Small Change?
Adaptation and its Costs, in Experimental Evolution

Zachary N. Ardern

A thesis submitted in partial fulfilment of the
requirements of the degree of
Doctor of Philosophy in Biological Sciences
The University of Auckland
New Zealand

2016

THE UNIVERSITY OF AUCKLAND
SCHOOL OF GRADUATE STUDIES

Co-Authorship Form

Graduate Centre
The ClockTower – East Wing
22 Princes Street, Auckland
Phone: +64 9 373 7599 ext 81321
Fax: +64 9 373 7610
Email: postgraduate@auckland.ac.nz
www.postgrad.auckland.ac.nz

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work. Chapter 7 "Investigating Trade-offs in Sexual Populations with Gene-flow" - published in Pontarotti P. (ed) (2015). Evolutionary Biology: Biodiversification from Genotype to Phenotype. Springer International Publishing, Cham. pp 245-262	
Nature of contribution by PhD candidate	Lead author
Extent of contribution by PhD candidate (%)	80%

CO-AUTHORS

Name	Nature of Contribution
Matthew R Goddard	Article's main concept and structure, editing

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
Zachary Ardern		24th February 2015
Matthew Goddard		25th Feb 2017

Thesis Consent Form

This thesis may be consulted for the purposes of research or private study provided that due acknowledgement is made where appropriate and that permission is obtained before any material from the thesis is published. Students who do not wish their work to be available for reasons such as pending patents, copyright agreements, or future publication should seek advice from the Graduate Centre as to restricted use or embargo.

Author of thesis	ZACHARY N. ARDERN
Title of thesis	SMALL CHANGE? ADAPTATION AND ITS COSTS IN EXPERIMENTAL EVOLUTION
Name of degree	DOCTOR OF PHILOSOPHY IN BIOLOGICAL SCIENCES
Date Submitted	07/03/2017

Print Format (Tick the boxes that apply)	
☑	I agree that the University of Auckland Library may make a copy of this thesis available for the collection of another library on request from that library.
☑	I agree to this thesis being copied for supply to any person in accordance with the provisions of Section 56 of the Copyright Act 1994.
Digital Format - PhD theses	
I certify that a digital copy of my thesis deposited with the University is the same as the final print version of my thesis. Except in the circumstances set out below, no emendation of content has occurred and I recognise that minor variations in formatting may occur as a result of the conversion to digital format. Access to my thesis may be limited for a period of time specified by me at the time of deposit. I understand that if my thesis is available online for public access it can be used for criticism, review, news reporting, research and private study.	
Digital Format - Masters theses	
I certify that a digital copy of my thesis deposited with the University is the same as the final print version of my thesis. Except in the circumstances set out below, no emendation of content has occurred and I recognise that minor variations in formatting may occur as a result of the conversion to digital format. Access will normally only be available to authenticated members of the University of Auckland, but I may choose to allow public access under special circumstances. I understand that if my thesis is available online for public access it can be used for criticism, review, news reporting, research and private study.	
Copyright (Digital Format Theses) (Tick **ONE box only**)	
☐	I confirm that my thesis does not contain material for which the copyright belongs to a third party, ***(or)*** that the amounts copied fall within the limits permitted under the Copyright Act 1994.
☑	I confirm that for all third party copyright material in my thesis, I have obtained written permission to use the material and attach copies of each permission, ***(or)*** I have removed the material from the digital copy of the thesis, fully referenced the deleted materials and, where possible, provided links to electronic sources of the material.

Signature Z Arden Date 27/02/2017

Comments on access conditions

Faculty Student Centre / Graduate Centre only: Digital copy deposited ☐ Signature Date

Small Change?
Adaptation and its Costs, in Experimental Evolution

Zachary N. Ardern

A thesis submitted in partial fulfilment of the
requirements of the degree of
Doctor of Philosophy in Biological Sciences
The University of Auckland
New Zealand

2016

0.1 Abstract

Trade-off is an important feature of theory concerning adaptation across environments. This thesis develops concepts and tests theories concerning trade-offs, using tools from experimental microbial evolution in the yeast *Saccharomyces cerevisiae*. Data used includes variants from whole genome sequencing and phenotypic data from Biolog TM multi-well plates. Both of these techniques were used to analyse populations from previous experiments over 300 generations of adaptation in chemostat and serial transfer conditions conducted by Jeremy Gray and Matthew Goddard.

It is shown that sexual status impacts on molecular evolution, changing the spectrum of variants observed in adapting populations. Mutation rate and sexual status are also shown to play a role in shaping the evolution of phenotypic trade-offs across environments, and groundwork is laid for future research in this area, using the increasingly accessible tools for high throughput microbial population research.

This work also contributes to theory in the field. It is argued that trade-off model in biological systems is correlated with genomic complexity, and that antagonistic pleiotropy is the dominant mechanism of trade-off for microbes under 'normal' conditions, such as in the absence of a mutator phenotype. Theory concerning trade-off and adaptation is also briefly applied to issues underlying the development of antibiotic resistance in bacterial populations.

0.2 Acknowledgements

Matthew Goddard: Mat, your open-minded support, and modelling of academic excellence have both been invaluable.

SBS, University of Auckland: Thanks to members of the Goddard group for help with numerous practical details and NZGL staff Dan Jones and Peter Tsai for early help with the vagaries of bioinformatics. Special thanks to Jeremy Gray for doing much of the experimental work that is analysed in this thesis. Ryan Estep for crucial help with an analysis script.

Mum & Dad: I am so grateful for all you have done for me.

Friends: The Evangelical Union, Clapham, & Auckland EV. Reuel Baptista, Don & Karene Biggs, Simeon Brown, Tim Conder, Esther Dale, Sidney Diamante, Selwyn Fraser, Eunice Hiew, Annaliese Johnston, Wesley Webb, Choonwei & Susi Wee, Tom Yates. Thank you for encouraging me to pursue a Christian mind.

Special thanks to Christina McDonald for your help over the last few months of thesis writing.

0.3 Preface

"I believe … that the history of science proves that advance is not only due to the discovery of new facts, but also to their correct interpretation: a true conception of natural processes can only be arrived at in this way. It is chiefly in this sense that the content of these essays are to be looked upon as research."
August Weismann, 1889.

In this thesis I have focussed on conceptual and definitional questions, in the belief that empirical investigation is aided by clear theory. The main question examined is the molecular basis of adaptation in the baker's yeast *Saccharomyces cerevisiae* and the phenomenon of trade-off across environments, where adaptation commonly leads to corresponding fitness loss in alternate conditions.

I have been privileged to present aspects of this work and related research at conferences hosted by NZ Next Generation Sequencing, American Society for Microbiology, the International Conference on the Evolution and Transfer of Antibiotic Resistance (ICETAR), the Ian Ramsey Centre for Science & Religion at Oxford, the British Society for the Philosophy of Science, and the Australasian Mycological Society. Diverse elements of the thesis have therefore benefitted from feedback in these forums.

0.4 Table of Contents

Chapter 1 - Introduction

1.1 Summary of Chapters

I. The first chapter gives an overview of the history of trade-off research, and the particular background of the experimental work reported in chapters 3 and 4.

II. The second chapter discusses and defines concepts important in experimental evolution and debates over trade-offs.

III. The third chapter presents the data from whole genome sequencing of 21 populations from an experiment previously conducted by Jeremy Gray and Matthew Goddard. Some features of SNPs, INDELs, and CNVs in each population are reported.

IV. The fourth chapter presents phenotypic data concerning metabolic trade-offs in mutator and non-mutator sexual and asexual populations from a second experiment previously conducted by Jeremy Gray and Matthew Goddard. It shows differences in the evolution of trade-offs in metabolism, due to sexual status, selection pressure, and mutation rate and situates this in the context of prior work.

V. The fifth chapter reanalyses previous claims of mutation accumulation in the Gray-Goddard populations, presenting four new arguments for the conclusion that antagonistic pleiotropy was the dominant driving mechanism behind the inter-environmental trade-offs observed.

VI. The sixth chapter proposes and defends the hypothesis that the mechanism of trade-off observed in populations depends partly on their genomic complexity. Specifically, it is proposed that more complex genomes, such as those of multicellular eukaryotes, are associated with what has been called 'mutation accumulation', while we should expect indicators of 'antagonistic pleiotropy' in microbial genomes.

VII. The seventh chapter is an overview of theory and previous results concerning the impact of sexual reproduction and gene-flow on trade-offs between environments. This chapter has been published as an edited book chapter. (Ardern, Zachary N., and Matthew R. Goddard. "Investigating Trade-Offs in Sexual Populations with Gene Flow." In *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, edited by Pierre Pontarotti, 245–62. Cham: Springer International Publishing, 2015.)

VIII. The eighth chapter discusses environmental complexity and trade-offs, with application to fitness costs resulting from the development of antibiotic resistance in bacterial populations. Some of this material was presented at the International Conference on the Evolution and Transfer of Antibiotic Resistance (ICETAR) in Amsterdam, July 2015.

XI. The ninth chapter concludes the main body of the thesis, with a discussion of sex and trade-off in light of the previous chapters' research.

Appendix A contains additional methods, used in DNA extraction (preparation for DNA sequencing results presented in Chapter 3.

Appendix B contains Unix scripts I developed for the analysis of the whole genome sequence data presented in Chapter 3.

Appendix C contains additional results from whole genome sequencing (Chapter 3).

Appendix D contains data from the Biolog plate experiment (Chapter 4).

1.2 Relationships between the Chapters

The driving question behind all of this research is the appropriate conceptual basis for analysing fitness trade-offs between environments. Across the different chapters I argue that theory on the mechanisms involved needs to be clarified, and that trade-off due to selected variants (antagonistic pleiotropy) is more prevalent in microbes under the conditions of laboratory evolution than trade-off due to selectively neutral variants (mutation accumulation).

There are three major strands to this thesis. The first is empirical evidence on the molecular basis of adaptation and trade-offs across harsh environments, in sexually and asexually reproducing yeast populations. The second is a focus on the mechanism of 'mutation accumulation', particularly the contribution of mutation rates to the development of trade-offs in microbes under some conditions. The third is constituted by conceptual issues in evolutionary theory relating to intra-environment trade-off.
In the first strand, Chapter 3 presents empirical evidence on the genetic basis of adaptation from

whole genome sequencing, and some of the relationships between different types of variants and inter-environmental trade-offs observed in these populations. Chapter 4 presents phenotypic data on trade-offs in sexual and asexual mutators and non-mutator populations.

In the second strand, Chapter 5 interprets previously collected data from the Goddard lab and infers antagonistic pleiotropy rather than mutation accumulation as the molecular basis of trade-offs. Chapter 6 argues that mutation accumulation is generally more prevalent in organisms with more complex genomes and so is not expected in bacteria apart from populations with very high mutation rates.

In the third strand, Chapter 1 introduces key conceptual issues for trade-offs, Chapter 7 focuses on the interaction between trade-off and the evolution of sex, and Chapter 8 applies trade-off theory to bacterial populations and some issues in the evolution of antibiotic resistance.

1.3 Novel Research

The core of this thesis is the genomic analysis of experimental evolution, and its interpretation, against a conceptual backdrop of models for the genetic basis of trade-offs across environments. Here I briefly review what is novel about the research conducted.

Whole population sequencing at high depths in yeast has only been conducted in recent years, and until very recently (McDonald et al., 2016) has not been used to compare the evolution of sexually and asexually reproducing populations. The research presented in Chapter 3 is unique in comparing molecular evolution in diploid sexual and asexual populations founded from isogenic ancestors.

To my knowledge, phenotypic microarrays have only rarely previously been used to study trade-offs, and have not been used in studying the 'side effects' of adaptation in mutator and non-mutator populations. There is potentially wide scope for incorporating this technology into future experimental evolution studies, to help give more comprehensive sense of the complex phenotypic changes occurring in evolving microbial populations.

This research presents conceptual advances in clarifying what trade-offs are as distinct, for instance, from ecological specialisation. It also clarifies the proper boundaries of the antagonistic pleiotropy and mutation accumulation models of trade-off, gives examples of phenomena such as hitch-hiking of deleterious variants that fits neither model, and begins to elucidate what should count as appropriate

evidence for each of the two main models proposed.

1.4 Trade-offs

Trade-offs across environments occur when a population's adaptation to a new environment is associated with changes to its genome which are detrimental in an alternative environment (such as the ancestral environment) when compared to the fitness of the ancestral (pre-adaptation) genome in that alternative environment. They have tended to be particularly closely associated with, and sometimes identified with the model of 'antagonistic pleiotropy', where the variants responsible for adaptation to one environment are directly responsible for the fitness cost in the alternate environment. Adaptation and ecological specialisation without trade-off are apparently common (Fry, 1996), and as will be discussed, other mechanisms can cause trade-off, and these mechanisms are sensitive to genomic and environmental context. The model that is commonly discussed as an alternative to antagonistic pleiotropy is mutation accumulation, where variants that accumulate in an adapting population through neutral processes are the cause of a fitness cost in an alternate environment. Many concepts important in discussions of trade-off, including comparison of the two main models proposed are discussed in more depth in future chapters, including Chapters 2 and 7.

It is important to situate the conceptual work within the broader literature on trade-offs between traits. There is a large literature on trade-offs between traits within an organism, operating within a particular environment, particularly what are termed 'life history traits'. That is, there are often apparent trade-offs between different aspects of an organism's life history, such that not all traits can simultaneously attain their maximum possible value.

Here are two ways that the concepts of inter-trait and inter-environment trade-off could be related to each other. Firstly, fitness within a particular environment could be considered as one trait among many instantiated by an organism, just like life history traits or other phenotypic effects. On this view, inter-environmental trade-offs are essentially the same kind of phenomenon as inter-phenotype trade-offs such as in life history traits. On an alternative account, the two could be distinguished by pointing to a difference in trade-offs across-environments and trade-offs across life-stages or across other phenotypes.

For the typical examples of life history traits, both traits involved in a trade-off (e.g. senescence and fertility) are components of a single measure of fitness for a particular organism in a given context, whereas inter-environmental trade-offs affect two different fitness measures, in the different

environments. Fitness is not divided across the life-span of an organism (it makes little sense to speak of evolutionary fitnesses at different ages, as it is the overall reproductive success that is causal in Darwinian evolution), and is also not divided, as 'fitness', into trait components (such a division makes little sense, as again it is the overall success of the organism that is causal), whereas fitness in different environments is given different measures. In other words, there is a clear division possible for fitness in different environments, while such a division, while perhaps possible, is not clearly available in the case of different phenotypic traits of an organism within a single environment.

There are additional reasons for why I find it helpful to distinguish these types of trade-offs - firstly as it is conceivable that a particular genetic model of trade-off may apply to life-history traits but not to inter-environmental traits. Secondly, trade-offs at the level of traits may not correspond to trade-offs in fitness, which are potentially drivers of evolutionary change.

Trade-offs between phenotypic traits within a particular organism in a given context are restrictions on optimisation which are intra-genomic. Because both traits are instantiated in the life of the organism, if a single variant is responsible for optimising one trait while being deleterious for the other then the conflict is unavoidable. Trade-offs between environments on the other hand can be merely hypothetical if the environment where there is a fitness cost is not encountered or is able to be avoided. Trade-offs between life history traits, if both traits are components of fitness in the current environment, will always have a fitness cost. Trade-offs between environments have a hypothetical cost, but this needn't always be realised. Note, it is also possible that some trade-offs between traits will be actualised but not impact fitness if a trait is not a component of fitness in a particular ecological context.

The focus of this thesis, then, is on trade-offs which are between environments and which occur at the level of a whole organism (by virtue of being trade-offs specifically in fitness), and consequently have effects on the distribution of alleles in a population.

1.5 Costs of Sex

The evolution of sexual reproduction in spite of apparent costs is a topic broached at various points in this thesis, particularly in Chapter 7. Here I give an overview of the costs of sex, providing context for the general search for the benefits of sex, found in Gray (2011), and the search for the molecular basis of these in this thesis.

Sexual reproduction, or something analogous, is found in all major branches of the evolutionary tree.

Its definition, origin and maintenance has been a topic of lively discussion in the evolutionary biology literature for many decades. On first appearance, sex is a less efficient form of reproduction than asexual processes. Here I clarify the distinctions that can be made between sexual and asexual reproduction in the context of model microbial systems developed to test theories about the maintenance of sex, and highlight aspects of the differences in reproductive mode that can plausibly be associated with costs and benefits. This discussion is extended into considering trade-offs across environments.

Gray, (2011) following Goddard, (2007) defines sex as involving recombination, random assortment, and syngamy. These terms refer to meiotic recombination (occurring between non-sister chromatids), random assortment of chromosomes, and the event of 'fertilisation' (fusing of gametes). It is difficult to clearly delineate between sex and asex in the biosphere, as there are many different reproductive modes. These include some parthenogenic organisms which undergo meiosis, including random assortment of chromosomes but not syngamy (Bell, 1982).

A key feature of the view of sex that we are interested in that is not captured in the above definition is that two parents are generally involved. Out-crossing seems an integral part of sex, insofar as it is associated with genetic variance, but it should be kept in mind that selfing is very common in some systems. The actual proportion of reproductive events that involve meiotic recombination, random assortment of chromosomes and syngamy can also vary widely.

Yeast, having isogamous gametes, does not face the vaunted two-fold cost of sex. There is a cost to sex, but the basis for it has minimal connection to any literal 'two fold cost'. What then, are plausible costs of sex in the Goddard yeast system? Here are a range of factors listed by Gray (2011), summarised and with additional comments:

Incorporation into messy genomes. The potential for alleles to be incorporated into a less beneficial background genome following recombination. This is particularly detrimental if positive epistasis is widespread.

Mate finding – with very high population densities as in the system assessed, this is less of a barrier.

Cellular, i.e. energetic costs, such as maintaining extra functional systems for meiosis. Though, note Stearns (1990), who shows that sexual reproduction involves a cost of "little energy but much time". Meirmans et al. (2012) argue that "The biochemical cost of meiosis is only relevant in very

small organisms, where it can be up to ten times that of asexual reproduction through mitosis". Note this is higher than a two-fold cost. Whether energy cost will impact on evolutionary fitness measured by way of a rate of population growth will depend crucially on environmental conditions.

Number of divisions. Meiosis involves two divisions, while mitosis involves one. This seems a more appropriate process to posit as the basis of a "two-fold cost" of sex than an assumption of anisogamy. The actual evolutionary cost will depend on the frequency of reproduction by sexual and asexual populations.

In summary it is not clear that the costs of sex are particularly high in single-celled eukaryotes with isogamous gametes and high population densities, with a key question in microbes being the availability of cellular energy.

Another factor which could be labelled a 'two fold cost' is genome dilution – a parent only transmits half of its genome to any progeny formed (Lehtonen et al., 2012). From a 'selfish gene' perspective, this would be a cost, but from the perspective of the survival of the species, there is no cost here unless genes affecting rates of reproduction differ between the parents so as to decrease the fitness of the offspring compared to the average fitness of the parents.

An unusual feature of the yeast system used is that the 'fitness' being measured is mitotic fitness rather than a more comprehensive measure. In a more 'natural' system, it may be that sex is more costly, as the time taken to reproduce sexually could, in such a system, be used for asexual divisions, while no such divisions occur during the sporulation process in the Goddard-Gray system. One consequence of this situation is that a decrease in the efficiency of sexual reproduction (e.g. the meiosis component) in this system will have no direct effect on fitness.

Epistasis occurs when there are non-independent effects of genetic loci on a trait. The extent, sign, and magnitude of epistasis, and whether these can be generalised across genomes, are matters of debate. In order for recombination to be favoured under direct selection, negative epistasis must occur among the loci under selection, and the magnitude of epistasis must be relatively non-variable. Peters & Lively (2000) suggest "epistasis must be weak to select for increased recombination, but epistasis must be strong for a sexual population to outcompete an asexual one." This, and other commentary in the literature suggests that there is a dynamic relationship between the evolution of sex and the development of epistasis in genomes. Epistasis is one reason that it is difficult to single out particular mutations as definitively the cause of a particular fitness effect.
Future work comparing experimental evolution in sexual and asexual systems could disentangle

different elements to sex. For instance the costs and benefits of selfing versus outcrossing, isogamy vs anisogamy, and the role of ploidy in the relative costs and benefits of sexual reproduction.

1.6 Introduction to Experimental Evolution

History

Experiments capturing multiple generations of microbial population growth and adaptation in the laboratory have been conducted and interpreted for the last 130 years, since at least the Rev. William Dallinger's 7-year experiment with protists in an incubator ended with equipment failure in 1886. Dallinger was a Methodist minister and a leading microscopist. The experiment which may have been the first in microbial evolution involved slowly raising the temperature of a controlled environment containing protists (not bacteria, contra Kassen, 2002) from approximately 15 degrees Celsius to 70 degrees Celsius over the course of 7 years, with some time at various points to allow adaptation to the new higher temperature environment (Hass, 2000). As the end-point population was unable to survive in the initial conditions, this appears to also be an instance of trade-off, the phenomenon at the centre of this thesis. Dallinger wrote to Darwin about his research with adaptation in protists, but Darwin seems to have missed the importance of the work for observing evolution in real time, responding by proposing a different line of research into the presence of algae in hot springs (Kassen & Rainey, 2004).

Now I briefly review key contributions to experimental evolution in the work of four key figures subsequent to Dallinger, each with different approaches – Jacques Monod, Daniel Dykhuizen, Graham Bell, and Richard Lenski – and its relation to this thesis with its focus on trade-offs across environments.

Jacques Monod (1910-1976) is known for his work explicating the regulation of the lac operon in E coli and semi-philosophical work exploring neo-Darwinism. His main contribution to experimental evolution, in addition to his many contributions to molecular biology and microbiology which have indirectly been of use in the field, may have been to promote the chemostat as an experimental set-up for microbial evolution. This system of continuous flow culture allows for an approximately constant population growth rate.

Daniel Dykhuizen (currently at the State University of New York) has been influential in developing techniques used in experimental evolution. He has developed the use of chemostats to achieve

continuous flow cultures for experimental evolution with bacteria, and theories of population genetics that explain adaptation in this system. Evolutionary populationary genetics as applied to bacteria has been a focus of his research. For instance, he was co-author of an important early review of the population genetics of adaptation in chemostats (Dykhuizen & Hartl, 1983).

Graham Bell (currently at McGill University) has done extensive research in ecology and evolution and is perhaps best known for his work on the 'queen of problems' in evolutionary biology, the evolution of sexual reproduction. He has published many papers on experimental evolution in systems including the algae *Chlamydomonas reinhardtii*. These studies include important data on trade-offs across environments following adaptive laboratory evolution in this species.

Richard Lenski is the principal investigator on the well-known 'long term experimental evolution' project growing *E. coli* bacteria for more than 64,000 generations over the last 28 years. This relatively simple serial transfer study has produced important data relevant to many aspects of adaptation including the rate of evolution, parallelism across replicates, and the evolution of a new phenotype. Lenski's group and some of his students have been at the forefront of using next generation sequencing technologies in the genetic analysis of experimental evolution.

1.7 Applications and Questions for Experimental Evolution

Other than the inherent theoretical interest of the dynamics of and molecular processes undergirding microbial evolution in the laboratory, pressing practical challenges to which this work may apply play some role in motivating experimental evolution research. Adaptation is observed in the short-term in the response of organisms to anthropogenic changes in the environment. These include adaptation to chemicals widely used by humans such as antibiotics, antifungal agents, herbicides, and pesticides; adaptation to heavy metal contamination, as well as ecological shifts in response to anthropogenic climate change.

The development of antibiotic resistance, discussed further in Chapter 8, is a topic of great interest to the health sector globally. Adaptation of bacteria to antibiotic conditions, as well as of yeasts to antifungal drugs has been observed in experimental media controlled for this purpose. Recent research has investigated many aspects of the problem including adaptation to multiple drugs at once, responses in new antibiotic environments following adaptation to a particular drug, fitness costs in alternate environments, and genetic variants responsible for adaptation.

Studies relating to the stress conditions assessed in Gray and Goddard's 'sex and gene-flow' experiment are reviewed later in this introduction. Some others are described here. Adaptation to the unique environmental conditions of "Evolution canyon" in Israel has been studied for various organisms, including *S. cerevisiae*. These conditions include contamination by heavy metals, to which the yeast populations gained tolerance. Understanding adaptation to toxins, salt, and other stressful environmental conditions is important in heavily contaminated environments such as some former mining sites, and the molecular basis of each has been studied. For instance, strains collected from Evolution Canyon were found to have enhanced function in the metal efflux pump PCA1, and strains from other locations showed divergence in nucleotides critical for this function, an example of trade-off as the trait is beneficial in some environments and appears to be selected against in others (Chang & Leu, 2011). A possible application of this research, if generalisations can be made across species or if experiments can be up-scaled to plants, is efforts to expand agriculture into areas not traditionally available due to poor soil quality.

Adaptation to changes in temperature has been a focus of many experimental evolution studies, including projects with bacteria, yeast, and fruit flies. With increasing recognition of the effect that anthropogenic climate change is likely to have on ecosystems and agriculture, this research has gained a new pressing importance. Recent work relating experimental evolution to environmental change has been done in the research group of Graham Bell at McGill University (Samani & Bell, 2016).

In addition to the large-scale practical challenges of general human interest discussed, there are large questions in evolutionary theory for which insights can be gathered through microbial experimental evolution. Out of many that could be addressed, I will outline the relevance of experimental evolution to determining the molecular basis of adaptation.

The types of mutation occurring in evolution is a topic of intensive current research. Mutations observed in experimental evolution and potentially involved in trade-offs across environment range in size from point mutations to large-scale structural changes in chromosomes. Point mutations are well understood as a potential basis of adaptation in microbes. In approximate order of average relevance to adaptation they include non-coding mutations, synonymous mutations, non-synonymous (or mis-sense) mutations, and nonsense mutations (which introduce a stop codon). The introduction of a premature stop codon into a coding sequence is sometimes termed a 'loss of function' mutation, as it is assumed that such a mutation will result in a non-functional protein product. Point mutations have been extensively catalogued in experimental evolution, for instance in the important study by Lang et al. (2013) of 40 haploid yeast populations of different sizes where alleles introduced by point

mutation were tracked by high depth population genome sequencing across 1000 generations of adaptation. Indels are also quite often referred to in experimental evolution studies but may be of less functional importance, occurring less often and found, in my interpretation of the literature, as parallel adaptations to a lesser extent than single nucleotide changes across replicate populations.

Changes in the copy numbers of genes have received less attention in the experimental evolution literature, with the main exception until recently being recognition of the importance of copy number variants in nutrient-limited conditions for adaptation in the yeast *S. cerevisiae*. Copy number changes have previously been assessed using micro-array technology (Gresham et al., 2008). As shown in Chapter 3 of this thesis, these variants can now be observed with next generation sequencing, although very few studies have made use of this technology with pooled population sequencing for assessing copy number variation, perhaps based on the mistaken assumption that such analysis is only relevant for individual genomes. Copy number variants have also recently been observed in adaptive laboratory evolution in other organisms such as the nematode worm *C. elegans* (Farslow et al., 2015). Larger scale structural changes to microbial genomes are also observed during the course of adaptive laboratory evolution experiments. Karyotypic changes in non-pathogenic yeasts are reviewed in Gerstein & Berman (2015), who cite examples where polyploidy is beneficial and others where genome reduction to a haploid state is. Sunshine et al. (2015) show in *S. cerevisiae* the fitness effects under nutrient limitation of a range of large scale changes in chromosomal arrangement.

Experimental evolution may have direct bearing upon questions of a practical nature, but this project aims to provide insights on broader questions in evolutionary theory. With next generation sequencing it is possible to watch the spread of de novo mutations or standing variance through a population, and as such to observe the dynamics of the fixation and loss of alleles in populations, as well as the complexities of clonal interference, frequency-dependent selection and other mechanisms which maintain alleles at intermediate frequencies. The important study by Lang et al. (2013) discussed above has shown the importance of clonal interference in asexual populations. The replicability of microbial experimental evolution along with the capacity to tightly control experimental conditions allows for testing hypotheses concerning the effects of different variables on the adaptive process. Such variables include selection pressure, population size, and environmental variability. The recent increase in the ability for in-depth analysis of experimental evolution has occurred against a back-drop of extensive molecular and functional understanding of microbial genetics and metabolism, particularly in the experimental models *E. coli* and *S. cerevisiae*. This allows for inferences concerning the functional effects of genetic variants, and increasing ability to draw connections between genetic and phenotypic data.

1.8 Microbial Model Systems

The use of microbial populations in experimental evolution, and the application of experimental results from these systems to broader evolutionary questions require justification. The advantages of microbes for evolutionary analysis consist especially in their short generation times, compactness and cost-efficiency for experimental set-up, and ease of cultivation. Eukaryotic microbes such as the yeast *Saccharomyces cerevisiae* have an advantage over prokaryotes as a model system by virtue of a more similar genome structure to multicellular eukaryotes such as humans, with multiple chromosomes, as compared to a typical prokaryote's single chromosome. In addition the capability to undergo meiosis is another important similarity to multicellular eukaryotes. An advantage of microbes is that the experimental conditions in which they are evolved can be controlled relatively precisely, such that we can determine population size, growth rate, and selective conditions over hundreds of generations, as is typically not possible for multicellular eukaryotes.

Microbes have some disadvantages as models for evolution conceived of more broadly as well. Experiments with microbes have tended to assess evolution at the level of populations, with limited awareness of intra-population diversity; this could, however, be improved with different approaches to sampling from and sequencing populations in future. Microbes are not multicellular, and do not have phenotypic traits differentiated in the way in which traits in multicellular organisms can be.

As such, the transfer of microbial experimental results to theories about evolution in multicellular eukaryotes cannot be easily justified tout court, and must be approached case by case. The difficulties in extrapolating from microbes to macrobes is an example of the general situation arising from the fact that there are few if any biological laws analogous to the laws of physics. Biology as standardly conceived is a result of many historical contingencies, and frequent evolution of novelty over the tree of life – as such, extrapolation must always be tentative, as evolutionary novelties may mean that putative trends are not found to be continued in different taxonomic groups. A key use of models such as microbial systems will be to refine hypotheses which can be tested, at least in part, in multicellular eukaryotes.

1.9 Methods for Experimental Evolution

The methods for experimental evolution have developed since early experiments with bacteria in the 1970s. Here I briefly review the organisms that have been used, the experimental conditions, the

measures taken, recent developments in DNA sequencing, and the analysis of sequencing data, particularly focusing on areas that will be directly relevant to later data and discussion or proposed future work.

Organisms and Strains

The most commonly used organism for experimental evolution is probably the bacterium *Escherichia coli*, used in many studies including Richard Lenski's famous Long Term Evolution Experiment conducted by Lenski and many graduate students, at Michigan State University. Arguably the first experimental evolution research project, conducted by William Dallinger, made use of single celled organisms, and they have been used in the field since. However, for a time around the 1950s the fruit fly *Drosophila melanogaster*, prominent as a model organism in genetics, was perhaps more high-profile – e.g. a paper testing the contributions of drift and natural selection by Dobzhansky and Pavlovsky (1957). Since the 1980s, microbial populations have been used for longer-term experiments with higher replication of lineages when compared to earlier studies. These have included *E. coli* and the popular single-celled eukaryotic model organism *S. cerevisiae*, as well as many others referred to in this thesis, including important results from the bacteria *Pseudomonas aeruginosa*, *Pseudomonas fluorescens*, *Methylobacterium extorquens*, *Mycobacterium tuberculosis*, and the algae *Chlamydomonas reinhardtii*. Recent work has also included some using non-microbial systems such as *Drosophila* species, and the nematode worm *C. elegans* – on which there is a recent review co-authored by Jeremy Gray (Gray and Cutter, 2014). The use of multicellular organisms also allows for predator-prey co-evolution experiments, for instance between *C. elegans* and the bacterium *Baccillus thuringiensis* (Schulte et al., 2010).

With the genomics revolution in the 21st Century, the particular strain used of the organism decided upon is increasingly recognised as important. Different strains of the same microbial species can differ significantly in gene content, and genetic engineering and sequencing together allow for precise specification of 'ancestral' population genomes for experimental evolution projects. Genes for conjugation in bacteria or meiosis in yeast may be particularly important in determining evolutionary dynamics, and can be knocked out to prevent these processes from occurring. Comparing the fitness of ancestral and derived strains is most accurately measured through competition experiments, which require at least one population to carry a genetic marker with a neutral (or minimal known) fitness effect to differentiate them. One marker which may be particularly useful is the fluorescent red protein mCherry, used for instance in a seminal study of 40 haploid *S. cerevisiae* populations by Lang et al. (2013) – the resulting colour difference can be exploited for automated counting with flow cytometry.

Experimental Conditions

The key items of equipment in the experimental set-up for microbial experimental evolution are the flasks and media to be used for multi-generation culture growth. The central question is perhaps the decision between continuous flow and batch culture. In the chemostat, which is the most common continuous flow culture method, media is pumped through flasks at a constant controlled rate, allowing approximately constant population sizes to be maintained while keeping the microbes in the exponential growth phase yet still under selective pressure. In batch culture, which for experimental evolution requires repeated serial transfer (for instance one transfer per day) of a portion of culture to a flask of fresh media, the population size and growth rate vary greatly over the course of the experiment. Batch culture is much simpler and experimentally tractable, but creates complex population dynamics (particularly after taking into account changes over the course of many generations of an adapting population), which have perhaps received inadequate attention in discussions of experimental evolution results. Continuous flow culture is more complex to set up and run, but in theory provides simpler experimental conditions. This relative simplicity of the resulting population dynamics arguably makes the task of interpretation of the wider relevance of any results easier, assuming that most natural populations are under conditions more closely approximating the constant expansion observed in chemostats rather than the varying growth rates and repeated bottle necks of batch culture. A different kind of continuous flow culture is created in turbidostats. Flasks are similarly constantly fed a supply of nutrient-containing media, but the supply is controlled so that nutrient limitation is never experienced by the microbial population (Gresham & Dunham, 2014). Rather than the flow rate of media addition being held constant as in a chemostat, the variable that is set to be constant is the cell density, measured as the turbidity of the solution. Nutrients are all set in excess rather than one being at a growth-limiting concentration as in typical chemostat conditions.

The environments used in the particular populations analysed are discussed later in this chapter. In recent decades, microbial evolution experiments have been conducted in media which was nutrient limited, high in salt or heavy metal concentrations, under various temperature changes, fluctuating in environment temporally, or spatially differentiated in various ways. Each of these environmental conditions has a different selective effect. Other factors that have been the subject of controlled studies have included population sizes, sexual vs asexual reproduction in populations, and studies of sexually reproducing populations with and without gene-flow between environmental niches.

Experimental Measures

There are multiple informative quantitative results that can be measured from carefully controlled microbial experimental evolution projects. Historically the main measures were survival under

different conditions or population growth rates, but now measures include a range of ways to detect phenotypes, genome sequence variants, and gene expression levels.

Population growth rates in the selective environment are used to determine evolutionary fitness, with various particular methods for this conversion discussed in Chapter I. Evolutionary fitness in microbial populations is best measured through competition of samples of ancestral and derived populations, determining the relative fitness of these genomes in the selective environment. Comparison of fitness in alternate environments between populations with ancestral and derived genomes tests for fitness trade-offs.

Phenotypes affected by adaptation in culture media, whether achieved through batch culture or continuous flow methods, can be measured in a wide range of ways. Differences between strains in their growth rates on different substrates can indicate phenotypic differences. Software such as DuctApe (Galardini et al., 2014) is available for analysis of this data in conjunction with sequencing results. Metabolism on different substrates can be measured similarly to population growth, but with the addition of a dye such as tetrazolium which measures the extent of oxidation occurring within a particular well containing substrate and a population sample. The effect on a cell's protein composition or membrane properties can also be quantified, and perhaps the development of metabolomics will allow for even better understanding of the actual cellular effects of adaptation beyond changes in growth rate.

