001).

ness and lower genetic load than asexual populations at equilibrium. However, contrary to Azevedo et al., they found no evidence that negative epistasis can contribute to long- and short-term benefits emerging from sexual populations. Moreover, they found that the lower deleterious mutation rate evolving from sexual populations is not able to account sufficiently for the ability of sexual populations to resist invasion by asexual populations in the long term. Lohaus et al. argued that it is the continuously increasing recombinational robustness that minimises the cost of sexual reproduction, and ultimately evolves resistance to asexual invasion in the long term.

Wagner (2011b) broadly reviewed possible reasons for the low cost of recombination. He showed that 1) recombination can cause greater genotypic changes than mutation, 2) recombination facilitates creating new phenotypes, 3) recombination is able to preserve phenotypes in the context of genetic networks, 4) recombination can preserve protein structure and function, and 5) recombinational robustness can be substantially increased during evolution. Wagner therefore concluded that recombination can create new phenotypes whilst disrupting well-adapted phenotypes much less than mutation.

Le Cunff and Pakdaman (2014) studied the relationship between individual-level evolutionary dynamics and population-level survival probability in the face of genetic and demographic stochasticity. Here, genetic stochasticity refers to fluctuations in genetic composition (variability), whilst demographic stochasticity refers to fluctuations in population size. Different from previous studies which employed the Wagner GRN model with a fixed evolution space, the population size is not fixed in each generation and extinction could happen due to genetic and demographic stochasticity modelled in the system. Le Cunff and Pakdaman found that recombination rate, initial population size and mutation rates can all influence population survival probability.

2.2.4 Plasticity and evolvability

Evolvability is the capacity of a population to produce heritable phenotypic variation to rapidly adjust to certain types of environmental challenge or opportunity (Wagner and Altenberg, 1996; Wagner, 2007; Pigliucci, 2008; Masel and Trotter, 2010). This capacity, documented in nature, reflects phenotypic plasticity enabled by the capacity of evolution to capture and represent regularities not only in extant environments but also in the ways in which the environments tend to change (Callahan et al., 1997; Pigliucci et al., 2006; Wang et al., 2014a). The simplest form of evolvability is simply variation — the rate of evolution is determined by the number of variations in a population (Fisher, 1930; Price, 1972). More sophisticated evolvability can be achieved via hierarchical complex organisations, for example, genetic networks (Aldana et al., 2007; Landry et al., 2007; Crombach and Hogeweg, 2008; Greenbury et al., 2010; Torres-Sosa et al., 2012; Clune et al., 2013). Many previous studies have focused on reconciling

the antagonistic relationship between robustness[9] and evolvability by showing that living systems can sustain phenotypic stability whilst producing genetic variations that lead to evolutionary innovations (Wagner, 2008; Whitacre and Bender, 2010; Masel and Trotter, 2010; Garfield et al., 2013). However, the concept of evolvability is still controversial, and how genetic networks evolve and become evolvable remains an open question (Crombach and Hogeweg, 2008; Masel and Trotter, 2010).

Bergman and Siegal (2003) introduced gene 'knock-out' operation to Wagner's GRN model and assessed phenotypic diversity before and after evolution. They showed that when a random gene is deleted by zeroing its corresponding row and column of the regulatory matrix in Wagner's GRN model, environmental and genetic canalisation can both break down, but consequently the 'knock-out' operation increases the rate of adaptation to new environments. Moreover, they further conducted knock-out experiments on yeasts and found that they exhibit variations in phenotype which match their model predictions well. Bergman and Siegal highlighted their results that complex genetic networks enable the evolutionary capacity to buffer genotypic variations under normal conditions, whilst promoting the accumulation of hidden polymorphism that can facilitate new adaptations under stress.

Borenstein and Krakauer (2008) looked at micro- and macro-evolutionary patterns by evolving genotype-phenotype maps in genetic networks. They showed that many evolutionary patterns observed and identified from empirical studies can be attributed to epistatic interactions between genes in regulatory networks. Borenstein and Krakauer highlighted that their findings support the view that development is an essential component in the production of endless forms, and it is also critical for constraining biotic diversity and evolutionary trajectories.

Draghi and Wagner (2009) studied whether natural selection facilitates the evolution of evolvability, particularly focusing on sexual populations. By introducing fluctuating environments (periodically changing target phenotypes), they demonstrated that natural selection facilitates the capacity of genetic networks to quickly adapt to new environments. This pattern was observed regardless of asexual or sexual reproduction modes, which suggests recombination does not suppress the evolution of evolvability. Draghi and Wagner highlighted that the evolution of evolvability can be achieved by evolving a complex genotype-phenotype map.

Fierst (2010) investigated conditions under which a network may produce a more evolvable phenotype. Specifically, she modified Wagner's GRN model by introducing a sexually dimorphic trait which has an underlying network architecture that can affect evolvability. She showed that sexually dimorphic characters not only increase mutational robustness but also substantially facilitate evolvability. When she looked more

[9]Here, robustness refers to the capacity to withstand mutations and maintain phenotypic stability, function or structure.

closely at the results, Fierst further found that linkage disequilibrium within or between sex accounted for different levels of evolvability between sexually dimorphic and monomorphic populations.

Fierst (2011) studied the effect of a history of phenotypic plasticity on adaptability to new environments. She found that populations with a history of phenotypic plasticity are able to adapt to new environments more rapidly than populations without a history of phenotypic plasticity, but the magnitude of the increased adaptation rate is dependent on the strength of selection in the original environments — weak selection generally facilitates phenotypic plasticity and substantially increases the adaptation rate. Fierst suggested that the results predict that the relative invasive capacity of different traits could be assessed through phenotypic variance in the original environment.

Espinosa-Soto et al. (2011a) introduced non-genetic perturbations (changes in initial gene expression patterns), and explored whether conditions under which phenotypic plasticity facilitates adaptation can be fulfilled in the context of genetic networks. They showed that non-genetic perturbations such as gene expression noise, environmental changes or epigenetic modifications can substantially stimulate phenotypic plasticity and ultimately facilitate adaptation to new environments. Espinosa-Soto et al. concluded that phenotypic plasticity has an essential role in adaptive evolution.

Pinho et al. (2015) investigated how different levels of noise (changes in initial gene expression patterns as well as perturbations at the developmental stage) can affect the accessibility of phenotypic space that facilitates phenotypic diversity. They found that increased levels of noise typically decrease accessibility to phenotypic space if the gene expression is binary, but increase accessibility if there are more gene expression states. Pinho et al. concluded that under specific conditions, noise enables individuals to explore more phenotypic space.

Wilder and Stanley (2015) compared evolvability at the individual level[10] with evolvability at the population level[11] , focusing on the potential to generate phenotypic variations. Specifically, by introducing divergent selection — selection for phenotypic variations — they showed that divergent selection is able to produce evolvable populations and encourage phenotypic diversity, whereas evolvable individuals are more likely to be formed by adaptive selection to fluctuating environments. Wilder and Stanley hypothesised that non-adaptive mechanisms may be more important for shaping the emergence of evolvability.

[10]Individual-level evolvability refers to the ability of a single genotype to generate a total number of unique heritable phenotypes via mutations.

[11]Population-level evolvability refers to the ability of all genotypes in the population to generate a total number of unique heritable phenotypes via mutations.

2.2.5 Other applications

The Wagner GRN model has also been employed to study research topics which are not focused on in this dissertation, including the following research questions:

Wagner (1994) formally proposed a simple mathematical model to capture the key developmental process underlying transcriptional regulation and employed the proposed model to study the mechanism of gene duplication and its effect on phenotypic stability. He found that about 40%, at most, of genes in a network are duplicated, depending on the fraction of genes that are duplicated in a single duplication event. Wagner concluded that the evolution of gene networks should occur through gene duplications, and the most favourable two forms of genomic organisation are tight linkage[12] or strong dispersal[13]. Siegal et al. (2007) first employed Wagner's GRN model to thoroughly study the relationship between network topology and its functional evolutionary properties. They found that the degree of distribution (scale-free, power law distribution) of the node in networks does not have a major effect on functional properties associated with nodes. Moreover, there is weak or almost no correlation between network connectivity and genetic variations.

MacCarthy and Bergman (2007b) employed Wagner's GRN model to study the subfunctionalisation indicated by the theory of duplication-degeneration-complementation. They showed that, in contrast to previous theory predictions, subfunctionalisation and neofunctionalisation can coexist in biological networks following gene duplication. MacCarthy and Bergman hypothesised that this pattern is facilitated by evolutionary plasticity in combination with the phenotypic neutrality which prevails in biological systems.

Sevim and Rikvold (2008) studied the effect of the evolution of genetic robustness on the dynamical character of gene regulatory networks. Here, dynamical character refers to the phenotypic stability of genetic networks against perturbations such as mutations or noise. They showed that selection for phenotypic stability only weakly affects network dynamical properties, and the networks that are most robust to mutations and noise are highly chaotic. Sevim and Rikvold argued that the damage propagation analysis[14] does not provide much useful information about robustness to mutations or noise in the context of genetic networks.

[12]Tight linkage refers to the fact that genes whose loci are nearer to each other are less likely to be separated onto different chromatids during chromosomal crossover.

[13]Strong dispersal refers to the massive rewiring of regulatory circuits via recombination.

[14]The damage propagation analysis refers to the measurement that is used to determine the existence of a phase transition in RBNs and RTNs (Aldana et al., 2003).

Palmer and Feldman (2009) investigated the Bateson-Dobzhansky-Muller incompatibilities[15] and extended Orr's model[16] to account for the complex dynamics of incompatibility in the context of genetic networks. They showed that depending on certain model parameters, under a constant selection environment, three patterns of system dynamics can be observed: hybrid incompatibility between two allopatric populations 1) may not increase at all, 2) may increase to large values, and 3) may lead to a pair of populations 'drifting' in and out of compatibility.

Espinosa-Soto and Wagner (2010) investigated how modularity evolves in the context of genetic networks when the developmental process is explicitly modelled in the system. They showed that modularity is able to arise in genetic networks as a by-product of specialisation in gene activity. They also demonstrated that new gene activity patterns that share existing patterns of gene activity are more likely favoured by the evolution of modularity.

Rhoné et al. (2011) studied the impact of selection on genes at the phenotypic level in the context of regulatory networks. They showed that there is a positive relationship between the selection strength on the phenotype and the level of regulation between the loci. Moreover, they found that genes that strongly regulate other genes as well as those that are less regulated by other genes respond more profoundly to selection within the network.

Pinho et al. (2012) investigated how varying features and parameters of Wagner's GRN model affect network transition from oscillatory dynamics to phenotypic stability. They showed that the cyclical behaviour is mainly due to complex epistatic interactions between genes, but not due to connection strengths or patterns. Moreover, they showed that stability distribution is highly robust to various model parameters, and found that sparse networks are more likely to be stable.

2.3 Implementation details

In Chapters 3–6, I employ a gene regulatory network model similar to that originally proposed by Wagner (1994, 1996) and developed by Siegal and Bergman (2002). The model typically assumes that different or partially overlapping sets of transcription factors are expressed in different cells or different regions at any given stage of development of an organism (Wagner, 1994).

[15]The model aims to explain how incompatibilities between closely related species develop without either of them going through an adaptive valley (Orr, 1996).

[16]The model suggests that the fitness load on hybrids should initially accelerate, and continue to increase as the number of potentially incompatible substitutions increases (Orr, 1995, 1996).

2.3.1 Genotype

A gene regulatory network (GRN) is a collection of regulators that interact with each other, which together control the gene expression levels of mRNA and proteins (Karlebach and Shamir, 2008). In Wagner's GRN model, a genotype is represented as a network which contains interactions among transcriptional genes. This interaction network encapsulates epigenetic features such as protein-DNA-binding affinities and transcriptional activation or repression strengths (Wagner, 1994; Siegal and Bergman, 2002).

Formally, for each individual network in a finite population M, N $cis-$regulatory transcription factors are encoded by $N \times N$ matrix W (see an example network with five genes in Figure 2-1). Each element $w_{i,j}$ $(i, j = 1, 2, \ldots, N)$ represents the regulatory effect on the expression of gene i of the product of gene j.

Note that the matrix W is appropriate to be considered as a 'genotype' in the sense that it can be mapped to specific nucleotide sequences in the enhancer regions of the network genes (Siegal and Bergman, 2002). The network connectivity parameter c determines the proportion of non-zero elements in the network W. A zero entry means there is no interaction between two genes. Through gene interactions, the regulatory effect acts on each gene expression pattern.

2.3.2 Phenotype

In Wagner's GRN model, a phenotype for a given network W is denoted by a state vector $\mathbf{s}(t) = (s_1(t), s_2(t), \ldots, s_i(t), \ldots, s_N(t))$, where $s_i(t)$ represents the expression level of gene (or concentration of proteins) i at time t.

Each value of expression state $s_i(t)$ is within the interval $[-1, +1]$ that expresses complete repression (-1) and complete activation $(+1)$. Note that for reasons of computational convenience, the expression level or the admissible concentration range for each $s_i(t)$ can be normalised and restricted to the interval $[0, 1]$, as in Draghi and Wagner (2009). Note that the model typically assumes that mRNA transcripts and their corresponding protein products are directly proportional in concentration. In other words, there is no post-transcriptional regulation, and therefore, $\mathbf{s}(t)$ can be considered as either transcription or protein concentration (Wagner, 1996; Siegal and Bergman, 2002).

The initial phenotypic state $\mathbf{s}(0)$ is usually assigned random values from $[-1, +1]$ (or $[0, 1]$), and is fixed throughout an individual's lifetime. This is because the model typically assumes that the initial state is a response to an extracellular signal, such as a growth factor or a specific composition of nutrients in the medium (Wagner, 1994). Therefore, it is assumed that the initial state is determined by the products of one or more 'upstream' genes that are not themselves part of the network, and is not regulated by any factors in the network.

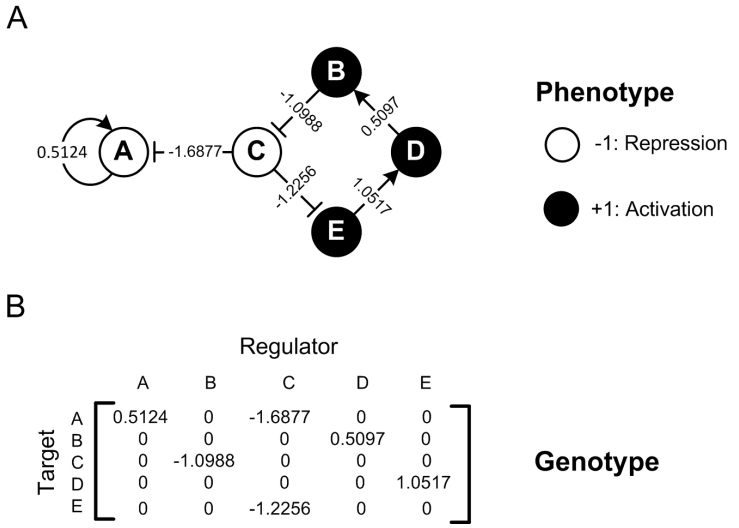


Figure 2-1: An example gene regulatory network. *(A) Network representation of regulatory interactions among five genes. Open and filled circles represent genes that are completely in activation (+1) or repression (−1). The initial gene expression pattern is* $\mathbf{s}(0) = (-1, +1, -1, +1, +1)$*. This example network is stable as it can reach an equilibrium pattern, which is* $\mathbf{s}_{\mathrm{EQ}} = (+1, +1, -1, +1, +1)$ *by iterating Equation (2.1) using the sigmoidal mapping function with* $a = 100$*. (B) The adjacency matrix (*W*) represents the network in (A). Each element in row* i *and column* j*, i.e.,* w_{ij} $(i, j = 1, 2, \ldots, 5)$*, represents the regulatory effect on the expression of gene* i *of the product of gene* j*. Note that a zero element means that there are no interactions between the two genes.*

2.3.3 Developmental process

In Wagner's GRN model, it is typically assumed that the expression of transcription factor genes is only in one developmental stage and only in one set of cells (nuclei), for example, a set of nuclei in a part of a *Drosophila* blastoderm expressing a specific subset of gap genes and pari-rule genes (Wagner, 1994). The basic idea of the developmental process is that an individual's phenotypic state changes over time due to cross-regulation and auto-regulation of the expression of member genes by their gene products (Wagner, 1994, see Figure 2-2). Formally, for a given gene regulatory network W, the dynamics of $\mathbf{s}$ for each gene i is modelled by a set of coupled difference equations:

$$s_i(t+1) = f\left(\sum_{j=1}^{N} w_{i,j} s_j(t) + \epsilon_i\right), \tag{2.1}$$

where $f(\cdot)$ is a sigmoidal function, and ϵ_i is a constant which reflects either a basal transcription rate of gene i or influences of upstream gene(s) on gene i. In all simulations, I set $\epsilon_i = 0$ and followed Siegal and Bergman (2002) and Azevedo et al. (2006) to define $f(x) = 2/(1 + e^{-ax}) - 1$, where a is the activation constant determining the rate of change from complete repression to complete activation. From Figure 2-3, we

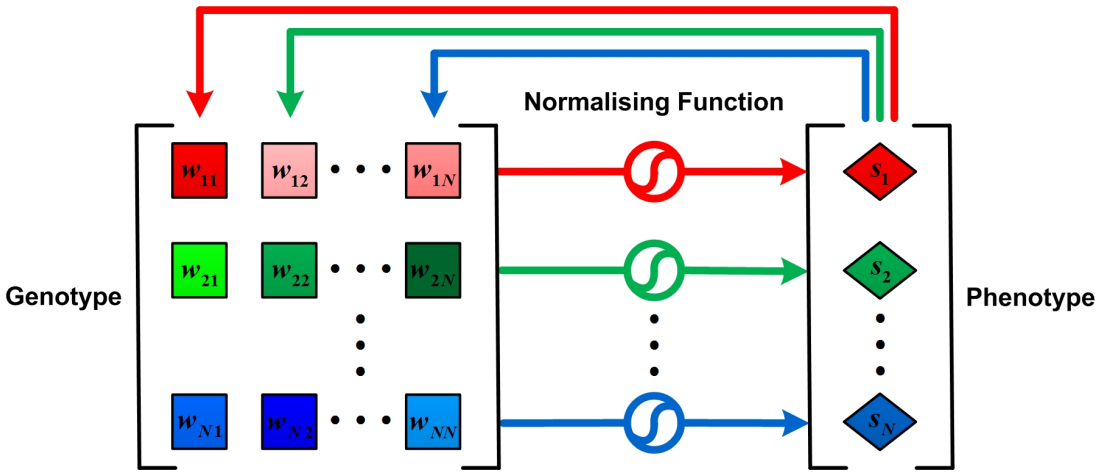


Figure 2-2: The developmental process in Wagner's GRN model. *Each gene phenotypic state at time $t+1$, $s_i(t+1)$ $(i = 1, 2, \ldots, N)$ (diamond boxes on the right), is regulated by the products of the other genes' phenotypic state at time t $s_j(t)$ $(j = 1, 2, \ldots, N)$ via upstream enhancer factors (square boxes on the left) whose strength and direction of regulation are depicted as different colour saturation levels. The result of additive regulation is then normalised by a mapping function, such as a sigmoidal or a step function. The figure is a modified version of Siegal and Bergman (2002).*

can see that when a is large, for example, $a = 100$, $f(x)$ is similar to a step function where $f(x) = -1$ for $x < 0$, $f(x) = +1$ for $x > 0$ and $f(0) = 0$. Therefore, it is quicker to produce extreme values (-1 or $+1$). The lower values of a, for example, $a = 1$, allow intermediate expression states (see Figure 2-3), but it is difficult to produce extreme phenotypic states. A detailed biological interpretation of parameter a can be found in Palmer and Feldman (2009), where the authors summarised that in terms of a metaphorical 'fitness landscape', larger values of a correspond to broad-based, sloping hills that are peaked rather than flat on top, whereas lower values correspond to narrow elevated areas with steep sides and a flat top.

2.3.4 Mutation

Generally, two kinds of mutation are usually modelled in the system. The first kind of mutation refers to changes in a given regulatory genotype, W. Specifically, such mutations can

1) cause changes in the existing interactions (non-zero entries in W) by replacing their original interaction strengths with new values drawn from the standard normal distribution $N(0, 1)$ (see Figure 2-4), and

2) form new interactions by setting new values drawn from $N(0, 1)$ to zero entries in W, or delete the existing interactions by setting their values to be 0.

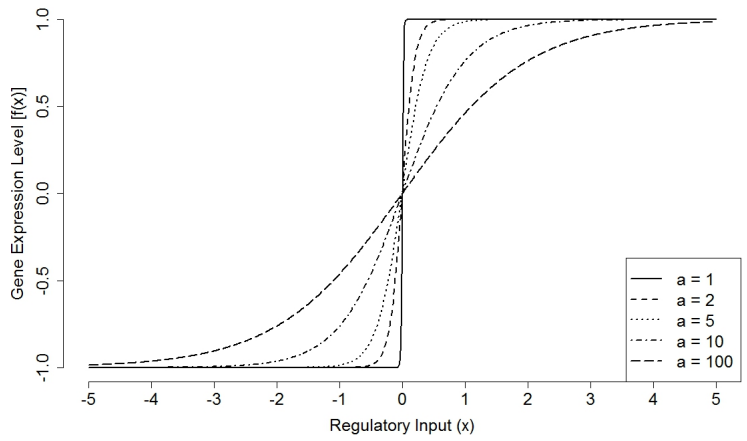


Figure 2-3: The sensitivity of parameter a to changing regulatory responses. *At each time step during the developmental stage, the expression level of a gene is determined by a filtering function, $f(x) = 2/(1 + e^{-ax}) - 1$, which normalises the sum of the regulatory effects from other genes. The activation constant a determines the rate of the transition between extreme expression states, -1 and $+1$.*

Here, I define the mutation rate in 1) as μ, and define the topological mutation rate in 2) as μ_{top} (typically $\mu \gg \mu_{\text{top}}$). In all simulations, I did not allow any topological mutation, i.e., $\mu_{\text{top}} = 0$, and, unless otherwise specified, the probability of an individual network acquiring k mutations in its non-zero entries was drawn from the Poisson distribution $P(x = k) = \mu^k e^{-\mu}/k!$ $(k = 0, 1, \ldots, \lfloor c \times N^2 \rfloor)$. Note that the model does not consider mutations in sequences that code for gene products — mutations that simultaneously affect the interaction for a given gene product with all its target enhancer or promoter regions (Siegal and Bergman, 2002).

The second kind of mutation refers to changes in the initial gene expression pattern, $\mathbf{s}(0)$. Such mutations have a non-genetic origin that could result, for example, from intracellular noise, environmental fluctuations or disturbances in the activity of genes upstream of the circuit (Espinosa-Soto et al., 2011a). However, for reasons of computational convenience, I do not consider any non-genetic mutation in my simulations.

2.3.5 Recombination

In the genotype W, because all entries in the i^{th} $(i = 1, 2, \ldots, N)$ row represent the promoter or enhancer regions of gene i, we can assume that the individual transcription factor binding sites on those regions are genetically closely linked to one another. Consequently, the recombination will occur only very rarely between them (Martin and Wagner, 2008). In contrast, different genes in a regulatory circuit are often assumed to be unlinked to one another as they can occur on different chromosomes (MacCarthy and Bergman, 2007a). In Wagner's GRN model, the recombination is modelled as a

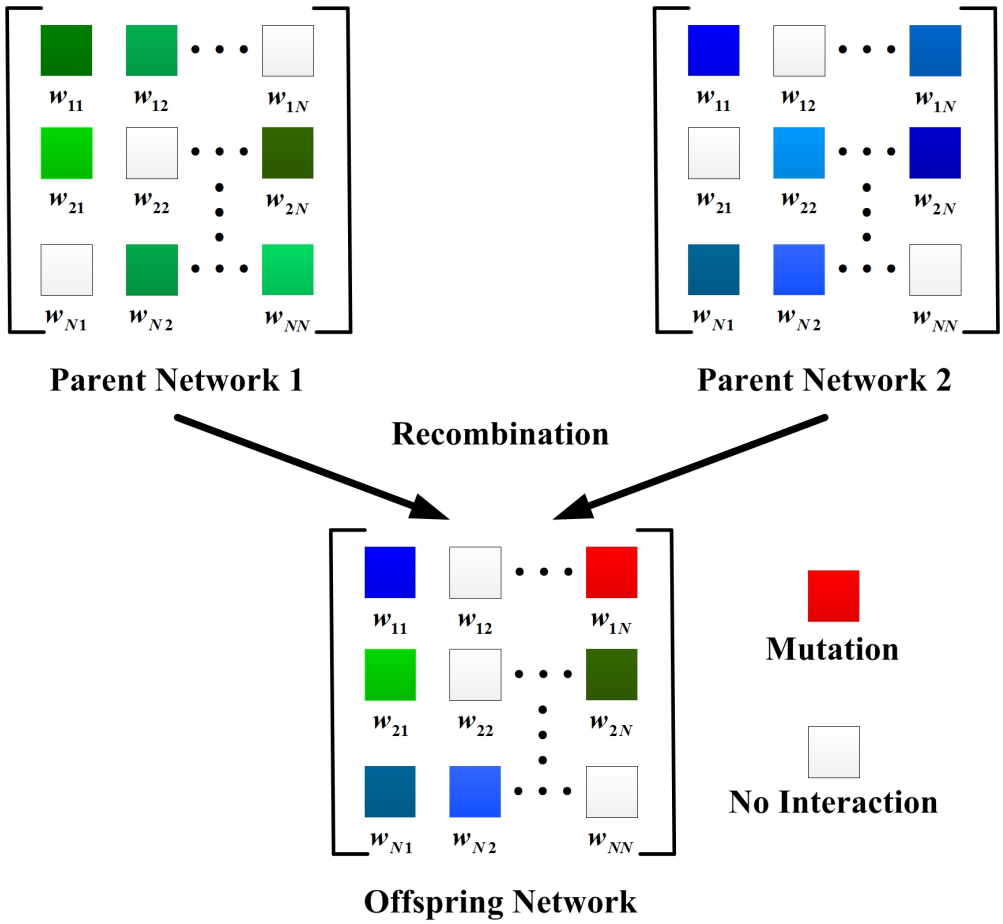


Figure 2-4: The operators of mutation and recombination in Wagner's GRN model. *Mutation (red box) as defined in this dissertation only occurs in non-zero entries in the genotype. Recombination occurs by choosing two parental networks (blue and green genotypes) at random to form a transient diploid, which then segregates rows of the matrix to form a single, haploid offspring network. Note that different colour saturation levels represent different strengths and directions of regulation.*

free recombination between circuit genes (see Figure 2-4), neglecting recombination within genes (promoters or enhancers). To be more specific, recombination occurs by randomly selecting two parental networks from the population pool to form a transient diploid. Then, for each pair of rows i in the parental networks, one of the two rows is chosen with an equal probability to form a single, haploid progeny (Wagner, 1994, 1996).

2.3.6 Selection for phenotypic stability

In all the simulations here, network phenotypic stability or developmental stability is defined as the progression from an arbitrary initial expression state, $\mathbf{s}(0)$, to an equilibrium expression state (reaching a fixed phenotypic pattern), $\mathbf{s}_{\mathrm{EQ}}(\infty)$, by iterating

Equation (2.1) a fixed number of times, $devT$. If a given network W can achieve stability over this developmental time period, it is termed *stable*; otherwise, it is labelled *unstable*. Note that this selection for phenotypic stability is also referred to as purifying selection, in which unstable networks will be eliminated. The equilibrium expression state can be reached when the following equation is met:

$$\frac{1}{\tau}\sum_{\theta=devT-\tau}^{devT} D\left(\mathbf{s}(\theta),\overline{\mathbf{s}}\right) \leqslant \xi, \tag{2.2}$$

where ξ is a small positive integer and set to be 10^{-4} in all the simulations presented in this dissertation, and $D(\mathbf{s},\overline{\mathbf{s}}) = \sum_{i=1}^{N}\left(s_i - {s'}_i\right)^2 \Big/ 4N$ measures the difference between gene expression patterns $\mathbf{s}$ and $\overline{\mathbf{s}}$ which is the average of the gene expression level over the time interval $[devT-\tau, devT-\tau+1, \ldots, devT]$, where τ is a time-constant characteristic for the developmental process under consideration, and depends on biochemical parameters, such as the rate of transcription or the time necessary to export mRNA into the cytoplasm for translation (Wagner, 1994).

2.3.7 Selection for target phenotype

In Wagner's GRN model, target selection refers to selection for a particular or optimal phenotype. For networks that can achieve phenotypic stability (reaching an equilibrium state, $\mathbf{s}_{\mathrm{EQ}}$), the phenotypic distance between the equilibrium state and the optimal state $D(\mathbf{s}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}})$, as defined in Equation (2.2), measures the degree of the Hamming distance by which the individual's equilibrium state ($\mathbf{s}_{\mathrm{EQ}}$) deviates from the optimal state ($\mathbf{s}_{\mathrm{OPT}}$). Note that this measurement normalises the distance to the interval $(0, 1)$. Using the distance D, the fitness of an individual is can be defined via a Gaussian function or a power function.

Specifically, two measurements are typically used in the model. The first exponential fitness evaluation function (see Figure A-1) is defined as in Wagner (1996) and Siegal and Bergman (2002):

$$F(\mathbf{s}_{\mathrm{EQ}}) = \exp\left(-\frac{D(\mathbf{s}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}})}{\sigma}\right), \tag{2.3}$$

where σ is the selection pressure that we impose on the population during evolution, and $\mathbf{s}_{\mathrm{OPT}}$ is usually set to be $\mathbf{s}(0)$. Unless otherwise specified, a zero fitness is assigned to individuals that cannot reach developmental equilibrium.

The second multiplicative fitness evaluation function (see Figure A-2) is defined as in Draghi and Wagner (2009):

$$F(\mathbf{s}_{\mathrm{EQ}}) = \frac{1}{(1+\sigma)^{D(\mathbf{s}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}})}} \tag{2.4}$$

where σ, $\mathbf{s}_{\mathrm{OPT}}$ and $\mathbf{s}_{\mathrm{EQ}}$ are defined similarly as in Equation (2.3). Note that for some variants of Wagner's GRN model, the w_{ij} is set to be a binary value, $w_{ij} \in \{0, 1\}$ (Draghi and Wagner, 2009; Fierst, 2010; Wilder and Stanley, 2015). Then, $D(\mathbf{s}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}})$ can be simply calculated as the number of gene equilibrium states that differ from the optimum.

Note that the fitness of both two measurements falls into the interval $(0, 1)$. During the selection process, a random value in $(0, 1)$ is first generated, and if an individual's fitness is greater than the random value, then it will be selected into the population pool for evolution in the next generation, otherwise the individual will be discarded. This selection procedure is known as roulette wheel selection, as widely used in genetic algorithms (Bäck, 1996).

2.3.8 The evolution process

The reproduction-mutation-selection life cycle is employed for *in silico* evolution (see Figure 2-5). In typical asexual evolution, an individual is chosen at random to reproduce asexually by cloning itself, and then subjected to mutation. Similarly, in typical sexual evolution, two individuals are chosen at random to reproduce sexually by recombining two genotypes, and then the offspring is subjected to mutation. The resulting offspring network is next exposed to selection for phenotypic stability (see Section 2.3.6). Unless otherwise specified in certain evolutionary scenarios, if the offspring network cannot reach an equilibrium state, then it will be wiped out from the population immediately. A stable offspring network is then exposed to selection for target phenotype (see Section 2.3.7), and can be selected into a new population pool for the next generation based on its fitness as calculated using Equations (2.3) or (2.4). In each generation, this process is repeated until M number of networks are produced.

2.4 Convergence analysis

In Wagner's GRN model, the evolution process has three operators as described in Sections 2.3.4–2.3.6: mutation, recombination and selection. Therefore, the evolution process of finding a target phenotype can be regarded as an optimisation process where the goal is to minimise $D(\mathbf{s}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}})$ such that all individuals' phenotypic state is close to the optimal phenotypic state.

Suppose that the initial population has M individual networks and the search space is in N dimensions. The phenotypes of individual networks at the g^{th} generation can be represented as $\mathbf{S}(g) = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_j, \ldots, \mathbf{s}_M]$, where $\mathbf{s}_j = (s_1, s_2, \ldots, s_N)$ is an individual's phenotype at equilibrium in an N-dimensional solution space. Let $\mathbb{S} = \mathbb{R}^N$ be the solution space, and $\mathbb{S}^M$ the population space. Without loss of generality, suppose that the optimisation goal of the evolution process described in Wagner's GRN model

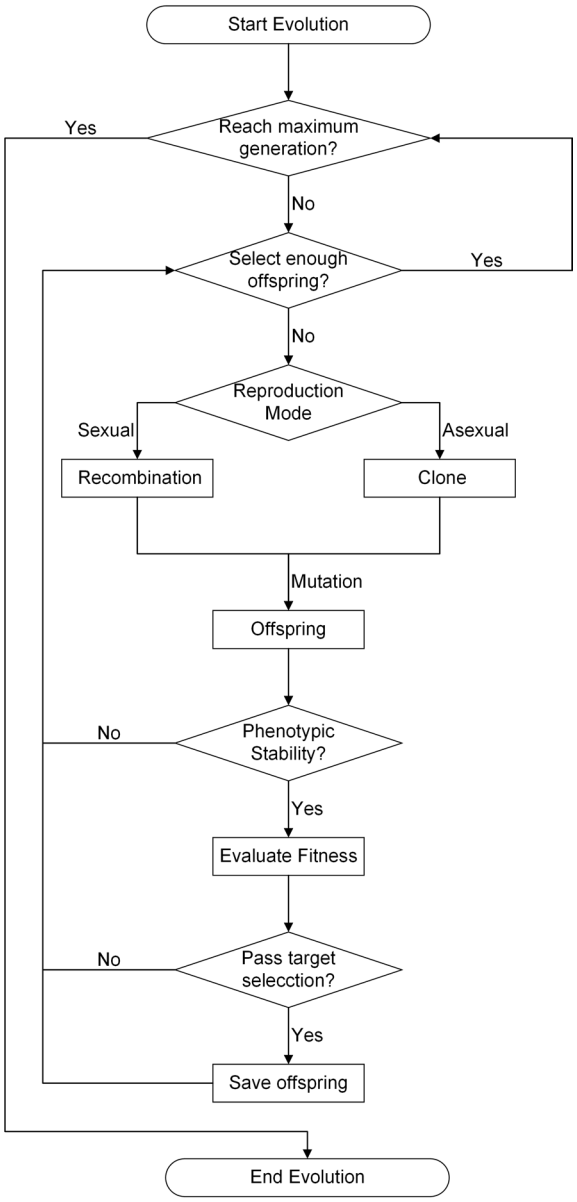


Figure 2-5: Flow chart of the evolution process.

is to find the target phenotype, formally defined as: Given f: $\mathbb{S} \to \mathbb{R}$ find $\mathbf{S}^* \in \mathbb{S}$ such that $f(\mathbf{S}^*) \leqslant f(\mathbf{S})$. Here, the objective function can be defined as $D(\mathbf{s}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}})$.

Using Markov chain theory, we can formally prove that the evolution process of the Wagner model can be regarded as an optimisation process searching for the target phenotype. The convergence analysis based on supermartingales can be easily adapted from the genetic algorithms (Rudolph, 1997; Reeves and Rowe, 2002). A similar convergence proof can be found in Yin et al. (2012a). The convergence analysis provides a mathematical foundation for applying the Wagner model to the machine learning field, as discussed in Section 2.7.2.

2.5 Initial population properties

To gain an impression of the properties of initial gene regulatory networks, in this section, I investigated their stability, robustness and path length.

2.5.1 Stability

I first tested the probability of stability in randomly generated networks. As illustrated in Figure 2-6, smaller networks are more likely to be stable. Moreover, the relative frequency of stability in networks with low levels of connectivity is higher than that of networks with high levels of connectivity. This is in general accordance with previous work (typically done at connectivity $c = 0.75$, e.g. Azevedo et al. (2006)) which indicates that larger networks with complex topology tend to be unstable. Similar patterns are also observed in networks with different values of activation constant a (see supporting information in Appendix A). Generally, when a is small ($a = 1$), networks have a higher initial stability. Note that the pattern is much more profound for networks with smaller sizes ($N = 5$, 10 and 15).

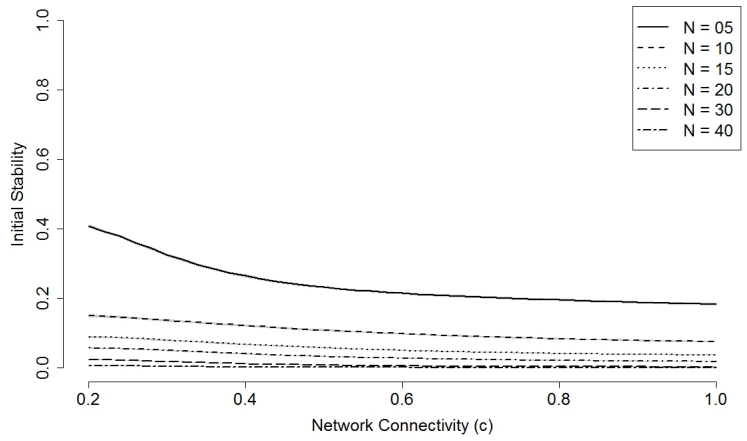


Figure 2-6: Stability of randomly generated networks. *For each network size ($N = 5$, 10, 15, 20, 30 and 40) with each connectivity given from a range of values in continuous intervals ($[0.2, 1]$, step size 0.02), the initial stability (proportion of randomly generated gene networks that were stable) was tested based on an initial $10,000$ randomly generated gene regulatory networks. The system-level parameters were set to be $a = 100$, $devT = 100$ and $\tau = 10$. The shaded areas represent 95% confidence intervals based on 100 independent runs.*

2.5.2 Robustness

Next, I explored the robustness of initially stable networks; that is, I investigated the probability that stable networks remain stable after a single round of mutation. Here, a single mutation means exactly one non-zero entry in an individual's genotype would

be mutated. Given that the initially stable networks were collected from the original randomly generated ones, it would seem reasonable to predict that the small stable networks are more likely to break after one mutation round, since they contain fewer pathways and a single mutation, therefore, has a greater proportional effect. However, the results in Figure 2-7 show the opposite effect: the stability of the small networks is still high (cf. Figure 2-6). The mutation operation is effectively an alternative way of generating new networks; thus, the mutated networks have the same properties as the initial ones. Similar patterns are also observed in networks with different values of activation constant a (see supporting information in Appendix A). Generally, when a is small, networks have a higher initial robustness.

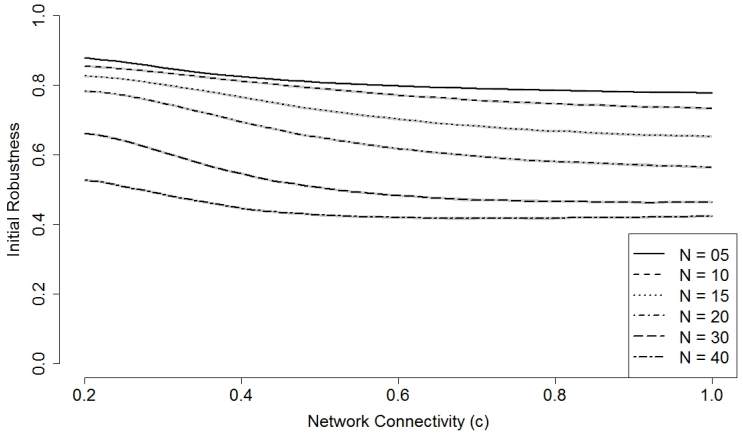


Figure 2-7: Robustness of initially stable networks. *For each network size ($N = 5$, 10, 15, 20, 30 and 40) with each connectivity given from a range of values in continuous intervals ($[0.2, 1]$, step size 0.02), the robustness (proportion of stable networks after exposure to a single round of mutation) was tested based on an initial $10,000$ randomly generated stable gene regulatory networks. The system-level parameters were set to be $a = 100$, $devT = 100$ and $\tau = 10$. The shaded areas represent 95% confidence intervals based on 100 independent runs.*

2.5.3 Path length

In the third set of experiments, I measured the path length of initially stable networks. Here, the path length, as defined in Wagner (1996), refers to the number of time steps[17], as used in Equation (2.1), that the network takes from an initial state $\mathbf{s}(0)$ to reach an equilibrium state $\mathbf{s}_{\mathrm{EQ}}$. From Figure 2-8, we can clear see that larger networks need more time to reach an equilibrium state. Moreover, networks with low levels of connectivity are able to stabilise faster than networks with high levels of connectivity, especially for networks with sizes of $N = 15$, 20 and 30. Similar patterns are

[17]Here, time steps refer to the minimum iteration times required for a network to reach an equilibrium state using Equation(2.1).

also observed in networks with different values of activation constant a (see supporting information in Appendix A). Generally, when a is small ($a = 1$), networks need much more time to reach an equilibrium state, especially for networks with a size of $N = 5$ in comparison with the results when a is large (cf. Figure 2-8). However, the path length slightly decreases for networks with sizes of $N = 10$, 15 and 20 when $a = 1$. Note that all initially stable networks that were used to measure path length were generated under the condition by which I fixed $devT$ to be 100.

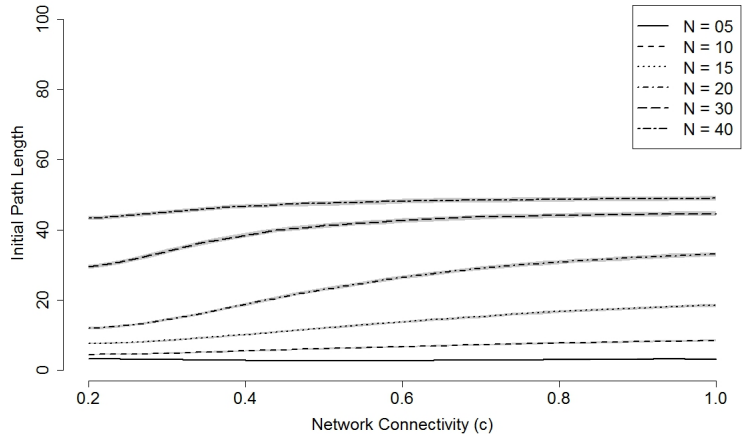


Figure 2-8: Path length of initially stable networks. *For each network size ($N = 5$, 10, 15, 20, 30 and 40) with each connectivity given from a range of values in continuous intervals ($[0.2, 1]$, step size 0.02), the path length (minimum time steps for reaching an equilibrium state) was tested based on an initial $10,000$ randomly generated stable gene regulatory networks. The system-level parameters were set to be $a = 100$, $devT = 100$ and $\tau = 10$. The shaded areas represent 95% confidence intervals based on 100 independent runs.*

2.6 Discussion

Networks of transcription factors are essential for forming developmental patterns in practically all organisms (Guelzim et al., 2002; Davidson et al., 2002; Siegal et al., 2007). The process of development reduces the effects of genetic or environmental perturbations due to the nonlinearity of genotype-phenotype mapping that enhances the robustness of the system, whilst constraining phenotypic diversity, and consequently inhibiting certain evolutionary pathways (Thomas et al., 2014; Pinho et al., 2015). Although many previous studies have shown that the process of development is critical for the study of evolution, the underlying mechanism, in particular of how the developmental process affects evolutionary dynamics that can drive evolutionary innovations, is still poorly understood.

Wagner's GRN model, which has mathematical roots originating from the Ising model (Ising, 1925) and neural networks (Hornik et al., 1989) (see the review article by

Fierst and Phillips (2015) on gene network family trees), has helped integrate network thinking into biology and motivated a new research theme focusing on the evolution of genetic networks (see the review of current papers in Section 2.2).

Mutations in Wagner's GRN model or other similar models of natural systems are shown to be an important source of innovation. Previous studies have focused on separating two sources of mutations, genetic and non-genetic (Masel, 2004; Sevim and Rikvold, 2008; Kimbrell and Holt, 2007; Ciliberti et al., 2007b; Martin and Wagner, 2008; Espinosa-Soto et al., 2011a; Pinho et al., 2015). On the one hand, genetic mutations refer to perturbations occurring to the genotypes[18]. These mutations usually have a weaker effect in altering a gene's phenotypic state or causing instability of the network, since the complex interactions among genes can buffer against mutations occurring at the genotype level. Non-genetic mutations, on the other hand, refer to perturbations caused by internal noise or environmental factors. These mutations may sometimes have a strong effect by causing oscillatory dynamics in phenotypic stability, especially changes occurring in initial gene expression patterns. Although previous studies have investigated many different types of mutation, it remains obscure as to how those mutations systemically affect phenotypic stability.

In addition to mutation, recombination is also believed to be critical to affecting the underlying evolutionary dynamics in the context of genetic networks. Recombination is modelled in Wagner's GRN in the manner of the free recombination of swapping rows between two parental genotypes. This operation follows the biological assumptions that recombination happens more often between genes, and tight linkage occurs among regulatory elements within a promoter (Wagner, 1994, 1996). Previous work has focused on the benefits and apparent low costs of recombination, to reconcile the traditional antagonistic view that recombination is more likely to damage well-adapted lineages due to massively shifting patterns of gene regulation (Azevedo et al., 2006; MacCarthy and Bergman, 2007a; Martin and Wagner, 2009; Lohaus et al., 2010; Wagner, 2011b; Le Cunff and Pakdaman, 2014; Wang et al., 2015). Although MacCarthy and Bergman (2007a) and Lohaus et al. (2010) previously introduced a modifier of recombination into the model, different recombination modes have not yet been studied thoroughly, given the variety of mating systems and strategies in nature (Shuster and Wade, 2003; Shuster, 2009).

In the seminal paper of Wagner (1994), the mathematical foundation of his GRN model was formally described. In the paper, Wagner showed that given an initial state $\mathbf{s}(0)$, the developmental process converges ultimately to a stable equilibrium state $\mathbf{s}_{\mathrm{EQ}}$. In this chapter, I have discussed that the evolution process modelled in Wagner's GRN model can be regarded as an optimisation process that converges on the target configu-

[18]In Wagner's GRN model, genetic mutations are assumed to be epistatic mutations that alter the gene's regulation strength to other genes, but not mutations that occur at the coding sequence at the lowest level.

ration. Besides the convergence analysis, a few other studies have employed theories for calculating periodic orbit to study the systematic behaviour of developmental processes (Pinho et al., 2012, 2015). However, it is still not clear how mutation and recombination operators modelled in the system change periodic orbit and ultimately affect the underlying evolutionary dynamics.

Previous work has shown that sparse networks are more stable than dense networks (Pinho et al., 2012). Here, I have observed a similar pattern by varying network sizes and activation constants (see Figure 2-6). Furthermore, I have shown that randomly generated stable sparse networks also have a higher robustness against mutations than dense networks (see Figure 2-7), although sparse networks may evolve to be more sensitive to mutations than networks that are more densely connected under selection for phenotypic stability (Wagner, 1996; Siegal and Bergman, 2002). However, Leclerc (2008) showed that if the costs of complexity are considered, then robust networks are more likely to be sparsely connected. This may help explain why sparse networks tend to be favoured by evolution in natural systems (Luscombe et al., 2004). As Wagner (1996) and Siegal and Bergman (2002) suggested, the path length or time to reach phenotypic equilibrium may partially account for the underlying mechanism of stability and robustness. This is because if the phenotypic stabilising process takes more time, then the network is more likely to be perturbed by internal noise or environmental factors. Here, I have observed a similar pattern, as in Wagner (1996), to support this likelihood by showing that sparse networks tend to have a shorter path length to reaching equilibrium (see Figure 2-8). Note that changing the activation constant, a, which indicates the sensitivity of the regulatory response to output phenotypes, can quantitatively affect initial stability, robustness and path length (see supporting information in Appendix A).

It should be emphasised that parameters used in Wagner's GRN model, such as population size, number of genes, network connectivity and the activation constant, will not typically change the qualitative results of general properties or patterns emerging from the evolved system (Wagner, 1996; Siegal and Bergman, 2002; Azevedo et al., 2006). In particular, previous studies have suggested that many biological networks have a scale-free topology; that is, the degree distribution of nodes follows a power law (Barabási and Albert, 1999; Newman et al., 2006). However, Wagner (1996), Azevedo et al. (2006), Siegal et al. (2007) and Pinho et al. (2012) have shown that the degree distribution itself does not have a major effect on functional properties associated with nodes. Therefore, although the networks I use in the following chapters are randomly generated and the parameter space has not been thoroughly explored, it is expected that the patterns or properties I have observed could be applied generally to most scale-free networks and the results presented in this dissertation are representative.

Finally, the main caveats of the model are summarised by Wagner (1994, 1996) as

below. These also apply to the general model assumptions made in the remainder of the dissertation:

(1) It is assumed that each gene expression pattern is regulated exclusively on the transcriptional level.

(2) It is assumed that each gene of the network produces only one species of an active transcriptional regulator.

(3) It is assumed that enhancer elements act independently from enhancer elements for other regulators of the same gene.

