Epistasis and the evolutionary process

Alex T. Kalinka

PhD

University of Edinburgh

2006

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated in the text. This work has not been submitted for any degree or professional qualification except as specified.

Acknowledgements

I would like to thank my supervisors, Nick Barton and Nick Colegrave, for their insights and support over the years, and for allowing me the freedom to pursue many varied research interests. My work has also benefitted from the many conversations and critical appraisals of a number of other people, including Angeles de Cara, Jonathan Coe, Andy Gardner, Toby Johnson, Denis Roze, and Jay Taylor. In particular, Andy has introduced me to many fascinating areas of evolutionary biology and continues to inspire me with his passion for the subject.

Through the years, my social life in Edinburgh has been enriched by a great many people, most of whom are formidable drinkers (some more than others): Alex K, Alex W, Allan, Andy, Angeles, the Dans, Dave, David, Ellie, Esther, Jamie, Jitka, Jon, Neelofer, Penny, Stu, Tim, and Tom, to name a few. Special thanks are due to Penny for helping me move flat on several occasions, for our many amusing chats and for her willingness to forego sobriety in the name of friendship. Thanks also to Neelofer and Jon for their crucial help with homelessness and broken glass respectively. Thanks to Dave for driving me around Scotland on an algal hunting expedition. And thanks to Alex W for help with all the printing!

I should like to acknowledge the continuing support of my parents, Nuala and Gyuri, who have encouraged and supported me throughout my lengthy education.

I would also like to acknowledge the support of the BBSRC.

"Anybody ever tells you how this most beautiful and most evil of planets is somehow homogeneous, composed only of reconcilable elements, that it all *adds up*, you get on the phone to the straightjacket tailor."

- Otto Cone in *The Satanic Verses* by Salman Rushdie

Abstract

The role of epistasis in the evolutionary process has been the subject of much debate ever since the inception of the neo-Darwinian synthesis. Fisher believed that the average additive effect of an allele was the object of selection and Wright believed that epistatic selection would act upon groups of co-adapted alleles. Despite disagreements over the fundamental importance of epistasis in the process of evolution there has been a great deal of interest in how epistasis may influence certain biological adaptations such as recombination and canalisation and how epistasis contributes to speciation and even the origins of life. In this thesis epistasis is examined in the unicellular, motile chlorophyte *Chlamydomonas reinhardtii* through its effects on the recombination load arising from sexual crosses. In an experiment that combined divergent selection to opposite environments with measures of between and within envinronment recombination loads it was found that the build-up of incompatibilities between populations was not enhanced by divergent selection, and was instead the product of allopatric divergence together with genotype-by-environment interactions. Theoretical analysis is presented that shows that recombination facilitates the evolution of mutational robustness in a two-locus model. A review of eukaryotic chromosomal linkage patterns is presented and it is argued, together with simple theoretical considerations, that epistatic selection will not cause general, large-scale patterning of gene order within chromosomes. Additionally, two more general non-linear models are developed. The first shows that maternal effects may be an important component of mate choice in sexual selection models and the second demonstrates that marine snow and algal toxin production are social adaptations that can be understood as individual selection as well as group selection.

Contents

1 Introduction **8**

1.1 Overview 8

1.2 Epistasis in a historical context 8

1.3 The Fisher-Wright debate 10

1.4 General considerations of epistasis 15

1.5 The evolution of recombination 16

1.6 Speciation and co-adaptation 17

1.7 Evolution of canalisation 18

1.8 The origins of life 18

1.9 *Chlamydomonas reinhardtii* and experimental evolution 20

1.10 Chapter outline 21

2 Experimental evolution in *Chlamydomonas reinhardtii* **22**

2.1 *Chlamydomonas*, from genus to division 22

2.2 The ecology of *C. reinhardtii* 25

2.3 The Life-cycle 27

2.4 The growth-cycle in cultures of limited volume 29

2.5 Rates of Adaptation: New mutation versus standing variation . . . 43

2.5.1 Introduction 43

2.5.2 Methods 44

2.5.3 Results and discussion 46

2.6 Growth in used media 53

2.6.1 Outline 53

2.6.2 Results and discussion 54

3 Divergent experimental selection and recombination load in *Chlamydomonas reinhardtii* 56

3.1 Introduction 57

3.2 Methods 62

3.3 Results 66

3.3.1 Adaptation of evolved populations 66

3.3.2 Offspring RLs 71

3.4 Discussion 79

4 Recombination and the evolution of mutational robustness 85

4.1 Introduction 85

4.2 Model and analysis 90

4.2.1 Two-locus model 90

4.2.2 Multilocus population statistics 91

4.2.3 Selection 94

4.2.4 Mutation 95

4.2.5 Transmission 96

4.2.6 Invasion analysis 96

4.2.7 Evolution of robustness 97

4.3 Discussion 103

5 Epistasis and the evolution of chromosomal linkage patterns 109

5.1 The transition from prokaryotes to eukaryotes 109

5.2 Fisher, Mather, Kimura et al. 111

5.3 Evidence of clustering from whole genome studies 114

5.4 Specific examples of gene clustering 116

5.5 The probability of inversions clustering co-adapted alleles 119
5.6 The waiting time for a successful inversion 126
5.7 Summary 133

6 Maternal effects and sexual selection **134**
6.1 Introduction 134
6.2 Models and analyses 138
6.2.1 Two-locus model 138
6.2.2 Three-locus model 145
6.2.3 Mate choice based on self-similarity 148
6.3 Discussion 154

7 Marine snow and algal toxins: An evolutionary perspective **160**
7.1 Introduction 160
7.2 Models and analyses 164
7.2.1 Marine snow 164
7.2.2 Algal toxins 171
7.3 Discussion 176

8 Summary **180**
8.1 Experimental work 180
8.2 Theory 181

References **183**

A Appendix **222**

1 Introduction

1.1 Overview

In this chapter we will begin by giving a brief history of the development of thought regarding the role of epistasis in genetics and in evolution. Following on from this, we briefly discuss some of the major roles that epistasis may play in the evolutionary process. *Chlamydomonas reinhardtii* is then introduced as a model organism for use in several experimental evolution studies and a rationale for its use is provided. Finally, we end by detailing a chapter outline for the thesis.

1.2 Epistasis in a historical context

The term 'epistasis' was coined by the English geneticist William Bateson in 1909 to describe the distortion of simple two-locus Mendelian segregation ratios (Bateson, 1909). Table 1.1 shows some examples of these ratio distortions. In this context epistasis was defined as the phenomenon whereby the allelic effects of one locus are masked by the allelic effects of another distinct locus, and indeed the word can be literally interpreted as meaning 'standing upon' (Phillips, 1998). The origin of intellectual interest in this phenomenon was therefore rooted in the study of discrete, relatively simple Mendelian traits. However, Ronald Fisher's demonstration that the inheritance of continuous (biometrical) traits was entirely consistent with the segregation of multiple loci following Mendel's laws of inheritance opened up the question of epistasis to the study of quantitative traits (Fisher, 1918).

The underlying implication for both discrete and continuous traits was that the phenotype of an individual cannot always be determined by summing the effects of separate loci. The inherent complexity of polygenic traits meant that epistasis had to encompass

Table 1.1. Distortion of the classic Medelian two-locus 9:3:3:1 segregation ratio by different forms of epistasis. Adapted from Phillips (1998).

Interaction Type	Diploid Genotype			
	$A-B-$	$A-bb$	$aaB-$	$aabb$
Classical ratio	9	3	3	1
Dominant epistasis	12		3	1
Recessive epistasis	9	3	4	
Duplicate genes with cumulative effect	9	6		1
Duplicate dominant genes	15			1
Duplicate recessive genes	9	7		
Dominant and recessive interaction	13		3	0

almost any type of interaction between groups of loci conceivable. A lack of information about the underlying genetics of polygenic traits also meant that epistasis was interpreted within a statistical framework. In practice this means that any genetic variance contributing to the phenotypic variance of a trait that cannot be attributed to the additive genetic variance of the trait is defined as non-additive genetic variance (e.g. Lynch and Walsh, 1998). Non-additive genetic variance can be further decomposed into dominance and epistatic genetic variance, which essentially comprise the non-selectable genetic effects.

Developing general models of epistasis with a manageable number of parameters is particularly difficult when epistasis can refer to any type of gene interaction. The complexity of this situation was first tackled by Cockerham (1954) and Kempthorne (1954) by partitioning the epistatic variance into terms describing different types of

gene interaction such as additive-by-additive, additive-by-dominance and dominance-by-dominance epistasis.

These statistical descriptions of epistasis are dependent on the allele frequencies of the loci in question and therefore do not inform about the physiological nature of the gene interactions. Hence, populations will only exhibit substantial levels of epistatic variance when the alleles of the interacting loci are at intermediate frequencies. At mutation-selection balance, where alleles are near fixation (assuming a low per locus mutation rate), almost all of the variance will be expressed as additive genetic variance even if the genes underlying the phenotype are entirely epistatic (Whitlock et al., 1995). In this scenario, the epistatic variance is converted into heritable additive genetic variance (Goodnight, 1987, 1988, 1995; Cheverud and Routman, 1995). Therefore, models that separate additive and epistatic effects do not capture their non-independence (Wagner et al., 1998).

1.3 The Fisher-Wright debate

The role of gene interaction in evolution is at the centre of the debate between the Fisher and Wright paradigms of evolutionary change. Whilst the mathematical details of Fisher and Wright's theoretical studies were largely in agreement, the conceptual emphases of the two were in stark contrast. Much of what differed in their views can be attributed to differences in their respective academic backgrounds. Fisher approached the subject of natural selection with a formal training in physics and mathematics and this grounding in the more exact sciences heavily influenced his conceptual treatment of natural selection (Provine, 1971).

By drawing an analogy between the statistical mechanics of populations of gas molecules and that of populations of genes, Fisher was espousing his belief that evo-

lutionary change could be adequately understood and described without excessive attention paid to the details of the 'particles' involved (Fisher, 1922). Contained within those details are the effects of gene interaction and random genetic drift. Fisher's view of epistasis was a statistical one that laid emphasis on the generation of non-linear sources of non-heritable variation and was not concerned with the mechanistic or physiological details of gene interaction. Consistent with these beliefs Fisher envisioned large, panmictic populations and smooth adaptive landscapes, each with a single global optimum.

Fisher's fundamental theorem of natural selection states that the rate of increase of the mean fitness of a population is equal to the additive genetic variance for fitness of that population, and was likened by Fisher himself to the second law of thermodynamics in physics, indicating his belief that biology could be reduced to laws as universal and regular as those developed in physics (Fisher, 1930). In Fisher's view the constant deterioration of the environment would guarantee the continual replenishment of additive genetic variance for natural selection to act upon.

The lack of any formal modeling of the effects of epistasis by Fisher can therefore be interpreted as an omission of details, the effects of which Fisher considered negligible in large populations. In large, panmictic populations without very strong linkage between loci, the formation of gene combinations would be constantly deteriorated by recombination, making the additive effect of an allele across all genetic backgrounds the object of selection.

Sewall Wright's background was distinctly biological with his first position in the physiological genetics laboratory of William Castle (Provine, 1971). The work he conducted with Castle investigated the inheritance of coat colour in guinea pigs and strongly influenced his view that gene interactions and pleiotropy are universal, inevitable properties of genomes. The shifting balance theory (SBT) of evolution was developed by

Wright as an alternative to Fisher's mutation and mass selection model, and incorporated gene interaction, random genetic drift and population structure, reflecting Wright's biological approach (Wright, 1929, 1931, 1932).

The significance of epistasis in the SBT was explained by Wright using the conceptual tool of an adaptive landscape. On Wright's adaptive landscape a population can inhabit a single point described on the y-axis by its mean fitness and on the x-axis by its genic frequency, which is the array of frequencies of the different alleles of loci contributing to fitness. According to Wright, with extensive epistasis this landscape will be multiply peaked and with pleiotropy the peaks will be of differing heights (Wright, 1977).

In this scheme there is a global optimum, but populations tend to adapt locally and therefore reach local fitness peaks. The key process, therefore, is how populations cross adaptive valleys to reach the global optimum. In the SBT different genic combinations will be fixed in different demes due to the effects of random genetic drift and selection. Interdemic selection can then opearte via migration from the fittest peaks to demes on lower peaks resulting in the crossing of adaptive valleys towards the global optimum. The objects of selection here are not single genes, but combinations of genes or as Wright put it 'the set of genic frequencies' (Wright, 1963). The model illustrates Wright's concept of a complex genotype-phenotype map and the structuring of populations into many small demes within a metapopulation.

The contention between the models has still not been fully resolved, and advocates of both theories continue to debate the issue (Coyne et al., 1997; Goodnight and Wade, 2000; Coyne et al., 2000). Despite the disagreement over the theories the presence of reproductive barriers between two diverged species is generally believed to be reasonable evidence for the existence of alternative adaptive peaks separated by a valley of

lower fitness (Dobzhansky, 1936; Muller, 1939). The crucial question is whether the presence of an adaptive valley between two diverged populations is an indication of adaptive peaks and valleys within a population. Fisher argued that adaptive landscapes are likely to be smooth because in a high-dimensional space local stable equilibria are unlikely (Fisher, 1930). Wright proposed the SBT to explain how a population could traverse intra-population valleys in the adaptive landscape assuming the existence of a rugged landscape as a given, and in this sense the SBT is a contrived answer to an assumed problem.

The question of the nature of multidimensional adaptive landscapes has been resurrected, however, and most notably by Gavrilets (1997). 'Holey adaptive landscapes' (HALs) represent the fitness surface of a high-dimensional genotype space and are somewhat more intuitively appealing than adaptive landscapes defined by arrays of gene frequencies (Gavrilets, 1997). HALs are an alternative to the more usually discussed 'rugged landscapes', such as those envisioned by Wright (1963, 1977) and Kauffman and Levin (1987). HALs are multidimensional flat surfaces pitted with many holes where genotypic fitnesses are randomly assigned to either one (viable) or zero (non-viable) (Figure 1.1).

The main consequence of this scheme is that viable genotypes are connected by ridges of equally viable genotypes and for this reason the HAL has been compared to the Dobzhansky-Muller model of hybrid incompatibilities arising between diverged populations without either having to cross an adaptive valley (Dobzhansky, 1936; Muller, 1939). Speciation on HALs, therefore, involves no temporary reduction in fitness and no crossing of an adaptive valley. A major drawback of the model is that the landscape is uncorrelated, meaning that a single mutation between two genotypes can result in a total reversal of fitness on the landscape. This would seem to be at odds with the general

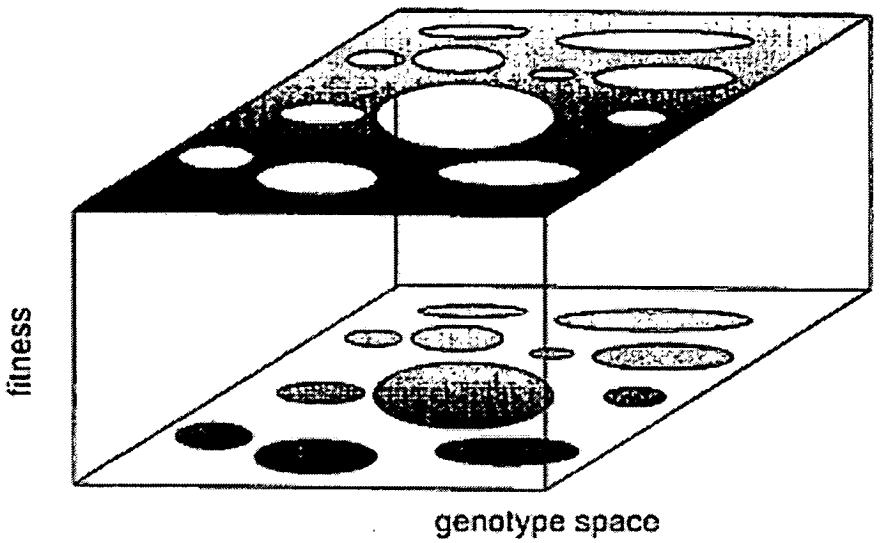


Figure 1.1. The holey adaptive landscape (HAL) proposed by Gavrilets (1997). See text for details.

observation that increasing phylogenetic distance tends to increase epistatic interactions for fitness, i.e. within population crosses reveal very little epistasis, but between population and species crosses reveal extensive epistasis (Whitlock et al., 1995). In this scenario one could imagine large, flat regions of genotype space where populations cluster, separated by gradual, but deep valleys, and also connected by ridges of fit genotypes (Figure 1.1).

Another simplifying assumption is that the fitnesses can be assigned as either zero or one, and the consequent flatness of the fitness surface, an assumption introduced to simplify the study of the large-scale dynamics of the model. The lack of hills would seem to imply that populations are free to drift across the surface and would not be increasing in mean fitness at any point. However, differing fitnesses can be incorporated into the model, but due to the multidimensional nature of the genotype space they do

not manifest as hills and are not easy to represent in just three dimensions. Further to this, the presence of just two fitness classes may be applicable to systems such as the secondary-structure folding dynamics of RNA sequences, where the folded molecules either form the correct structure and are therefore fit, or they do not and are unfit, thus forming what has been termed a neutral network (van Nimwegen et al., 1999).

1.4 General considerations of epistasis

There are two major effects of epistasis in natural populations, and these effects are central to most theories that are concerned with epistatic effects in evolution. The first is that epistasis will always generate linkage disequilibria (LD) between loci via selection. This is also true in quantitative genetic models of traits under stabilising selection, where individual loci combine additively, because stabilising selection favours combinations of plus and minus alleles at different loci (intermediate optimum) (e.g. Lewontin, 1964b).

The two forms of epistasis commonly discussed, synergistic (negative) and antagonistic (positive) epistasis, generate negative and positive LD respectively. In stabilising selection models the negative LD is adaptive, but in multi-locus models of directional selection negative LD retards adaptation by causing associations between beneficial and deleterious alleles thereby reducing the additive genetic variance for fitness. Positive LD arises when selection favours certain combinations of alleles over others and in this sense is considered adaptive.

The second effect of epistasis is that it makes the additive effects of genes depend on allele frequencies. This effect underpins much of the differences between Fisher's and Wright's views of evolution. In Fisher's scheme the average additive effect of an allele over all genetic backgrounds was what mattered, but Wright saw groups of alleles being co-selected in combination with genetic drift. The role of drift is key to changing

additive effects through essentially random allele frequency changes, supposedly acting to release additive genetic variance in bottlenecked populations from previously non-heritable epistatic genetic variance (Goodnight, 1987, 1988).

1.5 The evolution of recombination

Epistasis is central to a number of deterministic models for the evolution of recombination. Kimura and Maruyama (1966) showed that the mutation load of a sexual population will be less than that of an asexual when there are synergistic interactions between deleterious mutations. Synergistic interactions mean that the fitness of deleterious mutations are worse than expected when together than when they are on their own. This generates negative LD, and recombination acts to break this up, freeing up variation for selection to act on and remove deleterious mutations from the population resulting in a lower mutation load at equilibirum than an asexual population (Kondrashov, 1984; Charlesworth, 1990).

The same argument applies for populations that are adapting to an environment and so are subject to directional selection. If there is weak synergistic epistasis between beneficial alleles that are sweeping through a population then their progress will be retarded by the build-up of negative LD. Once again recombination frees up additive genetic variance for fitness and a modifier that increases the rate of recombination can hitchhike with the high fitness variants that it will generate (Barton, 1995). This basic idea was formulated by Weismann (1889) although without any knowledge of epistasis and LD. Maynard Smith (1980, 1988) and Charlesworth (1993) found that when a trait is under stabilising selection with a moving optimum modifiers of recombination will be able to track the optimum quicker as they will hitchhike on the variance that they generate.

1.6 Speciation and co-adaptation

Speciation involves antagonistic (positive) epistasis. Dobzhansky (1936) and Muller (1939) showed that genetic incompatibilities could arise between populations if they fixed separate alleles at different loci that were incompatible when brought together. These Dobzhansky-Muller incompatibilities prevent a population from having to go through an adaptive valley to achieve reproductive isolation. The incompatibility arises from antagonistic epistasis that generates positive LD between loci. That is, beneficial combinations of alleles will be found together more often than expected from their allele frequencies. These combinations are known as co-adapted gene complexes when they occur within populations (Dobzhansky, 1970).

More recently there has been a renewed interest in the subject of co-adaptation and selection for closer linkage. Following the work of Dobzhansky (1970) a study of the paracentric inversions on the third chromosome of *Drosophila pseudoobscura* suggests that there has been extensive epistatic selection acting to maintain linkage disequilibrium between loci within the inversions (Schaeffer et al., 2003).

Another study has examined chromosomal inversions in the apple maggot fly, *Rhagoletis pomonella*, and has identified six allozyme loci, already known to show host-associated differentiation, within three inverted chromosomal regions in the genome (Feder et al., 2003). There is a potentially compelling argument that the presence of these genes within recombination-suppressed chromosomal inversions has contributed to the sympatric speciation due to host-adaptation believed to be ongoing in this species. This does not necessarily imply that the inversions are composed of coadapted gene complexes since the findings are still consistent with the possibility that each inversion contains only one locus subject to selection with a host of neutral markers hitchhiking along and consequently exhibiting linkage disequilibrium.

1.7 Evolution of canalisation

Waddington (1940) and Schmalhausen (1949) recognised that the resistance of embryos to environmental and genetic perturbations during development was an adaptation to ensure the proper development of the embryo. It was realised that stabilising selection was acting on the developmental traits making any deviations from the optimum deleterious. It was their contention that this force benefits any mechanisms that prevent perturbations from resulting in deviations from the optimum. Waddington (1942) described the traits as 'canalised'.

This is closely related to the evolution of recombination because canalised traits are extreme forms of synergistic (negative) epistasis. Past some threshold in genetic perturbations (mutations) and fitness will drop off to zero, but above that threshold perturbations are tolerated.

1.8 The origins of life

Epistatic interactions between the earliest (probably RNA) replicators allowed for an increase in the evolution of information and complexity. The argument is as follows. Exponential replication of early replicators together with the constraint of a conservation of total population size enabled effective selection of the 'fittest' individuals with the highest replication rates, but also results in, what ecologists would term, 'competitive exclusion' of all but the fittest replicator quasi-species (Eigen and Schuster, 1979; Szathmary, 1999). Such a lack of species co-existence acts as a constraint on the evolution of primitive groups of co-operating replicators that is necessary for an increase in information which accompanies the evolution of early genomes. The 'error threshold' of replication imposes a limit on the maximum nucleotide length of the early replicators because of loss of the information through mutation.

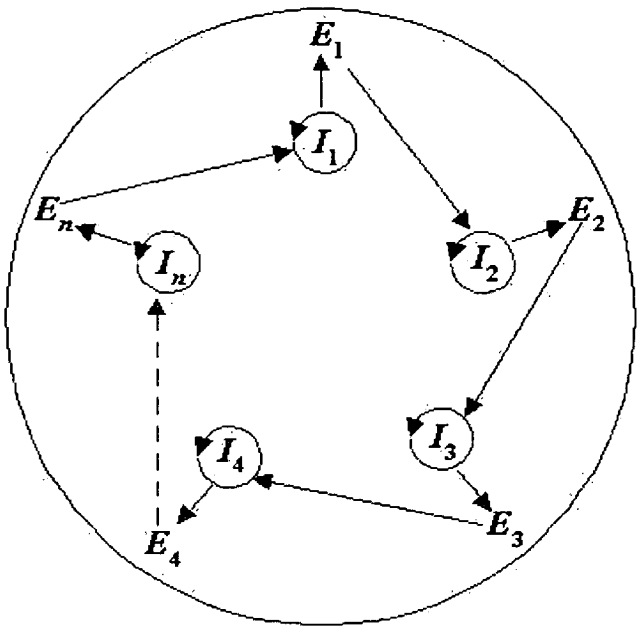


Figure 1.2. The hypercycle of Eigen and Schuster (1979). The I's are RNA molecules and the E's can either be coded enzymes or the enzymatic function of the RNA molecule. The system is a closed loop that ensures cooperation between replicators and does not violate the error threshold.

These problems combined together are known as 'Eigen's paradox', and can be summarised as: competitive exclusion between replicators prevents coexistence of different types and the error threshold forbids increasing information by simply ligating together independent replicators, so how may information increase? Eigen and Schuster (1979) attempted to solve it by recognising that for any increase in information to be possible there must be some form of functional linkage between replicators (Eigen and Schuster, 1979).

Functional contingency between individual replicators ensures the co-existence of different types, and thereby increases the information capacity of the co-operating group without stepping over the error threshold. The functional linkage proposed by Eigen is

in the form of a closed loop, where each replicator autocatalyses their own replication and also provides catalytic help for the replication of their neighbouring replicator in the cycle, and is called a 'hypercycle' (Figure 1.2).

An important consequence of the hypercycle is that whilst competition between distinct replicators within the hypercycle has been abolished, competition between quasi-species alleles of different hypercycles continues to ensure the maintenance of the information of the wild-type allele against mutation accumulation: therefore, both cooperation and competition are required. The positive interactions between the replicators could be thought of as antagonistic epistasis, which by coupling the fitnesses of the replicators together has ensured that cooperation is maintained and information can increase.

1.9 *Chlamydomonas reinhardtii* and experimental evolution

Two chapters in this thesis describe experimental evolution with populations of *Chlamydomonas reinhardtii*. The unicellular, motile chlorophyte is ideal for experimental evolution related to epistasis as the species is facultatively sexual and so allows us to mate different individuals together and make comparisons of offspring and parental fitnesses to find the signatures of epistasis.

Additionally, *C. reinhardtii* is haploid and has a simple life-cycle that is straightforward to manipulate. The species typically has a short generation time and the ability to keep them growing exponentially in very large populations ($>10^6$ ml^{-1}) means that adaptation and evolution will take place very rapidly. The species has minimal requirements: will grow on a medium of inorganic salts with light, and will even grow in the dark with acetate as an alternative carbon source.

1.10 Chapter outline

Chapters 2 and 3 are devoted to experimental evolution in *C. reinhardtii*. Chapter 2 describes the growth cycle of the alga in cultures of limited volume and in different vessels. Additionally, two pilot experiments are discussed. Chapter 3 describes an experiment that combines experimental evolution to divergent environments together with a measure of recombination load in crosses between evolved lines. In chapter 4, a multilocus model of the evolution of genetic robustness (canalisation) is analysed. In chapter 5, the evidence for epistasis causing clustering of loci in genomes is reviewed and some simple models are developed to support the conclusions. The last two chapters involve more general cases of non-linearity in evolution. In chapter 6, a model of maternal effects and sexual selection is analysed. Maternal effects are sources of non-linear interactions between maternal genotype and offspring phenotype. In chapter 7, a model of marine snow and algal toxin production is analysed. The social interactions between individuals in these models is akin to interactions between loci in epistatic models.

2 Experimental evolution in *Chlamydomonas reinhardtii*

Abstract

This chapter introduces the unicellular, motile chlorophyte *Chlamydomonas reinhardtii* as a model organism for experimental studies of the evolutionary process. A brief background of the evolutionary relationships of the genus and of the known ecology is outlined to develop the appropriate context in which to interpret experimental observations and results. Laboratory growth-cycles in different vessels and in different media are then described as a precursor to assaying fitness. Finally, the results of two pilot experiments are discussed. The first is a study of the effects of bottleneck size and base population heterogeneity on the rates of adaptation to a novel environment and the second is a cursory investigation of the growth-cycle in used media.

2.1 *Chlamydomonas*, from genus to division

Chlamydomonas is a genus of unicellular, motile chlorophytes that inhabit a diverse array of habitats ranging from ponds, soils and oceans, to the snow on mountain tops. The cells have a cell wall composed of glycoprotein, two anterior flagellae, a chloroplast containing chlorophyll *a* and *b*, a pyrenoid where starch is stored and a light-sensitive carotenoid-pigmented region ('the eye') that perceives light. There are more than 450 species described at present, but with many more likely to be discovered (Ettl, 1976).

Chlamydomonas is the type genus of the family Chlamydomonadaceae that contains around 800 species in 33 genera, over half of which belong to *Chlamydomonas* (Jakubiec, 1984). This family is a member of the order Chlamydomonadales which in turn belongs to the monophyletic class Chlorophyceae, a class which boasts a range of morphologies, both motile and the derived non-motile, with the coccoid form being the

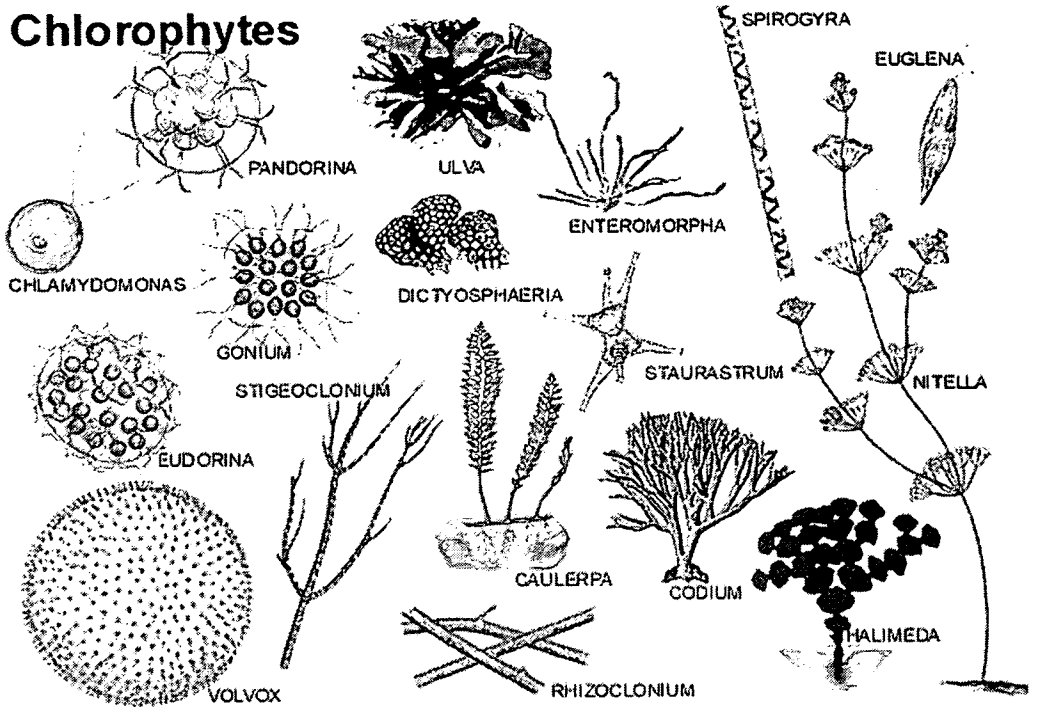


Figure 2.1. A sample of the morphological and evolutionary diversity of the Chlorophytes.

most abundant (Lewis and McCourt, 2004). Familiar genera in this class include the colonial *Volvox* and the coccoid *Chlorella*.

The Chlorophyceae belong to the division (= phylum) Chlorophyta (green algae *sensu stricto*) within which there are approximately 500 genera and more than 10^4 species identified to date (van Den Hoek et al., 1995; Bold and Wynne, 1985). The term 'algae' can be used to describe a myriad of different photosynthetic species that are not phylogenetically related, from the prokaryotic cyanobacteria (blue-green algae) to the eukaryotic dinoflagellates, diatoms, rhodophytes and giant kelp, all of whom are distantly related to one another and none of whom belong to the green algae.

The Chlorophytes are photosynthetic eukaryotes that utilise chlorophyll *a* and *b* as well as certain beta-carotenes and xanthophylls and store their energy in the form

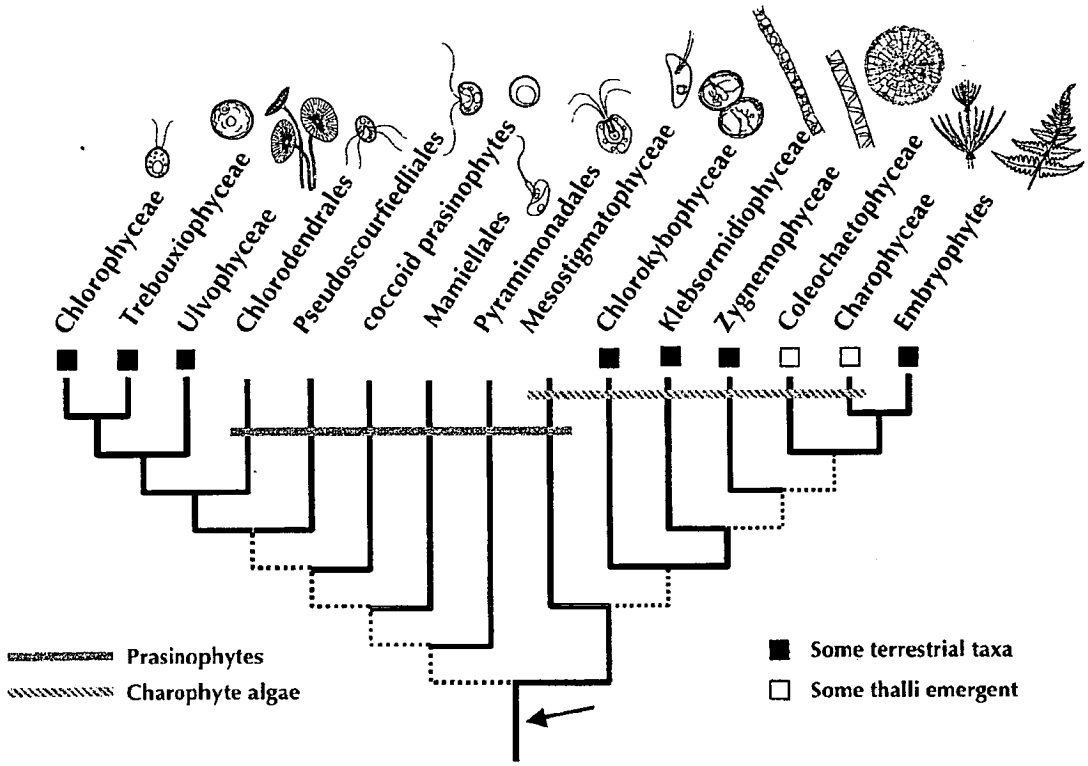


Figure 2.2. An illustrative overview of the phylogenetic relationships of the 'green algae' showing the two main lineages or clades, the chlorophytes, containing the prasinophytes, and the charophytes, which are paraphyletic to the embryophytes or higher plants (from Lewis and McCourt (2004)).

of starch, three characteristics that place them as close relatives of the embryophytes (higher plants). This division exhibits a great diversity of body morphology, ranging from free-swimming unicells to complex, multicellular thalli with colonial and filamentous types in-between (Figure 2.1). The majority of the species within the division inhabit freshwater environments. Some of the more familiar genera include *Ulva* (sea lettuce), *Desmidium* (desmids), *Spirogyra* and *Euglena*. The phylogenetic relationships of the green algae are summarised in Figure 2.2.

2.2 The ecology of *C. reinhardtii*

C. reinhardtii cells can commonly be found growing in soil or freshwater puddles and ponds. Species in this genus are also known to grow as phytoplankton in the euphotic region of the water column in lakes and the still waters of meandering rivers, and also on the surface of snow. The motility and positive phototaxis of the cells indicates that they prefer quiet waters where they are not at the mercy of turbulence and physical mixing (marine algae are often non-motile). The curious origin of the principal laboratory strains of *C. reinhardtii* has been determined as a single zygospore taken from a potato field in Massachusetts in 1945 (Harris, 1989). *C. reinhardtii* cells are capable of heterotrophic growth in the dark using acetate as a sole carbon source (Sager and Granick, 1953), and it is known that algae capable of growth in the dark are largely soil-derived species (Danforth, 1962; Lewin, 1963). For these reasons we can be reasonably confident that *C. reinhardtii* is a soil alga that is also content to grow in more water-logged conditions.