Genome sequence variants will be discussed in much greater depth throughout this thesis. The expression of genes and non-coding RNA sequences can be quantified and compared across treatments. There are two main methods for this, although there are many different specific protocols. Micro-array technology can show changes in the expression of particular targeted genes, through hybridisation to probes attached to a chip; thousands of such probes can be attached, allowing for analysis of expression levels of thousands of genes. In a set of techniques with potentially more precision, RNA can be sequenced and the reads aligned to a reference in approximately the same way as for normal DNA genome sequencing, after reverse transcription of RNA sequences into DNA.

Sequencing

Each of the experimental measures discussed could be explored in depth, however here I will briefly review sequencing technologies and methodological options available, as issues concerning sequencing were central to this project.

Prior to sequencing, DNA extraction is required. An important part of the design of a sequencing component to an experimental evolution project is determining whether to extract from a single clone expanded into a homogenous population sample or from the mixed population, either a large enough sample to extract sufficient DNA for sequencing or a smaller sample expanded in 'benign' media. As discussed further at the end of this introduction, each option is associated with positive and negative aspects. Sequencing individual clones (selected by plating out and then expanded into a larger population for extraction) has the advantage of producing whole individual genotypes, and more certainty that reads found at low frequency are errors rather than representative of low frequency alleles; sequencing errors are easier to isolate. Sequencing a whole mixed population at a reasonably high depth however allows for inferences concerning allele frequency with a single sample. A rational decision about whether to sequence a mixed population or multiple individual clones may depend on the cost of library preparation and sequencing with the technology chosen, unless the requirement for whole individual genotypes is taken as a decisive advantage for clonal sequencing.

Another experimental design decision is whether to sequence multiple time-points from the course of the laboratory adaptation (e.g. samples from times points 0 generations, 50 generations, 100, etc.) or just the derived sequences. Starting with known ancestral populations saves the need to sequence the start point, although this may be worthwhile anyway in many cases to confirm the exact genome sequence begun with – as such it is typical to sequence the start and end point of the adaptation experiment. However, with the dropping cost of sequencing and increasing automation of various aspects of adaptive laboratory evolution, it is increasingly feasible to sequence population samples or clones from multiple time-points. With sufficient sequencing depth of populations or numbers of clones sampled this allows tracking of changes in allele frequency during the course of the experiment. With population sequencing at multiple time points, thanks to mathematical work published in Lang et al. (2013), it is possible to use the time-course data to more confidently isolate probable sequencing errors, increasing the probability of successfully calling low frequency allele variants. Whether for the sake of error-minimisation it is in fact better to combine sequencing of multiple time points with multiple individual clone samples at low depth or DNA from mixed populations at higher depth should perhaps be the subject of statistical analysis in future. Further gains in error-minimisation can probably be achieved through careful experimental design.

The DNA sequenced can be the whole genome, a specifically targeted region with the use of primers, or derived from RNA with the use of a reverse transcriptase enzyme and hence used as a measure of transcription. These different techniques can be utilised to answer different experimental questions. Whole genome sequencing is best used at high depth across a few replicates, to detect

genome level changes or to attempt to catalogue all variants observed during a period of adaptive evolution. Targeted sequencing of a particular region (perhaps genomic regions known to be a target of selection, such as the HXT loci on chromosome IV of *S. cerevisiae*) lowers the cost of sequencing and may facilitate sequencing of a greater number of replicates. Sequencing the product of reverse transcription of RNA, again potentially in targeted regions, can show gene expression, an experimental measure which is one step closer to the protein phenotype.

Early genome analysis observed large-scale changes in chromosomes, such as changes in the banding patterns of *Drosophila melanogaster* chromosomes. Genetics was revolutionised with the origin and expanded accessibility of Sanger sequencing. Sanger sequencing of whole genomes is a fairly laborious process. Double strands of DNA are artificially terminated with fluorescently or radioactively labelled dinucleotides, and when four separate reactions are conducted for the four nucleotides, the sequences terminating at different points can be run on a gel together, separating by length and from the bands representing each nucleotide termination point the DNA sequence can be read. A later development was the range of 'next generation' sequencing technologies which have higher throughput and lower cost. Pyrosequencing proceeds by the sequential incorporation of nucleotides pairing with a single DNA strand. This incorporation releases a pyrophosphate molecule each time, and this event is made visible through the use of enzymes resulting in light emission, which indicate the incorporation of whatever nucleotide was added. A limitation of this process is that when runs of the same nucleotide occur, the precise number cannot be determined, i.e. the method is error-prone for repetitive sequences. Technology originally developed by Solexa and now distributed by Illumina is particularly popular. It works by the sequential addition of labelled nucleotides which pair with oligonucleotides bound to a chip. Controlled sequential addition is made possible by reversible termination of the polymerisation process. Other technologies are available and being created, such as Pacific Bioscience's technology allowing sequencing of very long reads, but Illumina is the current best standard for microbial genomes.

Sequence Data Analysis

A wealth of bioinformatics tools are available for analysing sequence data produced by technologies such as Illumina HiSeq or MiSeq. Here I give a brief overview of four levels of analysis possible, graded by difficulty and extent of customisability. I give examples of major online tools, pre-packaged pipelines that can be installed, tools for individually customising pipelines, and individually written programs.

The most popular online tool for next generation sequencing analysis is the Galaxy server, which allows for read alignment to a reference, variant calling and many other analyses of data uploaded to

their server. This is best for small data sets, and while not as customisable as other approaches, can be used for the main components of small projects. Other online tools which I have used in this project include websites which provide functional analysis including statistical tests for significant over-representation of particular functions of gene lists submitted. One example is 'FunSpec', available through the University of Toronto (Robinson et al., 2002).

Pipelines already created for the analysis of re-sequenced microbial genomes are available as well. Breseq from the Barrick lab is one example (Deatherage & Barrick, 2014), able to be downloaded as one package and designed for finding variants relative to a reference genome in re-sequenced bacterial genomes. Run on a Unix platform, the series of open source tools used can find single nucleotide variants, insertions and deletions and chromosomal rearrangements in small haploid genomes. A degree of customisation is possible for those who understand the processes being harnessed in the pipeline. As well as full pipelines, there are other collections of key bioinformatics tools available to be downloaded together, making the set-up easier than trying to install every package individually from scratch. For the analysis presented in Chapter 3 I have used the Linux distribution "Bio-Linux" (Field et al., 2006), a project supported by the National Environment Research Council (NERC) in the UK. Many less popular packages have to be individually installed however, which is not always a straightforward process.

There are many open source software tools available, and due to their customisability and easy access some of these are becoming the standard for analysis of resequencing with alignment to a reference genome or de novo sequencing, following experimental evolution. These can be combined together into customised pipelines, for instance with the use of shell scripting in a Linux environment – the pipeline I have created is described in Appendix B.

Finally, some labs with advanced bioinformatics capabilities create their own programs to run new bioinformatics analyses. There are a wealth of specific programs that are available, some of them made for standard platforms such as the statistics program 'R'. Many such programs written for use in R are available in the collection 'Bioconductor' (Gentleman et al., 2004). Use of these more advanced programs was outside the scope of this thesis, but worthy of investigation for future work.

Specifics of the Goddard/Gray System

1.10 Sex and Gene-flow Experiment

This project, in Chapter 3, includes genome analysis of 24 populations of *S. cerevisiae* derived from an experiment conducted by Jeremy Gray and Mathew Goddard, detailed in Jeremy Gray's PhD thesis (Gray, 2011) and in an article in Ecology Letters by Gray and Goddard (2012a). In the original adaptive laboratory evolution experiment conducted by Jeremy Gray and Matthew Goddard, treatment regimes were controlled for 60 populations of yeast grown in continuous flow culture for approximately 300 mitotic divisions. Continuous flow was maintained with use of a temperature-controlled incubator environment, containing chemostat flasks rotating at 125rpm. Sexual and asexual populations were grown in one of two pre-sterilised nutrient-limited selective environments; a base media altered to be carbon-limited kept at a high temperature of 37°C ('hot C'), and the base media altered to be nitrogen limited with high osmolarity kept at 30°C ('osmotic N'). The 300 generations of adaptive laboratory evolution were interspersed with 12 rounds of sex, and/or gene-flow between populations grown in the alternate environment, for the appropriate populations. Gene-flow ranged across five proportions: no gene-flow, transfers of 10^{-6}, 10^{-4}, 10^{-2}, and full-mixing (50% gene-flow). In order to ensure as few differences between the initial strains as possible, the asexual strain was one engineered from the wildtype sexually reproducing strain 'Y55', with two genes involved in recombination and division (SPO11 and SPO13) deleted.

Interpretation of the fitnesses of the resulting populations found that sex facilitated adaptation and that sex with gene-flow across the two environments enabled selection for the best collection of alleles suited to multiple environments. Further research (Gray and Goddard, 2012b) suggests that sex may both assist in removing detrimental mutations and in selecting alleles which are beneficial across more than one environment. Given that sex with even high levels of gene-flow did not slow local adaptation, it was inferred that trade-offs in fitness in these environments were likely due to mutation accumulation rather than antagonistic pleiotropy. If trade-offs in 0 gene-flow conditions were due to antagonistic pleiotropy, where the same variants responsible to adaptation to the selective environment caused trade-off in the other environment, it was assumed that this would retard adaptation in the full gene-flow conditions as alleles selected in one environment would tend to be deleterious when the population is exposed to the other.

There are other possible interpretations of this data, however. Perhaps there is a subset of mutations selected in the full gene-flow conditions that are not antagonistically pleiotropic between

the two environments, while the trade-offs observed without gene-flow are due to different mutations which do exhibit antagonistic pleiotropy.

1.11 Environments Used

The environmental conditions used (i.e. 'stress' conditions in the chemostats) and their likely impact on adaptation also deserve some attention. The combination of two stress conditions to comprise each selective environment were chosen by Matthew Goddard so as to create complex environments. The environmental conditions chosen are associated with many potential genomic targets for selection to act on any variation in. For instance, the effect of salt stress is known to be mediated by many different genes, as shown for instance in a study of the effect of salt stress on the transcriptome (Posas et al., 2000), and through laboratory evolution with microarray analysis and population sequencing by Dhar et al. (2011).

According to Gresham and Dunham (2014) in chemostat conditions it is the low concentration of the limited nutrient that "defines the selection imposed on cells" and adaptation is typically achieved through more efficient uptake or use of the limited nutrient. The effects of other stressful conditions on selective pressure in chemostats is perhaps worthy of more attention.

Salt Stress

A wide range of stress conditions have been used in experimental evolution before, but nearly always with the use of a serial transfer (batch culture) protocol. Somewhat surprisingly, it appears that adaptation to salt stress has not been assessed in a chemostat environment previously - searches on SCOPUS and Google Scholar turned up no such studies. **Table 1.1** below lists some studies of adaptation to salt stress in microbes using serial transfer, along with a summary of their results, as an example of the kinds of methods and findings in the field.

Table 1.1

Study	Organism	Method	Generations	Measured
Lachapelle et al., 2015	**Algae - *C. reinhardtii***	**Serial Transfer**	**500**	**Growth rates**
Freshwater algae were exposed to increasing concentrations of salt. Inducible response to salt shown to increase over the course of the experiment				
Wu et al., 2014	**Bacteria - E. coli**	**Serial Transfer**	**73 daily transfers**	**Growth rates & Genome sequence**
Adapted to increasing salt concentrations. None of the adapted mutants showed increased tolerance to the nonionic osmolyte sucrose Mutations to these genes common across replicates: drug resistance pump (emrR), ribosomal subunit (rpsG), and starvation-related protein (sspA)				
Horinouchi et al., 2014	**Bacteria - E. coli**	**Rapid Serial Transfer**	**70 6hr-ly transfers**	**Growth rates**
E coli evolved to various stressors including high salt. Used 96 well microplates. Automated system.				
Dhar et al., 2011	**Yeast - S. cerevisiae**	**Serial Transfer**	**300 generations**	**Growth, Expression, Genome sequence**
Changes in basal gene expression, regulation of gene expression, and a high frequency single nucleotide variant in MOT2.				
Stoebel et al., 2009	**Bacteria - E. coli**	**Serial Transfer**	**250 generations**	**Growth, Expression, Genome sequence**
Conditions of salt stress after deletion of a transcriptional regulator. Populations rapidly evolved fitness compensation The adaptive variant, insertion of a mobile genetic element, was common across all replicates.				

Raised Temperature

Adaptation to high temperature has been studied experimentally in microbes at least since William Dallinger's early long-term experiment in protists. Studies in bacteria have included one testing for a trade-off effect in *E coli.* at high temperature following adaptation to low temperature (Bennett & Lenski, 2007). Temperature fluctuations have been used for a more complex environment, for

instance in Ketola et al.'s study (2013) of the opportunistic pathogen *Serratia marcescens*' response to changing temperature – an advantageous generalist phenotype was observed to evolve. Other studies include industrial and commercial applications, using biotechnology to create heat resistant strains, and for studying the process of fermentation.

The key genetic results from adaptive evolution to conditions of nutrient limitation appear to be easier to interpret than adaptation in other stress conditions.

Carbon & Nitrogen Limited

The main genetic change consistently reported for studies of yeast adapting to low carbon conditions is an increase in the copy number of the hexose transporter genes HXT 6 and HXT 7 found in chromosome IV of *S. cerevisiae*. Nitrogen limitation has also been studied in some depth, for instance by Gresham & Hong (2014).

1.12 Sex and Mutators Experiment

In Chapter 4, phenotypic micro-array analysis of a similar experiment previously conducted by Gray and Goddard is presented. The metabolic rates of 24 populations plus four ancestral strains was tested with the use of Biolog™ phenotypic microarray plates in two sets of environments; 95 carbon and 95 nitrogen sources.

In the original experiment conducted by Jeremy Gray and Matthew Goddard, four ancestral *S. cerevisiae* populations – sexually or asexually reproducing genotypes with or without an increased mutation rate – were evolved to harsh or benign environmental conditions through daily serial transfer of culture media over the course of approximately 300 generations. The experiment is detailed in Jeremy Gray's PhD thesis (Gray, 2011).

The wildtype strain was sexually reproducing with a normal mutation rate. The asexually reproducing strain was created through gene knockout of SPO11 and SPO13 – these two strains were also used for the sex and gene-flow experiment. A 'mutator' strain was created from each of these genetic backgrounds through additional knock-out of the DNA repair gene MSH2.

It was found that in the benign environment no populations increased or decreased in fitness after 300 generations of the serial transfer protocol. In the harsh environment however, there were fitness changes and these also differed across treatment groups. Fitness was higher in sexual than

asexual populations, as had been expected. There was no significant effect of mutation rate.

It was concluded (Gray, 2011) that the main advantage of sex is in assisting with beneficial mutation fixation rather than deleterious mutation clearance. As argued in Gray & Goddard (2012b) however, the higher fitness of asexual wildtype over asexual mutator populations suggests a role for deleterious mutations in decreasing the fitness of asexual mutators. The relative success of sexual mutators suggests that they have cleared deleterious mutations better than asexual mutators.

The role of population size is worthy of further attention. I suggest that the much larger population sizes in the benign environment (approximately 10 fold greater than mutators in the harsh environment, and 200 fold greater than non-mutator populations in the harsh environment) may have been largely responsible for the clearance of deleterious mutations. Similarly, population size may explain differences between sexual mutator and sexual non-mutator populations. I assume that most mutations are deleterious, and that increasing the mutation rate increases the proportion of mutations that are deleterious. Whether the population was 'mutator' or not made no difference for sexual populations. On my assumption, sexual mutators however would have more deleterious mutations than sexual non-mutators, so the sexual mutators must have cleared deleterious mutations more effectively than sexual non-mutators – perhaps due to the mutators' larger population sizes.

Chapter 2
Molecular contributors to environmental trade-offs in microbes

2.1 What are Trade-offs?

Adaptation to a particular environment is often associated with changes in other traits. Trade-offs in adaptation among traits is a phenomenon that was noted by Charles Darwin, who in the context of artificial selection in domesticated animals referred to the "mysterious laws of the correlation of growth" (Darwin, 1859). He realised these might be due either to natural selection acting directly or a result of constraints on resources shared between traits. Here we consider the more specific concept of evolutionary trade-offs between environments. The molecular basis of this phenomenon has been investigated particularly rigorously in microbes, and there are potential implications for a range of issues in medical microbiology, including antibiotic resistance. When considered more broadly, the molecular events undergirding trade-offs are relevant in studying the spread of invasive species, cancer cells, and adaptive responses to climate change.

Trade-off is one instance of a genotype-x-environment interaction; where the phenotype or fitness effect of a genotypic variant differs according to the environment in which it is expressed. For the purposes of this article, 'trade-off' is used to refer to an effect seen at the level of the fitness of organisms, where adaptation to one niche is associated with decreased adeptness in another niche relative to the pre-adaptation ancestral state. In support of this account see e.g. Fry, (1996). The phenomenon of interest can be considered a 'cost of adaptation' that is hidden until the population is translocated to a new environment, or a shift occurs in the environment. Discussion of the molecular basis of trade-offs has often contrasted two models; antagonistic pleiotropy and mutation accumulation (e.g. Cooper & Lenski, 2000). Under antagonistic pleiotropy, a population experiences a fitness cost in an alternate environment because of deleterious pleiotropic effects of variants that were adaptive in the selective niche. Under mutation accumulation, the population experiences a fitness cost in alternate environment because of deleterious pleiotropic effects of variants that accumulated in the selective niche due to neutral processes such as genetic drift. Following recent work by Jerison et al. (2014), we refer to a broader concept of a joint distribution of fitness effects; considering more than one environment in assessing the fitness associated with genetic variants. The definition of trade-off used here allows that a multitude of different molecular mechanisms, including antagonistic pleiotropy and mutation accumulation may undergird the effect of trade-off at the level of the fitness of a microbial or other population.

2.2 The Concept of Trade-off Defined.

As introduced in chapter 1, the concept of 'trade-off' is most commonly associated with constraints on development of particular traits caused by physiological factors i.e. trade-offs between traits expressed in the same environment, either simultaneously or at different points in the organism's life history. For instance, the genetics underlying an apparent trade-off between longevity and fecundity has been a subject of research over many years (Remolina et al. 2012). This chapter is concerned with trade-offs experienced between different environments in microbes.

Even within the context of inter-environmental trade-off, the concept of trade-off has received a range of different definitions. Here these are reviewed, with explanation as to why the definition I offer is the best in light of recent developments in technology in microbial experimental evolution and genome sequence analysis. Briefly, trade-off occurs when adaptation to one environment is associated with a cost in another. The elements of trade-off as defined here follow: Firstly I define trade-off as a property of an individual genotype rather than a population. Secondly it is a property of a whole genotype rather than just one individual locus, gene, or region. Thirdly it is measured in terms of evolutionary competitive fitness rather than any individual element thereof. Fourthly, the 'cost' of trade-off is considered a decrease relative to the ancestor rather than a decrease relative to another evolved genotype that may in fact be an increase relative to the ancestor. Fifthly, this fitness is ideally measured in direct relation to a base or ancestral genome identical apart from the trait(s) or adaptation events under consideration, rather than a strain potentially having accumulated other genetic differences. Taking these elements together, I propose that a trade-off can definitively be said to occur when a genotype's evolutionary adaptation to one ('selective') environment is associated with a decrease in fitness in an alternate environment, measured relative to the ancestor. Now I consider each of these facets in turn, in the reverse order, beginning with the proper comparison class.

Adaptation to one environment can be associated with costs in others. In order to determine that the variants associated with the process of adaptation were responsible for the fitness costs observed in alternate environments, a strain free of the variants associated with adaptation should be used for comparison. The best comparison is the ancestral genome itself, which can be available from frozen samples in microbial adaptive laboratory evolution. Note the phrase "variants associated with", as whether particular variants were themselves adaptive or not is a separate question discussed later. The precise genotypes being compared become important when the phenomenon of epistasis, interactions amongst genetic variants, is considered – as discussed later in this chapter.

The decrease of trade-off is measured relative to the ancestor. Sometimes a broad definition of trade-off is given, such that it is identified with or defined to include specialisation, i.e. differential adaptation across niches (e.g. Kassen & Bell 1998, Elena & Lenski 2003). Both concepts are potentially useful in studying habitat shifts, but it is important to differentiate between them. Differential adaptation entails a merely relative difference amongst evolved genotypes, where the maximum possible level of adaptation is not achieved in alternative environments because of specialisation to a particular environmental niche. Under these conditions, 'maladapted' individuals could still outcompete the ancestral state if both were translocated to the alternative environment where the relative cost is experienced. By contrast, where trade-off in the sense used here occurs, in alternate environments the ancestor would outcompete any derived individuals showing trade-offs. This is an important example of the context-dependence of the concepts of 'fitness' and 'fitness cost'. Some claims concerning fitness costs in the context of the evolution of antibiotic resistance are examined in more depth in Chapter 8.

Trade-off is best measured in terms of competitive fitness rather than elements of fitness. Trade-off is used to try to explain evolutionary outcomes, in which the primary feature of organisms which is of interest is their competitive fitness. The evolutionary explanation for why one lineage survives and thrives and another does not will be in terms of comparisons of fitness under different conditions. There are other aspects of cellular life that can be of interest in seeing side-effects of adaptive processes, such as the measures of metabolism discussed in Chapter 4, but these need to be distinguished from fitness itself. As Futuyma & Moreno (1988) note, "a trade-off discerned by a reductionist analysis of a single character may not operate at the level of the organism, in which compensatory features come to play." Another consequence of this view is that because trade-off is measured in terms of fitness, it cannot be given a narrow definition where it is strictly identified with the particular model of antagonistic pleiotropy (e.g. Leiby, 2014), which makes a claim concerning the relationship of the particular causal variants to selective pressure in the adaptive environment. The same fitness effect could be observed whether the variants were maintained in the population through either selective or neutral processes. Trade-off is an effect at the level of fitness, i.e. concerns the fitness of the organism as a whole, and is usually inferred without knowing its molecular basis.

Trade-off, when defined as the effect of a particular process of adaptation, should be measured in terms of the net fitness of the whole genotype, rather than properties of an individual locus such as the expression of a gene or function of a particular protein. Particular phenotypic elements can be useful in determining what leads to a trade-off, but there may be many different elements that contribute to the overall fitness effect on a genotype, so the genotype-level fitness effect of trade-off

should be distinguished from various molecular factors contributing to it. Similarly, I take trade-off as defined here to be a property of a genotype rather than a 'trait' in general, unlike the usage of Fry (1996) and Kerstes & Martin (2013). A particular trait may well underlie a trade-of, but the effect will also depend on the broader genomic context.

Also, I take trade-off to primarily be a property of an individual, rather than a whole population. The mean effects of adaptation and its costs can be assessed at the population level and for practicality often must be, but the genotype is the ideal unit of analysis here. An individual genotype is a static entity that can be measured and at least in theory can be precisely replicated, whereas a particular population is more difficult to specify, shifting in various properties throughout time and as it is exposed to various conditions. Measuring the fitness of all individual genotypes in a microbial population is impractical, however this means that fitness results for whole populations, including claims of trade-off, need to be understood to be consequences of population-level effects due to complex interactions amongst individuals.

2.3 Further Concepts of Trade-off

Fry (1996) gives a starting definition of trade-off as being where a trait leading to increased fitness in one environment is deleterious for fitness in another environment, in the context of assessing fitness of phytophagous insects on various possible plant hosts. A more rigorous account of trade-off, taking into account the population as a whole, is offered in Appendix A of that paper (Fry, 1996). There, trade-off is defined as a situation where natural selection alone is unable to maximise the mean fitness of a population in two environments simultaneously. This latter definition is taken by Fry to be equivalent to the former if the definition is restricted to the inability of a single genotype to have maximal fitness in both environments. This means that the definition of 'deleterious' used implies a comparison with the optimal fitness rather than the ancestral fitness (fitness without the trait under consideration). The context, where various derived genomes (in fact, distinct species) are compared to each other, explains the definition used. The many genetic differences that have been accumulated between any two distinct species need to be taken into account in this kind of comparison – useful to remember even when comparing results across strains of *Saccharomyces cerevisiae*. Another interesting point from this study is the interest in population level effects in sexually reproducing diploid populations. Where variants of interest are heterozygous, we expect to find the occurrence of both alternate homozygotes as well; in general, frequency-dependent effects should be considered in microbial populations. This is not a focus of this research, but is worthy of future attention, in light of the advent of deep sequencing which allows precise tracking of allelic frequencies across time in

microbial population experiments.

Trade-offs are typically discussed in the context of ecological specialists and generalists, as explanation for why generalism is not more widespread. Specialisation is also referred to as 'local adaptation', on which there is a large literature. Local adaptation implies at least a relative cost – Kirkpatrick & Peischl (2012) note that "Because mutations are locally adapted, they cannot become fixed everywhere." However as has been noted, a relative cost amongst derived genotypes is not necessarily the same as a cost with reference to the baseline ancestral genotype. Trade-offs and specialisation both prevent the development of so-called "Darwinian demons" (Leiby, 2014), genotypes which are optimally fit across all environments. These conceptual opposites to trade-offs are also referred to as "super generalists", (Leiby, 2014) or "universally superior generalists" (Kassen, 2014). The term "superior generalist" has similarly been used for a situation where the generalist shows no decrease in fitness relative to the specialist (Gray & Goddard, 2012a). The fitness of a specialist after some time of adaptive evolution is often assumed to be equivalent to the optimal fitness for that genotype, but this needn't be the case, due to phenomena such as epistatic pleiotropy, discussed more below. Different genetic backgrounds can result in unexpected fitness results. Evolution to one environment can result in higher fitness in an alternate environment than achieved by 'specialists' in that environment, described as "roundabout selection" (Maclean & Bell, 2002). This is a particular instance of synclinal adaptation.

A review of local adaptation by Kawecki & Ebert (2004) limits the term local adaptation to situations where there is at least possible connection of the populations via gene-flow. This shows a particular focus on ensuring concepts have direct ecological relevance, which is an important aspect to consider when trying to clarify the terms used in this field. Applying trade-off theory tested in microbial laboratory experiments to various ecological contexts is another significant area for future work.

The concept of trade-off also has application in economics and engineering, and there are some similarities in biological systems. 'Pareto optimality' occurs when a further increase for any trait would necessarily result in a decrease in another trait. Bacterial metabolism has been determined to be close to the Pareto optimality front (Schuetz et al., 2012). In such a circumstance, the bacterial population lies on the Pareto curve mapping metabolic efficiency for different energy sources. In situations where such metabolic considerations are relevant, such as adaptation to different carbon sources in laboratory evolution, trade-offs could be illustrated as shifts along or off this Pareto optimality curve. In another application of mathematical modelling using the Pareto front concept, trade-offs between different morphological features in the shells of ammonoid fossils have been

rigorously assessed by Tendler et al. (2015). From the extremes of the phenotypic space mapped out, functional roles of the different morphological traits were hypothesized. A similar approach could conceivably be used with phenotypic data across related microbial species, to infer inherent limits to adaptation caused by trade-offs between traits. A better understanding of the phenotype space in different environmental conditions should assist with predicting responses to future environmental shifts. Perhaps trade-offs between traits can be used, for instance, to develop drug combinations which limit the development of resistance in human pathogens.

2.4 Models of Trade-off

As suggested earlier, some phenomena don't fit neatly into a dichotomy of causes underlying trade-offs in fitness across environments. Previous discussions in the literature have usually contrasted antagonistic pleiotropy, where the same locus responsible for fitness increase in one niche causes a decrease in fitness in alternate niches, with mutation accumulation, where mutations accumulated through neutral evolutionary processes are deleterious in alternate niches. These are both possible molecular bases of trade-off, but they are not exhaustive of possible causes. At least two additional possibilities can be distinguished. Firstly, trade-offs due to hitchhiking mutations, where mutations genetically linked to beneficial alleles can rise in frequency, potentially causing deleterious effects on fitness in alternate environments. Hitchhiking mutations can appear to have been selected, as they show the same population dynamics as the beneficial variation they are linked to. Remold (2012) includes trade-offs due to this process within the broader category of 'mutation accumulation', but we distinguish it here, as such variants are associated with some genomic indicators of selection rather than neutrality as in other cases of mutation accumulation. Whether hitchhiking mutations do in fact contribute to trade-offs between environments remains an open question. Researchers suggesting that they do include Kassen (2014), while those arguing otherwise include Maclean & Bell (2002), and Cooper & Lenski (2000). In addition to the phenomena of hitchhiking conditionally neutral mutations and epistatic pleiotropy, which don't fit into the idea of a simple dichotomy of causes underlying trade-offs, it is possible that some trade-offs are due to a combination of individual instances of both antagonistic pleiotropy and mutation accumulation (Bennett & Lenski, 2007).

Secondly, trade-offs could be due to epistatic pleiotropy, where the fitness effect of a variant in a gene depends on the genetic background, such as previous mutations in the gene or in other interacting genes (Remold, 2012). Epistatic pleiotropy could conceivably be the basis of trade-off involving variants which increased in frequency in a population through either neutral or selective processes. Some of the lessons from epistatic pleiotropy are that trade-off may not be due to the

adaptive variant itself, but an interaction between this variant and a particular genetic background that may be present due to either adaptive or neutral processes. This background may also only be present in a portion of the population, and/or due to the contingency of evolution may not be found with its corresponding trade-offs in subsequent replications of the adaptive process, even if the adaptive variant is found in parallel across replicates. Epistatic pleiotropy is one reason for the importance of assessing fitness compared to the ancestral genotype where possible, rather than merely comparing a strain possessing a trait (such as a single nucleotide variant or copy number increase of a gene) to one without the trait – the genetic background to that trait can differ markedly amongst strains.

Antagonistic pleiotropy has also been termed 'functional interference' (Bell, 2008), or 'sign pleiotropy' (Remold, 2012). It occurs where the same locus responsible for fitness increase in one niche causes a decrease in fitness in alternate niches. It is an instance of the common phenomenon of pleiotropy, where a single gene influences many traits in an organism. In pleiotropy which is 'antagonistic' across environments, the two 'traits' in question are the fitness of the organism in the environment of selection and in an alternate environment. Mutation accumulation has also been termed 'mutational degradation' (Bell, 2008), as it is envisaged that this kind of trade-off is due to conditionally neutral/deleterious mutations serving to degrade unused functions, with this degradation only becoming evident as a fitness cost in alternate environments. Moving away from 'mutation accumulation' is helpful, as this term has a well-established meaning as a kind of experiment for the purpose of establishing mutation rates and the distribution of fitness effects (Halligan & Keightley, 2009). However, it is conceivable that neutrally accumulated mutations could have a 'gain of function' rather than a degradative effect, so neither term is ideal. The concept of conditional neutrality is also not always used in the same way across the entirety of the evolutionary biology literature. Remold (2012) uses the term "conditionally deleterious", and conditional neutrality has been used to mean a mutation that is beneficial in one environment and neutral in another (Anderson et al., 2013; Savolainen et al., 2013). I therefore suggest the variants most relevant to trade-offs be termed 'conditionally neutral/deleterious' to indicate the respective effect in the selective and alternative environments. In another context 'conditionally neutral mutations' is a term used to mean "mutations that do not alter the fitness of the individual in which they arise, but that may alter the fitness effects of subsequent mutations" (Draghi et al. 2011), an instance of epistasis which is important for trade-offs in the case of epistatic pleiotropy discussed below.

2.5 Pleiotropy

Whatever the specific mechanism, trade-offs are a result of a kind of pleiotropy across different environments. As such, a key question relevant to the various models discussed is the relationship between pleiotropy across traits or functions expressed in a particular environment and pleiotropy across environments. Is a gene that is pleiotropic in the first sense necessarily, or even more likely to be pleiotropic across environments as well? Before this is answered, the relevance of much of the literature on pleiotropy to the genetic basis of environmental trade-offs will be unclear. How pleiotropy is inferred is an important question as well. Large scale inferences about pleiotropy in genomic networks are commonly made from data on the effect of gene knock-outs. The results have been a topic of debate; some suggest pleiotropy is highly restricted, such as a review of data from a wide range of organisms and experimental methods (Wagner & Zhang, 2011). Others, however, claim that pleiotropy is common, such as a recent study in *Drosophila melanogaster* (Blows et al., 2015). Perhaps microbial genomes are more modular than those in multicellular eukaryotes. As Wagner & Zhang (2011) have pointed out, the large increase in complexity across eukaryotes is disproportionate to the smaller increases in genome size, suggesting that the average pleiotropy of genes increases with complexity (see also Ardern & Goddard, 2016). The relevance of the data to pleiotropy across environments is also unclear. A study theoretically of direct relevance to the study of trade-offs, conducted by Qian et al. (2012), inferred on the basis of *Saccharomyces cerevisiae* gene knock-out mutants tested across multiple environments that antagonistic pleiotropy is widespread. However, the complete loss of genes is not common in adaptive laboratory evolution, so this theoretical feature of the genome may not correspond to the effect of variants seen in adaptation. The legitimacy of inferring the effects of a gene, including pleiotropy in general or antagonistic pleiotropy across environments in particular, from gene deletions is dubious. Wagner and Zhang (2011) argued that gene knockout studies give an upper limit on pleiotropy when conceived as a feature of the genome as a whole, but this is also not necessarily the case. In biological systems with redundancy incorporated into their design, absence of effect when a gene is knocked out does not imply that the gene plays no causal role at all – similarly, lack of obvious effect on a particular trait may be due to alternative processes covering up absences due to the deletion. One potentially helpful technique is the use of surveys of the effects of over-expression of particular genes alongside gene deletions. Further, modifications to a gene may have very different effects to gene deletion; testing all of the different combinations soon becomes experimentally intractable. These complications should be kept in mind before measures of the pleiotropy of a genetic network are given.

In order to understand pleiotropic effects, it will help to envisage different molecular processes

undergirding them. There are various different ways in which types of pleiotropy have been classified; here two with relevance to environmental trade-offs are discussed. Firstly a differentiation can be made between type I pleiotropy, where a single gene product serves in multiple functional roles at the molecular level, and type II, where a single molecular function has multiple morphological and physiological consequences (Wagner & Zhang, 2011). An example of the latter could be the action of the same signalling pathway or transcription factor having many different cellular roles and variants having different phenotypic consequences. As stated, the distinction between a molecular function and a physiological consequence is not precise, and the concept will need to be adapted into an analogous distinction for types of inter-environmental pleiotropy. In the first type, cross-environment pleiotropy of gene function may result from the same gene product playing different functional roles in different environmental conditions. For example, differences in expression of the gene or other related genes may change what roles the gene product plays under different conditions. In the second case, the functional role of the gene product (e.g. role as a transporter protein) even though the same across environments, may make a different contribution to fitness in different environments. Both cases could contribute to trade-offs, if variants change a gene's expression or the gene product's functioning under different environmental conditions. Another important distinction is given by Remold (2012). Following the nomenclature used by Weinreich et al. (2005) for epistasis, a distinction between magnitude pleiotropy and sign pleiotropy has been suggested to be useful in assessing the basis of specialisation and trade-off. Sign pleiotropy is a change in the 'sign' (positive or negative) of the fitness effect of a gene or variant across environments. When the effect relative to the ancestor is positive in the selective environment and negative in the alternate environment, this is antagonistic pleiotropy. Magnitude pleiotropy is where the sign does not change but the size of the effect does. Conditionally neutral ("conditionally deleterious") mutations are included within this category (Remold, 2012). Magnitude pleiotropy can underlie specialisation (local adaptation), or in the special case of conditionally neutral mutations can underlie trade-off. Considering the genetic bases of trade-off and specialisation as being positioned on a spectrum of pleiotropic effects across environments is a potential method of classification.