(4) It is assumed that strong cooperative effects of transcriptional activation by individual transcription factors are mainly responsible for the strong transcriptional activation or repression of a target gene.

2.7 Summary and future work

In this chapter, I have reviewed all currently available research papers that have used Wagner's GRN model. These previous research studies have been grouped into research topics such as robustness, sexual reproduction and evolvability, which are closely related to the research work presented in the following chapters. I have presented the implementation of Wagner's GRN model and its variants in details. Specifically, I have introduced the key operators in the model, such as mutation, recombination and two layers of selection (selection for phenotypic stability and target phenotype). I have also described the evolution process of the model. Using Markov chain theory, I have further discussed that the evolution process of the Wagner model can be considered as an optimisation process in which the probability of finding a target phenotype can converge to probability one. Finally, I have investigated network characteristics such as stability, robustness and path length in randomly generated initial populations. I have shown that networks with a small size and a sparse connectivity generally have a higher initial stability and robustness and a shorter path length. Some possible future research directions regarding improving the model and applying the model to a new application area are presented below.

2.7.1 Combining network stability and function for fitness evaluation

In Wagner's GRN model, there are typically two layers of selection — selection for phenotypic stability and selection for target (optimal) phenotype (Wagner, 1996; Siegal and Bergman, 2002; Azevedo et al., 2006). In most previous papers, if the network cannot achieve phenotypic stability with an equilibrium phenotypic state, then it will be labelled an 'unviable' network and will consequently be wiped out from the

population pool immediately. In other words, phenotypic stability is very restricted in the system, and evolutionary pathways may therefore be highly constricted, as per Pinho et al. (2012, 2015). However, in many biological organisations, such as proteins, there is a balance between stability and function. Many previous empirical studies have shown that new enzymatic functions of a protein are more likely to be accompanied by significant losses in protein stability, which suggests that there is a 'trade-off' between acquiring new enzymatic functions and retaining stability (Pakula and Sauer, 1989; Shoichet et al., 1995; Tokuriki et al., 2008). Therefore, relaxing selection for phenotypic stability would also be biologically realistic. In future work, evaluation of individual fitness is expected to take both network stability and function into consideration.

2.7.2 Application to artificial intelligence and machine learning

In this chapter, I have shown that Wagner's GRN model has been employed to study many fundamental evolutionary and ecological questions. However, would it be possible to introduce such a model derived from computational biology into another research field, especially artificial intelligence and machine learning? It has been found that Darwinian processes of mutation, recombination and selection are useful for generating complex adaptations via evolutionary computation, a subfield of machine intelligence (Wagner and Altenberg, 1996; Yin et al., 2012a; Spirov and Holloway, 2013). Many computational evolutionary algorithms have been used to solve real-world engineering optimisation problems (Yin et al., 2012b; Wang and Yin, 2014; Wang, 2015). For example, genetic algorithms (GAs) are methods well-suited for search and optimisation in non-linear and high-dimensional problems (Goldberg, 1989). Convergence to near-optimal solutions is often perceived as the goal for GAs. Since the goal of Wagner's GRN model is to find an optimal (target) phenotype, it should be possible to develop a similar system for discovering highly-evolvable genomes by exploiting genetic networks (van Dijk et al., 2012; Payne et al., 2014). The many-to-one mapping mechanism of genotype to phenotype explicitly modelled in Wagner's GRN model enables genes to buffer against and even exploit likely variations in the genome. In addition, such a dual learning system — coupled plasticity — is known to accelerate evolution in the right contexts (Hinton and Nowlan, 1987; Kashtan et al., 2007). Hinton and Nowlan (1987) focused on the interaction between evolution and learning, showing that coupled plasticity can solve a problem that is extremely difficult for an evolutionary process on its own. In particular, the genotype used in Wagner's GRN can be regarded as the hierarchical structure that controls the network output (phenotype), i.e., represented as a possible solution to the problem. Therefore, the aim is to explore how the robustness of genetic networks can improve the evolvability of evolutionary computation methods by exploiting genotypes to learn the structures required for rapid adaptations to environmental changes. Some preliminary results are presented in Wang et al.

(2014a).

Chapter 3

Characteristics of compensatory mutation in gene regulatory networks

3.1 Introduction

A significant open question in evolutionary biology is understanding how gene pathways evolve (Wilke and Adami, 2001; Wilke et al., 2003; Beerenwinkel et al., 2007; Lehner, 2011; Rokyta et al., 2011; Park and Lehner, 2013). There have been extensive studies on evolution models showing that gene regulatory networks can evolve by natural selection (Ciliberti et al., 2007a; Crombach and Hogeweg, 2008; Tsuda and Kawata, 2010; Cotterell and Sharpe, 2013). However, gene regulatory networks could also evolve through low-fitness intermediates (Wagner and Wright, 2007; Romero and Arnold, 2009; Olson-Manning et al., 2012), although this idea does not have much experimental support. The reason for the lack of empirical evidence is partly because gene regulatory pathways must go through low-fitness intermediates in order to pass through or shift from one fitness peak to another. This is unlikely to happen, given that people generally believe that low-fitness individuals will be immediately wiped out due to rigorous selection in nature. However, this general view may be biased, because it has not taken the frequency of selection into consideration. If the selection on particular networks or on particular parts of networks is sporadic or even relaxed, then it is possible that the function of broken networks, i.e., low-fitness individuals, can be restored by, for example, compensatory mutations, before the next round of rigorous selection is applied.

Most mutations are thought to be harmful in terms of decreasing individual fitness. However, not all mutations are deleterious or have the same detrimental effects on all individuals. There are occasionally beneficial mutations, for example, compensatory mutations (Kulathinal et al., 2004; Piskol and Stephan, 2008; Covert et al., 2013), which could potentially contribute to gene pathway evolution (Kimura, 1985; Moore

et al., 2000; Levin et al., 2000; Choi et al., 2005; Meer et al., 2010). However, previous work has assumed that compensatory mutation is not likely to play an important role in the evolution of independently acting genes. This is because the frequency of deleterious mutation is low and the frequency at which a new mutation compensates for the previous deleterious mutation is expected to be even lower. Furthermore, if the compensatory mutation restores fitness, then its probability of fixation in the population is the same as any allele under drift, the inverse of twice the effective population size (Wright, 1931a; Charlesworth, 2009). Therefore, compensatory mutations are expected to be very rare and assumed to be inconsequential, occurring only in low-fitness lineages which are eventually eliminated by natural selection. Thus, although compensatory mutation has long been considered to be of great potential significance (Parsch et al., 1997; Wagner, 2000; Crawford et al., 2007), existing theories indicate or assume that it is unlikely to contribute to the evolution of independently acting genes (Wright, 1931a,b; Stephan, 1996; Parsch et al., 1997; Whitlock and Otto, 1999; Whitlock et al., 2003; Zhang and Watson, 2009).

However, mutations do not only happen in independently acting genes but also in genetic networks where there are plenty of sites of complex interactions that could be mutated. If a deleterious mutation occurs at a locus that is not presently subjected to strong selective pressure, then as long as a compensatory mutation occurs before the lineage is driven to extinction, it may restore the lineage's fitness. Thus, the frequency and nature of compensatory mutations are of substantial importance for understanding their impact on pathway evolution.

Compensatory mutations could, therefore, be expected to play a key role in the formation of gene regulatory networks. The frequency at which deleterious mutations incapacitate gene regulatory pathways is likely to be substantially higher than that for an independently acting gene, because there will inevitably be many more possible sites to mutate. We do not know the frequency at which mutations in incapacitated networks can compensate for previous deleterious mutations. But because mutation, by definition, occurs in networks that were previously functional, it seems logical that there could be a wide range of mutational sites and magnitudes that might restore the function of a network. If the frequency of compensatory mutation is high and persistent enough over time, then there is a high probability that some compensatory mutations will be maintained, even if solely by drift.

In this chapter, for the first time, I address questions about the frequency, location and effect size of compensatory mutations using the evolutionary framework provided by gene regulatory network theory (Wagner, 1996; Siegal and Bergman, 2002; Azevedo et al., 2006). I show that the frequency of compensatory mutation is not only relatively high but is also relatively insensitive to the size and connectivity of the network. I find that compensatory mutations are likely to occur in genes at or adjacent to the site of a

previous deleterious mutation, in contrast to the more distributed locations of neutral mutations. The results also show that compensation is driven by mutations with a relatively large regulatory impact, whereas small-effect mutations are more likely to be neutral. These findings show that compensatory mutations have unique properties compared with neutral mutations, and indicate that gene pathway evolution may be far less constrained than previously considered.

3.2 Methods

In the modelling approach, I assumed either that time lags occur between bouts of strong selection for phenotypic stability (see Section 2.3.6) or that selection only acts sporadically on the networks we observe. Therefore, the timescale for accumulating mutations is longer than the timescale between rounds of strong selection for phenotypic stability. Compensation was simply defined as the property of recovering phenotypic stability after a single mutation in a compromised network. Therefore, individual fitness was simply assigned as either 1 (if the network was stable) or 0 (if the network was unstable). The system-level parameters were fixed to be $a = 100$, $devT = 100$ and $\tau = 10$ in all simulations. To simplify the analysis, the mutation operator (see more details below) was defined as replacing one non-zero entry selected at random in W with another random value drawn from the standard normal distribution $N(0, 1)$. Note that the recombination operation was not allowed in these simulations.

3.2.1 The computational model

The computational model (see Figure 3-1) includes three main stages: 1) the initialisation of the population pool, 2) the collection of unstable networks, and 3) the detection of compensatory mutations. The three stages are indicated by three columns in the figure. In the first stage, stable networks were generated randomly. For illustration, Figure 3-1 A shows ten gene regulatory networks in the initial population pool. These networks have been selected from a population pool of randomly generated networks meeting the criteria of phenotypic stability (see Section 2.3.6). The initial networks are all stable and, therefore, allocated high fitness (1). After one mutation round, four networks (indicated by red filled circles) have become unstable and are therefore designated as having low fitness (0). In the third phase, following another round of mutation, one of the low-fitness networks (the ninth) has recovered stability, but another (the second) has lost it. Note that circles with a dashed contour, as shown in Figure 3-1 A, are those networks not considered for the study. Each of the ten networks is composed of five genes ($N = 5$), indicated by five junctions, with varying connectivity. In the second stage (see Figure 3-1 B), each gene regulatory network has been mutated (red edge) and the resulting unstable networks have been collected for

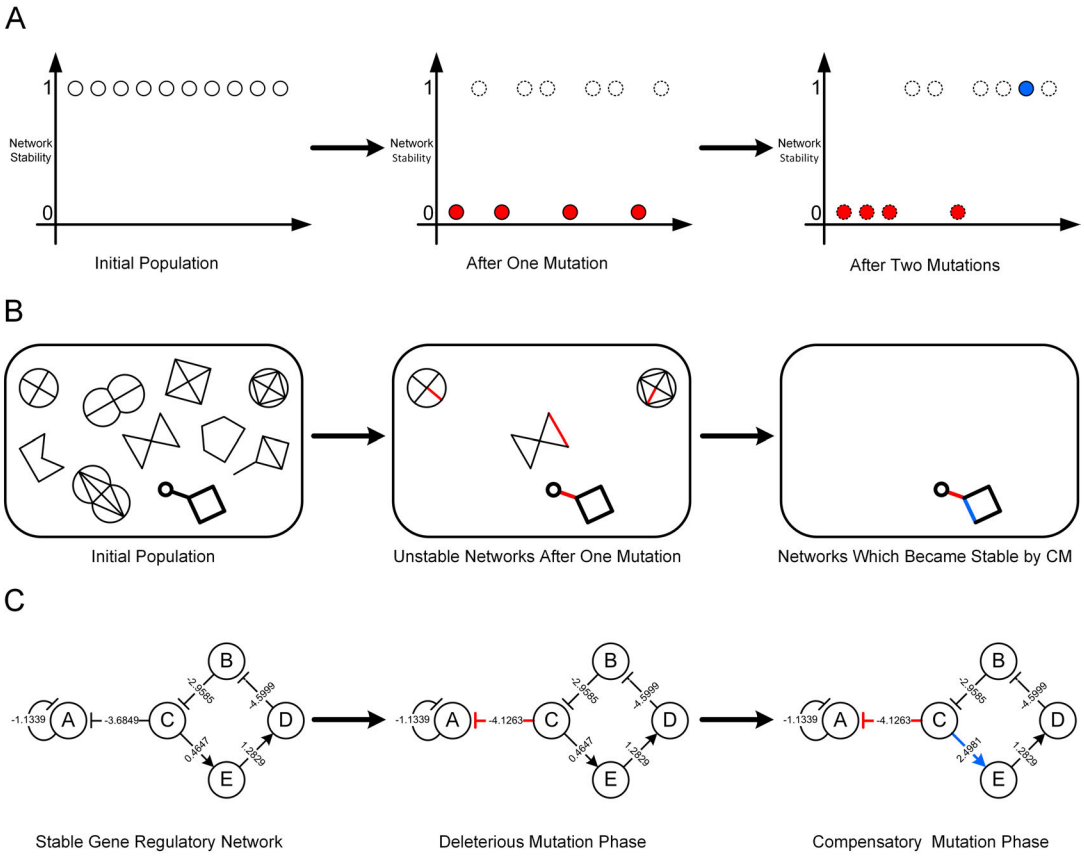


Figure 3-1: Overview of the computational model for exploring characteristics of compensatory mutation. *(**A**) Fitness of the gene regulatory network population. Note that dashed circles are networks no longer considered for the study. (**B**) The population pool of gene regulatory networks. Note that a red edge indicates a deleterious mutation and a blue edge a compensatory mutation. (**C**) View of a single network.*

further testing. In the third stage, the unstable networks have undergone a second round of mutation. I could then collect any newly stable networks. In this case, one network's mutation has been compensatory (blue edge). Figure 3-1 C shows an initially stable gene network which contains five genes: A–E. Each edge is directed and indicates the strength (weight) of the influence on one gene of another. In the Deleterious Mutation Phase, a mutation occurs on $\overrightarrow{CA}$ (red edge), which leads to the failure of stabilisation of the gene phenotypic states. In the Compensatory Mutation Phase, the compromised network is recovered by another round of mutation (blue edge), one occurring on $\overrightarrow{CE}$.

Initialisation

Each individual network in the population was generated with a gene regulatory matrix W associated with an expression state vector $\mathbf{s}(0)$. Specifically, the matrix

was generated by randomly filling W with $\lfloor c \times N^2 \rfloor$ non-zero elements $w_{i,j} \sim N(0,1)$ $(i,j = 1,2,\ldots,N)$. The associated initial expression state $\mathbf{s}(0)$ was also set by randomly choosing each $s_i(0) = +1$ or -1.

Mutation

In the mutation operation, exactly one element $w_{i,j}$ $(i,j = 1,2,\ldots,N)$ picked at random in each regulatory matrix W would be replaced by $w'_{i,j} \sim N(0,1)$. Note that the mutation only occurs among non-zero elements. In other words, the mutation process will not change the topology of the original network W in terms of forming new edges or deleting existing edges between two genes.

Selection for stable & unstable individuals

Selection for both stable and unstable individuals was required in all simulations. In the stable selection operation, only individuals which were able to attain phenotypic stability after the mutation process were selected. In contrast, only individuals which were incapable of reaching equilibrium were chosen in the unstable selection process. Note that accepting the possibility of unstable networks in viable individuals, and defining such individuals as 'impaired' rather than 'dead' is the primary departure from previously published models (Wagner, 1996; Siegal and Bergman, 2002; Azevedo et al., 2006).

3.2.2 Estimating the relative frequency of compensatory mutation

In this set of experiments, I investigated the compensatory mutation frequency in previously stable networks (see Figure 3-3). Specifically, I started from a population pool of $10,000$ sample networks where each stable network was randomly generated. I exposed these initially stable networks to a single round of mutation. Then, I focused on those unstable networks where each network contained a single deleterious mutation. Next, I exposed these compromised networks to an additional round of mutation. Finally, I tested the stability of the resulting networks. The stable networks at this point had experienced compensatory mutation. I then measured the frequency of individuals that experienced compensatory mutation.

3.2.3 Locating the compensatory mutations

In this set of experiments, I first sought to visualise locations at which the compensatory mutations are more likely to occur (see Figure 3-4). To this end, in a set of compromised networks (those stable networks that proved fragile to a single round of mutation), I marked the site of the deleterious mutation, then measured the relative frequency of compensatory mutation that occurred at each possible site, including the

site of the deleterious mutation, within this compromised network. For each possible site, I measured the outcomes over 1,000 simulated mutations on that site (so that only the extent of regulation was mutated randomly, not the location).

To quantify the distance between deleterious and (potentially) compensatory mutation, I first define *distance* as used in this chapter. Suppose a given gene regulatory network, denoted as W, has two marked edges denoted as $\overrightarrow{AB}$ (deleterious mutation) and $\overrightarrow{CD}$ (compensatory mutation), where A, B, C and D represent different genes in W and $\overrightarrow{\cdot}$ marks the edge direction. The distance between $\overrightarrow{AB}$ and $\overrightarrow{CD}$ can be calculated as

$$DIS\left(\overrightarrow{AB},\overrightarrow{CD}\right)=\begin{cases}0 & \text{if } A=C \text{ and } B=D\\ 1 & \text{if } A=D \text{ and } B=C\\ dis(A,C)+1 & \text{if } B \text{ and } D\notin path(A,C)\\ dis(A,C) & \text{if } B \text{ or } D\in path(A,C)\\ dis(A,C)-1 & \text{if } B \text{ and } D\in path(A,C)\end{cases} \tag{3.1}$$

where $dis(A,C)$ is the fewest edges possible from A to C and $path(A,C)$ includes the vertices on the shortest path between A and C in network W.

Figure 3-2 provides an example process of compensatory mutation in a gene regulatory network. This stable network can be compromised by a single deleterious mutation (marked in red) and compensated by an additional mutation (marked in blue). According to Equation (3.1), the distance from deleterious mutation site $\overrightarrow{CA}$ to compensatory mutation site $\overrightarrow{CE}$ can be calculated as: $DIS(\overrightarrow{CA},\overrightarrow{CE})=1$.

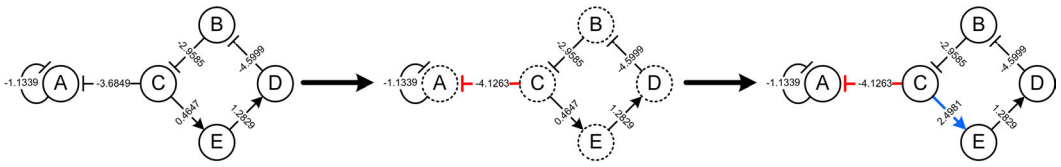


Figure 3-2: An example process of compensatory mutation in a gene regulatory network. *The initially stable gene network contains five genes: A, B, C, D and E. In the initial network (on the left side), each directional edge represents the strength (weight) of interaction between the linked two genes. The initial gene expression pattern is* $\mathbf{s}(0)=(-1,-1,+1,+1,+1)$. *In the compromised network (in the middle), a mutation occurs on* $\overrightarrow{CA}$ *(indicated in red), which leads to the failure of stabilising the gene expression patterns (marked by dashed circles). In the compensated network (on the right side), the compromised network is fixed by an additional mutation that occurs on* $\overrightarrow{CE}$ *(indicated in blue), reaching an equilibrium expression* $\mathbf{s}_{\mathrm{EQ}}=(-1,-1,+1,+1,+1)$.

Next, I compared the relative frequencies of compensatory mutation among gene networks whose marked edges (caused by additional mutation) were 0, 1, 2, 3, and 4 steps away from the deleterious mutation (see Figure 3-5). I also performed similar experiments for medium ($N=20$) and large networks ($N=40$), as shown in Figures B-

2 and B-3.

3.2.4 Exploring the effect size of gene regulation on compensatory mutation frequency

In this set of experiments, I investigated effective changes in gene regulation associated with these mutations (see Figure 3-6). Specifically, I conducted experiments to measure the frequency of compensatory mutation when the second mutation had an additional weight added to it. I studied a range of weight changes from ($w = [-5, 5]$) with a step size of 0.05. For each step size, I first performed one mutation round as usual on the initial population of stable networks, creating a sub-population of $10,000$ compromised networks. Then, for these mutated networks I performed a second mutation round; however, this time instead of replacing one entry in the adjacency matrix with $N(0, 1)$, I added a fixed value w drawn from $[-5, 5]$ to the original value of the randomly picked site. Then, I measured the frequency of second mutations restoring the network stability. I also performed similar experiments for medium ($N = 20$) and large networks ($N = 40$), as shown in Figures B-4 and B-5.

3.2.5 Exploring the distribution of regulation in initially stable, compromised and restored networks

In this set of experiments, I investigated the distribution of regulation in initially stable, compromised and restored networks (see Figure 3-7). Specifically, I collected $10,000$ sample regulatory values each from edges of randomly generated stable networks, edges where deleterious mutations occurred (compromising network stability), and edges where compensatory mutations occurred (restoring previously compromised networks). I then measured their corresponding distributions, discriminating between self- and non-self-regulatory edges. Note that separating self- and non-self-regulatory edges helps investigate whether they have different properties, given that positive regulation is more likely to be observed on self-regulatory edges in nature (Fournier et al., 2007; Ramos et al., 2011; Sugár and Simon, 2014). I also performed similar experiments for medium ($N = 20$) and large networks ($N = 40$), as shown in Figures B-6 and B-7.

3.2.6 Exploring properties of location and size effects in neutral mutations

In this set of experiments, I investigated properties of location and size effects in neutral mutations which served as control groups for the experiments described in Sections 3.2.3 and 3.2.4. Specifically, to test the location effect, I collected a population pool of stable networks that had been subjected to one round of mutation (neutral). Then, I measured the probability of stable networks after performing a second mutation

that was 0, 1, 2, 3, and 4 steps away from the previous neutral mutation site based on 10,000 sample networks for each distance category (see Figure 3-8). Similarly, to test the mutation size effect, I collected a population pool of stable networks that had been subjected to one round of mutation (neutral). Then, I measured the probability of stable networks after performing a second mutation that had a particular shift in gene regulation from $[-5, +5]$ based on 10,000 sample networks for each shifted-weight category (see Figure 3-9). In both tests for location and size effects, I also performed similar experiments for medium ($N = 20$) and large networks ($N = 40$), as shown in Figures B-8 and B-9.

3.3 Results

Using the well-established synthetic Wagner model of gene regulatory networks described in Section 3.2.1, I was able to explore characteristics of compensatory mutation in the context of genetic networks. The gene regulatory theory is a particularly appropriate method because it explicitly incorporates genetic interactions in an evolutionary framework. Simulation allowed me to generate thousands upon thousands of networks of different sizes and connectivities, which we could not do with *in vivo* approaches, and made it relatively simple to identify, track and understand the properties of all of the compensatory mutations within those networks. A key insight of the model used in this chapter is that whether a mutation is deleterious, compensatory or neutral is entirely dependent on its context within a complex system — a regulatory network evolved for phenotypic stability.

3.3.1 Compensatory mutations are common and relatively scale invariant

I first tested whether compensatory mutation is frequent in the context of gene regulatory networks. I found that the frequency of compensatory mutation is largely scale invariant. From Figures 2-6 and 2-7, we can see that the stability and robustness in initial networks are quite different among varying sizes and levels of connectivity of gene regulatory networks. Which type of network, once compromised, more frequently experiences compensatory mutation? Figure 3-3 answers this question. As can be seen, the patterns of frequency of compensatory mutation depend on network size. For the smaller networks $N = 5$, 10, 15 and 20, the compensatory mutation rates continuously increase as the network connectivity increases, but very gradually. In contrast, for the larger networks $N = 30$ and 40, with the rise in connectivity, the compensatory mutation rates decrease slightly. However, overall the results indicate that the frequency of individuals that can be fixed by compensatory mutation is more sensitive to network size than to network connectivity. The implied probability of compensatory

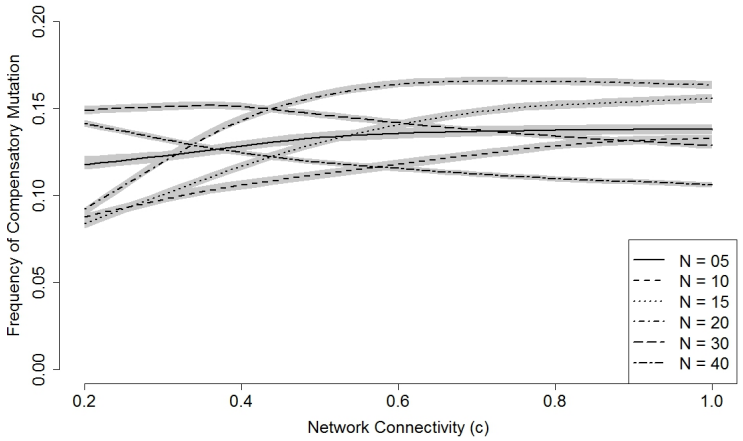


Figure 3-3: The influence of the size and connectivity of a gene regulatory network on its frequency of compensatory mutation. *For each network size (*$N = 5$*,* 10*,* 15*,* 20*,* 30 *and* 40*) with each connectivity given from a range of values in continuous intervals (*$[0.2, 1]$*, step size* 0.02*), the frequency of compensatory mutation was tested based on an initial* $10,000$ *randomly generated stable gene networks. The shaded areas represent* 95% *confidence intervals based on* 100 *independent runs.*

mutation from the relative frequencies observed ranges from 5% to 15% of compromised networks recovering, with the larger rates associated with larger networks. This is marked as relatively scale invariant (see Figure B-1, which is identical to Figure 3-3 but re-scaled), in contrast to the scale dependencies shown for deleterious mutations in Figures 2-6 and 2-7.

3.3.2 Compensatory mutations often occur close to the deleterious mutation's site

I next looked at where these compensatory mutations happened in compromised networks. I found that they are more likely to occur at or close to the site of the original, deleterious mutation. Note that the distance, as described in Section 3.2.3, is defined in terms of regulatory structure, not nucleotide sequence, due to Wagner's GRN model assumption (see more details in Section 2.3). In a typical small network with size $N = 5$ genes (see Figure 3-4 A), I found a 95.8% chance that a mutation that occurs on the exact site of a deleterious mutation compensates for it. The frequency of compensatory mutation is also high on most of the edges close to the original mutation site. Mutations on edges far away from the deleterious mutation site are much less likely to experience compensation. The same basic pattern is also seen in a larger network with size $N = 20$ genes (see Figure 3-4 B), where the frequency of mutations being compensatory, if they occur on the original deleterious site, is 85%. The percentages beside each edge in these figures indicate the proportion of mutations that occur on

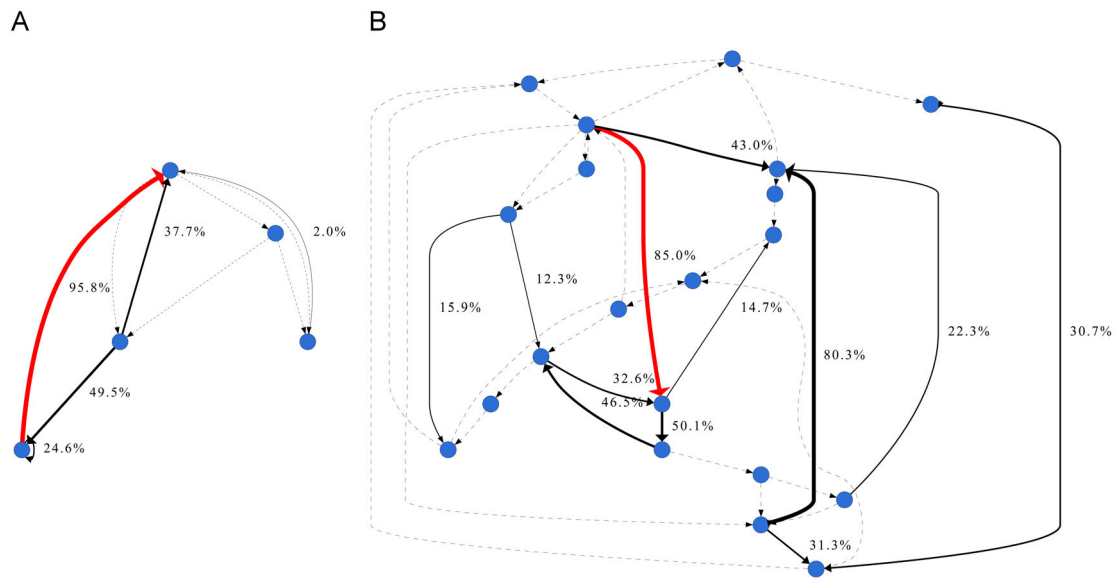


Figure 3-4: Examples of the spatial probability of compensatory mutation occurring on gene networks. *In both examples ($N = 5$ (**A**) and 20 (**B**)), for a particular compromised network that was stable initially, I executed one additional mutation round* $1,000$ *times on each edge. Then, the percentage of each broken edge that could be fixed (Note: the compensatory mutation occurred on this edge) after the mutation operation was measured. Finally, I marked each broken edge whose percentage was above* 0%*. Note the solid line with different widths to indicate different fixable probabilities and the dashed line to represent the edges that were unable to be fixed. The original deleterious mutation occurred on the edge marked in red. Note: The directed edge represents the interaction between two connected genes. But I do not distinguish negative or positive regulations in the provided examples.*

that edge that are compensatory, out of the $1,000$ simulated second rounds of mutation I ran on each edge for each network after it had previously suffered a single deleterious mutation. In general, as these representative figures indicate, the compensatory effect could happen in many positions in a broken network, but is more likely to be observed at sites that are close to a deleterious mutation's site.

Figure 3-5 (solid line) demonstrates the generality of the result indicated in Figure 3-4. It illustrates the frequency among $10,000$ initially stable gene networks of compensatory mutation against different spatial distances from the single deleterious mutation suffered by each network. As can be seen, compensatory mutations generally occur in edges between genes close to the deleterious mutation site. I restrict the analysis to these five categories because there is only a narrow range of distribution distances for randomly sampled mutations (see dashed line in Figure 3-5). Similar patterns are also observed in networks with different size and connectivity (see supporting information in Appendix B). These theoretical results predict that compensatory mutations are more likely to be observed at or adjacent to the original site of a deleterious mutation in nature. The results also indicate that compensatory mutations in networks may be

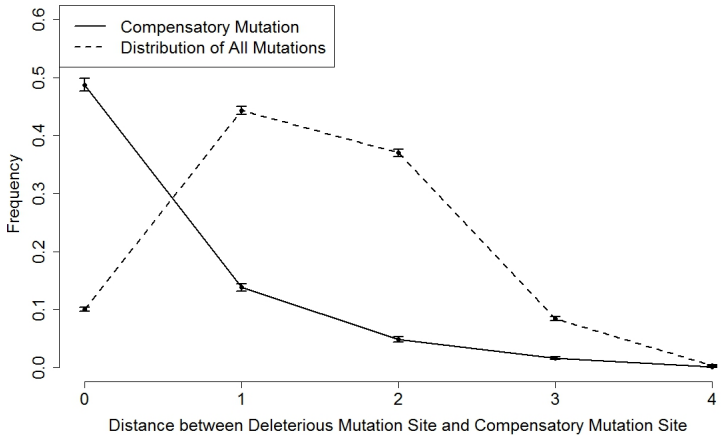


Figure 3-5: The compensatory mutation location and distance distribution of all mutations relative to the original deleterious mutation sites. *For initially stable networks with size* $N = 5$ *and connectivity* $c = 0.4$*, I first collected a pool of compromised networks with deleterious mutations after a single mutation round. I then forced second mutations, classifying these as being* 0 *(on the same site),* 1*,* 2*,* 3 *and* 4 *steps away from the original deleterious mutations. For each of these mutation-site-distance categories, I measured the probability that the mutation was compensatory (that it returned the network to stability), based on* $10,000$ *sample networks collected for each distance category as shown in the solid line. I also recorded the spatial distribution of second mutations (*$10,000$ *sample networks) occurring randomly in those compromised networks with respect to their original deleterious mutation sites, shown in the dashed line. The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

localised to particular areas or features of network topology.

3.3.3 Regulatory changes leading to compensation tend to be large-effect mutations

Next, I investigated how different mutation size influences the probability of compensation in compromised networks. I found that compensation is more likely to be driven by large-effect mutations. Figure 3-6 presents the frequency of compensatory mutation against various intensities of up or down regulation among $10,000$ randomly generated stable gene networks that had experienced a single deleterious mutation. For a randomly chosen site in each network, I experimented with mutations across a range of regulatory strengths. As can be seen, larger regulation changes, both positive and negative, are up to a point associated with an increased frequency of compensatory mutation. However, the shape of the curve for compensatory mutations across all edges is not a symmetrical 'V'. Rather, compensatory mutations occur more by positive changes to gene regulation than by negative changes. The explanation for this phenomenon is rooted in the fact that there are two edge types that can be affected

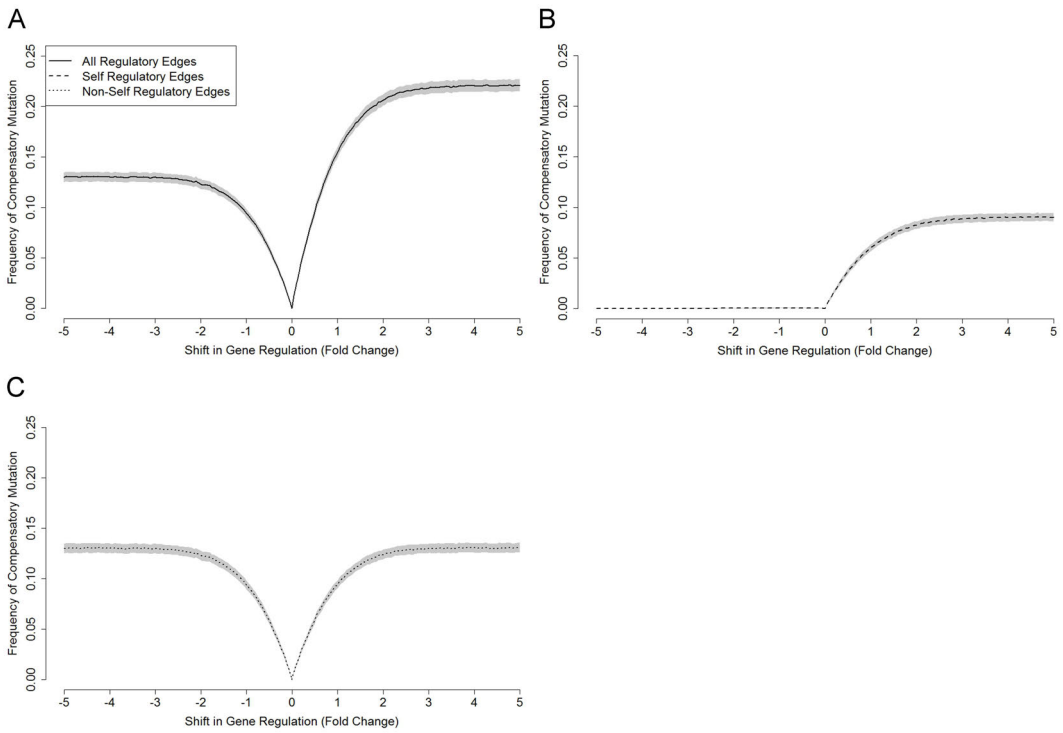


Figure 3-6: The influence of different intensities of gene regulations on the frequency of compensatory mutation. *I first collected* $10,000$ *sample networks that had been made unstable by a single mutation from a pool of initially stable networks with* $N = 5$ *and* $c = 0.4$*. Then, I experimented with how a new mutation of varying intensities of gene regulation altered the chances of restoring gene stability. Specifically, I performed new mutations to those compromised networks with deleterious mutations by adding a weight from* $[-5, +5]$ *(step size* 0.5*) to the original regulatory impact, then assessed the resulting patterns in all regulatory edges (**A**), in self-regulatory edges (**B**) and ignoring self-regulatory edges (**C**). The shaded areas represent* 95% *confidence intervals based on* 100 *independent runs.*

by compensatory mutation: inter-gene regulation connecting two different genes and self-regulating edges. In the simulations, almost no compensatory mutations are both negative and self-regulating (see Figure 3-6 B). Only the 'V' shape for inter-gene regulation is almost symmetrical (see Figure 3-6 C), suggesting that for these, negative and positive regulations are equally likely to be useful. It is true for both the negative and positive cases that compensatory mutation is increasingly likely with greater regulatory strength up to a certain extent. Similar patterns are also observed in networks with different size and connectivity (see supporting information in Appendix B).

Although I found that compensatory mutation tends to the positive, this is not a special property. I confirmed this by investigating the regulatory effects in randomly generated stable networks[1]. From Figure 3-7 A, we can clearly see that there are more positive regulations in both initially stable networks and networks with compensatory

[1]Note that the regulations in these networks are drawn from $N(0, 1)$.

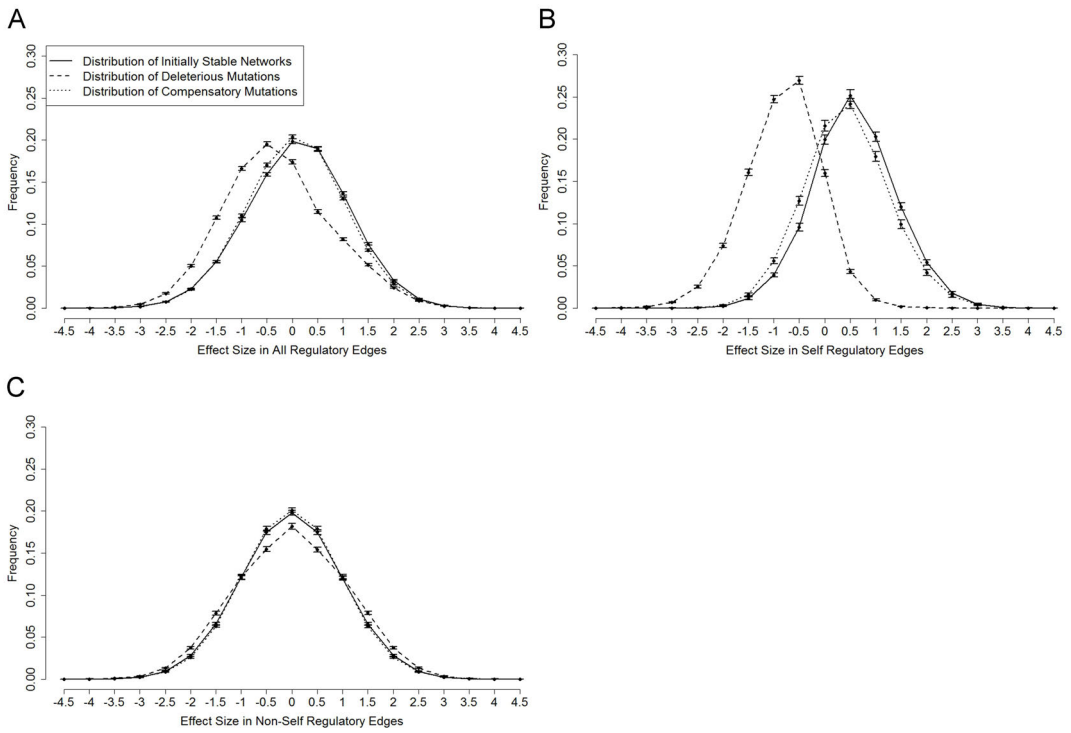


Figure 3-7: The distribution of regulation in initially stable, compromised and restored networks. *For randomly generated stable networks with* $N = 5$ *and* $c = 0.4$*, I collected* $10,000$ *sample regulations. I also collected* $10,000$ *sample regulation weights from deleterious mutations that compromised initially stable networks as well as from compensatory mutations that restored the stability of previously broken networks. I then measured the distributions in all regulatory edges (**A**), in self-regulatory edges (**B**) and ignoring self-regulatory edges (**C**). Given that the regulations are continuous values, I grouped them into* 19 *bins from* $[-4.5, +4.5]$ *(step size* 0.5*). The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

mutations, whereas deleterious mutations tend to be more negative in compromised networks. By separating self- and non-self-regulatory edges, I found that compensatory mutations have a larger effect (in terms of shifting gene regulation) on self-regulatory edges than non-self-regulatory edges (see Figures 3-7 B and C). Similar patterns are also observed in networks with different size and connectivity (see supporting information in Appendix B).

3.3.4 Networks with neutral mutations tend to have different location and size effect proprieties

In order to investigate whether compensatory mutations have any special property in terms of location and size effect, I further conducted similar experiments as described in Sections 3.3.2 and 3.3.3 for networks with neutral mutations. I found that, compared with the results of compensatory mutations, neutral mutations are more evenly

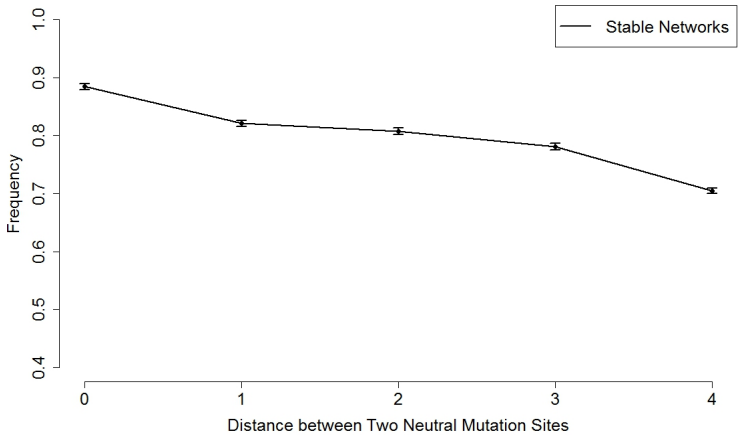


Figure 3-8: Location effect in networks with neutral mutations. *For networks with size $N = 5$ and connectivity $c = 0.4$, I first collected a pool of stable networks with neutral mutations after a single mutation round. I then forced second mutations, classifying these as being 0 (on the same site), 1, 2, 3 and 4 steps away from the previous neutral mutations. For each of these mutation-site-distance categories, I measured the probability that the mutation was neutral (did not impair network stability) based on $10,000$ sample networks collected for each distance category. The error bars represent 95% confidence intervals based on 100 independent runs.*

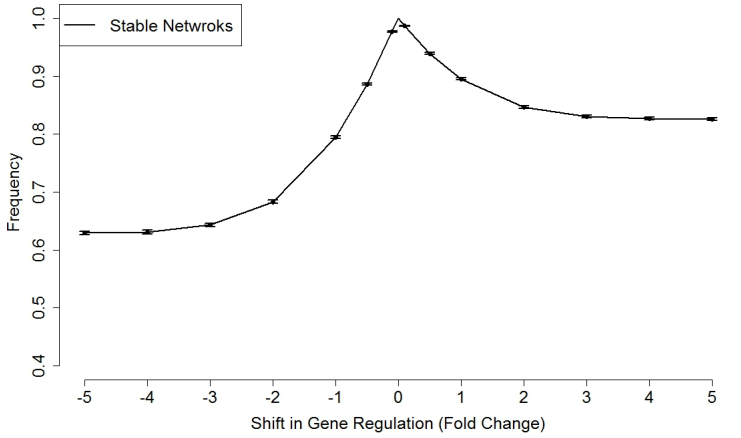


Figure 3-9: Mutation size effect in networks with neutral mutations. *I first collected $10,000$ stable networks with neutral mutations after a single mutation round from a pool of initially stable networks with $N = 5$ and $c = 0.4$. Then, I experimented with how new mutations of varying intensities of gene regulation altered the chance of retaining network stability. Specifically, I performed new mutations to those networks with neutral mutations by adding a weight from $[-5, +5]$ (step size 1 and with four additional regulation shifts: $-0.5, -0.1, 0.1$ and 0.5) to the original regulatory impact, then assessed the resulting patterns. The error bars represent 95% confidence intervals based on 100 independent runs.*

distributed in terms of location, and small-size mutations are more likely to be observed in networks with neutral mutations.

Specifically, instead of measuring the frequency of a second mutation (compensatory mutation) that can restore network stability for a compromised network (which has one deleterious mutation), I measured the frequency of a second mutation (neutral mutation) with different distance and size effects that can retain the stability for a network that has already had one neutral mutation. On the one hand, from Figure 3-8 we can see that the distance effect has a much less profound role in networks with two consecutive neutral mutations than in networks with one deleterious mutation and one compensatory mutation (cf. Figure 3-5). In fact, neutral mutations tend to be enriched if they are far apart in larger networks (see supporting information in Appendix B). On the other hand, from Figure 3-9 we can see that small-effect neutral mutations are more likely to retain the network stability (cf. Figure 3-6). Similar patterns are also observed in networks with different size and connectivity (see supporting information in Appendix B).

3.4 Discussion

Research on evolutionary gene pathways has attracted great attention for decades (Rison and Thornton, 2002; Orr, 2005; Fusco and Minelli, 2008; Iwasaki and Takagi, 2009). As observed in all forms of adaptation, from human development to machine learning, increasing the quality of individual performance sometimes requires radical changes to current strategies and, therefore, passing through phases of lower performance (Plunkett and Marchman, 1991). In the context of evolution, however, these lower-performing individuals might be expected to be 'selected out' before they can consolidate into useful innovation, outcompeted by other individuals holding the older and stable strategy.

In its simplest form, this concern about strong selection is not well-founded. Even strong natural selection is never deterministic, but rather stochastic, with weaker strategies less likely to reproduce, rather than being entirely blocked from it. Further, periods or spaces of strong selection often alternate with periods or spaces of very weak selection, for example, after an ecosystem population cycle or a climactic event that leaves an ecosystem well below carrying capacity for a particular species (Lambin et al., 1998; Liebhold et al., 2004; Sherratt and Smith, 2008). This phenomenon has been shown to promote the spread of initially maladaptive traits such as altruism (Čače and Bryson, 2007; Alizon and Taylor, 2008); similar logic applies here. Thus, low-fitness lineages previously thought to be inconsequential might be sustained long enough to be rescued or even improved upon by compensatory mutation — provided only that the probability of such mutation is great enough and that this rescue is likely to occur before

the lineages are eliminated by natural selection. The importance of the results is that they show that at least for one measure of fitness (phenotypic stability), compensatory mutation is in fact relatively likely. Not only that, its rate is highest in those larger, more fragile networks that are more likely to suffer deleterious mutations.

However, compensatory mutations have not been studied extensively. Many general properties of compensatory mutation are consequently still unknown. This is because compensatory mutations are thought to be rare in independently acting genes. However, mutations do not just happen in those genes. There is substantial molecular evidence for mutations in genes which exhibit complex interactions with other genes (Wilke and Adami, 2001; Wilke et al., 2003; Beerenwinkel et al., 2007; Lehner, 2011; Rokyta et al., 2011; Park and Lehner, 2013; Connelly et al., 2014). Moreover, there is extensive empirical evidence to show that compensatory mutations do occur and can occur quite frequently (Stephan, 1996; Mintseris and Weng, 2005; Poon and Chao, 2005; Poon et al., 2005; Davis et al., 2009; Comas et al., 2012). In this chapter, by adapting the Wagner GRN model presented in Chapter 2, I have demonstrated support for this possibility, that compensatory mutation could potentially be frequent and relatively insensitive to the size and connectivity of the network (Figure 3-3). These findings imply that mutations that are able to fix broken networks offer surprisingly little variation in the context where the mutation happens. This property may facilitate the further study of compensatory mutations, as the findings that are drawn from the model specified by the standard parameters could be representative.

In this chapter, compensatory mutation has been defined as mutation that can restore the phenotypic stability of the network. A compensatory mutation can only occur when the selection for phenotypic stability is relaxed, so that compromised networks can have opportunities to be restored. Note that the deleterious mutation I have modelled here is the lethal mutation that destroys a network's stability. In most previous papers, phenotypic stability is very restricted in the system, and evolutionary pathways may therefore be highly constricted, as per Pinho et al. (2012, 2015). However, in many biological organisations, such as proteins, there is a balance between stability and function. Many previous empirical studies have shown that new enzymatic functions of a protein are more likely to be accompanied by significant losses in protein stability, which suggests that there is a 'trade-off' between acquiring new enzymatic functions and retaining stability (Pakula and Sauer, 1989; Shoichet et al., 1995; Tokuriki et al., 2008). Therefore, relaxing selection for phenotypic stability would also be biologically realistic.

Many recent studies have shown that conventional *de novo* mutations are widely distributed throughout the genome and have a wide distribution of phenotypic effects, from complete lethality to weak benefit with respect to fitness (Sanjuán et al., 2004; Eyre-Walker and Keightley, 2007; Keightley and Eyre-Walker, 2007; Mezmouk and

Ross-Ibarra, 2014). Although there have been no predictive tests of the location of compensatory mutations, empirical studies show that compensatory mutations are often found in proteins that are in or interact with proteins that exhibit a deleterious mutation (Poon et al., 2005; Poon and Chao, 2005; Davis et al., 2009; Comas et al., 2012; Bhattacherjee et al., 2015). Regardless of size, any incapacitated network, as defined in this chapter, carries the network property that it is one mutational step away from stability. This implies a potential for the frequency of compensatory mutation to be relatively invariant to the size of the network, although of course the precise counter of a previous mutation would be increasingly unlikely with more potential mutation sites. In this chapter, I have shown that, compared with neutral mutations, compensatory mutations are much more likely to occur in genes that carry deleterious mutations or are closely linked in genetic pathways (Figure 3-5), although we do not know where those mutations happen in particular networks. This may provide a guide in principle that compensatory mutations are more likely to be observed in sites that are close to the original mutations.