The soil algal flora is typically less speciose than the freshwater epipelon (sediment surface) and phytoplankton (Round, 1981). It has been suggested that phototactic (and geotactic) motility may be important for epipelon species incase of burial (Round, 1981), and this may also be the case in (moist) soils. While the fast-growing cyanobacteria predominate on slightly alkaline, nutrient-rich soils, the green algae tend to prefer more acidic, nutrient-poor soils (Metting, 1981).

A major factor affecting soil algal growth is the level of moisture in the soil. Together with diurnal fluctuations of dew deposited overnight there will also be seasonal variations in rainfall, which overall provide an uncertain future environment. There is evidence from the field and the laboratory that diatoms are less tolerant to low water potential and are more vulnerable to dessication than green algae (Bristol-Roach,

1928). One account reported that dessicated *C. reinhardtii* spores survived a period of 66 years (Bristol, 1920). The resistance of the Chlorophyte zygospore stage (Figure 2.3) has been described as a potential adaptation to stressful, uncertain environmental conditions (Coleman, 1983). Soil algal populations must also tolerate a broad range of temperatures (compared to phytoplankton species) as freeze-thaw cycles and high mid-day temperatures can be common features of the soil environment (Metting, 1981).

There is little to no data regarding the global distribution of *C. reinhardtii* although it is known that the genus is of world-wide distribution. In general, freshwater algal species tend to have world-wide distributions with surveys suggesting that between 50 and 70% of unicellular, freshwater species may be considered cosmopolitan species (although many species that are distributed throughout the northern and southern temperate zones are absent from the tropics) (Round, 1981) . This is in contrast to marine species that tend to have far narrower, more distinct distributions, which may be attributed to the barrier to dispersal posed by high water temperatures (Round, 1981). Soil and freshwater algal species are likely to enjoy a great deal of aerial dispersal via resting spores being blown upwards from dried soils and transitory puddles or ponds. Unlike marine species, wind-borne freshwater spores are likely to survive high or low air temperatures and can remain viable for many decades. *Chlamydomonas* species were recovered from an altitude of 1100 metres after petri dishes were exposed to the air for just one minute (Brown et al., 1964). Typically, the majority of air-borne algae recovered from the atmosphere are soil algae (Maguire, 1963; Brown et al., 1964; Behre and Schwabe, 1970).

In summary, *C. reinhardtii* cells have evolved to cope with uncertain environmental conditions (fluctuating moisture levels and temperatures) and low levels of nutrients. The distribution of the species world-wide is likely to be wide, but within the extremes

of tropical and frozen habitats. Taken together these life-history details would suggest that *C. reinhardtii* is an algal generalist that is robust to environmental fluctuations at the same time as being able to exploit suitable conditions as an opportunistic disperser. Growth in nutrient-poor soils will have selected for the ability to efficiently acquire and assimilate limited resources in a competitive arena.

It is not clear, however, whether nutrient-limited environments will have presented a strict growth-reproduction trade-off in the life-history strategies of *C. reinhardtii* such as have been documented for many multicellular species (MacArthur, 1962; Cody, 1966; MacArthur and Wilson, 1967; Pianka, 1970; Roughgarden, 1971; Charnov and Schaffer, 1973). Low nutrient environments may place a demand on motility for the active recruitment of nutrients, the rigours of which place an upper limit on cell size. Added to this is the temporal instability of the soil environment (although the rate of change may be low relative to the algal generation time). High dispersibility and environmental instability are considered to be characteristics associated with the life-cycles of so-called 'fugitive' species that colonise available environments opportunistically, but such species are not necessarily r-strategists in the classical sense as they must allocate resources into non-reproductive pursuits to ensure dispersal and survival (Baker, 1965; Gadgil and Solbrig, 1972).

2.3 The Life-cycle

The life-cycle of *C. reinhardtii* is summarised in Figure 2.3. The haploid, vegetative cells will grow asexually (biphasic development: cell growth and motility followed by loss of motility and cell division) until they encounter a stressful environment (lacking available nitrogen) when they will undergo gametogenesis and differentiate into one of two isogametic mating types, designated pluses and minuses (heterothallic). This aspect

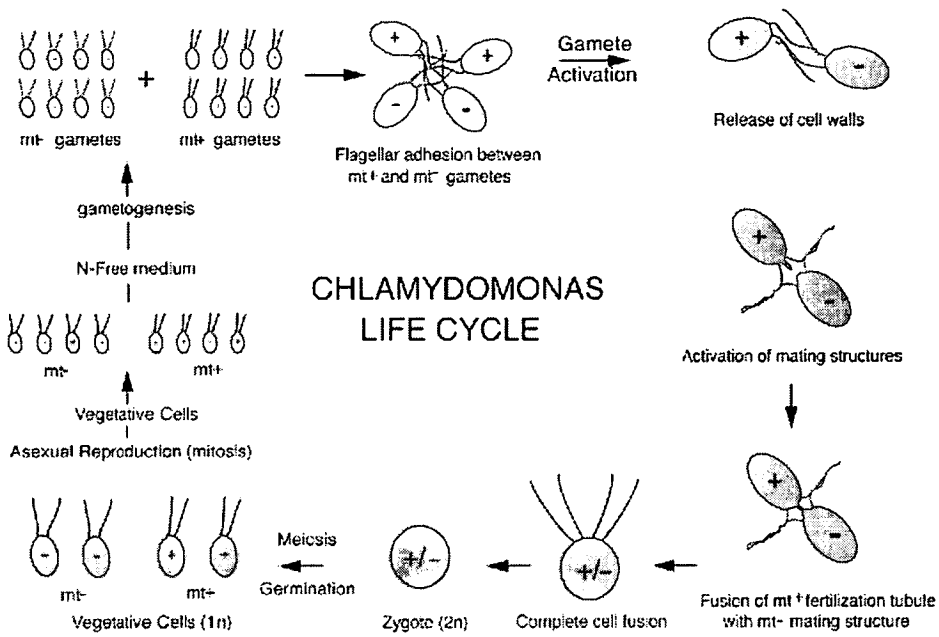


Figure 2.3. The life-cycle of *Chlamydomonas reinhardtii* showing both asexual (vegetative) and sexual phases.

of their life-cycle makes them a facultatively sexual species, which is an attractive characteristic for experimental evolutionists. Opposite mating types fuse forming a diploid zygote that sheds the flagellae forming a thick, resistant wall that can protect the zygote during adverse conditions (dessication, freezing). This stage can be viewed as a resting spore or akinete (Coleman, 1983). When the environment improves (available nitrogen is replaced in the media) the zygote undergoes meiosis forming four haploid zoospores with two of each mating type.

The mating type locus is located on linkage group VI, situated within a one megabase (Mb) long, recombination-suppressed region (Gillham, 1969; Ferris and Goodenough, 1994). In the centre of this region is the R-domain, a 200 Kb region that has undergone

numerous translocations and inversions resulting in recombination-suppression over the entire area (Ferris and Goodenough, 1997; Ferris et al., 2002). Three genes that are selectively transcribed in response to nitrogen-starvation have been identified within the R-domain: *fus1*, a glycoprotein necessary for gametotype fusion (Ferris et al., 1996), the *mid* gene, a transcriptional regulator that induces gametotype differentiation in response to nitrogen-starvation (Ferris and Goodenough, 1997), and *sad1*, a flagellar sexual agglutinin necessary for mating type recognition and fusion (Hwang et al., 1981). Linkage analysis in the nuclear genome of *C. reinhardtii* has identified 17 linkage groups in a total genome size of 10^8 bp (100 Mb) (Harris, 1989).

C. reinhardtii is also known to form palmelloid colonies, which are small groups of adhered, non-motile cells, and this stage is often associated with the production of gelatinous extracellular exudates (Harris, 1989). Curiously, such stages will sometimes revert to the motile stage upon the addition of fresh culture media or by using media that is low in phosphate levels and/or acidic (Bold and Wynne, 1985; Harris, 1989). The palmelloid stage may represent a semi-resting stage when nutrients do not need to be scavenged via energy-expensive motility (high phosphates) or when the nutrients have been depleted (old culture media). Phosphates tend to be most readily available in neutral or alkaline soils. Interestingly, it has been noted that almost all of the algal species that require the very scarce vitamin B_{12} are motile (Hutchinson, 1967).

2.4 The growth-cycle in cultures of limited volume

The growth of microbes in cultures of limited volume typically follows a characteristic pattern and unicellular algae are no exception to this rule. The pattern of growth is illustrated in Figure 2.4. In the lag phase (A), the algae are adapting physiologically to the change in environmental conditions experienced by the inoculated cells (Black,

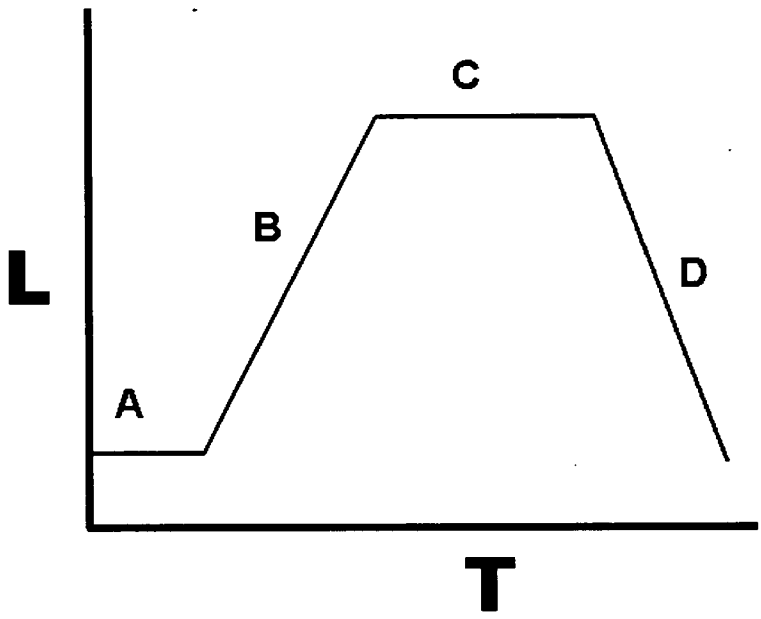


Figure 2.4. The characteristic pattern of microbial growth in cultures of limited volume. A - lag phase, B - log phase, C - stationary phase, D - death phase. T - time, L - log(cell no.).

1996; Fogg, 1975). In the log phase (B), the cells are growing exponentially, in the stationary phase (C) the carrying capacity of the culture is reached and in the death phase (D), the nutrients have been exhausted and cell death may be precipitated by the accumulation of toxic waste products.

C. reinhardtii will grow in a simple medium of inorganic salts together with a light source. For all of what follows cultures were grown in a Sanyo Orbital floor incubator with continuous shaking and illumination in sterile Bold's culture media (Bold, 1942). The growth of cultures can be measured using the optical absorbance of the culture in a spectrophotometer at 665 nm. However, before utilising this method it is necessary to calibrate the spectrophotometer for the culture vessel and media that will be used so that a conversion factor can be derived to convert optical absorbance into cell numbers.

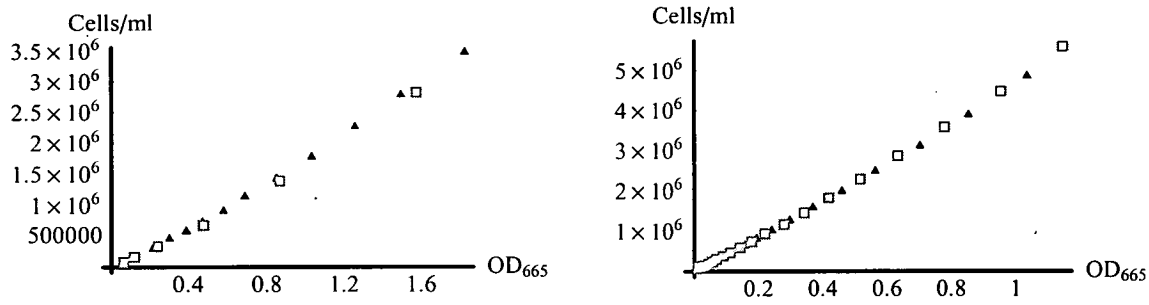


Figure 2.5. Spectrophotometer calibration curves for *C. reinhardtii*. On the left is the curve for 25 x 150 mm glass (borosilicate) culture tubes and on the right is the curve for 1.5 ml plastic cuvettes. OD_{665} - Optical Density at 665 nm.

Figure 2.5 shows calibration curves for 10 ml cultures in glass tubes and in cuvettes.

Two replicate 10 ml cultures were serially diluted and one of the lowest dilutions was counted three times on a haemocytometer under a microscope and the average was taken. All of the dilutions were vortexed and the absorbance at 665 nm was measured. The gradient of the resulting lines is taken as the conversion factor assuming a zero intercept. For 25 x 150 mm glass tubes the conversion factor is 1798560 cells/ml OD_{665}^{-1} and for 1.5 ml cuvettes the conversion factor is 4321900 cells/ml OD_{665}^{-1}. Two things can be noticed from these curves. Firstly, the cuvettes appear to give a more consistent estimate of cell density and secondly, the cuvettes can estimate the cell density over a larger range, from less than a million cells/ml up to and above 6 million cells/ml, whereas the glass tubes have reached the maximal absorbance of 2 at just 3.5 million cells/ml. The drawback of using cuvettes is that they are an invasive method that require the growing cultures to be opened to the air as well as necessarily reducing their volume.

The broad range of cell density estimates capable with cuvettes enables us to examine the growth cycle of *C. reinhardtii* over a longer period of time than would be pos-

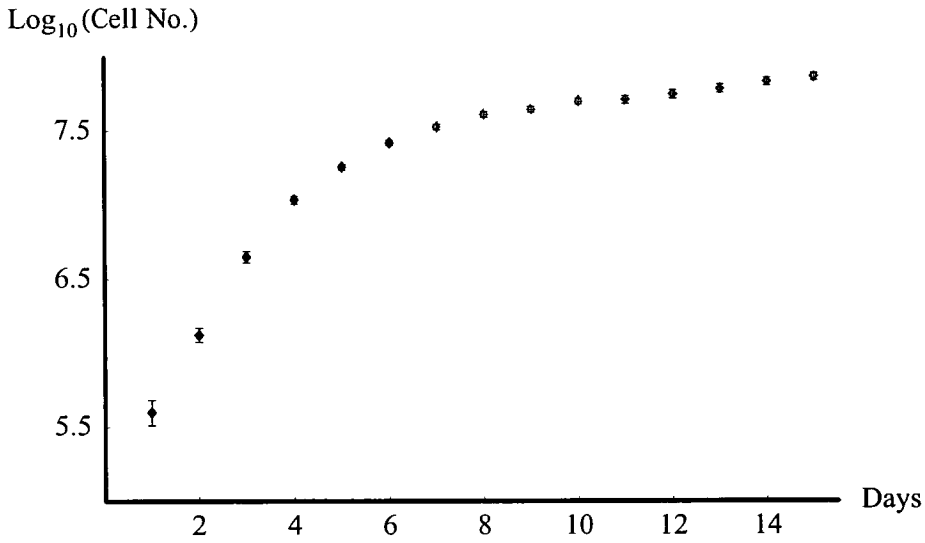


Figure 2.6. The growth curve of twelve *C. reinhardtii* clones over a period of fifteen days. Cell density estimates were measured in glass tubes from days one to six (black symbols) and in cuvettes from days seven to fifteen (grey symbols). Error bars are plus and minus the standard error of the mean.

sible with just glass tubes, and also allows us to compare our independently-estimated conversion factors. To this end, the growth of twelve different, independently-cultured clones (six plus and six minus) was measured over a period of fifteen days from a starting inoculum of 5 x 10^4 cells. Figure 2.6 shows the logarithm of the average cell density on each day, plus and minus the standard error of the mean.

The curve shows that there is a good agreement between the conversion factor estimated from glass tubes (days one to six) with the conversion factor estimated from cuvettes (days seven to fifteen). The curve also shows that, while exponential growth slows considerably over the first six days, there is still some growth after day six which is above that expected for a culture in stationary, or indeed, death phase. This may in-

dicate that the culture has not exhausted the nutrient supply and started to accumulate toxic waste products, but that there is perhaps self-shading occurring as a result of the high cell density, which results in a light-limited state of growth.

It is possible to measure the average exponential growth rate for these twelve clones from this data, but before doing so it may be helpful to explain the rationale. An exponentially growing population may be described with the following differential equation

$$\frac{dN}{dt} = rN \tag{1}$$

where r is the intrinsic growth rate, which is the per capita birth rate minus the per capita death rate, and N is the population size. This equation can be solved explicitly to give the population size, N_t, at any point in time in terms of the initial population size, N_0

$$N_t = N_0 e^{rt}. \tag{2}$$

This solution can be rearranged to find r

$$r = \frac{1}{t}(\ln N_t - \ln N_0). \tag{3}$$

Therefore, to measure r accurately requires that the time-points we choose for N_t and N_0 capture only the exponential portion of the growth curve. To determine this one could simply look for the most linear portion of the log plot in Figure 2.6, but this may not give the best results as at the lowest densities the spectrophotometer may not be sensitive enough to give accurate readings. It seems reasonable to expect a positive correlation between the growth rate and the final cell density (on day fifteen) and this measure can be used to determine the most appropriate days over which to calculate r:

	Days			
	1,2,3	1,2,3,4	2,3	2,3,4
Correlation	0.0795 ($p > 0.4$)	0.199 ($p > 0.25$)	0.247 ($p > 0.2$)	0.493 ($p < 0.05$)

It appears that including day one greatly reduces the correlation and this is most likely due to the lower sensitivity of the spectrophotometer at this density and this is reflected in the higher variance for this reading. Measuring r between days two and four gives the highest, and only significant, correlation. The particular days that give the best results will depend on the starting density.

Now that we have determined the best period over which to calculate r, it is of interest to compare our estimate with previously published estimates (Sager and Granick, 1953; Sorokin and Krauss, 1958). The most useful comparison to make is between the doubling times in hours. The doubling time, G, is the mean generation time for a cell to divide into two, and can be calculated from the growth rate per hour as follows

$$G = r^{-1} \ln 2. \tag{4}$$

The average doubling time for the twelve clones is 15.8 hours. Sager and Granick (1953) calculated a doubling time of 7.2 hours and Sorokin and Krauss (1958) found G to be just 6.3 hours.

So what is causing this apparent discrepancy? It is possible that the spectrophotometer is not sensititve enough to detect very low cell densities in glass tubes. To check this we can re-calculate r using the starting density of 5 x 10^4 and extend the time from day zero to day four. This gives a doubling time of 12.7 hours. Using day three this can be cut down to 11.5 hours and to 10.5 hours on day two, indicating that the spectrophotometer is missing some early portion of the exponential growth. Another contributing factor is that Sager and Granick (1953) and Sorokin and Krauss (1958) bubbled a mixture of dry air and 5% CO_2 through their growing cultures, which is necessary for optimal growth (Harris, 1989). Our cultures relied upon physical agitation to encourage atmospheric CO_2 to dissolve, a method which may be less efficient in glass tubes than directly bubbling the gas through, and may result in CO_2-limited growth.

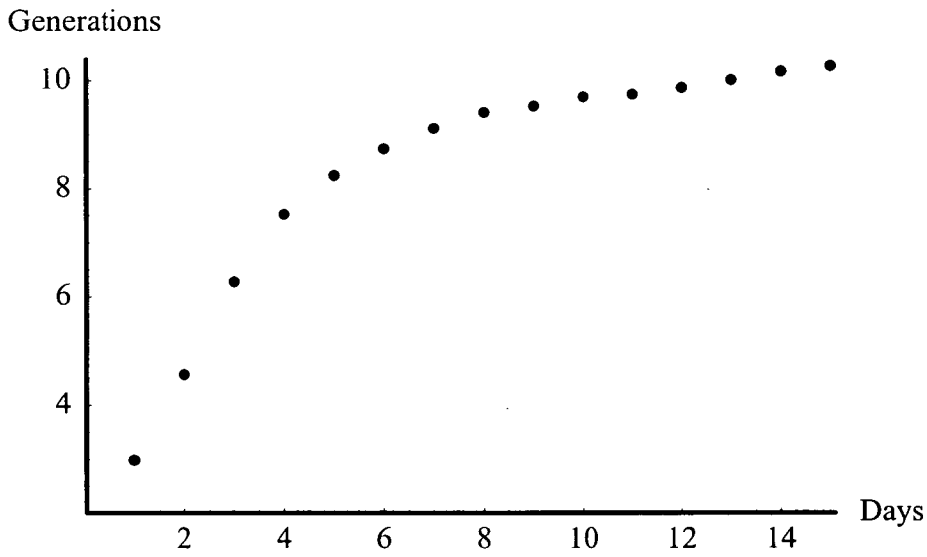


Figure 2.7. The number of generations that twelve clones of *C. reinhardtii* have passed through during a fifteen day period of growth.

However, for large numbers of tubes in experiments and fitness assays, bubbling CO_2 through the cultures would be impractical and expensive.

During a selection experiment it is useful to know how many generations the selected lines have passed through to gauge when an experiment can be stopped. It is also useful to know when cultures should be passaged into fresh media, so as to maintain the culture in exponential growth (which may the object of selection) and maximise the number of generations of selection. Figure 2.7 shows the number of generations the twelve clones have passed through over the fifteen days of growth. After three days the cultures have gone through more than six generations, showing that in ten weeks more than 120 generations can be achieved using 10 ml cultures. Larger culture sizes may reduce the necessary frequency of transfers without reducing the number of generations passed through.

There is an alternative growth vessel and growth measurement method available. *C. reinhardtii* cells can be grown in 96-well plates and their cell densities monitored via optical absorbance in an Enzyme-Linked ImmunoSorbent Assay (ELISA) plate reader. ELISA is a method used by immunologists to detect the presence of antigens and antibodies in sera and the plate readers they use are ideally suited for monitoring the cell density in a growing population of algae. Multiple replicates can be assayed on single plates, saving space and increasing efficiency and statistical power, and owing to the small volumes being used (0.25 ml) the algae grow in a shorter space of time, reducing the assay length. Additionally, the ELISA plate readers are high-precision machines that can detect much lower cell densities than was previously possible. One drawback is that the cells cannot be agitated during growth or assay, making it necessary to mix the wells prior to taking measurements.

A necessary first step towards using ELISA plates is to determine to what extent growth rates measured on the 96-well plates are correlated with those measured in glass tubes. With this intention in mind, the growth rates of ten different clones (five plus and five minus), grown separately in ELISA plates and glass tubes, were compared. For the glass tubes there were ten replicates for each clone, totaling 100 tubes, which were simultaneously grown in a single incubator with their positions in the incubator randomised daily. The inoculum density was 5×10^4. On the 96-well plate there were six replicates for each clone (the outside wells were not used) and the inoculum density was 2×10^3.

The positions of the clones on the plate was a constant, but the plate was rotated in the incubator every six hours and was 'blocked' in with empty plates to reduce light and evaporation variation across the plate. A blanking plate of sterile Bold's was also required so that the optical absorbance of the media could be taken into account. Growth

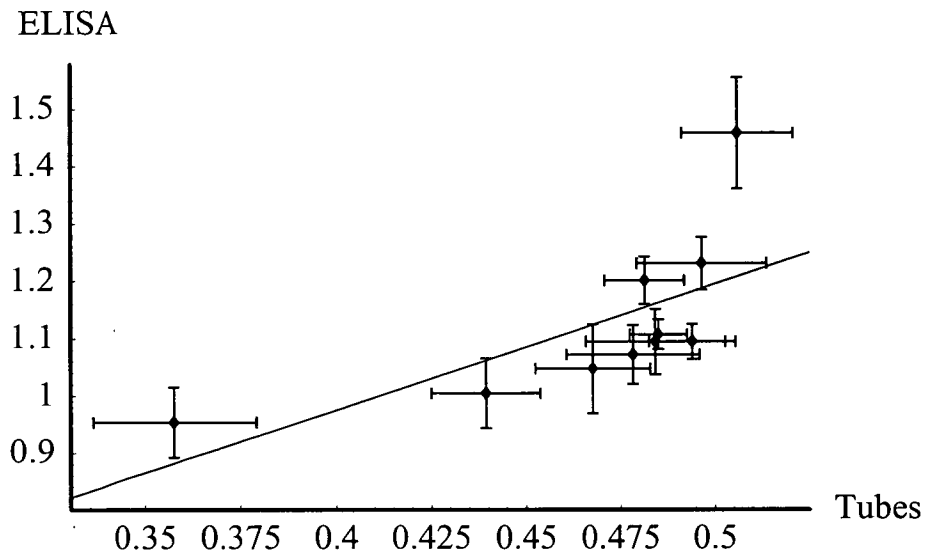


Figure 2.8. The correlation of the averaged growth rates ($\log_{10}$ d^{-1}) of ten different *C. reinhardtii* clones measured in glass tubes and on an ELISA plate showing the least-squares line of best fit. Error bars ±se.

in the glass tubes was measured over a four day period and the ELISA plate was monitored every six hours over a 48 hour period. The growth rate on the plate was calculated using the first reading after six hours and the final reading at the end of the 48 hours.

Figure 2.8 shows the average per day growth rates for each clone correlated between the glass tubes and the ELISA plate. The correlation coefficient is 0.67 ($t = 2.5$, $p < 0.02$, 8 df). Further to this, it is of interest to know how repeatable our measures of growth rates are in glass tubes compared to ELISA plates. To do this we define a measure of repeatability as follows

$$R = 1 - \frac{\bar{V}_w}{V_b} \tag{5}$$

where $\bar{V}_w$ is the average variance in growth rates for replicates within clones and V_b is the variance in growth rates between different clones. The greater this quantity is, the

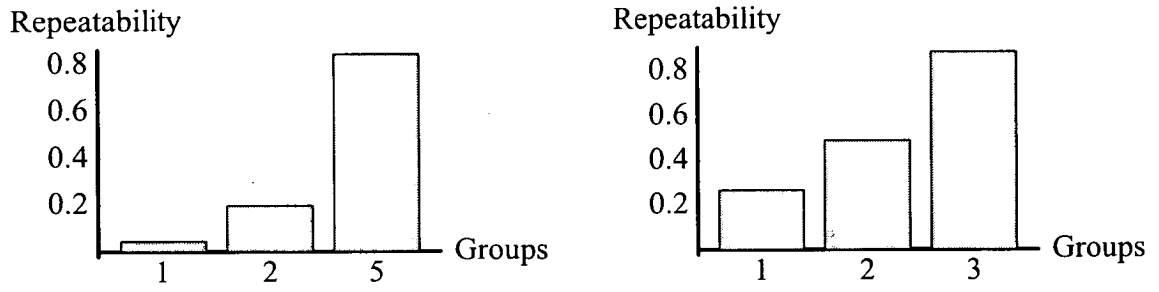


Figure 2.9. The repeatability of different sized groups of randomly chosen replicates within clonal measures from glass tubes (left) and from ELISA plates (right).

more repeatable the measures are for single clones . For the glass tubes we find an R value of just 0.045 and for the ELISA plate the R value is 0.27.

We can extend this method to ask how the repeatability increases as we take the averages of groups of randomly selected replicates within each clonal measure and use these to calculate $\bar{V}_w$. For example, we can take the averages of groups of two randomly selected replicates and then do the same for groups of three. These measures will indicate how many replicates are required to deliver repeatable results. Figure 2.9 shows the repeatability of different sized groups of replicates for glass tubes and for the ELISA plate. It can be seen that a repeatability of over 0.8 is achievable with just three replicates for the ELISA plate but requires five replicates for the glass tubes.

It is worth noting that using the ELISA plate data we can calculate a doubling time of 6.4 hours, which is in close agreement with the doubling times determined by Sager and Granick (1953) and Sorokin and Krauss (1958). This result suggests that the ELISA plate reader may be sensitive enough to monitor the early stages of exponential growth. It also appears that growth in 96-well plates is not limited by CO_2. This is due to an increased surface area to volume ratio. For glass tubes the surface area exposed to the air

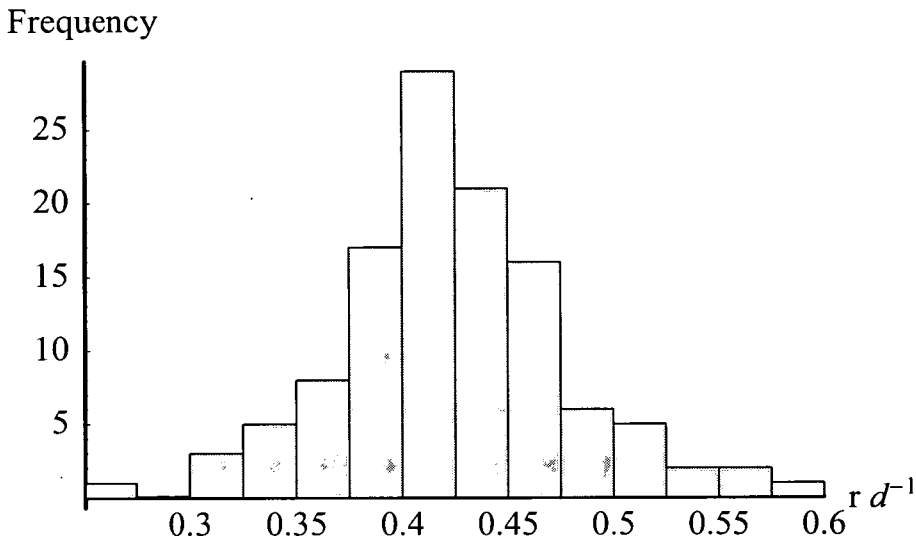


Figure 2.10. A histogram showing the numbers of replicates of a single clone that fall into different categories of growth rates ($\log_{10}$ d^{-1}). There are 116 measurements of groups of roughly 20 replicates each, measured in six independent trials over two different incubators (measured in glass tubes). The average growth rate is 0.42 and the variance is 0.003.

(πr^2) for a tube with a diameter of 25 mm is 4.91 cm^2 and for a 10 cm^3 culture volume this gives a surface area to volume ratio of 0.49. For a well in a Corning 96-well plate with a diameter of 6.86 mm the surface area exposed to the air is 0.37 cm^2 and for a 0.25 cm^3 culture volume this gives a surface area to volume ratio of 1.48, approximately three times greater than the glass tube.

During fitness assays it may not always be possible to simultaneously assay growth rates in a single incubator, especially if there are large numbers of lines or if there is a staggered design to the experiment. Therefore, it is of interest to know how comparable (and repeatable) growth rate measurements are between independently estimated measurements. A survey of the growth rates of single clones and groups of different clones

was conducted on separate occasions and across two separate incubators to get a feel for what the important sources of variation may be. All of the masurements were made in glass tubes.

Figure 2.10 shows the distribution of growth rates for a single clone based on 116 measurements of six groups of roughly 20 replicates each, measured on separate occasions in one of two incubators. The distribution appears to be leptokurtic. It is possible to analyse the sources of this variation using Fisher's intraclass correlation coefficient (Fisher, 1918, 1925), which in this instance will provide an estimate of the fraction of the total variance attributable to variation arising from separate incubator measurements. This allows us to weigh up the variance arising from within-incubator trials from position effects caused by slight temperature or light gradients across the incubator against variation arising from different ambient conditions, differences in the seeding cultures and potential temporal differences in the incubators. The intraclass correlation coefficient can be expressed as

$$\rho = \frac{MS_b - MS_w}{MS_b + (n' - 1)MS_w} = \frac{V_b}{V_t} \tag{6}$$

where MS_b is the between mean squares, MS_w is the within mean squares, V_b is the variance between trials, V_t is the total variance and n' is the within trial size for an unbalanced ANOVA,

$$n' = \frac{1}{N-1}\left(\sum_{i=1}^{N} n_i - \frac{\sum_{i=1}^{N} n_i^2}{\sum_{i=1}^{N} n_i}\right). \tag{7}$$

N is the number of separate groups and n_i is the sample size in the i'th group. For the single clone in Figure 2.10 we find a ρ value of 0.53 indicating that the variance between trials is marginally larger than the variance within incubator trials. It may seem somewhat surprising, but this result suggests that the variance from position effects within an incubator is almost equal to the variance from independently measured trials.

This highlights the importance of randomising the position of lines within an incubator during selection experiments and fitness assays.

The value for ρ is expected to increase as the genetic relation between the lines in separate trials decreases. To check this, eight clonal offspring from a single mating were assayed in separate trials. The resulting value of ρ increased modestly to 0.57. The same process was carried out with four offspring from four different matings. This increased ρ to 0.68.

It is possible to put these results together and partition the effects of the different sources of variation on the growth rates of the different clones. If we nest the effect of the incubator within the clone effect and treat it as a random factor we can carry out a mixed effects model to test whether any effect of the clone can be detected above the noise introduced by the incubator. The effect of the incubator is then tested against the position effect within each incubator which is the measurement error (residual) in this model. Table 2.1 shows the ANOVA results. The results show that when comparing clones in different incubator runs it is not possible to disentangle their effects from the effects of the incubator ($F_{12,11} = 1.70$, $p = 0.19$). This highlights that the most appropriate measures of comparison between clones should be carried out within the same incubator run. The effect of the incubator is, however, clear above the position effects from within the incubator ($F_{11,479} = 21.78$, $p = 6.3 \times 10^{-6}$). This suggests that comparisons made within incubator runs will not be obscured by measurement error.

The composition of the growth medium represents an important component of an algal lineage's selective environment and as such it may be worthwhile to identify some general trends in growth that different medium types elicit. Two clones of opposite mating type were grown in four different medium types: Bold's, Sueoka (1960), a so-called high-salt medium, and a modified Eversole (1956) medium using the Sager and

Table 2.1. The ANOVA results for the influence of different sources of variation on the growth rates of different clones in different incubator runs. Inc - different incubator runs. The ANOVA model is a mixed effects model with the effect of the incubator treated as a random factor and used to test the effect of the clones.

Between and within incubator effects on clonal growth rates				
Source of variation	df	Mean Sq	F-ratio	P-value
Clone	12	0.146	1.70	0.193
Inc (Clone)	11	0.086	21.78	$6.3 \times 10^{-6***}$
Residuals	479	0.004	-	-

Granick (1953) trace elements (to provide a source of Fe^{3+} ions), and utilising the Tris-HCl buffer to develop pH 6 and pH 8 varieties. Sueoka is high in phosphates relative to the other medium types and Bold's is the only medium with a non-reduced nitrogen source. Ten replicates of each mating type were carried out.

Figure 2.11 shows the average growth rate of the two clones in each medium type. This is by no means an exhaustive survey of the genetic variation that exists for growth in different medium types, but is intended to illustrate some possible coarse-grained trends. The clones grow best when supplied with a high phosphate medium with a reduced-nitrogen source (sueoka). It has been noticed that algae will preferentially use NH_4^+ to NO_3^- if they are both present (Syrett, 1962; Thacker and Syrett, 1972), possibly to save energy through switching off nitrate reductase expression. However, there is a large drop in growth rates when the pH is raised or lowered even in the presence of a reduced nitrogen source. Additionally, varying the pH has significantly affected the differences between the clones ($F = 49.75$, $p = 7.7 \times 10^{-7}$). In the following two sections, the results of two pilot experiments are discussed.

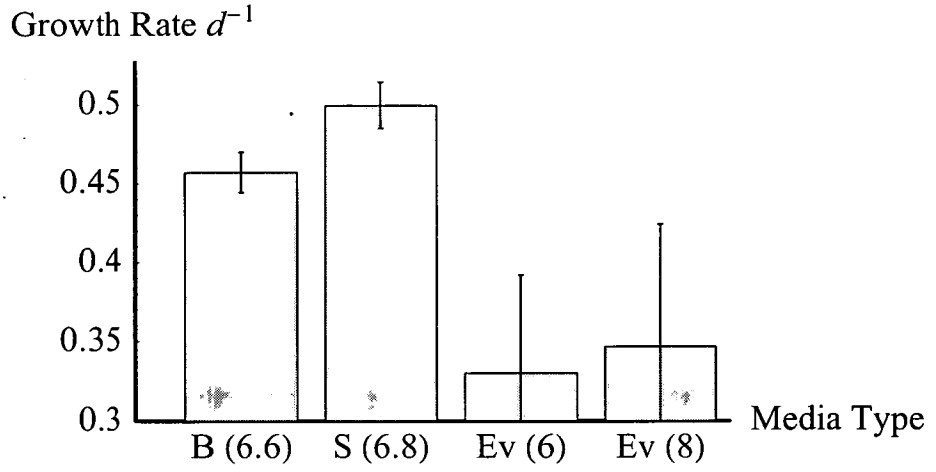


Figure 2.11. A comparison of growth rates ($\log_{10}$ d^{-1}) of two *C. reinhardtii* clones in four different medium types. B (6.6) - Bold's pH 6.6, S (6.8) - Sueoka pH 6.8, Ev (6), Ev (8) - Eversole pH 6 and pH 8. Error bars ± se.