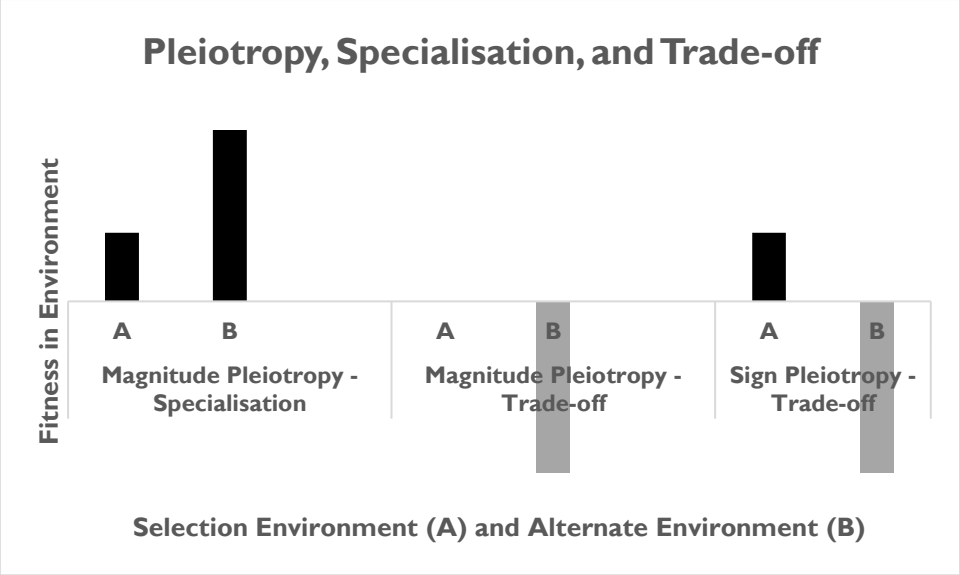


Figure 2.1: The effects of variants with different types of pleiotropy. Magnitude pleiotropy can underlie specialisation, or if the variant is conditionally neutral/deleterious or nearly neutral can underlie trade-off. Sign pleiotropy can underlie trade-off, as antagonistic pleiotropy.

2.6 Genomic Factors Underlying Trade-offs

In recent years, direct access to the genomic correlates of changes in fitness during experimental evolution has become available. These variants accrued during adaptive processes range in size from single nucleotide polymorphisms to large scale changes in cellular karyotype. Single nucleotide variants are perhaps the mutations occurring in experimental studies which are the most straightforward to catalogue. If occurring in a coding region of the genome these can categorised as synonymous, missense, or nonsense. Synonymous mutations rarely have a large effect on fitness, although they can have a fitness impact for instance through RNA folding; deleterious effects are discussed for viruses by Cuevas et al. (2012), and two beneficial synonymous mutations in the bacteria *Pseudomonas fluorescens* are discussed by Bailey et al. (2014). Regarding possible contributions to trade-offs, it is plausible that some of these mutations are conditionally neutral; for instance, in the plant *Arabidopsis thaliana* environmental relevance was found to be explanatory of a small proportion of variation in synonymous polymorphisms (Lee & Mitchell-Olds, 2012). Missense mutations are also typically not major contributors to adaptation; they frequently involve swaps between chemically similar amino acids and in such cases can be expected to have a modifying effect on protein function. Nonsense mutations in a gene can result in functionally important early truncation of the protein product. This is an example of 'loss of function' mutations, which are common in experimental evolution (Behe, 2010; Lazar et al., 2014). Similarly, Indels (insertions and deletions) can result in loss of function effects, or other changes to protein products. In contrast,

variants in the copy number of genes have been shown to contribute to large changes in fitness, for instance in *Saccharomyces cerevisiae*, transporter genes such as hxt6 are commonly increased in copy number in nutrient limited conditions (Payen et al., 2014) and in *Escherichia coli* a duplication in the citrate transporter citT was involved in the evolution of citrate metabolism in Richard Lenski's long term evolution experiment (Blount et al., 2012). Larger chromosomal changes also have fitness effects, and can result in trade-offs due to antagonistic pleiotropy (e.g. Avelar et al., 2013). The role of regulatory evolution has been less closely investigated than coding region changes, and is another important potential source of insight.

The effects of variants when taken together are not always additive in a simple fashion; there are epistatic effects, where the effect of new variants is dependent upon the genomic background. Experimental evolution has begun to show the prevalence of this phenomenon, and these genetic interactions have been suggested to be particularly common within genes and between genes within the same biochemical pathway. (Bank et al., 2014). If trade-offs are an instance of GxE interaction, then epistatic effects can be considered as GxGxE interactions – demonstrated in *E. coli* by de Vos et al. (2013) and Flynn et al. (2013); higher level interactions are also conceivable, but difficult to test. 'Diminishing returns' epistasis, where additional mutations have less effect in evolving populations than expected from the effect in a 'clean' ancestral genomic background, is common in long term experiments. It has been suggested that while epistasis within genes (intragenic epistasis) is comparatively unpredictable, epistasis between genes more consistently follows the diminishing returns pattern (Chou et al., 2011). A different use of the term 'GxGxE interaction' is when there are interactions between parasite and host populations, another important area where trade-offs are relevant, discussed by Méthot & Alizon (2014).

2.7 Measuring Fitness

Another important question is the experimental tractability of discerning the relative contributions of different concepts. Concepts should be defined such that their referents are empirically distinguishable. So, broadly, what kind of evidence will establish the existence of a trade-off? One common methodology makes use of reciprocal transplant experiments, where two populations adapted to different environments are tested for fitness by a short time of population growth in the alternative environment (Savolainen et al. 2013). On one definition, a trade-off can be inferred when both populations have higher relative fitness (competed with the other population) in their native environment than in the alternate environment. (Hereford, 2009). However, this kind of evidence does not distinguish between differential adaptation and trade-off in the stricter sense where there is

a fitness loss in the alternate environment when compared to the ancestral performance. Where possible, a more informative approach compares fitness in an alternate environment to that of the ancestral population. This is possible for instance in experimental microbial populations where a representative sample of the ancestral population has been stored as frozen stocks.

There is also not full consensus on the proper way to define and measure fitness. Different components of fitness can be measured, including viability and fecundity; alternatively a demographic factor such as population growth rate can be used as a summary fitness measure (Hereford, 2009). For microbial populations, this growth rate approach, a holistic measure of fitness, is more clearly the appropriate measure than in larger organisms where measuring population growth rates is less feasible due to long generation times. The sense of fitness of interest here is relative fitness as compared to an ancestor through competition experiments.

Relative fitness, w, is equal to the selection coefficient, s, plus 1; $w = 1 + s$ (Melnyk et al., 2014). However, there are a few different ways in which relative fitness and selection coefficients have been calculated; some of them are equivalent, while others are not. I explain the main options here, but do not give an in depth treatment – this is an issue that would benefit from more widespread agreement amongst key research groups. Let 'M' represent the 'mutant' evolved population, and 'A' the ancestral population, and N_i either the population size or proportions of competitor 'i' (so long as this choice is kept consistent within the equation). T_i is the time point I, for instance where i=0 at the start of the experiment and 1 at the end.

1)
The ratio of the growth rates of the two strains. (e.g. Dean & Dykhuizen, 2009)
$w = u_M/u_A$
This is often termed the "Darwinian fitness" (Hartl & Clark, 1997). It comes from a simplified discrete model of population growth where generations do not overlap.

If growth rates are derived from a continuous rather than discrete time model of population growth, then the Malthusian parameters of the strains are of relevance.
2)
$r_{MA} = m_M / m_A$, where $m_i = \ln[N_i(T_1)/N_i(T_0)]/T_1$
(Then subtract 1 to gain a selection coefficient rather than relative fitness.)

This measure is equivalent to the ratio of the number of doublings of the two strains. (Lenski et al. 1991) and is the typical measure of relative fitness in microbial selection experiments. However, it is

persuasively argued in an important article by Chevin (2010) that it has no clear evolutionary meaning.

3)
A direct calculation of the selection coefficient "s" is from the difference of the Malthusian parameters. (Hansen et al., 2007)
$r_{MA} = m_M - m_A$, where $m_i = \ln[N_i(T_1)/N_i(T_0)]/T_1$

This is equivalent to the slope of $\ln(N_M/N_A)$ over time, as found in Bohannan & Lenski (2000), and equivalent equations in Friedenberg (2003), and Dykhuizen and Hartl (1983). This measure has also been termed the 'selection rate constant'. An important advantage claimed for this measure is that it is less affected by sampling error than a ratio (Travisano & Lenski, 1996). Chevin (2010) suggests multiplying it by the generation time, in order to get a unitless measure comparable across studies with different generation times.

4)
Difference in numbers of doublings of two competitors.
Cowen et al. (2001) incorrectly equates this with the difference between the Malthusian parameters, citing Travisano & Lenski (1996). The usefulness of this 'difference in doublings' measure is dubious. However, an important point is that in comparisons of competitions across treatment conditions the differences for each competition can be standardized by the total number of doublings across all assays.

Melnyk et al (2014) give two ways in which the selection coefficient has been measured in competition experiments which are equivalent to equation 2 (after subtracting 1 to get a selection coefficient), and equation 3 above. They note that "Both estimates of fitness are widespread in the literature, and we see no principled reason to prefer one to the other". However, I take the arguments of Chevin to be persuasive that the difference in malthusian parameters (equation 3) is the more appropriate measure. Melnyk et al (2014) though, provide a helpful conversion factor, whereby the results of equation 2 can be approximated to those of equation 3 after subtracting 1 and dividing by 1.7, for experiments of 4 or more generations length.

As Chevin (2010) notes based on a number of arguments, "[o]verall, the available data make it difficult to compare selection coefficients across species and studies" – and the same goes for relative fitnesses. The use in experimental evolution of model organisms such as *E. coli* and the use of

the dimensionless ratio of Malthusian parameters as the typical measure of relative fitness in these studies allows for some comparison across similar studies, but cross-study comparisons should be treated with care. The inclusion of generation times and growth rates of the ancestral population would greatly assist cross-species comparisons (Chevin, 2010).

2.8 Future Work on Trade-offs

There is much more work that can be done on the molecular bases of trade-offs, in both microbes and multicellular organisms. In particular, the integration of high throughput phenotypic assays with high throughput genomic data collection may reveal new insights of the relative contributions of factors at the level of the gene and genome, as well as any epigenetic factors. Here I briefly review five areas of technology and some important questions that they show promise of being able to answer if applied in combination with microbial experimental evolution. Next generation sequencing, gene knockout libraries, transcriptomics, high throughput phenotypic assays, and surveys of metabolites can potentially be used together contributing to a much richer understanding of the contours of the genotype-phenotype map during adaptation and other evolutionary processes.

The first two technologies concern our ability to explore or exploit different genotypes in experimental microbial populations. Next generation sequencing is now approximately 15 years old, and in the last 5 years in particular has become an accessible tool for large scale microbial sequencing projects. A number of different platforms are clustered together under this term; the dominant is the Illumina short read sequencing platform. The application of NGS to experimental evolution has rapidly gained attention, with a number of recent reviews of the developing field. (e.g. Dettman et al., 2012). Next Generation Sequencing can be used to determine the relative contributions of SNPs vs larger rearrangements under different conditions, and differences between organisms such as bacteria and fungi, as well as multicellular eukaryotes. Methods involving manipulation of the genomes of model organisms can be paired with the use of sequencing technology. For instance, gene knockout and overexpression libraries are key tools in determining the functional effect of genetic variation, by recreating variants of interest and testing their effects on fitness and/or other phenotypic measures. As an example of the kind of research that is now possible, a recent study in *Saccharomyces cerevisiae* used overexpression collections in combination with deletion libraries to establish the fitness effects of many genetic changes and compare them to those observed in other experiments (Payen et al., 2015). Future research using these genetic tools of sequencing and genetic manipulation will be able to further probe important aspects of the genome such as the complex web of epistatic interactions between genes, as well as epistatic

interactions between mutations within single genes.

Transcriptomics is conceptually ideally positioned for exploration of the border between genotype and phenotype. DNA, particularly stable as it is double-stranded and not liable to base-catalysed hydrolysis, is more easily extracted for sequencing than RNA. Following this initial extraction step, the RNA sequences are converted to complementary DNA with the use of a reverse transcriptase enzyme, and sequence libraries are prepared – this often involves filtering the relevant sequences down to only include messenger RNAs, that will go on to be translated into proteins. After these preparatory steps, RNA sequencing is able to be conducted similarly to DNA sequencing. It is my supposition that gene expression levels measured through transcriptomics would change during the course of adaptive evolution faster than gene sequences, but to my knowledge this question of evolutionary rate has not been investigated. The interpretation of results from technologies such as RNA-Seq can be difficult. Demonstrating correlations between genetic variants and transcriptional changes is not straightforward; as Sandberg et al. (2014) note in their experimental evolution study in *E. coli* which incorporated RNA sequencing that "Isolating the transcriptional effects of any individual mutation is clearly complicated by the presence of other genetic changes within the evolved strains."

Another set of technologies can assess organismal phenotype in depth. High throughput phenotype assays such as Biolog™ plates can test functional effects of genetic variation in microbes across many different environmental conditions. Large scale replication across multiple environments, using precisely controlled growth conditions and isogenic strains will allow the discovery of the contribution of the environment to different kinds of polymorphism & fixation events. For example, a study in the plant *Arabidopsis thaliana* showed that the environment had an influence even on the proportion of synonymous polymorphisms in the population (Lee & Mitchell-Olds, 2012). High throughput studies across multiple populations and environments, when combined with sequencing technologies, are able to discover more comprehensive distributions of fitness effects (DFEs). That is, it is now possible to determine which mutations are beneficial or deleterious under which conditions. Another developing field is metabolomics, the study of metabolites in cells. Along with the already-discussed transcriptomics, along with proteomics, and fluxomics, which respectively study the patterns of gene expression, protein interactions, and biochemical fluxes, these technologies allow exploration of the 'black box' processes connecting genotype and phenotype. Studying some of these biochemical events in microbes helps in determining the relative contributions of plasticity, genomic evolution, and other epigenetic factors to organisms' ability to adapt to new environments. One area where these technologies may produce crucial insight is research into the evolution of antibiotic resistance. Recent papers have begun to analyse the

relationship between adaptation to different antibiotics, termed the cross-sensitivity and cross-resistance network, with the aid of sequencing technologies (e.g. Lázár et al., 2014, Munck et al., 2014). The effects in alternate environments of variants conferring resistance are commonly deleterious, i.e. there are costs of resistance. As an example, one important finding relating genetic changes to the cellular mechanics underlying differences in resistance has been that resistance to aminoglycoside antibiotics is associated with a reduction in proton-motive force across the bacterial cell membrane, decreasing the activity of efflux pumps dependent on this, and therefore increasing sensitivity to other antibiotics which could otherwise be removed by these pumps. This example is discussed along with other potential molecular bases of cross-sensitivity by Pal et al. (2015)

2.9 Conclusion

Trade-off is an important concept with applications ranging from the evolution of drug resistance to cancer cell proliferation to plant ecology. Increasingly the molecular bases of this phenomenon are accessible to evolutionary biologists, in particular through rigorous experimentation with microbial evolution over hundreds or thousands of generations in the laboratory. The traditional dichotomy between 'antagonistic pleiotropy' and 'mutation accumulation' as a basis for trade-off is giving way to more nuanced concepts of distributions of fitness effects across environments, and complex interactions between variants in genomes adapted to particular environments. The molecular events underlying the phenotype-level processes can be investigated with new technologies in genomics, transcriptomics, metabolomics, and other molecular biological fields. New experimental techniques allow for more ambitious experimental designs incorporating multiple lines of evidence, contributing to the central question of the intricate connections between evolving organisms and evolving genomes.

Chapter 3
Genomic Basis of Adaptation to Nutrient-Limited Environments in Sexual and Asexual Populations of *Saccharomyces cerevisiae*

"There is nothing like looking, if you want to find something. You certainly usually find something, if you look, but it is not always quite the something you were after."
– Gandalf

The genetic basis of adaptation of microbial populations to complex environments is now accessible with next generation sequencing technology. In this study we report genomic correlates of fitness in sexual and asexual populations of *Saccharomyces cerevisiae* derived through 300 generations of experimental evolution from nearly isogenic ancestors. It is suggested that in particular, SNPs and copy number variants are contributors to adaptation in these populations.

21 whole populations derived from experimental evolution in nutrient limited chemostats were sequenced to an average depth of >70x, and three others were able to be sequenced at a lower depth, as well as the genome of the starting strains ("ancestor"). This is one of the first evolve-and-resequence studies reported for a diploid organism, and the second after Sunshine et al. (2015) published experimental evolution study for a diploid organism that we are aware of utilising sequencing of DNA samples from a mixed population. It is to our knowledge the first to report copy number variants (CNVs) based on whole population sequencing data and the first to analyse the differences between diploid sexual and asexual populations sourced from an isogenic common ancestor. Only very recently have genetic differences in the processes of evolution which are due to sexuality been investigated. Data from a recent study by McDonald et al (2016) are used for comparison.

3.1 Resequencing Studies and the Basis of Adaptation

Experimental evolution using microbial populations has been performed at least since the early 1880s, and particularly over the last 40 years. There have been two main techniques used for microbial evolution in recent decades; batch culture, where regular transfers into new media allow for population expansion, and the chemostat, which facilitates continuous exponential growth. The chemostat, originally called a 'Bactogen' was popularised by Jacques Monod (Adams & Rosenzweig, 2014), and has the advantages of an approximately constant population size in the culture flask, avoiding bottle necks, and a constant selective pressure (although the effect of this pressure will change as the population adapts). Changes in fitness are measured by competition between evolved clones and a sample from the ancestral population; relative growth rates constitute a measure of fitness.

The variants responsible for the phenotypic changes underlying shifts in fitness are accessible through high throughput genome sequencing technologies. Until recently, whole genome sequencing was prohibitively expensive, but is now achievable, particularly for microbes. Typically, populations have been grown from individual clones and these homogenous samples sequenced – sometimes, a few are taken from an evolved population to get a sense of intra-population variation. Alternatively, it is possible to pool DNA from an evolved population, and sequence this heterogenous mix of DNA, containing many alleles at various frequencies. The main advantage of single-clone sequencing

over whole-population sequencing is the possibility of ascertaining full genotypes, potentially including linkage between variants which are not fixed in the population. The advantage of sequencing whole populations is the potential for estimating population-wide allele frequencies and, with high enough depth of sequencing, this approach allows access to low frequency variants.

Discovering variants which were actually adaptive is not entirely straightforward. Many of the variants which show evidence of having been acted on by natural selection will have been adaptive. However, in microbial genomes with minimal recombination, some variants will increase in frequency in the population due to being linked to adaptive variants, and much adaptation will be to the general culture conditions rather than particularly in response to the environmental stressor focussed on. Variants which are derived across multiple populations exposed to the same treatment, however, can be assumed to be adaptive. Apart from some key examples, we are interested in this study not in specific adaptations, but in the numbers of different kinds of variants which have risen in frequency across populations exposed to different treatments; in other words, in the influence of environment and sexual status on the general mechanics of adaptation.

3.2 A Complex System

The study by Gray and Goddard analysed here involved a range of factors which are relatively complex compared to standard adaptive laboratory evolution experiments with microbes. As comparison, compare the long term evolution experiment (LTEE) from Lenski et al, that has produced a wealth of data on the mechanism of evolution in bacteria. The LTEE was in haploid asexual bacteria; the Goddard & Gray study was in distinct populations of a diploid sexual and diploid asexual eukaryote. The LTEE involved only one relatively benign static batch culture environment – in comparison, the Goddard & Gray study was conducted over two different continuous flow culture media which each imposed a strong selective pressure on the yeast populations; gene-flow between the environments was also added to the system for some replicates, as a further variable for comparison. The environments in the G/G study are complex, involving two stress conditions each, and the traits assessed are also complex, for instance response to salt stress is known to be polygenic. The genomic analysis conducted in this study makes use of whole population sequencing rather than sequencing of individual clones, and is the first to demonstrate the existence of gene copy number variants by this means.

3.3 The Adaptive Benefits of Sex

Sexual reproduction is prevalent across eukaryotic lineages, and given that it is a complex strategy with a number of associated costs (Lehtonen et al., 2012), it is presumed to be associated with some fitness advantage. Sex, on our definition, involves recombination, random assortment, and syngamy (Goddard, 2007). These terms refer to meiotic recombination occurring between non-sister chromatids, random assortment of chromosomes, and the event of 'fertilisation', i.e. the fusing of gametes. Here we add the requirement of outcrossing, to exclude self-fertilisation, which is common in some systems. The costs of sex are contributed to by many factors, and will differ between different systems (Lehtonen et al. 2012).

Hypothetical benefits of sexual over asexual reproduction have been proposed. These include that sex functions to increase variation for natural selection to act on, that it decreases deleterious variation, or that it is associated with a quite different benefit such as enhanced DNA repair. A fitness benefit for sexual populations in adapting to harsh environments has been directly shown in different systems, such as in yeast Goddard & Gray (2012a), and in rotifers Becks & Agrawal (2012). Findings about benefits of sexual reproduction are not necessarily generally applicable however, as different organisms differ in reproductive mechanisms, and there is not a simple dichotomy between

sexual and asexual populations, as discussed in Chapter 7.

In the system under investigation, some potential costs of sexual reproduction are reduced. In single-celled eukaryotes with isogamous gametes and high population densities, there is minimal energetic cost to mate finding. More generally, the use of 'mitotic fitness', i.e. clonal growth rates, at discrete time-points as a measure of evolutionary fitness eliminates consideration of costs involved in the process of meiosis. A change in the rate of mitotic division may not be definitive in determining a population's fitness change.

3.4 Results

Single Nucleotide Polymorphisms/Variants - SNPs

I initially hypothesised that the number of SNPs would differ significantly between sexual and asexual populations, with sexual populations having more SNPs, due to having adapted to a greater extent. The actual story is more complex. To begin with, I report results from 10 zero gene-flow populations, as these had better quality whole genome data available than some of the full gene-flow populations sequenced.

After filtering to remove SNPs judged likely to be ancestral rather than derived based on either being called as present in the ancestral genome or being present across nearly all replicate populations (see Methods), there is a strong, although not statistically significant, tendency for sexual populations to have fewer SNPs than asexual populations. For higher frequency SNPs, however, there is a clear difference between sexual and asexual populations, with sexual populations having fewer SNPs. This was tested with a two-tailed t test and is true both for SNPs present in 50% or more reads at a site, ($p=0.0048$) and SNPs present in 70% or more reads at a site ($p=0.0094$).

The carbon-limited populations, which increased more in fitness as measured by Jeremy Gray (Gray, 2011), also have fewer SNPs than N-limited populations, but the difference is not quite statistically significant ($p=0.054568$, two-tailed t-test).

The rest of the analyses for SNPs and Indels focusses on zero gene flow populations. Future work coming out of this project will include data from the full gene-flow populations.

Coding & Non-coding SNPs

Sexual populations, as in the recent study by McDonald et al. (2016) have proportionally fewer non-coding SNPs, and more coding SNPs. Sexual 'zero gene flow' populations had on average 17% of their SNPs non-coding, compared to 18.6% non-coding for asexual populations. This difference was consistent, in that all sexual populations had a smaller proportion of non-coding SNPs than any asexual population, but was not a statistically significant difference when assessed with a two-tailed t-test ($p=0.056911$).

Parallel SNPs

When counting SNPs is restricted to SNPs found in a few populations (3-5) in parallel, there are more in asexual than sexual populations. However, when assessing SNPs found in many zero gene-

flow populations (7 or more – initially discarded as likely being ancestral), there is a tendency for sexual populations to have more SNPs.

SNPs and Fitness

For the zero gene-flow N-limited populations there is a positive relationship between variants found in parallel across populations (in 6 populations out of the 10 zero gene-flow populations with adequate data) and fitness in the N-limited environment. For this positive relationship the correlation coefficient R^2=0.5954. For all populations taken together, there is a much weaker relationship, but a positive relationship relationship of R^2=0.4932 if instead the SNPs found in 3-5 populations are considered. There appears to be a contribution of parallel SNPs to fitness, but this cannot be stated definitively and is confounded by many factors such as other variants affecting fitness, difficulties in removing all ancestral SNPs from the analysis, genetic hitchhiking, and insufficient data.

Indels

There is no statistically significant difference in the numbers of indels observed in different samples when 'full gene flow' samples are grouped by either sexual status or environment.

Parallel Indels

A number of genes, approximately 10, were initially found to have indels across many zero gene flow populations, and frequently at the same site, however with a more conservative filtering of putative indels, removing sites which in the ancestor had indels nearby (within 2 nucleotides) or SNPs within 1 nucleotide distance, few sites were observed to have indels in parallel. This will be an area for further investigation.

To check that the pipeline prepared does work as intended and that putatively parallel indels were not present in the ancestral reads or due to an abnormality in read depth or alignment at the site at which they were called by GATK, the BAM files for the ancestors were individually checked around a few indel sites in the evolved populations.

FLO9 was of particular interest, with an indel called in all populations, however on closer assessment of the ancestral BAM alignment file, the region around the putative derived indel was found to contain an indel and SNPs relative to the reference, so this was discarded.

Coding & Non-coding

As would be expected if indels occurring in genes are more likely to be detrimental than those outside of genes, sites outside of annotated genes were heavily over-represented for indels:

The total genome length for S. cerevisiae Y55 is 11700636, and total genes sequence length is 8893415; 76% of the genome codes for proteins. In contrast, the proportion of total indels found in genes was much lower, ranging from 17% to 30% in different populations. This is a result that is consistent across treatment groups; there is no statistically significant difference in the proportion of indels in genes when 'full gene flow' samples are grouped by either sexual status or environment.

Copy Number Variants

All samples carry some sites at high copy number (at least twice the average read depth ratio to the ancestral genome) compared to the ancestral genome, however when this data was filtered to genes with at least 70% of the whole gene present at this high depth, only three genes were found.

Out of 18 population genomes with adequate depth for this whole genome analysis:

YDR342C (HXT 6) was found at very high depth in four samples evolved in a C-limited environment (two with zero gene-flow, and two with full gene-flow).
YNR075C-A was at very high depth in four samples, across both environments.
YDR354C-A was at very high depth in one C-limited population.

HXT 6 is a high-affinity glucose transporter, and will receive further attention in this study. YNR075C-A is a 30 amino acid long putative open reading frame, of unknown function and cellular location. YDR354C-A is, similarly, a short dubious open reading frame (putatively 47 amino acids long). Such non-standard genes should receive further research attention.

CNVs by Environmental Treatment, and Gene-flow

With the '1.5x ratio, 70% present' criteria, the only gene that is found at high copy in multiple replicate C-limited populations and no N-limited populations is YDR342C, i.e. the hexose transporter HXT6; all other 'parallel' genes are found in the other environment. Likewise, there are no genes found to be at high copy in parallel in the N-limited populations but not at least one C-limited populations – the only exception is the 29 amino acid long 'dubious open reading frame' YER090C.

All genes except one that are found copied in parallel across replicates in either zero gene-flow

environment are also found in the full gene-flow conditions. The exception is again the dubious open reading frame YER090C. Perhaps it has some role in the osmotic N conditions. The only gene found in parallel (2 out of 5 populations) in the N-limited full gene-flow populations that is not found in any of the C-limited populations is YHR189W, a peptidyl-tRNA hydrolase required for respiratory growth on minimal medium – however, the case for this being particularly advantageous in N-limited conditions is not strong given that it doesn't occur as a CNV in any of the zero gene-flow N-limited populations either. In summary, assessing the presence or absence of genes in this list doesn't show clear evidence of antagonistic pleiotropy due to genes that are copied in parallel across replicates, that have been shown to correlate with fitness. The 'most adaptive' variants able to be discovered at the level of copy number variants are not a cause of antagonistic pleiotropy.

Specific Gene Examples

The HXT6 (YDR342C) sites are found to have increased read depth in two out of four zero gene-flow C-lim populations with adequate data and two out of nine full gene-flow populations (one out of the five pairs of full gene-flow populations with at least one representative). The increase in read depth in the derived populations is less in the full gene-flow pair exhibiting increase than in either of the two zero gene-flow populations. On first site this could be taken as evidence for antagonistic pleiotropy, i.e. selection against expansion in HXT6 copy number in the N-lim environment, however due to the periodic 50:50 mixing in the full gene-flow populations if the duplication was deleterious in N-lim conditions an expansion of a maximum of approximately 50% of the typical zero gene-flow expansion would be expected to be seen in full gene-flow populations, and less in the populations most recently exposed to the N-limited environments. Instead, the full gene-flow expansion when we can confidently observe it is equivalent to at least 2/3 of the large expansions observed in the zero gene-flow populations.

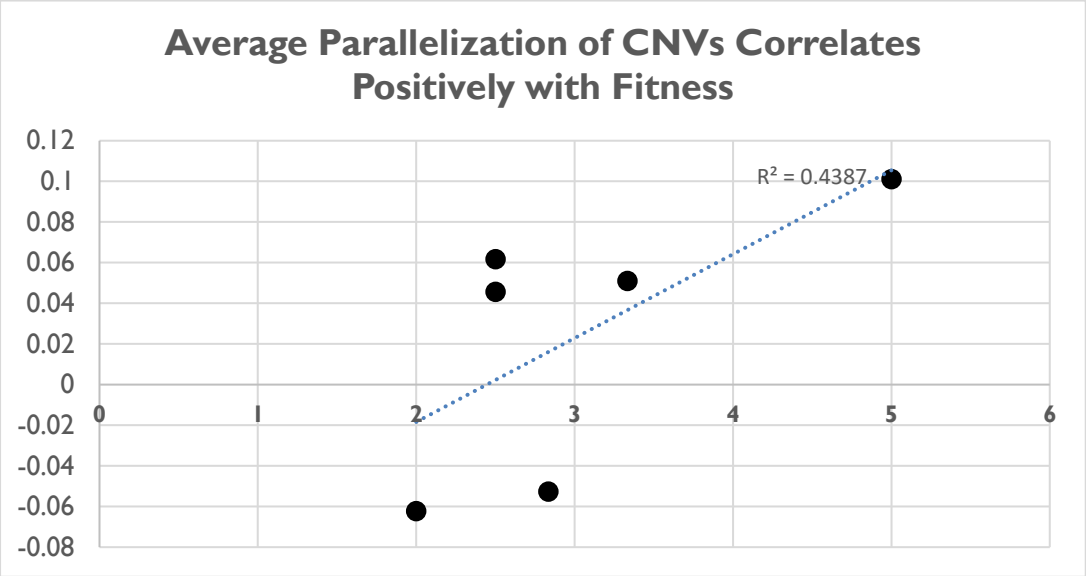
In addition to this, the difference in average read depth between the affected pair of full gene-flow populations (those evolved in the C and N-lim environments, with mixing from the paired population every 12 generations) is essentially indistinguishable. This suggests that the HXT copy number increase is not highly detrimental for fitness in the N-lim environment, otherwise, if clones with lower read depth were selectively favoured, a decrease in the average population copy number would be expected, and the N-lim full geneflow population should have noticeably less read depth at these sites. In conclusion, HXT6 is not a major contributor to any antagonistic pleiotropy of C-lim derived populations in the N-lim environment.

Measuring Parallelization of CNVs

Counting the number of times that a gene occurs in replicates within a treatment provides a score

of 'parallelisation' for a gene. It is hypothesised that the CNV-genes occurring in parallel are most likely to be adaptive, and hence those populations which have parallel CNV-affected genes are likely to be have higher fitness.

The copy number variants in zero gene-flow populations were therefore assessed, and it was determined that in both the hot C and osmotic N environments there is a positive correlation between the 'average parallelisation' of copy number variants and fitness. Populations with the very highest fitness contained few CNVs, with copied genes in these populations being frequently found across other replicates.



Future research should take into account such possible genes, as well as other intergenic regions, which were also overrepresented in high copy number sites – whether such sequences might have a function relating to gene regulation that could be a target of selection in adaptive laboratory evolution should be investigated in future.

Variants in Sexual versus Asexual Populations

On further reflection, there is little reason based on the fitness data obtained by Gray and Goddard to expect two distinct groupings in the data on genomic variants which correspond to sexual and asexual populations, when the entire data set is taken together. Firstly, fitness is always relative to an environment and it is plausible that the environment has a greater effect on molecular evolution than sexual status does.

If we take fitness as changes as transitive across environment for illustration's sake, then it is

observed that environment was a greater contributor to fitness differences between populations than sexual status was (**Figure 3.1**). Carbon sexual and asexual populations are indistinguishable (P=0.7603), while carbon and nitrogen populations are distinct, as their ranges do not overlap. While sexual status makes a larger difference within the N-lim environment, the ranges seen within sexual and asexual population fitnesses do just overlap.

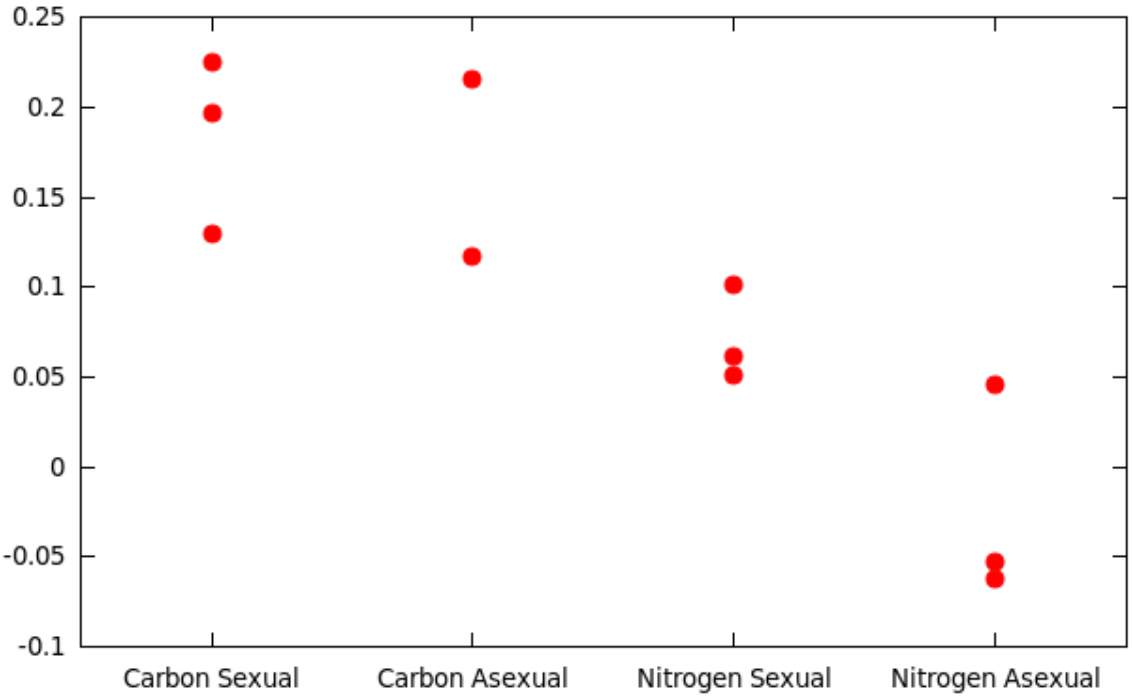


Figure 3.1 Data from Gray (2011) Fitness of Zero gene-flow populations, arranged by treatment group. Sexual status was particularly important for fitness in the Nitrogen-limited, high salt environment. It is possible that the genetic basis for adaptive traits here was more complex than in the C-limited environment; it is known than adaptation to salt stress has a multifaceted genetic basis. Also, removing deleterious mutations may be of particular importance in an environment with few beneficial variants available.