When we consider functional networks, Fisher's geometric model of adaptive evolution argues that adaptive evolution should generally result from the substitution of many mutations of small effect, because advantageous mutations of small effect should be more common than those of large effect (Fisher, 1930; Burch and Chao, 1999). However, when I study these compromised networks, Fisher's rule may not apply. In this chapter, I have shown that, compared with neutral mutations, compensatory mutations with a small size effect are unlikely to repair networks, whereas large-effect mutations are more likely to be able to restore unfunctional networks (Figure 3-6). This may suggest that the broken networks are far away from fitness peaks, so that they need a larger mutation step to be facilitated towards the phenotypic optimum. Although compensation can be caused by both positive and negative weight changes, previous work on levels of gene regulation has provided considerable circumstantial evidence that there are more positive, rather than negative, self-regulations in gene networks. The theoretical simulation result shows that, at least where the compensatory mutation is self-regulatory, it is far more likely to be driven by up regulation. However, it should be noted that this is not a special property of compensatory mutation. In fact, there are more positive regulations in self-regulatory genes in functional networks that have never been through compensation (Figure 3-7). This may account for the high amount of positive self-regulation observed in nature (Fournier et al., 2007; Ramos et al., 2011; Sugár and Simon, 2014).

3.5 Summary and future work

In this chapter, I have studied characteristics of compensatory mutations in a network context using Wagner's GRN model. Specifically, compensatory mutation is defined as a mutation that can restore the stability of a network compromised by a previous deleterious mutation. I have shown that the frequency of compensatory mutation is not only relatively high but is also relatively insensitive to the size and connectivity of the network. When I looked at compensatory mutations more closely, I found that they are likely to occur in genes at or adjacent to the site of a previous deleterious mutation. I have also found that compensation is likely to be driven by large-effect mutations. The characteristics of neutral mutations have been observed to be different from compensatory mutations. Specifically, neutral mutations tend to be more evenly distributed or even enriched when they are far apart. Moreover, small-effect mutations are more likely to be observed in networks with neutral mutations. Some possible future research directions regarding exploring characteristics of compensatory mutations in complex fitness-associated evolutionary scenarios and in scale-free networks are presented below.

3.5.1 Modelling compensatory mutation in complex evolutionary scenarios

In this chapter, the compensatory mutation has been modelled in such a way that it restores a network's phenotypic stability. In other words, the fitness of a network is simply defined as 1, if the network can retain its phenotypic stability when it is subjected to mutation, otherwise 0. The binary value of the fitness substantially helps simplify the computational model, making tracking compensatory mutations much more easier. However, it would be more biologically realistic if compensatory mutation was associated with a complex fitness function that could have continuous values. Thus, the evaluation of phenotypic stability is expected to develop more intermediate values besides the two extreme cases — the network is either stable or unstable. In addition, compensatory mutation can also be modelled in an evolutionary scenario where it improves an individual's phenotypic state to close to the optimum when the network is subjected to target selection (as defined in Section 2.3.7). In such a case, we can further investigate, for example, characteristics of super-compensatory mutation which not only restores an individual's fitness but further increases its fitness to be higher than its original value. Those super-compensatory mutations, although rare, may have a huge impact when they finally emerge, as indicated by Covert et al. (2013). It would be interesting to explore whether these super-compensatory mutations would have the same or different characteristics to the ones I have discovered in this chapter.

3.5.2 Exploring characteristics of compensatory mutation in scale-free networks

In this chapter, I have investigated characteristics of compensatory mutation in randomly generated networks. It would be interesting to explore whether similar patterns could also be observed, for example, in scale-free networks. Although previous studies have indicated that the degree of distribution itself does not have a major effect on functional properties associated with nodes (Wagner, 1996; Azevedo et al., 2006; Siegal et al., 2007; Pinho et al., 2012), the frequency of compensatory mutation may relate to the topology of the network. For example, a hub node may be essential for maintaining network stability. Therefore, networks regulated by one or several hub nodes are expected to be robust. However, if a deleterious mutation occurs on the edge associated with the hub node, then subsequent mutations may be unlikely to restore network stability, since the hub node is regulated by many other genes. It would also be interesting to explore the likelihood of compromised networks being restored by multiple mutations acting simultaneously.

Chapter 4

Compensatory mutation generates regulatory complexity through non-adaptive processes

4.1 Introduction

Although gene regulatory networks underlie all stages of life, from development to adult homeostasis to senescence, we do not understand how they evolve (Davidson, 2010; Hasty et al., 2001; Barabasi and Oltvai, 2004; Boiani and Scholer, 2005; Levine and Davidson, 2005). Many previous studies have suggested that complex genetic architectures are shaped by competitive adaptive processes, which occur when novel gene combinations increase in the population because they confer differential reproductive success (Barabasi and Oltvai, 2004; Madan Babu et al., 2006). However, a substantial body of genomic evidence indicates that gene regulatory networks arise through non-adaptive processes such as genetic drift, mutation and recombination, which can influence how genetic variation is lost but do not alter competitive ability (Lynch, 2007a,b; Lusk and Eisen, 2010; Fernández and Lynch, 2011; Sorrells and Johnson, 2015; Payne and Wagner, 2015). However, it is still poorly understood how non-adaptive processes might generate regulatory complexity.

Compensatory mutation could play an essential non-adaptive role in generating regulatory complexity. During periods of relaxed selection, regulatory networks with lethal mutations have the potential to be compensated by additional mutations. If compensatory mutation occurs frequently enough and generates different patterns of gene regulation than networks with neutral mutations, then it could alter which types of network are lost through purifying selection[1]. Systematic biases in the loss of partic-

[1]Here, purifying selection refers to selection for phenotypic stability, as defined in Section 2.3.6.

ular network configurations could allow network features associated with compensatory mutation to accumulate in the population, even when the features do not confer differential reproductive success. In addition, the combination of recombination, deleterious mutation and compensatory mutation under moderately effective population sizes could then permit the evolution of increased regulatory complexity.

The previous literature supports this hypothesis. The theory indicates that compensatory mutation is not likely to play an important role in adaptation, because mutations that simply restore fitness to the mean of the population have the same low probability of fixation as any neutral allele under drift (Wright, 1931b,a; Stephan, 1996; Parsch et al., 1997; Whitlock and Otto, 1999; Whitlock et al., 2003; Zhang and Watson, 2009). Similarly, molecular evidence indicates that although they are associated with rapid divergence, compensatory mutations do not alter how proteins function (Povolotskaya and Kondrashov, 2010). Thus, there is likely to be a non-adaptive source of compensatory mutations to explain the growing biophysical and molecular evidence for their existence (Kimura, 1985; Moore et al., 2000; Levin et al., 2000; Choi et al., 2005; Meer et al., 2010; Kulathinal et al., 2004; Piskol and Stephan, 2008; Covert et al., 2013; Wang et al., 2014b; Tedbury et al., 2015).

Relaxed selection is likely to be critical to the frequency of compensatory mutation. When selection against deleterious mutation is relaxed, the frequency of compensatory mutation in organisms carrying deleterious mutations is surprisingly high (Maisnier-Patin et al., 2002; Gifford and MacLean, 2013). A recent empirical study by Sloan et al. (2014) has also suggested that relaxed selection facilitates compensatory mutation. Specifically, the authors investigated the patterns of molecular evolution in genes expressed in cytosolic, plastid and mitochondrial ribosomes in two different types of *Silene* species. In the paper, Sloan et al. found that *Silene* species with fast-evolving plastid and mitochondrial DNA exhibited increased amino acid sequence divergence in organelle genomes but not in cytosolic ribosomes. Moreover, Sloan et al. found no evidence that the observed pattern was driven by positive selection. They therefore concluded that rapid organelle genome evolution has selected for compensatory mutations in nuclear-encoded proteins.

It has also already been established that drift and weak effect compensatory mutations can shape non-adaptive processes that govern regulatory evolution (Lynch and Abegg, 2010; Payne and Wagner, 2015). However, the critical unresolved problem is the mechanisms by which compensation for lethal mutations contributes to biased purifying selection. Previous studies have suspected that key features of non-adaptive mechanisms are, for example, gene duplication and degeneration (Force et al., 1999), deletion bias (Hare et al., 2008) and biased gene conversion (Smith and Eyre-Walker, 2001). Perhaps, non-adaptive processes could simply contribute to gene regulatory complexity. In particular, compensatory mutation could potentially drive gene regulatory complexity

by generating biases in purifying selection. If networks with compensatory mutations exhibit co-localised mutations or have greater effect sizes than neutrally evolving networks, then it is plausible that they could alter network robustness and connectivity. Also, if this happens frequently over long time scales, then gene regulatory complexity could evolve via biased compensatory mutation through a similar non-adaptive process such as biased gene conversion. However, to date there is no available genomic dataset to test for what biases can create gene regulatory complexity. Moreover, it has been difficult to distinguish between adaptive and non-adaptive processes from biological data due to, for example, genetic linkage. One way in which we could potentially test this hypothesis is through an *in silico* network modelling approach.

Many previous computational studies have focused on the evolution of gene regulatory networks under constant selection (Azevedo et al., 2006; Ciliberti et al., 2007a; Crombach and Hogeweg, 2008; Tsuda and Kawata, 2010; Cotterell and Sharpe, 2013). However, as discussed in Chapter 3, constant selection necessarily constrains pathway evolution because it removes the low-fitness individuals who carry incapacitated gene networks. This in turn eliminates the potentially significant mechanism of compensatory mutation. Compensatory mutation is impossible under one of the dominant modelling frameworks, where unstable networks — networks whose phenotype never reach an equilibrium state — are always labelled as 'unviable' and therefore never subjected to further rounds of mutation (Wagner, 1996; Siegal and Bergman, 2002; Azevedo et al., 2006; Lohaus et al., 2010). However, other previous studies show complex dynamics can be observed if we allow multiple different types of mutations to occur simultaneously (Masel, 2004; Draghi and Wagner, 2009; Fierst, 2010; Misevic et al., 2010; Espinosa-Soto et al., 2011b). In this context, compensatory mutation is possible and able to allow lineages access to a greater variety of evolutionary pathways.

In this chapter, I present the first demonstration that compensatory mutation could contribute to the evolution of regulatory complexity. I find this to occur even in the absence of conventional adaptive selective forces using the evolutionary framework provided by gene regulatory network theory (Wagner, 1996; Siegal and Bergman, 2002; Azevedo et al., 2006). By only including purifying selection (here, selection for phenotypic stability), I can eliminate any possibility of conventional directional selection in terms of providing individuals with reproductive advantages relative to a specific environment. I first show that compensatory mutation can occur regardless of patterns of selection. I then find that purifying selection generates biased compensatory mutations that consequently form networks with a biased distribution of robustness in terms of location and mutation size. Finally, I show that compensatory mutation can play an important role in facilitating regulatory complexity without adaptive responses to directional selection. These findings are important because they provide an explanation of how major features of genome organisation, development and biodiversity can

emerge through non-adaptive processes.

4.2 Methods

The modelling approach was similar to that described in Section 3.2. The system-level parameters were fixed to be $a = 100$, $devT = 100$ and $\tau = 10$ in all simulations. By adopting such a network modelling approach, I was able to investigate how compensatory mutations could drive the formation of regulatory complexity through the incorporation of non-adaptive processes. Here, the non-adaptive portions of the process were the periods of the model where I relaxed purifying selection to tolerate deleterious mutations. This is by definition for compensatory mutation to occur. In the model set-up, I randomly selected individual networks with an equal probability in terms of reproductive success. The population was only subjected to periods of purifying selection, i.e., selection for phenotypic stability. In other words, the selection for the target (optimal) phenotype was not included in all simulations, to prevent any adaptive response being added to the results.

4.2.1 The computational model

The computational model (see Figure 4-1) was similar to that described in Section 3.2.1. However, instead of only looking at the characteristics of compensatory mutation that I showed in Chapter 3, in this chapter, I further examine their evolutionary consequences. Specifically, in the computational model, I started with a collection of randomly generated stable networks (see Figure 4-1 A). If the population was subjected to relaxed selection, both stable networks (solid edges, Figure 4-1 B) with neutral mutations (marked in green) and unstable networks (dashed edges, Figure 4-1 B) with deleterious mutations (marked in red) were allowed to stay in the population pool. Otherwise, if the population was subjected to strong purifying (phenotypic stability) selection, only networks that were able to either retain network stability by neutral mutations (marked in green, Figure 4-1 C) or restore the network stability by compensatory mutations (marked in blue, Figure 4-1 C) were allowed to stay in the population pool. In other words, those compromised networks with deleterious mutations (marked in red, Figure 4-1 B) that could not be restored by additional mutations (compensatory mutations) would be wiped out immediately from the population pool.

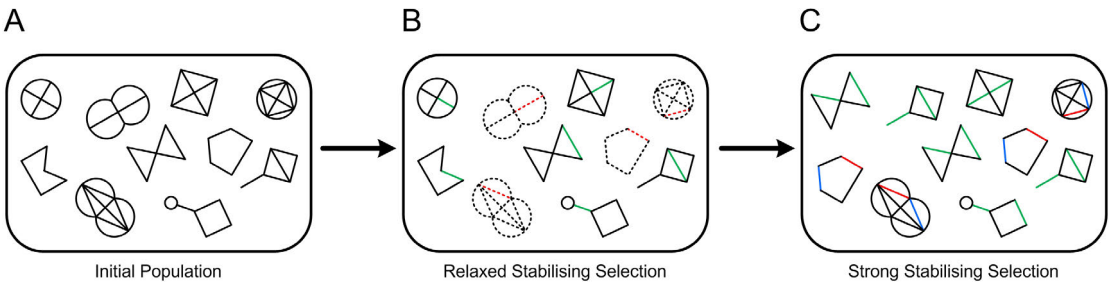


Figure 4-1: Overview of the computational model for testing compensatory mutation in generating regulatory complexity. *During evolution, the initial population (**A**) was either subjected to relaxed selection for phenotypic stability (**B**) or strong selection for phenotypic stability (**C**). Note that a green edge indicates a neutral mutation by which the network can retain its stability, a red edge indicates a deleterious mutation by which the network stability is lost, and a blue edge indicates a compensatory mutation by which the network can restore its stability. Dashed edges represent networks that are not able to reach an equilibrium state.*

Initialisation

The initialisation process was the same as described in Section 3.2.1.

Mutation

The mutation operator was the same as described in Section 3.2.1.

Recombination

In some simulations presented in this chapter, I allowed individual networks to recombine with each other. A recombinant was produced by picking two individuals and selecting rows of the W matrices from each parent with an equal probability (see Section 2.3.5). This process is similar to free recombination between units formed by each gene and its *cis*-regulatory elements, but with no recombination within regulatory regions.

Relaxed and strong selection for phenotypic stability

As illustrated in Figure 4-1, when the population was evolved under a relaxed selection regime, both unstable and stable networks were able to survive in the next generation, whereas compromised networks would be wiped out immediately from the population pool if they were evolved under a strong selection for phenotypic stability regime.

Evolution

The evolutionary simulations were performed under the reproduction-mutation-selection life cycle. The population size M was fixed in every generation throughout the evolution in all simulations. In typical asexual evolution, an individual was chosen at random to reproduce asexually by cloning itself and was then subjected to a single mutation. Similarly, in typical sexual evolution, two individuals were chosen at random to reproduce sexually by recombining two parent networks and then subjected to a single mutation. Depending on different patterns of selection, unstable networks were excluded (under the strong selection for phenotypic stability regime) or allowed to stay in the population (under the relaxed selection for phenotypic stability regime). This process was repeated until M number of networks were produced.

4.2.2 Exploring strong and relaxed selection for phenotypic stability on compensatory mutation frequency

In this set of experiments, I investigated the frequency of compensatory mutation after many generations of both strong and relaxed selection for phenotypic stability to test whether compensatory mutation continues to occur even after lengthy evolution (see Figures 4-2 and 4-3). Specifically, under the strong selection for phenotypic stability regime, I collected $10,000$ stable networks at each generation where each network in the population was subjected to one single mutation. Then, I performed another round of mutation, focusing on the unstable networks that resulted from the previous round, and measured the probability of a second mutation that could restore the network stability of those compromised networks. Similarly, under the relaxed selection for phenotypic stability regime, I collected $10,000$ networks at each generation where each network in the population was subjected to one single mutation. However, for each relaxed selection generation, there were both stable and unstable networks when the population was subjected to one single mutation, since I did not restrict for networks being stable. For those stable networks, I measured the frequency of compensatory mutation in a similar way to that mentioned above in the strong selection for phenotypic stability regime, whereas for unstable networks, I just performed another round of mutation, and measured the probability of a second mutation that could restore network stability. The overall frequency of compensatory mutation for the population during each relaxed selection generation was averaged over the results of stable networks and unstable networks that were calculated separately.

4.2.3 Exploring population diversity for highly stable networks

In this set of experiments, I investigated whether the population diversity would be highly reduced in networks that have been exposed to many generations of selection

for phenotypic stability (see Figure C-1). Specifically, I tested whether the increased compensatory mutation frequency shown in Figure 4-2 was due to the property of particular networks that had been selected for, or whether it was the property of a diverse population. Following the measurement used in Azevedo et al. (2006), the genetic diversity is defined as:

$$H = 1 - \sum_{i=1}^{n} p_i^2, \tag{4.1}$$

where n is the total number of alleles, i.e., the unique values contained in the same site crossing all individual networks, and p_i is the frequency of allele i. The genetic variation in a population is calculated as the mean gene diversity over non-zero sites of the adjacency matrix for a given genotype W. Note that when the total number of unique alleles is large, the diversity fast approaches to 1.

4.2.4 Exploring the frequency of compensatory mutation in seriously damaged networks

In this set of experiments, I measured the frequency of compensatory mutation among unstable networks during each relaxed selection event to further confirm that compensatory mutation can occur even in seriously damaged networks (see Figure 4-4). Specifically, I collected $10,000$ unstable networks at each generation where each network in the population was subjected to one single mutation, so really in this case I had selected against network stability. Then, I performed another round of mutations and measured the probability of a second mutation that could restore network stability. Note that this set of experiments is similar to those experiments described in Section 4.2.2, but here I only focused on unstable networks, whereas I considered both stable and unstable networks in the relaxed selection for phenotypic stability regime in Section 4.2.2.

4.2.5 Exploring the frequency of relaxed selection for phenotypic stability in stimulating compensatory mutations

In this set of experiments, I tested whether frequent relaxed selection for phenotypic stability can generate more compensatory mutations (see Figure 4-5). Specifically, I collected a population pool of $10,000$ stable networks that were generated randomly. The initial population was then evolved under a relaxed selection for phenotypic stability regime with a frequency of 1/2, 1/5, 1/10, 1/25, 1/100, 1/200 and 1/500 for a total of $1,000$ generations. Note that during a relaxed selection event, both stable and unstable networks could appear when the population was subjected to one single round of mutation. The number of compensatory mutations was recorded immediately after each relaxed selection event (the population was exposed to strong selection for

phenotypic stability) when the population was subjected to another single round of mutation. The reported results are the total (see Figure 4-5) and mean frequency of compensatory mutations (per relaxed selection event, see Figure C-2) arising over $1,000$ generations.

4.2.6 Exploring the impact of distance and size effects on network robustness

In this set of experiments, I explored the effects of location and mutation size on robustness in networks with one deleterious mutation and one compensatory mutation and in networks with two consecutive neutral mutations, to investigate whether networks with compensatory mutations have a different evolutionary consequence compared with networks with neutral mutations (see Figures 4-7 and 4-8).

Specifically, to test the distance effect, I collected $10,000$ sample networks at each distance (between deleterious mutation and compensatory mutation). Then, for each category of distance, I measured the proportion of stable networks after one additional round of single mutation. The reported results are both actual robustness (see the solid line in Figure 4-7 A) and percentage change in robustness (see the solid line in Figure 4-7 B). Similarly, for the control group, instead of collecting networks that were subjected to one deleterious mutation and one subsequent compensatory mutation, I collected $10,000$ sample networks that were subjected to two consecutive neutral mutations at each distance (between two neutral mutations), and then assessed the actual robustness (see the dashed line in Figure 4-7 A) as well as the percentage of robustness change (see the dashed line in Figure 4-7 B). I also performed similar experiments for medium ($N = 20$) and large networks ($N = 40$), as shown in Figures C-3 and C-4.

Likewise, to test size effect, I collected $10,000$ sample networks that were compensated by mutations with different shifts in gene regulation. Then, for each category of mutation size, I measured the proportion of stable networks after one additional round of single mutation. The reported results are both actual robustness (see the solid line in Figure 4-8 A) and percentage change in robustness (see the solid line in Figure 4-8 B). Similarly, for the control group, instead of collecting networks that were subjected to one normal deleterious mutation and one subsequent compensatory mutation with different shifts in gene regulation, I collected $10,000$ sample networks that were subjected to two consecutive neutral mutations, one normal neutral mutation and the other neutral mutation with different shifts in gene regulation, and then assessed the actual robustness (see the dashed line in Figure 4-8 A) as well as the percentage of robustness change (see the dashed line in Figure 4-8 B). I also performed similar experiments for medium ($N = 20$) and large networks ($N = 40$), as shown in Figures C-5 and C-6.

4.2.7 Exploring how network connectivity evolves under a relaxed selection regime

In this set of experiments, I investigated whether regulatory complexity (increased network connectivity) could arise under a relaxed selection for phenotypic stability regime where compensatory mutations could occur and accumulate (see Figures 4-9 and 4-10).

In the first set of experiments, I tested whether we could observe greater complexity arising using a population pool of $10,000$ stable networks of $N = 10$ genes with a simple 'Star' topology (see Figure 4-9). Specifically, the initial population pool was generated using the following rules:

- Randomly select a gene to be the hub node.
- There is at least one edge between the hub node and non-hub nodes (either inward or outward); there is a possibility (0.5) of having both inward and outward edges.
- Each node has a possibility (0.5) of having a self-regulatory edge (including the hub node).
- The value (interaction strength) of each edge is drawn from the standard normal distribution $N(0, 1)$.

In theory, for network size $N = 10$, the minimum connectivity is $c_{\min} = 0.09$ (9 edges) and the maximum connectivity is $c_{\max} = 0.28$ (28 edges). In the randomly generated initial population pool used in this chapter, the minimum connectivity was $c_{\min} = 0.10$ (10 edges), the maximum connectivity was $c_{\max} = 0.26$ (26 edges), the median connectivity was $\tilde{c} = 0.17$ (17 edges) and the average connectivity was $\bar{c} \approx 0.17$. Then, the initial population was evolved for $5,000$ generations under strong and relaxed selection for phenotypic stability regimes: In four scenarios with strong selection for phenotypic stability, the initial population was evolved under a no mutation and no recombination regime, a mutation but no recombination regime, a recombination but no mutation regime, a mutation and recombination; in three other scenarios, the initial population was evolved under a relaxed selection for phenotypic stability regime with a frequency of 1/10, 1/25, and 1/50. The statistical details for connectivity in initial and evolved populations can be found in Table C.1. Note that compensatory mutation could only occur during each relaxed selection event.

In order to make a stronger argument that relaxed selection can facilitate regulatory complexity, in the second set of experiments, I further investigated how network connectivity evolves under a relaxed selection regime using randomly generated networks (see Figure 4-10). Specifically, for a network size $N = 40$ with connectivity $c = 0.15$, I collected $10,000$ stable networks, each of which had the same initial gene expression

pattern, all activation, i.e., $\mathbf{s}(0) = (+1, +1, \ldots, +1)$. This population was then evolved for $5,000$ generations, in this case allowing for recombination with other individuals from the same generation. Note that in the previously described experiments in this chapter, a mutation could not change the topology of an individual network; that is, it could not change zero elements into non-zero or *vice versa*. In contrast, recombination can alter the topology if the non-zero sites are different in individual networks. In addition, I further performed an additional simulation that served as the control group to investigate how network connectivity evolves when two layers of selection (selection for phenotypic stability and target phenotype) are absent (see Figure C-7). The reported results are the mean network connectivity of all individuals in the population in every 200 generations under different frequencies of relaxed selection. Note that network connectivity was measured in the next generation of selection for phenotypic stability immediately after the previous relaxed selection; therefore, I only report the results in stable networks[2].

4.2.8 Exploring the effect of selection for phenotypic stability on network connectivity

In this set of experiments, I performed Price equation (Price, 1970) analysis to investigate the effect of selection for phenotypic stability on network connectivity (see Figures 4-14). Specifically, I employed the same population pool as described in Section4.2.7 — $10,000$ stable networks ($N = 40$ and $c = 0.15$), each of which had the same initial gene expression pattern, all activation, i.e., $\mathbf{s}(0) = (+1, +1, \ldots, +1)$. Then, the population was evolved for $5,000$ generations under a relaxed selection for phenotypic stability regime with a frequency of $1/50$. The population at the end of evolution was saved for Price equation analysis. I measured the network connectivity as the trait value for each individual, and assessed its robustness as reproductive success. Note that to assess the robustness of each individual network, I performed 100 perturbation tests to record the probability that the network remained stable after a single round of mutation. Here, a single mutation means exactly one non-zero entry in an individual's genotype would be mutated. The scatter plot between the network connectivity (trait vale) and robustness (reproductive success) was reported.

4.2.9 Exploring the impact of repeated compensatory mutations on network robustness

In this set of experiments, I measured the robustness of networks with compensatory mutations and networks with neutral mutations (see Figures 4-12 and 4-13). Specifically, I collected $10,000$ sample stable networks which had been exposed to one

[2]For the additional simulation when two layers of selection are absent, I only measure the network connectivity for stable networks.

to five cycles with compensatory mutation (S → U → S). For comparison analysis, I also collected 10,000 sample stable networks which had been exposed to one to five cycles without compensatory mutation (S → S → S); that is, that had been through the same number of rounds of mutation but had never become unstable until the final round before testing. For both of these populations, I then measured the robustness. Note that I only focused on selecting stable networks in the 'S' round, unstable networks in the 'U' round, and '→' means the population is subjected to one round of single mutation.

4.3 Results

Using the well-established synthetic Wagner model of gene regulatory networks described in Section 4.2.1, I was able to uncover how purifying selection generates biased compensatory mutations that consequently drive regulatory complexity through non-adaptive processes. An overview of the computational model can be found in Figure 4-1.

4.3.1 Compensatory mutation can occur regardless of different patterns of selection

In Section 3.3.1, I showed that new mutations can restore network stability in 5–15% of low-fitness lineages in the initial population. In this chapter, I further investigated whether compensatory mutations would be expected to be able to occur in a population that had been exposed to bouts of generations of relaxed and strong selection for phenotypic stability. I found that compensatory mutation occurs in both evolutionary scenarios.

From Figure 4-2, we can see that compensatory mutation is able to occur even in highly stable networks that have been subjected to strong selection for phenotypic stability for many generations. In addition, the compensation probability tends to be constant after many rounds of mutation. Furthermore, I found that, across network sizes, all populations still maintain a high diversity in the presence of strong selection for phenotypic stability, and for many generations (see Figure C-1).

It is not surprising therefore to see that, as shown in Figure 4-3, compensatory mutation can occur in the mixed populations (stable and unstable networks) that result from a relaxed selection for phenotypic stability regime, although it is less pronounced there and declines significantly over rounds of selection. Interestingly, I found that compensatory mutation can still fix seriously damaged networks, if we only select for those broken networks at each mutation round, as shown in Figure 4-4, where compensatory mutations restore, for example, about 14% of networks for $N = 5$ that are broken by one round of mutation, but the frequency quickly drops to compensatory

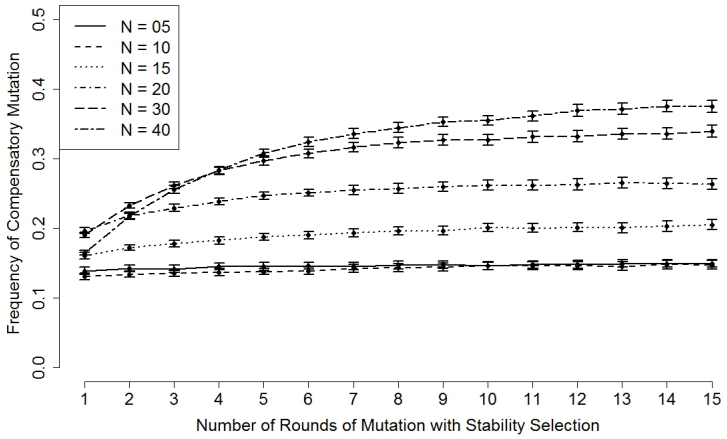


Figure 4-2: The frequency of compensatory mutation in networks that have been subjected to bouts of strong selection for phenotypic stability. *For each network size (N = 5, 10, 15, 20, 30 and 40) with network connectivity $c = 0.76$, I collected $10,000$ stable networks with one to fifteen rounds of mutation. For each round of mutation, each network was subjected to one single mutation. Then, I measured the frequency of compensatory mutation in each set of collected networks. The error bars represent 95% confidence intervals based on 100 independent runs.*

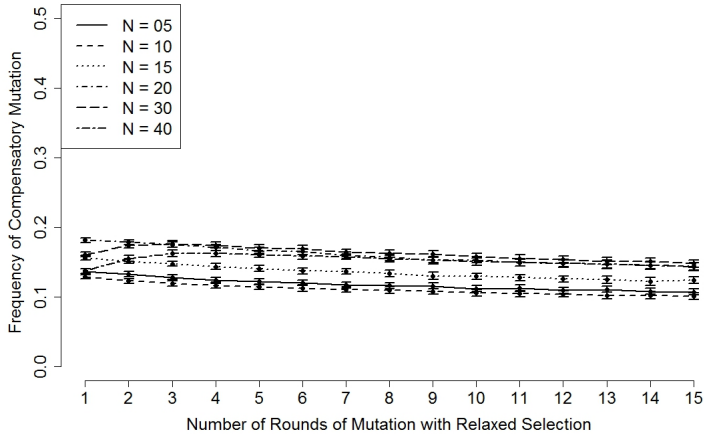


Figure 4-3: The frequency of compensatory mutation in networks that have been subjected to bouts of relaxed selection for phenotypic stability. *For each network size (N = 5, 10, 15, 20, 30 and 40) with network connectivity $c = 0.76$, I collected $10,000$ networks (both stable and unstable) with one to fifteen rounds of mutation. For each round of mutation, each network was subjected to one single mutation. Then, I measured the frequency of compensatory mutation in each set of collected networks. The overall frequency of compensatory mutation for the population during each relaxed selection generation was averaged over the results for stable and unstable networks, which were calculated separately. The error bars represent 95% confidence intervals based on 100 independent runs.*

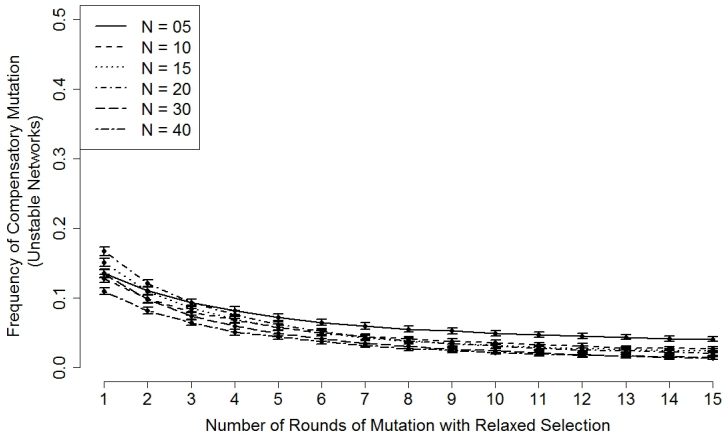


Figure 4-4: The frequency of compensatory mutation in networks with cumulative deleterious mutations. *For each network size (N = 5, 10, 15, 20, 30 and 40) with network connectivity $c = 0.76$, I collected* $10,000$ *unstable networks with one to fifteen rounds of mutation. For each round of mutation, each network was subjected to one single mutation. Then, I measured the frequency of compensatory mutation in each set of collected networks. The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

mutations being able to restore the stability of 5% broken networks that have had many deleterious mutations up to 15 generations. This means that compensatory mutations are cure-alls even for seriously damaged networks.

4.3.2 Relaxed selection stimulates compensatory mutations

Next, I investigated the possibility that relaxed selection can stimulate compensatory mutations. I found that, as expected, we can observe more compensatory mutations in the presence of relaxed selection for phenotypic stability. Specifically, I performed simulations to measure the number of compensatory mutations in which the relaxed selection occurred in different frequencies. From Figure 4-5 (also see Figure C-2), we can clearly see that the number of compensatory mutations increases as the consequence of having more generations of relaxed selection. We can also see that smaller networks typically have more compensatory mutations compared with larger networks. This is because compromised networks with smaller sizes are more likely to experience compensation after lengthy evolution, although larger networks tend to have a higher frequency of compensatory mutation at early stages as indicated in Figure 3-3.

From the simulation results, we can speculate that genes with compensatory mutations are more likely to be those that have experienced periods of relaxed selection. In fact, this prediction is consistent with the empirical evidence from one recently published work by Sloan et al. (2014). I used Sloan et al.'s data to plot the d_N/d_S ratio in cytosolic ribosomes and organelle genomes of the two studied *Silene* species, as shown

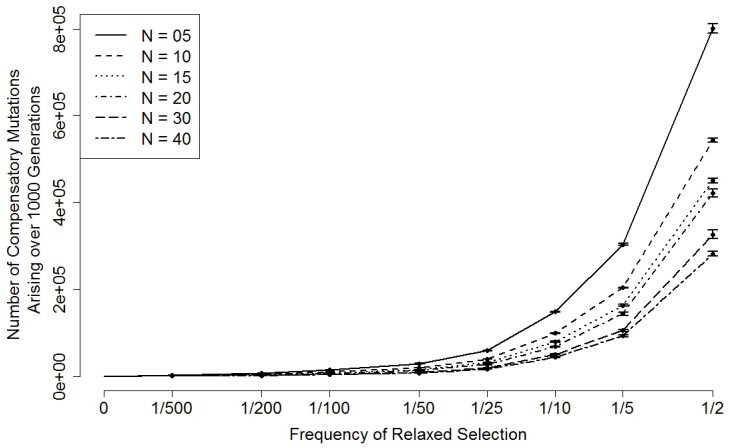


Figure 4-5: Total number of compensatory mutations occurring in each relaxed selection event. *For each network size (N = 5, 15, 10, 20, 30 and 40) with connectivity $c = 0.76$, I measured the number of compensatory mutations occurring after the previous relaxed selection for phenotypic stability, which happened in every* 2, 5, 10, 25, 50, 100, 200 *and* 500 *generations. The reported results are the total number of compensatory mutations occurring over a total of* $1,000$ *generations for populations with different network sizes. The error bars represent* 95% *confidence intervals based on* 10 *independent runs.*

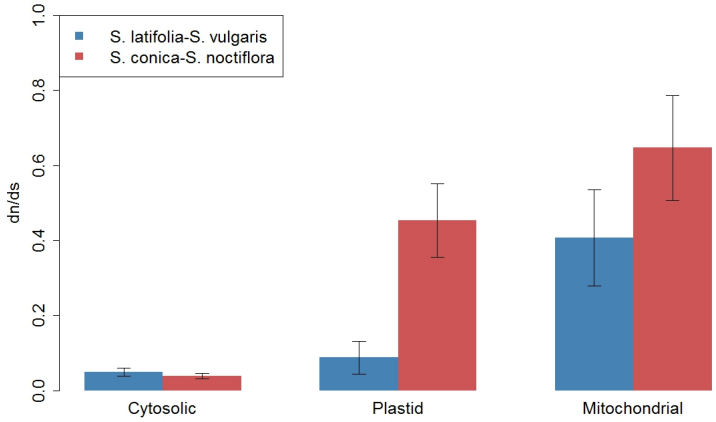


Figure 4-6: Compensatory mutations facilitate rapid organelle genome evolution in two *Silene* species. *The reported results are the d_N/d_S ratios of amino acid sequence divergence in cytosolic ribosomes and organelle genomes of the two different Silene species studied by Sloan et al. (2014). The error bars represent* 95% *confidence intervals.*

in Figure 4-6. Note that d_N/d_S ratio is an indicator of selective pressure acting on a protein-coding gene[3]. From Figure 4-6, we can clearly see that cytosolic ribosomes

[3]In the same given period of time, the ratio is calculated as the ratio of the number of non-synonymous substitutions per non-synonymous site (d_N) to the number of synonymous substitutions per synonymous site (d_S). Homologous genes with a d_N/d_S ratio above 1 are evolving under positive

show a slower rate of evolution. This confirms to the predictions of the model, given that cytosolic ribosomes do not experience many periods of relaxed selection. We would expect plastid and mitochondrial ribosomes to exhibit a much more rapid evolution due to the substantial periods of relaxed selection they have been exposed to, a prediction also supported by this chapter.

4.3.3 The robustness of networks with compensatory mutations exhibits bias in location

In Section 3.3.2, I showed that compensatory mutations are more likely to occur at or close to the site of the original, deleterious mutation. In this chapter, I further investigated the evolutionary consequence, i.e., robustness, of this location effect. I found that patterns of localisation-generating robustness are quite different. Specifically, I compared the robustness of stable networks following one round of deleterious and compensatory mutation with that of stable networks with two consecutive neutral mutations, as shown in Figure 4-7. In general, robustness is far higher when compensatory mutation occurs closer to the original deleterious mutation site (see the solid line in Figure 4-7 A), whereas after two neutral mutations, closer distances are not better associated with higher robustness (see the dashed line in Figure 4-7 A). By measuring the percentage change in robustness (see Figure 4-7 B), we can also see that compensatory mutations generate a profound increase in robustness. It should be noted that although networks with compensatory mutations exhibit a more profound biased change in robustness with respect to location, their actual robustness is much lower than that of networks with neutral mutations (see Figure 4-7 A). Similar patterns are also observed in networks with different size and connectivity (see supporting information in Appendix C). These theoretical results indicate that these co-localised compensatory mutations are more likely to be accumulated, whereas compensatory mutations that are far apart from the previous deleterious mutations are more likely to be lost, by subsequent purifying selection.

selection, indicating that some of the mutations concerned must be advantageous, whereas the ratio will be in the range 0 to 1 if all the mutations are neutral or disadvantageous.

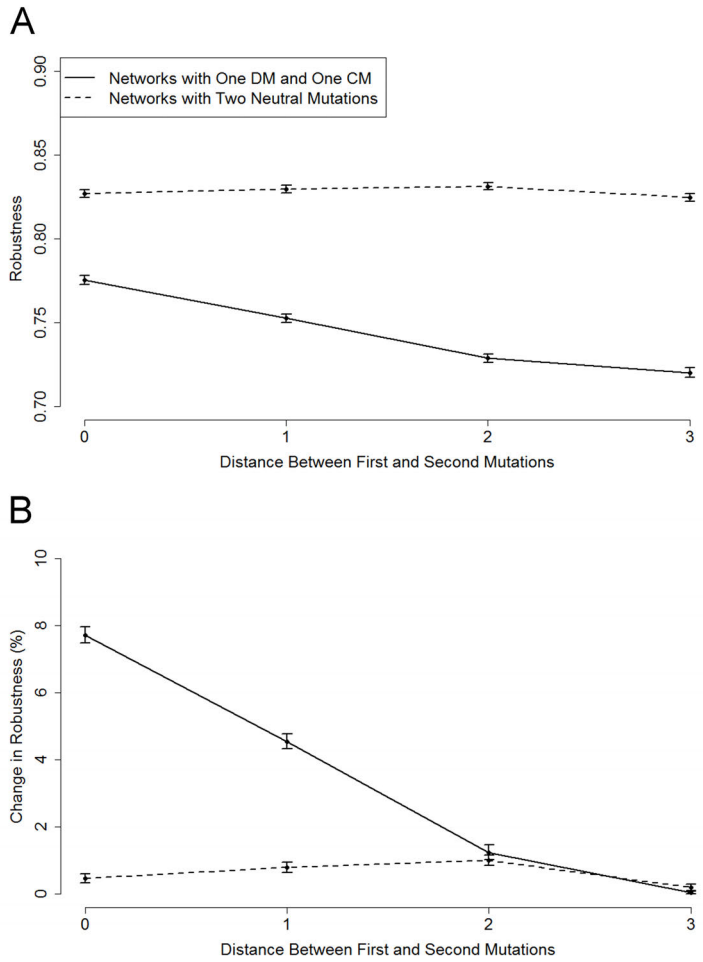


Figure 4-7: The impact of distance effect on network robustness. *For* $N = 5$ *and* $c = 0.4$*, I collected* $10,000$ *sample stable networks that were subjected one deleterious mutation and then restored by one subsequent compensatory mutation that was* 0*,* 1*,* 2 *and* 3 *steps away from the previous deleterious mutation. The sample networks for the control group were collected in a similar way, except that the networks were subjected to two consecutive neutral mutations. Then, I assessed the robustness of the sample networks at each distance step. The reported results are actual robustness (**A**), and change in robustness (**B**) (the actual robustness was normalised by subtracting the minimal value among all categories, and then dividing by the minimal value). The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

4.3.4 The robustness of networks with compensatory mutations exhibits bias in mutation size

In Section 3.3.3, I showed that compensatory mutations are more likely to be caused by mutations leading to larger shifts in gene regulation. In this chapter, I further investigate the evolutionary consequence, i.e., robustness, of this mutation size effect. I found that patterns of shifting regulation-generating robustness are also quite different.

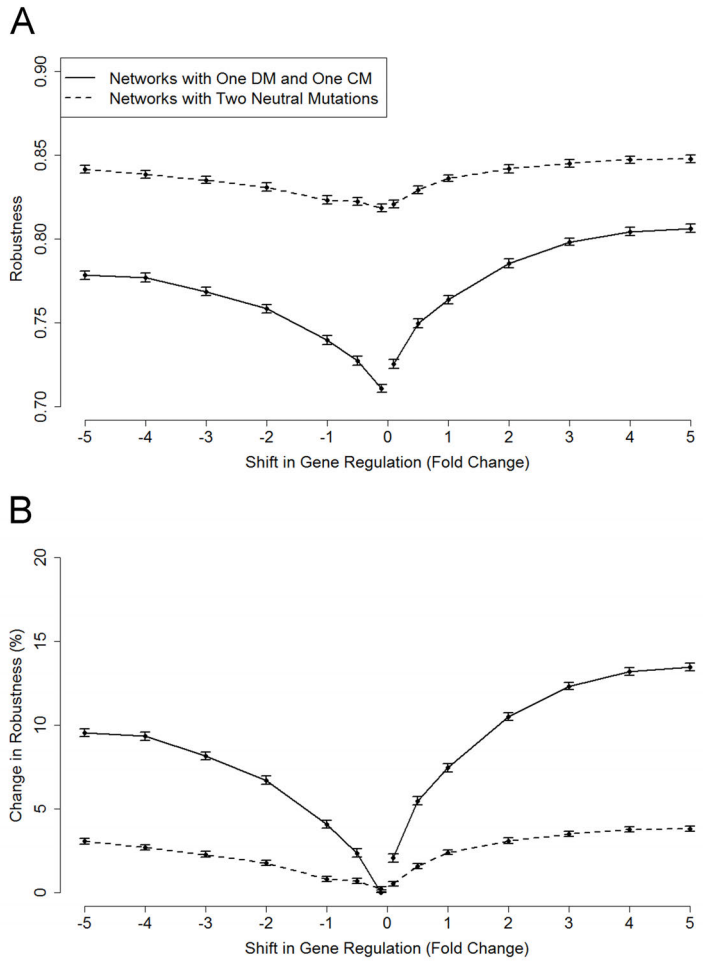


Figure 4-8: The impact of mutation size effect on network robustness. *For small networks ($N = 5, c = 0.4$), I collected* $10,000$ *sample stable networks that were subjected to one deleterious mutation and then restored by one subsequent compensatory mutation with different shifts in gene regulation from* $[-5, +5]$ *(step size* 1 *and with four additional regulation shifts:* -0.5, -0.1, 0.1 *and* 0.5*). The sample networks for the control group were collected in a similar way, except that the networks were subjected to two consecutive neutral mutations. Note that the second neutral mutation had different shifts in gene regulation to the compensatory mutation. Then, I assessed the robustness of the sample networks at each category. The reported results are actual robustness (**A**), and change in robustness (**B**) (the actual robustness was normalised by subtracting the minimal value among all categories, and then dividing by the minimal value). The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

Specifically, I compared the robustness of stable networks having one deleterious mutation and compensatory mutation with that of stable networks having two consecutive neutral mutations, as shown in Figure 4-8. In general, the robustness is higher when compensatory mutation has a larger shift in gene regulation (see the solid line in Figure 4-8 A). Although networks with neutral mutations tend to have a similar pattern (see the dashed line in Figure 4-8 A), by measuring the percentage change in robustness

(see Figure 4-8 B), we can clearly see that compensatory mutations generate a much greater increase in robustness. Again, it should also be noted that although networks with compensatory mutations exhibit a more profound biased change in robustness with respect to mutation size, their actual robustness is much lower than that of networks with neutral mutations (see Figure 4-8 A). Similar patterns are also observed in networks with different size and connectivity (see supporting information in Appendix C). These theoretical results indicate that these large-effect compensatory mutations are more likely to be accumulated, whereas small-effect compensatory mutations are more likely to be lost, by subsequent purifying selection.

4.3.5 Compensatory mutation generates regulatory complexity

Looking at the long-term evolutionary consequences of biased compensatory mutations, I might predict that the effects of the two fundamental network properties, location and size, facilitate a biased evolution through non-adaptive processes, or at least during periods of relaxed selection interspersed between bouts of strong purifying (phenotypic stability) selection. I then observed an increase in the complexity of gene regulatory networks, but only in a context where they have been withdrawn from the purifying selection for at least some proportion of generations. Specifically, I first generated a pool of $10,000$ stable networks ($N = 10$) with a simple 'Star' topology (see Figure 4-9 A), then evolved the population under different evolutionary scenarios (see details in Section 4.2.7). Figure 4-9 B shows four evolutionary scenarios where the population is exposed to strong selection for phenotypic stability in every generation such that there is no opportunity for compensatory mutation. From the typical results (networks with a median connectivity), I found that:

1) the median connectivity is the same as the initial population's if it is evolved without mutation or recombination (only by drift),

2) the median connectivity decreases if evolved under either a mutation but no recombination regime or a recombination but no mutation regime (although the network structures are greatly altered when invoking only recombination), and

3) the median connectivity increases to an intermediate level if evolved under a regime allowing both mutation and recombination.

Figure 4-9 C shows these three evolutionary scenarios where the population is evolved with periods of relaxed selection, invoking mutation (including compensatory mutation) and recombination. From these typical and individual results (networks with a median connectivity), we can see that the median connectivity greatly increases and is higher than in the case when the population is subjected exclusively to strong selection for phenotypic stability so that no compensatory mutation can occur.