2.5 Rates of Adaptation: New mutation versus standing variation

2.5.1 Introduction

When designing a selection experiment an important element of the design is the choice of base population. In short to medium term selection lengths most of the response to selection may be supplied by standing variation already present in the base population (e.g. Bell, 1997; Colegrave et al., 2002). In longer term selection experiments, however, the response to selection will eventually rely upon the input of new genetic variation from mutation (Collins and Bell, 2004; Bell, 2005). Therefore, the rate at which fitness-affecting mutations occur in a population acts to constrain the response to selection.

Goho and Bell (2000) measured the mutational heritability, defined as the ratio of mutational variance to environmental variance, in *C. reinhardtii* clones growing in Bold's medium in chemostats and found a value of 4 - 5 x 10^{-3} per generation based on the mean and variance of division rates. This figure includes both deleterious and beneficial mutations and accounts for environment-dependent mutations, which were neutral under chemostat conditions (liquid medium), but non-neutral on agar plates (solid medium). The chemostat cultures did not increase in mean fitness during the experiment and so from these results we cannot deduce how important new mutations are for the adaptation of a lineage to a novel environment.

The aim of this experiment is to compare the rates of adaptation to a novel environment of heterogeneous and clonal populations of *C. reinhardtii* over two different effective population sizes (bottleneck sizes). The results are intended to shed some light on how much heritable variation in fitness can be generated by sexually crossing randomly chosen clones and to what extent new mutations contribute to the response to selection over a short to medium period of selection. The *a priori* expectation is that heterogeneous populations will adapt more readily than clonal populations. It is also expected that larger effective population sizes will have a larger relative effect in clonal versus heterogeneous populations.

2.5.2 Methods

Heterogeneous base populations were formed by crossing four separately isolated clones (two plus, two minus). The population was then formed from the parents and F1 progeny from this cross. To induce gametogenesis and mating, each clone was grown up in full Bold's media to a density of 1 - 2 x 10^6 cells/ml, spun down at 2000 rpm and resuspended in autoclaved distilled water with equal proportions of each clone made up to

10 ml. This suspension was left in bright light overnight and the resulting zygotes were plated onto Bold's agar and left in total darkness for five days to allow zygote maturation, after which they were placed in bright light until the spores ruptured and started to grow ($\sim$ 3 - 4 days). The growing cells were then resuspended in Bold's liquid media prior to transfer to selective media for the first round of selection. Two clonal lines were used, one plus and one minus mating type.

The selective media was based on a modified Bold's medium, consisting of 20% of the standard concentration of Phosphate, 120% of Nitrate plus the addition of 500 mg/ml of sodium bicarbonate ($NaHCO_3$). Bicarbonate ions compete with CO_2 for uptake into the cell, thereby decreasing the growth rate (Colegrave, 2002). Populations were maintained in glass tubes in the exponential phase of growth by bottlenecking a fixed number of cells into fresh media at the end of seven days of growth. Two bottleneck sizes were used, 10^3 and 10^6, determined by optical absorbance in a spectrophotometer at 665 nm. Three replicates of each population treatment were maintained, totaling 18 lines. All of the lines were grown in a Sanyo Orbital floor incubator with continuous illumination and shaking at a temperature of 27°C, and their positions were randomised in the incubator daily.

The selection period lasted for eleven weeks after which the evolved lines were assayed together with their respective ancestors in a paired tube design, so that after tube randomisation, paired tubes were always next to one another (36 assay lines). All lines were inoculated at a density of 5 x 10^4 for the assay and their cell densities were monitored daily for a period of eight days. Growth rates and final cell densities were taken as separate measures of fitness and adaptation was computed as

$$A = \frac{f_E - f_A}{f_A} \tag{8}$$

where f_E is the evolved fitness and f_A is the ancestor fitness. Adaptation was then

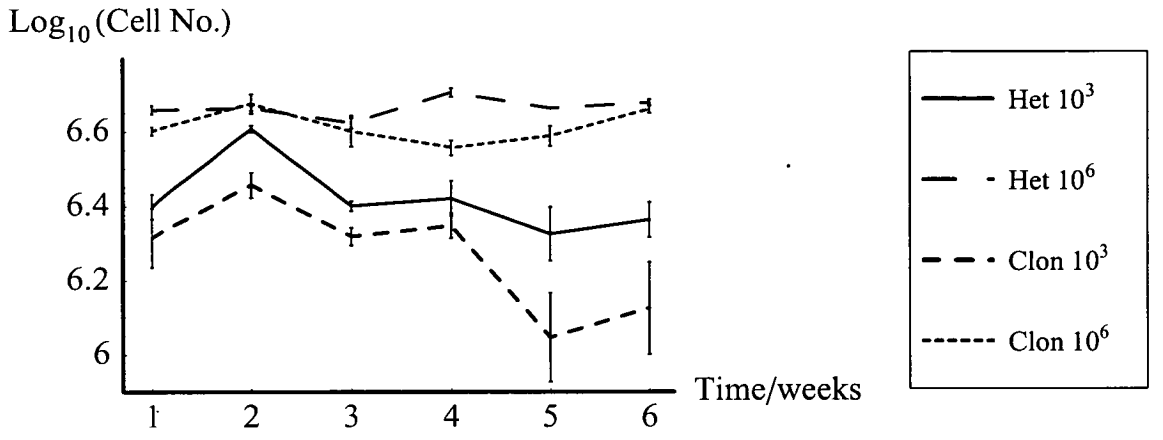


Figure 2.12. The average cell density of each population treatment, taken prior to bottlenecking, over a period of six weeks during a selection experiment. Error bars ± se.

converted into the rate of adaptation by dividing by the average number of generations that each bottleneck treatment had gone through.

2.5.3 Results and discussion

Figure 2.12 shows the average number of cells in each population treatment prior to bottlenecking into fresh media, during a six week selection period. The heterogeneous lines appear to be denser for both bottleneck sizes. A surprising, and unexpected, occurrence is that the density of the 10^3 treatments appear to be correlated between independently growing lines. The correlation coefficient for the 10^3 lines is 0.83 ($t = 3.01$, $p < 0.05$) and for the 10^6 lines is -0.19 ($t = 0.399$, $p > 0.35$).

At the lower bottleneck density, lines must be more sensitive to local variations in growth conditions, such as ambient temperatures and humidity, which may affect the

Table 2.2. Fitness estimates of the evolved and ancestor lines and adaptation of the evolved lines based upon the $\log_{10}$ cell density after 7 days of growth. A_u - adaptation prior to correction for the number of generations, A_c - adaptation after correction for the number of generations, Het - heterogeneous, Clon - clonal.

Final cell density ($\log_{10}$)						
Pop Treatment	Rep	f_A	f_E	A_u	A_c	$\bar{A}_c$ ($\pm se$)
Het 10^3	1	6.34	6.43	0.014	0.00011	0.00015 (1.7 x10^{-5})
	2	6.38	6.51	0.021	0.00017	
	3	6.34	6.47	0.020	0.00016	
Het 10^6	1	6.32	6.46	0.023	0.00093	0.00070 (1.5 x 10^{-4})
	2	6.36	6.43	0.010	0.00041	
	3	6.38	6.50	0.019	0.00077	
Clon (−ve) 10^3	1	6.42	6.50	0.0067	5.3 x 10^{-5}	3.9 x 10^{-5} (1.8 x 10^{-5})
	2	6.50	6.51	0.00025	2.1 x 10^{-6}	
	3	6.44	6.49	0.0078	6.3 x 10^{-5}	
Clon (−ve) 10^6	1	6.45	6.47	0.0035	0.00014	0.00012 (1.9 x 10^{-5})
	2	6.53	6.54	0.0013	5.2 x 10^{-5}	
	3	6.51	6.54	0.00072	0.00015	
Clon (+ve) 10^3	1	6.42	6.46	0.0069	5.5 x 10^{-5}	2.3 x 10^{-5} (1.8 x 10^{-5})
	2	6.46	6.47	0.0024	2.0 x 10^{-5}	
	3	6.46	6.45	−0.00073	−5.9 x 10^{-6}	
Clon (+ve) 10^6	1	6.47	6.42	−0.0067	−0.00027	0.00016 (2.3 x 10^{-4})
	2	6.46	6.54	0.012	0.00049	
	3	6.54	6.58	0.0067	0.00027	

Table 2.3. Fitness estimates of the evolved and ancestor lines and adaptation of the evolved lines based upon the $\log_{10}$ growth rates d^{-1}. A_u - adaptation prior to correction for the number of generations, A_c - adaptation after correction for the number of generations, Het - heterogeneous, Clon - clonal.

Growth rates ($\log_{10}$ d^{-1})						
Pop Treatment	Rep	f_A	f_E	A_u	A_c	$\bar{A}_c$ ($\pm se$)
Het 10^3	1	0.47	0.59	0.27	0.0022	0.00095 (6.2 x10^{-4})
	2	0.38	0.40	0.061	0.00049	
	3	0.45	0.46	0.024	0.00019	
Het 10^6	1	0.44	0.48	0.16	0.0064	0.0073 (0.0017)
	2	0.45	0.48	0.27	0.011	
	3	0.45	0.50	0.12	0.0049	
Clon (−ve) 10^3	1	0.44	0.48	0.083	0.00067	0.0014 (5.9 x 10^{-4})
	2	0.48	0.54	0.12	0.00094	
	3	0.39	0.51	0.32	0.0026	
Clon (−ve) 10^6	1	0.53	0.66	0.25	0.010	0.0077 (0.0018)
	2	0.47	0.52	0.10	0.0041	
	3	0.48	0.59	0.22	0.0089	
Clon (+ve) 10^3	1	0.411	0.47	0.15	0.0012	0.0011 (4.3 x 10^{-4})
	2	0.43	0.45	0.030	0.00024	
	3	0.43	0.53	0.21	0.0017	
Clon (+ve) 10^6	1	0.50	0.51	0.00025	9.8 x 10^{-6}	0.0050 (0.0049)
	2	0.50	0.69	0.37	0.015	
	3	0.45	0.46	0.0048	0.00019	

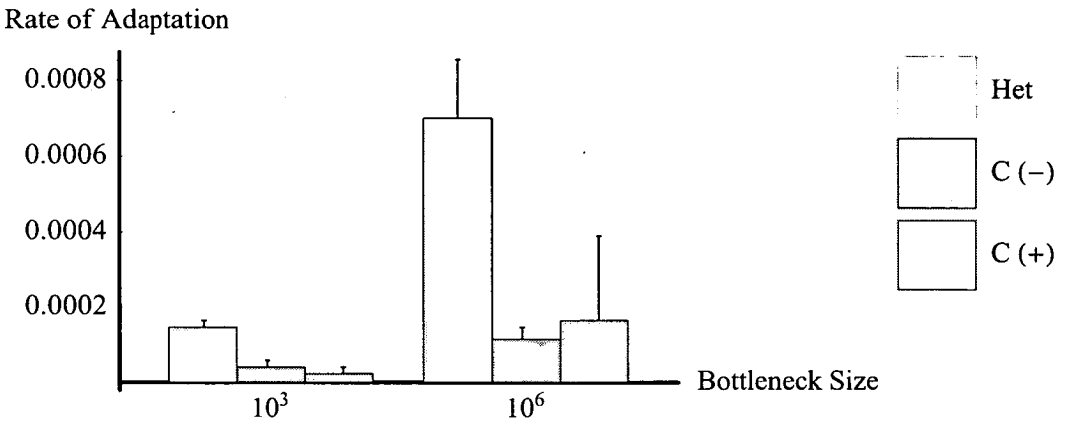


Figure 2.13. The rates of adaptation for different population treatments based upon the final cell density. Het - heterogeneous, C (-) - Clonal minus mating type, C (+) - Clonal plus mating type. Error bars + se.

incubator, and as such the variation feeds (and amplifies) into their final densities in a correlated fashion. At the higher bottleneck density, populations are more insensitive to local variations because they are closer to mid-log phase and are probably more sensitive to slight errors in their seeding densities.

Table 2.2 shows the fitness and adaptation estimates based upon the cell density at the end of the assay growth period (seven days). Table 2.3 shows the fitness and adaptation estimates based upon the per day growth rates.

Figure 2.13 shows the rates of adaptation for each population treatment derived from the final cell density measured on day eight of the assay. There appears to be an effect of both population size and base population treatment. The relative effect of population size on clonal versus heterogeneous treatments can be assessed simply as the ratio of

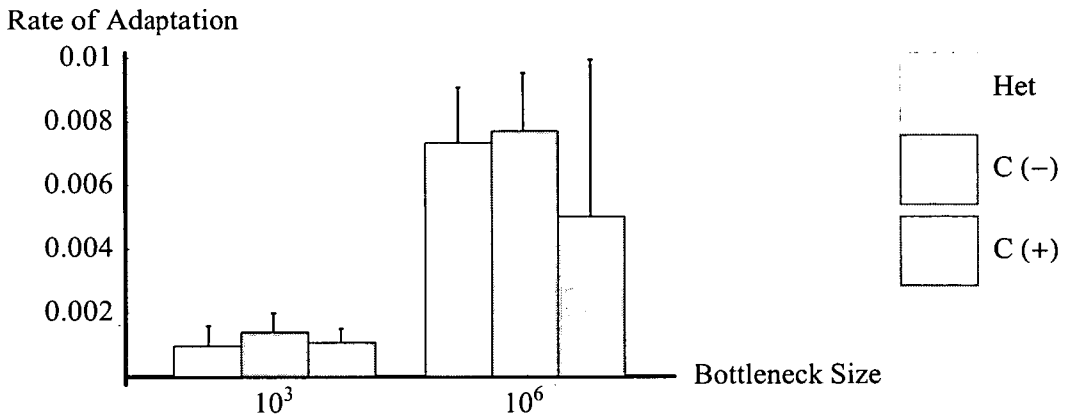


Figure 2.14. The rates of adaptation for different population treatments based upon the exponential growth rates. Het - heterogeneous, C (-) - Clonal minus mating type, C (+) - Clonal plus mating type. Error bars + se.

the rates of adaptation. The average ratio for the two clonal mating types is 5.03 and for the heterogeneous populations is 4.78, just a shade lower. There is a substantial effect of population size for the heterogeneous lines as well as for the clonal lines.

If we carry out a two-way ANOVA with bottleneck size and base population treatment as factors we find that both bottleneck size ($F = 7.72$, $p = 0.0166$) and base population treatment ($F = 5.94$, $p = 0.0160$) are significant in effect. The interaction between the two, however, is not significant ($F = 2.62$, $p = 0.113$). This result provides support for the *a priori* expectation that heterogeneity in the base population facilitates a faster response to selection and that larger bottleneck sizes will enjoy a higher rate of adaptation due to new mutation. This conclusion is not so straightforward when one examines the rates of adaptation derived from the exponential growth rates.

Figure 2.14 shows that, while the effect of bottleneck size remains important ($F = 8.83$, $p = 0.0116$), the effect of base population treatment evaporates when exponential growth rates are measured ($F = 0.23$, $p = 0.795$). So what might be causing this discrepancy between our measures of fitness?

One major difference between our base population treatments is that in the clonal populations cells compete with other clone mates, whereas in the heterogeneous cultures cells are competing with a mix of different genotypes. It is possible that unexpected competitive interactions are occurring in the heterogeneous populations that are not occurring in the clonal populations. Such interactions could be occurring directly in the fitness assay in both the ancestors and in the evolved lines, and could also have affected the selective period through frequency-dependent effects. This suggests that there may be a potential difficulty in comparing mixtures of genotypes with single clones and suggests that competitive assays may not give the same results as independent comparison assays.

Negative frequency-dependent interactions between different genotypes can result in rare genotypes being favoured. For example, predators may use visual or olfactory cues to preferentially attack more common genotypes (Allen and Clarke, 1968; Clarke, 1969; Soane and Clarke, 1973). More relevant to this study is negative frequency-dependence arising from resource competition. If different genotypes utilise different resources then rarer genotypes will be at an advantage as they will be competing with fewer individuals for their resource. Three genotypes of the esterase-6 locus in *Drosophila melanogaster* adhered to negative frequency-dependence when reared on media conditioned by different proportions of each genotype (Kojima and Yarbrough, 1967; Yarbrough and Kojima, 1967; Huang et al., 1971). A similar result was found for three genotypes of the alcohol dehydrogenase (ADH) locus also in *D. melanogaster* (Kojima and Tobari, 1969).

Synergistic interactions between different genotypes have been investigated in agriculture by comparing the yields of crop monocultures to crop mixtures. Frey and Maldonado (1967) found that specific combinations of oat-cultivar mixtures showed increased yields. However, in an independent study with more environmental variables oat-cultivar monocultures showed the highest yields (Shorter and Frey, 1979).

Bell (1991b) investigated the potential 'complementary' effects upon growth that may occur when different genotypes of *C. reinhardtii* are grown in mixtures compared to when they are grown in monocultures. The work concluded that more dominant genotypes were simply replacing the slower growing ones and that no genotype interactions of any kind were occurring. The assay only focussed on the long-term dynamics of the mixed populations (relevant for the agronomic use of crop mixtures) over a period of six weeks after an initial period of 35 days of growth. It may not be surprising that replacement had occurred by this time. The study may have neglected competitive interactions that occur much earlier in the growth cycle. An earlier study suggested that genotype interactions were important between different strains and species over the same time scale (Bell, 1990b).

There is also evidence for positive frequency-dependent interactions between genotypes where different genotypes compete with one another by producing pollutants that repress the growth of their competitors. Newman and Rovira (1975) found allelopathic interactions between different British grassland species, and also between members of the same species. All major bacterial lineages have been found to produce bacteriocins, which are antimicrobial proteins, and they are often targeted to conspecifics (Riley and Wertz, 2002; Riley et al., 2003). Buss and Grosberg (1990) demonstrated the existence of both offensive and defensive interactions between *Hydractinia* (marine hydrozoan) colonies of the same species. Thus, there seems to be a great scope for more actively

competitive interactions across many different species.

Competitive interactions may be important for *C. reinhardtii* cells as they have evolved to grow in nutrient poor conditions where an ability to acquire and assimilate limited resources more efficiently than your competitors may be selectively beneficial. More generally, the large densities that microbial populations can reach may provide a benefit to individuals investing in competitive behaviours.

2.6 Growth in used media

2.6.1 Outline

Whilst it may be difficult to directly investigate the growth dynamics of mixtures of different genotypes, it is possible to 'mimic' the conditions of growth in a mixed genotype arena by growing single clones in the used media of groups of genotypes. A group of ten clones (five plus, five minus) were grown in Bold's media at the same time as a single clone. At the end of seven days of growth, the cells were spun down at 2000 rpm, the supernatant decanted and then spun down again and carefully pipetted off before being retained. 5 ml mixtures of these used media were mixed with 5 ml of fresh media and used as the growth medium for a single clone.

Three treatments were setup: growth of the clone in full media, growth in media it had grown in previously and growth in media that ten unrelated genotypes had grown in previously. There were five replicates of each treatment. Lines were grown for a period of nine days with continual illumination and shaking, and their cell densities were monitored daily.

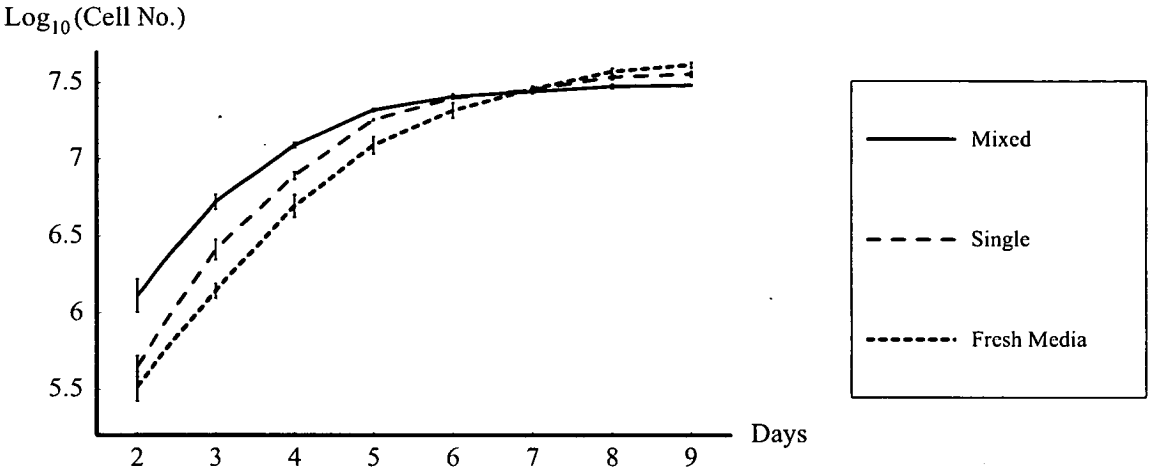


Figure 2.15. The growth of a single clone in three different media treatments: Fresh media, used media from itself (Single) and the used media from ten unrelated clones (Mixed). Error bars ±se.

2.6.2 Results and discussion

Figure 2.15 shows the growth of the single clone in each of the three media treatments. Growth in the mixed used culture was denser to begin with but eventually became the least dense. Growth in the fresh media followed the opposite pattern with the clone grown in its own media falling almost exactly half-way between. The growth rates are as follows: fresh media, 0.58 ± 0.042, single used media, 0.62 ± 0.023, mixed used media, 0.49 ± 0.046. The lower cell density at the end for the used media types may be explained by the lower nutrient composition of used media representing a lower carrying capacity. However, the higher cell density at the beginning for the mixed used media is quite unusual. It cannot be ruled out that some cells were transferred across from the older media, thereby increasing the starting cell numbers. This would still fail to account for the significant difference between the mixed used media and the fresh

media ($F = 13.37$, $p = 0.011$) and the lack of a difference between the single clone used media and the fresh media ($F = 1.18$, $p = 0.32$) since both used media treatments were prepared in the same way. If we discount this explanation then what else could result in such a pattern?

It could be that the algal clone is exiting lag phase earlier than normal as a result of a perceived high density of competitors. It is also possible, however, that the used media has abolished the need to physiologically adapt to the new environment. It could also be the case that some waste product of the algae is stimulatory for growth. This last possibility was explored by Sen and Fogg (1966) in experiments where they stimulated the early exit of lag phase of *Chlorella* by the addition of glycolate to the media. Glycolate is a waste product of *Chlorella* and many algae including *C. reinhardtii*, but it can also be utilised as a carbon source in the presence of light. Sen and Fogg (1966) championed the idea that glycolate acts as a stimulatory factor for small populations of growing algae. However, in their experiments they found that by exiting lag phase earlier the algae reduced their growth rate below what it would be without the additional glycolate. This could be indicative of a competitive behaviour, whereby the algae are willing to sacrifice a high growth rate because if they wait in lag phase for too long the nutrients may be substantially reduced by their competitors. In this scenario, glycolate acts as a signal that other algae are growing and not just as a stimulatory carbon source. Further work needs to be done to clarify the situation.

3 Divergent experimental selection and recombination load in *Chlamydomonas reinhardtii*

Abstract

The recombination load (RL) measures the deviation of the average offspring fitness from the mid-parental fitness and as such has been interpreted as implying the existence of epistatic interactions between loci. Even within this interpretation the RL could result from two different types of epistasis: antagonistic or synergistic. Studies in a number of species suggest that there is a measurable RL. Here, I describe an experiment aimed at comparing the RL in populations of *C. reinhardtii* that have been asexually selected in three different environments; two divergent, opposite extremes of pH (6 and 8) and the third an intermediate pH environment (6.6). Lines were selected in separate mating type mixtures for 140 generations. There was significant adaptation of lines for lines evolved in extreme environments but not for lines evolved in the intermediate environment. A significant increase in offspring growth rate variance was detected together with a significant decrease in their average growth rates consistent with a RL. The range measured was between 5% and 31%. The RL of offspring derived from crosses between lines selected in different environments was the same as or lower than the RL of offspring from crosses between lines selected in the same environment. It is argued that this surprising result demonstrates that the evolution of isolated populations in uniform environments is as potent a force for the build-up of between population incompatibilities as the evolution of isolated populations in divergent environments.

3.1 Introduction

Deterministic models for the evolution of recombination predict a role for epistasis through generating negative linkage disequilibria between loci via natural selection. A negative linkage disequilibrium implies that selectively beneficial alleles tend to be associated with deleterious alleles (or in the case of a quantitative trait, plus alleles are associated with minus alleles). Stabilising selection acting on a quantitative trait will generate such associations and so will synergistic (negative) epistasis between beneficial or deleterious mutations. This association acts to reduce the additive genetic variance in fitness for natural selection to act upon.

Following Weismann (1889), Mather (1943) recognised that, for quantitative traits subjected to partially stabilising and directional selection, the reduction of genetic variance will be both beneficial (under stabilising selection) and deleterious (under directional selection), thereby creating a tension between the two opposing selective forces. When a trait is under stabilising selection with a moving optimum this reduction of genetic variance provides a benefit for modifiers that increase the rate of recombination as they can hitch-hike on the beneficial variation they generate (Maynard Smith, 1980, 1988; Charlesworth, 1993). The same conclusions have been drawn from multilocus models with directional selection acting on beneficial mutations (selective sweeps) with weak synergistic epistasis (Barton, 1995). Such models are also relevant for mutation-selection balance with synergistic epistasis (Kimura and Maruyama, 1966; Kondrashov, 1984; Charlesworth, 1990).

Within this scheme it has been shown theoretically that the progeny from a single generation of mating are expected to show a reduction in fitness as a result of recombination creating unfit, extreme genotypes, known as a recombination load (RL) (Charlesworth and Barton, 1996). This argument allows for the maintenance of a non-

zero recombination rate. An alternative interpretation of RLs is that they represent the breaking up of co-adapted combinations of alleles built up by selection, where antagonistic (positive) epistasis predominates and selection acts to maintain polymorphisms at the loci involved, and in this scenario selection would be expected to favour a reduction in recombination (Fisher, 1930; Kimura, 1956; Turner, 1967; Dobzhansky, 1970; Maynard Smith, 1978; Feldman et al., 1980; Zhivotovsky et al., 1994; Barton, 1995).

These alternative causes of RLs are relevant to models of evolution in sexual versus asexual populations (Fisher, 1930; Muller, 1932; Crow and Kimura, 1965). Crow and Kimura (1965) argued that asexual populations are more likely to build up co-adapted gene complexes since mutations that are maximally beneficial on the genetic background that they arise on will be favoured by selection. Further to this, they suggested that in the extreme of freely recombining populations, where beneficial mutations at many loci are co-segregating, the alleles that invade are those that have a large additive component allowing them to have a higher fitness across all genetic backgrounds. Mayr (1963) has referred to these alternative allelic strategies as 'narrow specialists' and 'good mixers' respectively, and framed their evolution in terms of migration between locally-adapted sub-populations. These hypotheses predict that the RL will increase as the rates of recombination decrease, predicting, for example, higher RLs for facultatively sexual species compared to obligately sexual species. Therefore, the nature of the RL has implications for the evolution and maintenance of recombination in sexual populations and for the dynamics of adaptation and the genetic architecture of asexual and sexual populations.

Empirically, there has been some progress made in the measurement of RLs in different species and for different traits under selection. A series of results for chromosomal homozygotes of three different *Drosophila* species measured a mean RL of 14%

for egg-to-adult viability (Spassky et al., 1958; Spiess, 1958a,b; Dobzhansky et al., 1959). This is a surprisingly high RL for three obligately sexual species and is hard to reconcile with the high and heritable recombination rates maintained in higher organisms (Brooks, 1988). However, the results may not be so clear, especially when extrapolated to chromosomal heterozygotes (Charlesworth and Barton, 1996). Mukai and Yamaguchi (1974) measured a RL of 3% for chromosomal homozygotes but failed to detect a RL in heterozygotes. Charlesworth and Charlesworth (1975) measured a RL of 0.7% for chromosomal heterozygotes of *D. melanogaster*. Thus, there are a range of RLs presently measured in *Drosophila* species, possibly partially reflecting different measurement and chromosomal extraction techniques. However, all of the studies to date have found a measurable load.

In crosses between wild isolates of the filamentous fungus, *Aspergillus nidulans*, Jinks et al. (1966) and Butcher (1969) found significant RLs in the offspring. The average RL was 2.8% with a broad range from −8.6% to 18.9%. The wild isolates were taken from a number of different locations with a mixture of so-called 'heterokaryon incompatibilities'. Butcher (1969) showed that the incompatibility generally increased as the genetic dissimilarity between the parents increased. In an experiment designed to detect epistasis between deleterious mutations in a cross between two strains of *Chlamydomonas moewusii*, de Visser et al. (1996) found no RL in the original strains, but a large RL in UV-irradiated samples from the original strains. The average RL for r, the intrinsic rate of growth, was 22.4% and for K, the carrying capacity, was 10.9%.

While these are the only studies to explicitly measure RLs there are other studies that can shed light on the interaction between selection, epistasis and recombination rates. Malmberg (1977) selected populations of T4 bacteriophage for resistance to the drug proflavine, and during the selective process the level of recombination between bacte-

riophage was controlled by manipulating the 'multiplicity of infection' (MOI), which is the ratio of infecting bacteriophage to bacteria. At the end of the selective period, when lines were judged to have reached a fitness plateau, levels of epistasis were measured using a marker-based approach for evaluating interactions between different segments of the circular bacteriophage genome (akin to QTL mapping). This was accomplished by recombining the selected lines with a multiply marked unselected line carrying eight temperature-sensitive mutations. The T4 genome was broken into eight segments and the fitness of each segment was measured individually, summed together and compared to the fitness of the whole. Thus, any deviation from the whole is an indication of epistasis. Comparison of high and low recombination lines revealed significantly higher epistasis values for the low recombination lines relative to the high recombination lines consistent with the predictions of Crow and Kimura (1965).

A somewhat similar study was conducted by Crow (1956) investigating the genetics of DDT resistance in adult *Drosophila*. Resistant lines were selected by growing large, mixed laboratory populations in cages with DDT painted on the inside surfaces of the cage. Resistant lines were crossed to sensitive strains containing chromosomal markers and backcrosses were conducted with heterozygous males (to prevent recombination). By comparing different combinations of resistant and sensitive chromosomes for chromosomes 2, 3 and the X-chromosome the study revealed that DDT resistance was a polygenic trait with resistance genes spanning multiple chromosomes. Additionally, resistance showed an almost complete absence of epistasis with 95% of resistance attributable to additive effects of chromosomes (also F2 lines did not deviate significantly below the mean of the mid-parental and F1 lines).

Cavalli and Maccacaro (1952) investigated the polygenic nature of chloramphenicol-resistance in *Escherichia coli*. Resistant and sensitive lines were recombined and the

resulting cells were often highly sensitive and very rarely approached the resistant parent in resistance, indicating that there are epistatic interactions between the resistant alleles. All of these studies suggest an increasing level of epistasis between adaptive substitutions as levels of recombination decrease in keeping with the theory of Crow and Kimura (1965).

Here, I describe an experiment that combines experimental evolution with explicit measurements of RLs in the facultatively sexual, unicellular chlorophyte *Chlamydomonas reinhardtii*. Replicate heterogeneous base populations were selected in three different pH environments, two extreme environments (pH 6 and 8) representing divergent selection, and an intermediate environment (pH 6.6) representing the lab-adapted 'optimal' pH. The lines that made up the populations have been maintained in the intermediate environment for many generations and for this reason the strength of directional selection is expected to be weak in this environment relative to the extreme environments and enables a comparison of selection strength. Populations were separated into single mating type mixtures so that evolved populations could be exclusively crossed with each other at the end of selection so that comparisons of offspring from the same and different environments could be made.

This allows us to compare RLs between allopatric populations selected in the same and different environments. This is a key comparison motivating the study. The importance of antagonistic epistasis for the build-up of genetic incompatibilities between isolated populations has been established in theory (Dobzhansky, 1936; Muller, 1939, 1940; Templeton, 1981; Nei et al., 1983; Barton and Charlesworth, 1984; Charlesworth et al., 1987; Orr, 1995; Gavrilets, 1997). Studies such as that of Butcher (1969), which measure the RL of genetically diverged and spatially separated individuals in the wild, afford some limited insight into the importance of Dobzhansky-Muller incompatibilites

in natural populations.

The advantages of selecting populations in an experimental evolution setting are that it is possible to fully control both the selective environments that populations are diverging in, and the length of time that divergent selection lasts for. The ability to manipulate these features of the experiment allows us to make controlled comparisons between the matings of evolved lines with prior knowledge of their selective history. In this system it is possible to compare the RLs of populations selected in the same environment with the RLs of populations selected in different environments. This enables a controlled evaluation of the relative importance of differential selection versus random divergence under uniform selection for the build-up of genetic incompatibilities between populations. One might naively expect that crossing lines that have been selected in different environments will result in a larger RL than crossing lines that have been selected in the same environment. As far as we are aware, this is the first study to combine divergent experimental evolution with explicit measures of the RL.

3.2 Methods

The experiments were started with two base populations. The first (the plus population) was made up of a mixture of 14 plus mating type plus clones, drawn from a population of parental, first and second generation recombinants, whilst the second (the minus population) was made up of 14 minus mating type clones drawn from the same population. Selection lines were then started by drawing random samples from these base population, thus each line initially consisted of a random mixture of single mating type clones. Three replicates of each mating type population were selected in each environment totaling 18 lines (six lines in each environment composed of three mating type plus lines and three mating type minus lines). These lines were maintained in glass tubes in the

exponential phase of growth by serially passaging them every four days with a bottle-neck size of 5 x 10^4 cells determined from optical absorbance in a spectrophotometer at 665 nm. The selective period lasted for approximately 140 generations. The lines were shaken continuously under constant illumination at 27°C in a Sanyo Orbital floor incubator.

The pH environments were constructed as follows. The intermediate pH (6.6) was standard Bold's culture medium (Bold, 1942). The extreme pH environments (6 and 8) were made up using a modified Eversole (1956) medium with the trace elements provided by the Sager and Granick (1953) medium to provide a source of Fe^{3+} ions. The Tris-HCl buffer of the Eversole medium facilitated the construction of pH 6 and pH 8 varieties. A complete recipe for these media types can be found in the Appendix.

At the end of selection populations were kept in relative stasis under dim lights and on agar slants with a base of Bold's medium. To measure the RL the final populations were mass mated with each other and samples of 20 offspring clones were taken from the pool of recombinants to measure the RL (see Table 3.1).

Mating was achieved by growing each population in Bold's media to a density of 1 - 2 x 10^6 cells/ml, centrifuging them at 2000 rpm and resuspendeding in autoclaved distilled water with equal proportions of each population made up to 10 ml. This suspension was left in bright light overnight and the resulting zygotes were plated onto Bold's agar and left in total darkness for five days to allow zygote maturation. After this period the zygotes were left at −20°C overnight to ensure that the parents were killed and only the offspring were recovered (zygotes are resistant to freezing). It is possible that the freezing process may have selected for zygotes that are more resistant to being frozen. For this to influence the results there would have to be a significant correlated effect with the growth rate. Previous observations of frozen clones suggest that this is

not the case. After freezing zygotes were placed in bright light until the spores ruptured and started to grow ($\sim$ 3 - 4 days).

The growing cells were then resuspended in Bold's liquid media and serial dilutions were plated onto Bold's agar using a glass spreader. Clones were allowed to grow under full, continuous lighting and once visible were picked to form the offspring sample. A sample of roughly 20 offspring clones was taken for each mating.