At the level of fitness, statistically significant differences were not entirely as clear, as initially assumed when hypothesising that there would be obvious differences in the molecular evolution of sexual and asexual populations. There is overall a statistically significant difference between sexual and asexual treatments (P=0.027, two-tailed T-test). However within all populations exposed to the C-limited environment (all full gene-flow populations and C-limited zero gene-flow populations), the difference in fitness measured in these populations' selective environment is not statistically significant. The greater difference in the N-limited than the C-limited treatment due to sexual status is evident even in the full-gene flow populations, with the set exposed to the N-lim environment

having a statistically significant difference in the average fitness in the selective environment (Two-tailed t test, p= 0.0377) while the difference between the sexual and asexual full gene-flow sets exposed to the C-lim environment was not statistically significant (p=0.0591).

This data has bearing on the question of whether the primary adaptive consequences of sex in this system is in deleterious mutation clearance or in facilitating adaptive mutations. It is plausible that there are fewer beneficial mutations possible in the N-limited populations. The relative lack of adaptation in the N-limited population suggests that the populations began relatively closer to a fitness peak in this environment, compared to in the C-limited environment. It has been suggested that this closeness to the fitness peak is due to the highly osmotic conditions in grapes, where *S. cerevisiae* are found (Goddard & Gray, 2012a), however the problem of finding an environmental niche for *S. cerevisiae* remains unsolved (Goddard & Greig, 2015). The relative closeness to a fitness peak suggests that fewer beneficial mutations will be available in the N-limited compared to the C-limited environment, and as such a greater proportion of single step mutations will be deleterious or neutral. Sex was most advantageous in conditions where few beneficial mutations were available. This could be either because the primary advantage of sex is in removing deleterious mutations, or because there were so many beneficial mutations available in C-limited conditions that sex was not required in order to accumulate multiple adaptive variants in each population.

The data is limited, but it is tentatively suggested that evolving in a C-limited environment has a tendency towards a negative effect on fitness in N-lim for sexual, but not asexual populations. Evolving in an N-limited environment has no corresponding negative effect in the C-lim environment.

3.5 Comparisons to and analysis of McDonald et al. (2016)

A study was published in *Nature* earlier this year (McDonald et al., 2016) that bears some similarities to this project. Sexual and asexual populations of *Saccharomyces cerevisiae* sourced from a common ancestor were evolved in multiple independent lines through 1000 generations, and a few populations were subsequently sequenced. There are many similarities with our study, however, there are important differences as well. Perhaps crucially, the asexual strains used by McDonald et al were haploid, and here evidence is presented that this affected molecular evolution in the populations in line with expectations based on the literature.

In addition to this, McDonald et al. conducted laboratory evolution in batch culture (specifically microwell plates), compared to Goddard & Gray's use of chemostats. Copy number variants were

not assessed in the mixed population DNA samples from McDonald et al., however 8 clones were separately sequenced, and copy number variants investigated with a sophisticated mathematical approach. Sequence data for the McDonald et al. populations was obtained at lower coverage (approximately 40x) than the results for the Goddard populations (approximately 70x coverage), however multiple time-points were sequenced by McDonald et al., allowing removal of more likely sequencing errors through analysis of correlations between time points.

Haploid:

It is known that, at least under certain conditions, diploid populations adapt more rapidly than haploid populations (Paquin & Adams, 1983). It is also at least suspected that haploid populations accumulate more 'loss of function' mutations than diploid populations e.g. Raynes & Sniegowski (2014), although I am not aware of this having been tested directly in an isogenic background. The numbers, proportions and frequencies of SNPs that occur as LOF mutations in the data from McDonald et al. were tested, and reported here and compared to those for the Goddard-Gray populations.

The total numbers of SNPs observed in the McDonald et al data were considerably higher in sexual than asexual populations, as reported in McDonald et al. (2016). The number of LOF mutations compared to the common reference (W303 genome from SGD, Saccharomyces Genome Database) was also statistically significantly higher in the sexually reproducing populations, unsurprising given the large significant difference in numbers of SNPs. The proportion of SNPs that coded for a stop codon however was higher in asexual than sexual populations, despite the proportions of synonymous, nonsynonymous and intergenic mutations being similar in these groups (McDonald et al., 2016). The difference in total number of stop codons encoded by a SNP was statistically significant ($p = 0.025574$, two-sided t-test), however the difference in SNPs coding for a stop codon at a site that was not previously a stop codon was not quite significant with a two-tailed t-test ($p=0.067423$); this is significant for a one-tailed t-test testing the specific hypothesis that asexual populations have more LOF mutations ($p=0.033711$).

In the Goddard-Gray populations, there was no difference between sexual and asexual populations in the number of SNPs coding for a stop codon, according to a two-tailed t-test ($p=0.893065$). There is more work to do in future in further exploring such data sets, for instance investigating sites mutated in parallel across replicates. In McDonald et al's analysis, such sites were discarded, with what I take to be an overly conservative approach that is likely to miss key adaptive mutations, on the assumption that parallel evolution is very rare: "for each candidate mutation, we estimated the site-specific error rate by calculating the frequency of the alternate allele outside of the

population in which the mutation was called. We then excluded candidates with an estimated error rate above 0.05." (McDonald et al., 2016). Particularly interesting to compare may be those sexual and asexual populations including clones with near-identical fitness (McDonald et al., 2016 – Figure 1a).

Batch culture:

The use of batch culture, i.e. serial transfer, affects population dynamics, and may have a different effect on sexual and asexual populations than observed in chemostats. For instance, perhaps the relatively rapid adaptation of sexual populations, which do not need to overcome clonal interference to the same extent, is magnified under serial transfer conditions, where small population sizes could particularly hinder asexual populations. How difference in culture conditions affects sexual versus asexual populations could be an area of future research.

3.6 Concluding Discussion

Fitter populations tend to have fewer SNPs at medium to high frequencies. Is this a cause or symptom? Having fewer SNPs may mean less mutational load, making fewer SNPs a cause of fitness. Being a fitter population is associated with having been driven to a higher frequency in the population by selective processes, which may have selected clones with less mutational load, bringing them to higher frequency in the population, making higher fitness a cause of having fewer SNPs. Sequencing of individual clones over time may help to answer which is the case in a given situation.

The differences between our data and those of McDonald et al (2016) may also be worth further exploration. It appears likely that some of the sharp differences observed between sexual and asexual populations in the McDonald et al. data may be due to the asexual lines' haploid genomes as compared to the sexual lines' diploidy rather than a difference due to sexual reproduction per se. It is plausible that the large difference in LOF mutations (more in asexual) may be due to haploids tending to fix LOF mutations more. The difference in numbers of SNPs between sexual and asexual populations is also arguably expected, in that it is more of a challenge to fix a mutation in a diploid than a haploid genome, and so perhaps it is not too surprising that those variants that are fixed in the diploid populations are overwhelmingly non-coding.

There is much future work to do in this whole area. With the Goddard-Gray populations, future work could include testing variants & integrating genetic with phenotypic data, time-course data (as

in Lang et al., 2013; McDonald et al. (2016)), partial gene-flow populations, disentangling the different elements of sex vs asex. And exploring non-SNP variants to a greater extent.

Putative genes are another area of possible work. 672 out of the 6577 sequences included as nuclear genes in the genome features file downloaded from SGD for Y55 are described as 'dubious open reading frames', i.e. slightly more than 10%.

3.7 Methods

Populations were regrown from -80°C stocks stored after an adaptive evolution experiment by Jeremy Gray and Matthew Goddard. The 24 populations sequenced had been evolved for approximately 300 generations in nutrient limited chemostats. For further details of this adaptive laboratory evolution experiment, see Gray (2011), and Gray & Goddard (2012). For the creation of the sexual and asexual strains (through knockout of SPO11 and SPO13 to create an asexual strain) see Goddard et al. (2005). The conditions under which each population were evolved are given here:

DNA was extracted from the populations from -80°C frozen samples in glycerol (Gray, 2011), grown in 30mL of YPD in a 28°C warm room for 48hours, followed by use of a Qiagen genomics kit for purification.

Sequencing was conducted by BGI in Hong Kong and by NZGL. The majority of the data was from a Hi-Seq run conducted in early 2013. As some of the samples were of low quality, including contamination of the ancestor with bacteria [confirmed with Blast search of some of the many unaligned reads], a second sequencing run with samples of the ancestors was conducted by NZGL, on a MiSeq machine. This sequence reads for both ancestors (sexual and asexual) was then collated with the HiSeq pooled ancestor sample after mapping, and this merged file was then used in future as the baseline against which to detect evolutionary change over the course of the 300 generations.

The reference genome was chosen by comparing alignment rates against a few available genomes, and the Y55 genome from Saccharomyces Genome Database (rather than, for instance, the newer 2014 Stanford Y55 genome) gave the best coverage results after mapping. For the exact scripts used for SNP, Indel, and CNV detection, see Appendix A. Alignment against the reference was conducted using the short reads aligner Bowtie-2 (Langmead et al., 2009). Converting the binary alignment files giving full details for each read to 'mpileup' files arranged by genomic site, and subsequently calling SNPs involved using the programs Samtools (Li et al., 2009) and VarScan2 (Koboldt et al., 2012)

respectively.

Any SNPs found in sites with SNPs called in ancestral population genome files were removed, using both VarScan and Genome Analysis Toolkit (GATK) (McKenna et al., 2010) SNP calling programs, however on manually checking some putative derived SNPs against the ancestral bam files, still many were found to be present in the ancestor (i.e. calling of ancestral SNPs was not sensitive enough). Consequently, any SNPs at sites found with variants across many replicates were removed. For the 10 zero gene-flow populations, SNP sites found in six or more populations were removed, based on a plot of the frequency of putative SNP sites across populations, showing that putative SNP sites that were highly parallel across 3-7 populations were rare. Those found in six or more populations were assumed to be ancestral, for the initial conservative filtering.

Data quality:

The Y55 genome from SGD was chosen after checking alignment against the Stanford Y55 genome from SGD and S288c (SC 73) genome from Ensembl; alignment was best overall against the SCY55 genome.

The number of high read depth sites for each sample, were counted, by chromosome – low coverage chromosomes (e.g. 14a chromosome 4) were excluded from further analyses.

Chapter 4
The Role of Mutator Genotypes in Microbial Trade-offs

The effects of sexual reproduction and mutator genotypes on metabolism in *S. cerevisiae*, and genotype x environment interactions for this trait across multiple environments are reported.

4.1 Sex, Mutation, and Trade-off

The effects of sex and mutation on trade-offs have not previously received much attention. Trade-off, or decreased fitness relative to the ancestor in an alternate environment, is one 'side effect' of adaptation. Here we investigate a related side-effect, changes in rates of metabolism across environments, with the use of Biolog ™ multi-well plates to measure metabolism in populations from a laboratory evolution experiment previously conducted by Jeremy Gray, reported by Gray (2011) and (Gray & Goddard, 2012b).

Biolog plates have previously been used in the context of experimental evolution, for instance in the study of trade-off models in *E. coli*, with a comparison of two key papers here, Cooper & Lenski (2000) and Leiby & Marx (2014), discussed in Chapter 7. Biolog plates have also been used to study different yeast populations, for instance Samani et al. (2015) compared metabolic differences in various wild yeasts with this technique.

The original experiment, as detailed in Gray (2011), involved daily serial transfer of sexual (wildtype and asexual (two meiosis genes knock out strain) yeast populations with or without a mutator phenotype (DNA repair gene knockout) into fresh tubes containing 3mL of culture media. 40uL was transferred each time. Each of the four treatment conditions was grown in triplicate, i.e. three isolated lines for each treatment were maintained, over the course of approximately 300 generations of population growth. In addition to this transfer in benign media, the same process was conducted in 'harsh' culture media, which involved the addition of a high concentration of sodium chloride, adding osmotic pressure to the yeast cells. Eight treatments conducted in triplicate made for a total of 24 independent lines.

When the evolutionary fitness of each population was tested at the end of the experiment, it was

found that sexual populations were fitter than asexual populations, that asexual mutators had the least fitness and sexual mutators the highest fitness. (Gray, 2011). This suggested that there was a benefit to sexual reproduction that was able to overcome a negative effect of high mutation rate.

To assess other phenotypes affected by evolution, the populations were all tested for metabolism on carbon and nitrogen sources (Biolog™ PM1 and PM3 plates, respectively), and the change from the measurements obtained for the 'ancestor' strain (the strain prior to any evolutionary changes) compared.

4.2 Results

The carbon-limited wells (Biolog ™ PM1) are the main focus of most of this results section, as both the optical density of derived populations after four days to test for metabolism, and the changes in metabolism between ancestral and derived populations were greatest on these plates, providing more opportunities to see whether differences between treatments are real.

Numbers of Trade-offs:

For the nitrogen source plates (PM3), within a treatment (such as benign environment populations, non-mutator) sexual populations tended to exhibit fewer trade-offs than asexual populations – that is, multi-well plates from sexual populations tended to have fewer wells (as compared to those from asexual populations) which showed a decrease in metabolism relative to the average ancestral value for that well. Mutator populations showed fewer trade-offs in the benign environment, and more in the harsh environment than non-mutator populations. These tendencies however were not statistically significant, and rely to a large extent on some very large numbers of trade-off observed in asexual and mutator populations.

The general trend is the same for the carbon source plates, with the exception of the mutator populations adapted to the harsh environment. The asexual mutators in the harsh environment, when the data for each well was pooled and averaged, exhibited fewer trade-offs than the sexual mutators.

Metabolism on a certain number of nutrient sources would be expected to exhibit a trade-off effect by chance. For instance, if evolution in these populations does not have a direction on average, then we would expect approximately 50% of the wells to show higher metabolism relative to the ancestor, and 50% to show trade-off. There are some populations that have more than 50% of wells

showing trade-offs, for the nitrogen source plates. However, it is the significant differences amongst treatments rather than the absolute number of trade-offs that is of most interest here.

A three-way ANOVA was conducted in SPSS on the effect of environment, sexual status, and mutation rate on the number of metabolic trade-offs on Carbon sources (Biolog™ PM1). No significant interaction between these factors was discovered $F(1,15)=0.364$, $p=0.555$. However, a significant effect due to mutation rate was found $F(1,15)=5.399$, $p=0.035$.

The same analysis for the Nitrogen sources found no significant interaction between the factors ($F(1,15)=0.079$, $p=0.783$), and no significant effect due to any individual factor either. Interestingly, the factor closest to having a significant effect was sexual status rather than mutation rate ($F(1,15)=2.842$, $p=0.112$).
There is, surprisingly, no relationship between the numbers of trade-offs observed in each population on carbon and nitrogen sources; correlation coefficient $R^2=0.0024$ (data for populations with multiple technical replicates is removed).

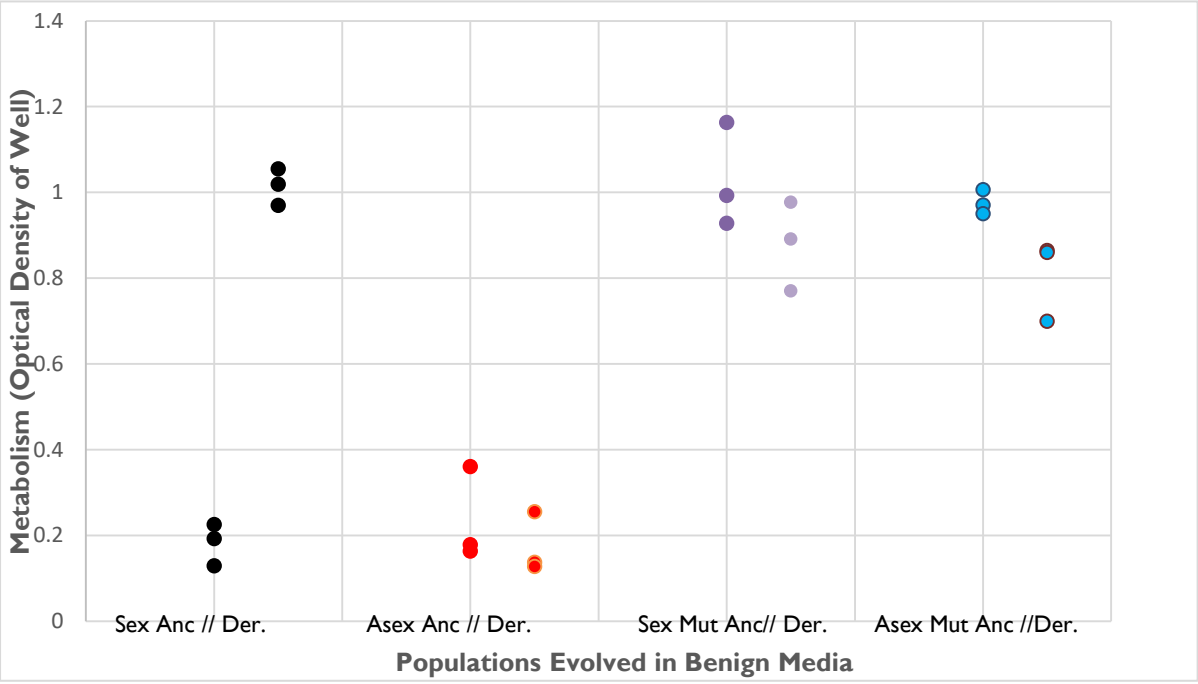


Figure 4.2a: Populations evolved in benign media show differences to the ancestors (measured in triplicate) in glucose metabolism, with a tendency towards trade-off in mutator populations. Mutator ancestors, both sex and asex, begin with much higher glucose metabolism than non-mutator ancestors do. Sexual wildtype derived populations evolve to achieve a similar metabolism to mutator ancestors.

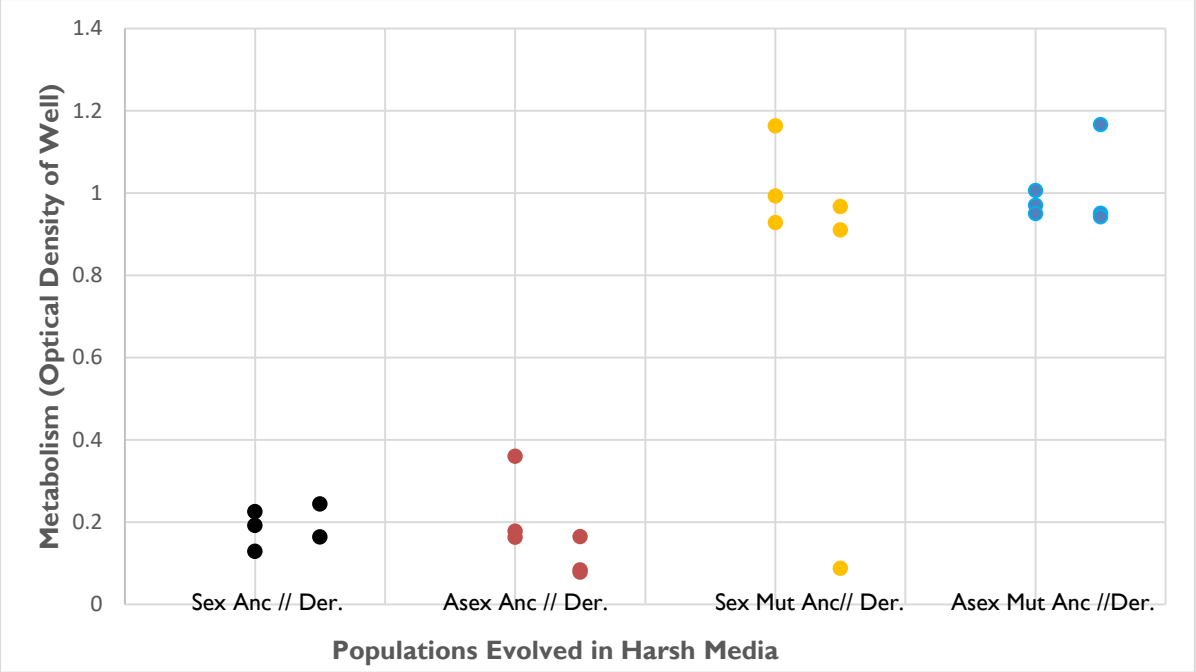


Figure 4.2b: Populations evolved in harsh media are less distinct from the ancestors (measured in triplicate) in glucose metabolism, with a tendency towards trade-off rather than adaptation. Mutator ancestors, both sex and asex, begin with much higher glucose metabolism.

There is a tendency for sexual populations to display fewer trade-offs than asexual populations (statistically significant for carbon sources) and for wildtype populations to show fewer trade-offs than mutator populations (non-significant).

When all populations are considered together, there is a statistically significant difference between sexual wildtype populations and asexual mutator populations, with sexual wildtype populations tending to have fewer trade-offs (when numbers of trade-offs for technical replicates are averaged, for a paired t-test comparing numbers of trade-offs by treatment group, p=0.037488.)

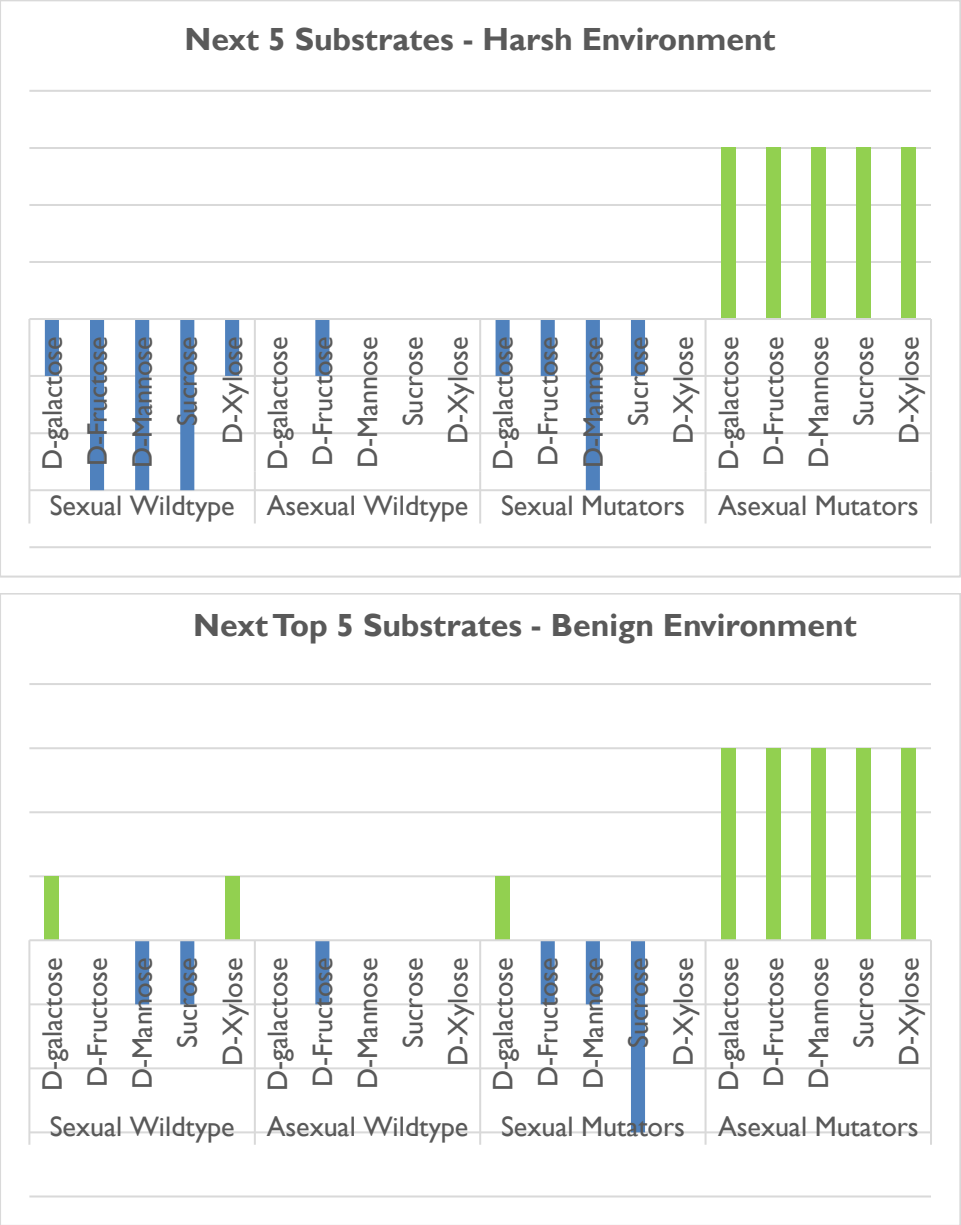
Metabolic Performance on Glucose

Complicating the analysis is the fact that the different ancestral populations did not begin at the same point in their metabolism on different substrates. Particularly noticeable were differences on glucose, illustrated above, in **Figures 4.2a and 4.2b**. The derived populations also clearly show the benefit of sexual reproduction for adaptation as it relates to glucose metabolism in this system.

Metabolic Performance Across 'High Respiration' Wells.

As illustrated in Figures **4.3a and 4.3b**, there is a strong tendency for evolved asexual mutator populations from the Gray-Goddard sex/mutator experiment populations (Gray, 2011) to exhibit

significantly greater metabolism than the ancestors from which they were derived on each of the five sugars listed. The pattern observed raises the possibility of something like the Warburg effect seen in cancer cells also happening here – cells unable to clear mutations have a high metabolism rate. It has in the past been debated whether this is a cause or effect of a cancer phenotype, but is typically taken now as an effect of high genotypic mutation.



Figures 4.3a and 4.3b: Asexual mutators all tend to show significant adaptation to these 'next top five' substrates (highest metabolisers after glucose) across both benign and harsh environments, on average. Other populations tend to show trade-offs. Tall bars indicate a significant change; green are positive changes relative to the ancestral replicates (adaptation), blue are negative (trade-off).

Relationship to Fitness

There were only very weak correlations between numbers of trade-off measured by changes in metabolism on the Biolog™ PM and PM3 plates and the fitness measured by Jeremy Gray (data not shown). These were sometimes positive and sometimes negative. This means there is not strong evidence for antagonistic pleiotropy underlying trade-off in this system.

4.3 Discussion:

This study was designed to get a new perspective on what I term the 'hidden effects of adaptation', and shows that for *Saccharomyces cerevisiae* populations following a period of evolution, both sexual status and mutation rate have an impact on metabolism in environments other than the environment of selection.

The difference seen in trade-offs following population growth in the benign environment contrasts with the fitness results measured by Jeremy Gray (2011). It was noted there that "It can be concluded from these results that in benign media under purifying selection, sex and recombination play little to no role in mutation clearance in *S. cerevisiae*". In light of the multi-well plate results, I argue that sex prevents the accumulation of conditionally deleterious mutations in these populations, but the effect only becomes visible when fitness is measured more holistically across multiple environments.

I hypothesise that a mutational ratchet is in fact operating in these populations, evident to an extent in wildtype asexual populations, and particularly in mutator populations. In many cases, sexual reproduction is able to limit the tendency towards trade-off due to the silent ratchet. The fact that asexual and mutator populations tend more towards trade-off than sexual and/or wildtype populations may suggest that the mechanism of trade-off in this system is mutation accumulation. Add to this the fact that in the benign environment, no significant fitness changes were seen, implying that few mutations fixing in that environment are adaptive, and the case becomes reasonably strong.

However, a number of features of the study mean that this conclusion cannot be confidently made. Firstly, the measure used was metabolism rather than fitness, and metabolism may increase in very unfit populations, as seen in the 'Warburg' effect in cancer cells. Secondly, the general applicability of

any finding of mutation accumulation is limited, as the serial transfer mechanism used in the experiment means that the populations passed through a bottleneck of approximately 1.3% of their population each day (40uL out of 3mL were transferred each time).

In conclusion, this study gives some evidence for mutation accumulation being particularly prevalent in asexual mutator populations, as expected, but the results are ambiguous as to whether this mechanism is also at work with sexual reproduction or in wildtype (non-mutator) populations, where few trade-offs and many gains in metabolism (metabolism increased on many sources) were generally seen.

4.4 Methods:

Biolog Plates:

Day 1: Inoculate 50 mL falcon tubes containing YPD with 5uL of the frozen sample in glycerol.
Day 2: prepare solutions for the Biolog plates
Day 3: Spin down falcon tubes (5 mins at 3000rpm), replace YPD with water, starve cells for 4hrs. Find appropriate cell concentration for approx. 62% transmittance. Create inoculation mixes using dye and fluid provided by Biolog™.
Inoculate Biolog plates with 0.25mL of the diluted culture, as well as Biolog™ dye D and the inoculation fluid supplied, following standard Biolog™ procedure. Use the same cell suspension for one PM1 and a PM3 plate, hence make at least double the mix required for one plate. (Note, this was not followed for all additional plates added later as potential replacements).
Read plates with EnSpire plate reader. Repeated at 24 hours, 48 hours, 72 hours, 96 hours.
The plates were inoculated and read in five separate subsets, to make the process manageable. Each treatment group was divided across these five runs, to prevent any potential batch effects from dominating the analysis.

Ancestral population samples (Sexual wildtype, asexual 'wildtype' (SPO11 and SPO13 knockouts; see Goddard et al., 2005), sexual mutators, and asexual mutators (both have MSH1 knocked out) were measured in triplicate for PM1 plates, and duplicate on PM3, based on initial readings showing that metabolism of these populations was higher and exhibited more variation on PM1 plates.

Analysis:

The measure of metabolism used was obtained by subtracting optical density at the first plate read (0hrs of growth on the nutrient source) from optical density after 72 hours. The empty wells (well number 1 out of 96) were used to control for initial differences for the carbon sources (PM1 plate) – change in optical density in these wells was subtracted all of the measures of metabolism (change in optical density from 0hrs to 72hrs in the plate), as there was a weak positive relationship between the measure of metabolism in the empty well and average results for other wells. When the empty well value was subtracted from the others, the strength of the relationship decreased from an R2 correlation coefficient of 0.379 to 0.0784, and further to 0.0061 when a single outlier data point was excluded. The same process was conducted for the PM3 data, although it showed no positive relationship between change in optical density in the empty well and average change in density across the other wells. Data available on request from Zachary Ardern.

Chapter 5
Antagonistic Pleiotropy and Trade-off

The molecular basis of evolutionary trade-offs across environments, where fitness decreases in alternate environments as a result of ecological specialisation, has generally been attributed to either selected or neutrally accumulated variants. As such, there are two commonly discussed mechanisms of trade-off termed 'antagonistic pleiotropy' and 'mutation accumulation' respectively. Previous analysis of experimental evolution in S. cerevisiae (Gray & Goddard, 2012b) suggested that the mechanism underlying trade-off between environments was mutation accumulation – mutations accumulated neutrally over the course of adaptation to one harsh environment were responsible for concomitant loss of fitness in the alternate environment. We argue that additional factors not taken into account in the previous analysis, including analysis of time-course fitness data and whole-genome sequencing results indicate a more important role for antagonistic pleiotropy undergirding the trade-offs observed, in line with most findings in other microbes with small genomes.

5.1 The Experiment

An experimental evolution project in *Saccharomyces cerevisiae* was previously conducted by Jeremy Gray and Matthew Goddard (Gray, 2011; Gray & Goddard, 2012a). 60 populations of yeast were grown from isogenic starting points in one of two nutrient-limited media environments in chemostats for 300 generations, with fitness tested at multiple time points throughout, and rounds of meiosis facilitated by sporulation. Five different levels of gene-flow between populations, ranging from full gene-flow (50% of paired populations mixed every 12 generations) to zero gene-flow (these populations are the main main focus of Chapter 3). 30 populations of asexually reproducing yeast were similarly grown, using a strain isogenic apart from two gene knock-outs. Of interest for the purposes of assessing trade-off were 12 sexual and 11 asexual populations for which the most fitness data was obtained, with environmental treatment conditions listed here:

Populations	Sexual Status	Environment	Migration Rate
1a,2a,3a	Asex	N-lim, high salt	Zero Gene-flow
13a,14a,15a	Asex	N-lim (mixed)	Full Gene-flow
17a,19a	Asex	C-lim, high temp	Zero Gene-flow
29a,30a,31a	Asex	N-lim (mixed)	Full Gene-flow
1s,2s,3s	Sex	N-lim, high salt	Zero Gene-flow
13s,14s,15s	Sex	C-lim (mixed)	Full Gene-flow
17s,18s,19s	Sex	C-lim, high temp	Zero Gene-flow
29s,30s,31s	Sex	C-lim (mixed)	Full Gene-flow

5.2 Arguments for Mutation Accumulation (MA)

In all sexual and four out of six asexual populations exposed to just one environment, trade-offs were observed in the alternate environment at some point over the course of the 300 generations. However, in all sexually reproducing populations with full gene-flow apart from one, trade-offs were not observed; these populations were 'superior generalists' (Gray & Goddard, 2012a). In asexually reproducing populations with full gene-flow, trade-offs were also observed; there is something different about adaptation in sexual populations with gene-flow. This distribution amongst the

populations of trade-offs according to gene-flow and sexual status was taken to imply that the cause of trade-offs was the accumulation of neutral mutations (Gray & Goddard, 2012a). The key observation was in the sexually reproducing strain, namely the existence of trade-offs in zero gene-flow populations but not in full gene-flow populations. This was a striking finding, but does not convincingly show that mutation accumulation undergirded these trade-offs, in light of the arguments below and the possibility of an alternative explanation.

It may be that 'antagonistically pleiotropic' sites responsible for trade-offs simply were not retained in the sexual populations with full gene-flow, due to selection pressure against them. Sexual reproduction and gene-flow were not added to the system as a perturbation, but instead the adaptation under investigation (in sexual full gene-flow populations) was conducted with the capacity for meiotic recombination and exposure to both niches always present. As such, as long as there is a subset of adaptive mutations which is not antagonistically pleiotropic, superior generalists will be able to evolve – from this observation, it seems that little can be legitimately be concluded about the basis of adaptation in the other populations (i.e. asexual and sex-without-gene-flow). A similar suggestion was made by Elena and Lenski regarding plant virus evolution; it may be that there are two classes of mutations and "[e]ven if mutations with host-specific benefits were more common than the generally beneficial mutations, the latter class would be differentially enriched in viral populations that evolved on alternating host types" (Elena & Lenski, 2003).

The evidence outlined above and in the previously published paper (Gray & Goddard, 2012a) is consistent with MA underlying the trade-offs observed in the asexual populations and sexual populations without gene-flow. However, it is also consistent with these populations experiencing trade-offs in accordance with antagonistic pleiotropy (AP) and is such is not contrastive evidence for MA over AP. The new analyses however favour antagonistic pleiotropy as the main contributor to environmental trade-offs in this system.

5.3 Evidence for Antagonistic Pleiotropy

Five new analyses all suggest a greater role for antagonistic pleiotropy than previously hypothesised. Firstly, in sexual populations, fitness data from both environments from multiple time points during the experiment suggests antagonistically pleiotropic effects of beneficial mutations. Secondly, a regression analysis of end-point fitness data supports antagonistic pleiotropy. Thirdly, reconsidering the difference in the costs of adaptation between sexual and asexual populations weakens the support for mutation accumulation as a cause of trade-off. Fourthly, consideration of the mutational resources available raises the question of whether the system has the capacity for experiencing significant fitness costs from neutrally accumulated mutations..