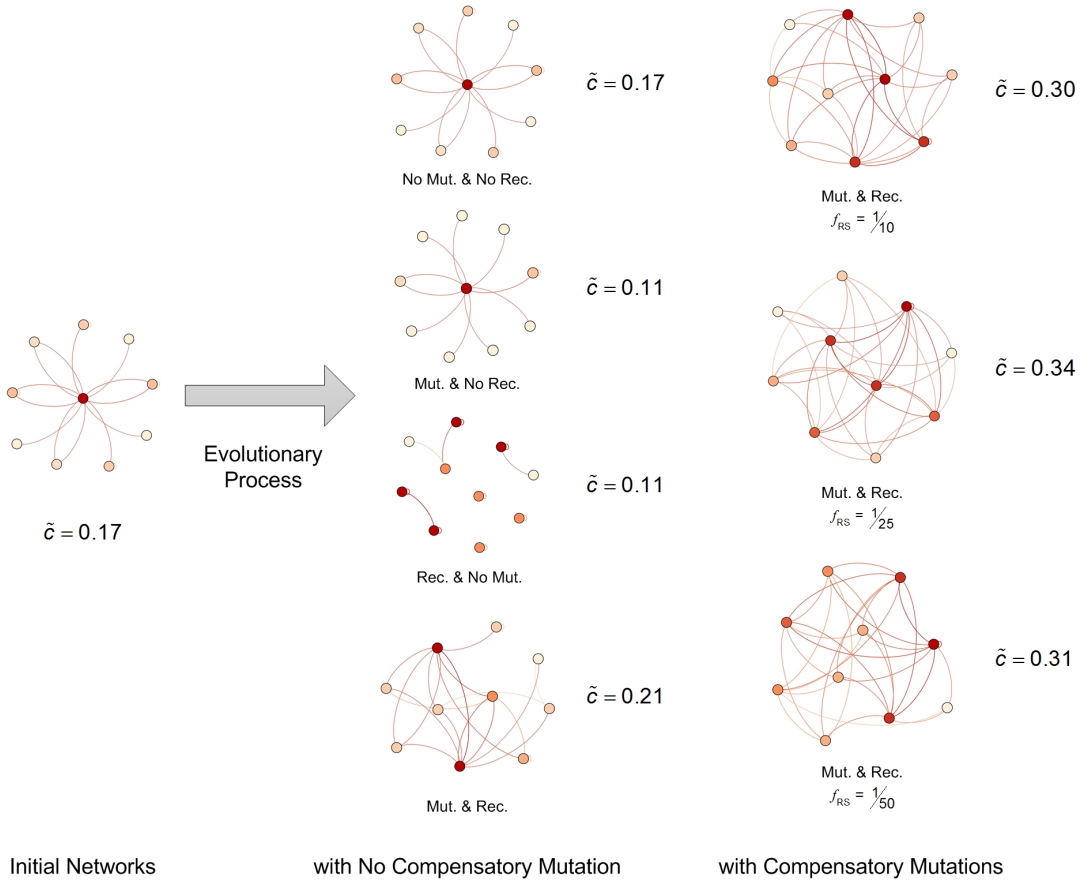


Figure 4-9: Compensatory mutation generates regulatory complexity in stable networks without an initial variation in network structure. *The initial population pool was composed of* $10,000$ *sample stable networks with* $N = 10$ *genes. These networks had a similar 'Star' topology (one hub node and nine non-hub nodes) and varying network connectivity* $[0.10, 0.26]$*. A detailed description of generating the initial population can be found in Section 4.2.7. (**A**) A representative network from the initial population. (**B**) The initial population was evolved for* $5,000$ *generations with strong selection for phenotypic stability (*$f_{RS} = 0$*) under a no mutation and no recombination regime, a mutation but no recombination regime, a recombination but no mutation regime and a mutation and recombination regime. (**C**) The initial population was also evolved for* $5,000$ *generations under a relaxed selection regime with different frequencies* $f_{RS} = 1/10$*,* $1/25$ *and* $1/50$*. Note that compensatory mutation cannot happen when the population is persistently subjected to selection for phenotypic stability, since there would then be no deleterious mutations to compensate. The plotted networks were selected randomly with the median connectivity,* $\tilde{c}$*, in each of the initial or evolved populations. The node's saturation is associated with its inward and outward degree.*

To quantify the impact of relaxed selection, in a separate experiment, I further investigated whether compensatory mutation could drive regulatory complexity in randomly generated networks. Specifically, I collected $10,000$ stable networks and then evolved them for $5,000$ generations, allowing both mutation and recombination. From Figure 4-10, we can see that if there is no relaxed selection at all, the mean connectivity of the population can be highly maintained during evolution, whereas the network

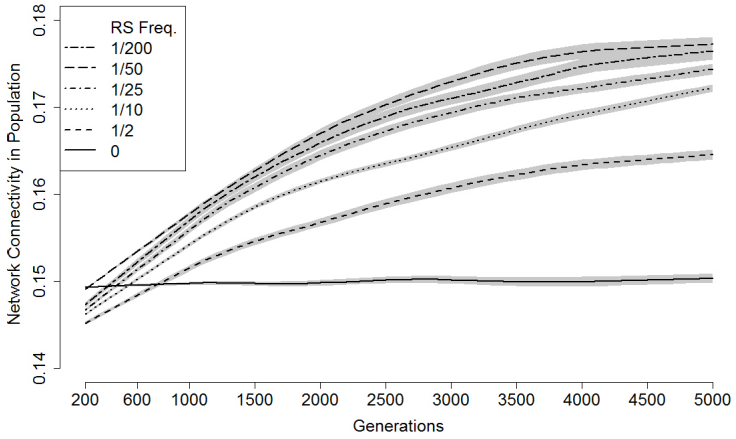


Figure 4-10: Compensatory mutation generates regulatory complexity in stable networks without an initial variation in network connectivity. *For network size* $N = 40$ *and connectivity* $c = 0.15$*, I collected* $10,000$ *stable networks, then evolved them for* $5,000$ *generations, allowing both mutation and recombination at each generation. In every* 200 *generations, I measured the network connectivity of the population (stable) in which the relaxed selection occurred in every* 2*,* 10*,* 25 50 *and* 200 *generations. I also measured the network connectivity of the population when there was no relaxed selection as the control group. The shaded areas represent* 95% *confidence intervals based on* 10 *independent runs.*

connectivity can increase if we allow compensatory mutations to occur in each relaxed selection event. It should be noted that in the first experiment, as shown in Figure 4-9, I fixed the network structure but varied the network connectivity in the initial population, whereas I fixed the network connectivity but varied the network structure in the second experiment (see Section 4.2.7). These results demonstrate that strong selection for phenotypic stability where it impedes deleterious and compensatory mutations constricts complexity, whereas compensatory mutations contribute to regulatory complexity as a part of a non-adaptive process.

4.3.6 Networks with compensatory mutations are evolved through non-adaptive processes

In Sections 4.3.3 and 4.3.4, I showed that compensatory mutation generates biases in location and mutation size, and consequently can drive regulatory complexity. But are networks with compensatory mutations evolved through non-adaptive processes? Here I have performed two sets of simulations to support the argument that compensatory mutations modelled in this chapter are non-adaptive.

First, let us consider some conceptual scenarios of adaptive and non-adaptive processes in the context of protein absorption, as shown in Figure 4-11. Figure 4-11 A shows that a trait (protein absorption) enhances the reproductive success (fitness) of

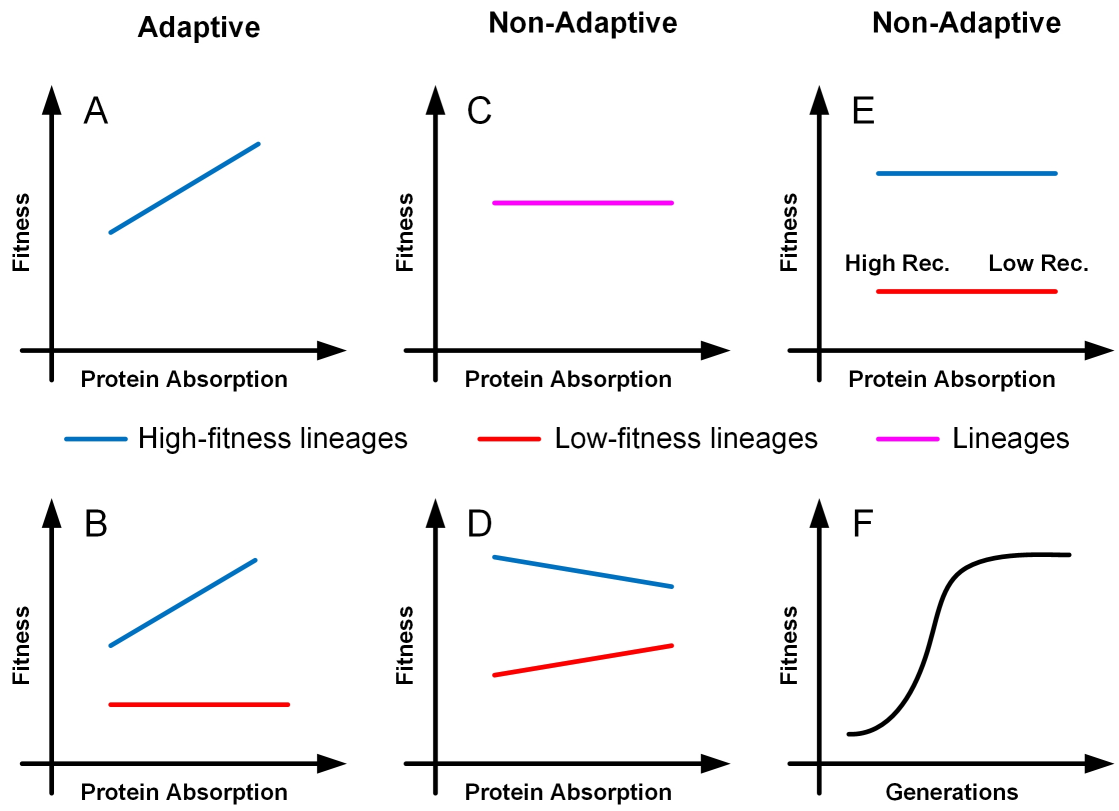


Figure 4-11: Conceptual scenarios of adaptive and non-adaptive processes in the context of protein absorption.

lineages which carry the trait. In such a scenario, we can consider this process to be adaptive because it increases the probability of individuals' own transmission to subsequent generations. The adaptation in this case is particularly strong when the trait engenders high reproductive success in individuals which are the most competitive, as lineages with low fitness are much more inclined to be removed by natural selection, as indicated in Figure 4-11 B. However, non-adaptive evolution can occur when the trait does not increase competitive success, as illustrated in Figure 4-11 C. In this scenario, more protein absorption does not render lineages a higher competitive ability in terms of reproductive success. When we look into these lineages and classify them according to whether they are reproductively competitive or not, as shown in Figure 4-11 D, we may find that the trait of protein absorption is maladaptive because it slightly reduces competitive ability (red line in Figure 4-11 D), and it only helps individuals whose reproductive success is below the mean reproductive success and are likely to be removed by natural selection (blue line in Figure 4-11 D). As a consequence, in subsequent generations, we would expect to see slightly higher protein absorption in each generation because there is a bias in purifying selection — out of the individuals who are more likely to be eliminated by natural selection, those with high protein processing

are slightly less likely to be wiped out. Moreover, the non-adaptive evolution of traits can also occur when the trait does not affect reproductive success, but does affect the rate of underlying evolutionary processes. As can be seen from Figure 4-11 E, protein absorption does not affect the reproductive success of individuals with high competitive ability or low competitive ability. However, if lineages with low protein absorption are nutrient limited, and end up with higher rates of recombination due to the fact that they cannot run DNA repair processes, then protein processing would likely evolve. This is because the ones that do it well can keep gene combinations that work for protein processing, but the ones with poor protein processing shuffle genes around, which is likely to reduce the success of subsequent generations because of recombination costs, until higher protein absorption arises. Therefore, higher protein absorption could evolve with it being adaptive, as illustrated in Figure 4-11 F.

In the first set of simulations, I designed two extreme evolutionary scenarios to test whether networks with compensatory mutation are evolved through non-adaptive processes. I found that generally evolved networks with compensatory mutations have a lower robustness than networks with neutral mutations. Specifically, I conducted experiments to force networks to evolve going through cycles of deleterious and compensatory mutations (S $\rightarrow$ U $\rightarrow$ S) or cycles of two neutral mutations (S $\rightarrow$ S $\rightarrow$ S) where the networks have never been compromised (no compensatory mutation). From Figure 4-12, we can see that although for $N = 5$ and $N = 10$ robustness tends to slightly decrease, whereas robustness tends to slightly increase for $N = 30$ and $N = 40$, generally robustness will largely not be evolved, whereas Figure 4-13 shows that robustness greatly increases in networks that persistently accumulate neutral mutations. Nevertheless, it should be noted that networks with compensatory mutations have a much lower robustness than networks with neutral mutations. This is due to the fact that compensatory mutations always happen in those fragile networks that have been compromised by deleterious mutations. Note that this pattern is also observed in Figures 4-7 A and 4-8 A. Taken together, these results suggest that robustness is generally lower in networks with compensatory mutations than in networks with neutral mutations, and, therefore, are evolved through non-adaptive processes.

In addition, it should be noted that there is no selection for regulatory complexity in all evolutionary scenarios. However, regulatory complexity may be coupled with selection for phenotypic stability, since such selection has been included in the simulations. Therefore, in the second set of simulations, I further applied Price equation analysis to investigate the effect of selection for phenotypic stability on network connectivity. I found that network robustness is not associated with network connectivity. Specifically, I measured the network connectivity of the population that had been evolved for $5,000$ generations under a relaxed selection regime with a frequency of $1/50$, and for each individual network the robustness was assessed based on 100 perturbation. Note

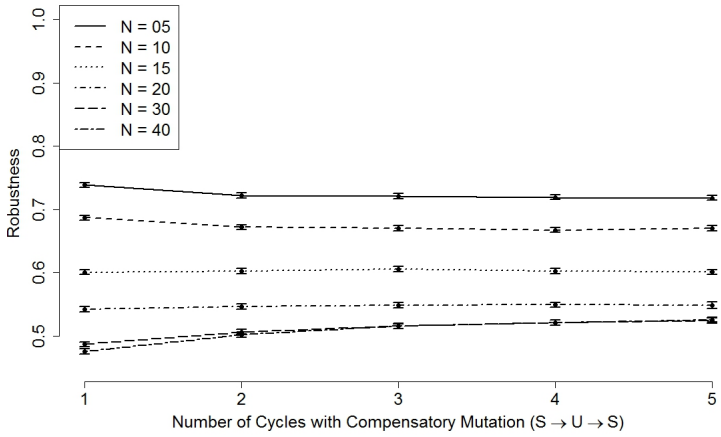


Figure 4-12: Robustness of networks with compensatory mutations. *For each network size ($N = 5$, 10, 15, 20, 30 and 40) with connectivity $c = 0.76$, I measured the robustness based on $10,000$ sample networks that had been through one to five cycles with compensatory mutation ($S \to U \to S$). Note that here a 'cycle' means two mutational steps, i.e., one deleterious mutation and one compensatory mutation. 'S' means only selecting stable networks in this cycle, 'U' means only selecting unstable networks, and '$\to$' means the population is subjected to one round of single mutation. The error bars represent 95% confidence intervals based on 100 independent runs.*

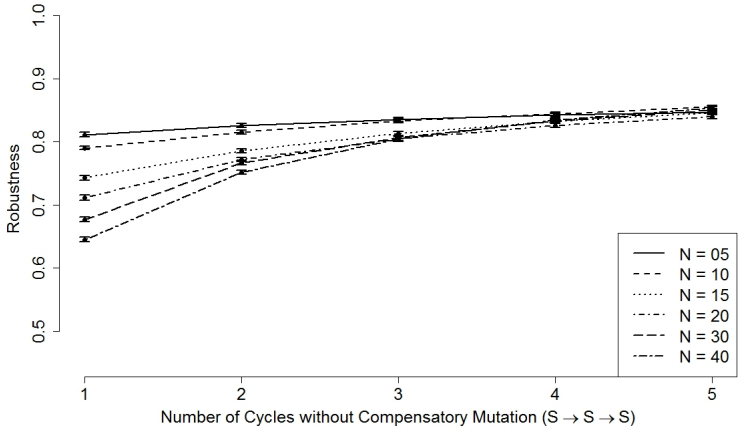


Figure 4-13: Robustness of networks with neutral mutations. *For each network size ($N = 5$, 10, 15, 20, 30 and 40) with connectivity $c = 0.76$, I measured the robustness based on $10,000$ sample networks that had been through one to five cycles without compensatory mutation ($S \to S \to S$). Note that here a 'cycle' means two mutational steps, i.e., two neutral compensatory mutations. 'S' means only selecting stable networks in this cycle, and '$\to$' means the population is subjected to one round of single mutation. The error bars represent 95% confidence intervals based on 100 independent runs.*

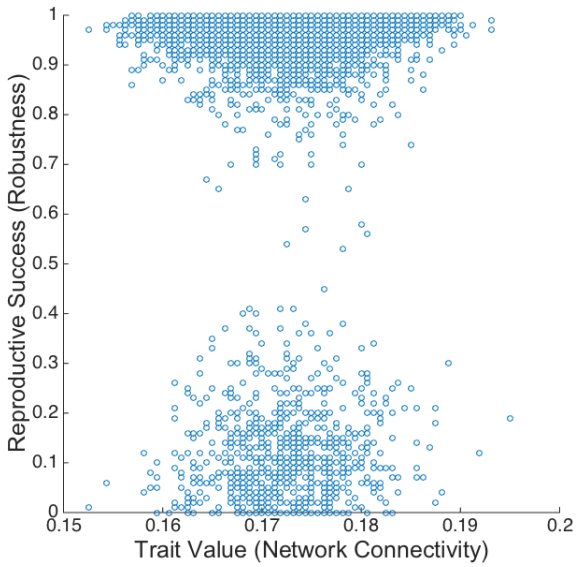


Figure 4-14: Relationship between robustness and network connectivity. *For network size* $N = 40$ *and connectivity* $c = 0.15$*, I collected* $10,000$ *stable networks, then evolved them for* $5,000$ *generations in which relaxed selection occurred in every* 50 *generations, allowing both mutation and recombination at each generation. At the end of evolution, the network connectivity was measured for each individual network. The robustness of each individual network was also assessed based on* 100 *perturbation tests. The reported result is the scatter plot of individuals' network connectivity (trait value) and their corresponding robustness (reproductive success). The slop base on the linear regression analysis is -0.2059 (F-test, df: 9998, p-value: 0.623).*

that in this case, the trait value defined in the Price equation is network connectivity, and the robustness is regarded as the reproductive success described in the Price equation. Then, I performed a linear regression analysis based on the scatter plot as shown in Figure 4-14. The reported slop of the linear regression is -0.2059 (F-test, df: 10000, p-value: 0.623), which indicates that there is no linear relationship between the robustness and network connectivity. Thus, regulatory complexity must be evolved through non-adaptive processes. What drives regulatory complexity may be the influx of very biased sets of low-performing networks with large-effect and closely-linked compensatory mutations.

4.4 Discussion

Compensatory mutations have long been considered the primary means by which low-fitness lineages might be able to be restored to high fitness (Levin et al., 2000; Crawford et al., 2007; Meer et al., 2010). However, the extent of their role has often been considered to be negligible because they were considered to be highly improbable and rare. Therefore, they have not been studied extensively, and many of their general

properties are consequently still unknown. If the results in simulation hold for *in vivo* regulatory networks, then compensation may be far more probable and frequent than has previously been considered. Stable networks may by their nature be surprisingly robust, such that a wide variety of alterations to a compromised network effect recovery.

Unfortunately, interactions in mutation *in vivo* are hard to measure and the results usually have weak statistical significance (West et al., 1998, 1999). In such situations, exploration of theoretical possibilities through simulation offers an ideal means to identify and test for logically coherent scientific hypotheses and to discover unanticipated consequences of these. These unanticipated consequences are predictions arising logically from the hypotheses the model expresses — predictions that can inform our search for evidence *in vivo* (Bryson et al., 2007). The ability to observe and manipulate thousands of individuals' models in a matter of hours allows for a systematic exploration of largely unknown theoretical territory. In this chapter, the extension of the previous simulation approaches, while primarily conceptual, is therefore of great theoretical importance, as unlike previous research, I have been able to assess the probability and impact of compensatory mutations (Wagner, 1996; Siegal and Bergman, 2002; Azevedo et al., 2006).

The use of binary fitness outcomes (0/1 or unstable/stable) that are only periodically tested by strong purifying (phenotypic stability) selection is operationally quite useful. This avoids making unrealistic assumptions about the selection coefficient distribution and proceeds on the assumption that very slightly deleterious mutations will be predominant and allow the accumulation of subsequent mutations (some of which are compensatory mutations). Periodic assessment of the functional operation of networks, i.e., periods of purifying selection, is a necessary practical consideration. In fact, the fluctuating selection regime (periods of strong purifying selection) modelled in this chapter is also biologically realistic. For example, Siepielski et al. (2009) concluded that selection usually fluctuates when they studied the temporal dynamics of selection in a database containing $5,519$ estimates of selection in wild populations. Similar arguments using empirical evidence can be found in Brachi et al. (2013), Gompert et al. (2014), Seppälä (2015) and Bijleveld et al. (2015).

Previous work has been taken to indicate that compensatory mutation is not likely to play an important role in the evolution of independently acting genes. However, when considering that mutations occur in genes which exhibit complex interactions with other genes, then the frequency at which deleterious mutation incapacitates gene regulatory pathways is likely to be substantially higher than that for an independently acting gene, because there will inevitably be many more possible sites to mutate. In addition to Chapter 3 where I showed that compensatory mutation could potentially be frequent, in this chapter, I have further shown that compensatory mutation can occur regardless of the patterns of selection that the networks have been through (Figures 4-

2 and 4-3). I have also shown that compensatory mutation can still occur even among seriously damaged networks (Figure 4-4). In a related recent empirical study, Sloan et al. (2014) found that two *Silene* species with fast-evolving plastid and mitochondrial DNA exhibited increased amino acid sequence divergence in organelle genomes but not in cytosolic ribosomes. Given that the authors found no evidence that the observed pattern was driven by positive selection, they concluded that the rapid organelle genome evolution had selected for compensatory mutations in nuclear-encoded proteins. In this chapter, I have demonstrated in support of this empirical study that compensatory mutations can be greatly increased if the population is evolved under a relaxed selection regime (Figures 4-5 and C-2).

In Chapter 3, I explored how compensatory mutations could restore compromised networks. I showed that there is a bias with respect to where compensatory mutations happen such that compensatory mutations tend to generate regulatory circuits that closely interact with each other (Figure 3-5), whereas neutral networks tend to accumulate mutations that are further apart from each other (Figure 3-8). I also found a bias with respect to the size of compensatory mutations in terms of shifting gene regulation, such that compensatory mutations generate regulatory circuits that have larger interactive impacts (Figures 3-6), compared to neutral mutations (Figure 3-9). Previous work has indicated that the origin of mutational robustness may come from the non-adaptive results of biophysical principles or non-adaptive evolutionary forces (Payne and Wagner, 2015). In this chapter, I have found the evidence to support this hypothesis by showing that stable networks formed by these biased compensatory mutations tend to generate a profound change in robustness compared to the impact on stable networks of neutral mutations (Figures 4-7 and 4-8). These results indicate that over time, compensatory mutations that occur during generations of relaxed selection for phenotypic stability could be biased such that regulatory circuits that closely interact and have larger interactive impacts are more likely to be maintained.

Previous work has also indicated that compensatory mutations might facilitate the transition of the regulatory network to new fitness peaks. In particular, compensatory mutations have been observed to have a positive correlation with drug resistance mutations, where low-fitness lineages can create intrinsic selection pressure to mitigate their deleterious effects through compensatory mutations (Comas et al., 2012; Brandis et al., 2012; de Vos et al., 2013; Brandis and Hughes, 2013; Song et al., 2014). There is some evidence that compensatory mutation can even help the transition of lineages towards new fitness peaks (Martinez et al., 2014; Ivankov et al., 2014; Szamecz et al., 2014). Moreover, some recent studies have also shown that compensatory mutations can help increase plasmid stability, and thus facilitate adaptation (San Millan et al., 2014; Porter et al., 2015; Harrison et al., 2015). Despite suggestions in the literature that peak shifts must occur through low-fitness genotypes (Wagner and Wright,

2007; Romero and Arnold, 2009; Olson-Manning et al., 2012; Osada and Akashi, 2012; Barreto and Burton, 2013), few studies have focused on how the formation of regulatory networks could be influenced by this process. Given that regulatory networks could be evolved through compensatory mutations (Martinez et al., 2014) and the non-adaptive process could facilitate regulatory complexity (Ruths and Nakhleh, 2013), compensatory mutations are expected to play an essential role in driving regulatory complexity through non-adaptive processes.

In the work presented here, I have assessed the evolutionary consequences of these biased compensatory mutations. I have shown that compensatory mutation can facilitate regulatory complexity in terms of increasing network complexity in initial networks with connectivity variance but fixed structure (Figure 4-9), as well as networks with structure variance but fixed connectivity (Figure 4-10). It should be noted that the model set-up enables a non-adaptive evolution of compensatory mutation even if it brings an individual with fitness 0 (unstable) to fitness 1 (stable). First, let us consider how we test for evidence of adaptation with molecular data. An excess of non-synonymous substitutions ($d_N/d_S > 1$) suggests adaptive or diversifying selection, no difference between synonymous and non-synonymous mutation rates ($d_N/d_S = 1$) is taken as evidence for neutrality, and an excess of synonymous mutations ($d_N/d_S < 1$) indicates purifying selection. The interpretation of the result of the d_N/d_S value is that adaptation ($d_N/d_S > 1$) is evident when a beneficial mutation occurs in a coding part of a gene and then increases in the population to such a point that it is disproportionate to silent site mutations. However, although the compensatory mutation modelled in this Chapter is beneficial, since it restores an individual's fitness from 0 to 1, we would never expect networks with compensatory mutations to substantially increase in the population. This is because compensatory mutation does not increase competitive success relative to the reproductively active individuals in the population. If it happens frequently enough, and compensatory mutation does not introduce bias in the networks that are lost by purifying selection, then it is possible that a small fraction of compensatory mutations could increase in the population through random genetic drift. Therefore, if we were to sample the population for compensatory mutations, we would expect to find evidence of neutrality, i.e., $d_N/d_S = 1$. Likewise, if compensatory mutation can alter patterns of purifying selection, then although the mutation is beneficial, we would never expect it to increase in the population. Similarly, this is because compensatory mutation does not increase relative competitive success (Figures 4-12 and 4-13). However, if it happens frequently enough, and when it happens in particular patterns it is less likely to be removed by purifying selection, then it is possible that networks with compensatory mutations which increase robustness could increase by drift. If we were to sample the population for compensatory mutations, we would still expect to find evidence of neutrality ($d_N/d_S = 1$), but there also might

be evidence of weak purifying selection ($d_N/d_S < 1$), because only particular combinations of deleterious mutations and compensatory mutations would be maintained. Note that this explanation does not apply to super-compensatory mutations (which not only restore fitness, but also increase fitness to the point where it gives a competitive advantage). It should also be noted that there is no selection for regulatory complexity imposed in the simulations, and regulatory complexity is not coupled with selection for phenotypic stability (Figures 4-14). Therefore, taken together we can hypothesise that it is the two network properties I discovered — the location and regulatory impact biases observed in compensatory mutations — that drive the evolution of regulatory complexity through non-adaptive forces. These results are important, as they provide a better mechanistic understanding of how regulatory complexity arises through non-adaptive evolution. Compensatory mutations are essential in driving regulatory complexity via a biased purifying selection.

4.5 Summary and future work

In this chapter, I have further examined the evolutionary consequences of characteristics of compensatory mutations discussed in Chapter 3. Specifically, I have shown that compensatory mutation can occur under both strong and relaxed selection for phenotypic stability. In particular, compensatory mutation is still able to restore the stability of seriously damaged networks that have accumulated deleterious mutations, even for many generations. I have further shown that the number of compensatory mutations increases as the consequence of experiencing bouts of relaxed selection. This result is also supported by a recent empirical study by Sloan et al.. I have observed that robustness is higher when compensatory mutation occurs closer to the original deleterious mutation site or has a larger shift in gene regulation. These patterns are different in networks with neutral mutations. Specifically, robustness tends to be higher when neutral mutations are far apart. Moreover, large-effect mutations cannot generate a profound change in the robustness of networks with neutral mutations. However, robustness has been observed to be much higher in networks with neutral mutations than in networks with compensatory mutations. Finally, I have shown that compensatory mutations can drive regulatory complexity in terms of increasing the network connectivity of the population in two separate cases — initial networks with connectivity variance but fixed structure, and initial networks with structure variance but fixed connectivity. Some possible future research directions regarding exploring the phenotypic complexity generated by compensatory mutation and conditions under which regulatory complexity can arise are presented below.

4.5.1 Exploring the phenotypic complexity generated by compensatory mutation

In this chapter, I have shown that compensatory mutations can drive regulatory complexity in terms of increasing network connectivity. However, I have not yet tested whether the compensatory mutation can generate other aspects of complexity, for example, phenotypic complexity. Here, phenotypic complexity means that individuals can exhibit more different phenotypes. Networks with compensatory mutations typically have lower robustness, but they are expected to access greater phenotypic space. Therefore, It would be interesting to compare the number of unique phenotypes generated by networks with compensatory mutations and that of unique phenotypes generated by networks with no compensatory mutation. If networks with compensatory mutation could exhibit more different phenotypes, then these networks are more likely to survive and be maintained, facilitating adaptation to new environments. Thus, it would also be interesting to explore the role of compensatory mutations in improving individuals' evolvability.

4.5.2 Exploring conditions under which regulatory complexity can arise

In this chapter, I have provided two cases where we can observe the regulatory complexity arising through networks with compensatory mutations. However, I have not yet rigorously explored the conditions, such as relaxed selection frequency, initial network connectivity and number of genes, under which compensatory mutation can drive regulatory complexity in terms of increasing network connectivity and/or accessing greater phenotypic space. It would be interesting to explore the patterns or underlying mechanisms for those cases where regulatory complexity cannot arise. It would also be interesting to explore how the compensatory mutations with a biased robustness shown in this chapter have evolved over time for both cases when compensatory mutation is able or not able to drive regulatory complexity.

Chapter 5

Recombination is constructive in the context of selection for phenotypic stability

5.1 Introduction

Recombination is ubiquitous in multicellular plants, animals and even fungi. However, even basic questions such as explaining the costs and benefits of sexual reproduction are still unknown to both biological and computational sciences. Sex implies recombination — the reshuffling of parental genetic information, which generates heritable innovations (Eshel and Feldman, 1970; Feldman et al., 1996; Otto and Feldman, 1997; West et al., 1999). However, sexual reproduction is also considered to be very costly, since it may damage well-adapted lineages, and necessarily produces fewer directly reproductive offspring, since it also produces males. Evolution should favour defection to a lower-cost strategy, such as asexual reproduction. How then can sexual reproduction be beneficial?

For decades, researchers have been making tremendous efforts and proposing numerous theories for explaining the advantages of sex and recombination (Eshel and Feldman, 1970; Hurst and Peck, 1996; West et al., 1999; Otto and Lenormand, 2002; Meirmans and Strand, 2010; Wagner, 2011b). Two classic benefits of sexual reproduction are nevertheless still controversial: 1) purging deleterious mutations more efficiently, and 2) creating novel gene combinations (Kondrashov, 1993; Otto and Feldman, 1997; Otto and Gerstein, 2006; Kouyos et al., 2007; Barton, 2009; Martin and Wagner, 2009). An important third possibility is that the process of recombination, by allowing the localisation of both coherence and variation across the genomes of a population, is able to both improve robustness and facilitate evolutionary adaptation, a process

known as *evolvability* (Wang et al., 2014a). Although robustness and facilitated adaptation are observed phenomena and are often attributed to sexual reproduction, the underlying mechanisms are still poorly understood (Wagner, 2011b).

Recently, Wagner's GRN model (Wagner, 1994, 1996) has been employed as a powerful computation tool to study recombination in a network context (Azevedo et al., 2006; MacCarthy and Bergman, 2007a; Lohaus et al., 2010; Le Cunff and Pakdaman, 2014). An interesting feature of studying evolution in gene regulatory networks is that we can find clear evidence that evolution is not a simple optimisation process. The shifting genetic characteristics of the population resulting from mutation and selection optimise and innovate, but in nature the optima they track are also transient. Consequently, goals for an evolutionary genome always include robustness and agility, not only gradient ascent.

In a previous study, Siegal and Bergman (2002) designed evolutionary scenarios where they measured the phenotypic distance of evolved populations in the presence of mutation perturbations under different selection pressures for the optimal phenotype[1]. Siegal and Bergman reported that networks can evolve greater insensitivity to mutation (canalisation) even without directional selection for this property. In their paper, the authors described this property as lineages moving towards the optimum so long as the population is under selection for phenotypic stability; that is, selection for the optimal phenotype is largely absent. Although this suggests that selection for phenotypic stability is an important evolutionary force, the role of recombination is left unclear, since the earlier simulations ignored the possibility of asexual populations. In two later studies, Azevedo et al. (2006) and Lohaus et al. (2010) discovered that sexually reproducing organisms evolved higher mutational and recombinational robustness than asexual lineages. However, these authors did not explicitly measure the phenotypic distance of evolved asexual and sexual populations from the optimum. Therefore, it is an open question as to whether recombination is still able to sustain sexual reproduction as lineages near the optimum.

To get a better intuition on the questions of evolution in gene regulatory networks presented here, I employ a simple three-state descriptive model, as shown in Figure 5-1. Considering the probabilities ($p_{\mathrm{C,F}}$ and $p_{\mathrm{F,C}}$) of a lineage's bidirectional movements from being **c**lose to the optimum to being **f**ar from the optimum ($\mathbf{C} \rightarrow \mathbf{F}$) and *vice versa* ($\mathbf{F} \rightarrow \mathbf{C}$), I expect both to be fairly high in the absence of substantial selection for the optimal phenotype. This is because recombination in particular is a strong force that can substantially alter gene regulation in offspring networks. Therefore, the state transition probabilities for asexual populations may differ from those for sexual populations because mutation usually has a weaker effect. This also indicates that one

[1]Here, the optimal phenotype is not a phenotype picked at random but specifically refers to the initial gene expression pattern of a founder network, i.e., $\mathbf{s}_{\mathrm{OPT}} = \mathbf{s}(0)$.

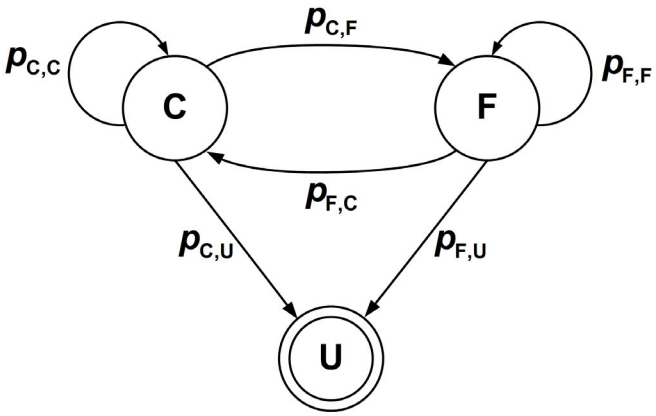


Figure 5-1: State transitions in a gene regulatory network. *There are three states in the system:* ***C****: individuals that are close to the optimum,* ***F****: individuals that are far from the optimum, and* ***U****: individuals that are unable to achieve phenotypic stability as defined in Section 2.3.6.* ***U*** *is an absorbing state in the system, since unstable genomes cannot reproduce and will be eliminated from the population.*

of the observations made in Siegal and Bergman (their Figure 2) may not be correct for asexual lineages. If we used such a conceptual model as shown in Figure 5-1 to analyse the behaviour of asexual and sexual populations, the state transition probabilities in asexual populations may not be the same as in sexual populations. In addition, the arguments for the benefits of sex and recombination in Azevedo et al. (2006) and Lohaus et al. (2010) are also incomplete, since these studies only show that $p_{\mathrm{C,U}}$ becomes smaller due to greatly increased mutational and recombinational robustness, leaving it still unclear whether recombination is able to retain sexual lineages in **C**, since $p_{\mathrm{C,F}}$ and $p_{\mathrm{F,C}}$ are largely unknown.

In this chapter, I hypothesise a new possible explanation for the widespread existence of sexual rather than asexual lineages by exploring systematically the approach to optima with only negligible selection ('no selection' as per Azevedo et al. (2006)) for the optimal phenotype, but only when there is selection for phenotypic stability. Using evolutionary simulations under the Wagner GRN model, I show that it is the evolutionary force of recombination together with developmental selection for phenotypic stability that drives populations towards the optimum. I further develop mathematical expressions for the conditions under which this process can be maintained. I find, quite surprisingly, that recombination does not frequently disrupt well-adapted lineages as conventionally expected. Rather, it facilitates finding good genetic combinations that are robust to disruption, although it also rapidly disrupts weaker configurations. These results indicate a fundamental difference between recombination and hypermutation, which has important implications for the role of gene regulation in the evolution of sex, and for the use of structured representations in machine learning.

5.2 Methods

In the modelling approach, asexual and sexual populations were evolved under purifying selection, i.e., selection for phenotypic stability (see Section 2.3.6), and no purifying selection. The system-level parameters were fixed to be $a = 100$, $devT = 100$ and $\tau = 10$ in all simulations. Note that periods of purifying selection were not allowed in simulations presented in this chapter. This means that the population will be either exposed to purifying selection at each generation or no purifying selection at all during its entire evolution. This is different from the previous Chapters 3 and 4, where populations were subjected to bouts of purifying selection.

5.2.1 The computational model

The computational model was similar to the model introduced in Section 2.3. Populations were either reproduced asexually by cloning themselves or sexually by recombining with each other during the entire life cycle. When the population was subjected to purifying selection, unstable individuals were wiped out immediately from the population pool, whereas both stable and unstable individuals could survive when purifying selection was absent.

Fitness evaluation

Fitness was evaluated by measuring the phenotypic distance between the equilibrium state and the optimal state. Specifically, for networks that were able to achieve phenotypic stability (reaching an equilibrium state, $\mathbf{s}_{\mathrm{EQ}}$), fitness was calculated as in Equation (2.3). For networks that were not able to achieve phenotypic stability under a no purifying selection regime, fitness was calculated as

$$F(\bar{\mathbf{s}}_{\mathrm{EQ}}) = \exp\left(-\frac{D(\bar{\mathbf{s}}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}})}{\sigma}\right), \tag{5.1}$$

where σ is the selection pressure, $\mathbf{s}_{\mathrm{OPT}}$ is the optimal phenotypic state, $\bar{\mathbf{s}}_{\mathrm{EQ}}$ is the approximated equilibrium phenotypic state and can be calculated by averaging the phenotypic state over $devT = 100$ iterations during an individual's developmental process, $D(\bar{\mathbf{s}}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}})$ is the phenotypic distance between the approximated equilibrium state and the optimal state and can be calculated as in Equation (2.2). Note that $\bar{\mathbf{s}}_{\mathrm{EQ}}$ was only used to calculate an individual's fitness when the individual was unstable and evolved under a no purifying selection regime. Otherwise, zero fitness was assigned to an individual that could not reach developmental equilibrium when evolved under a selection for phenotypic stability regime. This guaranteed that individuals with zero fitness would not be selected in the subsequent generation.

Initialisation

The initial population contained $M = 10,000$ identical clones of a founder network, which was generated by randomly filling W with $\lfloor c \times N^2 \rfloor$ non-zero elements $w_{i,j}$ $(i, j = 1, 2, \ldots, N)$ was drawn from the standard normal distribution, $N(0, 1)$. The associated initial expression state $\mathbf{s}(0)$ was also set by randomly choosing each $s_i(0) = +1$ or -1. The optimal phenotypic state was simply set to be the same as the initial expression state, i.e., $\mathbf{s}_{\mathrm{OPT}} = \mathbf{s}(0)$. This is because the model typically assumes that an individual's phenotype should be able to buffer against mutations in its genotype such that the initial gene expression pattern, $\mathbf{s}(0)$, could be maintained. In the conducted simulations, ten randomly generated stable networks were used as the founder networks. All founder networks had the same initial phenotypic distance from their corresponding optimum, $D(\mathbf{s}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}}) = 0.2$, although the genotype W and the associated initial expression state $\mathbf{s}(0)$ were different.

Mutation

For an individual network, each non-zero entry in the W adjacency matrix was replaced by $w'_{i,j} \sim N(0, 1)$ $(i, j = 1, 2, \ldots, N)$ with mutation rate μ. The expected number of mutations in W was drawn from the Poisson distribution as described in Section 2.3.4. In all simulations, I used $\mu = 0.1$, which meant on average there was an 0.1 non-zero entry in W that would be mutated per network per generation. Note that the mutation rate was different from that used in the previous Chapters 3 and 4 where there was one and only one non-zero entry mutated per network per generation.

Recombination

The recombination operator was the same as described in Section 2.3.5.

Stability and fitness selection

If individuals evolved under a regime of selection for phenotypic stability, I only allowed those that could reach developmental equilibrium to stay in the population. Otherwise, if individuals evolved without selection for phenotypic stability, I allowed both stable and unstable individuals to stay in the population. Unless otherwise specified, I set $\sigma = 10^9$ as used in Siegal and Bergman (2002) and Azevedo et al. (2006) in all simulations to evaluate individual fitness. Note that using such a large value $\sigma = 10^9$ in Equation (2.3), all individuals have a fitness greater than 0.9999, very close to 1. This means, in the conducted simulations, all populations were evolved under extremely weak or even absent selection for the optimal phenotype.

Evolution

The evolutionary simulations were performed under the reproduction-mutation-selection life cycle similarly to how it was described in Section 4.2.1. In typical asexual evolution, an individual was chosen at random to reproduce asexually by cloning itself, and then subjected to mutation, then extremely weak selection for the optimal phenotype. Similarly, in typical sexual evolution, two individuals were chosen at random to reproduce sexually by recombining two parent networks, and then subjected to mutation, then extremely weak selection for the optimal phenotype. This process was repeated until M number of networks were produced. Depending on whether or not the population evolved under the selection for phenotypic stability regime, I either excluded unstable networks or allowed these compromised networks to stay in the population, accordingly.

5.2.2 Measuring the phenotypic distance for asexual and sexual populations

In order to estimate the distance within the population or the distance between the population and the optimal phenotype, I employed a similar perturbation test to that described in Siegal and Bergman (2002): I defined a perturbation as a single mutant, i.e., exactly one non-zero entry in W was replaced by a random value drawn from $N(0,1)$. For each individual in the population, if its perturbed network could still reach an equilibrium state, I defined it as $\mathbf{s}_{\mathrm{P}} = \mathbf{s}_{\mathrm{EQ}}$, otherwise $\mathbf{s}_{\mathrm{P}} = \bar{\mathbf{s}}_{\mathrm{EQ}}$, where $\bar{\mathbf{s}}_{\mathrm{EQ}}$ is calculated as in Equation (5.1). The distances between the perturbed individual and its unperturbed one or the optimal phenotypic state are defined as $D(\mathbf{s}_{\mathrm{P}}, \hat{\mathbf{s}}_{\mathrm{EQ}})$[2] and $D(\mathbf{s}_{\mathrm{P}}, \mathbf{s}_{\mathrm{OPT}})$. For each individual in the population, the distance was estimated by averaging 10 perturbations. The reported results were averaged over $10,000$ individuals in the population, a total of $100,000$ perturbations for asexual and sexual populations under phenotypic stability or no stability selection regimes and the simulation was replicated using 10 randomly generated founder networks as shown in Figures 5-2 and 5-3.

To further confirm that recombination is fundamentally different from hypermutation, I designed two additional sets of simulations (see Figures D-3 and D-4). Specifically, two hypermutation strategies were modelled: random mutation and row mutation. In the simulations with random mutation, N non-zero sites[3] that were generated randomly were mutated (replaced with random values drawn from the standard normal distribution) for each individual since N sites were changed simultaneously in

[2]Note that $\hat{\mathbf{s}}_{\mathrm{EQ}}$ is the equilibrium phenotypic state of the unperturbed individual if it is stable; otherwise if the unperturbed individual is unstable, $\hat{\mathbf{s}}_{\mathrm{EQ}}$ is the approximated equilibrium phenotypic state as calculated in Equation (5.1).

[3]In the simulations performed in this Chapter, N is set to be 10.

recombination. Given that recombination was swapping rows among parent networks, in simulations with row mutation, each row of the parent network was mutated (all non-zero sites of the row were replaced with random values drawn from the standard normal distribution) with a probability of 0.5. It should be noted that the second mutation strategy was similar to recombination in terms of number of mutated sites, except that in recombination the mutated sites form a regulatory circuit that worked well together, whereas mutated sites were replaced with random values that may not work well together in hypermutation.

5.2.3 Measuring transition probability for asexual and sexual populations

To measure all state transition probabilities p_{ij} ($i, j \in \{\mathbf{C}, \mathbf{F} \text{ and } \mathbf{U}\}$) as shown in Figure 5-1, we need to find two populations that are in state **C** and state **F** for both asexual and sexual populations. Specifically, as Figure 5-2 indicates, I used the below four evolved populations for further analysis: For individuals close to the optimum, I chose the asexual and sexual populations that had been evolved for $1,000$ generations with selection for phenotypic stability; For individuals far away from the optimum, I chose the stable individuals from the asexual and sexual populations that had been evolved for $1,000$ generations without selection for phenotypic stability. Note that the chosen sexual and asexual populations were not perfect but reasonable approximations of two sets of individuals that were in state **C** and state **F**.

For these four populations, each individual experienced either asexual or sexual reproduction, and then was subjected to one single mutation (replacing exactly one non-zero w_{ij} with a random value drawn from $N(0, 1)$). Next, I could easily take the proportion of stable offspring as $1 - p_{\mathrm{C,U}}$ and $1 - p_{\mathrm{F,U}}$. Those stable mutants were saved to further measure the remaining four parameters in Figure 5-1. I took the mean and standard deviation of the phenotype distance from the optimum at the $1,000^{th}$ generation from the sexual population as a criterion for **C** to test whether the mutants were closer to the optimum. Similarly, I took the mean and standard deviation of the phenotype distance away from the optimum at the $1,000^{th}$ generation from the asexual population as a criterion for **F**[4] to test whether the mutants were further away from the optimum. For each mutant derived from **C**, if its phenotypic distance was smaller than one standard deviation, I counted it as in $p_{\mathrm{C,C}}$, otherwise it was counted as in $p_{\mathrm{C,F}}$. Similarly, for each mutant derived from **F**, if its phenotypic distance was greater than one standard deviation, I counted it as in $p_{\mathrm{F,F}}$, otherwise it was counted as in $p_{\mathrm{F,C}}$. Note that the boundary between $p_{\mathrm{C,C}}$ and $p_{\mathrm{C,F}}$, and the boundary between $p_{\mathrm{F,F}}$ and $p_{\mathrm{F,C}}$ are

[4]Since asexual and sexual populations behave similarly (see Figure 5-2) when there is no selection for phenotypic stability, it does not matter whether I took the mean and standard deviation of the phenotype distance from the asexual population or the sexual population as a criterion for an estimated **F**.

arbitrarily defined. If we slightly increase the range from one standard deviation to two standard deviations, then both $p_{C,F}$ and $p_{F,C}$ will be reduced (see Figures D-1 and D-2). It should be noted that the observation that the sexual population's $p_{C,F}$ are smaller than the asexual population's $p_{C,F}$ holds for any arbitrarily defined boundary.

5.3 Results

Using an established model of gene regulatory networks as described in Section 5.2.1, I was able to test the possibility that the maintenance of the system might depend on how well recombination generates lineages that could be maintained close to the optimum. Specifically, I designed two sets of simulations to investigate how mutation and recombination influence evolutionary dynamics in asexual and sexual populations. Unlike the experimental set-ups in Siegal and Bergman (2002), I took both sexuality and phenotypic stability into consideration. For each set of experiments, the results are presented for simulations under four different evolutionary scenarios: 1) population with asexual reproduction under a selection for phenotypic stability regime, 2) population with asexual reproduction under a no selection for phenotypic stability regime, 3) population with sexual reproduction under a selection for phenotypic stability regime, and 4) population with sexual reproduction under a no selection for phenotypic stability regime. Note that in all four cases, I set $\sigma = 10^9$, which means the selection for the optimal phenotype was extremely weak or even absent in the conducted simulations.

5.3.1 Recombination and selection for phenotypic stability drive lineages towards the optimum

In the first set of experiments, I measured the phenotypic distance between the evolved populations and the optimum. I found that it is the combination of recombination and selection for phenotypic stability that can drive the population towards the optimum. Specifically, I compared the results of the asexual and sexual populations under phenotypic stability or no stability selection regimes. From Figure 5-2, we can see that when there is no selection for phenotypic stability, both asexual and sexual populations rapidly move away from the optimum at a similar increasing rate. In contrast, when selection for phenotypic stability is imposed on the sexual population, the phenotypic distance continuously decreases. Although under a selection for phenotypic stability regime the phenotypic distance of the asexual population slightly decreases first and then slightly increases later, selection for phenotypic stability greatly impedes deviation, compared with the situation when selection for phenotypic stability is absent. These results suggest that the two forces of recombination and purifying selection (selection for phenotypic stability) are both critical for a population to evolve towards the optimum. Note that this phenomenon has been similarly reported in Siegal and

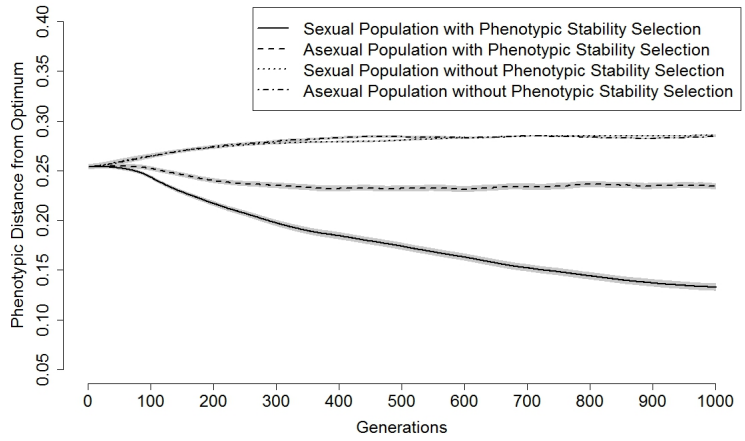


Figure 5-2: Comparison of the phenotypic distance between the evolved populations and the optimum. *The initial population (*$10,000$*) was cloned from a randomly generated stable founder network with size* $N = 10$ *and connectivity* $c = 0.75$*. The population was then evolved asexually or sexually under phenotypic stability or no stability selection regimes. In each generation, each individual in the population was subjected to a perturbation test in order to calculate the phenotypic distance between the evolved populations and the optimum (see Section 5.2.2). The shaded areas represent* 95% *confidence intervals based on* 10 *randomly generated stable founder networks.*

Bergman (2002), except there it could only be observed in sexual lineages rather than asexual lineages. It should also be noted that, as shown in Figure D-3, hypermutation does not help asexual lineages to move close to the optimum even when selection for phenotypic stability is included.