Table 3.1 gives an overview of the assay and mating design. Approximately 300 clones were picked by the end of the experiment and roughly 1260 lines were assayed. There are more lines than clones because for the between-pH matings the same 20 clones were assayed in both environments, and in each assay 20 plus parent and 20 minus parent replicates were also assayed in a triplet-tube design with their offspring such that each offspring was paired with both parents. From this outline it can be seen that replicates were mated with their corresponding replicate numbers for simplicity. It can be seen that within this setup a fully factorial analysis would be infeasibly large.

Assays were carried out in 25 x 150 mm glass tubes with 10 ml of the assay media and the growth rates were measured over a period of four days from an inoculum size of 5 x 10^4. Tube positions were randomised daily with offspring-parent triplets kept together, but their positions relative to each other was randomised. Growth rates were measured via absorbance in a spectrophotometer at 665 nm. Raw growth rates were converted into estimates of the RL, expressed as a percentage of the maximum, using

$$\Delta = \frac{\bar{G}_p + \bar{G}_m}{2} - \bar{G}_o \tag{1}$$

and

$$RL = \frac{\Delta}{\frac{1}{2}(\bar{G}_p + \bar{G}_m)} \times 100\% \tag{2}$$

where $\bar{G}_o$ is the average offspring growth rate, $\bar{G}_p$ is the average growth rate of the plus parent and $\bar{G}_m$ is the average growth rate of the minus parent. Thus, a positive

Table 3.1. An overview of the assay and mating design. '20' indicates that 20 offspring clones were assayed in the conditions specified. Sel - selection, Reps - replicates.

				Plus parent Sel pH								
				6			8			6.6		
		Assay pH	Reps	1	2	3	1	2	3	1	2	3
Minus parent Sel pH	6	6	1	20	-	-	20	-	-	-	-	-
			2	-	20	-	-	20	-	-	-	-
			3	-	-	20	-	-	20	-	-	-
		8	1	-	-	-	20	-	-	-	-	-
			2	-	-	-	-	20	-	-	-	-
			3	-	-	-	-	-	20	-	-	-
	8	6	1	20	-	-	-	-	-	-	-	-
			2	-	20	-	-	-	-	-	-	-
			3	-	-	20	-	-	-	-	-	-
		8	1	20	-	-	20	-	-	-	-	-
			2	-	20	-	-	20	-	-	-	-
			3	-	-	20	-	-	20	-	-	-
	6.6	6.6	1	-	-	-	-	-	-	20	-	-
			2	-	-	-	-	-	-	-	20	-
			3	-	-	-	-	-	-	-	-	20
Total (approximate) No. Clones												300
Total (approximate) No. Assay Lines												1260

RL indicates that the offspring are less fit than the parents and vice versa for a negative RL. Additionally, the growth rates of evolved lines were compared to those of their ancestors in the environment they were selected in, and these growth rates were measured in ELISA plates over a period of 48 hours. A significantly positive correlation between growth rates in ELISA plates and glass tubes has been found previously. Two plates were set up with randomised positions for the ancestor and evolved lines, and there were six replicates for each replicate evolved line over the two plates, and there were eight replicates of the ancestors in each environment over both plates totaling 156 measurements. The plates were rotated in the incubator every six hours and the wells were mixed via pipetting prior to assaying. Henceforth, G refers to $\log_{10}$ growth rates d^{-1} and $\bar{G}$ to their averages.

3.3 Results

3.3.1 Adaptation of evolved populations

Tables 3.2 and 3.3 show the growth rate data for evolved and ancestor lines in their respective pH environments. Figure 3.1 shows the growth rates of the evolved lines with their respective ancestors in each pH treatment. It can be seen that the average growth rates for the ancestors is lower for both mating types at pH 6 and at pH 8, but not for the ancestors in the intermediate environment, pH 6.6.

We can test for significant differences between the evolved lines and the ancestors by carrying out pair-wise one-way ANOVAs for each ancestor and evolved replicate pair within each pH environment. For the plus lines evolved at pH 6 (top left in Figure 3.1) all of the replicate line means were significantly higher than the ancestor mean. Carrying out a Tukey-Kramer *post hoc* test for unequal sample sizes to take the multiple comparisons into account we find that the lowest mean of the evolved lines (replicate 1)

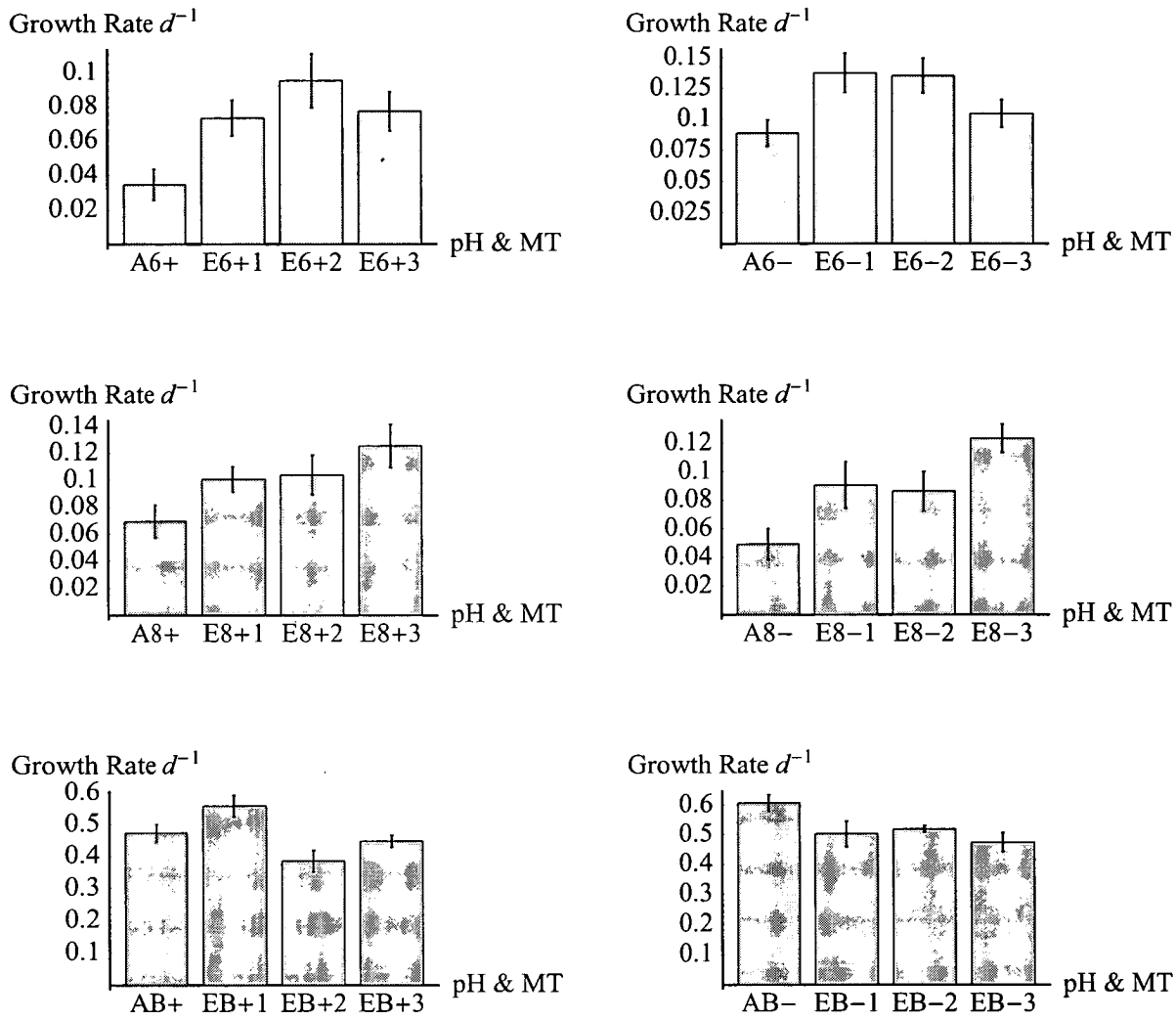


Figure 3.1. A comparison of growth rates for ancestors versus evolved lines in pH 6, pH 8 and pH 6.6 (B) environments for each mating type with the three evolved replicates. A - ancestor, E - evolved, MT - mating type. pH refers to the selective pH and the assay pH. Error bars ± se.

Table 3.2. The $\log_{10}$ growth rate d^{-1} data for the evolved lines and for the ancestor lines for the pH 6 and pH 8 selection environments. 'pH' refers to the assay pH and the selection pH. A - ancestor, Pop - population, Rep - replicate, G - $\log_{10}$ growth rate d^{-1} for each replicate, $\bar{G}_{Rep}$ - average growth rate of all of the replicates.

pH	Pop	Rep	G								$\bar{G}_{Rep}$ ($\pm se$)
			1	2	3	4	5	6	7	8	
6	6+	1	0.122	0.066	0.060	0.075	0.067	0.052	-	-	0.073 (±0.010)
		2	0.128	0.143	0.119	0.068	0.052	0.062	-	-	0.095 (±0.016)
		3	0.089	0.094	0.030	0.090	0.060	0.105	-	-	0.077 (±0.011)
	A+	-	0.058	0.023	0.030	0.006	0.080	0.030	0.008	0.045	0.035 (±0.008)
	6−	1	0.183	0.185	0.103	0.091	0.124	0.139	-	-	0.137 (±0.016)
		2	0.147	0.128	0.085	0.191	0.119	0.141	-	-	0.135 (±0.014)
		3	0.088	0.106	0.124	0.146	0.094	0.070	-	-	0.105 (±0.011)
	A−	-	0.091	0.081	0.082	0.098	0.128	0.123	0.035	0.073	0.090 (±0.010)
8	8+	1	0.073	0.150	0.090	0.097	0.102	0.091	-	-	0.100 (±0.011)
		2	0.134	0.053	0.083	0.101	0.081	0.170	-	-	0.103 (±0.017)
		3	0.117	0.086	0.143	0.136	0.072	0.199	-	-	0.125 (±0.018)
	A+	-	0.068	0.106	0.044	0.033	0.028	0.106	0.112	0.058	0.070 (±0.014)
	8−	1	0.160	0.076	0.132	0.034	0.070	0.073	-	-	0.091 (±0.019)
		2	0.151	0.042	0.101	0.081	0.091	0.051	-	-	0.086 (±0.016)
		3	0.147	0.117	0.126	0.162	0.106	0.083	-	-	0.123 (±0.012)
	A−	-	0.033	0.081	0.010	0.078	0.012	0.060	0.072	0.057	0.050 (±0.010)

Table 3.3. The $\log_{10}$ growth rate d^{-1} data for evolved lines versus ancestor lines for the pH 6.6 selection environment. 'pH' refers to the assay pH and the selection pH. A - ancestor, Pop - population, Rep - replicate.

pH	Pop	Rep	G								$\bar{G}_{Rep}$ ($\pm se$)
			1	2	3	4	5	6	7	8	
6.6	6.6+	1	0.708	0.600	0.587	0.504	0.454	0.493	-	-	0.557 ($\pm$0.038)
		2	0.494	0.432	0.444	0.334	0.300	0.302	-	-	0.383 ($\pm$0.034)
		3	0.435	0.483	0.517	0.435	0.411	0.398	-	-	0.446 ($\pm$0.018)
	A+	-	0.512	0.578	0.394	0.481	0.456	0.415	0.378	0.566	0.473 ($\pm$0.027)
	6.6−	1	0.645	0.577	0.46	0.370	0.405	0.470	-	-	0.502 ($\pm$0.043)
		2	0.563	0.508	0.535	0.512	0.486	0.506	-	-	0.518 ($\pm$0.011)
		3	0.501	0.526	0.593	0.411	0.406	0.403	-	-	0.473 ($\pm$0.032)
	A−	-	0.704	0.688	0.651	0.581	0.603	0.596	0.577	0.455	0.607 ($\pm$0.032)

remains significantly higher than that of the ancestor (q = 10.75, p < 0.01, 12 df).

For each of the remaining groups of lines selected in pH 6 and pH 8 there was one replicate line that was not significantly different from their respective ancestor in the pair-wise ANOVAs. None of the plus lines were significantly greater than the ancestor in the intermediate environment (pH 6.6). For the minus lines the ancestor was even significantly higher than two of the replicate evolved lines.

If we use the three means of the evolved lines and compare them to the mean of the eight ancestral measures within each pH we can test whether there is on average a difference between our evolved and ancestor lines across all of the lines. To do so we carry out a three-way fixed effects ANOVA with ancestor-evolved, pH and mating type

Table 3.4. ANOVA results for the data in Tables 3.2 and 3.3 (analysis details explained in the text). MatType- mating type, Anc-Evol - ancestors vs evolved lines, Reps - within-pH replicates

pH 6, pH 6.6 and pH 8				
Source of variation	df	Mean Sq	F-ratio	P-value
Anc-Evol	1	0.0003	0.093	0.76
pH	2	1.57	478	<<0.001
MatType	1	0.029	8.72	0.0046**
Anc-Evol x pH	2	0.012	3.53	0.036*
Anc-Evol x MatType	1	40.0017	0.51	0.48
MatType x pH	2	0.014	4.34	0.018*
Anc-Evol x MatType x pH	2	0.0015	0.44	0.64
Residuals	36	0.0033	-	-

as the three factors. The results are shown in Table 3.4.

The results show that there is a marginal interaction effect between the ancestors and the averages of the evolved lines and the assay pH ($F_{2,36} = 3.53$, $p = 0.036$). This suggests that taken over all the lines there is evidence of a difference in the response to selection in our different environments. Examining Figure 3.1 we can see that at pH 6 and 8 the mean of the ancestors is consistently lower than that of the evolved lines, but this is not the case at pH 6.6. There was also an interaction between mating type and pH across all environments ($F_{2,36} = 4.34$, $p = 0.018$), although it is not possible to infer anything about mating type differences in pH growth rates as the effects are confounded with the different base populations. However, it does indicate a significant, and unexpected, difference between the mating type base populations in their growth

rates in each environment for both the ancestors and the evolved lines. Such differences are rarely documented in experimental evolution studies.

3.3.2 Offspring RLs

The data for offspring and mid-parental growth rates are summarised in Tables 3.5, 3.6 and 3.7. Here, we introduce an additional statistic of comparison, the ratio of variance (RV), which is calculated as

$$RV = \frac{V_O}{\frac{1}{2}(V_P + V_M)} \tag{3}$$

where V_O is the variance of the offspring, V_P is that of the plus parent and V_M is that of the minus parent. If the RV is equal to one then the offspring and parents have equal variance expressible in the environment, and if the RV is greater than one the offspring have an increased variance relative to the average of their parents. There is a significant increase in growth rate variance in the offspring from pH 6 and 8 selected parents ($F_{1,34}$ = 12.99, p = 0.0009), but no increase in variance for offspring from pH 6.6 selected parents ($F_{1,4}$ = 1.51, p = 0.29).

Figure 3.2 shows the distribution of growth rates for one case (8+/8− replicate 3) to illustrate the increase in variance of offspring relative to their parents. The tables show positive average RLs for all of the crosses from pH 6 and 8 selected parents bar 6+/8− when assayed at pH 6, which has a negative average RL (−8.48%). The offspring from pH 6.6 selected parents also have a (small) negative average RL (−2.73%). Additionally, all of the pH 6 and 8 derived lines have larger average RVs than the lines derived from pH 6.6 parents.

The triplet design of our experiment allows us to calculate a mid-parental value for each triplet, which can be used in a comparison with the offspring. Figure 3.3 shows the distribution (leptokurtic) of growth rates for offspring and mid-parents selected in pH 6

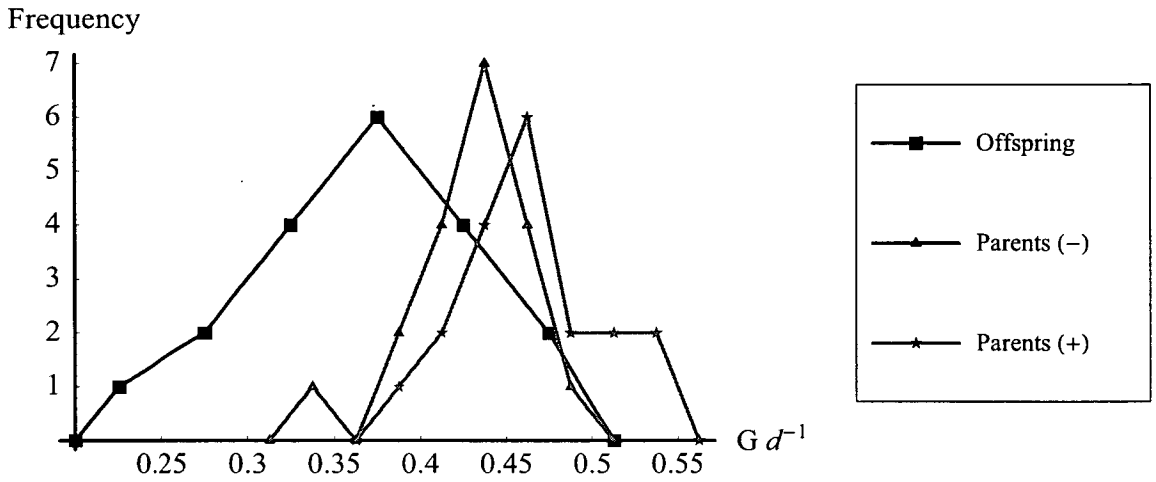


Figure 3.2. Frequency polygons showing the distribution of $\log_{10}$ growth rates d^{-1} for offspring and parents of replicate 3 of the 8+/8− cross summarised in Table 3.5.

and pH 8. Figure 3.4 shows that the distribution of ANOVA residuals for pH 6 and pH 8 offspring and mid-parents is normal. With these considerations in mind we proceed with the analysis.

First of all, we can ask whether the offspring have significantly different growth rates from their mid-parents using the raw growth rate data. To do so we separately pool together each measure for the mid-parents and then each measure for the offspring in each pH and regard this as a random factor allowing us to nest it within the treatment of midparent-offspring for a mixed effects ANOVA. Then across all of the mating combinations and for both assay pHs we find a marginal difference between the offspring and the mid-parents selected in pH 6 and pH 8 ($F_{1,34} = 4.19$, $p = 0.048$). Hence, there is a measurable decrease in the average growth rates of the offspring versus the mid-parents across all of the treatments. However, for offspring and parents in pH 6.6 there is no difference ($F_{1,4} = 2.54$, $p = 0.19$).

Table 3.5. Summary of growth rate data from pH 6 and pH 8 within environment matings and a 6−/8+ (assay pH 6) cross. Rep - replicate, *RV* - ratio of variance, O - offspring, MidP - mid-parent.

Cross (pH)	Rep	$\bar{G}$	Var(G)	*RV*	$\bar{RV}$ ($\pm se$)	*RL*	$\bar{RL}$ ($\pm se$)
6+/6− (6)	O1	0.410	0.0081	2.10	3.87 (1.45)	7.31%	14.07% (3.87)
	MidP1	0.443	0.0040				
	O2	0.378	0.010	2.76		14.17%	
	MidP2	0.439	0.0037				
	O3	0.343	0.016	6.75		20.71%	
	MidP3	0.433	0.0024				
8+/8− (8)	O1	0.443	0.014	2.40	3.21 (0.66)	10.00%	20.08% (6.20)
	MidP1	0.493	0.0072				
	O2	0.293	0.0091	4.52		31.38%	
	MidP2	0.427	0.0020				
	O3	0.363	0.0040	2.72		18.87%	
	MidP3	0.447	0.0015				
6−/8+ (6)	O1	0.456	0.0094	14.91	6.01 (4.48)	7.27%	4.88% (7.98)
	MidP1	0.491	0.00063				
	O2	0.410	0.0340	2.37		17.34%	
	MidP2	0.497	0.0143				
	O3	0.464	0.00456	0.723		−9.98%	
	MidP3	0.421	0.0063				

Table 3.6. Summary of growth rate data for the 6−/8+ cross assayed at pH 8 and for the 6+/8− cross assayed at both pH 6 and at pH 8. Rep - replicate, *RV* - ratio of variance, O - offspring, MidP - mid-parent.

Cross (pH)	Rep	$\bar{G}$	Var(G)	*RV*	$\bar{RV}$ ($\pm se$)	*RL*	$\bar{RL}$ ($\pm se$)
6−/8+ (8)	O1	0.413	0.0170	24.87	9.12 (7.88)	15.87%	10.34% (2.89)
	MidP1	0.491	0.00046				
	O2	0.486	0.0100	1.91		6.06%	
	MidP2	0.517	0.0052				
	O3	0.518	0.0061	0.589		9.10%	
	MidP3	0.570	0.010				
6+/8− (6)	O1	0.424	0.0183	3.32	2.32 (0.512)	−9.72%	−8.48% (0.621)
	MidP1	0.390	0.0055				
	O2	0.476	0.00196	2.02		−7.77%	
	MidP2	0.441	0.00097				
	O3	0.488	0.0239	1.62		−8.48	
	MidP3	0.452	0.0147				
6+/8− (8)	O1	0.362	0.010	6.75	4.27 (1.42)	12.54%	14.63% (3.23)
	MidP1	0.414	0.0015				
	O2	0.448	0.0069	1.84		10.38%	
	MidP2	0.499	0.0038				
	O3	0.382	0.0210	4.21		20.98%	
	MidP3	0.484	0.0050				

Table 3.7. Summary of growth rate data for pH 6.6. Rep - replicate, *RV* - ratio of variance, O - offspring, MidP - mid-parent.

Cross (pH)	Rep	$\bar{G}$	Var(G)	*RV*	$\bar{RV}$ ($\pm se$)	*RL*	$\bar{RL}$ ($\pm se$)
6.6+/6.6− (6.6)	O1	0.503	0.0036	1.90	1.42 (0.304)	−4.08%	−2.73% (1.23)
	MidP1	0.483	0.0019				
	O2	0.479	0.00166	0.857		−0.279%	
	MidP2	0.478	0.00193				
	O3	0.488	0.00394	1.51		−3.82%	
	MidP3	0.460	0.00261				

Table 3.8. The ANOVA results for the data in Tables 3.5, 3.6 and 3.7 and in Figure 3.5. Sel Comb - selection environment combination of the parents.

RL analysis				
Source of variation	df	Mean Sq	F-ratio	P-value
Sel Comb	4	260.05	4.40	0.016*
pH (Sel Comb)	2	429.23	7.27	0.0068**
Residuals	14	59.04	-	-

Figure 3.5 shows the average RLs for each cross in their respective assay pHs. Several things can be noticed from this graph. First of all, crosses within the extreme environments (pH 6 and pH 8) have large positive RLs compared to the cross within the intermediate environment(pH 6.6). Secondly, and more surprisingly, crosses between parents selected in different environments (pH 6 x pH 8) result in equal or smaller RLs

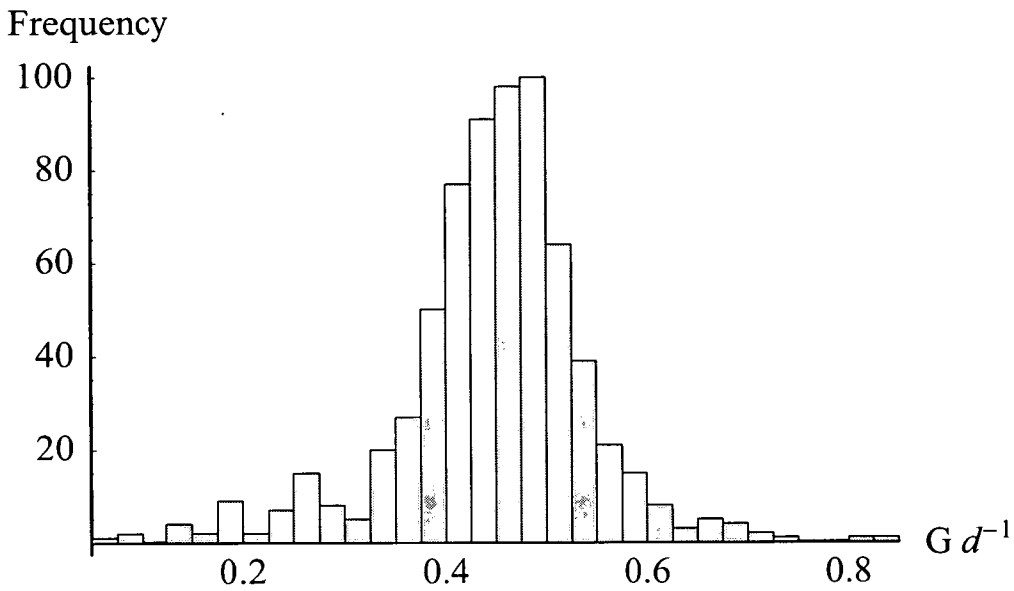


Figure 3.3. The distribution of $\log_{10}$ growth rates d^{-1} for offspring and mid-parents from pH 6 and pH 8 selected lines. The average is 0.447 and the variance is 0.009.

than crosses between parents selected in the same environments. Thirdly, there is a reversal in the sign of the RL for the 6+/8− cross depending on whether the offspring are assayed at pH 6 (−ve RL) or assayed at pH 8 (+ve RL).

To analyse this data we can use a fixed effects model with the assay pH as a nested fixed factor within the selection combination of the parents. The results are shown in Table 3.8. We find a weak, but significant effect of the selective environment combination of the parents ($F_{4,14} = 4.4$, $p = 0.016$) and a stronger significant effect of the assay pH ($F_{2,14} = 7.27$, $p = 0.0068$). The strong effect of pH arises partly from the reversal of the RL for the 6+/8− cross, and the weak effect of the selection environment combination is caused partly by the lack of a RL for the offspring from the pH 6.6 matings. Leaving this data out results in a weaker effect of the selective environment combination of the

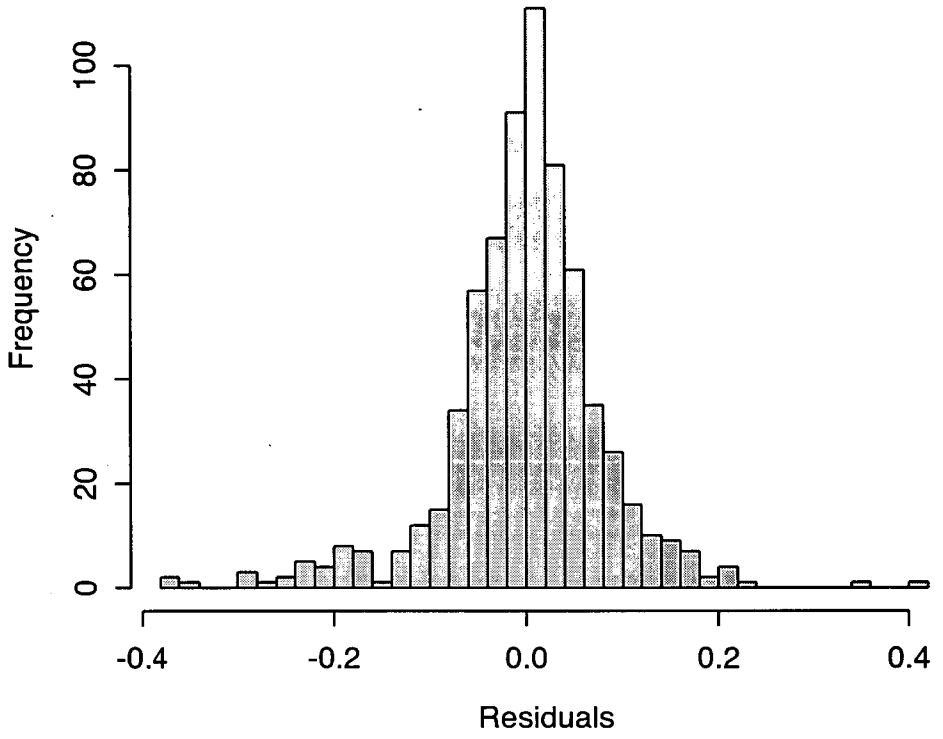


Figure 3.4. The distribution of ANOVA residuals for offspring and mid-parental growth rates from pH 6 and pH 8 selected lines.

parents ($F_{3,12} = 3.37$, $p = 0.055$).

Given that there was evidence for significant differences between the base populations in their responses to each environment we can examine the between environment crosses for pH 6 and for pH 8 to see if there was any carry-through to the offspring from these crosses. Figure 3.6 shows the correlation between offspring performance in both environments from both parental crosses. Plus parents tended to grow better at pH 8 and

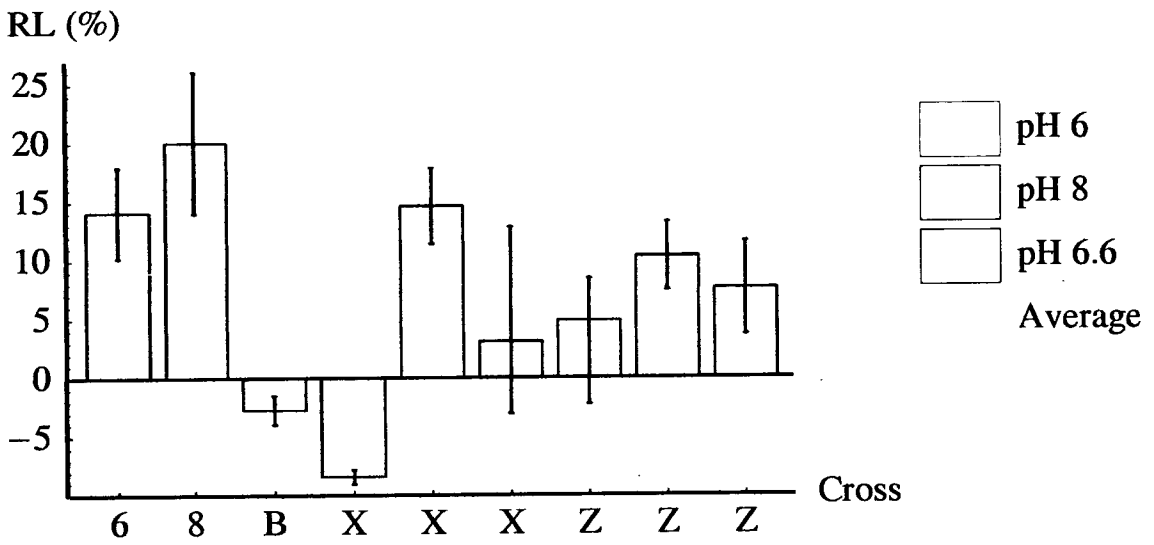


Figure 3.5. The RLs for within (6+/6−, 8+/8−, 6.6+/6.6−) and between (6+/8−, 6−/8+) matings for pH 6, pH 8 and pH 6.6 (B) selected lines. 6 - 6+/−, 8 - 8+/8−, B - 6.6+/6.6−, X - 6+/8-, Z - 6-/8+. 'pH' refers to the assay pH. Error bars ± se.

minus parents tended to grow better at pH 6. Hence, we might expect a negative correlation between the growth rates of the offspring across both environment types. The 6+/8− cross is between parental lines selected in the environments they do not grow best in and vice versa for the 6−/8+ cross. There is a significant negative correlation for the offspring from parents selected in the environment they grow best in (offspring from 6−/8+), but a weaker, non-significant negative correlation when the parents have been selected in environments they do not grow best in (offspring from 6+/8−). This may represent the signature of selection acting to reduce the inherent growth tendencies of the original base populations.

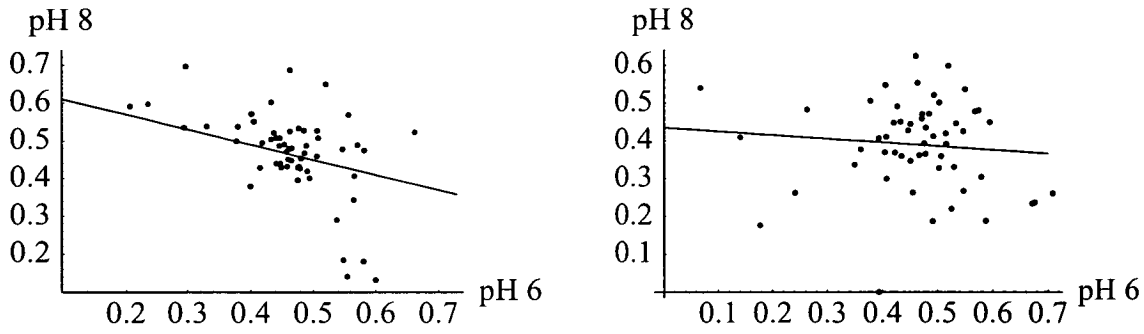


Figure 3.6. Correlations between the growth rates of 60 offspring clones in pH 6 environments with their growth in pH 8 environments. On the left are offspring from the 6−/8+ cross and on the right are offspring from the 6+/8− cross. For the 6−/8+ cross $r = -0.45$ ($t = 3.85$, $p < 0.0005$) and for the 6+/8− cross $r = -0.097$ ($t = 0.711$, $p > 0.05$).

3.4 Discussion

The experimental evolution of heterogeneous, replicate base populations of *C. reinhardtii* to three different environments and the RL of the offspring of the evolved populations has been described. Two of the environments were divergent with respect to each other (pH 6 and pH 8), and one of the environments was intermediate to the other two (pH 6.6). There are three main results in this study. The first is that lines subject to divergent selection showed evidence for adaptation relative to their ancestors when taken over all of the replicates, but that lines selected in the intermediate environment showed no evidence for adaptation. The second is that there were significant positive RLs measured in offspring from pH 6 and pH 8 selected lines, ranging from 5% to 31%, and that these were accompanied by significant increases in offspring growth rate variance relative to the parents. The third, and most interesting result, is that the RLs of offspring from parents selected in different environments were equal to or lower than

the RLs of offspring from parents selected in the same environment.

The third result shows that, in this system at least, incompatibilities between allopatric populations are just as likely to build-up when the populations are subject to uniform selection pressures as when they are subject to divergent selection pressures. That is, differential and divergent selection applied to isolated populations has not enhanced the potential genetic incompatibility between them any more than selection in the same environments. This may seem surprising, as one might naively expect one of two things. The first is that by recombining loci responsible for adaptations to opposite or different environments may be expected to be more detrimental than recombining loci that are responsible for adaptations to the same environment. Butcher (1969) found that the RL between wild isolates of *A. nidulans* showed no tendency to increase with the spatial distance between the isolates, but that the genetic dissimilarity between isolates was a good predictor of the RL. From this study we find that recombination between populations adapted to different environments can even result in fitter offspring in some cases. The second expectation is closely related to this, and is that one may expect uniform selective regimes to select for similar or identical adaptations and therefore prevent the build-up of adaptive variance between populations that could result in a RL when the populations are crossed.

Previous experiments have shown that this is unlikely to be the case. Cohan and Hoffmann (1989) found that when populations of *D. melanogaster*, *D. persimilis* and *D. pseudoobscura* were selected for 'knockdown' ethanol resistance, the variance in the adaptive characters between conspecific (and between different species) populations increased relative to unselected lines suggesting that uniform selection can act as a diversifying force (Cohan and Hoffmann, 1986; Hoffmann and Cohan, 1987; Cohan et al., 1989). Similar results have been found in long-term experimental evolution stud-

ies with *Escherichia coli* (Lenski, 1988; Lenski et al., 1991). This indicates that in our experiment, different adaptations for the same environment were built up in isolated populations and that recombining the loci responsible for these adaptations was equally or more detrimental than recombining loci responsible for adaptations to different environments.

What can this tell us about the adaptation of populations in the wild? This study suggests that the reproductive isolation of populations or species brought about by the build-up of genetic incompatibilities is influenced more by the extent of the isolation between the populations than by the environmental conditions that they are adapting to. This emphasises the difficulty for sympatric (or parapatric) populations to speciate compared to the relative ease with which allopatric populations will accumulate genetic incompatibilities even under uniform selection. As pointed out by Orr (1995) the rate at which incompatibilities build-up is at least as fast as the square of the time since the separation of the populations.