Time-Course Fitness Data

Fitness in both the selective and alternative environments over the course of the 300 generation experiment was measured by Jeremy Gray for the 12 sexually reproducing populations (Gray, 2011). Insufficient data points are available for the asexually reproducing populations, which in any case did not show a significant trade-off overall (Gray, 2011). Tracking correlated fitness changes in each over the course of the 300 generations shows that increases in the selective environment are generally associated with a decrease in fitness in the alternate environment. (Figure 5.1.) Further, there is no evidence of a steady accumulation of trade-off as expected under mutation accumulation. For instance, five out of six of the 0-gene-flow sexual populations exhibited their greatest trade-off within the first 200 generations, further implying antagonistic pleiotropy as the mechanism of trade-offs for the most significant trade-off causing mutations. In other words, for the most part, trade-offs in sexual populations were high early on in the course of the experiment, and decreased over time.

The dynamics of evolution are quite different in the sexual and asexual populations, including in the C-limited populations, which is masked by using limited fitness data from one time point. The difference is shown more clearly when data from 0 gene flow and 0.000001 proportion gene-flow populations are pooled to create a time series chart. (Figure 5.2).

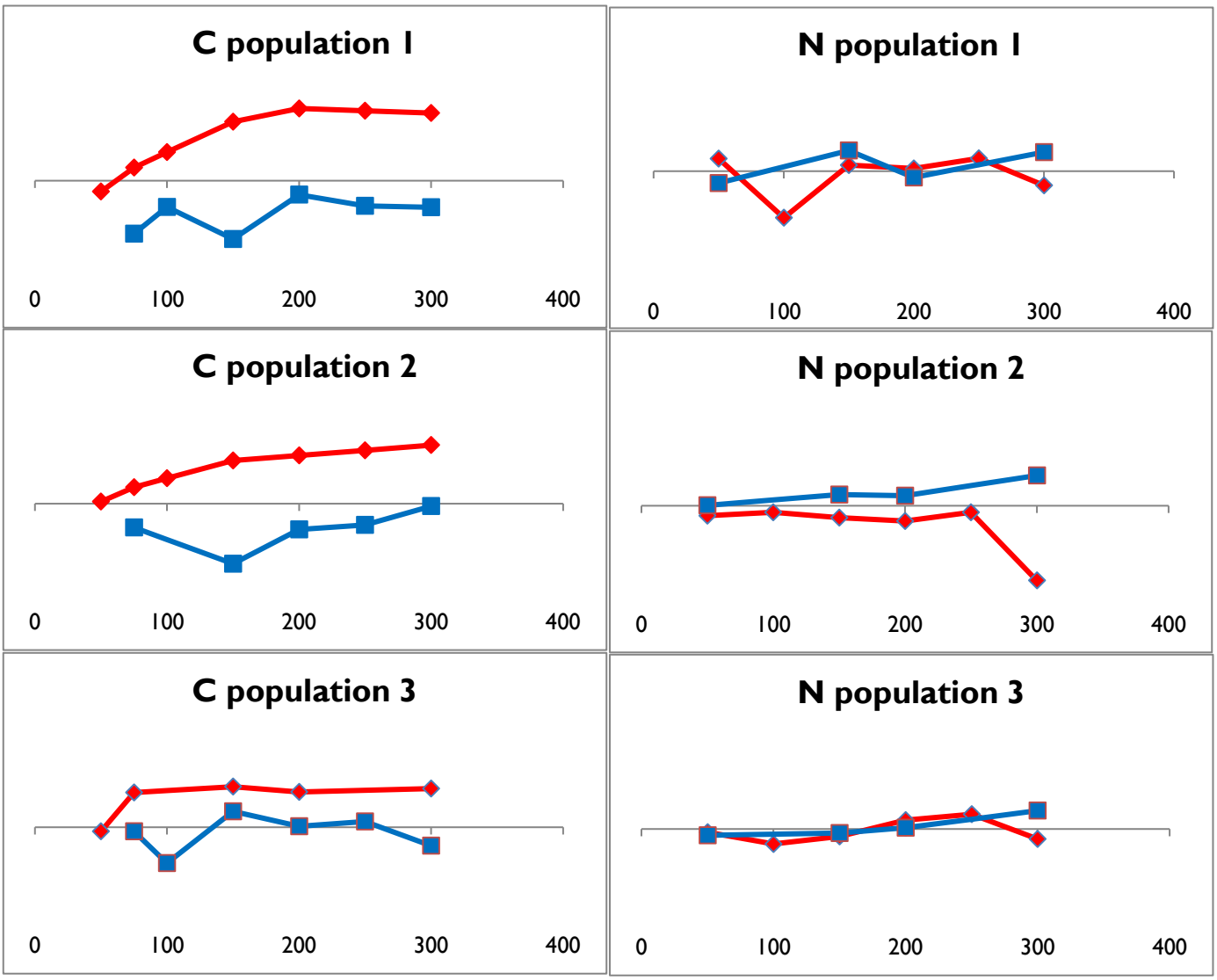


Figure 5.1 'Zero gene-flow' sexually reproducing population fitness over time. The y axis, fitness, has the same values in each case. Fitness in the Carbon limited (C) environment is in red, fitness in Nitrogen limited (N) is in blue. E.g. for C population one, C is the selective environment, and N is the alternative environment, showing trade-off over the whole course of the experiment.
Large increases in the selective environment tend to be associated with decreases in the alternate (larger trade-offs), and most adaptation occurs in the first half of the experiment.

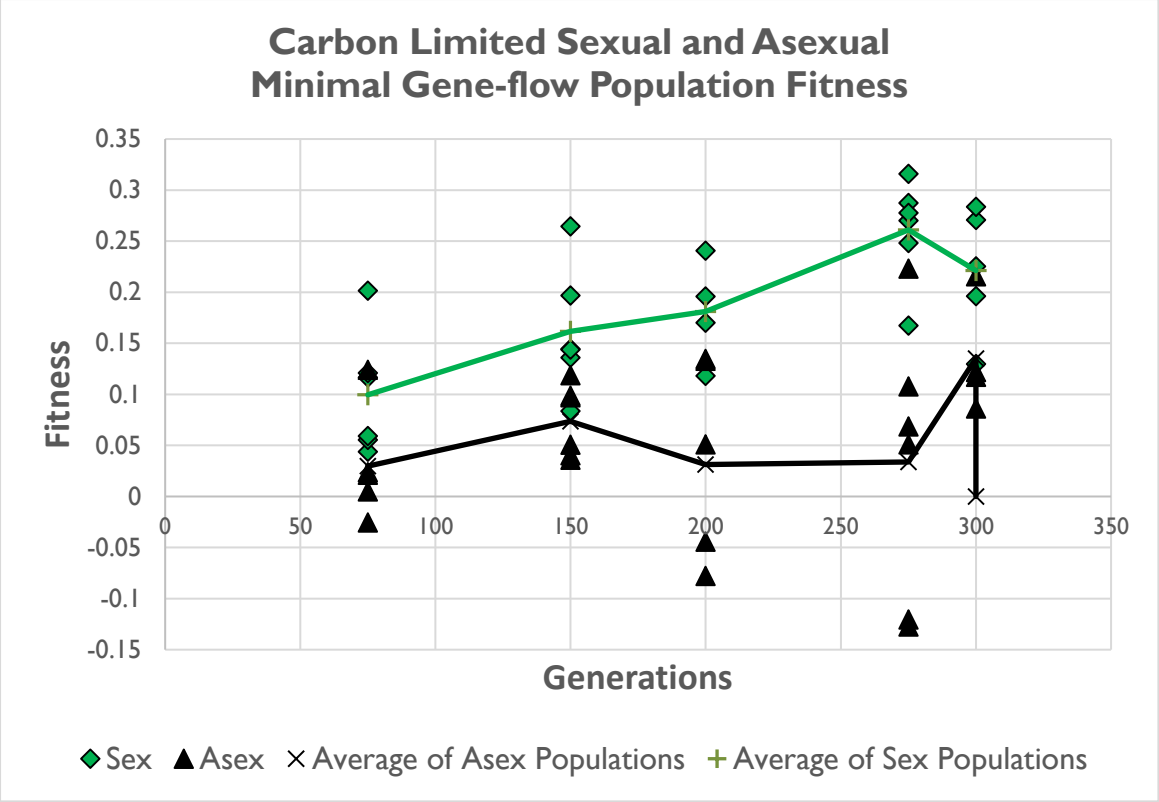


Figure 5.2 'Zero gene-flow' and '0.000001' proportion gene-flow sexually reproducing population fitness in their selective environment, over time, for both sexual (green diamond) and asexual (black cross) populations. The difference between sexual and asexual populations is statistically significant; for example, treating corresponding sexual and asexual populations as paired, the p value is 1.59E-05 (paired two-tailed t-test).

Regression

A regression of the cost of adaptation on the direct response to selection (Maclean & Bell, 2002) is one analysis not conducted by Gray and Goddard (2012a). With this approach, according to Maclean & Bell, if both slope and intercept are positive, there is evidence that both MA and AP contribute to the cost of adaptation. If the intercept is positive and slope is negative, only MA contributes to the trade-off; when the intercept is negative and slope is positive, only antagonistic pleiotropy is the cause of trade-off. To explain these predictions, the relationship between cost and fitness represented by these different graphs should be considered. A negative intercept shows that adaptation is required before any cost is observed, whereas a positive intercept shows that a fitness cost at least sometimes occurs without adaptation. A positive slope shows that increasing adaptation leads to increasing cost in the alternate environment, and a negative slope that decreasing adaptation leads to increasing cost.

The regressions for the sexual and asexual groups with zero gene-flow, as well as all with zero gene-flow clustered together, for each environment, are shown in Figure 3. This analysis demonstrates evidence of both AP and MA (for those population groups with positive slope and intercept). The asexual populations were not focussed on, as fewer showed trade-offs, and they were not the focus of the earlier paper (Gray & Goddard, 2012a).

Trade-offs in Both Sexual and Asexual Populations

Figure 5.3: Regression of the cost of adaptation on the direct response to selection for 0 gene-flow populations. Both show evidence of Antagonistic Pleiotropy, while N-limited populations also show evidence of Mutation Accumulation (positive intercept indicates that a fitness cost exists without a direct response to selection).

A trade-off between adaptation to the 'hot C' and 'osmotic N' conditions was not observed in asexual populations after 300 generations, considered as a group, although some populations did show a trade-off when considered individually. Trade-offs were observed in all 0 gene-flow sexual populations. There is no reason to expect large differences in mutation accumulation between sexual and asexual populations, but more rapid adaptation was observed in sexual populations than in asexual populations. If there is any difference in the accumulation of neutral mutations, we would expect more accumulation of non-adaptive mutations in asexual mutations through processes such as genetic hitchhiking. This larger accumulation of non-adaptive mutations is proposed from the analysis in Chapter 3, showing that asexual populations contain proportionately more low frequency mutations. As such, the existence of more trade-offs in sexual than in asexual populations favours antagonistic pleiotropy rather than mutation accumulation as the explanation of the trade-offs occurring.

Insufficient Mutational Capacity

In order for mutation accumulation to be the cause of trade-offs, sufficient conditionally neutral mutations must have occurred and risen in frequency in the populations through stochastic processes. The resources available are a genome size of approximately 12.5 million base pairs, a mutation rate of approximately 6.91×10^{-8} mutations per base pair per generation (Gray & Goddard, 2012b) and 300 generations of evolution. The fixation rate for neutral alleles is μ, the rate of neutral mutations.

The actual rate of neutral allele fixation is difficult to quantify. As Bank et al (2014) note, neutral mutations are expected to be those with $s < 1/Ne$, which is below the detectability of selection experiments when Ne is large. One comparison is given by a study in the wild yeast *Saccharomyces paradoxus*, where based on a population genetics model approximately 20% of spontaneous mutations were predicted to be effectively neutral in the natural environment (Koufopanou, 2015). In a direct empirical study comparing different kinds of mutations, approximately 20% of spontaneous mutations in *S. cerevisiae* were found to have a fitness effect of 5% or less (Wloch et al., 2001). The effective population size in S. paradoxus in the wild is approximately 8 million (Tsai et al, 2008), compared to over 100 million *S. cerevisiae* cells in the chemostat (Gray, 2011, p. 195). A larger population size means that fewer mutations will be effectively neutral; in this case, we would expect the relevant value of s to be a tenth or smaller the value in the Koufopanou (2015) study.

Taking both of these proportions of neutral mutations into account, and recognising that fewer mutations are likely to be effectively neutral in a harsh environment such as 'hot C' or 'osmotic N', we assume approximately 10% of spontaneous mutations in *S. cerevisiae* will be effectively neutral under the conditions of each of the chemostat experiment selective environments. As such, with a neutral mutation rate of 6.91*10^-9 (i.e. 10% of the total mutation rate), a genome size of 12.5 million base pairs, and 300 generations of evolution, we would expect approximately 26 mutations in each individual at the end to be present in the genome due to neutral evolution. The proportion of putatively neutral mutations are fitness reducing in alternate environments, along with the number of mutations that are neutral in large populations are important areas for future empirical studies.

Also important to take into account in this instance is the time-course data, which shows that trade-offs developed quickly in the populations, within 100 generations, within which time period an average of only approximately 9 mutations are present in each cell assuming the conservative parameters given above; too few to have a significant effect on fitness unless a large proportion of the genome has conditional fitness effects. Another consideration which may further reduce the

relevance of neutral mutations, through limiting the spread of neutrally accumulated mutations, at least in the asexual populations, is the operation of background selection – where selection against linked deleterious mutations (those co-occurring in a genome) reduces neutral variation (Charlesworth, 2012).

5.4 Discussion

The proposal of this chapter is that trade-offs observed between the 'hot C' and 'osmotic N' nutrient-limited environments following adaptation to the other environment in research conducted by Jeremy Gray and Matthew Goddard are predominantly due to antagonistically pleiotropic variants.

Sexual populations are more effective than asexuals in accumulating adaptive variants, which comes with the risk of trade-offs. The presence of trade-offs in sexual populations under zero gene-flow but not full gene-flow conditions suggests that trade-offs can be mitigated. A question remains however, concerning why asexual populations did not exhibit many trade-offs, even when they achieved similar levels of adaptation to the sexual populations. It is possible that the asexuals' larger population sizes enabled them to access different variants, which were not associated with trade-offs to the same extent as those fixed in sexual populations. One factor may be that the asexual populations, which had larger population sizes, may accumulate fewer nearly neutral mutations, and thereby avoid some neutral mutation accumulation which influenced the sexual populations; however in light of the other arguments offered this is unlikely to be the dominant factor. Perhaps clonal interference slowed both the adaptation and the development of trade-offs in asexual populations. I propose that whatever the precise reason, different evolutionary dynamics meant that antagonistically pleiotropic variants were fixed to a greater extent in sexual than in asexual populations.

Chapter 6
Genomic complexity drives differential mechanisms underlying trade-offs among taxa

Adaptation to a new environment is correlated with a wide range of fitness responses in other environments, and we currently have a very limited ability to predict the nature and extent of these consequences. The molecular basis underpinning competitive fitness in different environments is a pressing issue with application to understanding the evolution of drug resistance, metastasis in cancer, species responses to climate change, and shifts in ecological niches more generally. Using simple logic supported by data, we argue for a general rule predicting that the nature of these responses will be affected by genomic complexity. We predict that for adaptive genetic variants, the average antagonistic fitness effect in alternative environments will be greater in prokaryotes than eukaryotes and that the basis of environmental trade-offs in multicellular eukaryotes will be predominantly due to mutation accumulation. In support of this hypothesis we assess experimental evidence on the basis of environmental trade-offs in microbes, aspects of microbial genomes that result in antagonistic pleiotropy, and factors contributing to genomic complexity that favour mutation accumulation.

6.1 Trade-offs and Complexity

The distribution of fitness effects across environments is a key question in developing our understanding of adaptation. Populations are routinely exposed to changing environments, and their evolutionary response to each environment is dependent upon the distributions of fitness effects. The molecular bases of these distributions, while often complex, are increasingly accessible through genome sequencing and high throughput phenotypic screening – particularly for microbes – and these phenomena have important implications in research areas including medical microbiology, cancer progression, ecology and conservation. The inherent evolvability of the genomic networks may be constrained by trade-offs, where adaptation to one environment is associated with fitness decreases in alternative contexts. We expect genomic complexity to contribute to the importance of selective processes in the molecular evolution of trade-off. We propose that insofar as genomic complexity can be quantified, it will be positively correlated with trade-offs to adaptation driven by maladapted alleles accrued through neutral rather than selective processes.

The effect of genetic complexity on adaptation and the distribution of the fitness effects of mutations

within a selective environment have been discussed widely in the literature. For example, Martin & Lenormand (2006a) discuss differences in mutational effects across species, and show that the average deleterious effect of mutations varies by two orders of magnitude from viruses to higher organisms such as vertebrates. In another paper Martin & Lenormand (2006b) also investigate the distribution of the fitness effects of mutations across environments. However, combining these concepts, to consider the possible effects of organismal complexity on the distribution of fitness effects across environments has received little attention. In this article we introduce concepts important to environmental fitness trade-offs and explore the likely contributions of increasing genome complexity on this phenomenon.

6.2 Trade-off Mechanisms

When adaptation to one environment decreases fitness in an alternate environment compared to the ancestor, there is said to be a 'trade off' in fitness between environments. (Cooper & Lenski, 2000). There are two main models of mechanisms underlying trade-offs, and these are termed 'antagonistic pleiotropy' (AP) and 'mutation accumulation' (MA). In AP, the same beneficial mutations accumulating in a population through adaptive processes directly contribute to a decrease in fitness when displaced into another environment. Alternatively, in MA, any decrease in fitness in different environments is due to the accumulation of mutations which were selectively neutral in the initial environment, through stochastic processes such as genetic drift, but which are selectively disadvantageous in alternate environments. Such variants are also termed 'conditionally neutral' mutations. It is not clear how these two processes contribute to fitness trade-offs generally, or the relative magnitude of each, and whether there are differential patterns regarding the mechanisms underlying trade-offs among taxa. In a recent empirical study, it was suggested "trade-offs to local adaptation in higher taxa might be due to mutation accumulation and not generally due to antagonistic pleiotropy" (Gray & Goddard, 2012). Here we find support for this claim, detailing experimental evidence and some of the features of genomes that are likely to underlie this trend.

Evidence of the processes of molecular evolution underlying trade-offs between environments, while there are few detailed studies, suggests that antagonistic pleiotropy is prevalent in prokaryotes and mutation accumulation is prevalent in higher eukaryotes. Some examples that we take to be typical, showing antagonistic pleiotropy in microbes, and mutation accumulation in higher eukaryotes follow. From microbial experimental evolution studies, excluding viruses, where the trade-off mechanism was inferred, the different types of trade-off mechanisms are plotted against genome sizes and mutation rates per base-pair and shown in Figure 1. This plot suggests that mutation accumulation is associated with a larger mutation rate and/or genome size. There is a significant difference in

mutation rate per genome per generation between those microbial populations exhibiting each mechanism (t-test, p = 0.001762). Antagonistic pleiotropy has been commonly found in viruses, e.g. Garcia-Arenal & Fraile (2013) report for plant viruses that antagonistic pleiotropy is a major cause of trade-offs across hosts. It is also prevalent in bacteria, such as observed in *Escherichia coli* by Behrends et al (2014) where variants facilitating growth in phosphate-limited conditions decreased growth in carbon-limited conditions. Antagonistically pleiotropic effects across environments are also common in fungi. In yeast, Hong & Nielson (2013) showed antagonistic effects across environments, where mutations in RAS2 (a protein involved in glucose signalling) responsible for adaptation to galactose metabolism also caused decreased ability to metabolise glucose. A similar finding was reported by Kvitek & Sherlock (2013), who reported that signalling pathway mutations are common in adaptive laboratory evolution and result in antagonistic pleiotropy. Processes of the accumulation of conditionally neutral mutations (mutation accumulation) leading to trade-offs have been inferred in a number of eukaryotes, including multicellular organisms such as the plant *Arabidopsis lyrata* (Leinonen et al, 2013) and in protists such as the amoeba *Chlamydomonas reinhardtii* (Reboud & Bell, 1997).

There are also some counter-examples to the general trend we propose. In multicellular eukaryotes, antagonistic pleiotropy is sometimes observed. For instance, in the Brassicaceae *Boechera stricta* a mutation in the flowering locus nFT was responsible for an antagonistic effect on fitness between environments (Anderson et al, 2013), and in the dung fly *Sepsis punctum*, antagonistic pleiotropy likely undergirded trade-offs in body size and growth rate at different temperatures (Berger et al, 2014). With genome sizes of 197MB and 285MB respectively, *Boechera stricta* and *Sepsis punctum* are towards the lower end of the distributions of plant and insect genome sizes; we expect larger genomes would exhibit more mutation accumulation. In the case of the *Sepsis punctum* study (Berger et al, 2014), the inference to antagonistic pleiotropy was made on the basis of a lack of superior generalists in populations experiencing a varying environment; the particular genomic basis is yet to be discerned. Additionally, in the flowering locus study (Anderson et al, 2013), in line with our expectations more of the genome (8%) was shown to exhibit conditional neutrality for this trait than antagonistic pleiotropy (2.3%). On the other hand, microbial species have also in some cases exhibited trade-offs between environments due to the accumulation of neutral mutations with deleterious effects in alternate environments. For instance, in one of the most often referred to instances of antagonistic pleiotropy, observed in populations from Lenski's long term evolution experiment in *E. coli* (Cooper & Lenski, 2000), it has subsequently been revealed that most of the trade-offs were actually due to mutation accumulation (Leiby & Marx, 2014). Trade-offs were, though, predominantly seen in mutator populations i.e. those with abnormally high mutation rates.

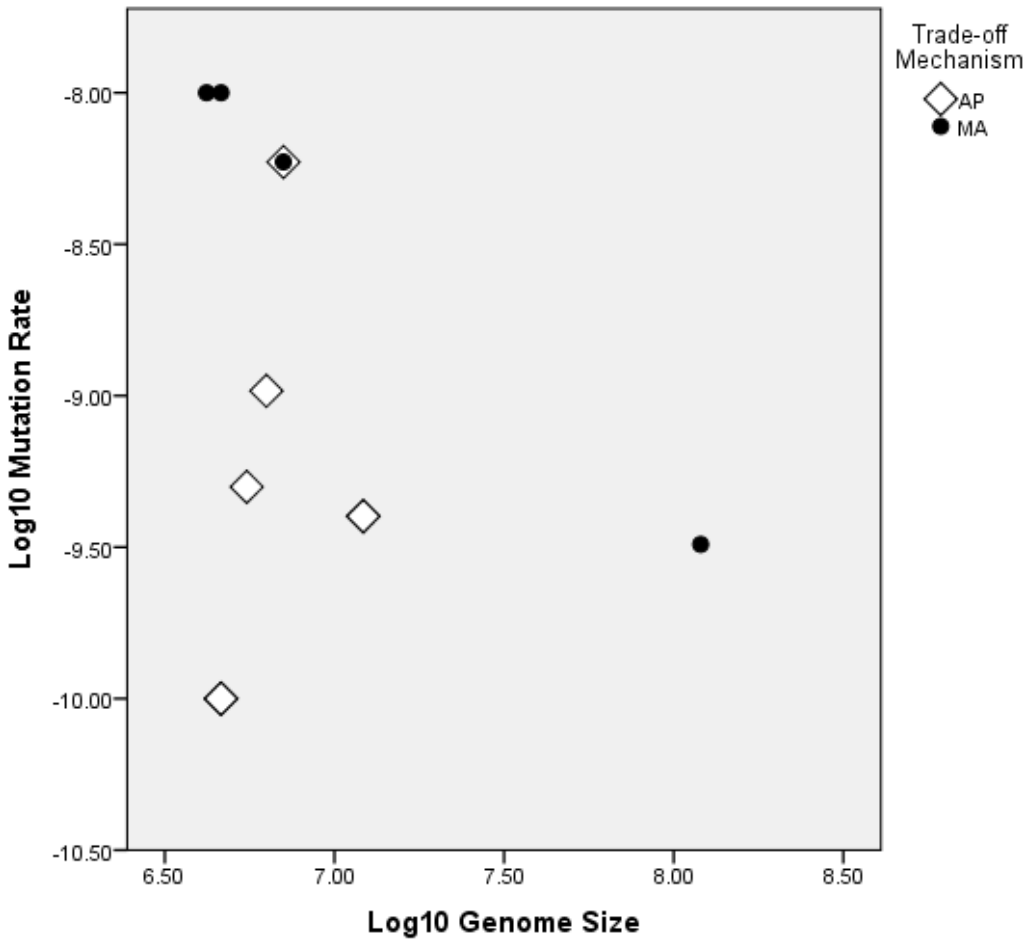


Figure 6.1: Trade-off models inferred in single celled organisms.

The data suggests that mutation accumulation only occurs in microbial populations with large mutation rates or with a large genome size. There is a significant difference in the mutation rate per genome per generation (mutation rate per nucleotide per generation x genome size) between populations showing antagonistic pleiotropy and those where mutation accumulation was found to be responsible for trade-offs. Student's t-test, $p = 0.001762$.

See **Appendix C, Supplementary Table 6.1** for data and references. Note that these mutation rates are indicatory, being the best estimates available; mutation rates per base-pair can differ widely across the genome and between strains, for instance a study in *Chlamydomonas reinhardtii* (Ness et al, 2015) demonstrated a 17-fold difference in rates across sites within a genome, seven-fold difference between strains, and nearly two-fold difference between individuals from a strain.

In addition to these high mutation rates, it is not clear whether the mutations responsible for trade-offs were actually conditionally neutral, or were perhaps mutations deleterious across multiple environments which hitchhiked along with adaptive mutations, or else accumulated due to a mutation rate exceeding natural selection's capacity to clear deleterious variants in these populations. It has been suggested that gene-loss in bacteria is commonly due to neutral factors (Lawrence & Roth, 1999), but there is evidence of both neutral and selective processes in the gene content of microbial genomes (Zinser et al, 2003; Lobkovsky et al 2013) and neutral processes may only be important in small populations, for instance of parasites (Wolf & Koonin, 2013), or over long time-spans.

On the basis of the above survey of the limited experimental evidence directly pertaining to the question available, we propose that antagonistic pleiotropy is typically, and mutation accumulation rarely, the basis of environmental trade-offs in microbes, at least during short-term adaptation with wildtype mutation rates. Further, we predict antagonistic pleiotropy of genetic variants across environments to be rare in higher eukaryotes, with the prevalence of mutation accumulation generally increasing with genome size and complexity.

6.3 Genome Complexity and Genome Length

We now turn to consider the implications of this prediction, and what constitutes complexity. That complexity has increased in some lineages has been described as a "rather boring conclusion" (Szathmary & Smith, 1995). Nonetheless, defining biological complexity is matter of contention and it can be compared across species in various ways. Two options are a measure that is relative to the environment, such as the sequence information in a particular organism that pertains to a particular environment (Adami, 2002), and a measure that concerns the number of different parts and the "irregularity of their arrangement" (McShea, 2000). We use a version of the latter in discussing genomic complexity, as an intuitive heuristic. Much biological complexity may be the result of neutral rather than selective processes, with evidence for this ranging from studies of bacterial endosymbionts of cicadas (Van Leuven et al, 2014) to differences between primates (Harris, 2010). As such, complexity should not be confused with adaptation or functionality. We consider complexity as a multi-level phenomenon, which increases with the number of parts and interactions between parts, and can be assessed at different levels of a biological system. Within organisms we can discern at least three levels of complexity – the length and internal structure of genes (intra-genetic complexity), interactions between genes including the number of genes (inter-genetic complexity), and interactions between cells and number of cell types (inter-cellular complexity). In this chapter we focus on relationships at the inter-gene level, which apply to microbes and multicellular organisms, and for which DNA sequence data is most apposite.

The most discussed element of molecular biological complexity is perhaps genome length, partly due to the famous 'C value paradox', where apparent complexity does not correlate well with genome length in nucleotides (Eddy, 2012). However, while a longer genome does not necessarily imply greater morphological complexity, greater complexity will often require a longer genome – for instance, prokaryotes are limited in both genome size and phenotypic complexity, arguably due to energetic constraints on genome size (Lane & Martin, 2010). An important lesson from genome sequencing is that complexity is not just about length of the genome; related issues are discussed in subsequent paragraphs. Longer genome sizes allow for more mutations to accumulate, particularly, in principle, amongst non-coding portions of the genome. If mutations in non-coding DNA are frequently conditionally neutral the apparent relationship between the proportion of the genome that is non-coding and organismal complexity (Taft et al, 2007) is relevant to the development of trade-offs through mutation accumulation. This is a question open to future investigation.

6.4 Genome Compactness, Epistasis, and Genome Complexity

The genomes of viruses, prokaryotes, and higher eukaryotes differ in structure. Virus genomes are generally relatively short, encoding just a few proteins. It is plausible that each virus protein contributes more directly to fitness than is the case in organisms with larger genomes and many proteins. Even when mutations are beneficial in one environment, we expect pleiotropic costs in alternate environments, at least as a general rule for loss of function mutations. Virus genomes are also particularly compact. Garcia-Arenal et al (2013) note that antagonistic pleiotropy is expected to be prevalent as a basis of in viruses because virus genes often encode multiple proteins. McGee et al (2014) also argue that the small genome size and overlapping genes in viruses contribute to the prevalence of antagonistic pleiotropy. Abrescia et al (2012) suggest that "the genetic economy of viruses does not permit most viruses to carry nonfunctional copies of genes." These features will ensure both a minimum of neutral mutations and a maximal 'side effect' in other environments for genetic variants, biasing the system heavily towards antagonistic pleiotropy. Similarly, bacterial genomes can have many genes overlapping other genes (Fukuda et al, 2003), as well as alternative transcripts within operons (Guell et al, 2011). Furthermore, prokaryotic genomes are typically more densely packed with genes than eukaryotic genomes. Eukaryotes by contrast have a weaker positive relationship between genome size and gene number (Hou & Lin, 2009; Friar et al, 2012) – see Figure 2 below. As Lane (2011) observes: "It is notable that eukaryotes support, on average, around 500 times more DNA than prokaryotes but only four times as many genes". We predict that high gene density is a causal factor in the development of antagonistic pleiotropy across environments for bacterial populations. As a nuclear compartment evolved, which separated transcription from translation, and larger and less compact genomes also evolved, opportunities for neutral mutations increased, allowing mutation accumulation as a possible basis of environmental trade-offs.

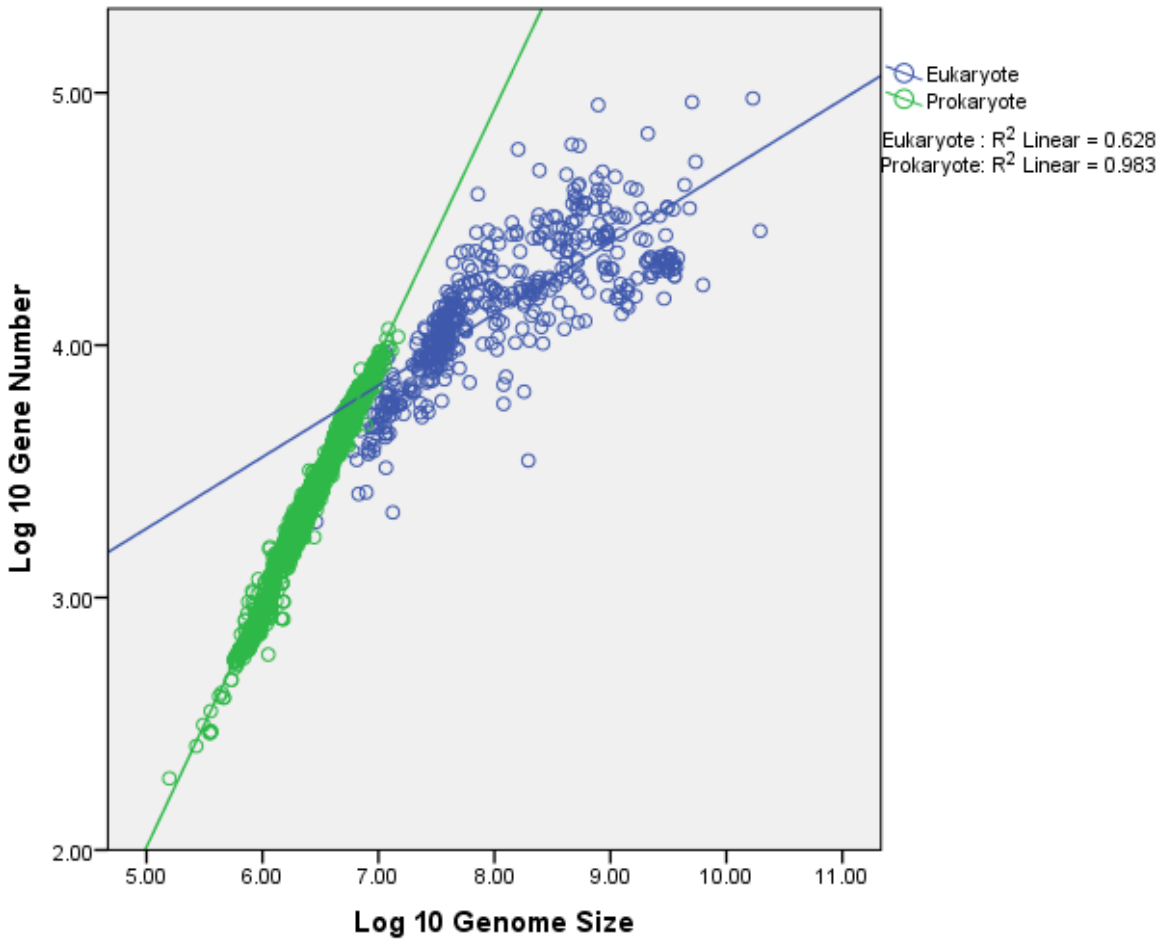


Figure 6.2 Genome size and gene number for sequenced genomes. Prokaryote and eukaryote genomes show clearly different trends, with eukaryotes having more genetic material per gene than prokaryotes. This potentially allows for more neutral, or conditionally neutral, mutations in non-coding DNA. Data from Elliott & Gregory 2015 and collated from NCBI Genomes

One feature of the genome that has previously been suggested to correlate with genomic complexity is epistasis, where the effects of genetic variants depend on genetic background (Sanjuán & Elena, 2006). RNA viruses display antagonistic epistasis, bacteria have minimal epistatic effects and multicellular eukaryotes show a trend towards synergistic epistasis. The trend towards greater effect of additional mutation in more complex organisms is suggested to be due to break-down of stability-conferring mechanisms in these mutationally-robust systems with additional mutations (Sanjuán & Elena, 2006). This property of mutational robustness may also undergird the greater possibility for the accumulation of neutral or nearly-neutral mutations in these organisms. Eukaryotic proteins are also longer on average than prokaryotic proteins (Brocchieri & Karlin, 2005). This may contribute to

a lesser effect of mutations in eukaryotic genomes, increasing the proportion of mutations that are conditionally neutral.

Discussion of mutation rate, genome size, and other genomic factors bear only indirectly on the proximate causes of trade-offs, variants which affect particular cellular systems. The relationships between genomic and cellular complexity and the bases of trade-offs are not entirely clear, but we can make some suggestions. More complex cellular systems may have an increased proportion of the genome susceptible to conditionally neutral mutations, for instance if these systems are more robust. In addition, more complex cell types may have more redundancy amongst sub-cellular systems. Within bacteria there is less redundancy in genomes that have undergone genome reduction, and there is more genetic redundancy in eukaryotes than prokaryotes (Mendonça et al, 2011). The details of the functional systems involved is an important area that is rarely explored but deserves more attention in discussions of trade-off. Examples in yeast and bacteria have been rigorously demonstrated. Kvitek & Sherlock (2013) found that trade-offs between a constant environment with a predictable nutrient supply and starvation conditions with unpredictable nutrients were due to loss of signalling networks. In the bacterium *Escherichia coli*, trade-offs in fitness between two different antibiotic environments were due to a loss of proton-motive force (Lazar et al, 2014). Specifically, this occurred as adaptation to the presence of aminoglycosides is commonly associated with cell membrane altering mutations which result in a loss of proton motive force. These mutations can increase susceptibility to other antibiotics, including classes inhibiting DNA and protein synthesis, by reducing the activity of transporter proteins which are typically involved in the development of resistance but require a proton-motive force. We suggest that microbial cells due to demonstrating less functional redundancy are more prone to such trade-offs than multicellular eukaryotes.