5.3.2 Recombination and selection for phenotypic stability facilitate lineages staying close

In the second set of experiments, I measured the phenotypic distance within the evolved populations. I found that recombination and selection for phenotypic stability can also help sexual lineages stay close to each other. Similarly to the first set of experiments, I compared the results for the asexual and sexual populations under phenotypic stability or no stability selection regimes. From Figure 5-3, we can see that in contrast to the results where I compared phenotypic distance between the evolved populations and the optimum, the phenotypic distance within the populations reduces when there is no selection for phenotypic stability in both asexual and sexual populations at a similar decreasing rate. Although there was a small change in phenotypic distance within the asexual population when I included selection for phenotypic stability, the phenotypic distance was highly reduced in sexual lineages. This indicates that the two forces of recombination and selection for phenotypic stability are also both critical for

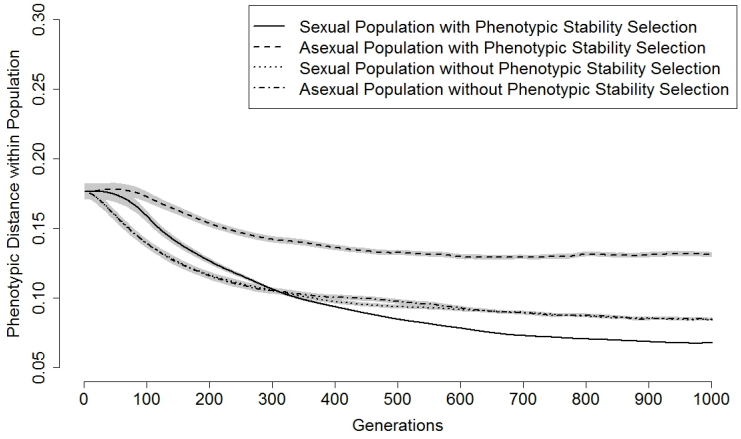


Figure 5-3: Comparison of the phenotypic distance within the evolved populations. *The initial population (*$10,000$*) was cloned from a randomly generated stable founder network with size* $N = 10$ *and connectivity* $c = 0.75$*. The population was then evolved asexually or sexually under phenotypic stability or no stability selection regimes. In each generation, each individual in the population was subjected to a perturbation test in order to calculate the phenotypic distance within the evolved populations (see Section 5.2.2). The shaded areas represent* 95% *confidence intervals based on* 10 *randomly generated stable founder networks.*

a population to evolve towards convergence. It should be noted that, as shown in Figure D-4, hypermutation also does not help asexual lineages to stay close to each other when selection for phenotypic stability is imposed.

5.3.3 Analysis

From Figures 5-2 and 5-3, we can clearly see that phenotypic distance is continuously decreasing in the sexual population as a consequence of recombination and selection for phenotypic stability. In order to fully understand how these two forces act together, here I further investigated the underlying evolutionary dynamics in the context of gene regulatory networks.

The evolutionary dynamics in the Wagner GRN model can be regarded as a Markov process, since the future of the evolution process is based solely on its present state, by definition in Section 2.4. To simplify the analysis, I define three states in the system as per Figure 5-1:

- **C**: Individuals that are close to the optimal phenotype
- **F**: Individuals that are far from the optimal phenotype
- **U**: Individuals that are unable to achieve phenotypic stability; thus will be eliminated from the population

It should be noted that although unstable networks will be removed from the population pool, the population size is fixed at each generation (see Section 5.2.1).

Suppose at the g^{th} generation, the frequency of individuals in state **C** is $f_{\mathrm{C}}(g)$, the frequency of individuals in state **F** is $f_{\mathrm{F}}(g) = 1 - f_{\mathrm{C}}(g)$. Then, at the $(g+1)^{th}$ generation, the frequency of individuals in state **C** and **F** are $f_{\mathrm{C}}(g+1) = (f_{\mathrm{C}}(g) \times p_{\mathrm{C,C}} + f_{\mathrm{F}}(g) \times p_{\mathrm{F,C}})/\lambda$ and $f_{\mathrm{F}}(g+1) = (f_{\mathrm{F}}(g) \times p_{\mathrm{F,F}} + f_{\mathrm{C}}(g) \times p_{\mathrm{C,F}})/\lambda$, where a normalising factor $\lambda = (f_{\mathrm{C}}(g) \times p_{\mathrm{C,C}} + f_{\mathrm{F}}(g) \times p_{\mathrm{F,C}}) + (f_{\mathrm{C}}(g) \times p_{\mathrm{C,F}} + f_{\mathrm{F}}(g) \times p_{\mathrm{F,F}})$. Therefore, the changing rate of frequency f_{C} in two consecutive generations can be described in the following differential equation:

$$\begin{aligned} \Delta \mathrm{C} &= f_{\mathrm{C}}(g+1) - f_{\mathrm{C}}(g) \\ &= \frac{f_{\mathrm{C}}(g) \times p_{\mathrm{C,C}} + f_{\mathrm{F}}(g) \times p_{\mathrm{F,C}}}{\lambda} - f_{\mathrm{C}}(g) \\ &= \frac{f_{\mathrm{C}}(g) \times p_{\mathrm{C,C}} + f_{\mathrm{F}}(g) \times p_{\mathrm{F,C}} - \lambda \times f_{\mathrm{C}}(g)}{\lambda}. \end{aligned} \tag{5.2}$$

As a population evolves towards the optimum, we expect to see a higher frequency of f_{C} in the population, i.e., $\Delta \mathrm{C} \geqslant 0$ Therefore, the below equation should be satisfied:

$$\begin{aligned} & f_{\mathrm{C}}(g) \times p_{\mathrm{C,C}} + f_{\mathrm{F}}(g) \times p_{\mathrm{F,C}} - \lambda \times f_{\mathrm{C}}(g) \\ &= f_{\mathrm{C}}(g) \times p_{\mathrm{C,C}} + f_{\mathrm{F}}(g) \times p_{\mathrm{F,C}} - f_{\mathrm{C}}(g) \times (f_{\mathrm{C}}(g) \times p_{\mathrm{C,C}} \\ &\quad + f_{\mathrm{F}}(g) \times p_{\mathrm{F,C}} + f_{\mathrm{C}}(g) \times p_{\mathrm{C,F}} + f_{\mathrm{F}}(g) \times p_{\mathrm{F,F}}) \\ &= f_{\mathrm{C}}(g) \times p_{\mathrm{C,C}} + f_{\mathrm{F}}(g) \times p_{\mathrm{F,C}} - f_{\mathrm{C}}^2(g) \times p_{\mathrm{C,C}} \\ &\quad - f_{\mathrm{C}}(g) \times f_{\mathrm{F}}(g) \times p_{\mathrm{F,C}} - f_{\mathrm{C}}^2(g) \times p_{\mathrm{C,F}} \\ &\quad - f_{\mathrm{C}}(g) \times f_{\mathrm{F}}(g) \times p_{\mathrm{F,F}} \\ &= f_{\mathrm{C}}(g) \times p_{\mathrm{C,C}} + (1 - f_{\mathrm{C}}(g)) \times p_{\mathrm{F,C}} - f_{\mathrm{C}}^2(g) \times p_{\mathrm{C,C}} \\ &\quad - f_{\mathrm{C}}(g) \times (1 - f_{\mathrm{C}}(g)) \times p_{\mathrm{F,C}} - f_{\mathrm{C}}^2(g) \times p_{\mathrm{C,F}} \\ &\quad - f_{\mathrm{C}}(g) \times (1 - f_{\mathrm{C}}(g)) \times p_{\mathrm{F,F}} \\ &= f_{\mathrm{C}}(g) \times p_{\mathrm{C,C}} + p_{\mathrm{F,C}} - f_{\mathrm{C}}(g) \times p_{\mathrm{F,C}} - f_{\mathrm{C}}^2(g) \times p_{\mathrm{C,C}} \\ &\quad - f_{\mathrm{C}}(g) \times p_{\mathrm{F,C}} + f_{\mathrm{C}}^2(g) \times p_{\mathrm{F,C}} - f_{\mathrm{C}}^2(g) \times p_{\mathrm{C,F}} \\ &\quad - f_{\mathrm{C}}(g) \times p_{\mathrm{F,F}} + f_{\mathrm{C}}^2(g) \times p_{\mathrm{F,F}} \end{aligned}$$

$$
\begin{aligned}
&= (p_{\mathrm{F,F}} + p_{\mathrm{F,C}} - p_{\mathrm{C,C}} - p_{\mathrm{C,F}}) \times f_{\mathrm{C}}^2(g) \\
&\quad + (p_{\mathrm{C,C}} - 2p_{\mathrm{F,C}} - p_{\mathrm{F,F}}) \times f_{\mathrm{C}}(g) + p_{\mathrm{F,C}} \\
&\geqslant 0.
\end{aligned}
$$

Therefore, the following condition should hold:

$$
(p_{\mathrm{F,U}} - p_{\mathrm{C,U}}) \times f_{\mathrm{C}}^2(g) + (2p_{\mathrm{F,C}} + p_{\mathrm{F,F}} - p_{\mathrm{C,C}}) \times f_{\mathrm{C}}(g) - p_{\mathrm{F,C}} \leqslant 0. \tag{5.3}
$$

It is reasonable to assume that individuals that are far away from the optimum (**F**) are more likely to become unstable than individuals that are close to the optimum (**C**), i.e., $p_{\mathrm{F,U}} \geqslant p_{\mathrm{C,U}}$. This is because there is always a selection for phenotypic stability that enables the population to move towards the optimum, whereas selection for phenotypic stability is absent, and consequently drags the population away from the optimum. Therefore, the quadratic equation Equation (5.3) is concave up. Given that $f_{\mathrm{C}}(g) \in [0, 1]$, Equation (5.3) will hold as long as it holds in $[0, 1]$. Therefore, the below two conditions should be satisfied:

$$
\begin{aligned}
-p_{\mathrm{F,C}} &\leqslant 0 \\
p_{\mathrm{F,U}} - p_{\mathrm{C,U}} + 2p_{\mathrm{F,C}} + p_{\mathrm{F,F}} - p_{\mathrm{C,C}} - p_{\mathrm{F,C}} &\leqslant 0
\end{aligned}
$$

Clearly, $p_{\mathrm{F,C}} \geqslant 0$ always holds, therefore the second condition, which is $p_{\mathrm{C,F}} \leqslant 0$, should hold. But we know that $p_{\mathrm{C,F}} \geqslant 0$. Therefore, in order to let Equation (5.3) hold, $p_{\mathrm{C,F}} \approx 0$ should hold. This suggests that as long as the population is continuously moving towards the optimum, the evolved lineages are unlikely to be deviated by mutation and recombination from the area close to the optimum (**C**) to the area far from it (**F**).

From the above analysis, we can speculate that as long as $p_{\mathrm{C,F}}$ is small enough, we should be able to see an increased frequency of lineages in the **C** state. From the observation of the evolutionary simulations, I further expect that $p_{\mathrm{C,F}}$ should be smaller in sexual lineages than in asexual lineages, since the sexual population is moving more quickly towards the optimum, whereas a similar pattern has not been observed in the asexual population.

It should be noted that the analysis presented in this section is based on the condition in which the selection for the optimal phenotype is extremely weak or even absent, and only selection for phenotypic stability is considered. However, as previous work has indicated, lineages are still able to move towards the optimum even if there is no such selection force imposed on the population (Siegal and Bergman, 2002). Here,

I have explored the mechanism of the underlying evolutionary dynamics and further analysed the condition under which we would expect to see an increase in the frequency of linages that move towards the optimum.

5.3.4 Simulations for measuring state transition probability

To verify the analysis, I conducted further simulations to measure the state transition probabilities p_{ij} ($i, j \in \{\mathbf{C}, \mathbf{F} \text{ and } \mathbf{U}\}$). To be more specific, I used the four evolved populations described in Section 5.2.3. From Figures 5-4 and 5-5, we can clearly see in that $p_{\mathrm{F,U}} \geqslant p_{\mathrm{C,U}}$ holds for both asexual and sexual populations. But only in the sexual population is $p_{\mathrm{C,F}}$ a small value, 0.0386, whereas $p_{\mathrm{C,F}}$ is a much larger value, 0.3810, in the asexual population. These results confirm that Equation (5.3) holds only for sexual populations when there is no selection for the optimal phenotype. However, this does not suggest that Equation (5.3) will never hold in asexual populations. When selection for the optimal phenotype is turned on, we can also observe a small value of $p_{\mathrm{C,F}}$ in the asexual population (results not shown).

From Figures 5-4 and 5-5, we can further calculate the stationary probabilities of lineages in states **C** and **F** for asexual and sexual populations: For the asexual lineages: 33.68% in the **C** state and 66.33% in the **F** state. For the sexual lineages: 86.90% in the **C** state and 13.10% in the **F** state. These results demonstrate that recombination substantially enables sexual lineages to sustain themselves near the optimum to a surprisingly high probability. This further indicates a fundamental difference between recombination and hypermutation, despite their superficial similarity in causing increased variations.

Taken together, from these state transition probabilities in sexual populations, we can also see that selection for phenotypic stability helps purge lineages more efficiently if they are far away from the optimum. However, those sexual lineages that have been able to move close to the optimum are evolved to be much more robust and can be highly maintained.

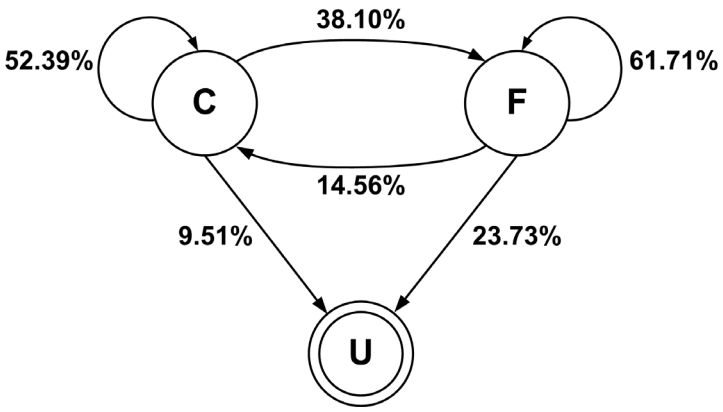


Figure 5-4: Estimated state transition probabilities in asexual populations. *$p_{C,U}$: 9.51% (SD: 0.86%), $p_{C,F}$: 38.10% (SD: 7.70%), $p_{C,C}$: 52.39% (SD: 7.57%), $p_{F,U}$: 23.73% (SD: 3.27%), $p_{F,C}$: 14.56% (SD: 6.54%), $p_{F,F}$: 61.71% (SD: 6.41%). For each population evolved from the founder network, the state transition probabilities were estimated based on* 50 *independent runs. The reported results are the mean probability averaged over* 10 *randomly generated stable founder networks. SD: Standard Deviation.*

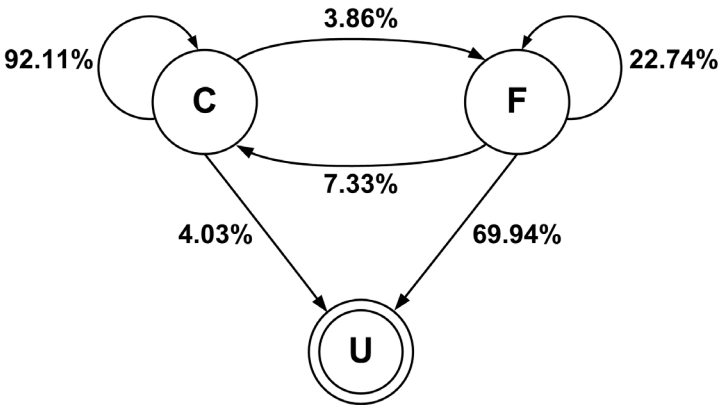


Figure 5-5: Estimated state transition probabilities in sexual populations. *$p_{C,U}$: 4.03% (SD: 0.68%), $p_{C,F}$: 3.86% (SD: 2.23%), $p_{C,C}$: 92.11% (SD: 1.92%), $p_{F,U}$: 69.94% (SD: 3.35%), $p_{F,C}$: 7.33% (SD: 3.12%), $p_{F,F}$: 22.74% (SD: 3.45%). For each population evolved from the founder network, the state transition probabilities were estimated based on* 50 *independent runs. The reported results are the mean probability averaged over* 10 *randomly generated stable founder networks. SD: Standard Deviation.*

5.4 Discussion

Mutation and recombination are two important evolutionary forces that provide heritable genetic innovations which ultimately stimulate adaptation for species to survive in nature. However, compared with mutation, recombination is thought to be much more mysterious because it leads to a fundamental evolutionary question: how can sexual reproduction, once evolved, be maintained in the long term? In particular,

we do not clearly understand why asexual lineages do not outcompete sexual lineages, given the substantial cost of recombination (disrupting good genetic combinations) and the twofold cost of sex (producing half as many lineages because only females reproduce). Although previous studies have posited that recombination has an important role in improving robustness and facilitating evolutionary adaptation, the underlying mechanism has remained unclear.

Wagner's gene regulatory network model has motivated research on the evolution of genetic networks (Fierst and Phillips, 2015), and attracted many researchers in different fields, since the model has both mathematical and biological roots (Payne et al., 2014; Hu et al., 2014; Wang et al., 2014a). Selection for phenotypic stability is a key feature in the Wagner GRN model. As Siegal and Bergman (2002) pointed out, it is difficult to find a case where such a developmental module is required in nature, but an individual's phenotype can be evolved independently from the selection for a particular optimum[5], especially when we imagine a scenario where a species colonises a new territory with abundant natural resources, so the selection for the optimal phenotype is extremely weak or even absent, but lineages are still able to continuously evolve.

Although Siegal and Bergman (2002) emphasised the importance of the 'phenotypic buffer' for genetic innovation provided by selection for phenotypic stability, they did not take sexuality into consideration. In later studies, Azevedo et al. (2006) and Lohaus et al. (2010) reported the observed mutational and recombinational robustness evolved from sexual lineages, but they still did not clearly explain why a greater benefit can be observed in sexual populations. Here, I have used a simple, three-state system to describe the evolutionary dynamics in the Wagner GRN model (Figure 5-1). With this I have shown that even if both the mutational and recombinational robustness are greatly increased (i.e. $p_{\mathrm{C,U}}$ is reduced), the underlying dynamics of sexual lineages have been poorly understood. One reasonable intuition would be a high transition probability of lineages moving from area **C** to **F**, since recombination is thought to greatly disrupt well-adapted lineages, and consequently it would be thought of as unlikely to sustain population in area **C**, close to the optimum. Therefore $p_{\mathrm{C,F}}$ would be expected to be high. Instead, I have found that while $p_{\mathrm{F,U}}$ is high, as similarly predicted, individuals of sexual lineages that are close to the optimum appear to be very robust against disruption by recombination. We can take this as clear evidence of evolvability (Wang et al., 2014a).

By comparing evolutionary simulations of asexual and sexual lineages under selection for phenotypic stability or in its absence, I have shown that the conclusion in Siegal and Bergman (2002) is not complete. It is the combination of two evolutionary forces — recombination and selection for phenotypic stability — that drives populations towards the optimum, not selection for phenotypic stability alone (Figures 5-2 and 5-3).

[5]Note that here the optimum specifically refers to the individual's initial phenotypic state.

I have thus clarified the benefit derived from sexual lineages. The numerical analysis has verified the assumptions and further validated the results. I have shown that recombination is surprisingly constructive for close-to-optimum sexual lineages and only destructive to populations far from the optimum. This indicates that recombination facilitates the evolution of evolutionary robustness — a form of evolvability — in sexual lineages. In contrast, mutations are more likely to be mildly deleterious and thus after accumulation in an asexual population, ultimately tend to move it away from the optimum. These findings indicate a fundamental difference between recombination and hypermutation, which has important implications both for structuring machine learning and for finally explaining the evolutionary stability of sex.

5.5 Summary and future work

In this chapter, I have provided a mechanistic understanding of why sexual lineages can evolve a greater benefit than asexual lineages in the context of genetic networks. Specifically, I have shown that it is recombination together with developmental selection for phenotypic stability that drives sexual lineages towards the optimum, not developmental selection for phenotypic stability alone, as indicated by previous work. The evolutionary forces of recombination and developmental selection for phenotypic stability have also been observed to help sexual lineages stay close to each other. Using a three-state conceptual model, I have found that in order to see an increased frequency of lineages in **C** state, the transition probability of $p_{\mathrm{C,F}}$ should be close to zero. I have further conducted simulations to measure transition probabilities in sexual and asexual lineages, and found that, as expected, the condition $p_{\mathrm{C,F}} \approx 0$ only holds for sexual lineages. I have shown that recombination facilitates finding good genetic combinations that are robust to disruption but rapidly disrupts weaker configurations. Some possible future research directions regarding exploring recombination benefits in diverse populations and examining how compensatory mutations benefit sexual lineages are presented below.

5.5.1 Exploring recombination benefits in diverse populations

In this chapter, I followed previous papers (Siegal and Bergman, 2002; Azevedo et al., 2006; Lohaus et al., 2010) in using identical copies of the founder network as sexual populations, simulating evolution in laboratory conditions, to avoid unfair comparison with the population under asexual reproduction. One limitation of using founder networks is that the recombination is not able to massively shift gene regulation and alter network topology. Therefore, it would be interesting to study whether network structures could also be evolved for particular properties that increase robustness or facilitate evolutionary adaptation more rapidly.

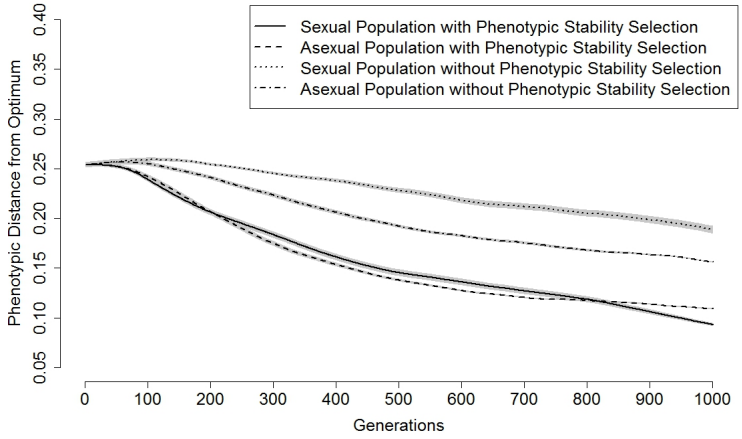


Figure 5-6: Comparison of the phenotypic distance between the optimum and the populations evolved under medium selection pressure for the target phenotype. *The initial population (*$10,000$*) was cloned from a randomly generated stable founder network with size* $N = 10$ *and connectivity* $c = 0.75$*. The population was then evolved asexually or sexually under phenotypic stability or no stability selection regimes, and the selection pressure for the target phenotype was set to* $\sigma = 10$*. In each generation, each individual in the population was subjected to a perturbation test in order to calculate the phenotypic distance between the evolved populations and the optimum (see Section 5.2.2). The shaded areas represent* 95% *confidence intervals based on* 10 *randomly generated stable founder networks.*

5.5.2 Examining the role of compensatory mutation in facilitating canalisation

In this chapter, I have included the fitness evaluation as calculated using Equations (2.3) and (5.1) which positively relate to the distance between the individual's phenotypic state and the optimal (target) phenotypic state. This is different from in the previous Chapters 3 and 4 where fitness was simply a binary value: 0 (unstable network) or 1 (stable network). However, I followed Azevedo et al. (2006) to set $\sigma = 10^9$ to simulate the evolutionary scenario where selection for a particular phenotypic state is largely absent. Thus, populations can be considered to be evolving similarly as in the previous Chapters 3 and 4 where the population is only subjected to purifying selection, i.e., selection for phenotypic stability. It is therefore natural to conduct further experiments to test whether the patterns observed in this chapter would be different if the population was subjected to both purifying selection (phenotypic stability) and target (fitness) selection.

Here, I have presented some preliminary results. Using a similar modelling approach to that described in Section 5.2, I have further measured the phenotypic distance between the optimum and the populations that have evolved asexually and sexually under phenotypic stability and no stability selection regimes. However, differently

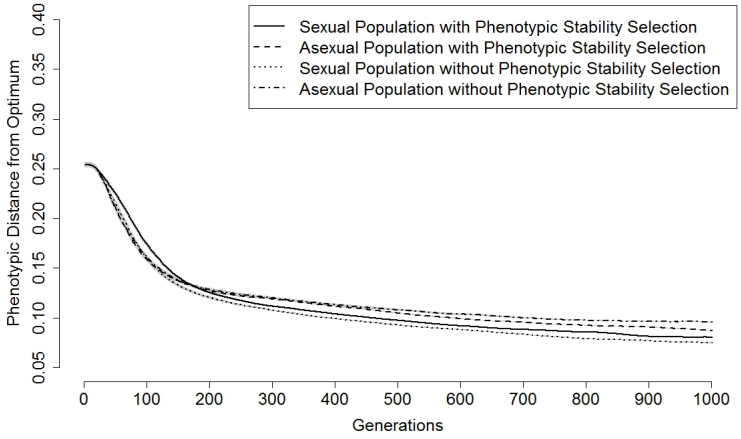


Figure 5-7: Comparison of the phenotypic distance between the optimum and the populations evolved under strong selection pressure for the target phenotype. *The initial population (*$10,000$*) was cloned from a randomly generated stable founder network with size* $N = 10$ *and connectivity* $c = 0.75$*. The population was then evolved asexually or sexually under phenotypic stability or no stability selection regimes, and the selection pressure for the target phenotype was set to* $\sigma = 0.5$*. In each generation, each individual in the population was subjected to a perturbation test in order to calculate the phenotypic distance between the evolved populations and the optimum (see Section 5.2.2). The shaded areas represent* 95% *confidence intervals based on* 10 *randomly generated stable founder networks.*

from the results in Section 5.3.1, I set $\sigma = 10$ and $\sigma = 0.5$ to simulate two evolutionary scenarios where populations were evolved under medium and strong selection for the optimal phenotype.

From Figures 5-6 and 5-7 (cf. Figure 5-2), it is not surprising to see that generally a stronger target selection will help speed up the evolution process of reducing phenotypic distance. However, it would be interesting to see whether populations without selection for phenotypic stability are able to move towards the optimum, in contrast to the situation where populations are evolved far away from the optimum if they are subjected to extremely weak or even in the absence of selection for the optimal phenotype. In particular, sexual populations without selection for phenotypic stability evolve even faster than the case when those populations are subjected to selection for phenotypic stability (see Figure 5-7). I then measured the proportion of stable networks in those asexual and sexual populations[6] which had been evolved for $1,000$ generations without selection for phenotypic stability. From Table 5.1, we can see that the proportion of stable networks substantially increases due to the increased selection pressure for the target phenotype, even if populations have never been subjected to selection for phenotypic stability. These results suggest that individuals that have been compromised (lost network stability) generally have lower fitness, and will be more

[6]The populations were taken from Figures 5-2, 5-6 and 5-7.

Table 5.1: *Proportion of stable networks in evolved populations without selection for phenotypic stability*

	No Target Selection ($\sigma = 10^9$)	Medium Target Selection ($\sigma = 10$)	Strong Target Selection ($\sigma = 0.5$)
Asexual Population	11.19% (SD: 3.01%)	55.16% (SD: 4.56%)	96.31% (SD: 1.07%)
Sexual Population	10.24% (SD: 2.25%)	16.09% (SD: 7.44%)	97.43% (SD: 0.90%)

The reported results are based on populations that have been evolved for $1,000$ generations using 10 randomly generated founder networks. SD: Standard Deviation.

likely to be wiped out when they are subjected to substantial selection pressure for the target phenotype. However, there may be occasionally compensatory mutations that could restore individual fitness. Consequently, those restored networks are likely to be maintained in subsequent generations in the presence of selection for target phenotype. Therefore, further experiments need to be conducted to rigorously investigate the role of compensatory mutations in benefiting sexual lineages.

Chapter 6

Selection pressure benefits low-fitness individuals and mitigates the costs of sex and recombination

6.1 Introduction

The maintenance of sex is one of the most mysterious unsolved problems in evolutionary biology. Sexual reproduction is widespread in nature, although asexual reproduction remains ubiquitously in single-celled organisms, many plants and fungi (Butlin, 2002). Individuals that have survived millions of years of evolution have increased their probability to well adapt to the current environment. Therefore, it is hard to explain why those individuals would still favour a risky strategy where they reshuffle their genes with other individuals via recombination (Otto and Lenormand, 2002).

On the one hand, recombination is considered to be very expensive because it is associated with several costs. First, sexual reproduction is believed to disrupt favourable gene combinations, and consequently reduces an individual's fitness (Stearns, 1987; Butlin, 2002). In addition, sexual lineages may have to pay for the substantial twofold cost of sex (Smith, 1978; West et al., 1999): in anisogamous species, only half of lineages are capable of bearing offspring, since males cannot themselves produce offspring, whereas asexual lineages are essentially all females and therefore able to produce twice as many offspring as sexual lineages. Moreover, sexual reproduction is also associated with costs of mating or conjugating. For example, many plant species spend substantial resources on the size of the floral display and nectar rewards (Willmer, 2011).

On the other hand, there is a large body of both theoretical and empirical work to explain the benefits of sex and recombination (Eshel and Feldman, 1970; Hurst and Peck, 1996; Höglund and Sheldon, 1998; West et al., 1999; Otto and Lenormand, 2002;

Butlin, 2002; Engelstädter, 2008; Meirmans and Strand, 2010; Wagner, 2011b; Wang et al., 2015). Most previous work can be classified into two major categories, although they are still controversial, to unravel the mechanisms of the maintenance of sex and recombination (Kondrashov, 1988, 1993; Otto and Feldman, 1997; Otto and Gerstein, 2006; Kouyos et al., 2007; Barton, 2009; Martin and Wagner, 2009). The first major benefit of sexual recombination, in contrast to the disruption of well-adapted lineages, is that recombination can facilitate adaptation by generating novel gene combinations, conferring sexual lineages with a better adaptive potential to new environments, and the second major advantage is that recombination prevents the accumulation of deleterious mutations.

However, the costs and benefits of sex and recombination are still equivocal. For example, the hypothesis that sex enhances the ability to purge deleterious mutations typically assumes synergistic (negative) epistasis. Keightley and Eyre-Walker (2000) tested this hypothesis by estimating genomic point mutation rates for protein-coding genes in a range of animal taxa, and found that sex is not maintained by its capacity to purge the genome of deleterious mutations. Lohaus et al. (2010) also argued that there is no evidence that the long- and short-term advantages to sex are explained by negative epistasis. In addition, Hörandl (2009) showed that the costs for the maintenance of meiotic recombination are expected to be lower. Wagner (2011b) also broadly reviewed mechanisms underlying sexual reproduction in the context of genetic networks, and showed that the destructive role of recombination can be mild or even non-existent. Many other explanations from previous studies have uncovered the maintenance of sex and recombination, such as ecological dynamics (Doncaster et al., 2000), complementation (Archetti, 2004), fluctuating epistasis (Gandon and Otto, 2007), co-evolution (Lively, 2009), fluctuating environments (Misevic et al., 2010) and multiple mating (Rueppell et al., 2012).

Selection is expected to be one of the key factors that help reconcile the paradox of the costs and benefits of sexual reproduction and genetic recombination under certain conditions (Charlesworth, 1993). Banner and Mc Lai (1991) showed the random nature of coronavirus RNA recombination in the absence of selection pressure, but found that RNA recombination is highly restricted due to selection for certain recombinants. Moutouh et al. (1996) showed similarly that the genetic recombinants derived from two distinct viruses can emerge rapidly under selective conditions, and ultimately contribute to the development of HIV-1 resistance to multiple drugs. Lefébure and Stanhope (2007) also emphasised the role of positive selection in the adaptation of the core-genome of different *Streptococcus* species to different hosts. A more recent study by Lumley et al. (2015) showed that sexual selection helps purify deleterious alleles to reduce mutation load, and consequently facilitates fixation of advantageous alleles, enhancing population survivability in the presence of genetic stress.

Although many existing studies have indicated that natural selection is critical to the maintenance of meiotic recombination, they have not explicitly considered how selection pressure affects the underlying evolutionary dynamics when recombination results in rewired gene regulatory networks. In this chapter, I hypothesise that selection pressure can shape the complex hierarchical representations found in the genome and facilitate a rate of evolution sufficient to compensate both the recombination cost and the twofold cost. Here, I use again the well-established computational approach of Wagner's GRN model to assess the costs and benefits of sex and recombination in a gene regulatory network context (Wagner, 1996; Siegal and Bergman, 2002; Azevedo et al., 2006), since traditional genetic models are unable to investigate multiple interactions simultaneously. In the first study, I find that low-fitness sexual lineages can gain a higher benefit when they are subjected to higher selection pressure, especially at the early stage. In the second study, I present a population-dynamics view of competition between asexual lineages (parthenogenetic species) and sexual lineages (anisogamous species), in which both recombination cost and twofold cost are explicitly modelled in the system. I find that although recombination is initially costly, it rapidly evolves — through rewiring gene regulation — to compensate in even a single bout for the costs of sex and recombination. I further explore the parameter space and find that sexual lineages with low levels of sex and recombination can outcompete strictly asexual populations under higher selection pressure and a lower mutation rate. These results indicate a key role of selection pressure in reducing mutation load as well as mitigating costs of sex and recombination, and have important implications for explaining the maintenance of sexual reproduction in the context of genetic networks.

6.2 Methods

In the modelling approach, asexual and sexual populations were evolved under both purifying selection, i.e., selection for phenotypic stability (see Section 2.3.6) and target selection, i.e., selection for target phenotype (see Section 2.3.7). The system-level parameters were fixed to be $a = 1$, $devT = 100$ and $\tau = 10$ in all simulations. Note that in all simulations presented in this chapter, individuals were subjected to purifying selection at each generation. In other words, networks that could not achieve phenotypic stability were eliminated from the population pool immediately. This is different from the previous Chapters 3, 4 and 5 where unstable networks could survive in certain evolutionary scenarios.

6.2.1 The computational model

The computational model was similar to that introduced in Section 2.3. Lineages were typically cloned to reproduce offspring or allowed to recombine with each other

during periodical sexual reproduction events. Here, an event of sexual reproduction refers to only having one generation of recombination in the population.

Fitness evaluation

Fitness was evaluated by measuring the phenotypic distance between the equilibrium state and the optimal state. Here, the optimal phenotype $\mathbf{s}_{\mathrm{OPT}}$ was set to be the initial gene expression pattern $\mathbf{s}(0)$. Unless otherwise specified, I used Equation (2.4) to calculate individual fitness[1]. For individuals that could not achieve phenotypic stability, a zero fitness was assigned to ensure that no unstable networks could survive in the subsequent generation.

Initialisation

Each individual network in population M was generated by randomly filling W with $\lfloor c \times N^2 \rfloor$ non-zero elements $w_{i,j}$ drawn from the standard normal distribution, $N(0, 1)$. The associated initial expression state for each network $\mathbf{s}(0)$ was simply setting $s_i(0) = +1$ $(i = 1, 2, \ldots, N)$.

Mutation

Unless otherwise specified, I used the same mutation operator as described in Section 5.2.1.

Recombination

The recombination operator was the same as described in Section 2.3.5.

Stability and fitness selection

All individuals were subjected to two layers of selection — selection for phenotypic stability (as defined in Section 2.3.6) and selection for target phenotype (as defined in Section 2.3.7).

Evolution

The evolutionary simulations were performed under the reproduction-mutation-selection life cycle similarly to how it was described in Section 4.2.1. In typical evolution, an individual was chosen at random to reproduce by cloning itself, if asexually,

[1] In some simulations, I used Equation (2.3) to calculate an individual's fitness. The difference between the two measurements can be found in Section 2.3.7. Generally, the multiplicative measurement, Equation (2.4), has a higher-resolution span of parameter σ than the exponential measurement, Equation (2.3). Note that the larger values of σ used in Equation (2.4) represent a stronger selection pressure, which is the opposite in Equation (2.3).

or by randomly recombining with another individual, if sexually, then the resulting network was subjected to mutation, followed by two layers of selection: selection for phenotypic stability and selection for the optimal phenotype. This process was repeated until M number of networks were produced.

6.2.2 Exploring effects of selection pressure on low-fitness individuals

In the first set of experiments, I investigated how different levels of selection strength benefit low-fitness individuals in both asexual and sexual lineages (see Figures 6-1 and 6-2). Specifically, both asexual and sexual lineages were derived from the same population pool which contained $10,000$ randomly generated stable networks. Note that all networks had the same initial gene expression pattern, all activation, i.e., $s_i(0) = +1, i = 1, \ldots, N$. Next, the population was evolved for one generation with asexual or sexual reproduction followed by one single mutation for each network. In other words, for each network, exactly one non-zero entry was mutated.

In the asexual population, for each individual, I recorded each individual's parental fitness at the initial generation as well as its offspring's fitness in the subsequent generation. Similarly, for the sexual population, I also recorded offspring fitness, but parental fitness was estimated as the mean fitness of the two parents at the initial generation. Next, each of the two (asexual and sexual) populations was grouped into ten bins according to parental fitness (in ascending order). Finally, the proportion of gained fitness for each individual's offspring relative to the corresponding parental fitness was measured and averaged for all individuals in each of the ten bins. For both the asexual and sexual populations, I also randomly selected $1,000$ individuals from one simulation run and plotted each individual's parental and offspring fitness on an actual phenotypic distance scale, as shown in Figures E-1 and E-2. Note that by calculating the phenotypic distance as described in Section 2.3.7, results obtained from different levels of selection strength can be displayed on the same scale.

In the second set of experiments, I further investigated how different levels of selection strength benefit evolved asexual and sexual lineages (see Figures 6-3 and E-3). Specifically, similar to the first set of experiments, the population was evolved asexually or sexually under selection pressure $\sigma = 100$, and I recorded each individual's fitness at the initial, 4^{th}, 9^{th} and 49^{th} generations, as well as its offspring's fitness in the subsequent generation, i.e., at the 1^{st}, 5^{th}, 10^{th} and 50^{th} generations. Finally, for each of four categories for both the asexual and sexual lineages, the proportion of gained fitness for each individual in its offspring relative to the corresponding parental fitness was measured similarly to the calculations in the first set of experiments.

6.2.3 Exploring effects of selection pressure on recombination cost

In this set of experiments, I tested whether sexual lineages are able to afford the recombination cost incurred by selection pressure (see Figures 6-4 and 6-5). Specifically, an initial population of $10,000$ randomly generated stable networks was evolved with different recombination frequencies (from recombination occurring at each generation to no recombination at all) for $1,000$ generations under extremely weak ($\sigma = 10^9$) or strong selection pressure ($\sigma = 0.5$). Individual phenotypic distance from the optimum, i.e., $D(\mathbf{s}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}})$, was measured at each generation in all evolutionary scenarios. Note that in this set of experiments, I used Equation (2.3) to calculate individual fitness.

6.2.4 Modelling recombination cost and twofold cost in a competitive regime

In addition to the recombination cost incurred by selection pressure, in this set of experiments, I introduced the twofold cost of sex in a competitive regime, and tested whether sexual lineages can outcompete asexual lineages under certain conditions (see Figures 6-6, 6-7 and 6-8). Specifically, the initial population contained $10,000$ randomly generated stable networks with an equal frequency of asexual and sexual lineages ($5,000$ individuals in each category[2]). Asexual lineages could only reproduce by cloning themselves. Sexual lineages, when there was no recombination event, followed the reproduction mode of asexual lineages. However, when recombination happened, sexual lineages were randomly divided in half and assigned transient 'female' and 'male' labels with an equal number. Only individuals with female labels were allowed to recombine with males to reproduce offspring. The asexual and sexual lineages competed against each other in a population pool which could hold a fixed number of $10,000$ individuals. In a typical competition round, an individual was randomly selected from the population pool. If the selected individual was from the asexual population, then the individual was cloned and subjected to mutation followed by selection (two layers of selection); whereas if the selected individual was from the sexual population and was also labelled as a female, it was allowed to recombine with a randomly selected male, then the recombinant was similarly subjected to mutation followed by selection. This process was repeated until $10,000$ offspring were selected. Note that in this set of experiments, I used Equation (2.3) to calculate individual fitness. The twofold cost of sex was modelled in a way such that sexual lineages only had half the chance to be selected to reproduce offspring than asexual lineages in the population. Note that when there was no recombination occurring in the sexual lineages, both individuals with 'female' and 'male' labels were allowed to reproduce offspring by cloning themselves. In other words, the twofold cost of sex was only considered whenever recombination

[2]Note that $5,000$ sexual lineages were cloned from asexual lineages, forming a total of $10,000$ individuals in the initial population.

occurred.

6.2.5 Exploring conditions for benefits of sex and recombination recouping costs of sex

In this set of experiments, I thoroughly explored how selection pressure along with recombination frequency and mutation rate affects the winning probability of sexual lineages competing against asexual lineages in the face of both recombination cost and twofold cost (see Figure 6-9). Specifically, in order to avoid the effects of perturbations such as drift on the competition results, instead of dividing the population into asexual and sexual lineages at the very first generation, as described in Section 6.2.4, two categories of lineage were differentiated at the first recombination event (by randomly selecting half of the population as asexual and the other half as sexual). In other words, the whole population was evolved by accumulating mutations regardless of sexuality before the first recombination event, and both the asexual and sexual populations had the same number of $5,000$ individuals when the twofold cost of sex was introduced into the model. For each competition trial, the whole population was allowed to evolve for a total of 500 generations. If the number of sexual lineages was greater than the number of asexual lineages at the end of evolution, the sexual population won, and otherwise the asexual population won. For each parameter combination (selection pressure, recombination frequency and mutation rate), the winning probability of sexual lineages was recorded based on 100 independent competition runs. Note that I used the same population pool for all competition trials. The complete results can be found in Table E.1.

6.3 Results

6.3.1 Strong selection pressure benefits low-fitness sexual lineages

I first investigated the effects of different levels of selection pressure on individuals' fitness. I found that low-fitness sexual lineages benefit most when the population is subjected to strong selection strength for the target phenotype. Specifically, I compared the gained fitness of offspring in proportion to their parental fitness for both asexual and sexual lineages evolved under different levels of selection pressure. From Figure 6-1, we can see that only lineages that have been classified into the group with the lowest fitness (the first bin) in the asexual population can slightly benefit when the selection pressure is sufficiently strong, whilst for the rest of the asexual lineages, the benefit of higher selection pressure is largely absent. In contrast, from Figure 6-2, we can clearly see that the group of lineages with the lowest fitness (the first bin) in the sexual population substantially benefits under a strong selection regime ($\sigma = 100$). We can also see that groups of sexual lineages with lower fitness generally gain a benefit

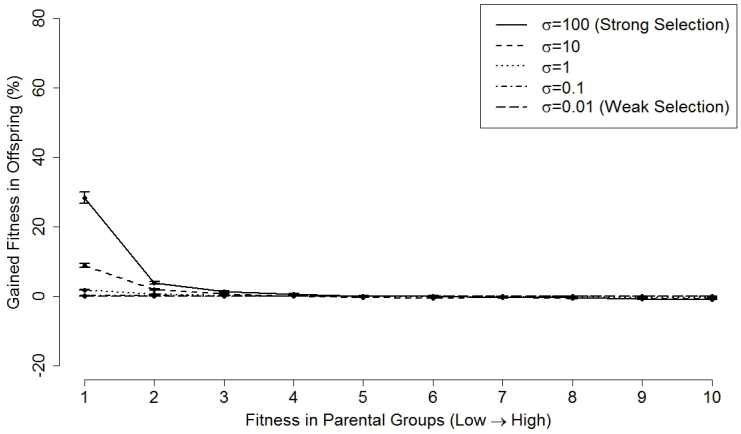


Figure 6-1: Comparison of effects of different levels of selection pressure on offspring fitness in asexual lineages. *I first collected an initial population pool of* $10,000$ *randomly generated stable networks (*$N = 10$ *and* $c = 0.75$*). Then, I recorded each individual's initial fitness and its offspring's fitness after evolving asexually for one generation under different selection pressure:* $\sigma = 100$ *(strong),* 10*,* 1*,* 0.1 *and* 0.01 *(weak), and grouped all individuals into ten bins based on their parental fitness in ascending order. Next, for each of ten bins, I calculated the mean gained fitness of offspring in proportion to their corresponding parental fitness. The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

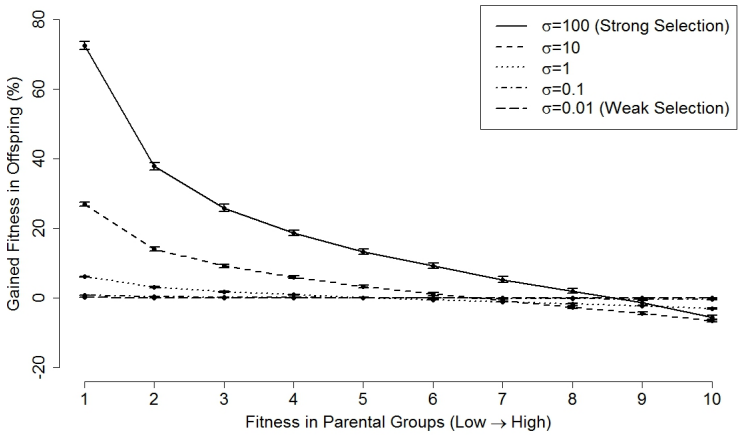


Figure 6-2: Comparison of effects of different levels of selection pressure on offspring fitness in sexual lineages. *I first collected an initial population pool of* $10,000$ *randomly generated stable networks (*$N = 10$ *and* $c = 0.75$*). Then, under different selection pressure:* $\sigma = 100$ *(strong),* 10*,* 1*,* 0.1 *and* 0.01 *(weak), the population was evolved sexually for one generation, and I recorded each offspring's fitness as well as the mean initial fitness of its two parents as the estimated parental fitness. All individuals were grouped into ten bins based on their parental fitness in ascending order. Next, for each of ten bins, I calculated the mean gained fitness of offspring in proportion to their corresponding parental fitness. The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

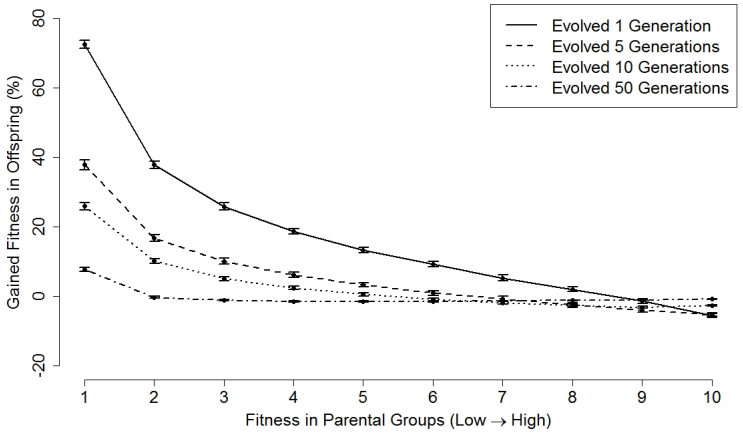


Figure 6-3: Comparison of gained fitness in evolved sexual lineages under strong selection pressure. *I used the same population pool of* $10,000$ *randomly generated stable networks with size* $N = 10$ *and connectivity* $c = 0.75$ *as described in Figure 6-2. The population was evolved sexually under selection pressure* $\sigma = 100$*. Then, I recorded each individual's fitness at the initial,* 4^{th}*,* 9^{th} *and* 49^{th} *generations as well as its offspring's fitness in the subsequent generation, i.e., at the* 1^{st}*,* 5^{th}*,* 10^{th} *and* 50^{th} *generations. I then calculated the mean gained fitness of offspring in proportion to their corresponding parental fitness for each of four categories in which all individuals were sorted and grouped similarly as described in Figure 6-2. The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

from selection to a magnitude depending on its strength. However, the magnitude of benefit for low-fitness sexual individuals generally reduced when I further studied the proportion of fitness gained in the evolved population (see Figure 6-3). This is because, after many generations of recombination, the sexual lineages have become well adapted to the environment, approaching the optimum. It should also be noticed that although strong selection strength slightly deteriorates high-fitness lineages at the early stage (see Figure 6-2), it becomes beneficial in the evolved population (cf. the last bin in Figure 6-3). This supports the pattern I showed in Chapter 5 that recombination will not disrupt well-adapted lineages when they are close to the optimum. Taken together, these results help explain why some species increase their recombination rate or switch from asexual reproduction to sexual reproduction mode when they are subjected to certain extreme environments such as in the face of pathogen infection (Haldane, 2006).