Another interesting result emerged from this study. In the 6+/8− cross the offspring exhibited a positive RL when assayed at pH 8, but a negative RL when assayed at pH 6. Such a dramatic reversal in offspring performance relative to the parents was wholly unexpected. The result is consistent with a genotype by environment interaction in the offspring. Genotype by environment interactions have been shown to play an important role in *Chlamydomonas* fitness components both within (Bell, 1991a) and between strains (Bell, 1990a). Further to this, they have been implicated as having an important role in the build-up of Dobzhansky-Muller incompatibilities (Bordenstein and Drapeau, 2001) and have been shown to play an important role in the fitness of hybrid flour beetles (Wade et al., 1999). The reversal in fitness indicates that there are unexpected positive effects of recombination in the parents. As we are dealing with

a haploid organism the positive interactions could not arise from overdominance, but would have to be the result of epistatic interactions. Epistasis has been identified as an important component of heterosis in rice hybrids from QTL studies (Yu et al., 1997; Li et al., 2001). However, without the benefit of more detailed genetic studies this must remain a speculation

The lack of adaptation of populations in the intermediate environment and the lack of a RL in the offspring of these lines suggests that the variation in the base population was neutral in the intermediate environment and that no beneficial mutations occurred during the selective period for these lines. This is surprising, but may indicate that the directional selection was very weak. Collins and Bell (2004) selected replicate lines of *C. reinhardtii* to high CO_2 levels for about 1000 generations and did not find any specific adaptations relative to their ancestors at the end of selection. The evolved lines had, however, a reduced ability to grow at ambient levels of CO_2 relative to their ancestors and it was concluded that relaxed stabilising selection on genes controlling the carbon concentration mechanism had allowed conditionally neutral mutations to accumulate. It could be that selection at elevated CO_2 levels was too weak to allow for the detection of any significant increase in growth rate and this may be the case in this experiment combined with too short a period of selection. The slight reduction in fitness of one of the pH 6.6 evolved lines could be due to a combination of very weak selection, a lack of selectively non-neutral variation and the fixation of slightly deleterious mutations through recurrent bottlenecking.

There was a significant increase in growth rate variance in the offspring from strong directional selection lines, but no detectable increase in the offspring from weak directional selection lines. This is consistent with the lack of adaptation seen in the weak directional selection lines suggesting, perhaps, that much of the genetic variance gener-

ated by recombination is neutral in this environment. Fitness variance has been shown to increase as a result of recombination in *Drosophila* studies of the RL (Mukai and Yamaguchi, 1974; Charlesworth and Charlesworth, 1975). The increase in variance indicates that there are linkage disequilibria between loci in the parental populations taken as a whole.

Together with an increase in variance there was a significant decrease in the average growth rates for the offspring from strong directional selection lines but not for offspring from weak directional selection lines. Taken together these results demonstrate that there is a measurable genetic RL and not simply a direct cost of recombination (Tucic et al., 1981; Kondrashov, 1984). The positive RLs we report range from about 5% up to 31%, which are comparable to the RLs first reported in *Drosophila* (Spassky et al., 1958; Spiess, 1958a,b; Dobzhansky et al., 1959). However, our estimates are far larger than more precise measurements in *Drosophila*: 3% measured by Mukai and Yamaguchi (1974) and 0.3% measured by Charlesworth and Charlesworth (1975). The measures of RL reported here are, however, comparable to previous estimates made in the closely related species *Chlamydomonas moewusii* of between 10.9 and 22.4% (de Visser et al., 1996). It is also interesting to note that both studies found a zero or small negative RL in some of their crosses. It could be that conditionally neutral mutations are a common feature of the genetic architecture of these species (Goho and Bell, 2000; Collins and Bell, 2004). In experiments designed to measure the potential benefit of sexual reproduction in *C. reinhardtii* for the rate of adaptation in novel environments (by generating novel variation in a Fisher-Muller scenario) Colegrave et al. (2002) and Kaltz and Bell (2002) found that there was an initial drop in the mean fitness of the populations after a round of sex, coupled with an increase in the variance of fitness. This effect, which is consistent with a RL, was found to disappear after three sexual generations and it was

concluded that it was a short-term effect, which was erased by selection and adaptation to a new environment, which presumably used up the fitness variance.

A major difference between the *Drosophila* studies and this one is that *C. reinhardtii* is facultatively sexual and for this experiment was maintained asexually, whereas *Drosophila* species are obligately sexual. As suggested by Crow and Kimura (1965) asexual populations may be expected to exhibit a higher degree of co-adaptation between loci and it may be that we are detecting antagonistic epistasis between loci in our crosses. This is far from clear, however. It could be that our asexual lines were limited by an inability to bring together beneficial mutations in a Fisher-Muller scenario (Fisher, 1930; Muller, 1932). The resulting negative linkage disequilibria would then arise from finite asexual populations and not from epistasis between loci (Colegrave, 2002). A fuller investigation of the dynamics of RL evolution through time starting from a zero measure in the ancestors may yield more insight into the potential cause of the RL.

In this study we have shown that a large and significant RL can be detected in populations of *C. reinhardtii*. Additionally, it has been demonstrated that in this system the magnitude of the positive RL does not increase as the selective environments of populations diverges. The propensity for populations to diverge from one another and become reproductively isolated under uniform selective regimes is an important implication of this result. Further studies are required to elucidate the genetic causes of the RL and to detail its evolution through time.

4 Recombination and the evolution of mutational robustness[1][2]

Abstract

Mutational robustness is the degree to which a phenotype, such as fitness, is resistant to mutational perturbations. Since most of these perturbations will tend to reduce tness, robustness provides an immediate benet for the mutated individual. However, robust systems decay due to the accumulation of deleterious mutations that would otherwise have been cleared by selection. This decay has received very little theoretical attention. At equilibrium, a population or asexual lineage is expected to have a mutation load that is invariant with respect to the selection coefcient of deleterious alleles, so the benet of robustness (at the level of the population or asexual lineage) is temporary. However, previous work has shown that robustness can be favoured when robustness loci segregate independently of the mutating loci they act upon. We examine a simple two-locus model that allows for intermediate rates of recombination and inbreeding to show that increasing the effective recombination rate allows for the evolution of greater mutational robustness.

4.1 Introduction

It has long been observed that many developmental traits display a high degree of phenotypic robustness, that is, the phenotype is remarkably immune to environmental and genetic perturbations (Waddington, 1940; Schmalhausen, 1949). Waddington (1942) described the phenomenon as canalization, and proposed an adaptationist explanation. He reasoned that traits under stabilising selection towards some intermediate optimum

[1] In collaboration with A. Gardner.

[2] Published in *J Theor Biol* (2006), 241:707-715.

should benefit from any mechanism that prevents deviation from that optimum due to either genetic or environmental perturbations. Within the class of genetic perturbations are those that are due to deleterious mutations. In recent years, mutational robustness has attracted renewed interest, on both theoretical and empirical fronts (Wagner et al., 1997; Wilkins, 1997; Rutherford and Lindquist, 1998; van Nimwegen et al., 1999; Kawecki, 2000; Wagner, 2000; Wilke, 2001; Wilke et al., 2001; Queitsch et al., 2002; Wilke and Adami, 2003; de Visser et al., 2003; Proulx and Phillips, 2005). Most attention has been given to adaptationist explanations, although some researchers have speculated that mutational robustness is a by-product of adaptation for environmental perturbations (Wagner et al., 1997; Burch and Chao, 2004) or simply an emergent property of genetic systems (Kacser and Burns, 1981; von Dassow et al., 2000; Meir et al., 2002; Shen-Orr et al., 2002).

All genetic models for the evolution of robustness require some form of gene interaction or epistasis between the loci involved. In this respect a distinction can be drawn between the two different models of robustness that are commonly discussed; whether the epistasis is exclusively between the loci involved in the trait or between the trait loci and an unrelated locus (a modifier). There is evidence for the former in RNA and protein folding where there is often extensive degeneracy between the primary sequence and the secondary or tertiary structure (Maynard Smith, 1970; Lau and Dill, 1990; Schuster et al., 1994; van Nimwegen et al., 1999; Wilke, 2001) and in metabolic and developmental pathways in prokaryotes and eukaryotes, where distributed network architectures facilitate robustness via internal pathway degeneracy (Edwards and Palsson, 1999, 2000a,b; von Dassow et al., 2000; Meir et al., 2002; Shen-Orr et al., 2002; Ingolia, 2004; Wagner, 2005). Wagner (2005) refers to this form of robustness as distributed robustness and argues, based on empirical evidence, for the primacy of its role in mu-

tational robustness in favour of gene redundancy arising from gene duplicates (Wagner, 2000, 2001, 2005). There is evidence for a modifier of robustness in higher eukaryotes where a potential canalizing gene, Hsp90, has been shown to buffer genetic variation for a wide-range of phenotypic traits in both *Drosophila* (Rutherford and Lindquist, 1998) and *Arabidopsis* (Queitsch et al., 2002), and GroEL in *Escherichia coli* (Fares et al., 2002), behave as modifiers of mutational robustness. For the remainder of this article we will restrict attention to the modifier view of robustness.

The evolution of mutational robustness is conceptually similar to the adaptive evolution of dominance proposed by Fisher (1928). In both cases it is the heritable deviation from the wild type that is being buffered, and the selective advantage of the modifier is of the order of the mutation rate (Wright, 1929). Fisher believed that although the selective advantage is weak, in a large population with a number of recessive mutations the accumulated selective pressure would drive the evolution of dominance. Wright took the view that dominance emerged as an intrinsic property of metabolic pathways and proposed an alternative physiological theory of dominance (Wright, 1934). Kacser and Burns (1981) provided considerable support for Wright's argument with a model of a multienzyme system that showed that the flux of the enzyme pathway is insensitive to concentration changes in the enzymes involved, suggesting that dominance is an inevitable property of such systems. More recently, however, Bagheri and Wagner (2004) have shown that this may only be the case when one neglects nonlinear enzyme interactions. Currently, empirical evidence appears not to support Fisher's adaptationist hypothesis (Orr, 1991), although it may be relevant in situations involving strong selection (Haldane, 1956; Mayo and Burger, 1997). The debate continues. Another related phenomenon that has received much attention is the evolutionary transition fromhaploidy to diploidy. A benefit may be afforded by an extended diploid phase due to the

masking of recessive or partially recessive deleterious mutations (Crow and Kimura, 1965). Here, the adaptations view appears to have a plausible theoretical foundation (Kondrashov and Crow, 1991; Perrot et al., 1991) although, interestingly, it is incompatible with Fisher's view of dominance since it requires that newly arisen deleterious mutations are always (at least) partially recessive (Perrot et al., 1991). Together with the evolution of mutational robustness, these scenarios involve evolutionary modification of the genetic system itself driven by the immediate benefit of alleviating the effects of deleterious mutations, which are of course a ubiquitous evolutionary phenomenon.

A classic result that motivates the present study is that at equilibrium the mutation load (L^*) of the population is invariant with respect to the fitness consequences of deleterious alleles. Assuming fitnesses combine multiplicatively across loci, an allele which arises by recurrent irreversible mutation at rate μ and incurs a fitness decrement s will equilibrate at frequency $\frac{\mu}{s}$ in a haploid population (mutation-selection balance). Hence the average fitness contributed by this locus is $(1 - \frac{\mu}{s} \times 1 + (\frac{\mu}{s} \times (1 - s))) = 1 - \mu$; the mutation load at this locus is then μ, and not a function of s (Haldane, 1937). The result has been generalised for all loci in the genome, giving a population load of mutations $L^* = 1 - e^{-U}$(Kimura and Maruyama, 1966; Kondrashov, 1984), where U is the per genome per generation mutation rate, and hence the decrement to fitness due to individual mutationsis again irrelevant. The reason for this is intuitive: if mutations are more harmful they are more readily removed from the population by selection. Those mutations with large deleterious effects are held at low frequency at mutation-selection balance, and thus cause the same decrement to the mean fitness of the population as less harmful, and hence more frequently encountered, mutations.

The action of mutational robustness is to reduce the magnitude of a mutations fitness effect. Whilst it may be temporarily advantageous to reduce the selection coefficient as-

sociated with the deleterious mutation, this leads to the accumulation of mutations that would otherwise have been cleared by selection, and so a closed population (i.e. no flow of genetic material between populations) with enhanced robustness does not improve its equilibrium mutation load. Thus there is no long-term benefit for being robust. This mutational decay of robust systems has received only limited attention (Frank, 2003). If robustness has an intrinsic cost, then in the long term it will cause a net disadvantage for the population. Therefore, in an asexual population, we predict eventual loss of robust lineages. However, robustness might be favoured in a sexual population. Since the benefit of robustness (a reduced impact of mutations) accrues to the robust lineage, whilst the cost (an increased load of mutations) is shared by the whole population, robust lineages may have a relative advantage. It seems that this will increasingly be the case as the rate of recombination (in particular, between robustness genes and those genes undergoing mutational perturbation) is increased. This has received some attention, and the hypothesis is supported by contrasting the predictions of models of complete linkage, in which costly robustness is never favoured (Hermisson et al., 2002), with those which assume free recombination, in which costly robustness can evolve (Wagner et al., 1997; Dawson, 1999). However, results for robustness evolution with the intermediate recombination rates are lacking (de Visser et al., 2003).

We examine a simple model that captures the essence of this problem. The dynamics of the system are described using a multi-locus methodology (developed by Barton and Turelli, 1991; Kirkpatrick et al., 2002) that highlights allele frequencies and linkage disequilibria, which is more natural than following genotype frequencies. Also, it provides a general notation that neatly partitions the various causes of evolutionary change, and allows for arbitrary complexity so that the model is readily extensible within this single framework. Specifically, we develop exact analytical recursions describing the dynam-

ics of a costly robustness modifier and its association with a mutating locus, and from this we generate an invasion condition to determine when this modifier will increase in frequency when vanishingly rare. We then make an assumption of minor robustness variants to examine how the robustness phenotype evolves in the longer term, moving from multi-locus population genetics to an evolutionary game theoretic analysis.

4.2 Model and analysis

4.2.1 Two-locus model

We consider a simple model which captures all the important features of this problem a large population of sexual haploids, with a lifecycle which involves (i) selection, followed by (ii) mutation, and finally (iii) mating to form diploid zygotes, which undergo meiosis to form the next generation of haploid individuals. All notation used in this article is summarized and defined in Table 4.1 and Table 4.2. A locus i suffers recurrent, irreversible mutation, from the wildtype allele with value $X_i = 0$ to mutant allele with value $X_i = 1$, at a rate μ. The fitness of the wildtype is 1, and the fitness of the mutant is $1s$ in the absence of robustness. The frequency of the mutant is denoted p_i, and thus the frequency of the wildtype is $q_i = 1 - p_i$. A second locus j controls the expression of the deleterious mutant, when it occurs at the first locus: with robustness k, the fitness contributed by the first locus is $1 - (1 - k)s$. Two alleles, with varying robustness effect, are present. The resident allele has value $X_j = 0$ and robustness effect kx, and the 'variant allele has value $X_j = 1$ and robustness effect ky. The robustness locus also incurs a direct cost, with the resident contributing $1 - cx$, and the variant $1 - cy$, to an individuals fitness. The frequency of the variant is denoted pj and the frequency of the resident is $q_j = 1 - p_j$. We will assume that the direct effects of the loci multiply to give genotype fitness. The four genotype fitnesses are summarised in Table 4.3. The effects of linkage

and inbreeding are described by an effective rate of recombination parameter, r_e.

Following the above model, an individual's fitness may be written in the form:

$$\begin{aligned} w =& (1-X_i)(1-X_j)w_{00} + X_i(1-X_j)w_{10} + (1-X_i)X_j w_{01} \\ &+ X_i X_j w_{11} = (1-X_i)(1-X_j)(1-c_x) \\ &+ X_i(1-X_j)(1-(1-k_x)s)(1-c_x) \\ &+ (1-X_i)X_j(1-c_y) + X_i X_j(1-(1-k_y)s)(1-c_y), \end{aligned} \tag{1}$$

where $w_{X_iX_j}$ is the fitness of the (X_i, X_j) genotype (see Table 4.3). This fitness function is analogous to Eq. (7) in Barton and Turelli (1991).

4.2.2 Multilocus population statistics

The multilocus framework of Kirkpatrick et al. (2002) describes individuals and populations according to deviations from average values. An allelic deviation ($\zeta_i = X_i - p_i$) is defined for a generic gene position **i**, and describes the deviation of the allelic value (X_i = 0 or 1) from the population average (p_i) at that position. Thus, the population average allelic deviation for a single gene position is zero. A corresponding deviation term ($\zeta_A = \prod_{i \in A} \zeta_i$) may be assigned to a set **A** of gene positions, and is the product of the allelic deviations for all the positions in that set. Note that the average deviation for a set of two gene positions (**i** and **j**) is equal to the allelic covariance between these positions ($E[\zeta_{ij}] = E[(X_i - p_i)(X_j - p_j)] = Cov[X_i, X_j]$), and thus is equivalent to the linkage disequilibrium (D_{ij}) between these gene positions. In general, the population average deviation for a set **A** of gene positions will be denoted D_A. Thus, the population composition with respect to a set of gene positions **B** may be fully described by the set of allele frequencies (p_i, $\mathbf{i} \in \mathbf{B}$) at these positions, and the statistical associations (D_A, $\mathbf{A} \subseteq \mathbf{B}$) between these positions. If an association term corresponds to a set of

Table 4.1. Summary of notation used. Continued in Table 4.2.

Notation	Definition
μ	Deleterious mutation rate at a single locus
s	Selection coefficient associated with deleterious mutation
U	Per genome per generation mutation rate
L^*	Equilibrium mutation load
i	Locus under recurrent mutation
j	Locus controlling robustness
$\mathbf{i},\mathbf{j}$	A genetic gene position
$\mathbf{A},\mathbf{B}$	A generic set of gene positions
$\mathbf{W}$	The set of all gene positions contributing to fitness
X_i	Allelic value for gene position i (0 or 1)
p_i	Frequency of the $X_i = 1$ allele
$q_i = 1 - p_i$	Frequency of the $X_i = 0$ allele
$\zeta_i = X_i - p_i$	Allelic deviation for gene position i
$\zeta_A = X_i - p_i$	Allelic deviation for set A of gene positions
$D_A = E[\zeta_A]$	Association for set A of gene positions
w	Fitness of an individual
$\bar{w}$	Population mean fitness
a_A	Multilocus selection coefficient for set A of gene positions
z	A generic robustness strategy
x	Resident robustness strategy
$y = x + \delta x$	Variant robustness strategy
x^*	Equilibrium robustness strategy

Table 4.2. Notation continued from Table 4.1.

Notation	Definition
k_z, $k[z]$	Robustness effect associated with strategy z
c_z, $c[z]$	Cost of robustness associated with strategy z
r_e	Effective rate of recombination
$\lambda = 1 + \delta\lambda$	Invasion fitness of robustness variant; its asymptotic rate of increase

Table 4.3. Genotype fitness (w) as a function of allelic value ($X = 0, 1$) at the mutating locus (i) and the robustness locus (j).

		X_j	
		0	1
X_i	0	$1 - c_x$	$1 - c_y$
	1	$(1-(1-k_x)s)(1-c_x)$	$(1-(1-k_y)s)(1-c_y)$

gene positions in which a particular position features several times, for example D_{iiA}, then a reduction formula may be applied to re-express this as $p_i q_i D_A + (1 - 2p_i)D_{iA}$, as outlined by Kirkpatrick et al. (2002).

We may now describe how sets of gene positions impact upon an individual's fitness. Making the substitution $X_i = \zeta_i + p_i$ into the fitness function (1), this may be rearranged into the form

$$\frac{w}{\bar{w}} = 1 + a_i(\zeta_i - D_i) + a_j(\zeta_j - D_j) + a_{ij}(\zeta_{ij} - D_{ij}), \tag{2}$$

where a_A is the contribution of the deviation for a set of gene positions A to relative fitness $\frac{w}{\bar{w}}$. This is analogous to Eq. (6) of Barton and Turelli (1991) and Eq. (7) of Kirkpatrick et al. (2002). The a_A terms provide selection coefficients for a multilocus

analysis (Kirkpatrick et al., 2002). For the present model, we have

$$a_i = -s(1-((1-p_j)(k_x+c_x(1-k_x))+p_j(k_y+c_y(1-k_y))))/\bar{w} \tag{3}$$

and

$$a_j = (sp_i((k_y+c_y(1-k_y))-(k_x+c_x(1-k_x)))-(c_y-c_x))/\bar{w} \tag{4}$$

and

$$a_{ij} = s((k_y+c_y(1-k_y))-(k_x+c_x(1-k_x)))/\bar{w}. \tag{5}$$

Mean fitness is found by taking an average of w over the population:

$$\bar{w} = 1-sp_i(1-k_x)(1-c_x)-(p_jc_y+(1-p_j)c_x)-s((1-k_y)(1-c_y)-(1-k_x)(1-c_x))(p_ip_j+D_{ij}). \tag{6}$$

Having extracted multilocus population fitness statistics from the model, we can now use them to make some remarkably elegant statements about how selection moulds the allele frequencies and linkage disequilibrium of this system. After selection we will consider mutation and then transmission.

4.2.3 Selection

The multilocus methodology provides simple recursion expressions for the change in allele frequencies and genetic associations due to selection. The basic equation is

$$\Delta_S D_A = \sum_{B\subseteq W} a_B(D_{AB}-D_AD_B) \tag{7}$$

where W is the set of all gene positions contributing to fitness. For the case of a single gene position ($A=i$), we may use expression the general equation to describe the change in allele frequency (p_i) due to selection:

$$\Delta_S p_i = \sum_{B\subseteq W} a_B D_{iB}. \tag{8}$$

A complication arises in that the association after selection ($D'_A = D_A + \Delta_S D_A$) is described with respect to allele frequencies before selection. It will usually be helpful to correct for this, and the procedure is described in Kirkpatrick et al. (2002). No correction is necessary for the expressions describing allele frequency change. In the context of the present model, the change in the frequency of the deleterious mutation that is due to selection is described by

$$p'_i = p_i + a_i D_{ii} + a_j D_{ij} + a_{ij} D_{iij} = p_i + a_i p_i q_i + a_j D_{ij} + a_{ij}(1 - 2p_i) D_{ij}. \quad (9)$$

This notational framework makes clear the causes of evolutionary change: here we see that the response to selection ($\Delta_S p_i = p'_i - p_i$) is given by the product of the strength of selection opearting directly on the focal locus (a_i) and the variation at that locus ($p_i q_i$), plus the product of selection operating directly on the other locus (a_j) and the association between the two loci (D_{ij}), plus the product of selection due to epistatic interaction between the two (a_{ij}) and the appropriate association ($(1 - 2p_i)D_{ij}$). Similarly, the change in allele frequency due to selection at the robustness locus is given by

$$p'_j = p_j + a_i D_{ij} + a_j D_{jj} + a_{ij} D_{ijj} = p_j + a_i D_{ij} + a_j p_j q_j + a_{ij}(1 - 2p_j) D_{ij}. \quad (10)$$

From expression (5), the change in the association between loci is described by

$$\begin{aligned} D'_{ij} &= D_{ij} + a_i D_{iij} + a_j D_{ijj} + a_{ij}(D_{iijj} - D_{jj}^2) - (p'_i - p_i)(p'_j - p_j) \\ &= D_{ij} + a_i(1 - 2p_i) D_{ij} + a_j(1 - 2p_j) D_{ij} + a_{ij}(p_i q_i p_j q_j + (1 - 2p_i)(1 - 2p_j) D_{ij} - D_{ij}^2) \\ &\quad - (p'_i - p_i)(p'_j - p_j), \end{aligned} \quad (11)$$

where the trailing term corrects for change in allelic frequency (Kirkpatrick et al., 2002).

4.2.4 Mutation

The change in frequencyof the deleterious allele after mutation is described by

$$p''_i = p'_i + \mu(1 - p'_i). \quad (12)$$

Since the j locus does not undergo mutation, $p''_j = p'_j$. From Kirkpatrick et al. (2002), the change in the linkage disequilibrium due to mutation is given by

$$D''_{ij} = (1-\mu)D'_{ij}. \tag{13}$$

4.2.5 Transmission

Transmission - the union of gametes, crossing over, and fair meiosis - does not alter the allele frequencies in this model (so $p'''_i = p''_i$ and $p'''_j = p''_j$, where triple primes denote the variable is measured after transmission), but it does impact on the linkage disequilibrium. This is reduced by a fraction equal to the effective rate of recombination (Crow and Kimura, 1970), and so we have

$$D'''_{ij} = (1-r_e)D''_{ij}. \tag{14}$$

4.2.6 Invasion analysis

We have obtained recursions describing the change in frequencies of the deleterious mutation (p_i) and robustness modifier (p_j) and the linkage disequilibrium (D_{ij}) over a single generation incorporating selection, mutation and transmission. We now consider that the variant robustness allele is vanishingly rare ($p_j \to 0$, and hence $D_{ij} \to 0$), and examine the conditions under which this rare allele will increase in frequency (invasion). We will assume that μ is sufficiently small for us not to have to worry about fixation of the deleterious mutation i.e. $\mu < (1-k_x)s$. We will assume that the deleterious mutation is initially at its equilibrium point, $p^* = \mu/((1-k_x)s)$. While the variant is rare, the evolutionary dynamics at the j locus has vanishing impact on dynamics at the i locus, so in any generation we may express the frequency of the deleterious mutation as $p_i = p_i^* + \delta p_i$, where $\delta p_i \to 0$. Making this substitution, and summarising the changes in the allele frequency at the robustness locus and linkage disequilibrium due to selection,

mutation and recombination, obtains $p_j''' = \alpha_1 p_j + \alpha_2 D_{ij} + O(\delta p_i^2, p_j^2, D_{ij}^2)$ and $D_{ij}''' = \alpha_3 p_j + \alpha_4 D_{ij} + O(\delta p_i^2, p_j^2, D_{ij}^2)$, where

$$\begin{aligned}
\alpha_1 &= \frac{(1-c_y)(1-k_x-\mu(1-k_y))}{(1-\mu)(1-c_x)(1-k_x)} \\
\alpha_2 &= -\frac{s(1-c_y)(1-k_y)}{(1-\mu)(1-c_x)} \\
\alpha_3 &= \frac{(1-r_e)(k_y-k_x)(1-c_y)\mu((1-k_x)s-\mu)}{(1-c_x)(1-k_x)^2 s(1-\mu)} \\
\alpha_4 &= \frac{(1-r_e)(1-c_y)((1-k_x)(1-(1-k_y)s)-\mu(k_y-k_x))}{(1-\mu)(1-c_x)(1-k_x)}.
\end{aligned} \tag{15}$$

Neither of these recursions are functions of δp_i, therefore we need not explicitly follow the frequency of the deleterious mutation, so long as we assume it is close to its equilibrium. The asymptotic rate of increase of the rare variant, its 'invasion fitness', is given by the leading eigenvalue for the above system. This is the solution λ to the characteristic equation $(\alpha_1 - \lambda)(\alpha_4 - \lambda) - \alpha_2\alpha_3 = 0$ that has the largest magnitude. The condition for invasion of the robustness variant is $\lambda > 1$.

4.2.7 Evolution of robustness

We have obtained a condition for the invasion of a given resident population by a given variant robustness allele.We now ask the following questions: (1) is there a resident allele that cannot be invaded by any robustness variant? (2) Will the population converge on this evolutionary stable state? In other words, we are interested in identifying the endpoint of robustness evolution in the longer term. To address this, we will now consider a continuum of robustness strategies (z), from zero robustness ($z = 0$) to full robustness ($z = 1$), each encoded by an allele at the j locus. the cost and effect of robustness parameters from the previous sections are now considered as functions of the robustness strategy ($c[z]$ and $k[z]$; where $c[0] = k[0] = 0$ and dc/dz, $dk/dz > 0$ for all z). The resident allele encodes the robustness strategy $z = x$, and the variant encodes

$z = y$. Thus, $c_x = c[x]$, $k_x = k[x]$, $c_y = c[y]$ and $k_y = k[y]$. For the ease of analysis, we will consider only local stability, restricting our attention to $y = x + \delta x$ where $\delta x \to 0$. Since this represents near-neutrality, the invasion fitness of the variant will be of the form $\lambda = 1 + \delta\lambda$, where $\delta\lambda \to 0$. Upon this assumption, we may solve the characteristic equation from earlier to obtain $\delta\lambda \to ((1-\alpha_1)(1-\alpha_4) - \alpha_2\alpha_3)/(\alpha_1 + \alpha_4 - 2)$ as $\delta x \to 0$, or

$$\delta\lambda \approx \frac{\mu r_e(1-c[x])k'[x] - (1-k[x])(r_e + (1-r_e)(1-k[x])s - \mu)c'[x]}{(1-c[x])(1-k[x])(r_e + (1-r_e)(1-k[x])s - \mu)}\delta x, \qquad (16)$$

where the primes denote derivatives evaluated at the resident robustness strategy, i.e. $c'[x] = dc[z]/dz|_{z=x}$ and $k'[x] = dk[z]/dz|_{z=x}$. Marginal invasion fitness is given by $\partial\lambda/\partial y|_{y=x} = \delta\lambda/\delta x$. Setting $r_e = 0$, marginal invasion fitness reduces to $\partial\lambda/\partial y|_{y=x} = -c'[x]/(1-c[x])$, which is negative for all x: over the whole range of resident strategies, selection favours variants with reduced robustness. Hence, the only equilibrium point in the absence of recombination is at $x^* = 0$. This means that when the effective rate of recombination is zero, costly robustness cannot evolve (Hermisson et al., 2002). We now ask, for $r_e > 0$, what is the end point of robustness evolution? We are therefore looking for a strategy that is both evolutionarily stable once attained (an ESS Matynard Smith and Price, 1973; Maynard Smith, 1982), and is also attainable (i.e. convergence stable, so that when x is close to x^*, y closer to x^* will invade) (Eshel and Motro, 1981; Taylor, 1996). A strategy that is both an ESS and is convergence stable is termed a 'continuously stable strategy' (CSS) (Eshel, 1983; Christiansen, 1991). If a strategy x^* is evolutionarily stable, then it must satisfy $\partial\lambda/\partial y|_{y=x=x^*} = 0$. Thus,

$$\frac{d}{dr_e}\left[\frac{\partial\lambda}{\partial y}|_{y=x=x^*}\right] = \frac{\partial}{\partial r_e}\left[\frac{\partial\lambda}{\partial y}|_{y=x=x^*}\right] + \frac{\partial}{\partial x^*}\left[\frac{\partial\lambda}{\partial y}|_{y=x=x^*}\right]\frac{dx^*}{dr_e} = 0, \qquad (17)$$

which can be re-arranged to give

$$\frac{dx^*}{dr_e} = -\frac{\partial/\partial r_e[\partial\lambda/\partial y|_{y=x=x^*}]}{\partial/\partial x^*[\partial\lambda/\partial y|_{y=x=x^*}]}. \qquad (18)$$

Noting that convergence stability implies $\partial[\partial\lambda/\partial y|_{y=x=x^*}]/\partial x^* < 0$ (Taylor, 1996), the CSS satisfies

$$sgn\left[\frac{dx^*}{dr_e}\right] = sgn\left[\frac{\partial}{\partial r_e}\left[\frac{\partial\lambda}{\partial y}|_{y=x=x^*}\right]\right] \tag{19}$$

(Pen, 2000), where the function sgn returns the sign, or sense, of its real argument, i.e. positive or negative or zero. The partial derivative on the RHS is

$$\frac{\partial}{\partial r_e}\left[\frac{\partial\lambda}{\partial y}|_{y=x=x^*}\right] = \frac{\mu((1-k[x^*])s-\mu)k'[x^*]}{(1-k[x^*])(r_e+(1-r_e)(1-k[x^*])s-\mu)}. \tag{20}$$

Since RHS of (20)>0, it follows from (18) that the CSS x^* (when it exists) is a monotonically increasing function of r_e. Thus, we expect the endpoint of evolution to be a greater degree of robustness the higher the effective rate of recombination. Applying the same procedure to the selection coefficient (s) and mutation rate (μ) obtains

$$\frac{\partial}{\partial s}\left[\frac{\partial\lambda}{\partial y}|_{y=x=x^*}\right] = -\frac{\mu r_e(1-r_e)k'[x^*]}{(r_e+(1-r_e)(1-k[x^*])s-\mu)^2} < 0 \tag{21}$$

and

$$\frac{\partial}{\partial \mu}\left[\frac{\partial\lambda}{\partial y}|_{y=x=x^*}\right] = \frac{r_e(r_e+(1-r_e)(1-k[x^*])s)k'[x^*]}{(1-k[x^*])(r_e+(1-r_e)(1-k[x^*])s-\mu)^2} \tag{22}$$

i.e. the CSS x^* is a decreasing function of s and an increasing function of μ, so we expect the endpoint of evolution to be a greater degree of robustness as we decrease the magnitude of the deleterious effect of mutations and as we increase the mutation rate. From (16), the exact value of the CSS x^* can be found by solving the equation

$$\mu r_e(1-c[x])k'[x] - (1-k[x])(r_e+(1-r_e)(1-k[x])s-\mu)c'[x] = 0. \tag{23}$$

Some representative numerical examples are given in Figures 4.1 and 4.2. The assumption of vanishing variation is somewhat artificial, and so we have used simulations to test the predictions using a similar two-locus model that allows for continuum alleles which are simultaneously extant (simulation results are presented in Figures 4.3 and

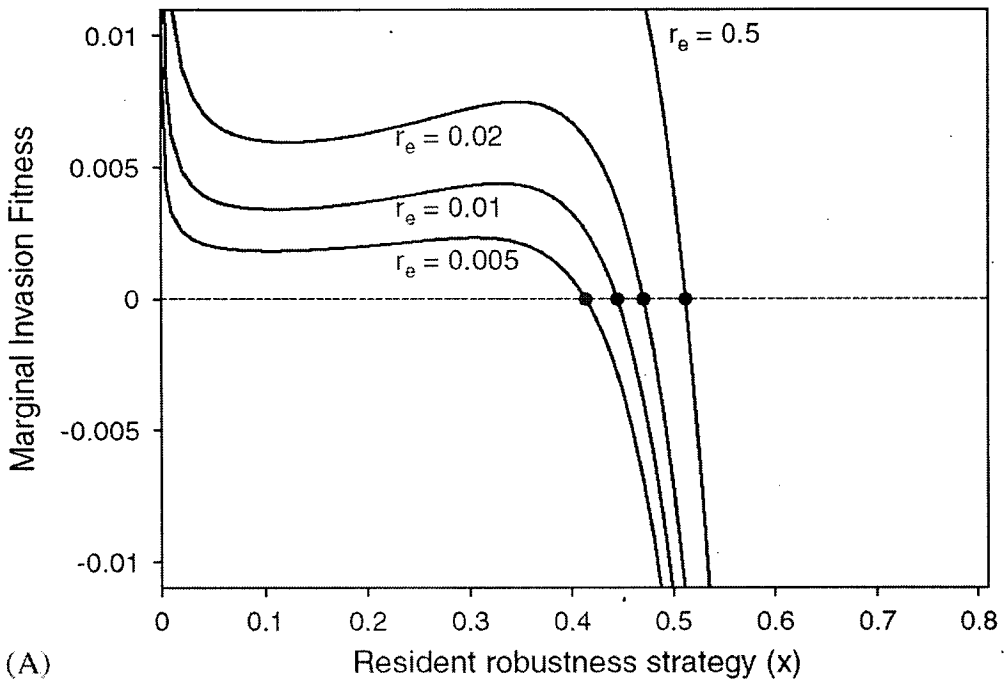


Figure 4.1. Marginal invasion fitness $\partial\lambda/\partial y|_{y=x}$ as a function of the resident robustness strategy (x) and the effective recombination rate (r_e) assuming $\mu = 0.01$, $s = 0.1$, $c[z] = z^{10}$, $k[z] = z^{1/2}$. The sign of marginal invasion fitness determines the direction of selection; if it is positive then variant strategies increasing robustness are favoured, and if it is negative then variant strategies reducing robustness are favoured. The CSS is an increasing function of the effective rate of recombination.