6.5 Mutation Rates and Types of Mutation

All else remaining equal, a higher mutation rate should contribute to a greater extent of trade-offs in alternate environments due to mutation accumulation. Over all domains of life, including RNA viruses, DNA viruses, microbes and multicellular eukaryotes, genome size and mutation rate per nucleotide covary, with an inverse relationship that has become known as 'Drake's rule' (Bradwell et al, 2013). Drake (1998) suggested that the total mutation rate per genome per replication was approximately constant at 0.003 across taxa, when only considering 'effective' genomes, i.e. the coding region. However, the non-coding region of the genome is commonly relevant to function, as shown by the ENCODE project's functional annotation of large swathes of the human genome (Stamatoyannopoulos, 2012) and many studies showing signatures of natural selection outside of coding DNA (e.g. Smith et al, 2013). Recent data shows that larger genome sizes are fairly

consistently associated with higher mutation rates per genome per replication. (**Figure 6.3**) While the trend has some important exceptions – for instance, RNA viruses, not included in the data set used, can have very high mutation rates (Drake et al 1998) – the mutation rate per genome per generation is generally higher in more complex organisms. An example supporting this is given by work from Keightley & Eyre-Walker (2000), showing that within various animals, organisms with longer generation times had larger genome point mutation rates (deleterious mutations).

Figure 6.3. Larger genomes have a higher mutation rate per genome, in the range of genome sizes relevant to microbes.
Data from Sung et al, 2012.

It is well established from adaptive laboratory evolution experiments with haploid organisms that many mutations fall within the 'loss of function' (LOF) category. It has been argued (Behe, 2010) that

most adaptive mutations result in the loss or degradation of functional elements. Adaptation often proceeds by populations fixing these loss of function mutations. It appears that most LOF mutations in diploids will be recessive; if this is the case, then such mutations will result in an actual loss of function effect more often in haploids than in diploids (Raynes & Sniegowski, 2014). Haploids should therefore experience more large-effect LOF mutations than diploids. A preponderance of LOF variants may contribute to the high occurrence of antagonistic pleiotropy in haploids, when the loss of function proves costly in alternate environments.

Loss of function mutations are perhaps particularly important in haploid genomes (as discussed here in Chapter 3, as well). As one example, a recent experiment by Lazar et al (2014) found 27% of mutations in *E. coli* evolved in antibiotic conditions resulted in a truncated protein, i.e. a clear loss of function effect. Lind et al (2015) have proposed that adaptation generally proceeds through a series of three different classes of variant; "evolution proceeds firstly via pathways subject to negative regulation, then via promoter mutations and gene fusions, and finally via activation by intragenic gain-of-function mutations." In this work, the pathways under negative regulation were affected by loss of function mutations, as the first step in adaptation, and this may be a common strategy – but perhaps only in haploid genomes.

Polyploidy is an extreme case of the more general phenomenon of the potential functional redundancy of gene products following a duplication event. Another way of measuring biological complexity is in terms of functional redundancy of parts. Complexity can be added by adding redundancy to a system, and this can increase possibilities for neutral mutations. As Nei (2007) notes, "If a character is controlled by a large number of interacting genes, it is possible that the genetic networks involved are robust and resistant to the effects of deleterious mutations". The effects of beneficial mutations may also be less, resulting in many mutations being neutral or close to neutral, altering the kinds of processes that tend to lead to their fixation. A similar effect has been modelled in artificial life-like networks, for instance mutation accumulation was observed as a cause of ecological specialization in digital organisms (Ostrowski et al 2007), and in another artificial network, Valverde and Solé (2012) note that "the more complex organisms were also more robust against the effect of mutations than the simpler ones."

Microbes and multicellular eukaryotes differ in typical genome structure and means of replication. The structure of eukaryotic chromosomes is not able to be differentiated from that of prokaryotic chromosomes with complete precision (Bendich & Drlica, 2000); for instance there are a number of bacteria with linear chromosomes (Galperin, 2007). However, there are still numerous differences between prokaryotic and eukaryotic chromosomes in general, including their typical shape and

number, and the roles of plasmids and centromeres (Kuzminov, 2014). The relationship between chromosomal structure and the accumulation of genetic variation has not been studied in depth to our knowledge, but differences in reproduction between eukaryotes and prokaryotes have been. Prokaryotic sexual reproduction differs widely from the processes of meiosis and syngamy in eukaryotes. We suggest that sexual reproduction, which is nearly ubiquitous across eukaryotic lineages (Burt, 2000), reduces the prevalence of chromosomal rearrangements. In general, the processes of segregation and recombination of chromosomes involved in meiosis function to reduce non-SNP genetic variation (Gorelick & Heng, 2011). Chromosomal variation has been observed in experimental evolution of *Escherichia coli* (Raeside et al, 2014) and found to contribute to local adaptation in the monkey flower *Mimulus guttatus* (Lowry & Willis, 2010). It may be that this scale of variation, contributing to reproductive isolation, is limited in sexually reproducing eukaryotes. It is also possible that the relative lack of mutator phenotypes in eukaryotes is due, in sexually reproducing populations, to recombination separating out the causative variants (Ness et al, 2015). We have argued elsewhere that sexual reproduction in eukaryotes serves to remove an important contributor to antagonistic pleiotropy across environments (Ardern & Goddard, 2015).

6.6 Pleiotropy, and a Cost of Complexity

There are reasons to expect an evolutionary 'cost of complexity', i.e. a higher cost for mutations in more phenotypically complex organisms (Orr, 2000). If this is a general trend, then mutations that do pass the initial filter of the current environment will be smaller in complex organisms. It may also be that increasingly complex epigenetic programs contribute to such a cost of complexity (Huang, 2009). At least in the case of large populations, there is evidence that haploids adapt faster than diploids (Zeyl et al, 2003). It may be that more rapid adaptation in one environment leads to more trade-offs in alternate environments, similarly to how, with an extension of Fisher's Geometric Model to two environments, we expect to see more trade-offs with large effect mutations beneficial in one environment than with small effect mutations (Martin & Lenormand, 2015; Ardern & Goddard, 2015). As such, we have additional reason beyond that already discussed to expect more trade-offs in haploid than diploid organisms.

Additional Factors Associated with Greater Genomic Complexity

A potentially confounding factor is that antagonistic pleiotropy may be more difficult to demonstrate in larger genomes, both due to greater difficulty in measuring selection in organisms with longer lifespans and difficulty in separating out the effect of any single variant in populations with large genomes and small populations. For a rigorous inference to this mechanism, a significant fitness effect of a variant needs to be shown across more than one environment. However, if not detectable, the potential evolutionary relevance of such effects in the typically small populations of higher

eukaryotes is also questionable.

It is also plausible that both cryptic variation and phenotypic plasticity increase in organisms with more complex genomes, contributing to the difference in the processes undergirding trade-offs across environments. Cryptic variation is variation that is only phenotypically apparent in some environments (Paaby & Rockman, 2014). We suggest that such variation is more common with greater genomic complexity; for instance, more genes and traits allow for more areas of the genome which are not subject to selection at any given time. In general, weakened selection facilitates the accumulation of cryptic variation (Masel, 2005). If it is also the case that variations in the genomic locations responsible for this cryptic variation are conditionally neutral, i.e. only manifest deleterious fitness effects in particular environments, then cryptic variation will contribute to trade-offs through mutation accumulation. A similar phenomenon is phenotypic plasticity, which is recognised in microbes, but discussion of which has particularly focussed on multicellular eukaryotes. Plasticity is the capacity to produce different trait values in different environments. If opportunities for plasticity of expression do increase with genomic complexity, this may decrease the deleterious effects in alternative environments of adaptive variants.

6.7 Conclusion

We have argued that a number of genomic features associated with greater complexity are likely to facilitate the accumulation of conditionally neutral/deleterious mutations, resulting in fitness trade-offs in alternate environments through mutation accumulation. Similarly, features of 'simpler' genomes predispose them towards exhibiting trade-offs due to antagonistic pleiotropy between variants. Overall, more complex genomes are likely to develop trade-offs due to variants accrued via neutral rather than selective processes. We predict that as high-throughput DNA sequencing techniques become more accessible, quantifying the environment-pleiotropic effects of neutrally accumulated and selected mutations under different circumstances will be a growing area of research. Better understanding genome-level contributors to trade-off is important in many areas in which genetic knowledge is applied, including evaluating risks from the evolution of drug resistance in pathogens and pest species, and potential challenges from shifts in environmental conditions due to climate change. Future work must test the role of various contributors to trade-offs across environments, in order to avoid or utilise them in different contexts. In the era of easily accessible genomic and increasingly post-genomic data and high-throughput assessment of phenotype, there are opportunities for rigorous experimental testing of these evolutionary hypotheses, to begin to disentangle the relative contribution of natural selection, stochastic processes, and structural constraints to shifts in biological function.

Chapter 7
Investigating Trade-offs in Sexual Populations with Gene-flow

Chapter 8
Genotype-by-Environment Interactions, Antibiotic Resistance, and Experimental Evolution

The evolution and spread of resistance to antibiotic drugs are widely acknowledged as major areas of concern for human health. The de novo development of resistance following exposure to these drugs is also commonly cited as a prime example of evolution through natural selection observable over short time spans. The process has been studied through experimental evolution, and many insights into the molecular events underlying the phenomenon of resistance have been gained. This article highlights insights which this research and the development of associated theory can provide regarding evolutionary costs of resistance. Such costs of resistance are an instance of the broader evolutionary phenomenon of trade-off between environments, and recent work on trade-offs enables predictions about the evolution of bacterial resistance to antibiotics.

8.1 Costs of Adaptation and Costs of Resistance

The phenotypic effect of new genotypic variants can differ widely across the environments to which an organism may be exposed. These phenotypic effects can translate into fitness differences. Genotype x environment interactions can include 'trade-offs', where a population's adaptation to one environment is associated with decreased adeptness in an alternate environment when compared to the non-adapted ancestor.

The development of antibiotic resistance is usually associated with a fitness cost in the absence of the drug, and higher levels of resistance are correlated with higher fitness costs (Melnyk et al., 2014). Many examples of the cost of resistance have been demonstrated through experimental evolution of resistance, with various cellular targets, involved for instance in RNA or DNA replication or in protein synthesis, shown to be altered in the presence of antibiotics with an effect of increased fitness (Melnyk et al, 2014). While most attention in molecular evolution studies tends to be focused on single nucleotide polymorphisms, copy number variants are also important to the development of resistance. These variants too, under various environmental conditions, have been shown to frequently be associated with a fitness cost (Tang & Amon, 2013).

However, subsequent to initial costs, 'compensatory' mutation, reducing fitness cost, has been demonstrated in many instances of resistance, including for instance in a clinical isolate of Mycobacterium tuberculosis (Meftahi et al, 2016). The effects of compensatory mutations have been demonstrated both *in vitro* and *in vivo* (Hall et al., 2015). Possible compensatory mechanisms 'rescuing' fitness include a subsequent mutation within the affected gene, a mutation increasing gene dosage, a mutation in in an alternate gene such as a neighbouring molecule in a protein complex, or a bypass mechanism such as compensation by a regulatory factor (Hughes & Andersson, 2015).

Compensatory mutations can refer to mitigating costs in the presence of the antibiotic or to costs in alternate environments, particularly in the absence of the antibiotic. It is possible, indeed seems likely, that compensatory mutations will carry hidden costs for the bacterium in alternate environments, a topic which has received limited study although Björkman et al (2000) have shown that the compensatory mutations fixed depends on environmental conditions. Similarly, Maclean & Vogwill (2015) discuss the relative lack of compensatory mutation in clinical versus laboratory culture studies testing for loss of resistance, suggesting that there are costs associated with compensatory mutations not typically captured in laboratory studies. Mutations labelled as 'compensatory' in a clinical setting can also, alternatively, increase fitness by increasing resistance to the drug (Vestergaard et al., 2016). The physiological effect on bacterial cells of compensatory mutations is thus worthy of study, and the specific environment in which there is a 'compensatory' effect needs to be kept in mind when considering their potential relevance to limiting the spread of resistance.

Some resistance mutations do not result in obvious fitness costs, they are 'no cost' mutations; a number of examples are given in a recent meta-analysis of costs of resistance (Melnyk et al, 2014). In some cases, then, compensatory mutations are not needed for cost-free resistance. One estimate suggests that in prokaryotes there are an average of approximately 13 possible compensatory mutations for each deleterious mutation (Poon et al, 2005), and later work shows that this makes mutations which compensate for the functional defect involved in resistance rather rare compared to mutations that are 'generally beneficial' i.e. which are adaptive regardless of the resistance (Qi et al., 2016). It has been shown in baker's yeast *Saccharomyces cerevisiae* that compensatory mutations subsequent to gene deletion are more likely to be fixed after very costly rather than low-cost resistance mutations (Szamecz et al, 2014). This may be partly due to an increased proportion of mutations having compensatory effect subsequent to larger fitness losses. If compensatory mutants are less fit than wildtype reversion, as is presumably typical, fixation of these mutants is possible in small populations, but a greater incidence of compensatory over reversion mutants is required in order for it to be expected. Poon (2005) mention, for instance, that the 8% fitness cost of

compensatory compared to reversion mutations observed in streptomycin resistant *E coli* by Levin et al (2000) means that a 10-fold greater incidence of compensatory mutations is required for them to fix.

In light of all of these things, further study is needed on 'no cost' resistance mutations, compensatory mutations, and the distribution of resistance mutation fitness effects across environments – both in vitro and in vivo, and both with and without antibiotics.

8.2 Complexity of Environments and Trade-off

In a more 'complex' multifaceted environment, we might expect that more phenotypes will become relevant to fitness as more cellular functions become useful to survival. As such, much as Maclean and Vogwill (2015) have suggested that researchers "overestimate the efficacy" of compensatory mutations, which are found to be more costly in vivo than in vitro, it is reasonable to expect costs of resistance to be higher in vivo than in vitro. Increasing the complexity of the environment can be seen as equivalent to increasing the complexity of the phenotypic space considered. Within the core of Fisher's Geometric Model, this is the only measure of environmental complexity – the number, n, of phenotypes which contribute to fitness in the environment considered. The differences between laboratory culture and an infection in a host are commonly remarked on. Realistic infection scenarios, for instance, are likely to involve spatially structured environments, rather than the well-mixed ones of lab evolution (Hall et al, 2015). In general, it seems reasonable to assume greater environmental complexity in human hosts than in flasks.

8.3 Cross-sensitivity and Cross-resistance

In studying adaptation and trade-offs in antibiotic-resistant pathogenic bacteria the environments of interest are different host conditions, in particular the presence or absence of one or multiple drugs. 'Cross-resistance' occurs when variants causing adaptation to the presence of one drug and thus conferring resistance, are associated with increased resistance to other drugs. This can happen when the same cellular mechanisms are involved in resistance to multiple drugs. Alternatively however, and more helpful in the fight against resistance, 'cross-sensitivity' can also occur, when adaptation to one drug is associated with a cost of resistance in the presence of another drug or drug combination. The best studied example is the relationship between resistance to aminoglycoside

antibiotics and some other classes of antibiotics. Resistance to aminoglycosides commonly is achieved through membrane-altering mutations that reduce the proton motive force required for uptake of these drugs into the cell. This has the side effect of also decreasing some efflux pumps dependent on this force, which are important in resistance to other classes of antibiotic. Resistance to other drug classes is then often hindered by developing resistance to aminoglycosides (Pál et al 2015).

Furthermore, the relationship a population demonstrates between resistance to different drugs can depend upon the order in which the drugs are administered; the response can be nonreciprocal (Richardson, 2015). Detailed knowledge of fitness landscapes allows for modelling of evolutionary trajectories, and predicting the best order in which to administer drugs in order to prevent the rise of resistance, with an early demonstration of this in a recent study analysing fitness landscapes in *E. coli* administered with *β*-lactam antibiotics (Nichol et al, 2015). This particular approach requires detailed fitness landscapes, which may not be feasible in other systems with many different mutations and epistatic relationships among them. Comprehensive landscapes of evolution over longer time periods may not be feasible, but perhaps generalisations will be able to be made about responses to antibiotic selection on the basis of a few such in-depth studies of local fitness landscapes.

8.4 Fisher's Geometric Model and Environment

A more general representation of short-term evolutionary change is Fisher's Geometric Model. As originally formulated this is a model of phenotypes under selection, where an individual organism is represented as a point in a multidimensional space, shifting through this space according to phenotypic effects of mutations accumulated. There is much support for the applicability of Fisher's Geometric Model to evolution in microbial populations. A recent review paper lists some of these findings: "Fisher's model successfully predicts the distribution of selective coefficient of random mutations (Martin et al. 2006), levels of epistasis (Martin et al. 2007, Gros et al. 2009, Rokyta et al. 2011), levels of dominance (Manna et al. 2011), and the drift load (Tenaillon et al. 2007, Gros and Tenaillon 2009)." (Blanquart et al., 2014). One reason that this model has become more popular in recent years is that genetic sequencing studies have shown the large number of mutations of small effect relevant to fitness, justifying its use of a continuous rather than discrete adaptive landscape (Tenaillon, 2014). The model has recently been extended to multiple environments by Guillaume Martin (Martin, 2015), and a simplified form of this development can be used to illustrate trade-offs. Considering an extention with two fitness optima allows prediction of the effect of a mutation in an alternate environment as well as the environment of selection (Kassen, 2014; Ardern & Goddard,

2015 – i.e. Chapter 7 of this thesis).

The basic model relates the probability of a beneficial mutation to the size of the mutation, the distance to the point of optimum fitness, and the number of phenotypes relevant to fitness. The complexity parameter 'n' has been discussed in relation to organismal complexity. For instance, it has been argued that there may be an evolutionary cost of complexity (Orr, 2000). The suggestion from the model has not itself been challenged, but its application to organismal complexity has been. Subsequent empirical studies on pleiotropy (the numbers of phenotypes affected by a gene) through deletion mutants have shown that pleiotropy is not high in complex organisms as genes typically affect few traits, and so perceived organismal complexity does not have the anticipated limiting effect on adaptation (Wagner et al, 2008; Wang et al, 2010). The extent of pleiotropy, however, has been a matter of debate; whether pleiotropy has been under-estimated (Paaby & Rockman, 2013) or not (Zhang & Wagner, 2013). The particular contribution of environmental complexity seems to have received less attention, but the number of phenotypes contributing to fitness is potentially highly dependent on the environment – it is a result of genotype-environment as well as intra-genomic interactions.

8.5 Population Genomes

Another factor in antibiotic resistance evolution that has received relatively little attention is the potential contribution of non-fixed mutations to phenotypes. Experimental microbial populations with selective pressure frequently, and more often than expected, fail to fix a single genotype (Burke, 2012). There are a number of reasons why populations may retain high levels of allelic diversity. These include clonal interference, frequency-dependent adaptation, and spatially structured microcosms. Clonal interference occurs when multiple clones exist within a population simultaneously, and the inability of asexual populations to recombine beneficial mutations helps ensure that multiple clones are retained for many generations; selection is slowed. The most comprehensive study of this to date was conducted in haploid yeast populations of varying sizes, with sequencing data obtained at multiple time-points (Lang et al, 2013). Frequency-dependent effects are closely related, and may help to maintain different genotypes within a population. An example is cross-feeding behaviour, where one variant in a population produces a metabolite, perhaps as a byproduct of its own metabolism, which another then specialises to grow on. This occurred in the Lenski long-term evolution experiment in E coli after the evolution of a subpopulation that could metabolise citrate (Cit+), facilitating the survival of Cit- cells without that trait (Turner et al, 2015). Spatially structured microcosms, as virtually all real-world environments must be, may similarly select for multiple phenotypes, allowing multiple lineages in what may appear a single microbial population.

The genomic foundation upon which evolution proceeds is a key aspect of the dynamics of population genomes. This genomic background has been shown to be important to evolutionary outcome; different strains evolve differently (Vogwill et al, 2014). The impact of epistasis, compensatory mutations, and regulatory mechanisms influencing gene expression in the development of antibiotic resistance have all been studied in recent years (Hall et al, 2015), and the role of each of these across different environmental conditions deserves further study.

Chapter 9

Conclusion: Sex, Gene-Flow, Mutation, and Trade-offs

9.1 In Hindsight: Different Methods

If the experiments were to be repeated, here are some things that I would do differently in hindsight, or that are worthy of consideration in future experimental design. Methods discussed include the use of the chemostat in growing populations, measuring fitness with phenotypic microarrays, counting cells, extraction of DNA, sequencing population DNA, and the bioinformatics pipeline for sequence analysis.

Sequencing Populations:

Two ways to sequence populations using 'next generation' sequencing with millions of short reads are either to pool the DNA from a representative sample of a population and sequence this to a reasonably high depth (as in the experiment reported here in Chapter 3), or to sequence a few individuals at a lower depth and compile the results together; i.e. 'pooled whole-population' or 'compiled individuals'. Both allow for the estimation of allele frequencies if there is high enough fold coverage of the population or enough sequences of individuals are compiled. The pooled whole-population method has the significant disadvantage of not allowing whole genome genotypes of individuals to be inferred. That is to say, there is no way to determine from such data whether any two variants are found within the same individual genome unless either the variants are both found in more than 50% of the population or the variants are very close together in the same chromosome region and happen to be observed in the exact same short read. The question of the nature of individual genotypes is perhaps a particularly important question when considering differences between asexually and sexually reproducing populations. It would have been interesting to find out whether recombination had occurred such that populations in the full gene-flow conditions showed 'mixed genomes'.

The lack of recombination in asexually reproducing populations means that it is easier for neutral or even maladaptive variants to 'hitch-hike' along with adaptive ones, whereas it is presumed that adaptive variants in sexually reproducing populations would be more 'isolated' insofar as fewer other variants should be pulled to higher frequency in a population by being associated with them –

knowledge of individual genotypes would allow fuller comparison of evolution in sexual and asexual populations.

It is difficult to infer actual allele freq with either the 'pooled whole population' or 'compiled individual sequences' method. Compiled individual sequences may give a more rigorous estimate of allele frequency. As suggested above, a statistical analysis of these approaches in future, to determine the read depth and/or number of individual sequences required for a good estimate of allele frequency under each method would be very useful for researchers in the field.

Regrowth for DNA Extraction

In order to extract DNA from the frozen samples stored in glycerol at -80°C and for later fitness testing (for instance testing for trade-offs in an alternate environment), the stocks must be regrown in 'stress-free' media – for this purpose, YPD was chosen, to maximise cell growth. This process will influence the distribution of the alleles in the population – and how much is not known; evidence available on this is currently ambiguous. It may be that some key alleles favoured in a selective stressful environment are selected against in the 'stress free' medium, and that proportions of various alleles in the populations change through this treatment. While this condition is applied to all populations, it is not ideal as it is likely to at least decrease the frequencies of adaptive alleles in the population to some extent.

If I were repeating the whole experiment I would attempt to minimise the need for population regrowth. Here is a brief assessment of this issue:

DNA extraction for Illumina Hi-Seq library prep requires 1.5ug of genomic DNA and each cell contains a minimum of approx. 1×10^{-15} grams of DNA, meaning that at least 1.5×10^{9} cells are required.

The cultures subsequent to adaptation had a cell density of at least 1×10^{7} colony forming units per mL (Gray, 2011), so assuming this value as the minimal density, 150mL of one of these cultures would be required for adequate extraction without regrowth, clearly far too large a requirement, given total chemostat volumes of approx. 40mL in this experiment. In actuality 1 mL of the outflow at each time-point was retained for creation of stock, i.e. approx. 10^{7} cells. As such, even if the whole stock was used a population expansion of at least 150 fold, greater than 7 population doublings, was required for sufficient DNA to be extracted for sequencing. If excess cells are produced to guarantee a high yield of DNA, then more doublings occur. For DNA extraction in the experiment reported in Chapter 3, cells were regrown from 100uL of stock (10^{6} cells) in 20mL of

YPD, which can sustain a higher cell density than the selective media, of approx. $2x10^8$ cells per mL (Gray, 2011), a total of $4x10^9$ cells if the YPD culture achieved saturation, equivalent to 12 doublings. After 12 population doublings an allele that begins with a frequency of 90% and decreases in relative population size by 5% each generation (i.e. has a fairly sizable fitness cost) will reduce in frequency to less than 50% of the population.

This size of fitness cost is within the range of what is possible in these kinds of experiments, but any actual fitness costs for most alleles in YPD are likely to be much smaller. For instance, adaptation to galactose in yeast was shown to increase in their maximum specific growth rate by +0.04/hr, and this adaptation lead to a decreased maximum specific growth rate of -0.02/hr in an environment where glucose was the carbon source (Hong and Nielson, 2013). The maximum specific growth rate of yeast is approximately 0.4/hr (van Hoek et al., 1998), so these are large effects on growth rate. More directly relevant to this experiment, adaptation to carbon limited conditions has been shown to result in a trade-off under carbon sufficiency (Wenger et al., 2011). When tested in chemostat conditions that were both limited for nitrogen and sufficient in carbon, decreases in the competition coefficient to below 0.8 were observed for some evolved clones. Importantly however, growth in glucose-rich batch culture showed only very minor changes in competition coefficient. When the effect of the HXT6/7 duplication by itself was assessed, small but significant differences in specific growth rate were observed. (See Figure S3 in Wenger et al., 2011). Changes in allele frequency under regrowth conditions is particularly likely to be problematic if the basis of fitness trade-offs between environments in this experimental system is predominantly antagonistic pleiotropy, as I argue in chapter 5. If variants adaptive under selective conditions are maladaptive in at least some other environments, then it is reasonable to expect some will be maladaptive in YPD regrowth conditions and so the need for regrowth should be minimised where possible.

Minimising population regrowth would involve storing larger samples, and/or choosing a media for adaptation that allows for larger population sizes. If the goal is $2x10^9$ cells for DNA extraction, this can be achieved under different conditions, such as media which allows a larger population size. For instance, samples collected from 10mL of a media with a capacity of $2.5*10^7$ cells would require only 3 population doublings. Preferably such doublings could be conducted under the same chemostat conditions as the adaptive evolution, or conditions approximating these, rather than in 'batch culture like' YPD regrowth.

Phenotypic Micro-Arrays

As reported in Chapter 4, phenotypic microarrays (BIOLOG, Inc. PM1 and PM3A 96 well plates) were used to assay phenotypes. Specifically, they were used in order to determine some of the

phenotypic effects across environments associated with previously measured fitness estimates in populations of sexually and asexually reproducing populations of *S. cerevisiae*, with or without a 'mutator' genotype. Tetrazolium dye (Biolog™ dye mix D) is added to the plates. The optical density of wells in these plates, increasing with greater reduction of the dye, is used to measure metabolic activity, with a sample of the population. NADH produced by the cell during respiration reduces the dye, causing a colour change to purple with high cellular respiration. Measuring cellular respiration in this way is not a direct measure of a population's evolutionary fitness or even the growth rate, although a recent study of wild yeast samples by Samani and Bell (2015) using Biolog™ plates showed a high correlation of R^2=0.76 between optical density after 72 hours of growth with and without the addition of the tetrazolium dye. This was taken to show that optical density following the addition of the dye is an accurate measure of cellular yield.

An alternative would have been to measure optical density in the wells without the addition of the tetrazolium dye; from this, population growth rates and hence evolutionary fitness could be more directly inferred. Growth on plates has previously been measured and compared to growth in chemostat conditions, and the plates have been determined to be a poor proxy for fitness in chemostat-evolved populations (Gray, 2011), but this is likely less of a problem when measuring evolutionary fitness of cultures evolved through serial transfers in batch culture, as in the 'sex and mutator' experiment assessed with the phenotypic microarray plates. Both population growth rates and metabolism under different conditions are aspects of microbial cellular phenotype which are potentially affected by adaptation. In hindsight I would have chosen to assay growth rate rather than metabolism, with the condition that it could first be shown that optical density in such plates is a good way to measure cell number and that evolution under the conditions assessed does not change the optical properties of the populations enough to significantly affect the results.

Counting Cells

The current method to determine fitness in the Gray-Goddard system tests competitive fitness with the use of replica plating (Gray, 2011). This involves spreading a diluted sample of *S. cerevisiae* cells, following competition between samples of the ancestral population and the population derived through experimental evolution, and counting the relative number of each cell type. One strain has an antibiotic resistance marker and the other does not (e.g. ancestral asexual population with kanamycin resistance gene inserted, and derived population without), so that comparing spread plates with and without the antibiotic allows for estimations of the competitive fitness of the derived population. The results from this method are inexact, particularly if conducted by researchers inexperienced at replica plating. The process of diluting the sample to achieve an appropriate number of cells on the plate is time-intensive, and the initial proportions of each strain to be used

for the competition experiment (required to be changed from the standard 1:1 ratio where derived fitness is high), so as to get an adequate number of cells of each strain on the plates at the end of the competition, are not always easy to determine.

As such, to facilitate future work in this area, methods for counting yeast cells more easily and accurately should be investigated. Flow cytometry is one approach that has been used for a wide range of applications. Cells stained with a fluorescent dye and held in suspension are made to pass sequentially through a high intensity beam of light, and the resulting fluorescence and light scattering are measured. Mixed populations (e.g. ancestor and derived) can be differentially stained and the proportion of each type compared. Many applications are discussed in the review by Kron et al. (2007), which includes a short section on the then-recent application to microbial experimental evolution. One important project in which this technique was used, with haploid yeast populations engineered to express the fluorescent protein mCherry was the pioneering study by Lang et al. (2013) tracking 40 populations over 1000 generations of evolution. One option to investigate in future for small-scale solutions may be collaboration with physicists on spectrophotometry and flow cytometry techniques.

Another approach to counting is sequence 'barcoding', using nucleotide sequence labels on strains and deep sequencing of populations at these sites to track allele frequencies. An example of this type of study was conducted by Payen et al. (2015). Thousands of strains with a single copy of a gene overexpressed or deleted were tagged with unique barcodes, and competed against each other for 20 generations, sequencing samples (involving amplification, with specific primers) of the appropriate barcode sites every 3 generations. This method could be used for instance in competitions between ancestral and derived genomes and should be compared to flow cytometry as a counting option.

Chemostat Failures

Gathering more data on the evolutionary fitness of the Gray-Goddard 'sex and gene-flow' populations was originally intended. Originally the Malthusian fitnesses of the populations with intermediate gene-flow were considered of high interest and attempts were made to measure these. However as the project developed it became clear that of more relevance to this project would be fitnesses measured in the alternate environment, for samples of the populations from time points prior to the 'end point' of 300 generations. This would allow more fine-grained time-course data on trade-offs, which would give a clearer picture of whether antagonistically pleiotropic fixation events occur. Combining the multi-time-point fitness data with sequence data at multiple points may allow for observing the particular alleles responsible for changes in fitness in each environment.

Measuring fitnesses using the chemostat system was attempted at an early stage during this project, however technical difficulties prevented enough useful data from being produced. The chemostat system requires careful attention to a number of factors over the course of the experiment, including flow rates, and ensuring fittings are sterile. The particular difficulties observed were difficulties with the pumps, blocked or severed tubing, and repeated contamination by other fungi and bacteria. Particular attention to the tubing – ensuring no splits or blockages – and a more easily sterilisable system for introducing media to the chemostats would have probably prevented most issues that did occur in the attempts at measuring fitness in this system that were conducted.

Bioinformatics Pipeline

The development of a workflow for analysing the whole genome sequence data presented in Chapter 3, with the final pipeline script in Appendix B, would have benefitted from being better informed in the initial stages.

Bioinformatics training in the use of R, and the ability to write scripts in Python or Perl would have greatly accelerated the process of creating the pipeline and resulted in a more efficient set of transformations of the sequence data. In hindsight using Bioconductor in R as the basis of a similar pipeline would probably be more efficient and allow for easier customisation and greater flexibility in analyses available.

9.2 Future work

Future work with Gray/Goddard populations:

Future work expanding on the results reported in Chapters 3 and 4:

Firstly, more trade-off fitness data (i.e. data on evolutionary fitness of derived populations in the environment in which they were not evolved) for both experiments could be obtained. For instance, for the 'sex and gene-flow' experiment, trade-off fitness data for those populations with different extents of gene flow (migration rates of 10^{-2}, 10^{-4}, and 10^{-6}) may be informative.

Additional sequence data for this sex and gene-flow experiment, obtained from samples at multiple time points, and confirmation of copy number variants reported in chapter 3, would also allow other questions to be answered.

For the 'mutator' experiment, actual trade-off fitness data (rather than merely using phenotypic microarrays to measure metabolism) would help give a more complete picture of the effect of adaptation on evolutionary fitness in alternate environments. In addition, sequence data for this experiment would show the actual mutations occurring in sexual and asexual populations with and without a mutator genotype.

Future Work Utilising Experimental Evolution Sequence Data:

Understanding population level effects in adaptation is I think a very important field of study. Frequency-dependent effects in populations are one example of this, along with clonal interference. In general, it will in future be possible to conduct informative meta-analyses of 'evolve and resequence' microbial experimental evolution studies' sequence data, to begin to make generalisations about molecular evolution in experimental contexts.

For instance, the different types of genetic mutations which increase in frequency in microbial populations under different conditions would be of significant interest. Are 'Loss of Function' mutations as prevalent as some have argued? (e.g. Behe, 2010). What is the relative contribution of copy number variants, point mutations, indels, and other chromosomal rearrangements to genetic diversity in these evolving populations? There has been increasing recognition of non-point mutations in general – see for instance Payen et al. (2013), on segmental amplifications in *S. cerevisiae* in sulfate-limited conditions. There is now the opportunity for this insight to be explored further in experimental evolution studies. A recently published study of sexually and asexually reproducing experimental populations of *S. cerevisiae* (McDonald et al., 2016) has some similarity to the work reported in this thesis but non-point-mutations only get a brief mention. Previous major studies such as Lang et al. (2013) have typically not mentioned gene amplification or other genome rearrangements.

Is there a typical sequence of mutation types observed in adaptation? Exploring this question could expand on the observations of Lind et al. (2015) on the sequence of mutation types observed in experiments with *P. aeruginosa.* This has important implications in understanding evolutionary responses to new environments. It may be, for instance, that drugs altering selective pressure so as to not favour commonly selected mutations may limit an adaptive response, for instance the adaptation of pathogenic bacteria to antibiotic environments.

Finally, future work will be better able to answer key questions touched on here with the aid of increased statistical power. This would be assisted by increased automation of batch culture (e.g. the

use of liquid transfer robots, as mentioned in Bell, 2016) and/or chemostat experiments. Chemostats are, however, difficult to automate or run for large numbers of independent culture environments. Incorporating transcriptomics data and/or other measures relating to phenotype such as phenotypic microarray data would provide a more holistic view of microbial evolutionary change.

9.3 Summary

The evolution of sex is one of the classic difficult questions of evolutionary biology. In light of the research presented here it is proposed as a hypothesis for future work that sexual reproduction is of particular advantage in complex environments, as it facilitates selection against antagonistically pleiotropic mutations, facilitating generalism, i.e. high fitness across multiple environments.