6.3.2 Benefits of sexual reproduction make the recombination cost incurred by selection pressure affordable

In Chapter 5.3.1, I showed that sexual lineages evolve to be insensitive to mutational perturbations even when selection for the optimal phenotype (the individual's initial expression state) is largely absent. Here, I further investigated the recombination cost incurred by selection pressure. I found that selection pressure can increase the benefits

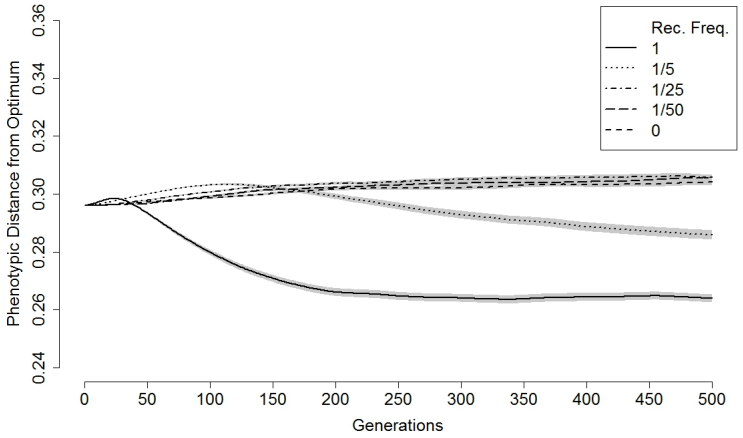


Figure 6-4: Phenotypic distance of the population evolved under extremely weak selection pressure. *I first collected an initial population pool of* $10,000$ *randomly generated stable networks with size* $N = 10$ *and connectivity* $c = 0.75$*. Then, the population was evolved with a recombination frequency at* 1 *(recombination occurring at each generation),* $1/5$*,* $1/25$*,* $1/50$ *and* 0 *(no recombination at all) under extremely weak or even absent selection (*$\sigma = 10^9$*) for the target phenotype. Note that, for each generation where there was no recombination happening, individuals reproduced asexually. Individual fitness was calculated using Equation (2.3), and the mutation rate was set to be* $\mu = 0.1$*. The shaded areas represent* 95% *confidence intervals based on* 10 *independent runs.*

of sexual reproduction, which are sufficient to compensate for the recombination cost. Specifically, I measured the phenotypic distance between the optimum and population that was evolved with different recombination frequencies under extremely weak selection ($\sigma = 10^9$) and strong selection ($\sigma = 0.5$) regimes. From Figure 6-4, we can see that when the selection pressure is extremely weak or even absent, the recombination should be sufficiently frequent (occurring at each generation or every 5 generations) to be able to drive the population towards the optimum. Otherwise, if the recombination is less frequent or absent, then the population is unable to move towards or even slightly deviate away from the optimum. Note that when the population is evolved under extremely weak selection, there is no recombination cost, or it can be largely neglected. This is because the differences in phenotypic distance between the individual and the optimum will not affect its fitness calculated using Equation (2.3), since the selection pressure is set to be $\sigma = 10^9$. However, as shown in Figure 6-5, when selection strength is strong, we expect to see the population is able to move more rapidly towards the optimum. We can also see that periods of recombination in sexual lineages are sufficient to drive evolution faster than asexual lineages (no recombination). Note that the ragged curves with recombination frequency at 1/5, 1/25 and 1/50 appearing in Figure 6-5 clearly show the recombination cost, which is the disruption of well-adapted lineages. These results suggest that bouts of recombination are enough to offset the cost incurred by

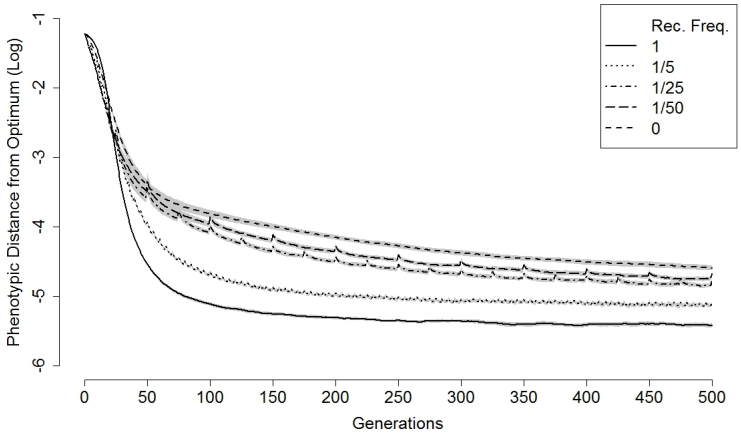


Figure 6-5: Phenotypic distance of the population evolved under strong selection pressure. *I first collected an initial population pool of* $10,000$ *randomly generated stable networks with size* $N = 10$ *and connectivity* $c = 0.75$*. Then, the population was evolved with recombination frequency at* 1 *(recombination occurring at each generation),* $1/5$*,* $1/25$*,* $1/50$ *and* 0 *(no recombination at all) under strong selection (*$\sigma = 0.5$*) for the target phenotype. Note that, for each generation where there is no recombination happening, individuals reproduced asexually. Individual fitness is calculated using Equation (2.3), and the mutation rate is set to be* $\mu = 0.1$*. The shaded areas represent* 95% *confidence intervals based on* 10 *independent runs.*

selection pressure.

6.3.3 Selection pressure can be beneficial for affording the costs of sex under certain conditions

In Section 6.3.2, I showed that the benefits of sexual reproduction are sufficient to afford the recombination cost incurred by selection pressure. Here, I explored whether the benefits are enough to accommodate the twofold cost of sex in a competitive regime. I found that, under certain conditions, sexual lineages can outcompete asexual lineages despite the recombination cost and the twofold cost. Figure 6-6 shows the frequency of sexual lineages in the population in the first 150 generations. Note that this is part of the results presented in Figure 6-7, where asexual and sexual lineages competed against each other for a total of 500 generations. From Figure 6-6, we can see that when a single bout of recombination occurred at the 50^{th} and 100^{th} generations, the frequency of sexual lineages immediately reduced due to the recombination cost (indicated by a red arrow) and the twofold cost (indicated by a blue arrow). To be more specific, on the one hand, the recombination cost was caused by disrupting well-adapted sexual lineages. On the other hand, the twofold cost was explicitly modelled in the competition where only half of sexual lineages were able to reproduce offspring. This mimics the phenomenon in most multicellular sexual species where only females are capable

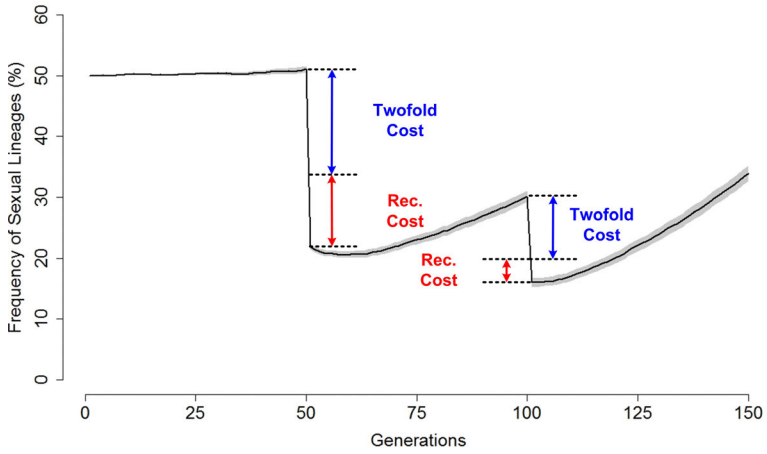


Figure 6-6: Visualising recombination cost and twofold cost in a competitive regime. *A total number of* $10,000$ *individuals (*$5,000$ *asexual lineages and* $5,000$ *sexual lineages) were evolved and competed against each other for* 500 *generations (see Figure 6-7 for more details). When recombination occurred at the* 50^{th} *and* 100^{th} *generations, the reduced frequency of sexual lineages in the population was due to two costs — recombination cost (in blue) and twofold cost of sex (in red). The recombination cost was modelled in the situation where recombination disrupts well-adapted sexual lineages. The twofold cost of sex was modelled in the situation where only half of sexual lineages, if selected, were allowed to reproduce offspring. Selection strength* $\sigma = 1$*, and mutation rate* $\mu = 10^{-4}$*. Note that I used Equation (2.3) to calculate individuals' fitness. The shaded areas represent* 95% *confidence intervals based on* 46 *sexual winning trials of total* 50 *independent competition runs.*

of bearing offspring, whilst males cannot themselves produce offspring (Smith, 1978; Stearns, 1987). However, we can also clearly see that after the first single bout of recombination, the frequency of sexual lineages increased, although there were only about 20% of sexual lineages in the population at the 51^{st} generation. Both recombination cost and twofold cost became smaller in the second bout of recombination happening at the 100^{th} generation. It should be noted that the twofold cost of sex is modelled constantly associated with recombination, but this can be reduced, because the reproductive output (fitness) is higher in sexual lineages than in asexual lineages (see Figure 6-8). In other words, although asexual lineages have a higher chance to be selected for reproduction, especially at the earlier stage, whereas only half of sexual lineages can be selected for reproduction, sexual lineages are still likely to survive in the subsequent generation if the recombinants generally have a higher fitness than asexual offspring. From Figure 6-8, we can also notice that the recombination cost in reducing the fitness indicated by the immediate drops in sexual lineages also decreased during evolution. Taken together, these results suggest that both of recombination cost and twofold cost can be minimised, and benefits arising from sexual reproduction are able to facilitate a rapid adaptation and ultimately help sexual lineages resist invasion by

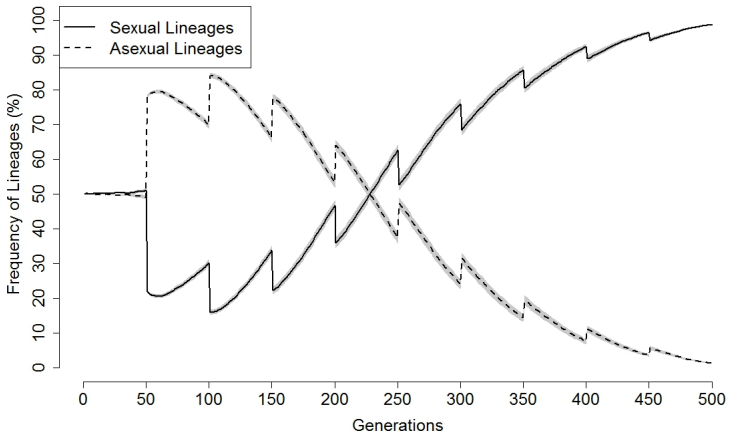


Figure 6-7: Frequency of asexual and sexual lineages in competition. *Both asexual and sexual lineages were cloned from a pool of* $5,000$ *randomly generated stable networks (*$N = 10$ *and* $c = 0.75$*), total* $10,000$ *individuals in the initial population for competition (*$\sigma = 1$ *and* $\mu = 10^{-4}$*). Then, asexual and sexual lineages competed against each other for* 500 *generations. When recombination occurred (in every* 50 *generations) in sexual lineages, only half of lineages were allowed to reproduce offspring, whereas when there was no recombination, both asexual and sexual lineages could be selected with a probability in proportion to their total number in the population to reproduce offspring by cloning themselves. The resulting offspring were then subjected to mutation followed by selection until* $10,000$ *individuals were selected for the next generation. The frequency of asexual and sexual lineages was recorded at each generation. The shaded areas represent* 95% *confidence intervals based on* 46 *sexual winning trials of total* 50 *independent competition runs.*

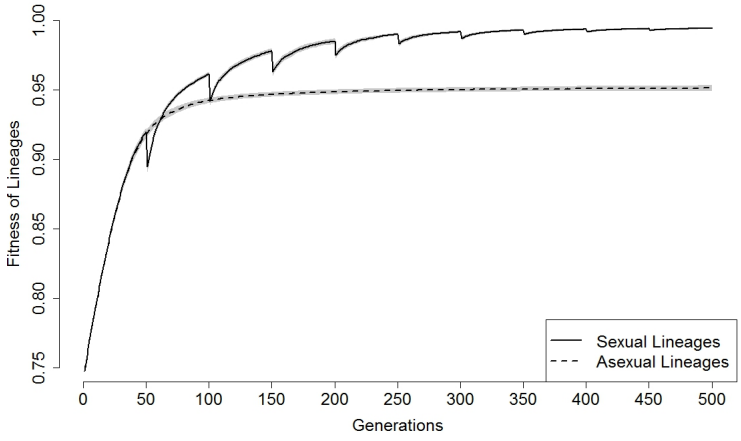


Figure 6-8: Fitness of asexual and sexual lineages in competition. *As with the results shown in Figure 6-7, I also measured the fitness of lineages during competition. Note that individual fitness was calculated using Equation (2.3). The shaded areas represent* 95% *confidence intervals based on* 46 *sexual winning trials of total* 50 *independent competition runs.*

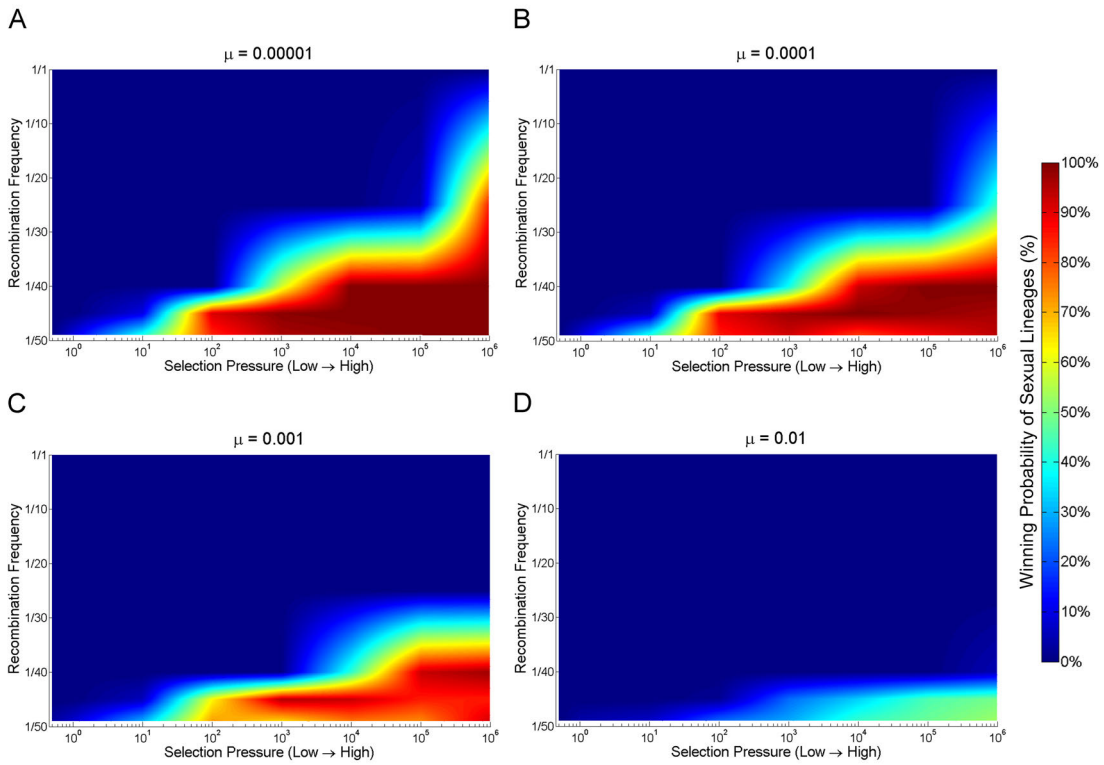


Figure 6-9: The influence of selection pressure and recombination frequency on competition outcomes. *The asexual population contained* $5,000$ *randomly generated stable networks (*$N = 10$ *and* $c = 0.75$*), and was cloned to form the same number of sexual population, total* $10,000$ *individuals in the initial population pool. Then, asexual lineages competed against sexual lineages for total* 500 *generations under different selection pressures:* $\sigma = 0.5$ *(weak),* 1*,* 10*,* 10^2*,* 10^3*,* 10^4*,* 10^5 *and* 10^6 *(strong), and different recombination frequencies:* $f_{Rec.} = 1/50$*,* $1/25$*,* $1/10$*,* $1/5$ *and* $1/1$*. I also performed similar competition simulations using different mutation rates:* $\mu = 10^{-5}$ ***(A)****,* $\mu = 10^{-4}$ ***(B)****,* $\mu = 10^{-3}$ ***(C)*** *and* $\mu = 10^{-2}$ ***(D)****. If the number of sexual lineages was greater than the number of asexual lineages at the end of evolution, then sexual lineages won, otherwise, asexual lineages won. The winning probability of sexual lineages was recorded based on* 100 *independent competition runs. The surface was generated using linear interpolation. The complete results can be found in Table E.1.*

asexual lineages.

Next, I explored the parameter space to investigate how the recombination cost incurred by selection pressure and twofold cost incurred by recombination frequency affect competition outcomes. I also examined the competition results under different mutation rates. I found that generally asexual lineages are more likely to outcompete asexual lineages when selection pressure is higher and recombination is less frequent under a lower mutation rate. Specifically, starting with an equal frequency (50%), asexual lineages and sexual lineages competed against each other in a fixed space which can hold $10,000$ individuals for a total of 500 generations. Figure 6-9 shows the competition outcomes for each combination of parameters (selection pressure, recombination

frequency and mutation rate) based on 100 independent competition runs. As can be seen from Figure 6-9, recombination benefits facilitated by selection pressure are generally able to afford both the recombination cost caused by selection pressure itself and the twofold cost caused by recombination frequency under higher selection pressure and lower recombination frequency. However, the results show that asexual lineages can outcompete obligate sexual lineages where recombination occurs in every generation, i.e., $f_{Rec.} = 1/1$. It should be also noted that a lower mutation rate can also help sexual lineages to outcompete asexual lineages. This is probably because higher mutation rates are more likely to disrupt well-adapted recombinants.

6.4 Discussion

Sexual reproduction prevails in animals, plants and even fungi. Although a large number of theories have been proposed to explain the maintenance of sex and recombination, it remains a great puzzle in evolutionary biology (Lehtonen et al., 2012). Previous work has shown that recombination rates can be increased in organisms when they are subjected to higher selection pressure. For example, Zhong and Priest (2011) and Zhong (2013) exposed *Drosophila melanogaster* to mating stress, heat shock and cold shock, and found that each stress treatment can increase the rate of recombination. Jackson et al. (2015) also showed that the recombination rate is increased in *Drosophila melanogaster* in response to parasite infection. In this chapter, I have shown that low-fitness sexual lineages can greatly benefit from recombination in the presence of strong selection pressure (Figure 6-2), especially at the early stage. This may help explain the benefits of recombination in terms of facilitating low-fitness sexual lineages to adapt to new environments under stress.

In Chapter 5, I showed that recombination together with selection for phenotypic stability can drive sexual lineages towards the optimum, even in the absence of selection for an optimal phenotype, but this pattern can only be observed when recombination is sufficiently frequent (Figure 6-4). However, it is still not clear whether these benefits can compensate for the recombination cost, since selection pressure for the target phenotype is extremely weak or even absent in the simulations presented in Chapter 5. When a population evolves under high selection pressure, the recombination cost cannot be neglected. If the recombinant deviates away from the optimum, then its fitness reduces dramatically if the individual is subjected to high selection pressure. In this chapter, I have shown that the benefits of recombination are able to offset the recombination cost (Figure 6-5). In fact, periods of recombination are sufficient to afford such a cost inured by selection pressure in sexual lineages.

In the later competition study, I explicitly modelled both the recombination cost and the twofold cost into the system to investigate whether the benefits of recom-

bination are sufficient to accommodate the two costs. Specifically, the competitive advantage of asexual lineages relative to sexual lineages[3], i.e., the twofold cost of sex, is associated with recombination frequency — wherever recombination happens in sexual lineages, they have to pay for the cost such that only half of the population is allowed to produce offspring. I have shown that sexual lineages with less frequent recombination can outcompete asexual lineages under high selection pressure (Figures 6-7 and 6-9). In addition, higher mutation rates also reduce the winning probability of sexual lineages (Figure 6-9). This may be consistent with previous work stating that sexual reproduction will be favoured with a lower level of mutation rate (Agrawal, 2002; Agrawal and Wang, 2008). This also suggests that although recombination can massively alter patterns of gene regulation, it is essentially different from hypermutation, as shown in Chapter 5. It should be noted that population size, although it has not been thoroughly explored in this chapter, is expected to affect the winning probability, as indicated in Le Cunff and Pakdaman (2014). Note that sexual lineages modelled in this chapter only have periodic recombination. Here, the model mimics alternation between sexual and asexual reproduction, which is biologically realistic. For example, the freshwater *Daphnia magna* reproduces by parthenogenesis in the spring, then switches to sexual reproduction mode when the intensity of competition or predation increases (Ebert, 2005).

The deterministic mutation hypothesis for explaining the maintenance of sexual reproduction speculates that recombination can help purge deleterious mutations more effectively (Kondrashov, 1988). This is because the theory typically assumes that deleterious mutations display synergistic epistasis, causing a profound reduction in fitness via recontamination, and consequently are more likely to be eliminated by natural selection. Azevedo et al. (2006) reported supportive simulation results that synergistic epistasis can evolve as a by-product of selection for genetic robustness in sexual lineages in the context of genetic networks. However, many studies have challenged this deterministic mutation hypothesis. For example, MacCarthy and Bergman (2007a) introduced a recombination modifier to the Wagner GRN model and found that the emergent synergistic epistasis cannot explain the maintenance of sexual reproduction. Lohaus et al. (2010) also examined the hypothesis, and confirmed that there is no evidence that the long- and short-term advantages of sex and recombination cannot be explained by synergistic epistasis. In fact, in Chapter 5, I also showed that recombination can rapidly purge weaker configurations even when selection is largely absent. This pattern is expected to be particularly evident when the mutation rate is higher, as indicated in Figure 6-9. In the competition simulations presented in this chapter, the epistasis has not been explicitly measured. However, it is expected that the competition results cannot be explained by synergistic epistasis, since sexual lineages only

[3]Note that here I do not consider sexual hermaphrodites.

have periods of recombination, so synergistic epistasis may not exist or can be largely neglected.

If the capability to effectively reduce mutation load in sexual lineages cannot be explained by synergistic epistasis due to the lack of evidence that it can be evolved to a sufficient level, then alternative explanations for costly sexual reproduction are needed. Becks and Agrawal (2012) used experimental populations of a facultatively sexual species of rotifer *Brachionus calyciflorus* to show that although recombination breaks up well-adapted gene combinations, and consequently reduces the mean fitness in offspring, sexual reproduction can generate offspring with more variable fitness, allowing for faster adaptation. In this chapter, I have also provided simulation results to support this empirical study (Figures 6-8 and E-2). Many previous studies have also indicated that non-random mating can alter reproductive success in the face of competition or choice to help purge deleterious mutations (Agrawal, 2001; Siller, 2001; Whitlock and Agrawal, 2009; Lumley et al., 2015). The competition results from Figure 6-9 may also imply that it is non-random mating that helps sexual lineages to outcompete asexual lineages. This is because one of the reasons that the sexual population is more likely to win under substantial selection pressure is that only certain recombinants are able to reach the threshold imposed by selection, whereas it is impossible for asexual lineages to pass through the selection barrier via mutation only.

6.5 Summary and future work

In this chapter, I have investigated how selection pressure benefits sexual lineages and mitigates recombination cost and twofold cost. Specifically, I have shown that strong selection pressure can greatly help sexual lineages with a lower fitness, especially at the early stage, whereas low-fitness asexual lineages generally will not gain benefits from selection. I have also shown that bouts of recombination can substantially increase the benefits of sexual lineages and sufficiently compensate for the recombination cost incurred by selection pressure. I have designed an evolutionary scenario where sexual lineages compete against asexual lineages in a fixed space, and found that, under certain conditions, although recombination is initially costly, it can rapidly evolve to compensate for the costs of sex and recombination. I have further explored the parameter space and found that sexual lineages with low levels of sex and recombination can outcompete strictly asexual populations under higher selection pressure and a lower mutation rate. Some possible future research directions regarding measuring transition probabilities in the competitive regime and investigating conditions under which sexual lineages can still outcompete asexual lineages when asexual lineages have gained the same benefits as sexual lineages after evolution are presented below.

6.5.1 Measuring transition probabilities in a competitive regime

In Chapter 5, I measured the transition probabilities of evolved populations to show how sexual lineages can evolve to have a greater benefit than asexual lineages. It would be interesting to perform similar simulations to measure the transition probabilities of asexual and sexual lineages in the competition presented in this chapter. We could then rigorously explore how these transition probabilities evolve under conditions in which sexual lineages win or lose. We would also be able to investigate how mutation rates affect these transition probabilities in asexual and sexual lineages. As I have shown in Chapter 5 that recombination can help purge weaker configurations more rapidly, it is reasonable to conjecture that the reason that asexual lineages are more likely to win under a higher mutation rate is that more low-fitness (unstable) sexual lineages are purged by purifying selection (selection for phenotypic stability) at the early stage, such that the benefits of sexual reproduction cannot afford the substantial twofold cost of sex.

6.5.2 Exploring how sexual reproduction can be maintained once evolved

In this chapter, I have shown that sexual lineages can outcompete asexual lineages under certain conditions. However, it is still not clear how sexual reproduction can be favoured in the face of invasion by asexual lineages that are derived from sexual lineages. In other words, if asexual lineages have gained the same benefits from the evolved sexual lineages, then how can sexual reproduction still be maintained? It is natural to envision that if both sexual and asexual lineages still compete against each other once they have evolved in the same environment, the asexual population is more likely to win, since both asexual and sexual lineages have evolved close to the optimum, but sexual lineages still have to pay for the twofold cost of sex. Therefore, it would be interesting to explore how selection pressure, frequency of recombination and mutation rate affect the maintenance of sexual reproduction in changing environments, since previous work has indicated that fluctuating environments can facilitate rapid adaptation (Draghi and Wagner, 2009; Misevic et al., 2010; Tsuda and Kawata, 2010; Le Cunff and Pakdaman, 2014; Wang et al., 2014a). It would also be interesting to perform simulations using different mating strategies or track successful recombinants to thoroughly examine, for example, the role of sexual selection on the maintenance of sex in the context of genetic networks.

Chapter 7

Conclusions

> Laws of variation were barely conjectured; the different types of variability were only imperfectly distinguished. The breeders' conception was fairly sufficient for practical purposes, but science needed a clear understanding of the factors in the general process of variation.
>
> Hugo Marie de Vries

In this dissertation, I have mainly focused on studying two important evolutionary forces, mutation and recombination, in the context of genetic networks. Both mutation and recombination operate on the genotype which represents regions of non-coding DNA (*cis*-regulatory elements) that regulate the transcription of nearby genes. By manipulating the genotype, both mutation and recombination can change patterns of gene activities or expression concentrations encaptured in the phenotype, and consequently alter the underlying evolutionary dynamics and drive new evolutionary innovations, forming novel macroscopic traits or physiological states.

In Chapter 2, all currently available research papers using the Wagner GRN model have been reviewed in Section 2.2. I have described the implementation details of Wagner's GRN model as well as its variants in Section 2.3, because a similar version of the model has been extensively employed in the reminder of the chapters. The proprieties of stability, robustness and path length (stabilisation time) in initial populations have been investigated in Section 2.5. Generally, initial stability is higher in smaller networks than in larger networks. Networks with low levels of connectivity are more likely to be stable than networks with high levels of connectivity. A similar conclusion has also been applied to the robustness of initially stable networks — smaller networks with sparser connectivity have higher initial robustness. The path length of initially stable networks has also been observed to be shorter in smaller networks. Larger networks are more likely to have a longer path length when network connectivity is higher, but path length tends to be constant for smaller networks regardless of connectivity. The pa-

rameter a, which indicates the sensitivity of regulatory response to output phenotypes, can quantitatively affect the results of initial stability, robustness and path length, but general patterns still hold. From all of the conducted experiments, we can see that generally the results are insensitive to different parameters used in the Wagner GRN model.

The conclusions drawn from the rest of the chapters are summarised separately below to answer the research questions raised in Chapter 1.

What are the characteristics of compensatory mutations?

In Chapter 3, I have shown that there is a relatively high probability that a compensatory mutation will fix a broken network caused by a deleterious mutation in the previous generation. For smaller networks, the frequency of compensatory mutation continuously increases as network connectivity increases. In contrast, for larger networks, with the rise of connectivity, the frequency of compensatory mutation decreases slightly. The results indicate that the frequency of individuals that can be fixed by compensatory mutation is more sensitive to network size than to network connectivity. However, the overall result is marked as relatively scale invariant in contrast to the scale dependencies of deleterious mutation in initial stability and robustness, as shown in Chapter 2. In addition, compensatory mutations are more likely to occur at or close to the site of the original, deleterious mutation, and are also more likely to be driven by large-effect mutations. These general patterns are very different from those observed in networks with neutral mutations. Specifically, on the one hand, neutral mutations are more likely to be distributed evenly in terms of location in smaller networks, and the neutral mutations tend to be enriched if they are far apart in larger networks. On the other hand, small-effect mutations are more likely to be observed in networks with neutral mutations. These findings show that compensatory mutations have unique properties compared with neutral mutations, and indicate that gene pathway evolution may be far less constrained than previously considered.

How do compensatory mutations contribute to evolutionary complexity?

In Chapter 4, I have shown that compensatory mutation can continue to occur even after evolving for many generations under both strong and relaxed selection for phenotypic stability. Even in seriously damaged networks that have accumulated deleterious mutations for many generations, compensatory mutations are still able to fix those compromised networks. In fact, the more bouts of relaxed selection the population has been exposed to, the more compensatory mutations can be found. The characteristics of compensatory mutations discovered in Chapter 3 are also expected to affect the evolutionary consequences of networks with compensatory mutations. Specifically,

robustness is far higher both when the compensatory mutation occurs closer to the original deleterious mutation site and when the compensatory mutation has a larger shift in gene regulation. These general patterns, however, are observed differently in networks with neutral mutations. Specifically, on the one hand, closer distances are weakly associated with higher robustness, and in fact, robustness tends to be higher in neutral mutations when they are far apart. On the other hand, large-effect mutations cannot generate a profound change in robustness. This location- and size-specific robustness systematically biases which networks are lost by selection for phenotypic stability in subsequent generations, which, over time, can drive regulatory complexity in terms of increasing the network connectivity of the entire population. This pattern has been observed in two independent cases: 1) initial networks with connectivity variance but fixed structure, and 2) networks with structure variance but fixed connectivity. These findings are important because they provide an explanation of how major features of genome organisation, development and biodiversity can emerge through non-adaptive processes.

Why can sexual lineages evolve greater benefits?

In Chapter 5, I have shown it is the combination of recombination and selection for phenotypic stability that can drive the population towards the optimum, even in the absence of selection for such an optimal phenotype (an individual's initial expression state). This conclusion completed the previous work in which the role of recombination was largely overlooked. Only having selection for phenotypic stability is not sufficient to help asexual lineages to move towards the optimum. In addition, recombination and selection for phenotypic stability have been observed to help sexual lineages stay close to each other. The reason that sexual lineages can evolve greater benefits than asexual lineages can be explained by differences in the underlying evolutionary dynamics in sexual and asexual lineages. Specifically, recombination can more efficiently help purge sexual lineages with deleterious mutations that are far away from the optimum. However, those sexual lineages that have been able to move close to the optimum, i.e., well-adapted lineages, are evolved to be much more robust to disruption and can be highly maintained. In other words, recombination is surprisingly constructive for close-to-optimum sexual lineages, and only destructive to populations far from the optimum. This indicates that recombination facilitates the evolution of evolutionary robustness in sexual lineages. In contrast, mutations are far more likely to be deleterious, and thus after accumulation in asexual lineages, ultimately tend to move them away from the optimum. These results indicate a fundamental difference between recombination and hypermutation — although they have a superficial similarity in causing increased variations, recombination involves swapping regulatory circuits that work well together, whereas mutated sites are randomly generated in hypermutation. These findings have

important implications for the role of gene regulation in the evolution of sex, and for the use of structured representations in machine learning.

When can sexual lineages resist invasion considering the substantial costs incurred by sex and recombination?

In Chapter 6, I have presented two case studies to show that the selection pressure acting on rewiring gene regulation is critical to increasing the benefits whilst mitigating the costs of sex and recombination. In the first analysis, I have shown that low-fitness sexual lineages benefit most when the population is subjected to strong selection pressure for the target phenotype, especially at the early stage. In contrast, the benefit of evolving under strong selection pressure is largely absent in asexual lineages. These results have important implications for explaining why some species increase their recombination rate or switch from asexual reproduction to sexual reproduction mode when they are subjected to certain extreme environments such as in the face of pathogen infection. In the second analysis, I have shown that selection pressure can increase the benefits of sexual reproduction, which are able to compensate for the recombination cost. In fact, bouts of recombination in sexual lineages are sufficient to drive the population evolving faster than asexual lineages. In a competition analysis, I have shown that recombination is initially costly, but it can rapidly evolve to compensate for the costs of sex and recombination. I have further explored the parameter space to investigate how the recombination cost incurred by selection pressure and the twofold cost incurred by recombination frequency affect competition outcomes. I have shown that generally sexual lineages are more likely to outcompete asexual lineages when selection pressure is higher and recombination is less frequent under a lower mutation rate. These results indicate a key role of selection pressure in reducing mutation load as well as mitigating costs of sex and recombination, and have important implications for explaining the maintenance of sexual reproduction in the context of genetic networks.

Possible future work

Future work has been discussed in each chapter throughout the dissertation. Here, I only provide some general discussion.

The Wagner GRN model has been extensively used to explore many fundamental research questions in evolutionary biology and ecology (Fierst and Phillips, 2015). However, only a few studies have focused on analysing the system per se (Wagner, 1994; Pinho et al., 2012, 2015). In particular, due to the non-linear mapping from the regulatory response to the output phenotype at each time step during the developmental stage, it has been difficult to determine whether the network is able to reach an equilibrium phenotypic state, or, if it could, what its equilibrium state would be. This

is critical to many research questions. For example, the robustness assessed in most studies aims to examine if the network remains stable when it is subjected to certain perturbations (Azevedo et al., 2006; MacCarthy and Bergman, 2007a; Ciliberti et al., 2007b; Espinosa-Soto et al., 2011b; Payne and Wagner, 2015). For research work on evolvability, researchers focus on studying whether the individual network can generate novel and inheritable phenotypes in the face of, for example, fluctuating environments (Draghi and Wagner, 2009; Wilder and Stanley, 2015). In almost all current studies, the equilibrium phenotypic state is examined or calculated through iterating the difference equations within certain time steps. This is, however, an extremely time-consuming solution, especially for evolving a large-size population for a very long time. Since the developmental process can greatly slow down the simulation, it is worth exploring how to efficiently calculate or estimate the equilibrium phenotypic state analytically. In the meantime, high performance computing techniques need to be employed to further speed up the simulation process.

The currently available studies have implemented many different types of mutation or noise. For example, on the one hand, mutations happen in the genotype where they can change existing regulations by altering non-zero entries, or change the network topology by creating new regulations in zero entries or deleting existing regulations from non-zero entries (Siegal and Bergman, 2002). Mutation can also occur in an individual's initial expression state and consequently alter its equilibrium state (Espinosa-Soto et al., 2011a). Noise, on the other hand, is normally modelled at each time step during the developmental stage (Masel, 2004; Ciliberti et al., 2007b; Pinho et al., 2015). However, to my best knowledge, the recombination operator has not yet been thoroughly explored. Almost all current studies follow the 'free recombination' strategy (Wagner, 1996; Siegal and Bergman, 2002; Azevedo et al., 2006). But we know that offspring may not inherit equal information from their parents, and there are many different mating strategies in nature. By implementing different recombination operators, we may be able to gain a better understanding of the origin and maintenance of sex and recombination for different species in nature. For example, by differentiating males and females in sexual lineages, we may be able to rigorously examine the role of sexual selection. We could also implement different features, such as different mutation rates, for males and females together with varying mating strategies, to test whether that would affect the underlying evolutionary dynamics.

Most of the current studies have also strictly required that each individual in the population is capable of achieving developmental equilibrium (Wagner, 1996; Siegal and Bergman, 2002; Azevedo et al., 2006). In other words, networks with oscillating phenotypic states will be wiped out immediately from the population. However, although this requirement is a reasonable biological assumption, it largely impedes alternative pathway evolution through, for example, compensatory mutations. In fact,

many empirical studies have indicated that a fluctuating selection regime (periods of purifying selection) is also biologically realistic (Siepielski et al., 2009; Brachi et al., 2013; Gompert et al., 2014; Seppälä, 2015; Bijleveld et al., 2015). As I have shown in this dissertation, networks that have regained phenotypic stability have very different properties compared with networks that have never been compromised. Those properties are expected to affect the underlying evolutionary dynamics, which have been largely overlooked in the current studies. Therefore, future studies are encouraged to consider including those networks that have been through compensation in the simulation, and examine the different evolutionary consequences, if any.

However, if we allowed compromised networks to stay in the population for a while, then the problem would be, for example, how to calculate their fitness. The fitness evaluation functions used in current studies measure the phenotypic distance between the individual's equilibrium state and a given target state (Wagner, 1996; Siegal and Bergman, 2002; Azevedo et al., 2006). However, if the individual is not able to achieve phenotypic stability, then we cannot calculate individual fitness because there is no equilibrium state. In some studies, instead, the average expression state has been used to calculate fitness for networks with oscillating phenotypic states (Siegal and Bergman, 2002). In this dissertation, I have also shown that if we use such a method to calculate fitness, we will find that most of the evolved networks are stable under a strong (target) selection regime, even if we do not include selection for phenotypic stability, which suggests that networks with oscillating phenotypic states generally have lower fitness. This is, however, a temporary expedient. In fact, fitness evaluation should consider both an individual's ability to reach developmental equilibrium and its distance away from the target. This is also biologically realistic, as in many biological organisations, for example, proteins, there is a balance between stability and function. Therefore, future work should take both network stability and its function into consideration when evaluating an individual's fitness.

The Wagner GRN model also has great potential to be used to solve optimisation problems in the machine learning field, since the model can converge to a target phenotype, as shown in this dissertation (Wang et al., 2014a). Then, the problems are 1) how to encode a solution into the model, and 2) how to evaluate the solution, i.e., designing new fitness functions tailored for particular optimisation problems in the real world. It is natural to consider that an individual's phenotype can be encoded as the candidate solution. By evaluating candidate solutions using a designed fitness function, an optimal solution, in theory, can be found at the end of the evolution process. If we used the discrete Wagner GRN model where the gene expression state is either $-1/+1$ or $0/1$, then the model could be easily modified to solve combinatorial optimisation problems such as the knapsack problem, travelling salesman problem, etc. However, it would be difficult to develop the Wagner GRN model to be used for solving more complicated

continuous optimisation problems. Future work should rigorously test the performance of the Wagner GRN model for solving real optimisation problems in comparison with other evolutionary computation methods such as the ant colony optimisation, artificial bee colony algorithm, artificial immune systems, differential evolution, genetic algorithm, particle swarm optimisation, etc.

As George E. P. Box said, 'Essentially, all models are wrong, but some are useful.' The model developed in this dissertation makes no attempt to fully cover the biochemical processes of the underlying transcriptional regulation in real biological systems. Instead, the abstraction of regulatory systems, as well as the developmental process, are explicitly modelled and emphasised. The conclusions drawn from this dissertation using such an abstract model aim at providing useful high-level explanations and predictions for general patterns or properties that we would observe in natural systems. In particular, for the findings regarding to compensatory mutation presented in this dissertation, it is expected to use appropriate real datasets to test the properties of compensatory mutation such as where and when compensatory mutation occurs in context of genetic networks, and examine the role of compensatory mutation on non-adaptive evolution. For the findings regarding to recombination presented in this dissertation, it is expected to use appropriate real datasets to examine the underlying evolutionary dynamics where sexual lineages compete against asexual lineages under the condition when both cost of recombination and twofold cost of sex are considered.

Bibliography

Agrawal, A. F. (2001). Sexual selection and the maintenance of sexual reproduction. *Nature*, 411(6838):692–695.

Agrawal, A. F. (2002). Genetic loads under fitness-dependent mutation rates. *Journal of Evolutionary Biology*, 15(6):1004–1010.

Agrawal, A. F. and Wang, A. D. (2008). Increased transmission of mutations by low-condition females: Evidence for condition-dependent DNA repair. *PLoS Biology*, 6(2):e30.

Aldana, M., Balleza, E., Kauffman, S., and Resendiz, O. (2007). Robustness and evolvability in genetic regulatory networks. *Journal of Theoretical Biology*, 245(3):433–448.

Aldana, M., Coppersmith, S., and Kadanoff, L. P. (2003). *Perspectives and Problems in Nolinear Science: A Celebratory Volume in Honor of Lawrence Sirovich*, chapter Boolean Dynamics with Random Couplings, pages 23–89. Springer New York, New York, NY.

Alizon, S. and Taylor, P. (2008). Empty sites can promote altruistic behaviour. *Evolution*, 62(6):1335–1344.

Archetti, M. (2004). Recombination and loss of complementation: A more than two-fold cost for parthenogenesis. *Journal of Evolutionary Biology*, 17(5):1084–1097.

Azevedo, R. B. R., Lohaus, R., Srinivasan, S., Dang, K. K., and Burch, C. L. (2006). Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature*, 440(7080):87–90.

Bäck, T. (1996). *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press, Oxford, UK.

Banner, L. R. and Mc Lai, M. (1991). Random nature of coronavirus RNA recombination in the absence of selection pressure. *Virology*, 185(1):441–445.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Barabasi, A.-L. and Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113.

Barreto, F. S. and Burton, R. S. (2013). Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. *Molecular Biology and Evolution*, 30(2):310–314.

Barton, N. H. (2009). Why sex and recombination? *Cold Spring Harbor Symposia on Quantitative Biology*, 74(0):187–195.

Becks, L. and Agrawal, A. F. (2012). The evolution of sex is favoured during adaptation to new environments. *PLoS Biology*, 10(5):e1001317.

Beerenwinkel, N., Pachter, L., Sturmfels, B., Elena, S., and Lenski, R. (2007). Analysis of epistatic interactions and fitness landscapes using a new geometric approach. *BMC Evolutionary Biology*, 7(1):60.

Bergman, A. and Siegal, M. L. (2003). Evolutionary capacitance as a general feature of complex gene networks. *Nature*, 424(6948):549–552.

Bhattacharya, S., Zhang, Q., and Andersen, M. (2011). A deterministic map of Waddington's epigenetic landscape for cell fate specification. *BMC Systems Biology*, 5:85.

Bhattacherjee, A., Mallik, S., and Kundu, S. (2015). Compensatory mutations occur within the electrostatic interaction range of deleterious mutations in protein structure. *Journal of Molecular Evolution*, 80(1):10–12.

Bijleveld, A. I., Twietmeyer, S., Piechocki, J., van Gils, J. A., and Piersma, T. (2015). Natural selection by pulsed predation: Survival of the thickest. *Ecology*, 96(7):1943–1956.

Boiani, M. and Scholer, H. R. (2005). Regulatory networks in embryo-derived pluripotent stem cells. *Nature Reviews Molecular Cell Biology*, 6(11):872–881.

Borenstein, E. and Krakauer, D. C. (2008). An end to endless forms: Epistasis, phenotype distribution bias, and nonuniform evolution. *PLoS Computational Biology*, 4(10):e1000202.

Brachi, B., Villoutreix, R., Faure, N., Hautekèete, N., Piquot, Y., Pauwels, M., Roby, D., Cuguen, J., Bergelson, J., and Roux, F. (2013). Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in Arabidopsis thaliana. *Molecular Ecology*, 22(16):4222–4240.

Brandis, G. and Hughes, D. (2013). Genetic characterization of compensatory evolution in strains carrying rpoB Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. *Journal of Antimicrobial Chemotherapy*, 68(11):2493–2497.

Brandis, G., Wrande, M., Liljas, L., and Hughes, D. (2012). Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. *Molecular Microbiology*, 85(1):142–151.

Bryson, J. J., Ando, Y., and Lehmann, H. (2007). Agent-based models as scientific methodology: A case study analysing primate social behaviour. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362(1485):1685–1698.

Burch, C. L. and Chao, L. (1999). Evolution by small steps and rugged landscapes in the RNA virus phi6. *Genetics*, 151(3):921–927.

Butlin, R. (2002). Evolution of sex: The costs and benefits of sex: New insights from old asexual lineages. *Nature Reviews Genetics*, 3(4):311–317.

Čače, I. and Bryson, J. J. (2007). Agent based modelling of communication costs: Why information can be free. In Lyon, C., Nehaniv, C. L., and Cangelosi, A., editors, *Emergence of Communication and Language*, chapter 16, pages 305–321. Springer London.

Callahan, H. S., Pigliucci, M., and Schlichting, C. D. (1997). Developmental phenotypic plasticity: Where ecology and evolution meet molecular biology. *BioEssays*, 19(6):519–525.

Charlesworth, B. (1993). Directional selection and the evolution of sex and recombination. *Genetical Research*, 61:205–224.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3):195–205.

Choi, S. S., Li, W. M., and Lahn, B. T. (2005). Robust signals of coevolution of interacting residues in mammalian proteomes identified by phylogeny-aided structural analysis. *Nature Genetics*, 37(12):1367–1371.

Ciliberti, S., Martin, O. C., and Wagner, A. (2007a). Innovation and robustness in complex regulatory gene networks. *Proceedings of the National Academy of Sciences of the United States of America*, 104(34):13591–13596.

Ciliberti, S., Martin, O. C., and Wagner, A. (2007b). Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS Computational Biology*, 3(2):e15.

Clune, J., Mouret, J.-B., and Lipson, H. (2013). The evolutionary origins of modularity. *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1755):1–9.

Comas, I., Borrell, S., Roetzer, A., Rose, G., Malla, B., Kato-Maeda, M., Galagan, J., Niemann, S., and Gagneux, S. (2012). Whole-genome sequencing of rifampicin-resistant mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. *Nature Genetics*, 44(1):106–110.

Connelly, C. F., Wakefield, J., and Akey, J. M. (2014). Evolution and genetic architecture of chromatin accessibility and function in yeast. *PLoS Genetics*, 10(7):e1004427.

Cotterell, J. and Sharpe, J. (2013). Mechanistic explanations for restricted evolutionary paths that emerge from gene regulatory networks. *PLoS ONE*, 8(4):e61178.

Covert, A., Lenski, R., Wilke, C., and Ofria, C. (2013). Experiments on the role of deleterious mutations as stepping stones in adaptive evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 110(34):E3171–E3178.

Crawford, H., Prado, J. G., Leslie, A., Hue, S., Honeyborne, I., Reddy, S., van der Stok, M., Mncube, Z., Brander, C., Rousseau, C., Mullins, J. I., Kaslow, R., Goepfert, P., Allen, S., Hunter, E., Mulenga, J., Kiepiela, P., Walker, B. D., and Goulder, P. J. (2007). Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. *Journal of Virology*, 81(15):8346–8351.

Crombach, A. and Hogeweg, P. (2008). Evolution of evolvability in gene regulatory networks. *PLoS Computational Biology*, 4(7):e1000112.

Davidson, E. H. (2010). *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. Elsevier Science.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z. j., Schilstra, M. J., Clarke, P. J. C., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L., and Bolouri, H. (2002). A genomic regulatory network for development. *Science*, 295(5560):1669–1678.

Davis, B. H., Poon, A. F. Y., and Whitlock, M. C. (2009). Compensatory mutations are repeatable and clustered within proteins. *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1663):1823–1827.

de Vos, M., Mueller, B., Borrell, S., Black, P. A., van Helden, P. D., Warren, R. M., Gagneux, S., and Victor, T. C. (2013). Putative compensatory mutations in the rpoC gene of rifampin-resistant Mycobacterium tuberculosis are associated with ongoing transmission. *Antimicrobial Agents and Chemotherapy*, 57(2):827–832.

Doncaster, C. P., Pound, G. E., and Cox, S. J. (2000). The ecological cost of sex. *Nature*, 404(6775):281–285.

Draghi, J. and Wagner, G. P. (2009). The evolutionary dynamics of evolvability in a gene network model. *Journal of Evolutionary Biology*, 22(3):599–611.

Ebert, D. (2005). *Ecology, Epidemiology, and Evolution of Parasitism in Daphnia*. National Center for Biotechnology Information (US).

Engelstädter, J. (2008). Constraints on the evolution of asexual reproduction. *BioEssays*, 30(11-12):1138–1150.

Eshel, I. and Feldman, M. W. (1970). On the evolutionary effect of recombination. *Theoretical Population Biology*, 1(1):88–100.

Espinosa-Soto, C., Martin, O. C., and Wagner, A. (2011a). Phenotypic plasticity can facilitate adaptive evolution in gene regulatory circuits. *BMC Evolutionary Biology*, 11(1):5.