4.4). We find that numerical solutions to the analytical prediction given by (23) and the results of the simulations are generally in good agreement.

Depending on the choice of parameters and robustness functions, there may be: (1) a single internal equilibrium, which is a CSS (Figure 4.1); (2) an unstable equilibrium in addition to the CSS (Figure 4.2, e.g. lines for $r_e = 0.1, 0.05, 0.01$); or (3) no internal equilibria (Figure 4.2, e.g. $r_e = 0.5$). The simulations confirm that a population initial-

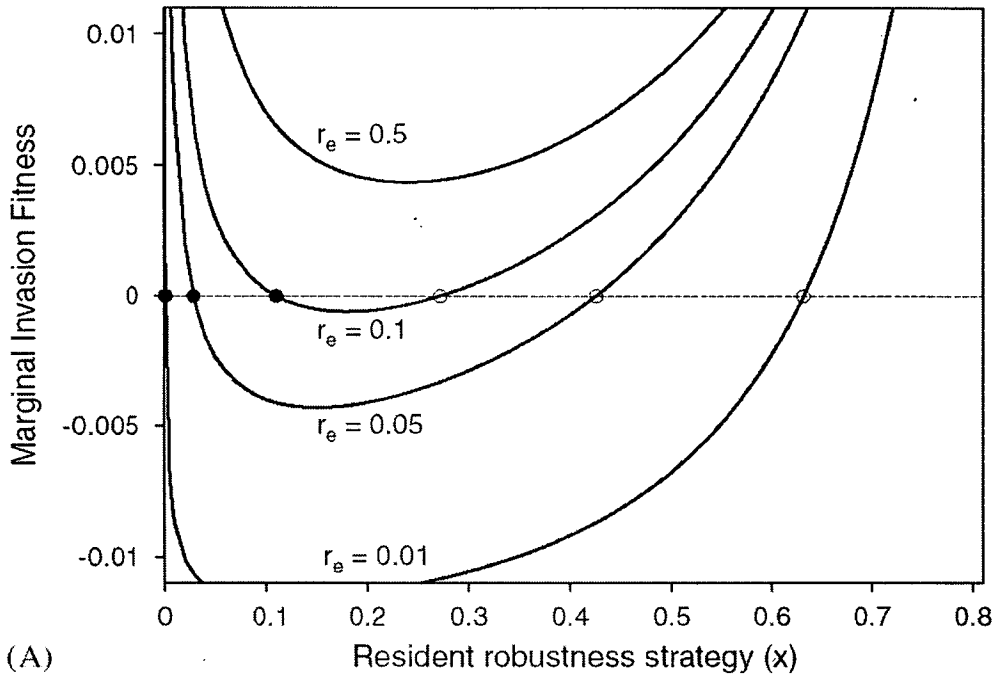


Figure 4.2. Marginal invasion fitness $\partial\lambda/\partial y|_{y=x}$ as a function of the resident robustness strategy (x) and the effective recombination rate (r_e) assuming $\mu = 0.01$, $s = 0.1$, $c[z] = 0.15z$, $k[z] = z^{1/2}$. The sign of marginal invasion fitness determines the direction of selection; if it is positive then variant strategies increasing robustness are favoured, and if it is negative then variant strategies reducing robustness are favoured. The CSS (where it exists, e.g. for $r_e = 0.01, 0.05, 0.1$, marked by a filled circle) is an increasing function of the effective rate of recombination; there is also sometimes an unstable equilibrium (marked by an empty circle).

ized at close to zero robustness will ultimately find itself trapped at the CSS, where this exists (Figures 4.3 and 4.4).

Due to the assumption that mutation-selection balance holds the deleterious mutation at intermediate frequency (i.e. $\mu < (1 - k_x)s$, which may also be written as $k_x < 1 - \mu/s$) the present analysis does not allow for examination of the evolution of

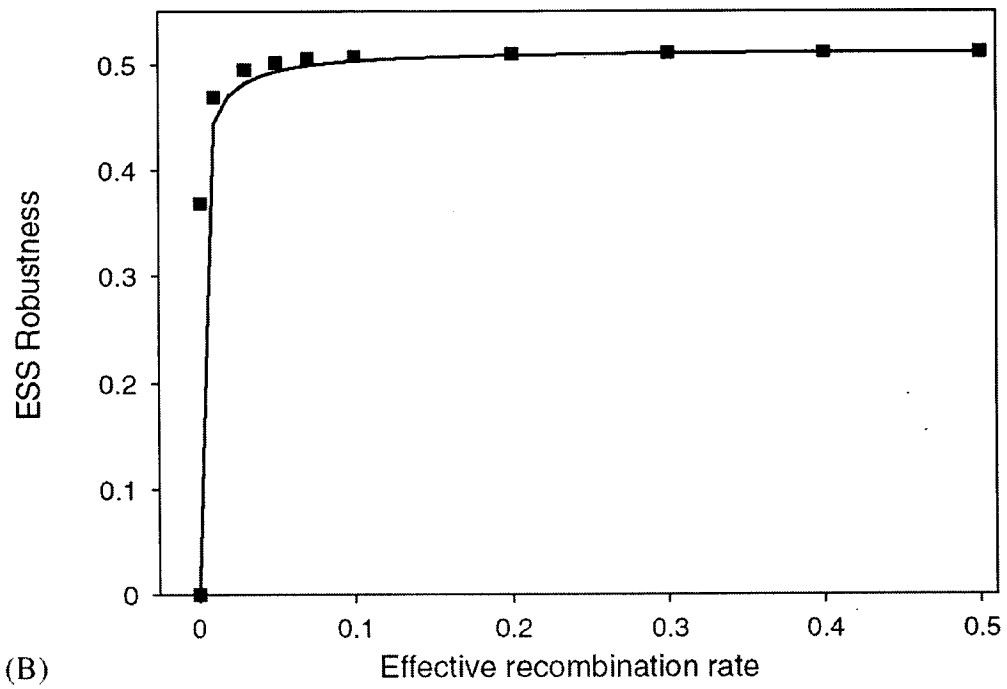
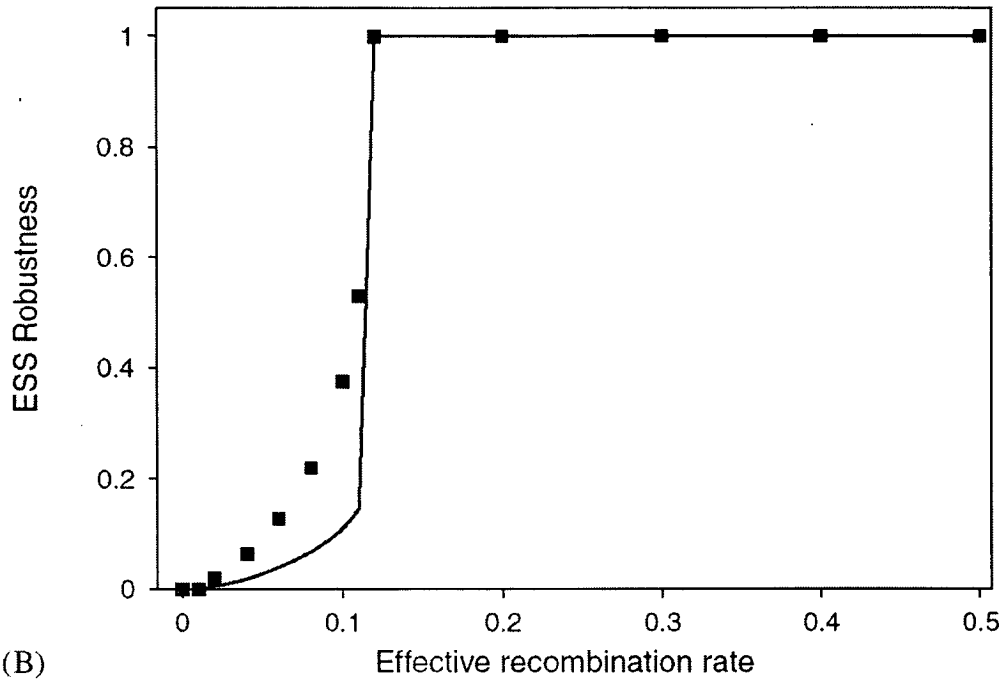


Figure 4.3. Simulations results for $\mu = 0.01$, $s = 0.1$, $c[z] = z^{10}$, $k[z] = z^{1/2}$ (Figure 4.1). Simulation results (squares) confirm the analytical prediction (line) that robustness increases with the effective rate of recombination.

almost complete robustness ($k_x = 1$).

Although the simulations and analytical predictions have a very good fit, they are not perfect. In particular, the simulations tend to give an end-point of robustness evolution that is higher than predicted from the game theoretic approach. There is reason to suspect that the invasion analysis underestimates the benefits of robustness. For example, consider a variant robustness allele with associated cost such that its asymptotic rate of increase is exactly 1. Initially, this allele will increase in frequency, and will eventually settle at a neutral equilibrium. To some extent then, this allele has been favoured, although technically it does not invade.



(B)

Figure 4.4. Simulation results (squares) confirm the analytical prediction (line) that robustness increases with the effective rate of recombination, and that there is no internal stable endpoint for $r_e > 0.11$. $\mu = 0.01$, $s = 0.1$, $c[z] = 0.15z$, $k[z] = z^{1/2}$ (Figure 4.2).

More generally, deviations from the analytical predictions will occur due to slowness in attaining the evolutionary endpoint, as selection acts weakly upon robustness and only a finite number of generations are simulated.

4.3 Discussion

We have examined the evolution of costly robustness in a simple two-locus model for when recombination (r_e) between the two loci is intermediate. Previously, only the extremes of zero recombination (Hermisson et al., 2002) and freely recombining loci

(Wagner et al., 1997; Dawson, 1999) have been considered. A multilocus methodology has been employed to obtain recursions for allele frequencies at the robustness locus and the association between this locus and the locus that is under mutation. The result is an analytical condition for when the robustness variant invades a population. Restricting attention to minor variants, we have used this condition to determine how the end point of robustness evolution varues with the effective rate of recombination (r_e), the intrinsic deleterious effect of the mutation (s) and the mutation rate (μ). Consistent with previous theory, we find that costly robustness cannot be favoured when the effective rate of recombination is zero. In addition, we show that, where one exists, the internal stable endpoint of robustness evolution is an increasing function of effective recombination rate and the mutation rate, and is a decreasing function of the intrinsic deleterious effect of the mutation. Although the analysis assumes vanishing robustness variation in the population at any time, simulations that relax this assumption reveal the analytical treatment is robust.

Why do we predict enhanced robustness with increasing effective rate of recombination? Recombination favours robustness in two ways: (1) by allowing a robust lineage to discard the excess of deleterious mutations it has accumulated, and (2) these deleterious mutations are inflicted upon non-robust lineages where they cause enhanced damage to fitness, thus increasing the relative fitness of the robust lineages. Put another way, by breaking down the association between the robustness gene and the target of the robustness effect, recombination decouples the immediate benefit of robustness (enhanced fitness in the context of a mutant genotype) from the long-term cost (increased frequency of mutations at mutation-selection equilibrium); the former accruing only to robust individuals, and the latter being paid by the population as a whole. This benefit for robustness is mirrored in the Perrot et al. (1991) model for the evolution of

diploidy, which features gene flow between haploids and diploids. With this in mind, the model predicts increased maladaptation in the genomes of sexual, outbred populations, whereas the genomes of asexual or inbred populations should be less afflicted with the mutationally decayed remains of robust networks.

The endpoint of robustness evolution is predicted to be a decreasing function of the selection coefficient associated with deleterious mutations. This is because as the strength of selection upon the mutating locus increases, so does the epistasis between the robustness and mutating loci, which results in a greater association between these. Since it is the build-up of this linkage disequilibrium which acts to disfavour the evolution of robustness, weaker selection against the deleterious mutant favours enhanced robustness. An analogous result emerged from the Perrot et al. (1991) model for the evolution of diploidy. Interestingly, this is responsible for the run-away selection for robustness observed in some of the simulations (4.4), because as the population becomes more robust to deleterious mutations there is a reduced selection acting upon the mutating locus, and thus a lower build-up of linkage disequilibrium between the two loci. Intuitively, it would seem that mutational robustness should be increasingly favoured as deleterious mutations become more, rather than less, harmful. Yet it is the mutation load and not the mutation effect that is crucial (Proulx and Phillips, 2005), so the endpoint of robustness evolution is an increasing function of the mutation rate and not the deleterious mutational effect. This is consistent with the view of Wright (1929) that selection for mutational robustness will be of the order of the mutation rate.

It is of interest to compare the present results with previous models for the evolution of mutational robustness. Wagner et al. (1997) investigated the evolution of a modifier of mutational robustness impacting upon a number of loci underlying a quantitative trait under Gaussian stabilising selection. Individual-based simulations showed that the

selection coefficient acting on the modifier tends to increase with the intensity of the stabilising selection, a result that was verified in deterministic simulations by Kawecki (2000). This is in contrast to the result we report here where the modifier invades most easily at lower strengths of selection against the deleterious allele because it is here that the modifier experiences weakest linkage disequilibrium with the deleterious allele. Wagner et al. (1997) attribute their result to stronger stabilising selection enabling stronger selection for canalisation. Their result crucially depends on the possibility of back mutation, plus high mutation rates, so that back mutation is strong relative to selection. The present analysis assumes no back mutation, and this appears to be the reason for the disparity. The neglecting of back mutation seems reasonable if there are many alleles that give rise to defective gene products and only a few that code for a correctly functioning protein. However, further work is needed to clarify the impact of back mutation on the evolution of robustness.

We note some possibilities for the evolution of synergistic epistasis, where an individual's fitness declines more rapidly with increasing numbers of deleterious mutations than predicted by a multiplicative fitness scheme. In many models, mutational robustness is synonymous with synergistic epistasis de Visser et al. (2003); Michalakis and Roze (2004). For example the classic 'neutral network' (van Nimwegen et al., 1999) models of robustness - involving individuals with less than some threshold number of mutations having wild-type fitness, and individuals exceeding that threshold being inviable - presents an extreme form of synergistic epistasis. A substantial amount of theory has been devoted to the evolution of sex and recombination given synergistic epistasis between deleterious mutations (Kimura and Maruyama, 1966; Kondrashov, 1984; Charlesworth, 1990). Inferring from the present analysis, we suggest that synergistic epistasis can be an evolutionary outcome of sex and recombination, insofar as the lat-

ter processes promote the evolution of robustness, and synergistic epistasis emerges as a consequence. This hypothesis has some empirical support - there is a general trend towards weak synergistic epistasis between deleterious mutations among eukaryotes, but no trend in prokaryotes (de Visser et al., 1997; Elena and Lenski, 1997; de Visser and Hoekstra, 1998; Elena, 1999; Burch and Chao, 2004). This is beyond the scope of the present analysis, which has maximally one deleterious mutation in each individual, though it presents an interesting problem for the future.

Currently, no convincing empirical evidence has been published that demonstrates that genetic robustness exists as an adaptation. One reason for this is that, while it is possible to demonstrate that heritable variation is buffered in particular organisms, it is not easy to determine whether genetic robustness is the primary function, merely a side-effect of evolution for environmental robustness (Rutherford and Lindquist, 1998; Ancel and Fontana, 2000; Queitsch et al., 2002; Burch and Chao, 2004), or perhaps simply an emergent property of gene networks (Kacser and Burns, 1981; von Dassow et al., 2000; Edelman and Gally, 2001; Meir et al., 2002; Shen-Orr et al., 2002). A closely related problem is that the selection coefficient for a modifier of genetic robustness will be very weak, typically of the order of the mutation rate itself. However, the evolution of genetic robustness as a primary function may be plausible if there is migration between subpopulations in a heterogeneous environment (Mayr, 1963; Otto and Bourguet, 1999; Stearns, 2002). Migration rates can be much higher than mutation rates and therefore provide a stronger pressure for the buffering of (locally) maladapted alleles. Additionally, genetic robustness may evolve when selection fluctuates over time (Kawecki, 2000) and when selective sweeps take a population out of equilibrium (Mayo and Burger, 1997). It is with a view to extending the analysis to more complicated multilocus models that we have employed the methodology of Kirkpatrick et al. (2002),

which permits arbitrary complexity within a single notational framework.

5 Epistasis and the evolution of chromosomal linkage patterns

Abstract

In the dawn of the post-genomic era it has become clear from whole genome studies that previously held notions of random gene order across eukaryotic chromosomes are no longer relevant. Conspicuous cases of non-random gene order have resurrected population genetics theories as way of explanation. The theory focussed on here is that of antagonistic epistasis favouring the clustering together of loci within a chromosome. The probability of an inversion clustering together co-adapted alleles is derived and used to calculate expected waiting times for successful inversions to occur. Throughout, calculations are made using human data to illustrate the argument. It is shown that the expected waiting time will probably be too large for species with long generation times and small N_e to allow the clustering of loci with lineage-specific interactions. That is, it seems unlikely that selection will not fluctuate over the time-scales involved. Additionally, a numerical analysis suggests that higher order epistatic interactions will elicit weaker selection for chromosomal rearrangements than lower order interactions. Finally, we end with a discussion of the current empirical evidence for gene clustering from whole genome studies and for specific gene cluster cases.

5.1 The transition from prokaryotes to eukaryotes

In a series of elegant experiments, Jacob and Monod (1961) elucidated the co-ordinated regulation of protein synthesis in the *lac* operon of *Escherichia coli*. It soon became clear that the operon, a cluster of functionally related genes that are transcribed as a single polycistronic mRNA, is a common feature of prokaryotic genome organisation.

At the same time it also became clear that in eukaryotes the protein-coding genes are monocistronic and, therefore, that the operon as a unit of genome organisation could not be applied to eukaryotic genomes. What emerged from these early forays into the nature of genomes was a generally unquestioned acceptance of the null hypothesis of eukaryotic gene organisation that posits a random distribution of genes along linear chromosomes. Although counter-examples were known about at the time, such as the globin and Hox clusters, they were correctly identified as the products of tandem gene duplications. The question remained, why should there be such a radical difference between the organisation of prokaryotic and eukaryotic genomes?

A crucial difference between prokaryotes and eukaryotes arose as a result of the origin of mitosis in eukaryotic cells. In prokaryotic cells there is a single origin of replication for the DNA strand, but in eukaryotes chromosomes are home to multiple replicons that are attached to the nuclear matrix. In exponentially growing prokaryotes multiple, simultaneous rounds of replication may be initiated in a staggered fashion, and since the time for completion of DNA replication is longer than that for cell duplication (DNA replication is the rate-limiting step) this results in an increased gene dosage of genes near the single origin of replication by between a factor of four and eight relative to genes at the terminus of the DNA strand (Chandler and Pritchard, 1975). The effect has selected for highly expressed genes to be located near the origin of replication and lowly expressed genes to be near the terminus in fast-growing bacteria (Sharp et al., 1989; Couturier and Rocha, 2006). It has been suggested that operons and polycistronic mRNA enable an equality in the protein synthesis of co-ordinated, functionally related genes by clustering the genes together to counteract a potential inequality generated by the gene dosage effect (Cavalier-Smith, 1981, 1993). This inequality is not an issue for eukaryotic cells where there are multiple replicons, a relatively short S-phase and

no gene dosage effect. The consequently diminished selection to preserve operons may have allowed operon break-up through the accumulation of neutral, degradative mutations resulting in a more random distribution of functionally related genes in eukaryotes.

There is another major difference between prokaryotes and eukaryotes that may have influenced the break-up of operons. For co-ordinated translation rates of operon genes to enable equivalent enzyme levels in a particular metabolic pathway, prokaryotic ribosomes use internal polycistron sites to initiate translation. In eukaryotes, however, where mRNA is capped and polyadenylated, ribosomes use particular cap structures to initiate translation. By translating monocistronic mRNAs with a common cap structure equivalent levels of enzymes for a metabolic pathway can be achieved without the need for operons in eukaryotes (Cavalier-Smith, 1993).

Capped and polyadenylated mRNA is also much more stable than its uncapped counterpart and reduces variability in mRNA levels in a cell. This reduction in variation will also act to reduce the need for polycistronic mRNA to ensure similar levels of functionally related enzymes in particular pathways (Cavalier-Smith, 1993). Given these considerations one can ask whether there are any theoretical reasons to expect non-random gene order within eukaryotic chromosomes?

5.2 Fisher, Mather, Kimura et al.

Fisher (1930) was the first to propose a genetic scenario that might lead to closer linkage between loci in eukaryotes. He suggested that a symmetrical system of two epistatically interacting loci, such that two of the genotypes are equally fit and two equally unfit, would benefit any mechanism reducing the rate of recombination between the loci (Table 5.1a). The key process here is the effect of recombination reducing the positive linkage disequilibrium (LD) built up by selection and generating a positive recombina-

Table 5.1. Fitnesses for the 9 genotypes in the two locus epistatic model for the evolution of closer linkage between loci for (a) the Fisher (1930) model and (b) the Kimura (1956) model (where t > s).

(a)

	AA	Aa	aa
BB	1+s	1	1-s
Bb	1	1	1
bb	1-s	1	1+s

(b)

	AA	Aa	aa
BB	1+s	1+t	1-s
Bb	1	1+t	1
bb	1-s	1+t	1+s

tion load (RL). The form of epistasis acting here is positive or antagonistic epistasis.

Haldane (1931) was quick to point out that this scenario is not sufficient for the maintenance of a stable polymorphism at both loci, without which there would be no selection to reduce the rate of recombination between the loci. Kimura (1956) extended the model, based upon a verbal model suggested by Sheppard, to determine under what circumstances stable polymorphisms could be maintained and, therefore, select for closer linkage. In his model one of the loci is maintained in a stable polymorphism by overdominance, and together with antagonistic epistasis this, under certain conditions, can enable a stable polymorphism to be established at the second locus (Table 5.1b). The importance of overdominance for maintaining stable polymorphisms was reiterated by Kojima (1959a,b).

Mather (1943) introduced a verbal model of a quantitative trait that may benefit from a reduction in recombination between the loci contributing to its phenotype. Any polygenic trait subject to stabilising selection will exhibit epistasis between its loci on the fitness scale (e.g. Lewontin, 1964b). This epistasis will generate negative LD between the loci. Recombination between the loci will act to break-up this LD, increasing the additive genetic variance in fitness and resulting in a RL, and so by reducing the rate of

recombination the mean fitness of the population will increase (or the fitness of a modifier of recombination). Mather (1943) proposed that intra-chromosomal adjustments, such as inversions, could act to optimise the rates of recombination between different loci depending on whether they are involved in directional or stabilising selection, the forces of which will act in opposite directions with regard to the optimal recombination rates. The form of epistasis acting here is negative or synergistic epistasis although the argument is couched in the same terms: selection acts to put beneficial combinations of alleles together and recombination breaks them apart.

This same basic argument has resurfaced in a number of theoretical works and all of them have concluded that, without any other forces acting (such as directional selection in combination with negative epistasis or stabilising selection with a fluctuating optimum), selection will favour a reduction in the rates of recombination between the loci involved (Lewontin and Kojima, 1960; Lewontin, 1964a,b; Turner, 1967; Nei, 1967, 1968; Karlin and Feldman, 1969; Dobzhansky, 1970; Lewontin, 1971; Maynard Smith, 1978; Feldman et al., 1980; Zhivotovsky et al., 1994; Barton, 1995; Pepper, 2003). Somewhat similar considerations even motivated Franklin and Lewontin (1970) to propose that the chromosome be considered the unit of selection in favour of the gene.

A key component of these models is the maintenance of stable polymorphisms at the loci involved. While there is abundant additive genetic variance for the majority of quantitative traits in natural populations, this fact alone does not guarantee that recombination reducing modifiers will enhance the linkage (either physically or by reducing the crossover rates between specific loci) of functionally related genes. Stable polymorphisms must be maintained in a stable state for as long as it takes to modify recombination rates and for as long as is required to fix the modifications. Even after this point degradative mutations could disrupt the changes if the mode of selection acting on the

loci were to change.

Given that no population or species is ever evolutionarily static this may seem like a redundant point, but what matters is whether there is sufficient time for changes to occur and be fixed within species before conditions change and degradative mutations interfere. Once conditions have changed one might expect degradation to be rapid since there will be no selective sieve to prevent the neutral modifications from accumulating. Hence, when we examine the genomes of extant species will sufficient evolutionary time have elapsed for us to find clusters of transiently co-adapted alleles? To put it another way, can modifications of the genome occur rapidly enough to keep up with evolution?

5.3 Evidence of clustering from whole genome studies

In this section we focus on the empirical evidence for non-random gene order within chromosomes (i.e. not comparisons of patterns between chromosomes). Early studies of whole genomes found striking evidence for the clustering of co-expressed genes. For example, yeast mitotic cell-cycle genes tend to be clustered together into groups that are expressed in the same phase of the cell cycle (Cho et al., 1998). In *Caenorhabditis elegans* a great deal of clustering of co-expressed genes is simply due to the unusual presence of operons in this species (Blumenthal et al., 2002), although even when this is controlled for there appears to be clusters of co-expressed genes more than expected by chance (Roy et al., 2002; Lercher et al., 2003). There is also evidence for the clustering of co-expressed genes in *Drosophila melanogaster* (Boutanaev et al., 2002), in *Arabidopsis thaliana* (Williams and Bowles, 2004), in rodents (Ko et al., 1998), and in humans (Bortoluzzi et al., 1998; Gabrielsson et al., 2000; Dempsey et al., 2001; Yang et al., 2002).

Lee and Sonnhammer (2003) found evidence for the clustering of genes involved in the same metabolic pathways across a range of species from yeast, *C. elegans*, *A. thaliana* and *Drosophila* to humans. The extent to which this is true varies considerably between the species with yeast boasting a remarkable 98% of metabolic pathways with significant clustering between genes. Further to this, in yeast there is also significant clustering between genes whose protein products interact in stable protein complexes (Teichmann and Veitia, 2004). This is suggestive of epistatic interactions between loci favouring gene clustering although it cannot be ruled out that co-expression and co-regulation at a transcriptional level is the selective force favouring their clustering.

An important level of organisation for chromosomes is that of the chromatin. Heterochromatin tends to be silent transcriptionally, as indicated by the failure to express transgenes that integrate into heterochromatin (Milot et al., 1996). Chromatin must be unwound for gene expression to occur, a process which is energetically laborious and involves modification of the histone proteins (Eberharter and Becker, 2002). There is growing evidence that chromosome blocks can be switched on via chromatin modification to allow the co-expression of genes involved in functionally-related pathways, such as gluconeogenesis in yeast (Robyr et al., 2002). The inactivation of blocks of tissue-specific genes has also been implicated as playing a crucial role in cell differentiation in humans (Lunyak et al., 2002). Interestingly, it has been noted that in mammals chromosomes with a high density of genes tend to be located more centrally in the nucleus, which may suggest that there is a position effect of chromosomal gene expression within the nucleus (Cremer and Cremer, 2001; Tanabe et al., 2002).

Lercher et al. (2002) showed that there is a striking tendency for housekeeping genes to be clustered together in the human genome. Housekeeping genes are ubiquitously and constitutively expressed, and their gene products are required for the maintenance of the

cell. Additionally, Caron et al. (2001) found that in humans highly expressed genes are clustered together into chromosomal domains. Both of these lines of evidence point to the importance of transcriptionally active chromosomal domains for the expression of highly or constitutively expressed genes. This is suggestive of a general explanation for patterns of gene clustering as housekeeping genes are highly conserved between species and so will have had vast periods of time to wait for clustering chromosomal rearrangements to occur.

Interestingly, Pal and Hurst (2003) found that essential genes tend to be clustered together in regions of low recombination in both yeast and *C. elegans*, and the finding was corroborated in a separate study of *C. elegans* (Kamath et al., 2003). Nei (2003) was quick to suggest that essential genes are likely to exhibit antagonistic epistasis between their mutations, and therefore that their clustering is evidence in favour of the Fisher (1930) hypothesis. Since essential genes by definition have highly deleterious mutations it seems unlikely that they will harbour stable polymorphisms in natural populations, making the Fisher (1930) scenario redundant as an explanation. An alternative explanation is that genes with reduced allelic diversity, as essential genes would be expected to have, will not be affected by recombination and so can be optimally (in energetic terms) placed into regions of low recombination (Pal and Hurst, 2003; Gessler and Xu, 1999). Essential genes will also be highly conserved between species and so will have had sufficient time to accumulate chromosomal rearrangements.

5.4 Specific examples of gene clustering

Now we turn our attention to specific examples of gene clusters. The Hox gene clusters contain homeobox transcription factor genes that help to pattern the body plan of animals, fungi and plants during development. The loci are arranged in a colinear fashion

with regard to gene expression and chromosomal positioning and the lack of inversions within the clusters suggests that purifying selection is acting to maintain the gene order within the cluster (Zhang and Nei, 1996). The clusters have almost certainly arisen from both gene and whole cluster duplications (Ruddle et al., 1994), and so it seems unlikely that Hox clusters are examples of epistatic gene clustering.

The vertebrate MHC is a large, gene-dense cluster of highly polymorphic genes that encode cell-surface glycoproteins that present peptides to T-cells. There are suggestions that the polymorphisms are maintained by overdominant selection (Hughes and Yeager, 1998), although the evidence is not entirely clear. It is clear, however, that MHC loci are nearly ubiquitously expressed as are other unrelated genes mapped to the MHC region such as t-RNA synthetase, TAP transporter genes and histone H1.D and H2.A genes (Yeager and Hughes, 1999). Within the region there are also a number of large, multi-exonic genes known to undergo alternative splicing. Such observations are in keeping with the economy of expression explanation for gene clustering. Additionally, it has been observed that there is a certain amount of gene order conservation between species in the MHC region, suggesting that the gene order may be of functional significance (Yuhki et al., 2003).

The vertebrate Surfeit locus is a very tight cluster of six sequence-unrelated house-keeping genes that are ubiquitously expressed (Duhig et al., 1998). Additionally, within the locus there are overlapping genes (Williams and Fried, 1986), a bidirectional transcriptional promoter (Lennard and Fried, 1991), and one of the genes is alternatively spliced (Garson et al., 1995). All of these observations are indicative of co-regulation, slow transcription and high levels of expression. There are also a number of ubiquitously expressed genes mapped to the region, including the ribosomal protein L12 and the large, multiexonic COL5A1 (Yeager and Hughes, 1999). Therefore, it would seem

that the surfeit locus is a prime example of expression-associated clustering. Although it is interesting to note that, while the surfeit locus is exceedingly tightly linked in vertebrates, it is entirely absent in invertebrates with the gene homologs being widely dispersed through both the *Drosophila* and *C. elegans* genomes (Armes and Fried, 1996).

More convincing examples of epistatic gene clustering are for so-called 'supergenes' that have low recombination rates. The t-complex segregation dostorter in mice is a good example (Lyttle, 1991). The t-complex carries toxin-producing and anti-toxin genes, which have multiple alleles in natural populations. When the correct combination of alleles are not present in sperm the toxin kills the sperm and so there is strong antagonistic epistasis operating. The loci are also located near to centromeres and have inversions which cause a low rate of recombination. The Sd locus in *Drosophila* is also a good example of a segregation distorter supergene with a low recombination rate (Lyttle, 1991).

The mating type locus in *Chlamydomonas* fits into the category of a supergene with multiple inversions and a low rate of recombination (Ferris et al., 2002). The locus is maintained in a permanent state of polymorphism and it has been speculated that the system has evolved in step with alleles that enforce uniparental inheritance of chloroplasts and mitochondria by destroying the chloroplasts or mitochondria of the other mating type (Hurst et al., 2004). Given that uniparental inheritance is selectively beneficial (through avoiding organelle competition) this would lead to the formation of a mating type-organelle inheritance association, and eventual reduction in recombination between the two. The sex chromosomes of animal, fungal and plant species are more developed versions of this phenomenon, and are perhaps some of the clearest examples of epistatic clustering (persistent, ancient polymorphisms).

The self-incompatibility locus of the mustard family (Brassicaceae) is a region of

highly polymorphic genes involved in the prevention of self-fertilisation in a number of plant species (Nasrallah and Nasrallah, 1993). The polymorphisms are maintained by balancing selection brought about by the frequency-dependence of the system. There is growing evidence that recombination in the region is suppressed by genomic rearrangements (Boyes et al., 1997; Awadalla and Charlesworth, 1999; Charlesworth et al., 2003). Thus, this system may be subject to epistatic selection for linkage, and the ancient origins of the S locus fit well with this idea.

5.5 The probability of inversions clustering co-adapted alleles

Henceforth, we restrict our attention to intra-chromosomal inversions that occur on linear chromosomes. We are interested in the probability that inversions physically cluster together previously unclustered loci on a single chromosome in a single meiosis. We make four important assumptions, which we shall return to again when pertinent to the argument. Firstly, both paracentric and pericentric inversions are considered equally possible. Secondly, it is assumed that there is no interference between chromosomal break-points. Thirdly, the pre-inverted loci are assumed to not be within x bp (base pairs) of the ends of the chromosome where x bp is the maximum post-inversion distance between adjacent pairs of loci (the clustered distance). Finally, we assume that the minimum number of inversions required to bring the loci together have occurred in the meiosis.

This final assumption is made simply for completeness so that the general case of n loci being clustered can be calculated. However, since inversions are very rare events what we are interested in is the probability that a single inversion has occurred and clustered together two loci. This probability can then be used to calculate a waiting time before such an inversion will occur in a population.

Table 5.2. The minimum number of inversions required to cluster together different numbers of loci on a linear chromosome (# Inv.) together with the minimum number of chromosome breaks (# C/some Breaks), the number of possible successful inversions per locus in the pre-inverted state (# Pre-Inv./loc) and the number of ordered ways of picking inversions (permutations without repetition) per locus (# Perm. Inv./loc).

# loci	# Inv.	# C/some Breaks	# Pre-Inv./loc	# Perm. Inv./loc
2	1	2	1	1
3	2	4	4	4
4	3	6	9	36
5	4	8	16	576
...	...	...	...	...
n	$n-1$	$2(n-1)$	$(n-1)^2$	$\prod_{i=1}^{n-1}(n-i)^2$

Table 5.2 shows that the minimum number of inversions required to bring n loci together is $n-1$ since all of the loci are assumed to be unclustered prior to the meiosis. This in effect means that the loci cluster around one non-inverted locus to make a minimum number of inversions. Now we consider a linear chromosome L bp long containing n unclustered loci. Break-points occur between bp and a single inversion involves two separate break-points, inversion of the segment of DNA and reintegration in the inverted orientation. For a single inversion, the probability that it successfully clusters together the two loci of interest is taken to be $\frac{S}{T}$, where S is the total number of successful inversions possible and T is the total number of possible inversions across the entire length of the chromosome. Allowing edge inversions, but subtracting the trivial inversion of the whole chromosome, the total number of inversions is equal to the unordered number of combinations (without repetition) of two break-points across the

L bp of the chromosome,

$$T = \binom{L+1}{2} - 1 = \frac{1}{2}(L+2)(L-1). \tag{1}$$

The number of successful inversions depends on how close together the loci have to be brought to be considered clustered. For the time being we shall consider the maximum distance between pairs of loci to be x bp and any inversion that brings the loci within that distance is successful. To simplify things, we assume that the unclustered loci are not within x bp of the ends of the chromosome so that we can neglect their position on the chromosome from the calculation.