Sex has previously been shown to speed adaptation, and the molecular basis for this has begun to be explored, but this is the first study to directly compare nearly isogenic strains of sexual and asexual organisms with the same ploidy.

This is also the first study to compare metabolism in nearly isogenic sexual and asexual evolving populations, and has shown that some differences in metabolism are a repeatable result of evolutionary processes, and demonstrate high correlation with sexual status and mutation rate.

Following significant research, suggestions have been made for a newly developing 'best practice' in microbial experimental evolution going forward – of particular interest are techniques for high throughput screening of fitness, that may further increase the data available for analysis. Effective computer programs for the analysis of the flood of data appearing, which are able to capture variants not of so much traditional interest as single nucleotide changes (SNPs), such as copy number variants, will also be important. The analysis of multiple individual clones which is able to provide insights on whole genotypes of individuals will have to be carefully compared with pooled population genome sequencing for microbial populations.

A number of theoretical proposals have been made, including arguing that mutation accumulation is of particularly importance in more complex genomes, and in microbes is rare outside of mutator populations. This can be tested further in future. An important next step in this area, once some more data is available, will be for comprehensive meta-analyses across studies, to find general trends, and test theories proposed here, and other claims such as found in Lind et al. (2015) of trends in the kinds of variants involved in evolutionary adaptation. If such trends exist, making use of them could have large benefits in understanding and counteracting evolutionary processes detrimental to human well-being, such as antibiotic resistance and cancer tumour growth.

Appendix A

Qiagen Yeast Lysis & DNA extraction protocol, used for the sequencing in Chapter 3.

Inoculation:

Have YPD prepared beforehand (including autoclaving) .

Sample defrosted and then homogenised/mixed by pipetting.
50mL 'falcon tube' used for growth.100uL of sample in 20mL of YPD.
Tubes inverted 3x, then vortexed briefly.
28°C warm room for 48hrs.

Counting cells: [1.5hrs]

Dilute sample from the culture – take 10uL and add 90uL of distilled water (in an eppendorf tube). Then take 2x 15uL aliquots for the haemocytometer.

Lysis: [3.5hrs]

(2.) Have prepared beforehand:
RNase A
Zymolyase
TE buffer

(3.) Pellet cells from culture – 3220g, 4°C, at 12 mins.
Discard supernatant.

(4.) Resuspend in 2mL TE buffer by vortexing.

(5.) Pellet cells by centrifuging - 3220g, 4°C, at 12 mins.
Discard supernatant and resuspend the pellet in 1mL of buffer Y1, vortexing at top speed.

(6.) Add 100uL of zymolase stock solution.
Incubate at 30°C for 60mins.

(7.) Pellet the spheroplasts by centrifuging at 3220g for 15mins.

(8.) Resuspend pellet in 2mL of pre-prepared buffer G2 withRNase A (from bottle in fridge). Mix by inverting and vortexing.

(9.) Add 45uL of proteinase K stock solution (from fridge).
Incubate for 60mins.

(10.) Pellet cellular debris by centrifuging at 3220g, 4°C, for 15mins.
Retain the supernatant in new falcon tubes.

Extraction (Genomic Tip protocol) [4hrs]:

(1.) Have prepared beforehand:
corex tubes, clean for eluting DNA into.
Genomic tip – equilibrated with 2 mL of buffer QBT; emptied by gravity flow.

(2.) Vortex the sample (clear supernatant) for 10s at maximum speed.
Apply to genomic tip.

(3.) Wash the genomic tip with 3 x1 mL of buffer QC.

(4.) Elute the DNA into clean chorex tubes with 2 x 1mL of buffer QF.

(5.) Precipitate the DNA in the chorex tubes with 1.4mL of room temperature isopropanol.

(6.) Recover the DNA by mixing and centrifuging at 10000g for 15mins, at 4°C. *Have tube marked to help see the pellet when it forms.*
Carefully remove the supernatant – into a new eppendorf tube for each sample just in case pellet is lost in this process.

(7.) Wash the centrifuged DNA pellet with 1mL of cold 70% ethanol. Vortex briefly.
Centrifuge at 10,000g for 10mins at 4°C.
Carefully remove supernatant without disturbing the pellet (as in step (6.)).
Air-dry for 10mins.
Resuspend the DNA in 130uL (? – or perhaps a little more?) of water
Freeze DNA sample.

Appendix B

Some of the scripts used for processing of whole genome data, written in Bash – for use in a Linux command terminal

SORTED BAM FILE TO SNPs:

```
for f in *.sorted.bam ;
do      f2=${f%.*}.mpileup ;
        f3=${f%.*}.vcf ;
        samtools mpileup -f genome.fa $f > $f2 ;
        cat $f2 | java -jar VarScan.v2.3.6.jar mpileup2snp --output-vcf --min-coverage 10 --min-avg-
qual 25 --min-var-freq 0.00 -> $f3 ; done ;


for f in *.bam ;
do      f2=${f%.*}2.mpileup ;
        f3=${f%.*}2.vcf ;

        samtools index $f ;
        samtools mpileup -B -f genome.fa $f > $f2 ;
        cat $f2 | java -jar VarScan.v2.3.6.jar mpileup2snp --output-vcf --min-coverage 10 --min-avg-
qual 25 --min-var-freq 0.00 -> "$f3" ; done ;
```

##

FIND ZERO GENE-FLOW (ZGF) SITES

#"chrms-with-split.txt"is a list of all chromosomes, including any parts of chromosomes that have been split for data processing

```
        for i in `cat chrms-with-split.txt` ;
do
python3 parse_match.py "$i"_matched_ZGFsplitreaddepths *.$i-splitreaddepths ; done ;

#REPLACE NULL CELLS IN TABLE WITH DASHES

for f in *matched_ZGFsplitreaddepths ; do
```

```
sed -i 's|\t\t|\t-\t|g' $f ;  done ;

for f in *_matched_ZGFsplitreaddepths ;
do
cat $f | sort -V > "$f"_sorted ; done ;

#FIND SITES COMMON TO ALL ZGF POPLNS

for f in *_matched_ZGFsplitreaddepths_sorted ;  do

awk '{ if (NR!=1 && $0 !~ "-") print $0}' $f > "$f"_filteredsites ;

done ;

cat *_filteredsites | awk '{print $1}' > All_ZGFfilteredsites

cat All_ZGFfilteredsites | awk '{ print $1 "\t" "-"}' > All_ZGFfilteredsites_2col ;

cat All_ZGFfilteredsites | awk -F '_' '{ print "Y55."$1 "\t" $2}' > All_ZGFfilteredsites_sites ;

cat All_ZGFfilteredsites | awk -F '_' '{ print "Y55."$1 "\t" "-"}' > All_ZGFfilteredsites.Y55 ;

############ FILTER SNPs

#split ZGF sites file

for f in All_ZGFfilteredsites.Y5* ; do

for i in `cat chrms.txt` do

f2=${f%.*}.$i ;

awk -F '[.|_]' '{if ($2 ~ "'$i'") print $0}' $f > $f2 ;

done ;  done ;

#MATCH TO .VCF

for f in *.vcf ;

do awk '{ if (NR <= 24) print $0}' $f  > "$f"_header ; awk '{if (NR >= 25) print $1 "_" $2 "\t" $4 "_"
$5 "_" $10}' $f > "$f"2  ;   done ;
```

```
for f in *vcf2 ; do

for i in `cat chrms.txt` ;

do python3 parse_match.py "$f"_ZGF."$i" All_ZGFfilteredsites."$i" $f ;

done ;

cat "$f"_ZGF.chr* > "$f"_ZGF_all ; done ;

#REMOVE ANC SNPS

awk '{ print $1 "_" $2 "\t" "-" }' Anc-merged-realigned.vcf > Anc-merged-sites1 ;

awk '{ print $1 "_" $2 "\t" "-" }' Anc-merged-varscan-snps.vcf > Anc-merged-sites2 ;

cat Anc-merged-sites1 Anc-merged-sites2 | sort > Anc-merged-sites-all ;

cat *.vcf2_ZGF_all | awk '{ print $1}' | sort | uniq -c | awk -F ' ' '{if ($1>=6) print $2 "\t" "-"}' >
highly_parallel.txt ;

cat Anc-merged-sites-all highly_parallel.txt > filter.txt ;

for f in *_ZGF_all ; do

awk 'FNR==NR {a[$1]; next}; !($1 in a)' filter.txt $f > "$f"_non-filter ; done ;

# Count SNPs at different frequencies (>50% & >70%)

for f in *non-filter ; do echo "$f" ; awk -F '[:|%]' '{if ($7>=50) print}' $f | wc -l ; done ;

for f in *non-filter ; do echo "$f" ; awk -F '[:|%]' '{if ($7>=70) print}' $f | wc -l ; done ;

# COUNT PARALLEL SNPScat *.vcf2_ZGF_all | awk '{ print $1}' | sort | uniq -c | awk -F ' ' '{if
($1>=3) print }' | awk -F ' ' '{if ($1 <=5) print $2 "\t" $1 }' > parallel.txt

for f in *non-filter ; do

echo "$f" ;
awk 'FNR==NR {a[$1]; next}; ($1 in a)' parallel.txt $f | wc -l ;
done ;
```

#RUN 'SNPdat' FOR SNP ANNOTATION

```
for f in *non-anc ;  do

perl SNPdat.pl -i $f -g SCY55genes-chrms.gtf -f genome.fa -o "$f"-snpdat ;

done
```

##3

FILTER INDELS

#PRINT INDELS WITH FREQUENCY, RECALCULATED , AND INDEX FILE

```
 for f in *indels.vcf ;  do

f2=${f%.*}-freq.vcf ;

cat $f | awk '{ if (NR <= 49) print $0}' > header.txt ;

cat $f | awk '{ if (NR >= 50) print $1 "\t" $2 "\t" $4 "\t" $5 "\t" $10 }' | awk -F '[\t|:|,]' '{print $1 "_" $2 "\t" $3 "\t" $4 "\t" ($7 / ($6 + $7))}' > main.txt ;

cat header.txt main.txt > $f2 ; done ;


# for f in *-freq.vcf ; do  awk '{ if (NR <= 49) print $0 }; ELSE {print $0 "\t" "." "\t" "." "\t" "." "\t" "." "\t" "." }' $f > "$f"9 ; cat "$f"9 > $f | java -jar '/media/sf_shared/2016/IGVTools/igvtools.jar' index $f ; done

###MATCH TO ZGF SITES

#split ZGF sites file

for f in All_ZGFfilteredsites.Y5* ;   do
for i in `cat chrms.txt`
do

f2=${f%.*}.$i ;

awk -F '[.|_]' '{if ($1 ~ "'$i'") print $0}' $f > $f2 ;

done ; done ;
```

#MATCH TO .VCF

```
for f in *-freq.vcf ;

do awk '{ if (NR <= 49) print $0}' $f  > "$f"_header ; awk '{if (NR >= 50) print $1 "\t" $2 "_" $3 "_" $4}' $f > "$f"2

for i in `cat chrms.txt` ; do

python3 parse_match.py "$f"_ZGF."$i" All_ZGFfilteredsites."$i" "$f"2 ;  done ;

cat "$f"_ZGF.chr* > "$f"_ZGF_all ; done ;
```

#REMOVE ANC INDELS, FILTERING AGAINST ANC SITES AND -2,-1,+1, and +2 nucleotides compared to Anc sites.

```
cat Anc-merged-realigned_indels-freq.vcf_ZGF_all | \

 awk -F'[_|\t]' '{print $1 "_" ($2+2) "\n" $1 "_" ($2+1) "\n" $1 "_" $2 "\n" $1 "_" ($2-1) "\n" $1 "_" ($2-2)}' \

 > Anc-merged-realigned_indels-freq.vcf_ZGF_all_conservative ;
```

COMBINE WITH SNPS +1, 0, -1 FILE FOR MORE CONSERVATIVE FILTERING, AND ADAPT REST OF SCRIPT, RECOUNT INDELS

```
for f in *nc-merged-realigned.vcf ; do cat $f | \

 awk  '{ if (NR >= 25) print $1 "_" ($2+1) "\t" "-" "\n" $1 "_" $2 "\t" "-" "\n" $1 "_" ($2-1) "\t" "-" }' \

> "$f"_SNPs-filter ;

cat Anc-merged-realigned_indels-freq.vcf_ZGF_all_conservative "$f"_SNPs-filter | sort > "$f"_conservative-withSNPs ;

done  ;
```

```
cat Anc-merged-realigned.vcf_conservative-withSNPs Anc_SCY55.sorted_indels-freq.vcf2 | sort | awk '{ print $1 }' > Anc-merged-realigned.vcf_conservative-withSNPs_final ;
```

```
for f in *_ZGF_all ;

do

awk 'FNR==NR {a[$1]; next}; !($1 in a)' Anc-merged-realigned.vcf_conservative-withSNPs_final $f >
"$f"_non-anc ;

done ;

#MATCH WITH GENE

for f in *non-anc ; do awk '{ print $1 "\t" $3}' $f > "$f"2 ; done ;

awk '{print "Y55." $0 }' Y55.genes-full > Y55.genes-full2 ;

for f in *non-anc2 ; do python3 parse_match.py "$f"_genes $f Y55.genes-full2 ; done
```

```
##########################################################################################
```

FIND COPY NUMBER VARIANTS

```
#STATE INPUT FILES REQUIRED

echo "Input files required:

bam files

chrms.txt [list of nuclear chrms - replace names of large chrms with split forms]

parse_match.py [Python script for matching, written by Ryan Estep]

Y55.genes-test"

###
# CREATE MPILEUP FILES FROM SORTED BAM FILES#
##
for f in *.bam ; do samtools index $f ; done ;

for f in *.bam ; do  f2=${f%.*}.mpileup ;

samtools mpileup -f genome.fa $f > $f2 ;

rm $f ; done ;

###
```

```
# CONVERT MPILEUP TO READ DEPTH FILES

###

for f in *.mpileup; do

f2=${f%.*}.readdepths ;

cat $f | sed "s|Y55.||g" | awk '{print $1 "_" $2 "\t" $4}' > $f2 ; done ;

###

# FILTER TO HIGH READ DEPTH SITES ONLY

###

for f in *.readdepths ; do

f2=${f%.*}.highreaddepths ;

cat $f | awk '{if ($2>=10) print $1 "\t" $2}' > $f2 ; done ;

###

# split all read depth files by chromosome

###

cat Anc-merged_SCY55.sorted.highreaddepths | awk -F '_' '{print $1}' | uniq > chrms-all.txt ;

for f in *.highreaddepths ; do

for i in `cat chrms-all.txt` do

f2=${f%.*}.$i-splitreaddepths ;

awk '{if ($1 ~ "'$i'") print $1 "\t" $2}' $f > $f2 ; done ;  done ;

### SPLIT LARGE CHROMOSOMES IN HALF (chrms 4, 7, 15)

for f in *.chr04-splitreaddepths ;  do

f2=${f%.*}.chr41-splitreaddepths ;

f3=${f%.*}.chr42-splitreaddepths ;

cat $f | sed 's|_|\t|g' | awk '{if ($2 <= 700000) print $1 "_" $2 "\t" $3}' > $f2 ;
```

```
cat $f | sed 's|_|\t|g' | awk '{if ($2 > 700000) print $1 "_" $2 "\t" $3}' > $f3  ; done ;

for f in *.chr07-splitreaddepths ;  do

f2=${f%.*}.chr71-splitreaddepths ;

f3=${f%.*}.chr72-splitreaddepths ;

cat $f | sed 's|_|\t|g' | awk '{if ($2 <= 700000) print $1 "_" $2 "\t" $3}' > $f2 ;

cat $f | sed 's|_|\t|g' | awk '{if ($2 > 700000) print $1 "_" $2 "\t" $3}' > $f3  ; done ;

for f in *.chr15-splitreaddepths ;  do

f2=${f%.*}.chrf1-splitreaddepths ;

f3=${f%.*}.chrf2-splitreaddepths ;

cat $f | sed 's|_|\t|g' | awk '{if ($2 <= 700000) print $1 "_" $2 "\t" $3}' > $f2 ;

cat $f | sed 's|_|\t|g' | awk '{if ($2 > 700000) print $1 "_" $2 "\t" $3}' > $f3  ; done ;

###

# MATCH ALL FILES ON SITES

###

for i in `cat chrms-all.txt` ; do

python3 parse_match.py "$i"_matched_splitreaddepths *.$i-splitreaddepths ; done ;

#REPLACE NULL CELLS IN TABLE WITH DASHES

for f in *matched_splitreaddepths ; do

sed -i 's|\t\t|\t-\t|g' $f ;

done ;

### SORT MATCHED FILES TABLE

for f in *_matched_splitreaddepths ;

do cat $f | sort -V > "$f"_sorted ;

done ;
```

LIMIT MATCHED LIST TO ROWS WITH ANCESTRAL SITES

```
for f in *_matched_splitreaddepths_sorted ;  do

i=$(cat $f | awk '{ if (NR ==1) print NF}') ;

awk '{ if ($"'$i'" >= 10) print $0}' $f  > "$f"_ancsites ; done ;
```

REJOIN LARGE CHROMOSOMES TOGETHER

```
cat chr4B_matched_splitreaddepths_sorted_ancsites | awk '{ if (NR!=1) print $0}' >
chr4B_matched_splitreaddepths_sorted_ancsites2 ;

cat chr4A_matched_splitreaddepths_sorted_ancsites
chr4B_matched_splitreaddepths_sorted_ancsites2 >
chr04_matched_splitreaddepths_sorted_ancsites ;

cat chr7B_matched_splitreaddepths_sorted_ancsites | awk '{ if (NR!=1) print $0}' >
chr7B_matched_splitreaddepths_sorted_ancsites2 ;

cat chr7A_matched_splitreaddepths_sorted_ancsites
chr7B_matched_splitreaddepths_sorted_ancsites2 >
chr07_matched_splitreaddepths_sorted_ancsites ;

cat chrm15B_matched_splitreaddepths_sorted_ancsites | awk '{ if (NR!=1) print $0}' >
chrm15B_matched_splitreaddepths_sorted_ancsites2 ;

cat chrm15A_matched_splitreaddepths_sorted_ancsites
chrm15B_matched_splitreaddepths_sorted_ancsites2 >
chr15_matched_splitreaddepths_sorted_ancsites ;
```

REMOVE HEADER FOR REATTACHING LATER

```
for i in `cat chrms.txt` ;  do for f in $i*_ancsites ; do cat $f | head -1 > "$f".header ;

cat $f | awk '{if (NR!=1) print}' > "$f".text ; done ; done ;
```

########## DIVIDE DERIVED READ DEPTH BY ANCESTRAL READ DEPTH AT ALL SITES

```
for f in *_matched_splitreaddepths_sorted_ancsites.text ;    do

cat $f | awk '{printf($1)"\t"} {for(i=2;i<=NF; i++) printf($i /$NF)"\t"};{print FS}' > "$f"_normalised ;
done ;
```

```
######## MEAN DER/ANC DEPTH RATIO (M)

######## AND SITES GREATER THAN (1.5 * M) FOR EACH POPLN

# FIND AVERAGES OF COLUMNS (AVERAGE RATIO DER:ANC DEPTH)

# (AND ADD "Averages" TITLE TO CHROMOSOME_SITES COLUMN)

for f in *.text_normalised ; do  echo "$f" ;

cat $f | awk '{for(i=2; i<=NF; i++){sum[i]+=$i}} END {for(i=2; i<=NF; i++) { if (i<=NF) printf
sum[i]/NR "\t"; else printf sum[i]/NR}}' | awk '{print "Averages" "\t" $0 }' > "$f"_averages ; done ;

### APPEND HEADER AS FIRST LINE, AVERAGES AS SECOND#

for f in *.text_normalised ;  do

f2=${f%.*}.header ;   f3="$f"_averages ;  f4="$f"_combined ;  cat $f2 $f3 $f > $f4 ; cat $f4 | head -5 ;

done ;

### SPLIT BY COLUMN, AND NAME EACH OUTPUT ACCORDING TO FIRST LINE

for f in *_combined ;  do

chrm="$(echo "$f" | awk -F '_' '{print $1}')";

cat $f | awk -v chrm="$chrm" 'NR==1{for (i=2;i<=NF;i++) a[i]=$i; next}

{for (i=2;i<=NF;i++) {print $1 "\t" $i > chrm"-"a[i]".popln"}}' ;  done ;


### PRINT LINES WHERE $2 VALUES GREATER THAN (1.5 x AVERAGE) [and 3 x AVERAGE]

for f in *sorted.popln ;  do

i="$(cat $f | head -1 | awk '{print $2}')" ;

cat $f | awk '{ print $1 "\t" $2/"'$i'"}' > "$f"_compared ;

cat "$f"_compared | awk '{ if ($2 >= 2) print $0 }' > "$f"_2x_cnvs ;

cat "$f"_compared | awk '{ if ($2 >= 4) print $0 }' > "$f"_4x_cnvs ;

done ;
```

```
# SPLIT GENE-SITES FILE BY CHRM TO LIMIT RAM USE

for i in `cat chrms.txt` do

f2="$(echo "Genesites.""$i")" ;

cat Y55.genes-full | awk -F '[_|\t]' '{if ($1 ~ "'$i'") print $1 "_" $2 "\t" $3}' > $f2 ; done ;

# MATCH ON GENE SITES - uses python script parse_match.py [Ryan Estep]

for f in *.popln_*_cnvs ; do

i="$(echo "$f" | awk -F '-' '{print $1}')" ;

j="$(echo "Genesites.$i")" ;

python3 parse_match.py "$f"_genes $j $f ; done ;

for a in `cat samples.txt` ;

do echo *"$a"*4x_*genes | sed 's| |\n|g' | while read -r line ; do

awk '{ if (NR!=1) print $2}' $line ; done > "$a"_combined-4x ; cat "$a"_combined-4x | sort | uniq -c
> "$a"_combined_uniq-4x ; done ;

for a in `cat samples.txt` ;
do echo *"$a"*_2x_*genes | sed 's| |\n|g' | while read -r line ; do
awk '{ if (NR!=1) print $2}' $line ; done > "$a"_combined-2x ; cat "$a"_combined-2x | sort | uniq -c
> "$a"_combined_uniq-2x ; done

#sort gene results files by counts (highest to lowest number of sites) and put gene name in first
column

for f in *_uniq-2x ; do sort -g -r $f | awk '{print $2 "\t" $1}' > "$f"_counts ;

done ;

for f in *_uniq-4x ; do sort -g -r $f | awk '{print $2 "\t" $1}' > "$f"_counts ;

done ;

# create gene lengths table

cat Y55.genes | awk '{print $6 "\t" $5-$4}' > Y55.genes-lengths ;
```

```
#match gene results with counts to gene lengths table

for f in *uniq*counts ;  do

python3 parse_match.py "$f"_lengths $f Y55.genes-lengths ;

done ;


#filter med files to genes that are represented by sites representing at least 50% of total length

for f in *_lengths ;  do

echo "$f" ; awk '{ if (NR!=1) {if ($2/$3 >= 0.70) print $0 }}' $f > "$f"filtered70 ;

done ;

############################################################

# match genes to midpoints file

cat Y55.genes | awk '{ print $6 "\t" $1 "\t" ($4 + $5) /2 }' > Y55.genes-midpoints ;

for f in *filtered 90 ;

python3 parse_match.py "$f"_midpoints Y55.genes-midpoints $f

#### Parallelisation of genes, for each sample

# Combine filtered70 files by env-treatment

echo "Nitrogen Limited, Zero Gene Flow" > NLim.txt ;

for f in 0*1-5x*filtered70 ; do

echo "$f" >> NLim.txt ;

cat $f >> NLim.txt ;  done ;

echo "Carbon Limited, Zero Gene Flow" > CLim.txt ;

for f in 17*1-5x*filtered70 ; do

echo "$f" >> CLim.txt ; cat $f >> CLim.txt ;

done ;
```

```
for f in 18*1-5x*filtered70 ; do  echo "$f" >> CLim.txt ;

cat $f >> CLim.txt ;  done ;

for f in 19*1-5x*filtered70 ; do  echo "$f" >> CLim.txt ;

cat $f >> CLim.txt ;  done ;

echo "Carbon Limited, Full Gene Flow" > CLimGF.txt ;

for f in 2*1-5x*filtered70 ; do  echo "$f" >> CLimGF.txt ;

cat $f >> CLimGF.txt ;  done ;

for f in 3*1-5x*filtered70 ; do  echo "$f" >> CLimGF.txt ;

cat $f >> CLimGF.txt ;  done ;

echo "Nitrogen Limited, Full Gene Flow" > NLimGF.txt ;

for f in 13*1-5x*filtered70 ; do  echo "$f" >> CLimGF.txt ;

cat $f >> CLimGF.txt ;  done ;

for f in 14*1-5x*filtered70 ; do  echo "$f" >> CLimGF.txt ;

cat $f >> CLimGF.txt ;  done ;

for f in 15*1-5x*filtered70 ; do

echo "$f" >> CLimGF.txt ;

cat $f >> CLimGF.txt ;  done ;


##########  EXTRACT HXT6 sites from read depths file, for comparison

for f in *chr04*splitreaddepths ; do i="$(cat $f | awk '{if ($1=="chr04_1155656") print NR}')" ; sed -
n "$i",+1712p $f > "$f"_hxt6 ; done
```

Appendix C
Supplementary Tables

Table 6.1 – Models of Trade-off, Genome Size, and Mutation Rate

Antagonistic Pleiotropy	**Genome size**	**Mutation rate**	
Pentermann et al 2014 - Pseudomonas aeruginosa biofilm growth	6300000	1.04E-09	
E coli - Lenski lines - non-mutators	4629812	1E-10	
Pseudomonas fluorescens - Maclean & Bell 2004	7074893	3.16E-08	10 ^ -7.5
Saccharomyces cerevisiae - Kvitek & Sherlock 2011	12156677	4E-10	4 x 10 ^ -10
E coli - temperature adaptation [AP more likely]	4629812	1E-10	10 ^ -10
Methylobacterium extorquens 2.8 x 10-3 per genome per generation	5511322	5E-10	5 x 10 ^ -10
Hong & Nielson, March 2013 - S cerevisiae	12156677	4E-10	4 x 10 ^ -10
Mutation Accumulation			
Lenski lines - mutators	4629812	1E-08	10 ^ -8
Pseudomonas fluorescens - Maclean & Bell 2002	7074893	3.16E-08	10 ^ -7.5
Chlamydomonas	1.2E+08	3.23E-10	3.23 x 10 ^ -10

Bibliography

Abrescia NGA, Bamford DH, Grimes JM, Stuart DI (2012) Structure Unifies the Viral Universe. Annual Review of Biochemistry 81:795–822. doi: 10.1146/annurev-biochem-060910-095130

Adami C (2002) What is complexity? BioEssays 24:1085–1094. doi: 10.1002/bies.10192

Adams J, Rosenzweig F (2014) Experimental microbial evolution: history and conceptual underpinnings. Genomics 104:393–398. doi: 10.1016/j.ygeno.2014.10.004

Anderson JT, Lee C-R, Rushworth CA, et al (2013) Genetic Trade-Offs and Conditional Neutrality Contribute to Local Adaptation: Genetic Basis of Local Adaptation. Molecular Ecology 22:699–708. doi: 10.1111/j.1365-294X.2012.05522.x

Ardern ZN, Goddard MR (2015) Investigating Trade-offs in Sexual Populations with Gene Flow. In: Pontarotti P (ed) Evolutionary Biology: Biodiversification from Genotype to Phenotype. Springer International Publishing, Cham, pp 245–262

Bailey SF, Hinz A, Kassen R (2014) Adaptive synonymous mutations in an experimentally evolved Pseudomonas fluorescens population. Nature Communications. doi: 10.1038/ncomms5076

Ballard JWO, Pichaud N (2014) Mitochondrial DNA: more than an evolutionary bystander. Functional Ecology 28:218–231. doi: 10.1111/1365-2435.12177

Ball P (2013) DNA: Celebrate the unknowns. Nature 496:419–420. doi: 10.1038/496419a

Bank C, Ewing GB, Ferrer-Admettla A, et al (2014) Thinking too positive? Revisiting current methods of population genetic selection inference. Trends in Genetics 30:540–546. doi: 10.1016/j.tig.2014.09.010

Becks L, Agrawal AF (2010) Higher rates of sex evolve in spatially heterogeneous environments. Nature 468:89–92. doi: 10.1038/nature09449

Behe MJ (2010) Experimental evolution, loss-of-function mutations, and "the first rule of adaptive evolution." The Quarterly Review of Biology 85:419–445. doi: 10.1086/656902

Behrends V, Maharjan RP, Ryall B, et al (2014) A metabolic trade-off between phosphate and glucose utilization in Escherichia coli. Mol BioSyst 10:2820–2822. doi: 10.1039/C4MB00313F

Bell G (1982) The masterpiece of nature: the evolution and genetics of sexuality. Croom Helm, London

Bell G (2008) Selection: the mechanism of evolution, 2nd ed. Oxford University Press, Oxford ; New York

Bendich AJ, Drlica K (2000) Prokaryotic and eukaryotic chromosomes: what's the difference? BioEssays 22:481. doi: 10.1002/(SICI)1521-1878(200005)22:5<481::AID-BIES10>3.3.CO;2-K

Bennett AF, Lenski RE (2007) An Experimental Test of Evolutionary Trade-Offs During Temperature Adaptation. In: In the Light of Evolution. The National Academies Press, Washington DC,

Berger D, Walters RJ, Blanckenhorn WU (2014) Experimental evolution for generalists and specialists reveals multivariate genetic constraints on thermal reaction norms. Journal of Evolutionary Biology 27:1975–1989. doi: 10.1111/jeb.12452

Bernstein H, Bernstei C (2013) Evolutionary Origin and Adaptive Function of Meiosis. In: Bernstein C (ed) Meiosis. InTech,

Björkman J (2000) Effects of Environment on Compensatory Mutations to Ameliorate Costs of Antibiotic Resistance. Science 287:1479–1482. doi: 10.1126/science.287.5457.1479

Blanquart F, Achaz G, Bataillon T, Tenaillon O (2014) Properties of selected mutations and genotypic landscapes under Fisher's geometric model: GENOTYPIC LANDSCAPES UNDER FISHER'S MODEL. Evolution 68:3537–3554. doi: 10.1111/evo.12545

Blount ZD, Barrick JE, Davidson CJ, Lenski RE (2012) Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature 489:513–518. doi: 10.1038/nature11514

Blows MW, Allen SL, Collet JM, et al (2015) The Phenome-Wide Distribution of Genetic Variance. The American Naturalist 186:15–30. doi: 10.1086/681645

Bohannan BJM, Lenski RE (2000) The Relative Importance of Competition and Predation Varies with Productivity in a Model Community. The American Naturalist 156:329–340. doi: 10.1086/303393

Bradwell K, Combe M, Domingo-Calap P, Sanjuan R (2013) Correlation Between Mutation Rate and Genome Size in Riboviruses: Mutation Rate of Bacteriophage Q. Genetics 195:243–251. doi: 10.1534/genetics.113.154963

Brocchieri L, Karlin S (2005) Protein length in eukaryotic and prokaryotic proteomes. Nucleic Acids Research 33:3390–3400. doi: 10.1093/nar/gki615

Burke MK (2012) How does adaptation sweep through the genome? Insights from long-term selection experiments. Proc R Soc B. doi: 10.1098/rspb.2012.0799

Burt A (2000) Sex, recombination, and the efficacy of selection - was Weisman right? Evolution 54:337–351. doi: 10.1111/j.0014-3820.2000.tb00038.x

Butlin R (2002) Evolution of sex: The costs and benefits of sex: new insights from old asexual lineages.

Carroll SM, Lee M-C, Marx CJ (2013) Sign Epistasis Limits Evolutionary Trade-Offs at the Confluence of Single- and Multi-Carbon Metabolism in *Methylobacterium Extorquens* Am1: Evolved Metabolic Trade-Offs Limited by Epistasis. Evolution n/a–n/a. doi: 10.1111/evo.12301

Chang, Shang-Lin, and Jun-Yi Leu. "A Tradeoff Drives the Evolution of Reduced Metal Resistance in Natural Populations of Yeast." Edited by Jianzhi Zhang. PLoS Genetics 7, no. 3 (March 31, 2011): e1002034. doi:10.1371/journal.pgen.1002034.

Charlesworth B (2012) The Effects of Deleterious Mutations on Evolution at Linked Sites. Genetics 190:5–22. doi: 10.1534/genetics.111.134288

Chevin L-M (2010) On measuring selection in experimental evolution. Biology Letters 7:210–213. doi: 10.1098/rsbl.2010.0580

Chou H-H, Chiu H-C, Delaney NF, et al (2011) Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. Science 332:1190–1192. doi: 10.1126/science.1203799

Colegrave N (2012) The evolutionary success of sex. EMBO reports 13:774–778. doi: 10.1038/embor.2012.109

Cooper VS (2014) The Origins of Specialization: Insights from Bacteria Held 25 Years in Captivity. PLoS Biology 12:e1001790. doi: 10.1371/journal.pbio.1001790

Cooper VS, Lenski R (2000) The population genetics of ecological specialization in evolving Escherichia coli populations.

Cowen LE, Kohn LM, Anderson JB (2001) Divergence in Fitness and Evolution of Drug Resistance in Experimental Populations of Candida albicans. Journal of Bacteriology 183:2971–2978. doi: 10.1128/JB.183.10.2971-2978.2001

Cuevas JM, Domingo-Calap P, Sanjuan R (2012) The Fitness Effects of Synonymous Mutations in DNA and RNA Viruses. Molecular Biology and Evolution 29:17–20. doi: 10.1093/molbev/msr179

Dean AM, Dykhuizen DE (2009) Experimental Evolution from the Bottom Up. In: Garland T, Rose MR (eds) Experimental Evolution: Concepts, Methods and Applications of Selection Experiments.

Deatherage, Daniel E., and Jeffrey E. Barrick. "Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using Breseq." In Engineering and Analyzing Multicellular Systems, edited by Lianhong Sun and Wenying Shou, 1151:165–88. New York, NY: Springer New York, 2014. http://link.springer.com/10.1007/978-1-4939-0554-6_12.

Dettman JR, Rodrigue N, Melnyk AH, et al (2012) Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. Molecular Ecology 21:2058–2077. doi: 10.1111/j.1365-294X.2012.05484.x

De Vos MGJ, Poelwijk FJ, Battich N, et al (2013) Environmental Dependence of Genetic Constraint. PLoS Genetics 9:e1003580. doi: 10.1371/journal.pgen.1003580

Dhar R, SäGesser R, Weikert C, et al (2011) Adaptation of Saccharomyces cerevisiae to saline stress through laboratory evolution. Journal of Evolutionary Biology 24:1135–1153. doi: 10.1111/j.1420-9101.2011.02249.x

Dobzhansky T, Pavlovsky O (1957) An Experimental Study of Interaction between Genetic Drift and Natural Selection. Evolution 11:311. doi: 10.2307/2405795

Draghi JA, Parsons TL, Plotkin JB (2011) Epistasis Increases the Rate of Conditionally Neutral Substitution in an Adapting Population. Genetics 187:1139–1152. doi: 10.1534/genetics.110.125997

Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148:1667–1686.

Dykhuizen DE, Hartl DL (1983) Selection in chemostats. Microbiol Rev 47:150–168.