Espinosa-Soto, C., Martin, O. C., and Wagner, A. (2011b). Phenotypic robustness can increase phenotypic variability after nongenetic perturbations in gene regulatory circuits. *Journal of Evolutionary Biology*, 24(6):1284–1297.

Espinosa-Soto, C. and Wagner, A. (2010). Specialization can drive the evolution of modularity. *PLoS Computational Biology*, 6(3):e1000719.

Eyre-Walker, A. and Keightley, P. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8):610–618.

Feldman, M. W., Otto, S. P., and Christiansen, F. B. (1996). Population genetic perspectives on the evolution of recombination. *Annual Review of Genetics*, 30:261–295.

Felix, M.-A. and Barkoulas, M. (2015). Pervasive robustness in biological systems. *Nature Reviews Genetics*, 16(8):483–496.

Fernández, A. and Lynch, M. (2011). Non-adaptive origins of interactome complexity. *Nature*, 474(7352):502–505.

Fierst, J. L. (2010). Sexual dimorphism increases evolvability in a genetic regulatory network. *Evolutionary Biology*, 38(1):52–67.

Fierst, J. L. (2011). A history of phenotypic plasticity accelerates adaptation to a new environment. *Journal of Evolutionary Biology*, 24(9):1992–2001.

Fierst, J. L. and Phillips, P. C. (2015). Modeling the evolution of complex genetic systems: The gene network family tree. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*, 324(1):1–12.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* Oxford-University-Press.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y.-l., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151(4):1531–1545.

Fournier, T., Gabriel, J., Mazza, C., Pasquier, J., Galbete, J., and Mermod, N. (2007). Steady-state expression of self-regulated genes. *Bioinformatics*, 23(23):3185–3192.

Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000). Using bayesian networks to analyze expression data. *Journal of Computational Biology*, 7(3-4):601–620.

Fusco, G. and Minelli, A. (2008). *Evolving Pathways: Key Themes in Evolutionary Developmental Biology.* Cambridge University Press.

Gandon, S. and Otto, S. P. (2007). The evolution of sex and recombination in response to abiotic or coevolutionary fluctuations in epistasis. *Genetics*, 175(4):1835–1853.

Garfield, D. A., Runcie, D. E., Babbitt, C. C., Haygood, R., Nielsen, W. J., and Wray, G. A. (2013). The impact of gene expression variation on the robustness and evolvability of a developmental gene regulatory network. *PLoS Biology*, 11(10):e1001696.

Gibson, G. and Wagner, G. (2000). Canalization in evolutionary genetics: A stabilizing theory? *BioEssays*, 22(4):372–380.

Gifford, D. R. and MacLean, R. C. (2013). Evolutionary reversals of antibiotic resistance in experimental populations of pseudomonas aeruginosa. *Evolution*, 67(10):2973–2981.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition.

Gompert, Z., Comeault, A. A., Farkas, T. E., Feder, J. L., Parchman, T. L., Buerkle, C. A., and Nosil, P. (2014). Experimental evidence for ecological selection on genome variation in the wild. *Ecology Letters*, 17(3):369–379.

Greenbury, S. F., Johnston, I. G., Smith, M. A., Doye, J. P. K., and Louis, A. A. (2010). The effect of scale-free topology on the robustness and evolvability of genetic regulatory networks. *Journal of Theoretical Biology*, 267(1):48–61.

Guelzim, N., Bottani, S., Bourgine, P., and Kepes, F. (2002). Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genetics*, 31(1):60–63.

Haldane, J. B. S. (2006). *Malaria: Genetic and Evolutionary Aspects*, chapter Disease and Evolution, pages 175–187. Springer US, Boston, MA.

Hare, E. E., Peterson, B. K., Iyer, V. N., Meier, R., and Eisen, M. B. (2008). Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. *PLoS Genet*, 4(6):e1000106.

Harrison, E., Guymer, D., Spiers, A. J., Paterson, S., and Brockhurst, M. A. (2015). Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. *Current Biology*, 25(15):1–6.

Hasty, J., McMillen, D., Isaacs, F., and Collins, J. J. (2001). Computational studies of gene regulatory networks: In numero molecular biology. *Nature Reviews Genetics*, 2(4):268–279.

Hinton, G. E. and Nowlan, S. J. (1987). How learning can guide evolution. *Complex Systems*, 1(3):495–502.

Höglund, J. and Sheldon, B. C. (1998). The cost of reproduction and sexual selection. *Oikos*, 83(3):478–483.

Hörandl, E. (2009). A combinational theory for maintenance of sex. *Heredity*, 103(6):445–457.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.

Hu, T., Banzhaf, W., and Moore, J. H. (2014). The effects of recombination on phenotypic exploration and robustness in evolution. *Artificial Life*, 20(4):457–470.

Huerta-Sanchez, E. and Durrett, R. (2007). Wagner's canalization model. *Theoretical Population Biology*, 71(2):121–130.

Hurst, L. D. and Peck, J. R. (1996). Recent advances in understanding of the evolution and maintenance of sex. *Trends in Ecology & Evolution*, 11(2):46–52.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258.

Ivankov, D. N., Finkelstein, A. V., and Kondrashov, F. A. (2014). A structural perspective of compensatory evolution. *Current Opinion in Structural Biology*, 26:104–112.

Iwasaki, W. and Takagi, T. (2009). Rapid pathway evolution facilitated by horizontal gene transfers across prokaryotic lineages. *PLoS Genetics*, 5(3):e1000402.

Jackson, S., Nielsen, D. M., and Singh, N. D. (2015). Increased exposure to acute thermal stress is associated with a non-linear increase in recombination frequency and an independent linear decrease in fitness in Drosophila. *BMC Evolutionary Biology*, 15(1):175.

Karlebach, G. and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9(10):770–780.

Kashtan, N., Noor, E., and Alon, U. (2007). Varying environments can speed up evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 104(34):13711–13716.

Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3):437–467.

Kauffman, S. A. (1993). *The Origins of Order: Self Organization and Selection in Evolution*. Oxford University Press.

Keightley, P. D. and Eyre-Walker, A. (2000). Deleterious mutations and the evolution of sex. *Science*, 290(5490):331–333.

Keightley, P. D. and Eyre-Walker, A. (2007). Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics*, 177(4):2251–2261.

Kimbrell, T. and Holt, R. D. (2007). Canalization breakdown and evolution in a source-sink system. *The American Naturalist*, 169(3):370–382.

Kimura, M. (1985). The role of compensatory neutral mutations in molecular evolution. *Journal of Genetics*, 64(1):7–19.

Klipp, E., Herwig, R., Kowald, A., Wierling, C., and Lehrach, H. (2008). *Systems Biology in Practice: Concepts, Implementation and Application*. John Wiley & Sons.

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature*, 336(6198):435–440.

Kondrashov, A. S. (1993). Classification of hypotheses on the advantage of amphimixis. *Journal of Heredity*, 84(5):372–387.

Kouyos, R. D., Silander, O. K., and Bonhoeffer, S. (2007). Epistasis between deleterious mutations and the evolution of recombination. *Trends in Ecology & Evolution*, 22(6):308–315.

Kulathinal, R. J., Bettencourt, B. R., and Hartl, D. L. (2004). Compensated deleterious mutations in insect genomes. *Science*, 306(5701):1553–1554.

Lambin, X., Elston, D. A., Petty, S. J., and MacKinnon, J. L. (1998). Spatial asynchrony and periodic travelling waves in cyclic populations of field voles. *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1405):1491–1496.

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.

Le Cunff, Y. and Pakdaman, K. (2012). Phenotype–genotype relation in Wagner's canalization model. *Journal of Theoretical Biology*, 314:69–83.

Le Cunff, Y. and Pakdaman, K. (2014). Reproduction cost reduces demographic stochasticity and enhances inter-individual compatibility. *Journal of Theoretical Biology*, 360:263–270.

Leclerc, R. D. (2008). Survival of the sparsest: Robust gene networks are parsimonious. *Molecular Systems Biology*, 4:213.

Lefébure, T. and Stanhope, M. J. (2007). Evolution of the core and pan-genome of Streptococcus: Positive selection, recombination, and genome composition. *Genome Biology*, 8(5):R71.

Lehner, B. (2011). Molecular mechanisms of epistasis within and between genes. *Trends in Genetics*, 27(8):323–331.

Lehtonen, J., Jennions, M. D., and Kokko, H. (2012). The many costs of sex. *Trends in Ecology & Evolution*, 27(3):172–178.

Levin, B. R., Perrot, V., and Walker, N. (2000). Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics*, 154(3):985–997.

Levine, M. and Davidson, E. H. (2005). Gene regulatory networks for development. *Proceedings of the National Academy of Sciences of the United States of America*, 102(14):4936–4942.

Li, F., Long, T., Lu, Y., Ouyang, Q., and Tang, C. (2004). The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14):4781–4786.

Liebhold, A., Koenig, W. D., and Bjørnstad, O. N. (2004). Spatial synchrony in population dynamics. *Annual Review of Ecology, Evolution, and Systematics*, 35(1):467–490.

Lively, C. M. (2009). The maintenance of sex: Host–parasite coevolution with density-dependent virulence. *Journal of Evolutionary Biology*, 22(10):2086–2093.

Lohaus, R., Burch, C. L., and Azevedo, R. B. R. (2010). Genetic architecture and the evolution of sex. *The Journal of Heredity*, 101(Supplement 1):S142–S157.

Lumley, A. J., Michalczyk, Ł., Kitson, J. J. N., Spurgin, L. G., Morrison, C. A., Godwin, J. L., Dickinson, M. E., Martin, O. Y., Emerson, B. C., Chapman, T., and Gage, M. J. G. (2015). Sexual selection protects against extinction. *Nature*, 522(7557):470–473.

Luscombe, N. M., Madan Babu, M., Yu, H., Snyder, M., Teichmann, S. A., and Gerstein, M. (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, 431(7006):308–312.

Lusk, R. W. and Eisen, M. B. (2010). Evolutionary mirages: Selection on binding site composition creates the illusion of conserved grammars in Drosophila enhancers. *PLoS Genet*, 6(1):e1000829.

Lynch, M. (2007a). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 104(Supplement 1):8597–8604.

Lynch, M. (2007b). The evolution of genetic networks by non-adaptive processes. *Nature Reviews Genetics*, 8(10):803–813.

Lynch, M. and Abegg, A. (2010). The rate of establishment of complex adaptations. *Molecular Biology and Evolution*, 27(6):1404–1414.

MacCarthy, T. and Bergman, A. (2007a). Coevolution of robustness, epistasis, and recombination favors asexual reproduction. *Proceedings of the National Academy of Sciences of the United States of America*, 104(31):12801–12806.

MacCarthy, T. and Bergman, A. (2007b). The limits of subfunctionalization. *BMC Evolutionary Biology*, 7(1):213.

Madan Babu, M., Teichmann, S. A., and Aravind, L. (2006). Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *Journal of Molecular Biology*, 358(2):614–633.

Maisnier-Patin, S., Berg, O. G., Liljas, L., and Andersson, D. I. (2002). Compensatory adaptation to the deleterious effect of antibiotic resistance in salmonella typhimurium. *Molecular Microbiology*, 46(2):355–366.

Martin, O. C. and Wagner, A. (2008). Multifunctionality and robustness trade-offs in model genetic circuits. *Biophysical Journal*, 94(8):2927–2937.

Martin, O. C. and Wagner, A. (2009). Effects of recombination on complex regulatory circuits. *Genetics*, 183(2):673–684.

Martinez, C., Rest, J. S., Kim, A.-R., Ludwig, M., Kreitman, M., White, K., and Reinitz, J. (2014). Ancestral resurrection of the Drosophila S2E enhancer reveals accessible evolutionary paths through compensatory change. *Molecular Biology and Evolution*, 31(4):903–916.

Masel, J. (2004). Genetic assimilation can occur in the absence of selection for the assimilating phenotype, suggesting a role for the canalization heuristic. *Journal of Evolutionary Biology*, 17(5):1106–1110.

Masel, J. and Trotter, M. V. (2010). Robustness and evolvability. *Trends in Genetics*, 26(9):406–414.

Meer, M. V., Kondrashov, A. S., Artzy-Randrup, Y., and Kondrashov, F. A. (2010). Compensatory evolution in mitochondrial trnas navigates valleys of low fitness. *Nature*, 464(7286):279–282.

Meirmans, S. and Strand, R. (2010). Why are there so many theories for sex, and what do we do with them? *Journal of Heredity*, 101(Supplement 1):S3–S12.

Mezmouk, S. and Ross-Ibarra, J. (2014). The pattern and distribution of deleterious mutations in maize. *G3 (Bethesda, Md.)*, 4(1):163–171.

Mintseris, J. and Weng, Z. P. (2005). Structure, function, and evolution of transient and obligate protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 102(31):10930–10935.

Misevic, D., Ofria, C., and Lenski, R. E. (2010). Experiments with digital organisms on the origin and maintenance of sex in changing environments. *The Journal of Heredity*, 101(Supplement 1):S46–54.

Moore, F. B., Rozen, D. E., and Lenski, R. E. (2000). Pervasive compensatory adaptation in Escherichia coli. *Proceedings of the Royal Society of London B: Biological Sciences*, 267(1442):515–522.

Moutouh, L., Corbeil, J., and Richman, D. D. (1996). Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. *Proceedings of the National Academy of Sciences of the United States of America*, 93(12):6106–6111.

Newman, M., Barabasi, A.-L., and Watts, D. J. (2006). *The Structure and Dynamics of Networks*. Princeton University Press, Princeton, NJ, USA.

Olson-Manning, C. F., Wagner, M. R., and Mitchell-Olds, T. (2012). Adaptive evolution: Evaluating empirical support for theoretical predictions. *Nature Reviews Genetics*, 13(12):867–877.

Orr, H. A. (1995). The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics*, 139(4):1805–13.

Orr, H. A. (1996). Dobzhansky, bateson, and the genetics of speciation. *Genetics*, 144(4):1331–1335.

Orr, H. A. (2005). The genetic theory of adaptation: A brief history. *Nature Reviews Genetics*, 6(2):119–127.

Osada, N. and Akashi, H. (2012). Mitochondrial-nuclear interactions and accelerated compensatory evolution: Evidence from the primate cytochrome C oxidase complex. *Molecular Biology and Evolution*, 29(1):337–346.

Otto, S. P. and Feldman, M. W. (1997). Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theoretical Population Biology*, 51(2):134–147.

Otto, S. P. and Gerstein, A. C. (2006). Why have sex? The population genetics of sex and recombination. *Biochemical Society Transactions*, 34(4):519–522.

Otto, S. P. and Lenormand, T. (2002). Resolving the paradox of sex and recombination. *Nature Reviews Genetics*, 3(4):252–261.

Pakula, A. A. and Sauer, R. T. (1989). Genetic analysis of protein stability and function. *Annual Review of Genetics*, 23:289–310.

Palmer, M. E. and Feldman, M. W. (2009). Dynamics of hybrid incompatibility in gene networks in a constant environment. *Evolution*, 63(2):418–431.

Park, S. and Lehner, B. (2013). Epigenetic epistatic interactions constrain the evolution of gene expression. *Molecular Systems Biology*, 9:645–645.

Parsch, J., Tanda, S., and Stephan, W. (1997). Site-directed mutations reveal long-range compensatory interactions in the Adh gene of Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, 94(3):928–933.

Payne, J. L., Moore, J. H., and Wagner, A. (2014). Robustness, evolvability, and the logic of genetic regulation. *Artificial Life*, 20(1):111–126.

Payne, J. L. and Wagner, A. (2015). Mechanisms of mutational robustness in transcriptional regulation. *Frontiers in Genetics*, 6(322).

Petri, C. A. (1962). Kommunikation mit Automaten. Technical report, Universität Hamburg.

Pigliucci, M. (2008). Is evolvability evolvable? *Nature Reviews Genetics*, 9(1):75–82.

Pigliucci, M., Murren, C. J., and Schlichting, C. D. (2006). Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology*, 209(12):2362–2367.

Pinho, R., Borenstein, E., and Feldman, M. W. (2012). Most networks in Wagner's model are cycling. *PLoS ONE*, 7(4):e34285.

Pinho, R., Garcia, V., and Feldman, M. W. (2015). Phenotype accessibility and noise in random threshold gene regulatory networks. *PLoS ONE*, 10(4):e0119972.

Piskol, R. and Stephan, W. (2008). Analyzing the evolution of RNA secondary structures in vertebrate introns using Kimura's model of compensatory fitness interactions. *Molecular Biology and Evolution*, 25(11):2483–2492.

Plunkett, K. and Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition*, 38(1):43–102.

Poon, A. and Chao, L. (2005). The rate of compensatory mutation in the DNA bacteriophage phi X174. *Genetics*, 170(3):989–999.

Poon, A., Davis, B. H., and Chao, L. (2005). The coupon collector and the suppressor mutation: Estimating the number of compensatory mutations by maximum likelihood. *Genetics*, 170(3):1323–1332.

Porter, J. L., Collyer, C. A., and Ollis, D. L. (2015). Compensatory stabilizing role of surface mutations during the directed evolution of dienelactone hydrolase for enhanced activity. *Protein Journal*, 34(1):82–89.

Povolotskaya, I. S. and Kondrashov, F. A. (2010). Sequence space and the ongoing expansion of the protein universe. *Nature*, 465(7300):922–926.

Price, G. R. (1970). Selection and covariance. *Nature*, 227(5257):520–521.

Price, G. R. (1972). Fisher's 'fundamental theorem' made clear. *Annals of Human Genetics*, 36(2):129–140.

Pujadas, E. and Feinberg, A. P. (2012). Regulated noise in the epigenetic landscape of development and disease. *Cell*, 148(6):1123–1131.

Ramos, A. F., Innocentini, G. C. P., and Hornos, J. E. M. (2011). Exact time-dependent solutions for a self-regulating gene. *Physical Review E*, 83(6):062902.

Reeves, C. R. and Rowe, J. E. (2002). *Genetic Algorithms: Principles and Perspectives: A Guide to GA Theory*. Kluwer Academic Publishers, Norwell, MA, USA.

Rhoné, B., Brandenburg, J. T., and Austerlitz, F. (2011). Impact of selection on genes involved in regulatory network: A modelling study. *Journal of Evolutionary Biology*, 24(10):2087–2098.

Rison, S. C. G. and Thornton, J. M. (2002). Pathway evolution, structurally speaking. *Current Opinion in Structural Biology*, 12(3):374–382.

Rokyta, D. R., Joyce, P., Caudle, S. B., Miller, C., Beisel, C. J., and Wichman, H. A. (2011). Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. *PLoS Genetics*, 7(6):e1002075.

Romero, P. A. and Arnold, F. H. (2009). Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology*, 10(12):866–876.

Rudolph, G. (1997). *Convergence Properties of Evolutionary Algorithms*. Kovac.

Rueppell, O., Meier, S., and Deutsch, R. (2012). Multiple mating but not recombination causes quantitative increase in offspring genetic diversity for varying genetic architectures. *PLoS ONE*, 7(10):e47220.

Ruths, T. and Nakhleh, L. (2013). Neutral forces acting on intragenomic variability shape the Escherichia coli regulatory network topology. *Proceedings of the National Academy of Sciences of the United States of America*, 110(19):7754–7759.

San Millan, A., Peña-Miller, R., Toll-Riera, M., Halbert, Z. V., McLean, A. R., Cooper, B. S., and MacLean, R. C. (2014). Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids. *Nature Communications*, 5:5208.

Sanjuán, R., Moya, A., and Elena, S. F. (2004). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proceedings of the National Academy of Sciences of the United States of America*, 101(22):8396–8401.

Seppälä, O. (2015). Natural selection on quantitative immune defence traits: A comparison between theory and data. *Journal of Evolutionary Biology*, 28(1):1–9.

Sevim, V. and Rikvold, P. A. (2008). Chaotic gene regulatory networks can be robust against mutations and noise. *Journal of Theoretical Biology*, 253(2):323–332.

Sherratt, J. A. and Smith, M. J. (2008). Periodic travelling waves in cyclic populations: Field studies and reaction–diffusion models. *Journal of The Royal Society Interface*, 5(22):483–505.

Shin, J. and MacCarthy, T. (2015). Antagonistic coevolution drives Whack-A-Mole sensitivity in gene regulatory networks. *PLoS Computational Biology*, 11(10):e1004432.

Shmulevich, I., Dougherty, E. R., Kim, S., and Zhang, W. (2002). Probabilistic boolean networks: A rule-based uncertainty model for gene regulatory networks. *Bioinformatics*, 18(2):261–274.

Shoichet, B. K., Baase, W. A., Kuroki, R., and Matthews, B. W. (1995). A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences of the United States of America*, 92(2):452–456.

Shuster, S. M. (2009). Sexual selection and mating systems. *Proceedings of the National Academy of Sciences of the United States of America*, 106(Supplement 1):10009–10016.

Shuster, S. M. and Wade, M. J. (2003). *Mating Systems and Strategies*. Monographs in Behavior and Ecology. Princeton University Press.

Siegal, M. L. and Bergman, A. (2002). Waddington's canalization revisited: Developmental stability and evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10528–10532.

Siegal, M. L., Promislow, D. E. L., and Bergman, A. (2007). Functional and evolutionary inference in gene networks: Does topology matter? *Genetica*, 129(1):83–103.

Siepielski, A. M., DiBattista, J. D., and Carlson, S. M. (2009). It's about time: The temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, 12(11):1261–1276.

Siller, S. (2001). Sexual selection and the maintenance of sex. *Nature*, 411(6838):689–692.

Sloan, D. B., Triant, D. A., Wu, M., and Taylor, D. R. (2014). Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Molecular Biology and Evolution*, 31(3):673–682.

Smith, J. M. (1978). *The Evolution of Sex*. Cambridge University Press.

Smith, N. G. C. and Eyre-Walker, A. (2001). Synonymous codon bias is not caused by mutation bias in G+C-rich genes in humans. *Molecular Biology and Evolution*, 18(6):982–986.

Song, T., Park, Y., Shamputa, I. C., Seo, S., Lee, S. Y., Jeon, H.-S., Choi, H., Lee, M., Glynne, R. J., Barnes, S. W., Walker, J. R., Batalov, S., Yusim, K., Feng, S., Tung, C.-S., Theiler, J., Via, L. E., Boshoff, H. I. M., Murakami, K. S., Korber, B., Barry, C. E. I., and Cho, S.-N. (2014). Fitness costs of rifampicin resistance in Mycobacterium tuberculosis are amplified under conditions of nutrient starvation and compensated by mutation in the β' subunit of RNA polymerase. *Molecular Microbiology*, 91(6):1106–1119.

Sorrells, T. R. and Johnson, A. D. (2015). Making sense of transcription networks. *Cell*, 161(4):714–723.

Spirov, A. and Holloway, D. (2013). Using evolutionary computations to understand the design and evolution of gene and cell regulatory networks. *Methods*, 62(1):39–55.

Stearns, S. C. (1987). *The Evolution of Sex and Its Consequences*, volume 55. Birkhäuser.

Stephan, W. (1996). The rate of compensatory evolution. *Genetics*, 144(1):419–426.

Sugár, I. P. and Simon, I. (2014). Self-regulating genes. exact steady state solution by using poisson representation. *Central European Journal of Physics*, 12(9):615–627.

Szamecz, B., Boross, G., Kalapis, D., Kovacs, K., Fekete, G., Farkas, Z., Lazar, V., Hrtyan, M., Kemmeren, P., Groot Koerkamp, M. J. A., Rutkai, E., Holstege, F. C. P., Papp, B., and Pal, C. (2014). The genomic landscape of compensatory evolution. *PLoS Biology*, 12(8):e1001935.

Tedbury, P. R., Mercredi, P. Y., Gaines, C. R., Summers, M. F., and Freed, E. O. (2015). Elucidating the mechanism by which compensatory mutations rescue an HIV-1 matrix mutant defective for Gag membrane targeting and envelope glycoprotein incorporation. *Journal of Molecular Biology*, 427(6):1413–1427.

Thomas, P., Popović, N., and Grima, R. (2014). Phenotypic switching in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111(19):6994–6999.

Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D. S. (2008). How protein stability and new functions trade off. *PLoS Computational Biology*, 4(2):e1000002.

Torres-Sosa, C., Huang, S., and Aldana, M. (2012). Criticality is an emergent property of genetic networks that exhibit evolvability. *PLoS Computational Biology*, 8(9):e1002669.

Tsuda, M. E. and Kawata, M. (2010). Evolution of gene regulatory networks by fluctuating selection and intrinsic constraints. *PLoS Computational Biology*, 6(8):e1000873.

Tuch, B. B., Li, H., and Johnson, A. D. (2008). Evolution of eukaryotic transcription circuits. *Science*, 319(5871):1797–1799.

van Dijk, A. D. J., van Mourik, S., and van Ham, R. C. H. J. (2012). Mutational robustness of gene regulatory networks. *PLoS ONE*, 7(1):e30591.

von Dassow, G., Meir, E., Munro, E. M., and Odell, G. M. (2000). The segment polarity network is a robust developmental module. *Nature*, 406(6792):188–192.

Waddington, C. H. (1953). Genetic assimilation of an acquired character. *Evolution*, 7(2):118–126.

Waddington, C. H. (1959). Canalization of development and genetic assimilation of acquired characters. *Nature*, 183(4676):1654–1655.

Wagner, A. (1994). Evolution of gene networks by gene duplications: A mathematical model and its implications on genome organization. *Proceedings of the National Academy of Sciences of the United States of America*, 91(10):4387–4391.

Wagner, A. (1996). Does evolutionary plasticity evolve? *Evolution*, 50(3):1008–1023.

Wagner, A. (2000). Robustness against mutations in genetic networks of yeast. *Nature Genetics*, 24(4):355–361.

Wagner, A. (2007). *Robustness and Evolvability in Living Systems*. Princeton studies in complexity. Princeton University Press.

Wagner, A. (2008). Robustness and evolvability: A paradox resolved. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1630):91–100.

Wagner, A. (2011a). *The Origins of Evolutionary Innovations: A Theory of Transformative Change in Living Systems*. Oxford University Press.

Wagner, A. (2011b). The low cost of recombination in creating novel phenotypes. *BioEssays*, 33(8):636–646.

Wagner, A. (2014). *Arrival of the Fittest: Solving Evolution's Greatest Puzzle*. Penguin.

Wagner, A. and Wright, J. (2007). Alternative routes and mutational robustness in complex regulatory networks. *Biosystems*, 88(1-2):163–172.

Wagner, G. P. and Altenberg, L. (1996). Perspective: Complex adaptations and the evolution of evolvability. *Evolution*, 50(3):967–976.

Wang, Y. (2015). Contourlet-based multispectral image fusion using free search differential evolution. In *Proceedings of the 12th Biennial International Conference on Artificial Evolution*, pages 318–325, France. Artificial Evolution Association.

Wang, Y., Lan, Y., Weinreich, D. M., Priest, N. K., and Bryson, J. J. (2015). Recombination is surprisingly constructive for artificial gene regulatory networks in the context of selection for developmental stability. In *Proceedings of the European Conference on Artificial Life 2015*, pages 530–537. The MIT Press.

Wang, Y., Matthews, S. G., and Bryson, J. J. (2014a). Evolving evolvability in the context of environmental change: A gene regulatory network (GRN) approach. In *Artificial Life 14: International Conference on the Synthesis and Simulation of Living Systems*, pages 47–53. The MIT Press.

Wang, Y., Xing, H., Liao, L., Wang, Z., Su, B., Zhao, Q., Feng, Y., Ma, P., Liu, J., Wu, J., Ruan, Y., and Shao, Y. (2014b). Co-evolution of compensatory mutation K43E with mutation M41L in long-term HIV antiretroviral treatment. *Current HIV Research*, 12(1):22–31.

Wang, Y. and Yin, J. (2014). Intelligent search optimized edge potential function (EPF) approach to synthetic aperture radar (SAR) scene matching. In *Evolutionary Computation (CEC), 2014 IEEE Congress on*, pages 2124–2131.

West, Lively, and Read (1999). A pluralist approach to sex and recombination. *Journal of Evolutionary Biology*, 12(6):1003–1012.

West, S. A., Peters, A. D., and Barton, N. H. (1998). Testing for epistasis between deleterious mutations. *Genetics*, 149(1):435–444.

Whitacre, J. and Bender, A. (2010). Degeneracy: A design principle for achieving robustness and evolvability. *Journal of Theoretical Biology*, 263(1):143–153.

Whitlock, M. C. and Agrawal, A. F. (2009). Purging the genome with sexual selection: Reducing mutation load through selection on males. *Evolution*, 63(3):569–582.

Whitlock, M. C. and Davis, B. (2001). Genetic Load. In *eLS*. John Wiley & Sons, Ltd.

Whitlock, M. C., Griswold, C. K., and Peters, A. D. (2003). Compensating for the meltdown: The critical effective size of a population with deleterious and compensatory mutations. *Annales Zoologici Fennici*, 40(2):169–183.

Whitlock, M. C. and Otto, S. P. (1999). The panda and the phage: Compensatory mutations and the persistence of small populations. *Trends in Ecology & Evolution*, 14(8):295–296.

Wilder, B. and Stanley, K. (2015). Reconciling explanations for the evolution of evolvability. *Adaptive Behavior*, 23(3):171–179.

Wilke, C., Lenski, R., and Adami, C. (2003). Compensatory mutations cause excess of antagonistic epistasis in RNA secondary structure folding. *BMC Evolutionary Biology*, 3(1):3.

Wilke, C. O. and Adami, C. (2001). Interaction between directional epistasis and average mutational effects. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1475):1469–1474.

Willmer, P. (2011). *Pollination and Floral Ecology*. Princeton University Press.

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., and Romano, L. A. (2003). The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution*, 20(9):1377–1419.

Wright, S. (1931a). Evolution in mendelian populations. *Genetics*, 16(2):97–159.

Wright, S. (1931b). The genetical theory of natural selection: A review. *Journal of Heredity*, 21(8):349–356.

Yin, J., Wang, Y., and Hu, J. (2012a). Free search with adaptive differential evolution exploitation and quantum-inspired exploration. *Journal of Network and Computer Applications*, 35(3):1035–1051. Special Issue on Trusted Computing and Communications.

Yin, J., Wang, Y., and Hu, J. (2012b). A new dimensionality reduction algorithm for hyperspectral image using evolutionary strategy. *Industrial Informatics, IEEE Transactions on*, 8(4):935–943.

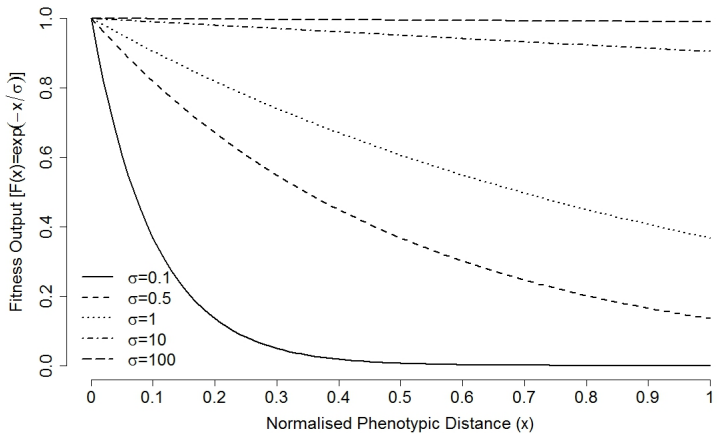
Zhang, L. and Watson, L. T. (2009). Analysis of the fitness effect of compensatory mutations. *HFSP Journal*, 3(1):47–54.

Zhong, W. (2013). *Ecological Immunology of Fungal Infections in Drosophila*. PhD thesis, University of Bath.

Zhong, W. and Priest, N. K. (2011). Stress-induced recombination and the mechanism of evolvability. *Behavioral Ecology and Sociobiology*, 65(3):493–502.

Appendix A

Supporting Information in Chapter 2



F

Figure A-1: Exponential selection curve for target phenotype. *The normalised phenotypic distance x is defined as $D(\mathbf{s}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}})$ (see Equation (2.3)). The fitness output was evaluated under different selection pressures: $\sigma = 0.1$ (strong), $\sigma = 0.5$, $\sigma = 1$, $\sigma = 10$ and $\sigma = 100$ (weak).*

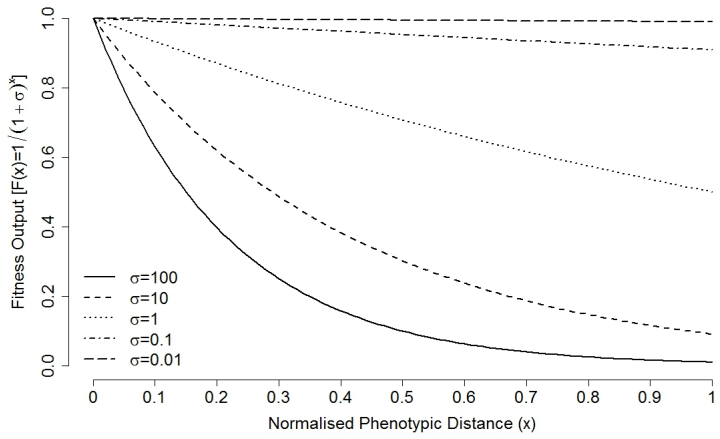


Figure A-2: Multiplicative selection curve for target phenotype. *The normalised phenotypic distance x is defined as $D(\mathbf{s}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}})$ (see Equation (2.4)). The fitness output was evaluated under different selection pressures: $\sigma = 100$ (strong), $\sigma = 10$, $\sigma = 1$, $\sigma = 0.1$ and $\sigma = 0.01$ (weak).*

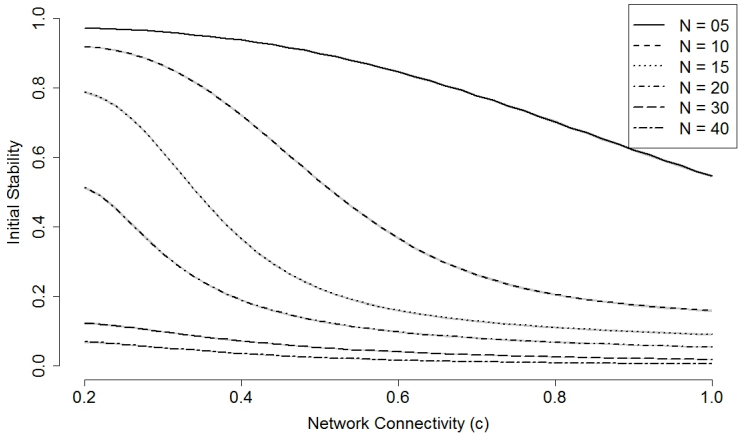


Figure A-3: Stability of randomly generated networks (a=1). *For each network size ($N = 5$, 10, 15, 20, 30 and 40) with each connectivity given from a range of values in continuous intervals ($[0.2, 1]$, step size 0.02), the initial stability (proportion of randomly generated gene networks that were stable) was tested based on an initial $10,000$ randomly generated gene regulatory networks. The system-level parameters were set to be $a = 1$, $devT = 100$ and $\tau = 10$. The shaded areas represent 95% confidence intervals based on 100 independent runs.*

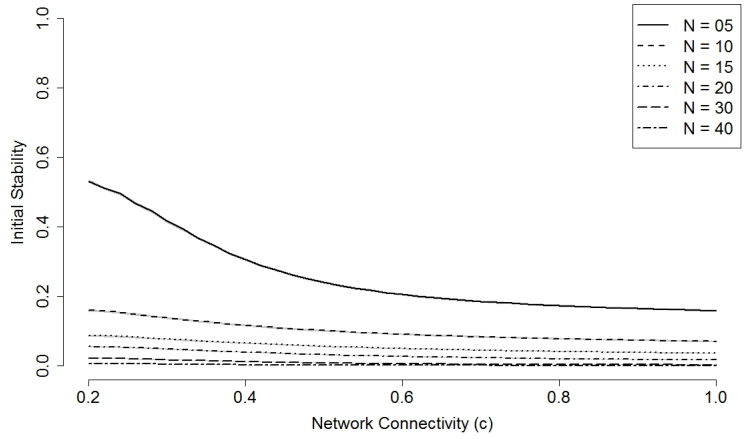


Figure A-4: Stability of randomly generated networks (a=5). *For each network size ($N = 5$, 10, 15, 20, 30 and 40) with each connectivity given from a range of values in continuous intervals ($[0.2, 1]$, step size 0.02), the initial stability (proportion of randomly generated gene networks that were stable) was tested based on an initial $10,000$ randomly generated gene regulatory networks. The system-level parameters were set to be $a = 5$, $devT = 100$ and $\tau = 10$. The shaded areas represent 95% confidence intervals based on 100 independent runs.*

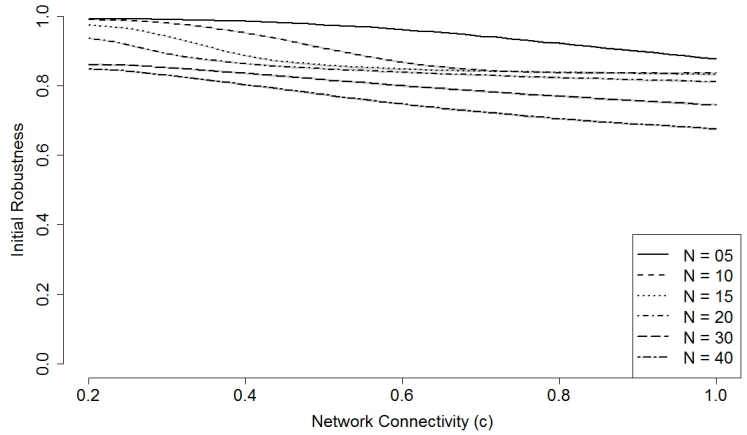


Figure A-5: Robustness of initially stable networks (a=1). *For each network size ($N = 5$, 10, 15, 20, 30 and 40) with each connectivity given from a range of values in continuous intervals ($[0.2, 1]$, step size 0.02), the robustness (proportion of stable networks after exposure to a single round of mutation) was tested based on an initial $10,000$ randomly generated stable gene regulatory networks. The system-level parameters were set to be $a = 1$, $devT = 100$ and $\tau = 10$. The shaded areas represent 95% confidence intervals based on 100 independent runs.*

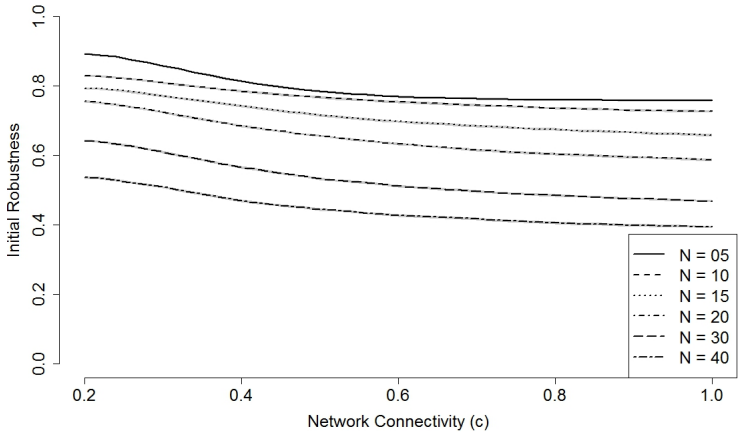


Figure A-6: Robustness of initially stable networks (a=5). *For each network size ($N = 5$, 10, 15, 20, 30 and 40) with each connectivity given from a range of values in continuous intervals ($[0.2, 1]$, step size 0.02), the robustness (proportion of stable networks after exposure to a single round of mutation) was tested based on an initial $10,000$ randomly generated stable gene regulatory networks. The system-level parameters were set to be $a = 5$, $devT = 100$ and $\tau = 10$. The shaded areas represent 95% confidence intervals based on 100 independent runs.*

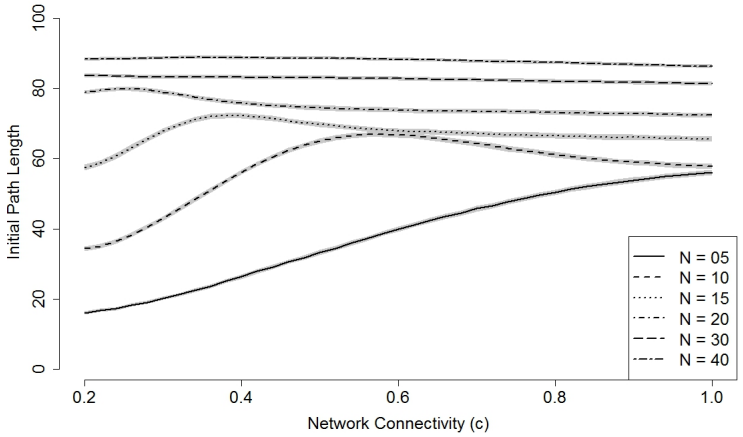


Figure A-7: Path length of initially stable networks (a=1). *For each network size ($N = 5$, 10, 15, 20, 30 and 40) with each connectivity given from a range of values in continuous intervals ($[0.2, 1]$, step size 0.02), the path length (minimum time steps for reaching an equilibrium state) was tested based on an initial $10,000$ randomly generated stable gene regulatory networks. The system-level parameters were set to be $a = 1$, $devT = 100$ and $\tau = 10$. The shaded areas represent 95% confidence intervals based on 100 independent runs.*

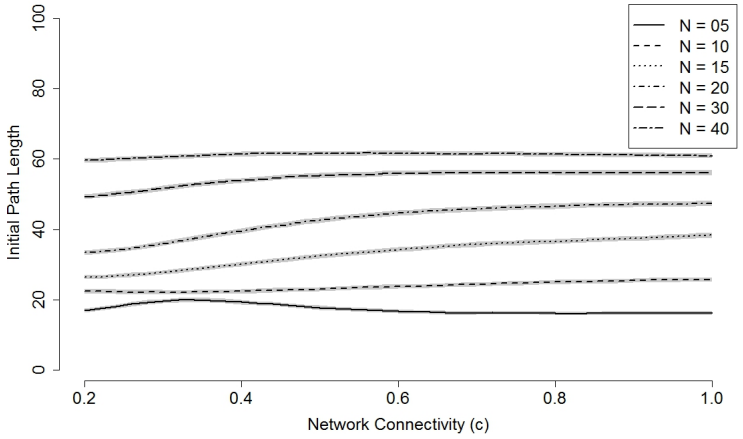


Figure A-8: Path length of initially stable networks (a=5). *For each network size ($N = 5$, 10, 15, 20, 30 and 40) with each connectivity given from a range of values in continuous intervals ($[0.2, 1]$, step size 0.02), the path length (minimum time steps for reaching an equilibrium state) was tested based on an initial $10,000$ randomly generated stable gene regulatory networks. The system-level parameters were set to be $a = 5$, $devT = 100$ and $\tau = 10$. The shaded areas represent 95% confidence intervals based on 100 independent runs.*

Appendix B

Supporting Information in Chapter 3

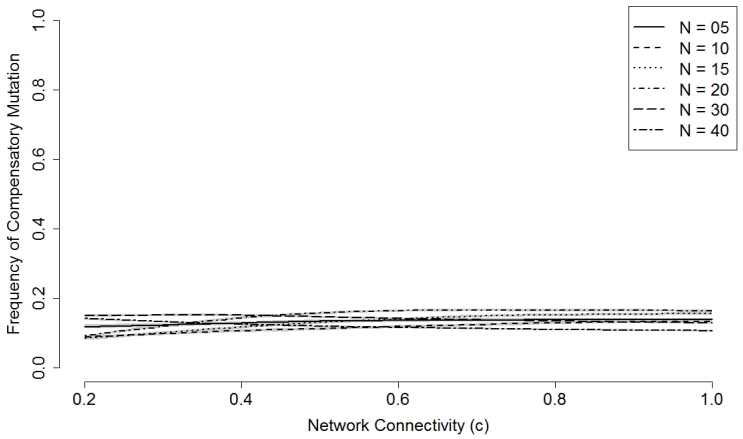


Figure B-1: The influence of the size and connectivity of a gene regulatory network on its frequency of compensatory mutation. (Re-scaled). *For each network size (*$N = 5$*,* 10*,* 15*,* 20*,* 30 *and* 40*) with each connectivity given from a range of values in continuous intervals (*$[0.2, 1]$*, step size* 0.02*), the frequency of compensatory mutation was tested based on an initial* $10,000$ *randomly generated stable gene networks. The shaded areas represent* 95% *confidence intervals based on* 100 *independent runs.*

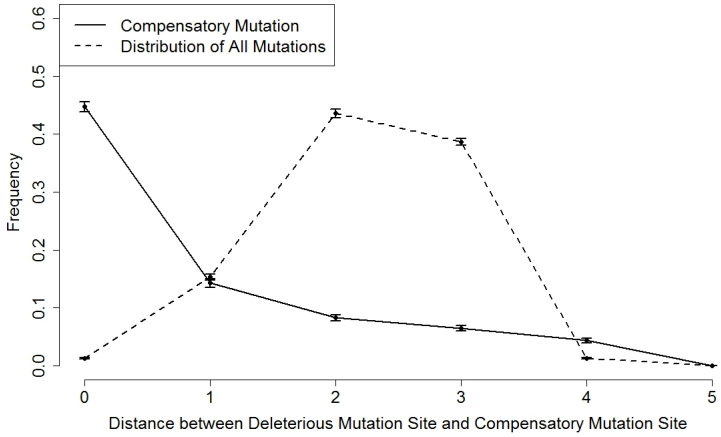


Figure B-2: The compensatory mutation location and distance distribution of all mutations relative to the original deleterious mutation sites (Medium Networks). *For initially stable networks with size* $N = 20$ *and connectivity* $c = 0.2$*, I first collected a pool of compromised networks with deleterious mutations after a single mutation round. I then forced second mutations, classifying these as being* 0 *(on the same site),* 1*,* 2*,* 3*,* 4 *and* 5 *steps away from the original deleterious mutations. For each of these mutation-site-distance categories, I measured the probability that the mutation was compensatory (that it returned the network to stability), based on* $10,000$ *sample networks collected for each distance category as shown in the solid line. I also recorded the spatial distribution of second mutations (*$10,000$ *sample networks) occurring randomly in those compromised networks with respect to their original deleterious mutation sites, shown in the dashed line. The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

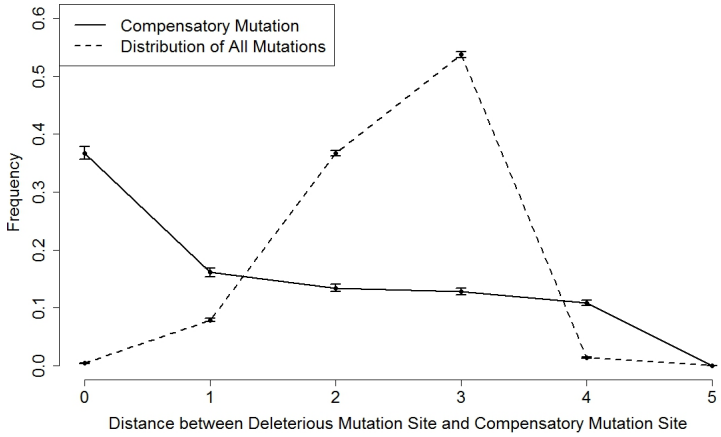


Figure B-3: The compensatory mutation location and distance distribution of all mutations relative to the original deleterious mutation sites (Large Networks). *For initially stable networks with size* $N = 40$ *and connectivity* $c = 0.15$*, I first collected a pool of compromised networks with deleterious mutations after a single mutation round. I then forced second mutations, classifying these as being* 0 *(on the same site),* 1*,* 2*,* 3*,* 4 *and* 5 *steps away from the original deleterious mutations. For each of these mutation-site-distance categories, I measured the probability that the mutation was compensatory (that it returned the network to stability), based on* $10,000$ *sample networks collected for each distance category as shown in the solid line. I also recorded the spatial distribution of second mutations (*$10,000$ *sample networks) occurring randomly in those compromised networks with respect to their original deleterious mutation sites, shown in the dashed line. The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