For a single inversion, successful break-points can only occur on the same side of both loci to ensure the inverted segment brings the loci together. If we imagine x bp on the right-hand sides of two unclustered loci (the inversion will bring the right-hand locus within x bp of the position of the left-hand locus) then the first break-point immediately adjacent to the left-hand locus can pair with $x+1$ break-points from the right-hand locus to bring the loci within x bp of each other (Figure 5.1). The second break-point, however, can only pair with x break-points from the right-hand locus to bring the loci within x bp of one another as it has a bp between itself and the left-hand locus (Figure 5.1). The third break-point can only pair with $x-1$ break-points and so on. In general we have

$$S = \sum_{i=0}^{x} (x+1-i) = \frac{1}{2}(x+2)(x+1). \tag{2}$$

The order in which different inversions cluster loci around one of the n loci matters and so we must take into account the number of permutations (without repetition) of inversions that cluster the loci together. Therefore, we are interested in the number of ordered ways of picking $n-1$ inversions from $n-1$ loci. In the pre-inverted state each locus can initially make $n-1$ successful inversions (an inversion that results in a cluster of two loci). Focussing on loci clustering around one non-inverted locus (minimum number of inversions) then there are $(n-1)^2$ possible successful inversions to

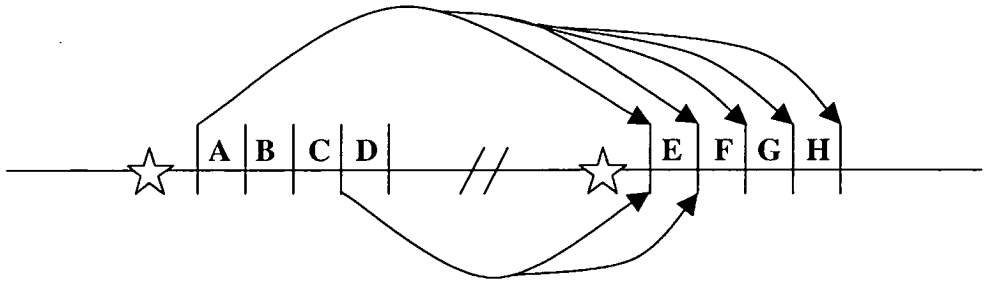


Figure 5.1. Schematic representation of the pairing of different break-points between bp (capital letters) to bring two loci (stars) within 4 bp of one another. In this example, the right-hand locus inverts towards the position of the left-hand locus. The first break-point adjacent to the left-hand locus can pair with 5 break-points $(x+1)$ from the right-hand locus, but the fourth break-point can only pair with 2 break-points $(x-2)$ from the right-hand locus to ensure that the loci are within 4 bp of one another.

begin with (Table 5.2), multiplied by n for clustering around each locus. After the first inversion occurs the subsequent number of possible successful inversions is reduced by the $n-1$ inversions that the inverted locus was capable of (each locus inverts only once - minimum number of inversions) plus the $n-2$ inversions that could be made towards the inverted locus, giving us

$$(n-1)^2-((n-1)+(n-2))=(n-2)^2. \tag{3}$$

After the second locus inverts the subsequent number of possible successful inversions is reduced by the $n-2$ inversions that were associated with this locus plus the $n-3$ inversions that could be made to this locus by the other loci giving us $(n-3)^2$ remaining successful inversions. In general we have

$$\wp(n-1,n-1)=\prod_{i=1}^{n-1}(n-i)^2 \tag{4}$$

where $\wp()$ stands for permutations. As the loci can cluster around any of the n loci

this gives us $n\prod_{i=1}^{n-1}(n-i)^2$ numbers of ways of clustering the loci together. Now we can express the conditional probability of $n-1$ inversions clustering n unclustered loci together in a single meiosis given that $n-1$ inversions have occurred as

$$P(C_n|I_{n-1}) = n\left(\prod_{i=1}^{n-1}(n-i)^2\right)\left(\frac{(x+1)(x+2)}{(L+2)(L-1)}\right)^{n-1}. \quad (5)$$

Therefore, assuming that the unclustered loci are not within x bp of the ends of the chromosome and that there is no interference between break-points their position and distance from each other does not influence the probability that they will become clustered.

It may prove useful to express the probability in terms of recombination fractions instead of bp. To this end we can make use of Haldane's mapping function (Haldane, 1919). The function expresses the recombination fraction as a function of the distance in Morgans between two markers assuming that there is no interference between crossovers, and derived assuming that the number of crossovers are distributed as a Poisson random variable,

$$r = \frac{1}{2}(1-e^{-2z}) \quad (6)$$

and rearranging for z

$$z = -\frac{1}{2}\ln(1-2r). \quad (7)$$

To convert the above estimate of Morgans (z) into bp we have to multiply by c, the conversion factor between Morgans and bp, which is species-dependent. Substituting this into (5) we have

$$P(C_n|I_{n-1}) = n\left(\prod_{i=1}^{n-1}(n-i)^2\right)\left(\frac{(c\ln(1-2b)-2)(c\ln(1-2b)-4)}{(c\ln(1-2r)+2)(c\ln(1-2r)-4)}\right)^{n-1} \quad (8)$$

where r is the recombinational length of the chromosome (i.e. the recombination rate between markers at the opposite ends of the chromosome) and b is the maximum recombination fraction between pairs of the clustered loci.

Table 5.3. The probabilities of clustering different numbers of loci in a single meiosis for different values of b, the maximum clustered recombination rate between adjacent pairs of loci. Values are based on the human conversion factor for Morgans and bp (1 cM = 10^6 bp, 1 cM = 1% recombination rate between two markers). The recombinational length of the chromosome, r, is taken to be 0.45.

	b (human bp)			
No. loci	0.02 (2 Mb)	0.01 (1 Mb)	0.005 (500 kb)	0.0025 (250 kb)
2	0.00063	0.00015	0.000038	9.5×10^{-6}
3	1.2×10^{-6}	7.1×10^{-8}	4.4×10^{-9}	2.7×10^{-10}
4	4.5×10^{-9}	6.6×10^{-11}	9.9×10^{-13}	1.5×10^{-14}
5	2.8×10^{-11}	1.0×10^{-13}	3.8×10^{-16}	1.5×10^{-18}
6	2.7×10^{-13}	2.3×10^{-16}	2.2×10^{-19}	2.1×10^{-22}

It can be seen that the probability of clustering drops off exponentially according to $n-1$ as the number of loci increases. This fact is illustrated in Table 5.3 where the probabilities of clustering different numbers of loci for different values of b are shown. The values are based on the human conversion factor for Morgans, c, which is 10^8 bp (100 Mb) per Morgan. Thus, even with an unusually high inversion rate, it is very unlikely that more than two loci will become clustered together in a single meiosis as these figures are the conditional probabilities and so do not take into account the probability of multiple inversions occurring in single meioses.

We can suppose that in the limit of a large number of meioses, each with a single inversion (and single chromosome), the number of successful inversions will be distributed as a Poisson random variable with equal mean and variance. Figure 5.2 illustrates this point, showing a binomial distribution of the number of successful inversions

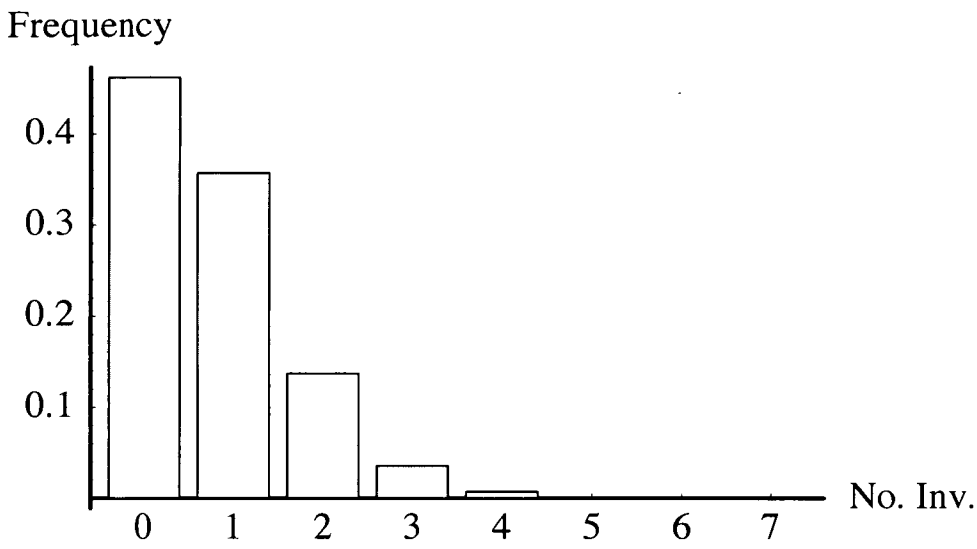


Figure 5.2. The distribution of the number of successful inversions based upon random samples from a binomial distribution ($Bin(10^4, \frac{S}{T})$). Parameter values are, $b = 0.01$ and $r = 0.45$. The mean and variance are both 0.772.

with 10^4 trials and a probability of $\frac{S}{T}$ that there is a success on each trial ($Bin(10^4, \frac{S}{T})$). The mean is 0.772 and the variance is 0.772.

To get a feel for the distances covered by clusters of functionally related genes in the human genome it may be useful to cite some familiar examples. The MHC (Major Histocompatibility Complex) spans a region of roughly 3.6 Mb on chromosome 6 and contains around 140 genes (MHC-Consortium, 2001), the four Hox clusters are around 100 kb in length each with between 5 and 10 homeobox genes (Acampora et al., 1989), and the Surfeit locus is just 60 kb in length and contains 6 sequence-unrelated genes (Duhig et al., 1998), and is the tightest known mammalian gene cluster (Huxley and Fried, 1990; Garson et al., 1995).

It seems likely that when more than two loci are to be clustered together via inver-

sions they will have to occur in separate meioses. The problem then faced by the population is how to get separately occurred inversions onto the same chromosome given that inversions suppress recombination in heterokaryotypes. This is akin to the Fisher-Muller argument for the benefit of sexual reproduction in large, finite populations that are adapting to an environment (Fisher, 1930; Muller, 1932; Crow and Kimura, 1965). It may be that inverted clusters will have to get to high frequency before waiting for the next inversion to occur and so on. However, it is still of interest to ask what the expected waiting time is for the appearance of an inversion that clusters together two previously unclustered loci.

5.6 The waiting time for a successful inversion

Thus far we have only dealt with the conditional probability of successful inversions occurring, but now we are interested in the probability that a successful inversion clusters two previously unclustered loci and the probability that a single inversion has also occured (the spontaneous inversion rate per chromosome per generation, μ_I), which is simply

$$P(C_2 \cap I_1) = P(C_2|I_1)\mu_I. \tag{9}$$

We will focus on a diploid population with two biallelic, epistatically interacting (antagonistic a la Fisher (1930)), unclustered loci (located on the same chromosome) that await the occurrence of an inversion that clusters them together. We shall assume that the frequencies of the alleles are at equilibrium in the population. In a single generation the probability that a successful inversion does not occur is

$$P(C_X) = 1 - 2N_e\hat{p}_\varepsilon P(C_2|I_1)\mu_I \tag{10}$$

where $\hat{p}_\varepsilon$ is the equilibrium frequency of both beneficial allelic chromosomal combinations in the population and N_e is the effective population size. Therefore, the probability

that a successful inversion does not occur in t generations is $P(C_X)^t$. To find the expected waiting time until a successful inversion occurs we can integrate this quantity with respect to t, between zero and infinity, giving us

$$E(t_I) = \int_0^\infty P(C_X)^t \, dt = -\frac{1}{\ln(1 - 2N_e \hat{p}_\varepsilon P(C_2|I_1)\mu_I)}. \tag{11}$$

The expected time will be sensitive to the parameters that are most likely to vary considerably between species, N_e and μ_I. Before we can explore numerical examples it is necessary to substitute in an estimate for the equilibrium frequency of the beneficial allelic chromosomal combination, $\hat{p}_\varepsilon$. We will use the equilibrium frequency derived by Kimura (1956) for overdominance in one locus and antagonistic epistasis between the loci

$$\hat{p}_\varepsilon = \left(\frac{1}{2} - \frac{wr_b}{s} + \sqrt{\frac{1}{4} + \frac{wr_b{}^2}{s}} \right) \tag{12}$$

where r_b is the recombination rate between the loci, s is the strength of epistatic selection and w is the reproductive rate of the double heterozygote (Kimura, 1956).

Figure 5.3 shows the expected waiting time as a function of the recombination rate and the strength of epistatic selection for parameter estimates based upon human data. There have been numerous estimates of N_e for humans using different methods and based on different periods in human history. Hayes et al. (2003) used a multilocus measure of LD, the chromosome segment homozygosity (CSH) method, on a human haplotype dataset to infer the N_e in the past. They estimated N_e to be around 5 x 10^3 at about 2000 generations ago, and 1.5 x 10^4 at about 180 generations ago, indicating recent exponential growth.

Estimates of ancient N_e have also been made based on current genetic variation. The so-called trichotomy method uses orthologous sequences from three closely related species with known phylogenies (Nei, 1987; Wu, 1991; Hudson, 1992). Based on this method, the human-chimpanzee ancestral N_e has been estimated to be in the range of

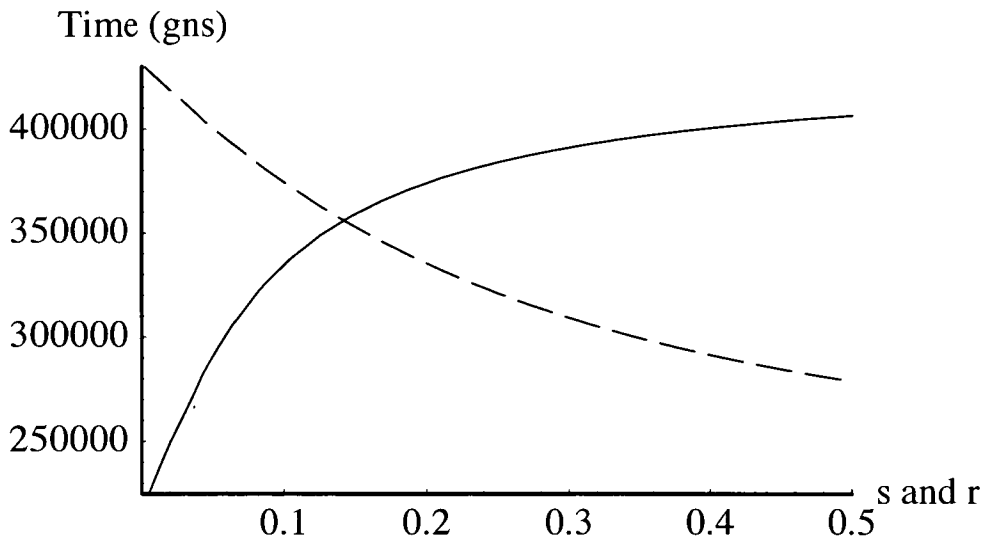


Figure 5.3. The expected waiting time in generations for an inversion to bring together two unclustered loci on the same chromosome as a function of s (dashed line), the strength of epistatic selection, and r_b (unbroken line), the rate of recombination between the loci. The graphs are plotted from (11) using parameter estimates based upon human data. $N_e = 5 \times 10^4$, $b = 0.01$ (1 Mb), $\mu_I = 9 \times 10^{-7}$, $s = 0.1$, $r_b = 0.2$, $r = 0.49$.

5×10^4 to 1.5×10^5 (Ruvolo, 1997; Chen and Li, 2001). Yang (2002) re-analysed the dataset of Chen and Li (2001) using a likelihood and Bayesian method to correct for variation in mutation rates among loci and for multiple substitutions at the same site, and found N_e to be in the range of 1.2×10^4 to 2.1×10^4. Therefore, the range for early human populations is between about 5×10^3 and 1.5×10^5 (depending on the estimation method, (Wang, 2005)), motivating the use of 5×10^4 in Figure 5.3 (assuming constancy through time).

Comparison of the chromosomes of the great apes has revealed that in the human lineage at least two large pericentric inversions have been fixed in the last several million

years, on chromosomes 4 and 17 (de Grouchy et al., 1973; Uzzell and Pilbeam, 1971). Radiation studies in human lymphocytes, both *in vitro* and *in vivo*, have suggested that the rate of occurrence of inversions is three to four times lower than that of reciprocal translocations (Seabright, 1973; Roman and Robrow, 1973). Given a human reciprocal translocation rate of 1.3 x 10^{-4} per gamete per generation (Jacobs, 1977), it has been inferred that the human spontaneous inversion rate is roughly 4 x 10^{-5} per gamete per generation (Lande, 1979). The rate per chromosome (46 for humans) is then roughly 9 x 10^{-7} (Figure 5.3).

Several things can be noticed from Figure 5.3. Firstly, the number of generations are in the 100s of thousands. With a generation time of 20 years these figures would put the expected waiting time at somewhere between 5 and 8 million years (and this does not take into account the probability of fixation). With the most recent estimate for the human-chimpanzee divergence at between 5 and 7 million years ago (Kumar et al., 2005), this is clearly too long to allow for the clustering of loci with epistatic interactions specific to the human lineage.

Secondly, the waiting time increases as the rate of recombination between the loci increases since an increased recombination rate will decrease the frequency of the beneficial allelic combination in the population. We might expect this to work in the opposite direction with respect to the selection coefficient acting on the newly inverted chromosome as the strength of selection will be proportional to the rate of recombination between the uninverted loci (Charlesworth and Charlesworth, 1973; Charlesworth, 1974). Hence, inversions capturing loci further apart will have larger selection coefficients but will also on average take longer before they appear in the population and vice versa for loci closer together.

Thirdly, the waiting time decreases as the strength of epistatic selection between the

Table 5.4. The expected waiting time for the occurrence of an inversion that clusters together two loci in generations and years for both humans and yeast (*S. cerevisiae*). μ_G - the spontaneous inversion rate per gamete per generation. Parameter values used: human gen time = 20 years, yeast gen time = 3 hours, N_e = 5 x 10^4, μ_G = 4 x 10^{-5}, s = 0.1, r_b = 0.2, r = 0.49, b = 0.01 (humans = 1 Mb, yeast = 3 kb), c (humans) = 10^8 bp, c (yeast) = 3 x 10^5 bp.

	Varying N_e			
	Humans		Yeast	
N_e	Generations	Years	Generations	Years
5 x 10^3	3.7 x 10^6	7.5 x 10^7	2.6 x 10^6	890
10^4	1.8 x 10^6	3.7 x 10^7	1.3 x 10^6	445
5 x 10^4	3.7 x 10^5	7.5 x 10^6	2.6 x 10^5	89
10^5	1.9 x 10^5	3.7 x 10^6	1.3 x 10^5	45
5 x 10^5	37410	7.5 x 10^5	25999	8.9
10^6	18704	3.7 x 10^5	12999	4.5
	Varying μ_I			
	Humans		Yeast	
μ_G	Generations	Years	Generations	Years
10^{-5}	1.5 x 10^6	3.0 x 10^7	1.0 x 19^6	356
5 x 10^{-5}	2.9 x 10^5	5.9 x 10^6	2.1 x 10^5	71
10^{-4}	1.5 x 10^5	2.9 x 10^6	1.0 x 10^5	36
5 x 10^{-4}	29928	5.9 x 10^5	20798	7.1
10^{-3}	14963	2.9 x 10^5	10399	3.6
5 x 10^{-3}	2992	59847	2079	0.71

loci increases since stronger epistatic selection will keep the beneficial alleles together at a higher frequency. It might be expected that this will work in the same direction with respect to the selection coefficient of the inversion, but this is not at all clear. Charlesworth (1974) found that the strength of selection acting on an inversion that simply suppresses recombination between beneficial combinations of alleles was higher for the loci with a higher epistatic selection coefficient only when recombination had abolished the LD between the loci with a lower epistatic selection coefficient. As epistatic selection will increase associations between beneficial alleles this will act to decrease the beneficial effect of reducing recombination between them. The relationship does not appear to be simple.

Table 5.4 shows the expected waiting times for a range of N_e and μ_I values for the different conversion factors, generation times and chromosome numbers of humans and of the haploid (assuming the polymorphism is maintained by something other than overdominance, but keeping the Kimura (1956) frequency for convenience), unicellular fungus, *Saccharomyces cerevisiae*. From these figures it appears possible for yeast, with a short generation time (3 hours) and large N_e, to optimise the linkage patterns of its genome in response to the beneficial interactions there may be between different alleles.

The figures for yeast are most likely overestimates since natural populations will not always be growing exponentially. Although it is not easy to measure how stable certain modes of selection are for yeast through time, and so not possible to guess at how stable their polymorphisms are, their continued persistence for many millions of years is suggestive of a persistent niche. One could look at their rates of non-synonymous to synonymous substitutions to gauge how much selectively driven change they have been subject to in the recent past. However, it is clear that of the extant eukaryotic species

examined to date, yeast have the most conspicuous patterns of functionally related gene clustering (genes involved in the same metabolic pathways) (Lee and Sonnhammer, 2003; Hurst et al., 2004). Interestingly, clusters of metabolically related genes tend not to be well conserved between yeast species (Lee and Sonnhammer, 2003). This may be the signature of differential selection acting on different lineages to increase or decrease the benefit of clustering particular loci together. In support of this interpretation it has been noticed that there are highly variable rates of chromosomal rearrangements between yeast lineages, possibly indicating a role for selection in optimising the genome with respect to differing selective demands (Fischer et al., 2006).

Differential rates of chromosomal rearrangement evolution have also been demonstrated between different mammal species (Murphy et al., 2005). For example, humans and chickens have far slower rates of chromosomal evolution than rodents (Pevzner and Tesler, 2003; Bourque et al., 2005). Lande (1984) proposed that periods of intense karyotypic evolution will go hand in hand with periods of intense selection or adaptive radiation in species. When there is no strong directional selection the deleterious effects of inversions in heterozygotes (due to a reduction in fecundity) will dominate, but when there is strong selection acting inversions may capture beneficial combinations of alleles and prevent them from recombining due to the suppression of recombination in heterokaryotypes.

The foundations of this idea have a long history (Sturtevant and Mather, 1938; Nei et al., 1967; Ohta and Kojima, 1968; van Valen and Levins, 1968; Dobzhansky, 1970; Charlesworth and Charlesworth, 1973; Charlesworth, 1974; Schaeffer et al., 2003; Kirkpatrick and Barton, 2006). The idea suggests that there may be extensive inversion polymorphisms in natural populations, but the evidence is so far only limited to certain species (Dobzhansky, 1970; Schaeffer et al., 2003; Feder et al., 2003). The mechanism

certainly will not suffer from the long waiting times that clustering inversions will, as the probability of inversions simply encompassing co-adapted alleles will be much larger than those for clustering alleles together. It may be that the optimisation of genome linkage patterns with regard to epistatic interactions is only feasible for a subset of eukaryotes with short enough generation times and large enough N_e.

5.7 Summary

The potential role of epistatic selection for generating patterns of gene clustering within chromosomes via inversions has been discussed. The likelihood that selection will not fluctuate over the time periods required for a successful inversion to occur seems small for species with long generation times and relatively small N_e. This analysis has not included the selective benefit of the newly arisen inversions and hence their fixation probabilities as this will not influence our conclusions. The empirical evidence to date is supportive of epistatic clustering only for specific cases where polymorphisms are maintained for long periods of time and where recombination is suppressed by chromosomal rearrangements. Clear examples of this phenomenon are the sex chromosomes of animals, plants and fungi. More generally, it seems likely that highly conserved genes that are constitutively expressed will tend to be clustered together into transcriptionally active chromosomal domains across eukaryotic species.

6 Maternal effects and sexual selection

Abstract

Various genetic models of sexual selection have been developed to explain the evolution of conspicuous secondary sexual characters and female preferences for them in natural populations. The models can be roughly split into neutral (or maladaptive) models and adaptive models with respect to the existence of a genetic benefit for female choice. Here, we develop and analyse an adaptive model of mate choice based on the benefit of maternal effects. A condition-dependent trait is assumed to develop in males as a result of maternal effects and females can select mates based on this trait. The model is compared against previous models and displays similar qualitative and to some extent quantitative behaviour. We discuss examples of potentially important maternal effects in polygynous species and it is concluded that maternal effects can be an important determinant of mate choice in polygynous species. These results support the notion that mate choice is often adaptive at the level of the individual.

6.1 Introduction

The evolution of mate choice, and of the often conspicuous traits that are chosen, has been the subject of much debate ever since Darwin (1859) first proposed sexual selection (Andersson, 1994). At the centre of the debate is the question over the adaptiveness of mate choice in polygynous species where there is no direct benefit of being choosy for females. It has been argued that the evolution of secondary sexual characters that are neutral or even detrimental with respect to the bearer's fitness is non- or maladaptive at the level of the individual and the population with no benefit for the choosy female (O'Donald, 1980; Lande, 1980, 1981; Kirkpatrick, 1982; Heisler, 1984; Kirkpatrick,

1987). In contrast to this stance others have argued that there can be genetic benefits of mate choice for choosy females that have preferences for males with 'good genes' (Fisher, 1915; Williams, 1966; Trivers, 1972; Zahavi, 1975, 1977; Hamilton and Zuk, 1982; Trivers, 1985; Seger and Trivers, 1986; Andersson, 1986; Pomiankowski, 1987, 1988; Grafen, 1990a,b).

Fisher (1930) was the first to propose a genetic scenario that could explain the evolution of a conspicuous, sexually dimorphic trait together with a female preference for the trait. In a verbal model Fisher (1930) suggested that an initially beneficial male trait may happen to be preferred by some females and over time the alleles for female preference and for the male trait would become associated with one another as a result of the mating preference. The mating advantage of males with the trait allows both preference and trait alleles to increase in frequency. Over time both the preference and the trait evolve to such extremes that the sexually selected trait in males becomes detrimental with regard to natural selection. This has been termed Fisher's runaway process. The key process here is the build-up of linkage disequilibrium (LD) between the loci such that the preference allele and the male trait allele are more often found together than expected from their allele frequencies as a direct result of the mating bias.

Genetic models based on this idea found that it could work even when the male trait was detrimental in terms of natural selection (Lande, 1980, 1981; Kirkpatrick, 1982). Kirkpatrick (1982) analysed a two-locus, haploid, bi-allelic model where there was a preference locus and a male trait locus. In models where choosy females preferred males with a detrimental trait an equilibrium analysis showed that there was a neutral line of equilibria that represented the balance between the opposing forces of sexual and natural selection acting on the male trait. Subsequent models have found that 'good genes' versions of this basic model can also work although it is difficult to evaluate the

role of Fisher's runaway versus the benefit of 'good genes'. Grafen (1990a,b) separated the effects and found that genetic benefits could indeed select for a male trait and female preference without the need for Fisher's runaway process to be operating.

Seger and Trivers (1986) studied a model based on the Kirkpatrick (1982) model, but included both Fisher's runaway and good genes as features of the model. The trait that is subject to choice is expressed in both males and females, but due to genetic or developmental constraints cannot be optimised simultaneously in both sexes. The model showed that, given a trade-off between choosing fitter daughters or fitter sons, preference alleles that prefer traits in males that signify high female fitness exhibit an asymmetry in their line of equilibria such that they can invade more easily than alleles that prefer traits for good male genes. The reason for this is that the preference allele is only expressed in females and so if it becomes associated with the high female fitness allele then there will be more females choosing males that signify good alleles for their daughters and so the preference for fitter daughters will increase in frequency.

Maternal effects are most likely a ubiquitous phenomenon (Mousseau and Fox, 1998a), and in many cases can have profound effects on the quality of an individuals offspring (Mousseau and Dingle, 1991; Ginzburg and Taneyhill, 1994; Mousseau and Fox, 1998b; Bjorklund, 2006). Mothers can determine the size and quality of their eggs together with where and when they are laid, and often they have significant roles in providing for and protecting their developing young. Since the nutritional status and health of males can influence the development of their secondary sexual characteristics (e.g. Wilkinson and Taper, 1999; Moczek, 2002; Cotton et al., 2004), it can be seen how maternal effects and sexual selection are intimately related to one another.

Quantitative genetic models of the interaction between parental effects, both maternal and paternal, have been developed (Wolf et al., 1997, 1999). The models are relevant

to cases where there are direct benefits provided by a mate that correlate positively with a sexually selected trait. It was shown that both indicator traits and parental investment can increase above their natural selection optimums showing that parental effects can have important consequences for adaptive mate choice.

Here, we study a multi-locus model that is rooted in Fisher's runaway process, but that includes a preference for good maternal genes at the expense of male fitness. The male trait that is subject to choice is determined by the maternal genotype. That is to say, there are maternal effects that influence the expression of a detrimental secondary sexual character that females can base their mate choice upon. The purpose of setting the model up like this is to investigate whether a preference allele for choosing a male who indicates that he carries good maternal genes via an indicator trait can increase in frequency via its association with a good maternal allele in a fitter female. In this sense our model presents females with the same trade-off as in the Seger and Trivers (1986) model. Both two-locus and three-locus models are analysed.

The difference between our model and those of Wolf et al. (1997, 1999), aside from the genetic assumptions, is that we focus on a polygynous species where there are no direct benefits to mate choice and therefore the only parental effect that is relevant is the maternal effect. Thus, our model is more pertinent to previous models of polygynous species and may shed some light on the role of good genes for mate choice in such species.

Additionally, we analyse a second multi-locus model of mate choice based upon a mate preference that is determined by the phenotypic similarity between the bearer of the preference allele and the male that is the subject of choice. In this model, females with the preference allele choose males that have the same phenotype as they do, determined by the allelic state at a second locus. This model is partly motivated by a recent

study that showed that female wolf spiders choose males that are of a familiar phenotype determined by their subadult experience (Hebets, 2003).

6.2 Models and analyses

6.2.1 Two-locus model

We will focus on a haploid model of a polygynous species (with a 1:1 primary sex ratio) with bi-allelic, autosomal loci and free recombination between the loci (Kirkpatrick (1982) showed that the equilibirum LD is independent of the recombination rate as its effect is counter-balanced by natural selection). There are two loci, a maternal care locus with two alleles, the G allele that is the good mother allele and the g allele that is the bad mother allele, and a preference locus with two alleles, the P allele that preferentially chooses males with an indicator trait and the p allele that mates randomly.

The maternal effects are manifested in such a way that any male genotype that had a mother with the G allele will develop the indicator trait that is detrimental to its survival but that gains preferential matings from females with the P allele. Thus, all of the males are assumed to be competent to develop the indicator trait regardless of their own genotype. The trait could be a pleiotropic effect of other fixed loci that when expressed is sub-optimal with regard to natural selection or genetic drift could allow the fixation of the trait allele that is neutral in the absence of the good mother allele.

The fitnesses for this model are shown in Table 6.1. It can be seen that the fitness of an individual is entirely determined by the genotype of their mother. It can also be seen that by choosing a male with an indicator trait a female cannot guarantee that he is carrying the good mother allele, G, but only that he had a good mother. In this model we assume that there is no direct cost for a female being choosy, such as loss of fecundity or survivorship due to a longer time spent searching for a suitable mate.

Table 6.1. Fitnesses for the two-locus maternal effects sexual selection model. The G allele is the good mother allele and results in higher fitness for daughters, but a fitness of $1-t$ for sons because they are competent to develop a costly indicator trait. The g allele results in lower fitness for both sons and daughters.

	Mother's genotype	
	G	g
Males (GP, Gp, gP and gp)	1-t	1-s
Females (GP, Gp, gP and gp)	1	1-s

After natural selection has acted on all of the genotypes, sexual selection occurs via female mate choice. Following Kirkpatrick (1982) we utilise the fixed relative preference model of mate choice, which is the conditional probability that a male of a certain type is accepted given that acceptance has occured (Seger, 1985). Thus, there is a fixed probability that a male of a certain type will get a mating, and this is not altered by the frequency of the male type in the population. Therefore, given that a female has accepted a male, the probability that it is a male with an indicator trait is

$$P(T|accept) = \frac{at}{((1-t)+at)} \tag{1}$$

where $\frac{a}{1+a}$ is the fixed probability of choosing the T male (a is the strength of the preference), and t is the frequency of this male type in the population. Biologically, this is relevant to a scenario where females keep sampling males until they find the one they want. An alternative form of mate choice model is the best of N males model, which is more relevant to lek mating systems, and the frequency of a particular preferred male will alter the probability that he is chosen (mate choice is frequency-dependent on the male type frequencies). However, we will not pursue this model of mate choice.

In a maternal effects model it is not straight-forward to calculate the effects of selection and mate choice because in both cases the effects do not depend on the individual's genotype, but depend on their mother's genotype. As we are interested in deriving genotype frequency recurrence equations for the four haploid genotypes without having to track phenotype frequencies, we will make use of the conditional probabilities of a particular genotype having a G or g mother given their own particular genotype. For example, the probability of a GP genotype having been reared by a female carrying a G allele is

$$P(G_{mother}|GP) = \frac{P(G_{mother} \cap GP)}{P(GP)} = \frac{P(G_{mother} \cap GP)}{f(GP)} \tag{2}$$

where $f(GP)$ is the frequency of GP genotypes in the population. The probability, $P(G_{mother} \cap GP)$, can be calculated from the mating frequencies in the previous generation together with the segregation rules of meiotic recombination (free recombination). These conditional probabilities have to be applied to calculate the fitness of different genotypes. For example, the fitness of a male with a GP genotype is

$$W(GP_{male}) = P(G_{mother}|GP)(1-t) + P(g_{mother}|GP)(1-s) \tag{3}$$

and the fitness of a GP female is

$$W(GP_{female}) = P(G_{mother}|GP) + P(g_{mother}|GP)(1-s). \tag{4}$$

These probabilities will not remain fixed, but will increase or decrease as the LD between the loci changes due to the mating preferences. The mating frequencies will also have to be calculated using conditional probabilities. For example, the frequency of matings between GP females (at a post-selection frequency of x') and GP males (at a post-selection frequency of x''), based upon the probability of a GP male being chosen (equation 1), is

$$f(GP \times GP) = \frac{x'x''(aP(G_{mother}|GP) + (1 - P(g_{mother}|GP)))}{B} \tag{5}$$

Table 6.2. The mating frequencies for a two-locus maternal effects model where females carrying the P allele have a preference for males with an indicator trait that is a result of having a mother with the G allele. The P_{ij}s refer to the probabilities of receiving both preferential and random matings for males in the 'ij' mating combination.

		♂			
		$GP\ (x'')$	$Gp\ (y'')$	$gP\ (z'')$	$gp\ (u'')$
♀	$GP\ (x')$	$\frac{x'x''P_{xx}}{B}$	$\frac{x'y''P_{xy}}{B}$	$\frac{x'z''P_{xz}}{B}$	$\frac{x'u''P_{xu}}{B}$
	$Gp\ (y')$	$y'x''$	$y'y''$	$y'z''$	$y'u''$
	$gP\ (z')$	$\frac{z'x''P_{zx}}{B}$	$\frac{z'y''P_{zy}}{B}$	$\frac{z'z''P_{zz}}{B}$	$\frac{z'u''P_{zu}}{B}$
	$gp\ (u')$	$u'x''$	$u'y''$	$u'z''$	$u'u''$

where

$$\begin{aligned} B =& x''(aP(G_{mother}|GP) + (1 - P(G_{mother}|GP)) + y''(aP(G_{mother}|Gp) + \\ & (1 - P(G_{mother}|Gp)) + z''(aP(G_{mother}|gP) + (1 - P(G_{mother}|gP)) + \\ & u''(aP(G_{mother}|gp) + (1 - P(G_{mother}|gp)) \end{aligned} \tag{6}$$

where y'', z'' and u'' are the post-selection frequencies of Gp, gP and gp males respectively. This demonstrates that there are non-zero probabilities of preferential matings for each male genotype. However, as LD between the loci builds up these probabilities will change with some increasing and some decreasing. Thus, the mating frequencies might be expected to start off weak, but get stronger as LD builds up.

Table 6.2 shows the mating frequencies for the sixteen combinations of male and female genotypes. The polygynous structure of the model is apparent now as the females mate at a frequency equal to their frequency in the population, but the males can mate above or below their own frequency (though the total sums to one).