Eddy SR (2012) The C-value paradox, junk DNA and ENCODE. Current Biology 22:R898–R899. doi: 10.1016/j.cub.2012.10.002

Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nature Reviews Genetics 4:457–469. doi: 10.1038/nrg1088

Elliott TA, Gregory TR (2015) What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philosophical Transactions of the Royal Society B: Biological Sciences 370:20140331. doi: 10.1098/rstb.2014.0331

El-Soda M, Malosetti M, Zwaan BJ, et al (2014) Genotype × environment interaction QTL mapping in plants: lessons from Arabidopsis. Trends in Plant Science. doi: 10.1016/j.tplants.2014.01.001

ER Jerison, S Kryazhimskiy, M Desai (2014) Pleiotropic consequences of adaptation across gradations of environmental stress in budding yeast.

Farslow JC, Lipinski KJ, Packard LB, et al (2015) Rapid Increase in frequency of gene copy-number variants during experimental evolution in Caenorhabditis elegans. BMC Genomics. doi: 10.1186/s12864-015-2253-2

Felsenstein J (1974) The Evolutionary Advantage of Recombination. Genetics 78:737–756.

Field, D., Tiwari, B., Booth, T., Houten, S., Swan, D., Bertrand, N. and Thurston, M. 2006. Open Software for biologists: from famine to feast. Nature Biotechnology 24, 801 – 803.

Flynn KM, Cooper TF, Moore FB-G, Cooper VS (2013) The Environment Affects Epistatic Interactions to Alter the Topology of an Empirical Fitness Landscape. PLoS Genetics 9:e1003426. doi: 10.1371/journal.pgen.1003426

Friar JL, Goldman T, Pérez–Mercader J (2012) Genome Sizes and the Benford Distribution. PLoS ONE 7:e36624. doi: 10.1371/journal.pone.0036624

Friedenberg NA (2003) Experimental evolution of dispersal in spatiotemporally variable microcosms. Ecology Letters 6:953–959. doi: 10.1046/j.1461-0248.2003.00524.x

Fry, James D. "The Evolution of Host Specialization: Are Trade-Offs Overrated?" *The American Naturalist* 148 (November 1996): S84–107. doi:10.1086/285904.

Fukuda Y, Nakayama Y, Tomita M (2003) On dynamics of overlapping genes in bacterial genomes. Gene 323:181–187. doi: 10.1016/j.gene.2003.09.021

Futuyma DJ, Moreno G (1988) The evolution of ecological specialization.

Galardini M, Mengoni A, Biondi EG, et al (2014) DuctApe: A suite for the analysis and correlation of genomic and OmniLog™ Phenotype Microarray data. Genomics 103:1–10. doi: 10.1016/j.ygeno.2013.11.005

Galperin MY (2007) Linear chromosomes in bacteria: no straight edge advantage? Environmental Microbiology 9:1357–1362. doi: 10.1111/j.1462-2920.2007.01328.x

García-Arenal F, Fraile A (2013) Trade-offs in host range evolution of plant viruses. Plant Pathology 62:2–9. doi: 10.1111/ppa.12104

Gentleman RC, Carey VJ, Bates DM, et al (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80. doi: 10.1186/gb-2004-5-10-r80

Gerstein AC, Berman J (2015) Shift and adapt: the costs and benefits of karyotype variations. Current Opinion in Microbiology 26:130–136. doi: 10.1016/j.mib.2015.06.010

Goddard MR (2007) Why Bother with Sex? Answers from Experiments with Yeast and Other Organisms. In: Taylor JW, Kronstad JW, Heitman J, Casselton LA (eds) Sex in Fungi. American Society of Microbiology, pp 489–506

Goddard MR, Godfray HCJ, Burt A (2005) Sex increases the efficacy of natural selection in experimental yeast populations. Nature 434:636–640. doi: 10.1038/nature03405

Goddard MR, Greig D (2015) Saccharomyces cerevisiae: a nomadic yeast with no niche? FEMS Yeast Research 15:fov009–fov009. doi: 10.1093/femsyr/fov009

Gorelick R, Heng HHQ (2011) Sex Reduces Genetic Variation: A Multidisciplinary Review. Evolution 65:1088–1098. doi: 10.1111/j.1558-5646.2010.01173.x

Gouyon P-H, de Vienne D, Giraud T (2015) Sex and Evolution. In: Heams T, Huneman P, Lecointre G, Silberstein M (eds) Handbook of Evolutionary Thinking in the Sciences. Springer Netherlands, Dordrecht, pp 499–507

Gray JC (2011) Testing the Major Theories Concerning the Evolution of Sex using Experimental Evolution. A thesis submitted to the University of Auckland.

Gray JC, Cutter AD (2014) Mainstreaming Caenorhabditis elegans in experimental evolution. Proceedings of the Royal Society B: Biological Sciences 281:20133055–20133055. doi: 10.1098/rspb.2013.3055

Gray JC, Goddard MR (2012a) Gene-flow between niches facilitates local adaptation in sexual populations. Ecology Letters 15:955–962. doi: 10.1111/j.1461-0248.2012.01814.x

Gray JC, Goddard MR (2012b) Sex enhances adaptation by unlinking beneficial from detrimental mutations in experimental yeast populations. BMC Evolutionary Biology 12:43. doi: 10.1186/1471-2148-12-43

Gresham D, Desai MM, Tucker CM, et al (2008) The Repertoire and Dynamics of Evolutionary Adaptations to Controlled Nutrient-Limited Environments in Yeast. PLoS Genetics 4:e1000303. doi: 10.1371/journal.pgen.1000303

Gresham D, Dunham MJ (2014) The enduring utility of continuous culturing in experimental evolution. Genomics 104:399–405. doi: 10.1016/j.ygeno.2014.09.015

Gresham D, Hong J (2014) The functional basis of adaptive evolution in chemostats. FEMS Microbiology Reviews n/a–n/a. doi: 10.1111/1574-6976.12082

Gros P-A, Le Nagard H, Tenaillon O (2009) The Evolution of Epistasis and Its Links With Genetic Robustness, Complexity and Drift in a Phenotypic Model of Adaptation. Genetics 182:277–293. doi: 10.1534/genetics.108.099127

Gros P-A, Tenaillon O (2009) Selection for Chaperone-Like Mediated Genetic Robustness at Low Mutation Rate: Impact of Drift, Epistasis and Complexity. Genetics 182:555–564. doi: 10.1534/genetics.108.099366

Güell M, Yus E, Lluch-Senar M, Serrano L (2011) Bacterial transcriptomics: what is beyond the RNA horiz-ome? Nature Reviews Microbiology 9:658–669. doi: 10.1038/nrmicro2620

Haag CR, Roze D (2007) Genetic Load in Sexual and Asexual Diploids: Segregation, Dominance and Genetic Drift. Genetics 176:1663–1678. doi: 10.1534/genetics.107.073080

Hall AR, Angst DC, Schiessl KT, Ackermann M (2015) Costs of antibiotic resistance - separating trait effects and selective effects. Evolutionary Applications 8:261–272. doi: 10.1111/eva.12187

Halligan DL, Keightley PD (2009) Spontaneous Mutation Accumulation Studies in Evolutionary Genetics. Annual Review of Ecology, Evolution, and Systematics 40:151–172. doi: 10.1146/annurev.ecolsys.39.110707.173437

Hansen SK, Rainey PB, Haagensen JAJ, Molin S (2007) Evolution of species interactions in a biofilm community. Nature 445:533–536. doi: 10.1038/nature05514

Harris EE (2010) Nonadaptive processes in primate and human evolution. American Journal of Physical Anthropology 143:13–45. doi: 10.1002/ajpa.21439

Hartfield M, Keightley PD (2012) Current hypotheses for the evolution of sex and recombination. Integrative Zoology 7:192–209. doi: 10.1111/j.1749-4877.2012.00284.x

Hartl, DL & Clark, AG (1997) Principles of Population Genetics, 3rd Edn. Sunderland, MA: Sinauer.

Hass JW (2000) The Reverend Dr William Henry Dallinger, F.R.S. (1839-1909). Notes and Records of the Royal Society 54:53–65. doi: 10.1098/rsnr.2000.0096

Hereford J (2009) A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. The American Naturalist 173:579–588. doi: 10.1086/597611

Hoekstra RF (2005) Evolutionary biology: Why sex is good. Nature 434:571–573. doi: 10.1038/434571a

Hong K, Nielson J (2013) Adaptively evolved yeast mutants on galactose show trade-offs in carbon utilization on glucose.

Horinouchi T, Minamoto T, Suzuki S, et al (2014) Development of an Automated Culture System for Laboratory Evolution. Journal of Laboratory Automation 19:478–482. doi: 10.1177/2211068214521417

Hou Y, Lin S (2009) Distinct Gene Number-Genome Size Relationships for Eukaryotes and Non-Eukaryotes: Gene Content Estimation for Dinoflagellate Genomes. PLoS ONE 4:e6978. doi: 10.1371/journal.pone.0006978

Huang S (2009) Inverse relationship between genetic diversity and epigenetic complexity. Nature Precedings. doi: 10.1038/npre.2009.1751.2

Hughes D, Andersson DI (2015) Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. Nature Reviews Genetics 16:459–471. doi: 10.1038/nrg3922

Jasmin J-N, Zeyl C (2013) Evolution of pleiotropic costs in experimental populations. Journal of Evolutionary Biology 26:1363–1369. doi: 10.1111/jeb.12144

ER Jerison, S Kryazhimskiy, and M Desai. "Pleiotropic Consequences of Adaptation across Gradations of Environmental Stress in Budding Yeast," 2014.

Judson OP, Normark BB (1996) Ancient asexual scandals. Trends in Ecology & Evolution 11:41–46. doi: 10.1016/0169-5347(96)81040-8

Kassen R (2002) The experimental evolution of specialists, generalists, and the maintenance of diversity: Experimental evolution in variable environments. Journal of Evolutionary Biology 15:173–190. doi: 10.1046/j.1420-9101.2002.00377.x

Kassen R (2014) Experimental evolution and the nature of biodiversity. Roberts and Company, Greenwood Village, Colorado

Kassen R, Bell G (1998) Experimental evolution in Chlamydomonas. IV. Selection in environments that vary through time at different scales. Heredity 80:732–741. doi: 10.1046/j.1365-2540.1998.00329.x

Kassen, Rees, and Paul B. Rainey. "The Ecology and Genetics of Microbial Diversity." Annual Review of Microbiology 58, no. 1 (October 2004): 207–31. doi:10.1146/annurev.micro.58.030603.123654.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. Ecology Letters 7:1225–1241. doi: 10.1111/j.1461-0248.2004.00684.x

Keightley PD, Eyre-Walker A (2000) Deleterious Mutations and the Evolution of Sex. Science 290:331–333. doi: 10.1126/science.290.5490.331

Kerstes NAG, Martin OY (2013) Insect host-parasite coevolution in the light of experimental evolution: Experimental host-parasite coevolution. Insect Science n/a–n/a. doi: 10.1111/1744-7917.12064

Ketola, Tarmo, Lauri Mikonranta, Ji Zhang, Kati Saarinen, Anni-Maria Örmälä, Ville-Petri Friman, Johanna Mappes, and Jouni Laakso. "Fluctuating Temperature Leads to Evolution of Thermal Generalism and Preadaptation to Novel Environments: Evolution of Thermal Tolerance and Preadaptation." Evolution, May 2013, n/a-n/a. doi:10.1111/evo.12148.

King T, Ishihama A, Kori A, Ferenci T (2004) A Regulatory Trade-Off as a Source of Strain Variation in the Species Escherichia coli. Journal of Bacteriology 186:5614–5620. doi: 10.1128/JB.186.17.5614-5620.2004

Kirkpatrick M, Peischl S (2012) Evolutionary rescue by beneficial mutations in environments that change in space and time. Philosophical Transactions of the Royal Society B: Biological Sciences 368:20120082–20120082. doi: 10.1098/rstb.2012.0082

Koboldt DC, Zhang Q, Larson DE, et al (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576. doi: 10.1101/gr.129684.111

Kondrashov AS (1988) Deleterious mutations and the evolution of sexual reproduction. Nature 336:435–440. doi: 10.1038/336435a0

Koufopanou V, Lomas S, Tsai IJ, Burt A (2015) Estimating the fitness effects of new mutations in the wild yeast Saccharomyces paradoxus. Genome Biology and Evolution. doi: 10.1093/gbe/evv112

Kron P, Suda J, Husband BC (2007) Applications of Flow Cytometry to Evolutionary and Population Biology. Annual Review of Ecology, Evolution, and Systematics 38:847–876. doi: 10.1146/annurev.ecolsys.38.091206.095504

Kuzminov A (2014) The Precarious Prokaryotic Chromosome. Journal of Bacteriology 196:1793–1806. doi: 10.1128/JB.00022-14

Kvitek DJ, Sherlock G (2011) Reciprocal Sign Epistasis between Frequently Experimentally Evolved Adaptive Mutations Causes a Rugged Fitness Landscape. PLoS Genetics 7:e1002056. doi: 10.1371/journal.pgen.1002056

Kvitek DJ, Sherlock G (2013) Whole Genome, Whole Population Sequencing Reveals That Loss of Signaling Networks Is the Major Adaptive Strategy in a Constant Environment. PLoS Genetics 9:e1003972. doi: 10.1371/journal.pgen.1003972

Lachapelle J, Bell G, Colegrave N (2015) Experimental adaptation to marine conditions by a freshwater alga: EXPERIMENTAL ADAPTATION TO MARINE CONDITIONS. Evolution 69:2662–2675. doi: 10.1111/evo.12760

Lane N (2011) Energetics and genetics across the prokaryote-eukaryote divide. Biology Direct 6:35. doi: 10.1186/1745-6150-6-35

Lane N, Martin W (2010) The energetics of genome complexity. Nature 467:929–934. doi: 10.1038/nature09486

Lang GI, Rice DP, Hickman MJ, et al (2013) Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature. doi: 10.1038/nature12344

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25. doi: 10.1186/gb-2009-10-3-r25

Lawrence JG, Roth JR (1999) Genomic Flux: Genome Evolution by Gene Loss and Acquisition. In: Charlebois RL (ed) Organization of the Prokaryotic Genome. American Society of Microbiology, pp 263–289

Lazar V, Pal Singh G, Spohn R, et al (2014) Bacterial evolution of antibiotic hypersensitivity. Molecular Systems Biology 9:700–700. doi: 10.1038/msb.2013.57

Lee C-R, Mitchell-Olds T (2012) Environmental Adaptation Contributes to Gene Polymorphism across the Arabidopsis thaliana Genome. Molecular Biology and Evolution 29:3721–3728. doi: 10.1093/molbev/mss174

Lehtonen J, Jennions MD, Kokko H (2012) The many costs of sex. Trends in Ecology & Evolution 27:172–178. doi: 10.1016/j.tree.2011.09.016

Leiby N (2014) Adaptation and Specialization in the Evolution of Bacterial Metabolism. Thesis submitted to Harvard University.

Leiby N, Marx CJ (2014) Metabolic Erosion Primarily Through Mutation Accumulation, and Not Tradeoffs, Drives Limited Evolution of Substrate Specificity in Escherichia coli. PLoS Biology 12:e1001789. doi: 10.1371/journal.pbio.1001789

Leinonen PH, Remington DL, LeppäLä J, Savolainen O (2013) Genetic basis of local adaptation and flowering time variation in *Arabidopsis lyrata*: GENETIC BASIS OF LOCAL ADAPTATION IN *A. LYRATA*. Molecular Ecology 22:709–723. doi: 10.1111/j.1365-294X.2012.05678.x

Lenski RE, Rose MR, Simpson SC, Tadler SC (1991) Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. The American Naturalist 138:1315. doi: 10.1086/285289

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

Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. doi: 10.1093/bioinformatics/btp352

Lieberman BS, Vrba ES (2005) Stephen Jay Gould on species selection: 30 years of insight. Paleobiology 31:113–121. doi: 10.1666/0094-8373(2005)031[0113:SJGOSS]2.0.CO;2

Lind PA, Farr AD, Rainey PB (2015) Experimental evolution reveals hidden diversity in evolutionary pathways. eLife. doi: 10.7554/eLife.07074

Lively CM, Morran LT (2014) The ecology of sexual reproduction. Journal of Evolutionary Biology 27:1292–1303. doi: 10.1111/jeb.12354

Lobkovsky AE, Wolf YI, Koonin EV (2013) Gene Frequency Distributions Reject a Neutral Model of Genome Evolution. Genome Biology and Evolution 5:233–242. doi: 10.1093/gbe/evt002

Lowry DB, Willis JH (2010) A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation. PLoS Biology 8:e1000500. doi: 10.1371/journal.pbio.1000500

MacLean RC, Bell G (2002) Experimental Adaptive Radiation in *Pseudomonas*. The American Naturalist 160:569–581. doi: 10.1086/342816

MacLean RC, Bell G, Rainey PB (2004) The evolution of a pleiotropic fitness tradeoff in Pseudomonas fluorescens. Proceedings of the National Academy of Sciences 101:8072–8077. doi: 10.1073/pnas.0307195101

MacLean RC, Vogwill T (2015) Limits to compensatory adaptation and the persistence of antibiotic resistance in pathogenic bacteria. Evolution, Medicine, and Public Health 2015:4–12. doi: 10.1093/emph/eou032

Manna F, Martin G, Lenormand T (2011) Fitness Landscapes: An Alternative Theory for the Dominance of Mutation. Genetics 189:923–937. doi: 10.1534/genetics.111.132944

Marini A, Matmati N, Morpurgo G (1999) Starvation in yeast increases non-adaptive mutation. Current Genetics 35:77–81. doi: 10.1007/s002940050435

Martin G, Elena SF, Lenormand T (2007) Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nature Genetics 39:555–560. doi: 10.1038/ng1998

Martin G, Lenormand T (2006a) A GENERAL MULTIVARIATE EXTENSION OF FISHER'S GEOMETRICAL MODEL AND THE DISTRIBUTION OF MUTATION FITNESS EFFECTS ACROSS SPECIES. Evolution 60:893. doi: 10.1554/05-412.1

Martin G, Lenormand T (2006b) THE FITNESS EFFECT OF MUTATIONS ACROSS ENVIRONMENTS: A SURVEY IN LIGHT OF FITNESS LANDSCAPE MODELS. Evolution 60:2413. doi: 10.1554/06-162.1

Martin G, Lenormand T (2015) The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. Evolution n/a–n/a. doi: 10.1111/evo.12671

Masel J (2005) Cryptic Genetic Variation Is Enriched for Potential Adaptations. Genetics 172:1985–1991. doi: 10.1534/genetics.105.051649

McDonald MJ, Rice DP, Desai MM (2016) Sex speeds adaptation by altering the dynamics of molecular evolution. Nature 531:233–236. doi: 10.1038/nature17143

McGee LW, Aitchison EW, Caudle SB, et al (2014) Payoffs, Not Tradeoffs, in the Adaptation of a Virus to Ostensibly Conflicting Selective Pressures. PLoS Genetics 10:e1004611. doi: 10.1371/journal.pgen.1004611

McKenna A, Hanna M, Banks E, et al (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. doi: 10.1101/gr.107524.110

McShea DW (2000) Functional Complexity in Organisms: Parts as Proxies.

Meftahi N, Namouchi A, Mhenni B, et al (2016) Evidence for the critical role of a secondary site *rpoB* mutation in the compensatory evolution and successful transmission of an MDR tuberculosis outbreak strain. Journal of Antimicrobial Chemotherapy 71:324–332. doi: 10.1093/jac/dkv345

Meirmans, Stephanie, Patrick G. Meirmans, and Lawrence R. Kirkendall. "The Costs Of Sex: Facing Real-World Complexities." The Quarterly Review of Biology 87, no. 1 (March 2012): 19–40. doi:10.1086/663945.

Melnyk AH, Wong A, Kassen R (2014) The fitness costs of antibiotic resistance mutations. Evolutionary Applications n/a–n/a. doi: 10.1111/eva.12196

Mendonça AG, Alves RJ, Pereira-Leal JB (2011) Loss of Genetic Redundancy in Reductive Genome Evolution. PLoS Computational Biology 7:e1001082. doi: 10.1371/journal.pcbi.1001082

Méthot P-O, Alizon S (2014) What is a pathogen? Toward a process view of host-parasite interactions. Virulence 5:775–785. doi: 10.4161/21505594.2014.960726

Munck C, Gumpert HK, Wallin AIN, et al (2014) Prediction of resistance development against drug combinations by collateral responses to component drugs. Science Translational Medicine 6:262ra156–262ra156. doi: 10.1126/scitranslmed.3009940

Nei M (2007) The new mutation theory of phenotypic evolution. Proceedings of the National Academy of Sciences 104:12235–12242. doi: 10.1073/pnas.0703349104

Ness RW, Morgan AD, Vasanthakrishnan RB, et al (2015) Extensive *de novo* mutation rate variation between individuals and across the genome of *Chlamydomonas reinhardtii*.

Nichol D, Jeavons P, Fletcher AG, et al (2015) Steering Evolution with Sequential Therapy to Prevent the Emergence of Bacterial Antibiotic Resistance. PLOS Computational Biology 11:e1004493. doi: 10.1371/journal.pcbi.1004493

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. Molecular Ecology 18:375–402. doi: 10.1111/j.1365-294X.2008.03946.x

Orr HA (2000) ADAPTATION AND THE COST OF COMPLEXITY. Evolution 54:13–20. doi: 10.1111/j.0014-3820.2000.tb00002.x

Orr HA (2002) THE POPULATION GENETICS OF ADAPTATION: THE ADAPTATION OF DNA SEQUENCES. Evolution 56:1317–1330. doi: 10.1111/j.0014-3820.2002.tb01446.x

Orr HA (2006) The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. Journal of Theoretical Biology 238:279–285. doi: 10.1016/j.jtbi.2005.05.001

Ostrowski EA, Ofria C, Lenski RE (2007) Ecological Specialization and Adaptive Decay in Digital Organisms. The American Naturalist 169:E1–E20. doi: 10.1086/510211

Otto SP (2009) The Evolutionary Enigma of Sex. The American Naturalist 174:S1–S14. doi: 10.1086/599084

Paaby AB, Rockman MV (2013) The many faces of pleiotropy. Trends in Genetics 29:66–73. doi: 10.1016/j.tig.2012.10.010

Paaby AB, Rockman MV (2014) Cryptic genetic variation: evolution's hidden substrate. Nature Reviews Genetics 15:247–258. doi: 10.1038/nrg3688

Pál C, Papp B, Lázár V (2015) Collateral sensitivity of antibiotic-resistant microbes. Trends in Microbiology. doi: 10.1016/j.tim.2015.02.009

Paquin C, Adams J (1983) Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations. Nature 302:495–500. doi: 10.1038/302495a0

Parera M, Martinez MA (2014) Strong Epistatic Interactions within a Single Protein. Molecular Biology and Evolution 31:1546–1553. doi: 10.1093/molbev/msu113

Payen C, Di Rienzi SC, Ong GT, et al (2013) The Dynamics of Diverse Segmental Amplifications in Populations of Saccharomyces cerevisiae Adapting to Strong Selection. G3: Genes|Genomes|Genetics. doi: 10.1534/g3.113.009365

Payen C, Di Rienzi SC, Ong GT, et al (2014) The Dynamics of Diverse Segmental Amplifications in Populations of Saccharomyces cerevisiae Adapting to Strong Selection. G3: Genes|Genomes|Genetics 4:399–409. doi: 10.1534/g3.113.009365

Payen C, Sunshine AB, Ong GT, et al (2015) Empirical determinants of adaptive mutations in yeast experimental evolution.

Peck JR (1994) A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. Genetics 137:597–606.

Pedersen SD (2013) Systems Biology Investigations of Pseudomonas aeruginosa Evolution in Association with Human Airway Infections. Technical University of Denmark

Penterman J, Nguyen D, Anderson E, et al (2014) Rapid Evolution of Culture-Impaired Bacteria during Adaptation to Biofilm Growth. Cell Reports. doi: 10.1016/j.celrep.2013.12.019

Peters A., Lively CM (2000) Epistasis and the Maintenance of Sex. In: Wolf, JB, Brodie ED, Wade MJ (eds) Epistasis and the Evolutionary Process.

Poon A (2005) The Coupon Collector and the Suppressor Mutation: Estimating the Number of Compensatory Mutations by Maximum Likelihood. Genetics 170:1323–1332. doi: 10.1534/genetics.104.037259

Posas F, Chambers JR, Heyman JA, et al (2000) The Transcriptional Response of Yeast to Saline Stress. Journal of Biological Chemistry 275:17249–17255. doi: 10.1074/jbc.M910016199

Poulton EB, Poulton EB, Schönland S, et al (1889) Essays upon heredity and kindred biological problems. Authorised translation, edited by Edward B. Poulton, Selmar Schönland and Arthur E. Shipley. Clarendon Press, Oxford

Qian W, Ma D, Xiao C, et al (2012) The Genomic Landscape and Evolutionary Resolution of Antagonistic Pleiotropy in Yeast. Cell Reports 2:1399–1410. doi: 10.1016/j.celrep.2012.09.017

Qi Q, Toll-Riera M, Heilbron K, et al (2016) The genomic basis of adaptation to the fitness cost of rifampicin resistance in *Pseudomonas aeruginosa*. Proceedings of the Royal Society B: Biological Sciences 283:20152452. doi: 10.1098/rspb.2015.2452

Raeside C, Gaffé J, Deatherage DE, et al (2014) Large Chromosomal Rearrangements during a Long-Term Evolution Experiment with Escherichia coli. mBio 5:e01377–14. doi: 10.1128/mBio.01377-14

Rainey PB, Travisano M (1998) Adaptive radiation in a heterogeneous environment. Nature 394:69–72. doi: 10.1038/27900

Rattray A, Santoyo G, Shafer B, Strathern JN (2015) Elevated Mutation Rate during Meiosis in Saccharomyces cerevisiae. PLoS Genetics 11:e1004910. doi: 10.1371/journal.pgen.1004910

Raynes Y, Sniegowski PD (2014) Experimental evolution and the dynamics of genomic mutation rate modifiers. Heredity 113:375–380. doi: 10.1038/hdy.2014.49

Reboud X, Bell G (1997) Experimental evolution in Chlamydomonas. III. Evolution of specialist and generalist types in environments that vary in space and time. Heredity 78:507–514. doi: 10.1038/hdy.1997.79

Remold S (2012) Understanding specialism when the jack of all trades can be the master of all. Proceedings of the Royal Society B: Biological Sciences 279:4861–4869. doi: 10.1098/rspb.2012.1990

Remolina SC, Chang PL, Leips J, et al (2012) Genomic Basis of Aging and Life-History Evolution in *Drosophila Melanogaster*: Genomics of Life-History Evolution. Evolution 66:3390–3403. doi: 10.1111/j.1558-5646.2012.01710.x

Richardson L (2015) Alternating Antibiotics Render Resistant Bacteria Beatable. PLOS Biology 13:e1002105. doi: 10.1371/journal.pbio.1002105

Robinson MD, Grigull J, Mohammad N, Hughes TR (2002) FunSpec: a web-based cluster interpreter for yeast. BMC Bioinformatics 3:35.

Rokyta DR, Joyce P, Caudle SB, et al (2011) Epistasis between Beneficial Mutations and the Phenotype-to-Fitness Map for a ssDNA Virus. PLoS Genetics 7:e1002075. doi: 10.1371/journal.pgen.1002075

Samani P, Bell G (2016) The ghosts of selection past reduces the probability of plastic rescue but increases the likelihood of evolutionary rescue to novel stressors in experimental populations of wild yeast. Ecology Letters 19:289–298. doi: 10.1111/ele.12566

Samani P, Low-Decarie E, McKelvey K, et al (2015) Metabolic variation in natural populations of wild yeast. Ecology and Evolution 5:722–732. doi: 10.1002/ece3.1376

Sandberg TE, Pedersen M, LaCroix RA, et al (2014) Evolution of Escherichia coli to 42 C and Subsequent Genetic Engineering Reveals Adaptive Mechanisms and Novel Mutations. Molecular Biology and Evolution 31:2647–2662. doi: 10.1093/molbev/msu209

Sanjuan R, Elena SF (2006) Epistasis correlates to genomic complexity. Proceedings of the National Academy of Sciences 103:14402–14405. doi: 10.1073/pnas.0604543103

Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. Nature Reviews Genetics 14:807–820. doi: 10.1038/nrg3522

Schluter D (2009) Evidence for Ecological Speciation and Its Alternative. Science 323:737–741. doi: 10.1126/science.1160006

Schuetz R, Zamboni N, Zampieri M, et al (2012) Multidimensional Optimality of Microbial Metabolism. Science 336:601–604. doi: 10.1126/science.1216882

Schulte, R. D., C. Makus, B. Hasert, N. K. Michiels, and H. Schulenburg. "Multiple Reciprocal Adaptations and Rapid Genetic Change upon Experimental Coevolution of an Animal Host and Its Microbial Parasite." Proceedings of the National Academy of Sciences 107, no. 16 (April 20, 2010): 7359–64. doi:10.1073/pnas.1003113107.

Schwander T, Crespi BJ (2009) Twigs on the tree of life? Neutral and selective models for integrating macroevolutionary patterns with microevolutionary processes in the analysis of asexuality. Mol Ecol 18:28–42. doi: 10.1111/j.1365-294X.2008.03992.x

Smith MA, Gesell T, Stadler PF, Mattick JS (2013) Widespread purifying selection on RNA structure in mammals. Nucleic Acids Research 41:8220–8236. doi: 10.1093/nar/gkt596

Stamatoyannopoulos JA (2012) What does our genome encode? Genome Research 22:1602–1611. doi: 10.1101/gr.146506.112

Stearns SC (1990) The evolutionary maintenance of sexual reproduction: The solutions proposed for a longstanding problem. Journal of Genetics 69:1–10. doi: 10.1007/BF02931662

Stoebel DM, Hokamp K, Last MS, Dorman CJ (2009) Compensatory Evolution of Gene Regulation in Response to Stress by Escherichia coli Lacking RpoS. PLoS Genetics 5:e1000671. doi: 10.1371/journal.pgen.1000671

Sung W, Ackerman MS, Miller SF, et al (2012) Drift-barrier hypothesis and mutation-rate evolution. Proceedings of the National Academy of Sciences 109:18488–18492. doi: 10.1073/pnas.1216223109

Sunshine AB, Payen C, Ong GT, et al (2015) The Fitness Consequences of Aneuploidy Are Driven by Condition-Dependent Gene Effects. PLOS Biology 13:e1002155. doi: 10.1371/journal.pbio.1002155

Szamecz B, Boross G, Kalapis D, et al (2014) The Genomic Landscape of Compensatory Evolution. PLoS Biology 12:e1001935. doi: 10.1371/journal.pbio.1001935

Szathmáry E, Smith JM (1995) The major evolutionary transitions. Nature 374:227–232. doi: 10.1038/374227a0

Taft RJ, Pheasant M, Mattick JS (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. BioEssays 29:288–299. doi: 10.1002/bies.20544

Tang Y-C, Amon A (2013) Gene Copy-Number Alterations: A Cost-Benefit Analysis. Cell 152:394–405. doi: 10.1016/j.cell.2012.11.043

Tenaillon O (2014) The Utility of Fisher's Geometric Model in Evolutionary Genetics. Annual Review of Ecology, Evolution, and Systematics 45:179–201. doi: 10.1146/annurev-ecolsys-120213-091846

Tenaillon O, Silander OK, Uzan J-P, Chao L (2007) Quantifying Organismal Complexity using a Population Genetic Approach. PLoS ONE 2:e217. doi: 10.1371/journal.pone.0000217

Tendler A, Mayo A, Alon U (2015) Evolutionary tradeoffs, Pareto optimality and the morphology of ammonite shells. BMC Systems Biology. doi: 10.1186/s12918-015-0149-z

Teresa Avelar A, Perfeito L, Gordo I, Godinho Ferreira M (2013) Genome architecture is a selectable trait that can be maintained by antagonistic pleiotropy. Nature Communications. doi: 10.1038/ncomms3235

Travisano M, Lenski RE (1996) Long-term experimental evolution in Escherichia coli. IV. Targets of selection and the specificity of adaptation. Genetics 143:15–26.

Tsai IJ, Bensasson D, Burt A, Koufopanou V (2008) Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. Proceedings of the National Academy of Sciences 105:4957–4962. doi: 10.1073/pnas.0707314105

Turner CB, Blount ZD, Mitchell DH, Lenski RE (2015) Evolution and coexistence in response to a key innovation in a long-term evolution experiment with *Escherichia coli*.

Valverde S, Solé RV (2012) Evolved Modular Epistasis in Artificial Organisms. MIT Press, pp 111–115

Van Hoek P, Van Dijken JP, Pronk JT (1998) Effect of specific growth rate on fermentative capacity of baker's yeast. Appl Environ Microbiol 64:4226–4233.

Van Leuven JT, Meister RC, Simon C, McCutcheon JP (2014) Sympatric Speciation in a Bacterial Endosymbiont Results in Two Genomes with the Functionality of One. Cell 158:1270–1280. doi: 10.1016/j.cell.2014.07.047

Vestergaard M, Paulander W, Marvig RL, et al (2016) Antibiotic combination therapy can select for broad-spectrum multidrug resistance in Pseudomonas aeruginosa. International Journal of Antimicrobial Agents 47:48–55. doi: 10.1016/j.ijantimicag.2015.09.014

Vogwill T, Kojadinovic M, Furio V, MacLean RC (2014) Testing the Role of Genetic Background in Parallel Evolution Using the Comparative Experimental Evolution of Antibiotic Resistance. Molecular Biology and Evolution 31:3314–3323. doi: 10.1093/molbev/msu262

Wagner GP, Kenney-Hunt JP, Pavlicev M, et al (2008) Pleiotropic scaling of gene effects and the "cost of complexity." Nature 452:470–472. doi: 10.1038/nature06756

Wagner GP, Zhang J (2011) The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. Nature Reviews Genetics 12:204–213. doi: 10.1038/nrg2949

Wang X, Elston RC, Zhu X (2010) The Meaning of Interaction. Human Heredity 70:269–277. doi: 10.1159/000321967

Weinreich DM, Watson RA, Chao L (2005) Perspective: Sign Epistasis and Genetic Constraint on Evolutionary Trajectories. Evolution 59:1165. doi: 10.1554/04-272

Wenger JW, Piotrowski J, Nagarajan S, et al (2011) Hunger Artists: Yeast Adapted to Carbon Limitation Show Trade-Offs under Carbon Sufficiency. PLoS Genetics 7:e1002202. doi: 10.1371/journal.pgen.1002202

Wloch DM, Szafraniec K, Borts RH, Korona R (2001) Direct estimate of the mutation rate and the distribution of fitness effects in the yeast Saccharomyces cerevisiae. Genetics 159:441–452.

Wolf YI, Koonin EV (2013) Genome reduction as the dominant mode of evolution: Prospects & Overviews. BioEssays 35:829–837. doi: 10.1002/bies.201300037

Wu X, Altman R, Eiteman MA, Altman E (2014) Adaptation of Escherichia coli to Elevated Sodium Concentrations Increases Cation Tolerance and Enables Greater Lactic Acid Production. Applied and Environmental Microbiology 80:2880–2888. doi: 10.1128/AEM.03804-13

Xu J (2012) Rapid evolution in experimental populations of major life forms. In: Singh RS, Xu J, Kulathinal RJ (eds) Rapidly Evolving Genes and Genetic Systems. Oxford University Press, pp 40–52

Zeyl C (2003) An Evolutionary Advantage of Haploidy in Large Yeast Populations. Science 299:555–558. doi: 10.1126/science.1078417

Zhang J, Wagner GP (2013) On the definition and measurement of pleiotropy. Trends in Genetics 29:383–384. doi: 10.1016/j.tig.2013.05.002

Zinser ER, Schneider D, Blot M, Kolter R (2003) Bacterial evolution through the selective loss of beneficial Genes. Trade-offs in expression involving two loci. Genetics 164:1271–1277.