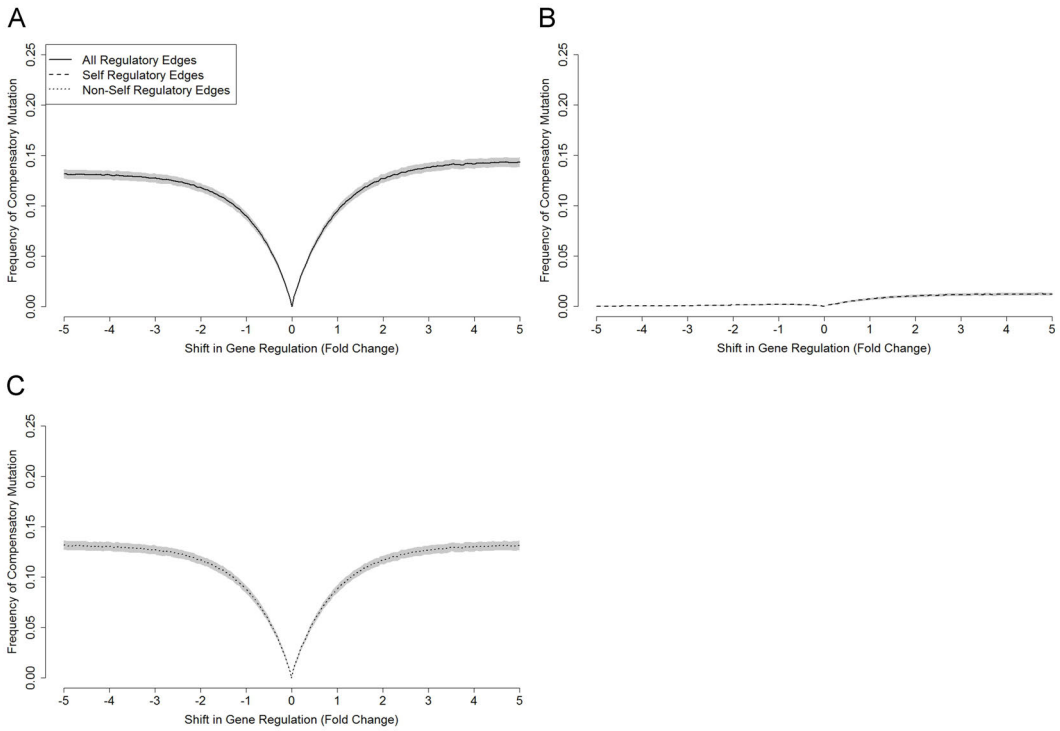


Figure B-4: The influence of different intensities of gene regulation on frequency of compensatory mutation (Medium Networks). *I first collected* $10,000$ *sample networks that had been made unstable by a single mutation from a pool of initially stable networks with* $N = 20$ *and* $c = 0.2$*. Then, I experimented with how a new mutation of varying intensities of gene regulation altered the chances of restoring gene stability. Specifically, I performed new mutations to those compromised networks with deleterious mutations by adding a weight from* $[-5, +5]$ *(step size* 0.5*) to the original regulatory impact, then assessed the resulting patterns in all regulatory edges (**A**), in self-regulatory edges (**B**) and ignoring self-regulatory edges (**C**). The shaded areas represent* 95% *confidence intervals based on* 100 *independent runs.*

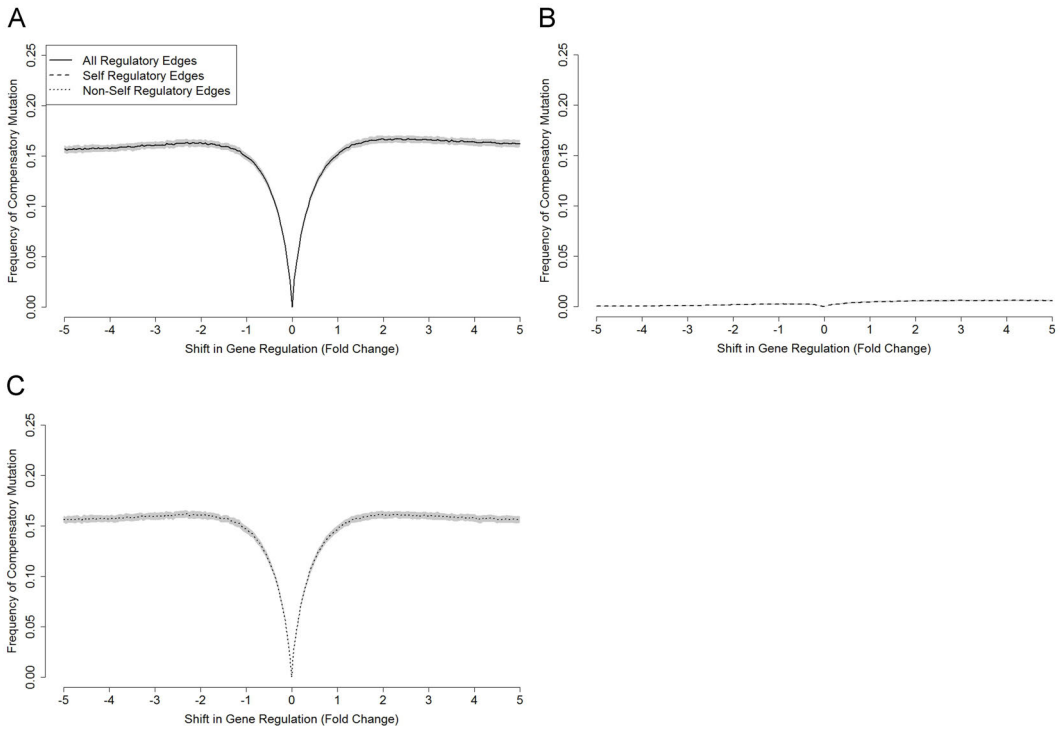


Figure B-5: The influence of different intensities of gene regulation on frequency of compensatory mutation (Large Networks). *I first collected* $10,000$ *sample networks that had been made unstable by a single mutation from a pool of initially stable networks with* $N = 40$ *and* $c = 0.15$*. Then, I experimented with how a new mutation of varying intensities of gene regulation altered the chances of restoring gene stability. Specifically, I performed new mutations to those compromised networks with deleterious mutations by adding a weight from* $[-5, +5]$ *(step size* 0.5*) to the original regulatory impact, then assessed the resulting patterns in all regulatory edges (**A**), in self-regulatory edges (**B**) and ignoring self-regulatory edges (**C**). The shaded areas represent* 95% *confidence intervals based on* 100 *independent runs.*

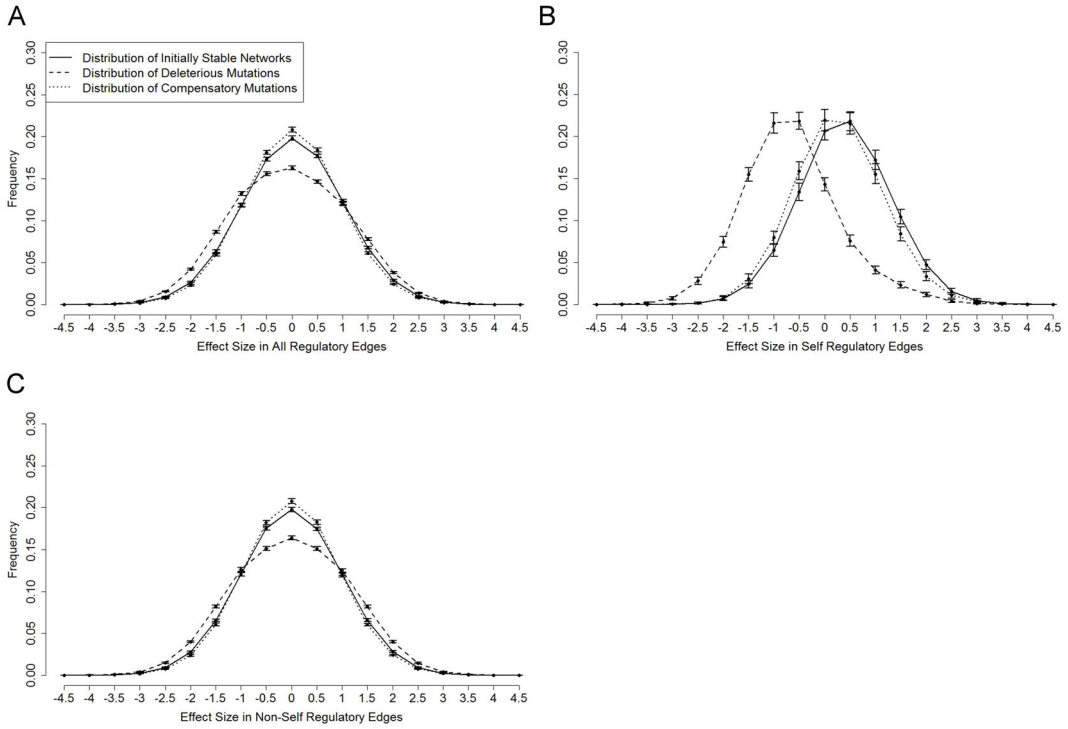


Figure B-6: The distribution of regulation in initially stable, compromised and restored networks (Medium Networks). *For randomly generated stable networks with* $N = 20$ *and* $c = 0.2$*, I collected* $10,000$ *sample regulations. I also collected* $10,000$ *sample regulation weights from deleterious mutations that compromised initially stable networks as well as from compensatory mutations that restored the stability of previously broken networks. I then measured the distributions in all regulatory edges (**A**), in self-regulatory edges (**B**) and ignoring self-regulatory edges (**C**). Given that the regulations are continuous values, I grouped them into* 19 *bins from* $[-4.5, +4.5]$ *(step size* 0.5*). The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

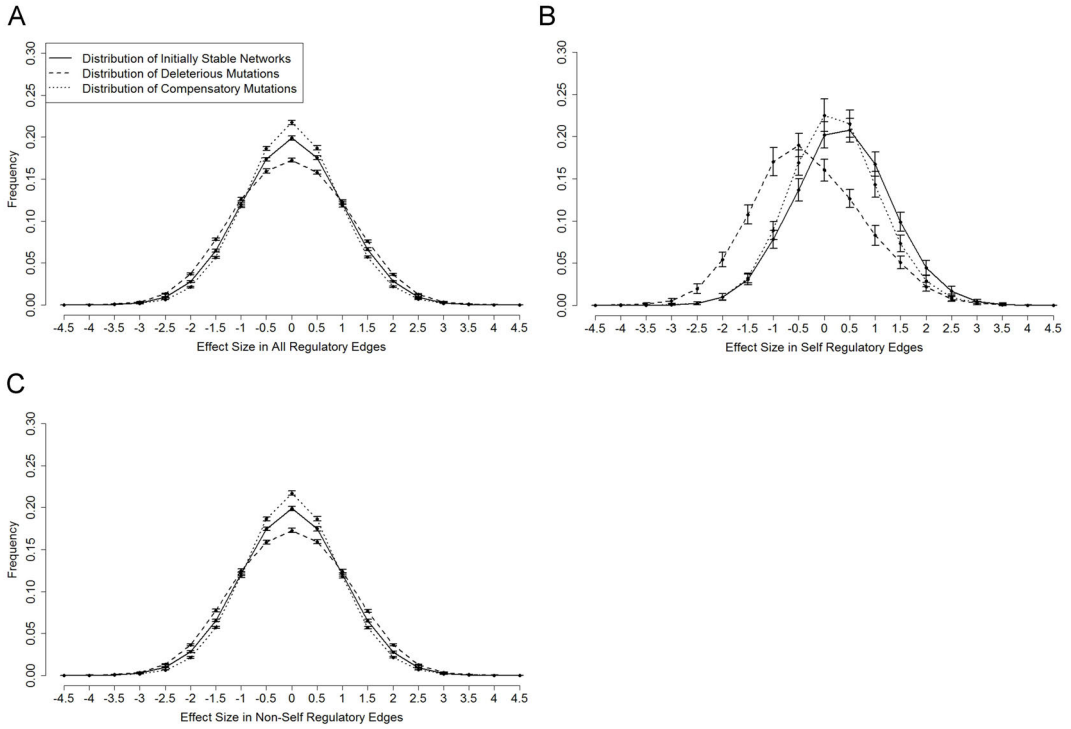


Figure B-7: The distribution of regulation in initially stable, compromised and restored networks (Large Networks). *For randomly generated stable networks with* $N = 40$ *and* $c = 0.15$*, I collected* $10,000$ *sample regulations. I also collected* $10,000$ *sample regulation weights from deleterious mutations that compromised initially stable networks as well as from compensatory mutations that restored the stability of previously broken networks. I then measured the distributions in all regulatory edges (**A**), in self-regulatory edges (**B**) and ignoring self-regulatory edges (**C**). Given that the regulations are continuous values, I grouped them into* 19 *bins from* $[-4.5, +4.5]$ *(step size* 0.5*). The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

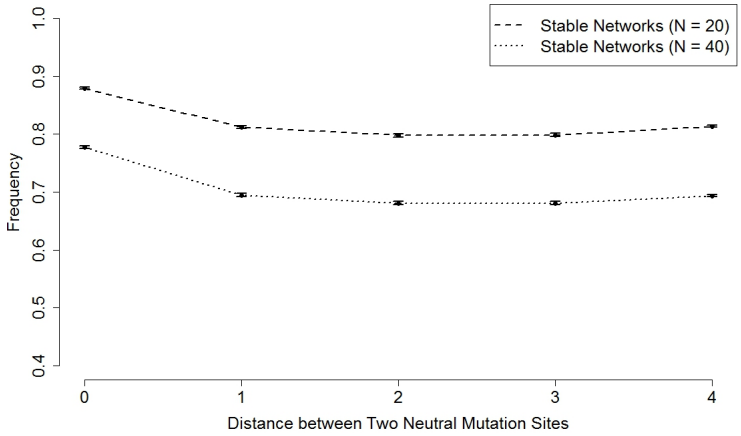


Figure B-8: Location effect in networks with neutral mutations (Medium and Large Networks). *For medium networks ($N = 20, c = 0.2$) and large networks ($N = 40, c = 0.15$), I first collected a pool of stable networks with neutral mutations after a single mutation round. I then forced second mutations, classifying these as being* 0 *(on the same site),* 1, 2, 3 *and* 4 *steps away from the previous neutral mutations. For each of these mutation-site-distance categories, I measured the probability that the mutation was neutral (did not impair network stability) based on* $10,000$ *sample networks collected for each distance category. The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

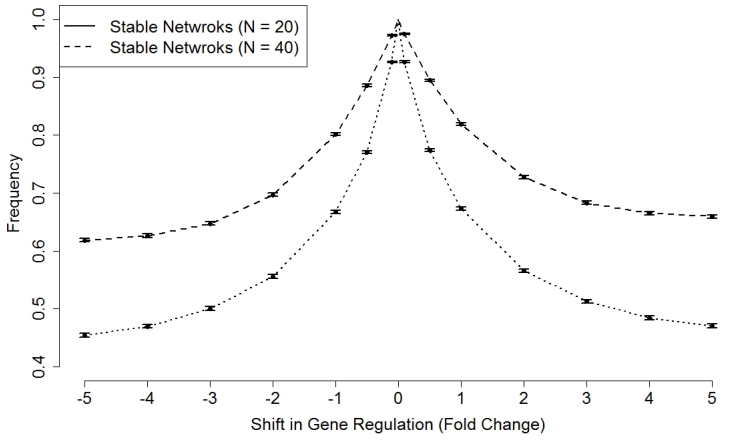


Figure B-9: Mutation size effect in networks with neutral mutations (Medium and Large Networks). *I first collected* $10,000$ *stable networks with neutral mutations after a single mutation round from a pool of initially stable medium networks ($N = 20, c = 0.2$) and large networks ($N = 40, c = 0.15$). Then, I experimented with how new mutations of varying intensities of gene regulation altered the chance of retaining network stability. Specifically, I performed new mutations to those networks with neutral mutations by adding a weight from* $[-5, +5]$ *(step size* 1 *and with four additional regulation shifts:* $-0.5, -0.1, 0.1$ *and* 0.5*) to the original regulatory impact, then assessed the resulting patterns. The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

Appendix C

Supporting Information in Chapter 4

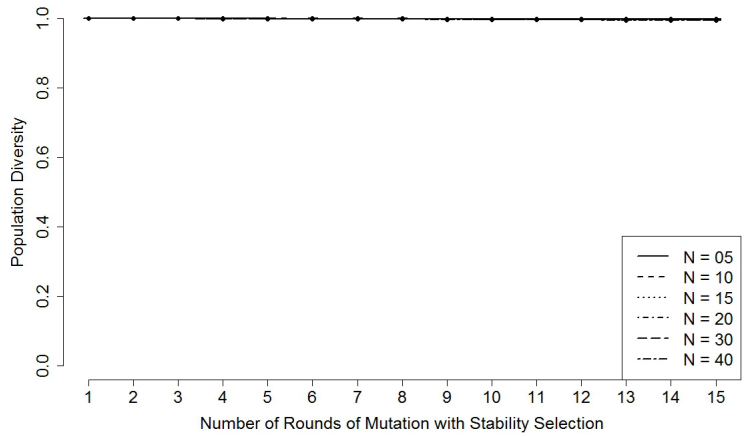


Figure C-1: Population diversity of highly stable networks. *For each network size ($N = 5$, 10, 15, 20, 30 and 40) with network connectivity $c = 0.76$, I tested population diversity for $10,000$ networks that had been exposed to strong selection for phenotypic stability with one up to fifteen rounds of mutation as described in Figure 4-2. The error bars represent 95% confidence intervals based on 100 independent runs.*

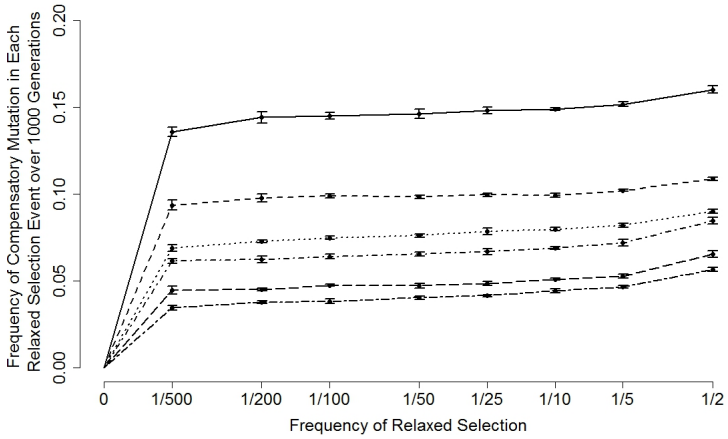


Figure C-2: The frequency of compensatory mutation occurring in each relaxed selection event. *For each network size ($N = 5$, 15, 10, 20, 30 and 40) with connectivity $c = 0.76$ ($W = 10,000$), I measured the number of compensatory mutations occurring after the previous relaxed selection, which happened in every* 2, 5, 10, 25, 50, 100, 200 *and* 500 *generations. The reported results are the mean frequency of compensatory mutations (per relaxed selection cycle) occurring over a total of* $1,000$ *generations for populations with different sizes. Error bars or shaded areas represent* 95% *confidence intervals based on* 10 *independent runs.*

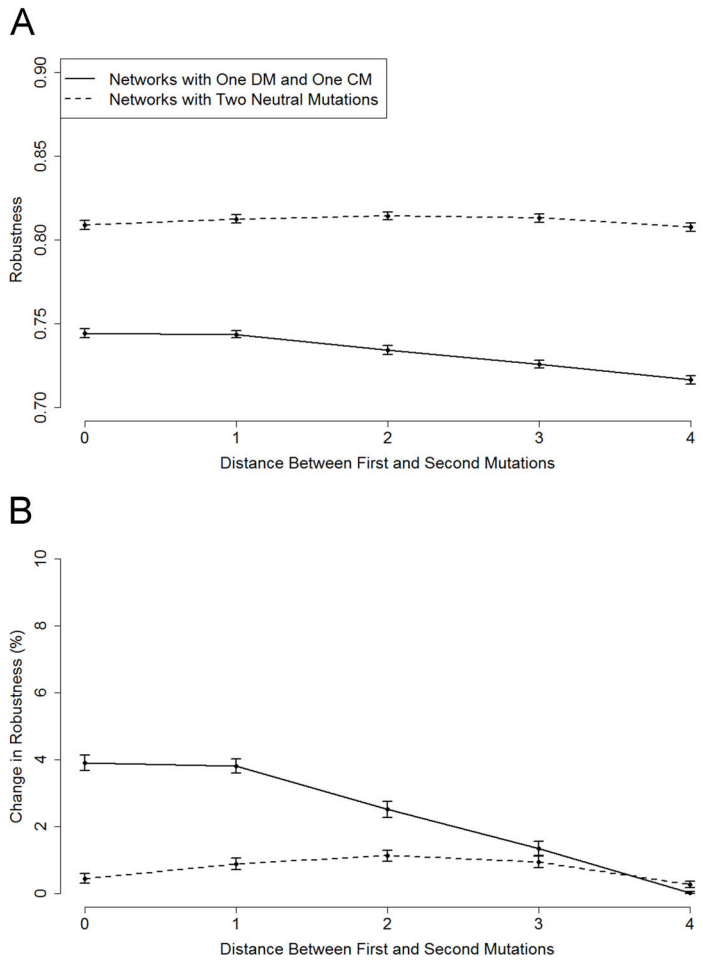


Figure C-3: The impact of distance effect on network robustness (Medium Networks). *For medium networks ($N = 20, c = 0.2$), I collected* $10,000$ *sample stable networks that were subjected one deleterious mutation and then restored by one subsequent compensatory mutation that was* 0, 1, 2, 3 *and* 4 *steps away from the previous deleterious mutation. The sample networks for the control group were collected in a similar way, except that the networks were subjected to two consecutive neutral mutations. Then, I assessed the robustness of the sample networks at each distance step. The reported results are actual robustness (**A**), and change in robustness (**B**) (the actual robustness was normalised by subtracting the minimal value among all categories, and then dividing by the minimal value). The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

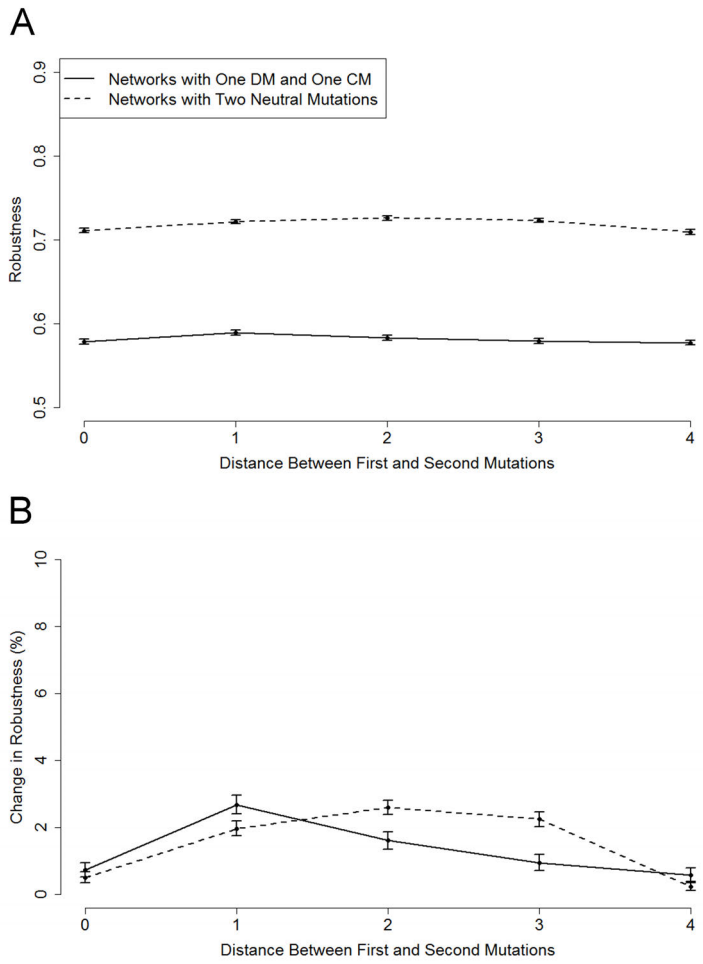


Figure C-4: The impact of distance effect on network robustness (Large Networks). *For large networks ($N = 40, c = 0.15$), I collected* $10,000$ *sample stable networks that were subjected one deleterious mutation and then restored by one subsequent compensatory mutation that was* 0, 1, 2, 3 *and* 4 *steps away from the previous deleterious mutation. The sample networks for the control group were collected in a similar way, except that the networks were subjected to two consecutive neutral mutations. Then, I assessed the robustness of the sample networks at each distance step. The reported results are actual robustness (**A**), and change in robustness (**B**) (the actual robustness was normalised by subtracting the minimal value among all categories, and then dividing by the minimal value). The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

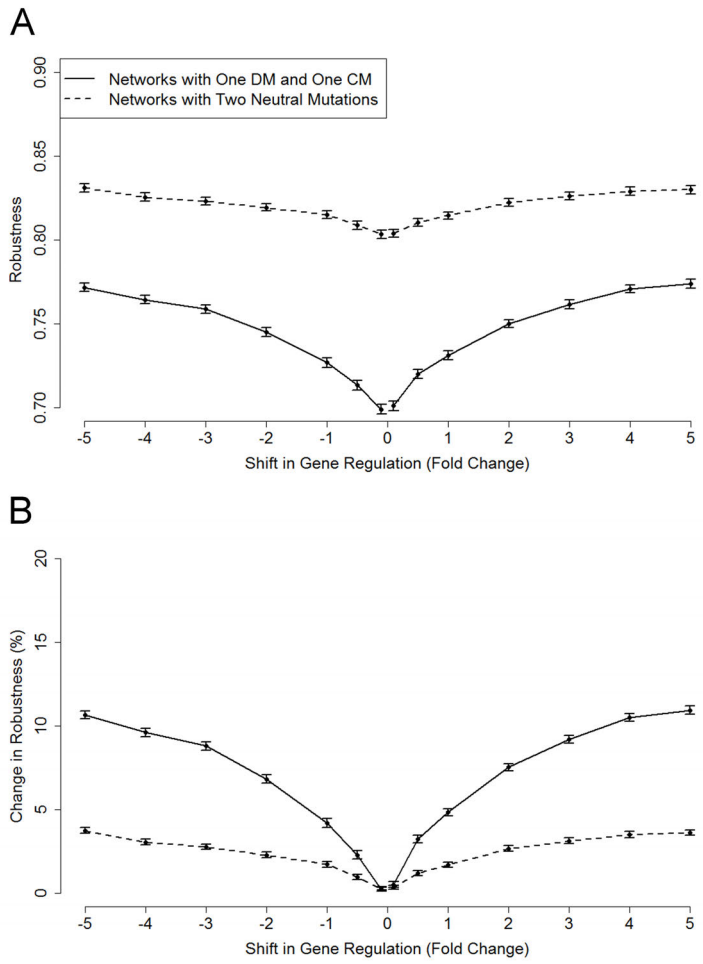


Figure C-5: The impact of mutation size effect on network robustness (Medium Networks). *For medium networks ($N = 20, c = 0.2$), I collected* $10,000$ *sample stable networks that were subjected one deleterious mutation and then restored by one subsequent compensatory mutation with different shifts in gene regulation from* $[-5, +5]$ *(step size* 1 *and with four additional regulation shifts:* -0.5, -0.1, 0.1 *and* 0.5*). The sample networks for the control group were collected in a similar way, except that the networks were subjected to two consecutive neutral mutations. Note that the second neutral mutation has different shifts in gene regulation to the compensatory mutation. Then, I assessed the robustness of the sample networks at each category. The reported results are actual robustness (**A**), and change in robustness (**B**) (the actual robustness was normalised by subtracting the minimal value among all categories, and then dividing by the minimal value). The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

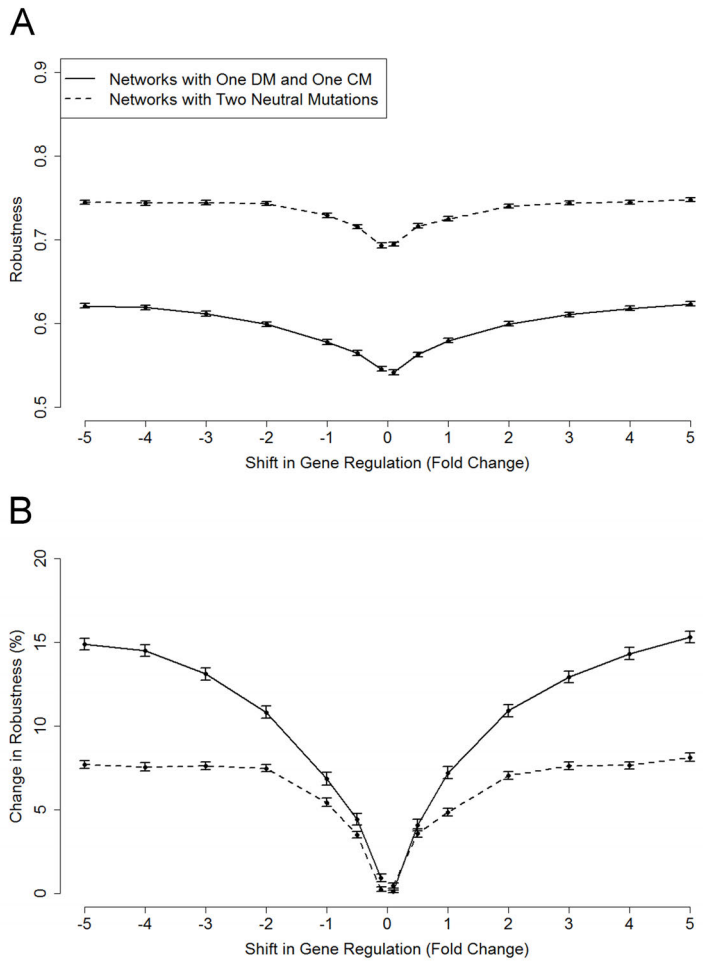


Figure C-6: The impact of mutation size effect on network robustness (Large Networks). *For large networks ($N = 40, c = 0.15$), I collected* $10,000$ *sample stable networks that were subjected one deleterious mutation and then restored by one subsequent compensatory mutation with different shifts in gene regulation from* $[-5, +5]$ *(step size* 1 *and with four additional regulation shifts:* -0.5, -0.1, 0.1 *and* 0.5*). The sample networks for the control group were collected in a similar way, except that the networks were subjected to two consecutive neutral mutations. Note that the second neutral mutation has different shifts in gene regulation to the compensatory mutation. Then, I assessed the robustness of the sample networks at each category. The reported results are actual robustness (**A**), and change in robustness (**B**) (the actual robustness was normalised by subtracting the minimal value among all categories, and then dividing by the minimal value). The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

Table C.1: *Basic statistics of evolved networks with a 'Star' topology*

	Medium	Mean	SD $(E-2)$
Init.	0.17	0.17	2.17
No Mut. & No Rec.	0.17	0.17	3.94
Mut. & No Rec.	0.11	0.11	$4.14E-13$
Rec. & No Mut.	0.11	0.11	2.19
Mut. & Rec.	0.21	0.20	4.42
Mut. & Rec. ($f_{RS}=1/10$)	0.30	0.30	0.43
Mut. & Rec. ($f_{RS}=1/25$)	0.34	0.34	0.47
Mut. & Rec. ($f_{RS}=1/50$)	0.31	0.31	0.51

SD: Standard Deviation

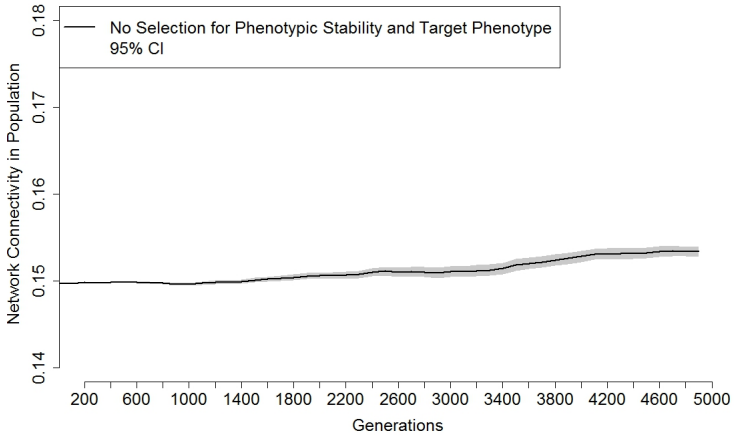


Figure C-7: The evolution of network connectivity in absence of selection. *For network size* $N=40$ *and connectivity* $c=0.15$*, I collected* $10,000$ *stable networks, then evolved them for* $5,000$ *generations, allowing both mutation and recombination at each generation. In every* 200 *generations, I measured the network connectivity of the population in which both selection for phenotypic stability and selection for target phenotype are absent. Note that I only measured the network connectivity for stable networks. The shaded areas represent* 95% *confidence intervals based on* 10 *independent runs.*

Appendix D

Supporting Information in Chapter 5

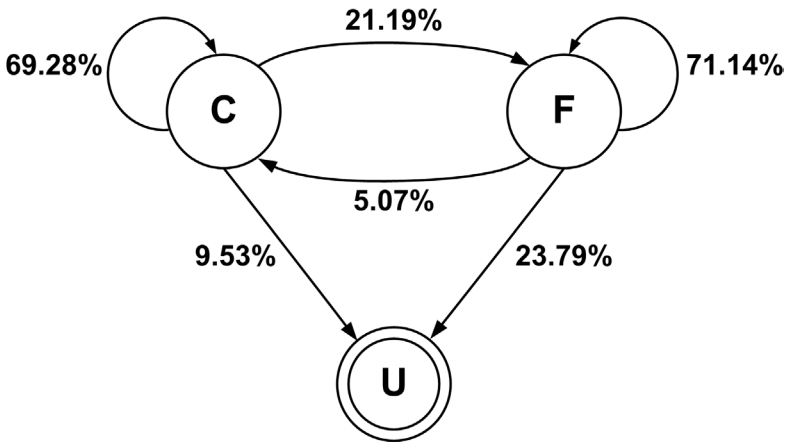


Figure D-1: Estimated state transition probabilities in asexual populations (Two Standard Deviation). *$p_{C,U}$: 9.53% (SD: 0.86%), $p_{C,F}$: 21.19% (SD: 6.13%), $p_{C,C}$: 69.28% (SD: 5.97%), $p_{F,U}$: 23.97% (SD: 3.41%), $p_{F,C}$: 5.07% (SD: 3.72%), $p_{F,F}$: 71.14% (SD: 4.95%). For each population evolved from the founder network, the state transition probabilities were estimated based on* 50 *independent runs. The reported results are the mean probability averaged over* 10 *randomly generated stable founder networks. SD: Standard Deviation.*

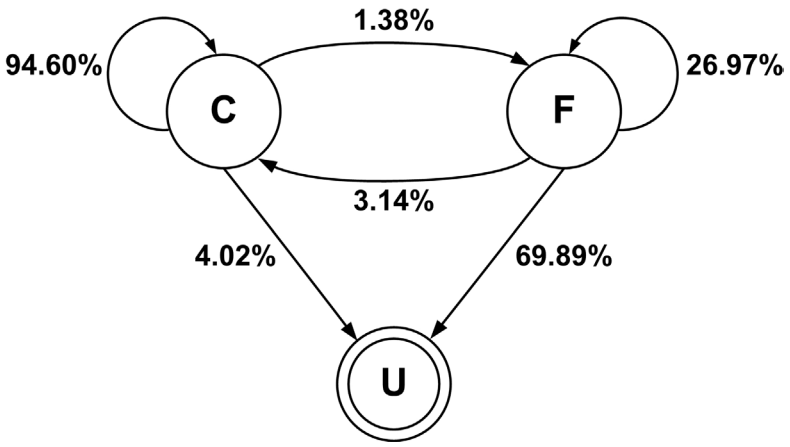


Figure D-2: Estimated state transition probabilities in sexual populations (Two Standard Deviation). *$p_{C,U}$: 4.02% (SD: 0.69%), $p_{C,F}$: 1.38% (SD: 0.76%), $p_{C,C}$: 94.60% (SD: 0.55%), $p_{F,U}$: 69.89% (SD: 3.29%), $p_{F,C}$: 3.14% (SD: 2.34%), $p_{F,F}$: 26.97% (SD: 3.28%). For each population evolved from the founder network, the state transition probabilities were estimated based on* 50 *independent runs. The reported results are the mean probability averaged over* 10 *randomly generated stable founder networks. SD: Standard Deviation.*

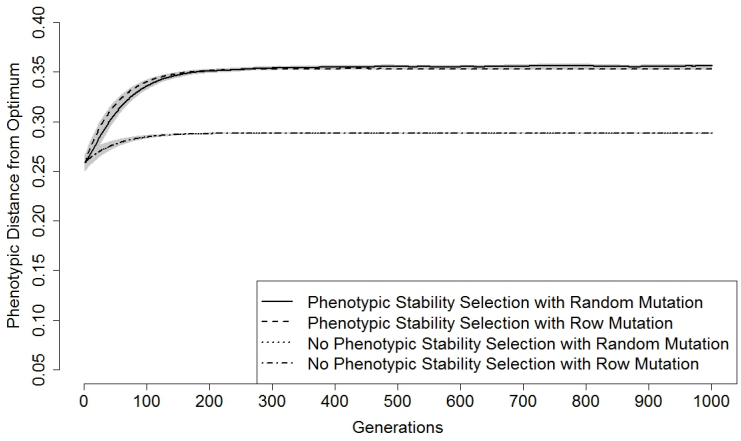


Figure D-3: The phenotypic distance between the evolved populations and the optimum in asexual lineages with hypermutation. *The initial population* $(10,000)$ *was cloned from a randomly generated stable founder network with size* $N = 10$ *and connectivity* $c = 0.75$*. The population was then evolved asexually under phenotypic stability or no stability selection regimes with random mutation or row mutation (see Section 5.2.2). In each generation, each individual in the population was subjected to a perturbation test in order to calculate the phenotypic distance between the evolved populations and the optimum. The shaded areas represent* 95% *confidence intervals based on* 10 *randomly generated stable founder networks.*

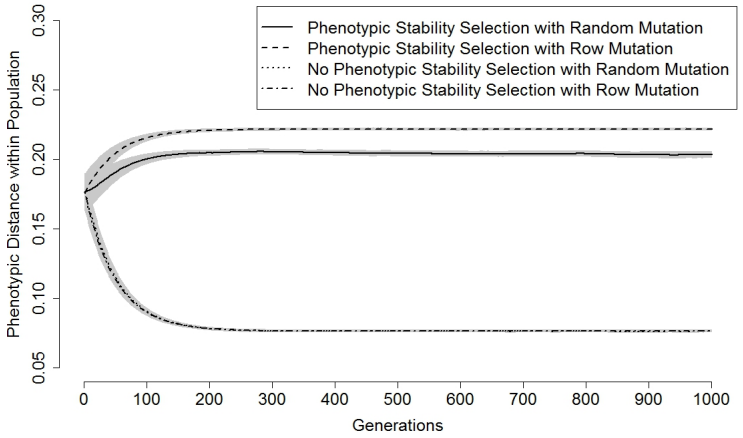


Figure D-4: The phenotypic distance within the evolved populations in asexual lineages with hypermutation. *The initial population (*$10,000$*) was cloned from a randomly generated stable founder network with size* $N = 10$ *and connectivity* $c = 0.75$*. The population was then evolved asexually under phenotypic stability or no stability selection regimes with random mutation or row mutation (see Section 5.2.2). In each generation, each individual in the population was subjected to a perturbation test in order to calculate the phenotypic distance within the evolved populations. The shaded areas represent* 95% *confidence intervals based on* 10 *randomly generated stable founder networks.*

Appendix E

Supporting Information in Chapter 6

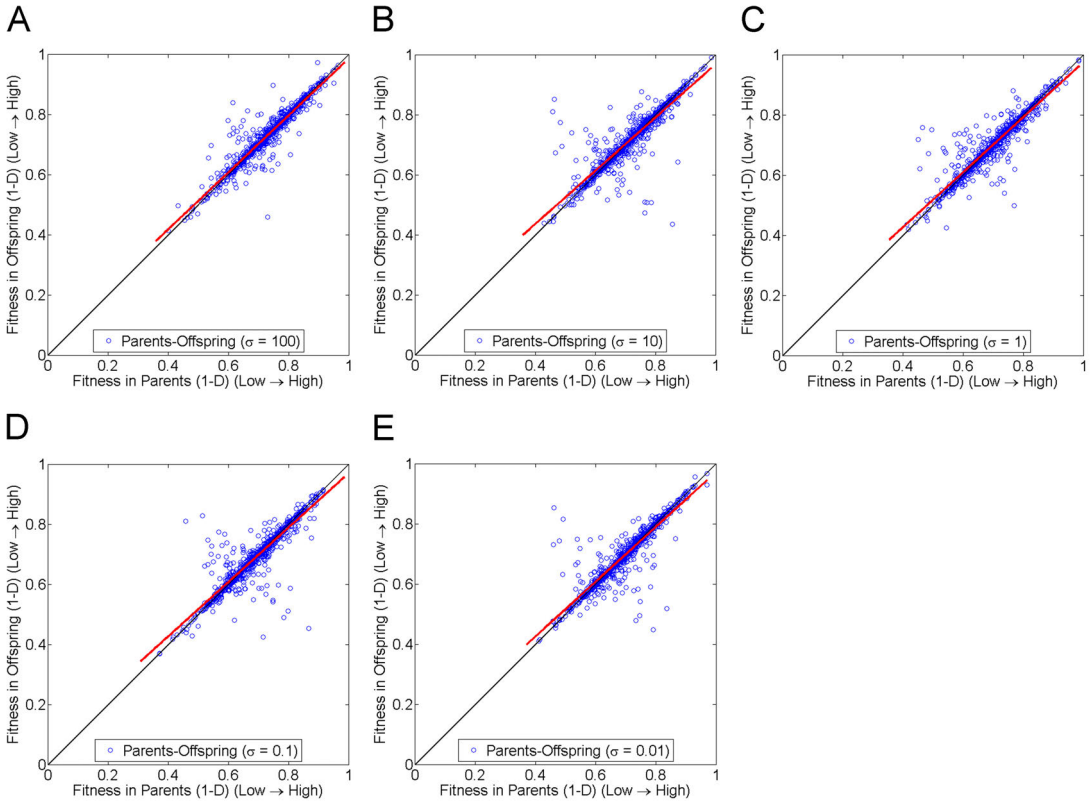


Figure E-1: Scatter plots of parental-offspring fitness in asexual population. *For different selection pressures:* $\sigma = 100$ *(strong)* ***(A)***, $\sigma = 10$ ***(B)***, $\sigma = 1$ ***(C)***, $\sigma = 0.1$ ***(D)*** *and* $\sigma = 0.01$ *(weak)* ***(E)****, I plotted the parental-offspring fitness based on* $1,000$ *randomly selected individuals in one simulation run from results presented in Figure 6-1. Note that the actual phenotypic distance from the optimum was used to calculate fitness* $(1-D(\mathbf{s}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}}))$ *presented in this figure. Each back line is the diagonal of the box. Each red line is linear regression line calculated using a generalised linear model with the normal distribution.*

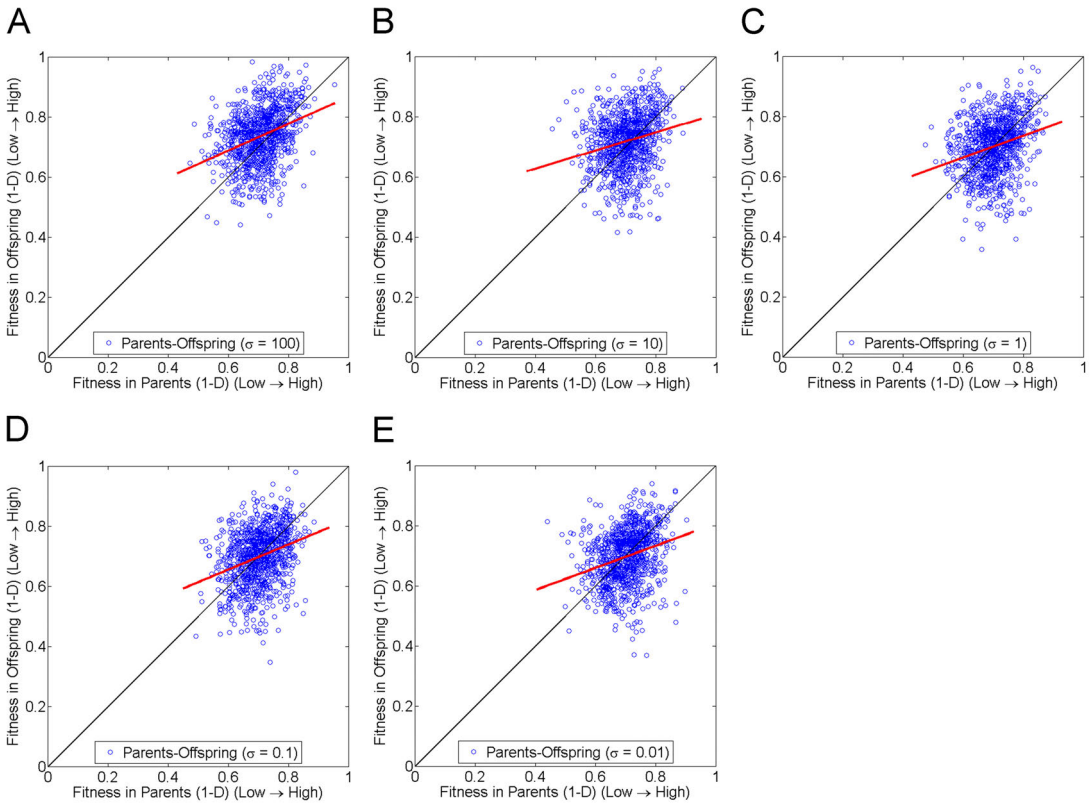


Figure E-2: Scatter plots of parental-offspring fitness in sexual population. *For different selection pressures:* $\sigma = 100$ *(strong)* ***(A)***, $\sigma = 10$ ***(B)***, $\sigma = 1$ ***(C)***, $\sigma = 0.1$ ***(D)*** *and* $\sigma = 0.01$ *(weak)* ***(E)****, I plotted the parental-offspring fitness based on* $1,000$ *randomly selected individuals in one simulation run from results presented in Figure 6-2. Note that the actual phenotypic distance from the optimum was used to calculate fitness* $(1-D(\mathbf{s}_{\mathrm{EQ}}, \mathbf{s}_{\mathrm{OPT}}))$ *presented in this figure. Each back line is the diagonal of the box. Each red line is linear regression line calculated using a generalised linear model with the normal distribution.*

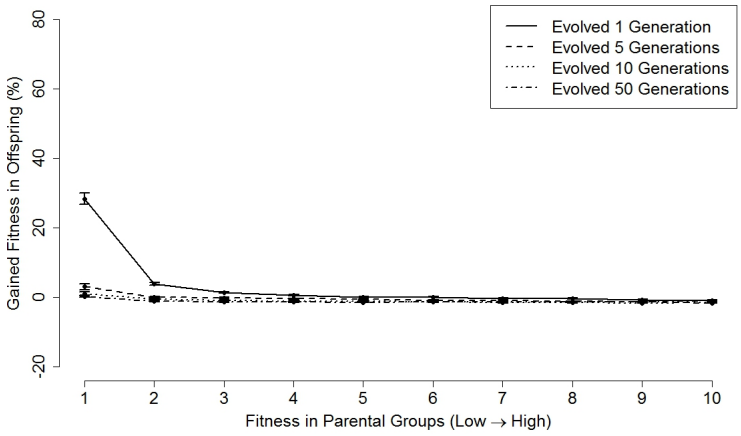


Figure E-3: Comparison of gained fitness in evolved asexual lineages under strong selection pressure. *I used the same population pool of* $10,000$ *randomly generated stable networks with size* $N = 10$ *and connectivity* $c = 0.75$ *as described in Figure 6-1. The population was evolved asexually under selection pressure* $\sigma = 100$*. Then, I recorded each individual's fitness at the initial,* 4^{th}*,* 9^{th} *and* 49^{th} *generations as well as its offspring's fitness in the subsequent generation, i.e., at the* 1^{st}*,* 5^{th}*,* 10^{th} *and* 50^{th} *generations. I then calculated the mean gained fitness of offspring in proportion to their corresponding parental fitness for each of four categories in which all individuals were sorted and grouped similarly as described in Figure 6-1. The error bars represent* 95% *confidence intervals based on* 100 *independent runs.*

Table E.1: *Winning probability of sexual lineages in* 100 *independent competition runs*

Mutation Rate	Rec. Freq.	Selection Pressure (σ)								
(μ)	($f_{Rec.}$)	0.5	1	10	10^2	10^3	10^4	10^5	10^6	
10^{-5}	1/50	0	10	40	84	96	98	100	100	Winning Probability of Sexual Lineages (%)
	1/25	0	0	10	91	99	100	99	100	
	1/10	0	0	1	0	58	100	100	100	
	1/5	0	0	0	0	0	0	5	85	
	1/1	0	0	0	0	0	0	0	0	
10^{-4}	1/50	0	9	41	85	92	81	88	94	
	1/25	0	0	6	90	97	100	98	96	
	1/10	0	0	0	0	39	93	100	100	
	1/5	0	0	0	0	0	0	1	44	
	1/1	0	0	0	0	0	0	0	0	
10^{-3}	1/50	3	8	30	72	66	76	74	92	
	1/25	0	0	6	63	95	93	85	85	
	1/10	0	0	0	0	1	42	93	97	
	1/5	0	0	0	0	0	0	0	0	
	1/1	0	0	0	0	0	0	0	0	
10^{-2}	1/50	0	4	7	23	29	43	48	53	
	1/25	0	0	0	2	23	33	45	49	
	1/10	0	0	0	0	0	0	0	5	
	1/5	0	0	0	0	0	0	0	0	
	1/1	0	0	0	0	0	0	0	0	