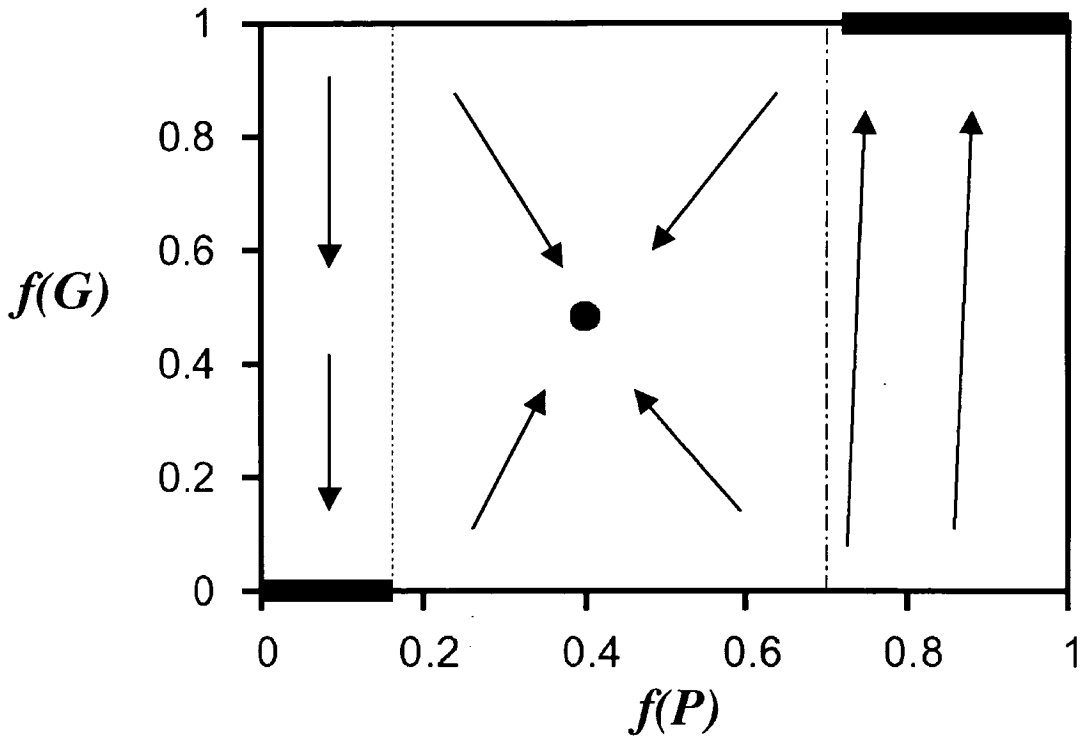


Figure 6.1. Schematic of the equilibrium results from a two-locus maternal effects model with parameter values that reflect those in the Kirkpatrick (1982) model. The heavy lines indicate the neutral lines of equilibria. Arrows indicate the direction and end points that allele frequency combinations from those vicinities reach. $f(P)$ - frequency of the female preference allele, $f(G)$ - frequency of the good mother allele (actually detrimental with these parameter values). $a = 4$, $s = 0$, $t = 0.4$.

Genotype frequency recurrence equations can be derived from Table 6.2 together with expressions for the change in frequency due to selection, which will be the relative fitness (a genotypes fitness divided by the mean fitness) of a genotype multiplied by its frequency. These equations prove hard to solve analytically because the conditional probabilities depend on the frequencies in the previous generation. The genotype frequency recurrence equations can be studied numerically, however.

The numerical simulations were started at linkage equilibrium and run until the system had reached equilibrium (250 -5000 generations depending on the parameter values). The parameters involved were a, the strength of the mating preference, s, the strength of selection against individuals reared by bad mothers and t, the strength of selection against the indicator trait in males.

We use our equations to compare against results from previous models. Figure 6.1 shows the results for parameter values intended to capture the Kirkpatrick (1982) model ($a = 4$, $s = 0$, $t = 0.4$). The model is not symmetrical in the sense that the p allele mates randomly and does not preferentially mate with males without the indicator trait. These parameter values mean that females with the preference allele are simply choosing males that develop a detrimental trait when they are reared by a G mother, but with no benefit for the female in terms of good mother genes ($s = 0$).

The graph shows that the evolution of preference alleles and trait alleles is possible when maternal effects determine the trait that is chosen. The threshold before the trait and preference can invade is $f(P) = 0.17$. The internal line of neutral equilibria that is present in the Kirkpatrick (1982) model has collapsed into a single point equilibrium with maternal effects.

Figure 6.2 shows the results for parameter values intended to capture some of the features of the Seger and Trivers (1986) ($a = 4$, $s = 0.1$, $t = 0.2$). With these parameter values females cannot co-optimise the fitnesses of their sons and daughters. Once again the preference allele and the good mother allele can increase in frequency together, but this time there is no threshold frequency for the invasion of the preference allele. This reflects the increased fitness of preferring females due to the build-up of LD between the loci. However, the good mother allele fixes much faster than the LD can develop and so the preference allele only increases modestly in frequency. This shows that the

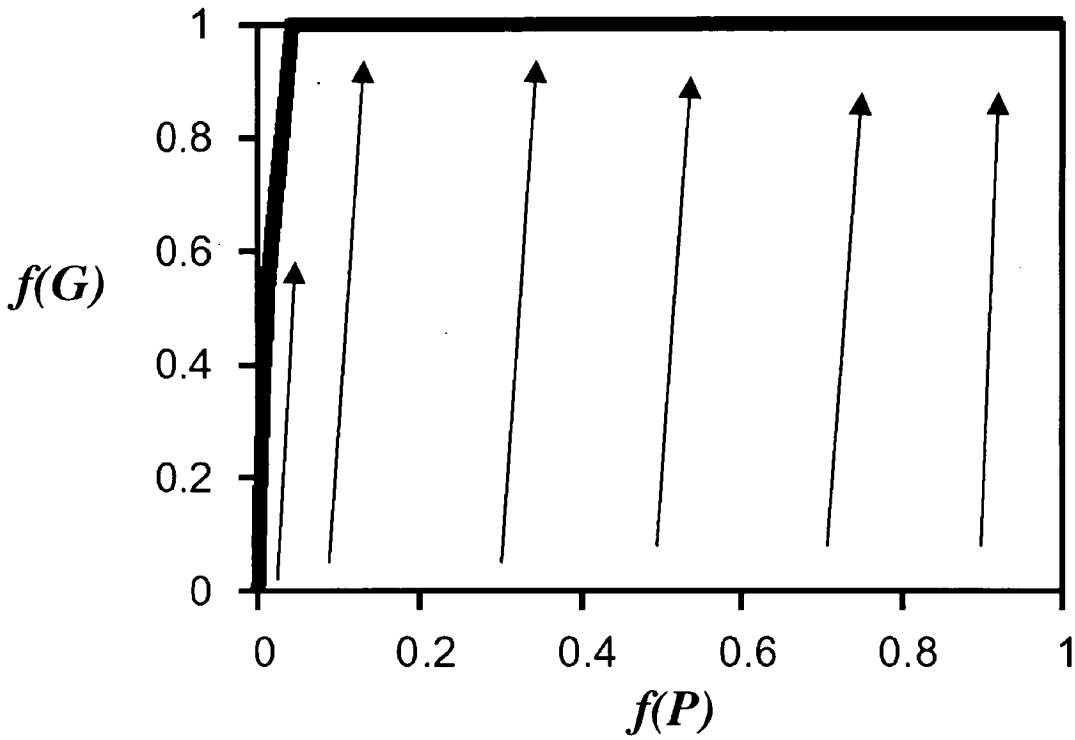


Figure 6.2. Schematic of the equilibrium results from a two-locus maternal effects model with parameter values that reflect those in the Seger and Trivers (1986) model. The heavy lines indicate the neutral lines of equilibria. Arrows indicate the direction and end points that allele frequency combinations from those vicinities reach. $f(P)$ - frequency of the female preference allele, $f(G)$ - frequency of the good mother allele. $a = 4$, $s = 0.1$, $t = 0.2$.

same process that operates in the Seger and Trivers (1986) model can also operate when maternal effects determine the expression of the indicator trait.

For both of these numerical studies all of the males are competent to develop the indicator trait if they are reared by a mother with the G allele. This will weaken the mating preferences, and therefore the build-up of LD between the loci, because males with the indicator trait do not always carry the G allele. This motivates a three-locus version of the model.

Table 6.3. Fitnesses for the three-locus maternal effects sexual selection model. The G allele is the good mother allele and results in higher fitness for both sons and daughters, but a fitness of $1-t$ for sons bearing the C allele because they are competent to develop a costly indicator trait. The g allele results in lower fitness for both sons and daughters ($1-s$).

		Mother's genotype	
		G	g
Males	C	1-t	1-s
	c	1	1-s
Females	C	1	1-s
	c	1	1-s

6.2.2 Three-locus model

Now we introduce a third locus (C, c) that determines the male competence to express the indicator trait when the mother has a G allele. The presence of this third locus means that females can now co-optimise the fitnesses of their sons and daughters. This can be seen in Table 6.3, which outlines the fitnesses in this model. It can be seen that the fitnesses are entirely determined by the G allele in mothers and by the C allele when present in a male.

The model is now more akin to previous models of a condition-dependent handicap (Andersson, 1986; Pomiankowski, 1987) based on the verbal models of Zahavi (1975, 1977). In these models, males that carry the appropriate trait allele will develop a conspicuous handicap that is dependent on their condition determined by another locus (viability). A three-way LD builds-up between the loci due to the female preferences for males that display the handicap. In this way the handicap is an honest signal of

Table 6.4. The mating frequencies for a three-locus maternal effects model where females carrying the P allele have a preference for males with a handicap trait that is a result of having a C allele and a mother with the G allele. The P_{ij}s refer to the probabilities of receiving both preferential and random matings for males in the 'ij' mating combination. Q is defined in the text.

		♂							
		GPC	GpC	Gpc	GPc	gPC	gPc	gpC	gpc
		h''	k''	u''	v''	w''	x''	y''	z''
♀	GPC (h')	$\frac{h'h''P_{hh}}{Q}$	$\frac{h'k''P_{hk}}{Q}$	$\frac{h'u''}{Q}$	$\frac{h'v''}{Q}$	$\frac{h'w''P_{hw}}{Q}$	$\frac{h'x''}{Q}$	$\frac{h'y''P_{hy}}{Q}$	$\frac{h'z''}{Q}$
	GpC (k')	$k'h''$	$k'k''$	$k'u''$	$k'v''$	$k'w''$	$k'x''$	$k'y''$	$k'z''$
	Gpc (u')	$u'h''$	$u'k''$	$u'u''$	$u'v''$	$u'w''$	$u'x''$	$u'y''$	$u'z''$
	GPc (v')	$\frac{v'h''P_{vh}}{Q}$	$\frac{v'k''P_{vk}}{Q}$	$\frac{v'u''}{Q}$	$\frac{v'v''}{Q}$	$\frac{v'w''P_{vw}}{Q}$	$\frac{v'x''}{Q}$	$\frac{v'y''P_{vy}}{Q}$	$\frac{v'z''}{Q}$
	gPC (w')	$\frac{w'h''P_{wh}}{Q}$	$\frac{w'k''P_{wk}}{Q}$	$\frac{w'u''}{Q}$	$\frac{w'v''}{Q}$	$\frac{w'w''P_{ww}}{Q}$	$\frac{w'x''}{Q}$	$\frac{w'y''P_{wy}}{Q}$	$\frac{w'z'}{Q}$
	gPc (x')	$\frac{x'h''P_{xh}}{Q}$	$\frac{x'k''P_{xk}}{Q}$	$\frac{x'u''}{Q}$	$\frac{x'v''}{Q}$	$\frac{x'w''P_{xw}}{Q}$	$\frac{x'x''}{Q}$	$\frac{x'y''P_{xy}}{Q}$	$\frac{x'z'}{Q}$
	gpC (y')	$y'h''$	$y'k''$	$y'u''$	$y'v''$	$y'w''$	$y'x''$	$y'y''$	$y'z''$
	gpc (z')	$z'h''$	$z'k''$	$z'u''$	$z'v''$	$z'w''$	$z'x''$	$z'y''$	$z'z''$

the male's quality. Since the high viability allele will always increase in frequency the handicap trait allele and the preference allele must hitchhike to higher frequencies with the high viability allele.

In our maternal effects version of this model, females do not receive an immediate benefit to mating with males demonstrating they carry good mother alleles, but their offspring's offspring will receive a fitness benefit. Table 6.4 shows the mating frequencies for the 64 mating combinations in this model. A difference can be seen compared to the mating frequencies in Table 6.2; not all of the males have non-zero probabilities of

receiving preferential matings from P females. This strengthens the mating preferences, and therefore, the build-up of LD between the loci.

In the Table the denominator, Q, is the three-locus equivalent of B from the two-locus model and is equal to

$$\begin{aligned} Q =& h''(aP(G_{mother}|GPC) + (1 - P(G_{mother}|GPC)) + k''(aP(G_{mother}|GpC) + \\ & (1 - P(G_{mother}|GpC)) + w''(aP(G_{mother}|gPC) + (1 - P(G_{mother}|gPC)) + \\ & y''(aP(G_{mother}|gpC) + (1 - P(G_{mother}|gpC)) + u'' + v'' + x'' + z''. \end{aligned} \quad (7)$$

Genotype frequency receurrence equations can be derived from the mating frequencies in Table 6.4 together with changes describing the effect of natural selection. Using these equations we can carry out numerical simulations in the manner described previously.

Figure 6.3 shows the results from a numerical simulation using the following parameter values: $a = 4$, $s = 0.1$ and $t = 0.4$. Once again there is a threshold frequency of the preference allele before it can invade, but now there is also a 'runaway' co-evolution of the trait and the preference alleles to high frequencies. This matches the qualitative behaviour seen in the Pomiankowski (1987) model for similar parameter values ($a = 4$, $t = 0.4$, and $H = 0.4$ - H reflects cyclical coevolution between parasites and hosts and cannot be compared directly with s in this model). However, it also matches remarkably closely the quantitative behaviour of the Pomiankowski (1987) model. The threshold frequency for invasion in the Pomiankowski (1987) version of the zahavi handicap model was 0.23, and was 0.21 for the 'revealing' handicap model. In our maternal effects model the threshold for invasion is 0.21 (although the qualitative behaviour matches the zahavi handicap model).

Runaway evolution occurs in the three-locus model because the mating preferences are not as weak as in the two-locus model and this allows LD between the loci to build up quicker. Therefore, we have shown that in three different models of sexual selec-

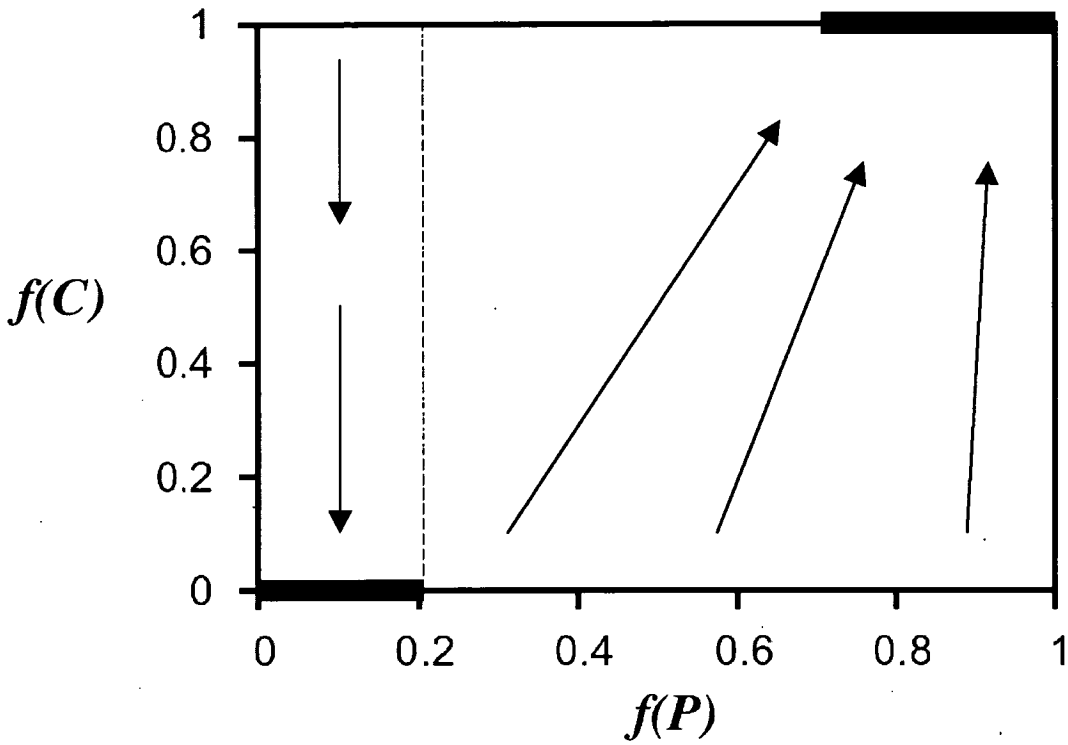


Figure 6.3. Schematic of the equilibrium results from a three-locus maternal effects model with parameter values that reflect those in the Pomiankowski (1987) handicap model. The heavy lines indicate the neutral lines of equilibria. Arrows indicate the direction and end points that allele frequency combinations from those vicinities reach. $f(P)$ - frequency of the female preference allele, $f(C)$ - frequency of the male handicap allele (only expressed when reared by G mothers - condition-dependent maternal effect). $a = 4$, $s = 0.1$, $t = 0.4$.

tion, maternal effects equivalents of the models produce very similar results thereby illustrating that maternal effects can be an important component of mate choice.

6.2.3 Mate choice based on self-similarity

Now we turn our attention to a different model of mate choice that does not incorporate maternal effects or detrimental indicator or handicap male traits. We are interested in the

Table 6.5. The mating frequencies for a two-locus mate choice model where females carrying the P allele have a preference for males with the same visible phenotype that they have, controlled by a second locus (V, v). Both L and M are defined in the text.

		♂			
		$PV\ (x)$	$Pv\ (y)$	$pV\ (z)$	$pv\ (u)$
♀	$PV\ (x)$	$\frac{ax^2}{L}$	$\frac{xy}{L}$	$\frac{axz}{L}$	$\frac{xu}{L}$
	$Pv\ (y)$	$\frac{yx}{M}$	$\frac{ay^2}{M}$	$\frac{yz}{M}$	$\frac{ayu}{M}$
	$pV\ (z)$	zx	zy	z^2	zu
	$pv\ (u)$	ux	uy	uz	u^2

evolution of mate choice when a female preference allele (P) prefers to mate with males that have the same visible phenotype as they do, determined by a second locus (V, v), which produces a visible phenotype. In essence, this is a model of positive assortative mating via female preferences. Once again we model a haploid population, but this time only sexual selection operates as there are assumed to be no viability differences between the genotypes for natural selection to act upon. The non-preference allele is assumed to mate randomly and there is no cost to mate choice. However, we no longer assume free recombination between the loci.

Table 6.5 shows the mating frequencies for the 16 mating combinations in this model. It can be seen that both female genotypes carrying the P allele give preferential matings to different males since they have different alleles at the visible phenotype locus. It might be expected, then, that any effect of preferential mating will cancel out. However, this will only be the case when the initial frequency of the visible phenotype is exactly 0.5. Either side of this and the more frequent phenotype will gain preferential matings more often than its counterpart, which in turn should allow LD to build-up

between the loci. L is then equal to

$$L = ax + y + az + u \tag{8}$$

and M is

$$M = x + ay + z + au. \tag{9}$$

Although this is straight-forward to analyse from the genotype frequencies we will pursue an approach based on the multilocus formalism developed by Barton and Turelli (1991) and Kirkpatrick et al. (2002). This method follows only the allele frequencies and the associations between (and within) loci and between individuals (using central moments). The advantage of using this method is that it allows the derivation of selection coefficients and association coefficients that detail the forces acting on single loci and groups of loci both within and between individuals in males and in females. The method also enables the use of the so-called quasi-linkage equilibrium (QLE) approximation that assumes weak selection relative to recombination so that the LD can be assumed to reach an approximate equilibrium to reduce the dimensions of the system and enable analytical solutions. The main features of this methodology have been outlined in Chapter 4.

First of all we would like to derive selection coefficients and coefficients of association (selection acting to favour certain associations between sets of loci) for the system. To do so it will prove useful to set the model up in the following way. Random mating is assumed to occur between males and females and the frequencies of the resulting mating pairs are modified by the mating preferences of the females. In this sense non-random mating is treated in the same way as natural selection. An important consequence of this scheme is that pre-selection (pre-mating) associations between individuals are random and so equal to zero. The preference locus will be referred to as p and the phenotype locus as v, and p_p is the preference allele frequency and p_v is the '1' phenotype allele

frequency. From Table 6.5 the fitness of an arbitrary mating pair can be expressed in terms of the female and male (starred) indicator variables as follows

$$W(X,X^*) = \frac{1}{L^*M^*}(M^*X_pX_t + L^*(M^* + aX_p - M^*X_p - aX_pX_t) + X_pX_v^*(a-1)(L^*(X_v-1) + M^*X_v)) \quad (10)$$

where

$$L^* = 1 + p_v(a-1) \quad (11)$$

and

$$M^* = p_v + a(1-p_v) \quad (12)$$

and are L and M with the genotypes expressed in allele frequencies and LD. Substituting $X_p = \zeta_p + p_p$, $X_v = \zeta_v + p_v$, $X_p^* = \zeta_p^* + p_p$ and $X_v^* = \zeta_v^* + p_v$ and rearranging we can find the selection coefficients. We will express the coefficients in the form $a_{U,V}$ where U is a set of loci in the females and V is a set of loci in the males. The four non-zero coefficients are

$$a_{\emptyset,v} = \frac{p_p(a-1)(1-2p_v)}{(a(p_v-1)-p_v)(1+p_v(a-1))} \quad (13)$$

$$a_{p,v} = \frac{(1-2p_v)(a-1)}{(a(p_v-1)-p_v)(1+p_v(a-1))} \quad (14)$$

$$a_{v,v} = \frac{(1-a^2)}{(a(p_v-1)-p_v)(1+p_v(a-1)} \quad (15)$$

$$a_{pv,v} = \frac{(1-a^2)}{(a(p_v-1)-p_v)(1+p_v(a-1))} \quad (16)$$

where $\emptyset$ refers to the empty set. $a_{\emptyset,v}$ refers to the sexual selection acting on the phenotype locus. $a_{p,v}$ refers to non-random mating selecting for the association between the female preference and male phenotype alleles. $a_{v,v}$ and $a_{pv,v}$ both refer to the effects of assortative mating. $a_{p,v}$, $a_{v,v}$ and $a_{pv,v}$ can only generate associations within gametes (LD) via recombination. The mean fitness is equal to one in this model.

Non-random mating is followed by transmission, which in this model is simply recombination between the loci. The change in allele frequency of the female preference allele due to non-random mating can be found from the general equation

$$D'_A = D_A + \sum_{B \subseteq W} a_B (D_{AB} - D_A D_B) \tag{17}$$

where W is the set of all gene positions, A and B are subsets and the Ds are the linkage disequilibria between or within the appropriate subsets. For our case, A is a single position, the preference allele, and the change due to non-random mating is

$$\Delta p_p = \tilde{a}_{\emptyset,v} D_{pv} + \tilde{a}_{pv,v} D^2_{pv} \tag{18}$$

and for the visible trait allele

$$\Delta p_v = \tilde{a}_{\emptyset,v} p_v (1 - p_v) + \tilde{a}_{pv,v} D_{pv} p_v (1 - p_v) \tag{19}$$

where $\tilde{a}_{U,V}$ is the symmetrised coefficient between sets U and V ($\frac{a_{U,V} + a_{V,U}}{2}$), which in this model means they are simply divided by two (genes expressed in only one sex). After non-random mating the LD becomes (from equation 17)

$$D'_{pv} = D_{pv} + \tilde{a}_{\emptyset,v}(1 - 2p_v) D_{pv} + \tilde{a}_{pv,v}(1 - 2p_v) D^2_{pv}. \tag{20}$$

Now we need expressions for the effect of recombination on the LD. After non-random mating the association between haploid individuals is no longer zero, as we assumed earlier, and recombination between associated mating pairs can increase the LD within gametes, therefore we want an expression for the between individuals post-non-random mating association, which (from (17)) is

$$\begin{aligned} D'_{p,v} =& a_{v,v} p_v (1 - p_v) D_{pv} + \tilde{a}_{p,v} p_p (1 - p_p) p_v (1 - p_v) + \tilde{a}_{p,v} D^2_{pv} + \\ & \tilde{a}_{pv,v} p_v (1 - p_v)(1 - 2p_p) D_{pv} + \tilde{a}_{pv,v}(1 - 2p_v) D^2_{pv}. \end{aligned} \tag{21}$$

Putting this together with (20) and updating the reference values after non-random mating (which changes allele frequencies, which are our reference values) the change of LD over one generation is

$$\Delta D_{pv} = (1-r)D'_{pv} + rD'_{p,v} - \Delta p_p \Delta p_v - D_{pv} \tag{22}$$

where r is the rate of recombination between the loci. Now we have a complete description of the dynamics of the system

$$\begin{aligned}
\Delta p_p &= \tilde{a}_{\emptyset,v} D_{pv} + \tilde{a}_{pv,v} D_{pv}^2 \\
\Delta p_v &= \tilde{a}_{\emptyset,v} p_v(1-p_v) + \tilde{a}_{pv,v} D_{pv} p_v(1-p_v) \\
\Delta D_{pv} &= (1-r)D'_{pv} + rD'_{p,v} - \Delta p_p \Delta p_v - D_{pv}.
\end{aligned} \tag{23}$$

These exact equations can be derived from Table 6.5 by converting genotype frequency recurrence equations into allele frequencies and LD. It can be seen that the preference allele relies upon its association with the visible trait allele (assuming the frequency is not 0.5) to increase in frequency. The visible trait allele that is dominant to begin with will sweep to fixation and the preference allele will always hitchhike to a higher frequency so long as $a > 1$, though very weakly.

The build-up of LD will be particularly slow in this model as the preference allele will always be giving preferential matings to both of the male types. Assuming that the system starts in linkage equilibrium, numerical simulations show that the increase of the preference allele will be negligible although one could imagine a number of modifiers hitchhiking on neutral variation as it sweeps to fixation. An alternative possibility is that there is already an association between the preference allele and a trait allele (due to drift/inbreeding, population structure etc).

Figure 6.4 shows that in this scenario there can be a large effect of the recombination rate as tightly linked loci will be able to preserve the association and hitchhike to higher

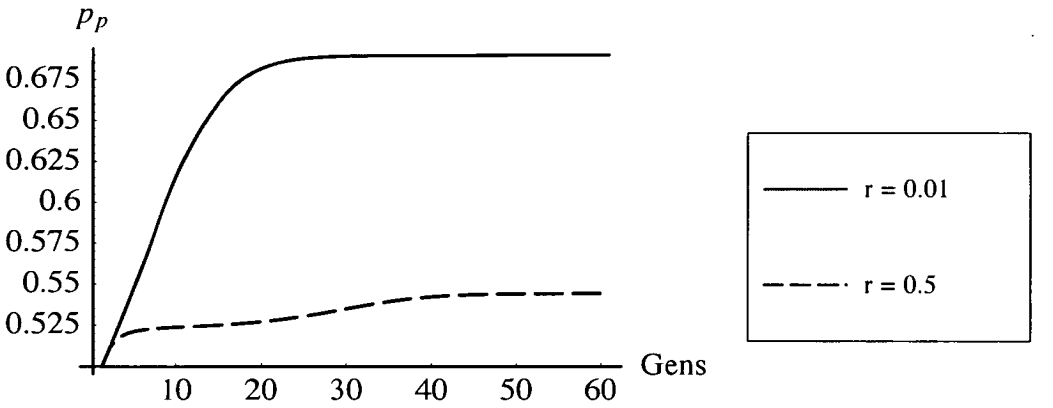


Figure 6.4. Numerical simulations based on exact allele frequency and LD equations for the dynamics of a positively assorting female preference allele. The graphs are plotted assuming an initial LD of 0.1 and an intital frequency of 0.5 for the preference allele and 0.501 for one of the visible trait alleles. $a = 4$, Gens - generations.

frequency. A good example of this may be the meeting of two recently diverged, but inter-fertile populations. In this instance a preference allele that mates assortatively will have a greater advantage over a randomly mating allele.

6.3 Discussion

Three models of the evolution of mate choice have been described and analysed. Two models sought to integrate maternal effects into two and three-locus models of sexual selection. The results showed that for all of the models investigated maternal effects produced the same qualitative, and in some cases quantitative, results as previous models. This supports the idea that maternal effects may be components of mate choice in natural populations.

In what species might maternal effects be important determinants of mate choice?

As we have modeled a polygynous species we will only explore these cases, i.e. not monogamous species where sexual selection will be confounded with direct benefits from choosing a mate. We would expect maternal effects to be important for species where there is maternal care of the eggs and to some extent the young.

A striking example of this is found in *Forficula auricularia*, the european earwig. In this species, and in earwigs more generally, the female looks after her small clutch of between 20 to 50 eggs after they have been laid in an underground nest. The female will remain with the eggs, protecting them, cleaning them and maintaining their moisture balance. Once hatched she forages for and protects the hatchlings until their first molt. This is a remarkable level of maternal care for an insect. One obvious effect of cleaning the eggs is that they will be free of parasites. It has been shown, in birds at least, that there is a trade-off between immunocompetence and growth, i.e. if a young offspring has to divert resources to fend off parasites they will have less resources to put into growth (Saino et al., 1998; Soler et al., 2003). Such a trade-off means that less parasitised eggs can give rise to larger male offspring with more resources to put into developing condition-dependent indicator traits.

Additionally, the species are sexually dimorphic for forceps size with males sporting larger forceps than females. There may be evidence that the forceps size is at least partly environmentally determined and therefore condition-dependent (Tomkins, 1999). There is also evidence that females prefer males with larger forceps (Tomkins and Simmons, 1998).

Added to this, there is also a dimorphism within males with a large (macrolabic) and small forceps variety. This dimorphism tends to be found more often in dense populations and it has been speculated that the large morph may benefit more from contests in dense populations thereby favouring the (costly) production of forceps (Tomkins and

Brown, 2004). However, there is an alternative explanation and that is that denser populations will be more competitive and healthier individuals (who have been reared well by their mothers) will do better thereby providing an incentive to select a mate on this basis. In a species where the level of maternal care plays such a crucial role in the survival and subsequent competitiveness of individuals it would not be surprising if it played a part in mate choice.

Another potentially important maternal role is that of oviposition. The decision of where and when to lay eggs can have a profound effect on the nutritional health and survival of the offspring. In the longhorn beetle (*Ceramycidae*) females oviposit eggs into rotting or live wood depending on the species. The nutritional quality of the host in terms of nitrogen content and its ability to resist parasitism is critical for the development of the larvae (Hanks, 1999). Additionally, the adults are sexually dimorphic for antennae length and females, in one species at least, have been shown to prefer males with longer antenna, which are a correlate of body size (Hughes and Hughes, 1985). Females are also known to respond to many different volatiles of freshly wounded trees (Hanks, 1999; Crook et al., 2004).

The oviposition site is likely to be important for stalk-eyed flies too. The development of the stalk-eyes in certain species has been shown to be condition-dependent (Cotton et al., 2004), and in the wild females oviposit into trees or roots. The same is true of stag beetles, carrion flies and a host of other insects where oviposition plays a crucial role in the fitness of the offspring.

Mate choice is a complex process in many animals with many variables feeding into it. It seems likely that there are a number of criteria that females use for finding a suitable mate and a number of causes of particular patterns of mate choice in nature. How might it be possible to determine the importance of maternal effects for mate choice? The

theory of Trivers and Willard (1973) predicts that females in good condition or when mated to a male in good condition will favour investing in male offspring because males in good condition will give her more grandoffspring due to the increased variance in reproductive success between males and females in many species. Typically, the idea is considered most relevant to species that have a substantial period of parental investment post-birth of the young. This may be relevant to the european earwig at least, but rather than expect a bias to male offspring we may expect a slight bias to female offspring when females are mated to a preferred male. Although when females are in good condition it would still make sense to have male offspring.

In general it will be difficult to separate the effects of male-male contest from those of maternal effects. The same problem was discussed by Seger and Trivers (1986) in relation to their female-male constraints model. The complication that arises is identical in this model. For example, in the european earwigs it is difficult to determine whether the forceps are simply the result of male-male contest for access to females or whether they are indicators of their mother's parenthood. Given that females actively choose macrolabic males (true also in stalk-eyed flies and longhorn beetles) it seems that there is more to the story than simply male-male contest.

It has been suggested that whenever males can 'monopolise' females via, for example, defending a critical resource against other males (resource defence polygyny) or through male dominance polygyny, then male 'emancipation' and polygyny will evolve (Emlen and Oring, 1977). The logic is that females will do better in polygynous systems when the benefits they receive from, for example, access to improved grazing pastures, outweigh the costs of polygyny (Orians, 1969). However, it seems that in all but the most extreme cases, females will have some active as opposed to passive choice in their mating partner.

We have also looked at a multi-locus model of positive assortative mating via female mate choice. Typically, when positive assortative mating is symmetrical there is no change of allele frequencies (e.g. Barton and Turelli, 1991). However, when positive assortment occurs in a polygynous species there will be allele frequency changes provided the loci being assorted with respect to are not equal to 0.5. We have shown that when this happens a female assortment allele can hitchhike to higher frequency with the assorted locus.

The build-up of LD is very weak in the system as preferential matings are given to both male types in the population. However, it has been shown that in the case when there is already LD present, linkage between the loci will enhance the spread of an assortment allele in females. For example, it could be that two recently diverged, but inter-fertile populations encounter each other and there are neutral phenotypic markers for males in each population. There will be LD between the loci and an assortment allele can increase in frequency.

Extensions to the model that are of interest are increasing the number of phenotype loci or modelling it as a quantitative trait so that females have more information to base assortment on. Also, the spread of an assortment allele may be enhanced in diploid populations where a dominant allele will receive more preferential matings than in the haploid case and will take longer to fix in the population due to recessive alleles 'hiding' in heterozygotes. Both of these facts may facilitate the enhanced build-up of LD between the loci. Developing diploid models of sexual selection with the Barton and Turelli (1991) and Kirkpatrick et al. (2002) methodology is of interest in itself. An obvious complication of such models will be the deleterious effects of inbreeding in diploids.

In summary, we have developed multilocus sexual selection models that take into

account maternal effects. In all of the published models looked at, maternal effects allows for the evolution of male traits and female preferences. These results support the notion that females may gain a genetic benefit from mate choice. Much ingenuity will be required to tease apart the effects of male-male contest from active female choice for signatures of good female parentage.

7 Marine snow and algal toxins: An evolutionary perspective

Abstract

There are many seemingly paradoxical biological adaptations, which when viewed from the perspective of the individual appear to be detrimental, but when viewed from the perspective of the group are beneficial. Intuitively appealing group selection arguments must assume that relatedness is high enough to select against individually selfish behaviours that would otherwise disrupt the evolution of the trait. Such models also usually make implicit assumptions about the scale of competition (soft or hard selection). Here, I relax these assumptions and describe evolutionary models of two important, and paradoxical, natural phenomena associated with phytoplankton: marine snow in diatoms and algal toxin production in a range of algae including dinoflagellates, cyanobacteria and raphidophytes. Conditions under which these behaviours can evolve are derived using a game theoretical approach, and it is shown that small asymmetries in life-history traits may be sufficient to bias the relative self-benefit of the behaviours and enable their evolution.

7.1 Introduction

Phytoplankton play a central and dynamic role in global nutrient and energy cycles with estimates suggesting that half of the planet's net primary production may be attributed to marine phytoplankton (Longhurst et al., 1995; Field et al., 1998). Additionally, the excretion of dimethylsulphide (DMS) by certain haptophyte species influences cloud formation, thereby helping to regulate the planet's temperature (Charlson et al., 1987; Hamilton and Lenton, 1998). Therefore, it is necessary to understand phytoplankton behaviour and ecology. Two important behaviours that require explanation are marine

snow and algal toxin production.

Marine snow occurs throughout the oceans and is a central component in the cycling of nutrients and energy through the marine ecosystem via the transport of primary production in the euphotic zone to the benthos where it can be degraded and released back into the environment (Alldredge and Silver, 1988). Diatom aggregation is one of the major contributors to marine snow (Thornton, 2002a). Paradoxically, diatom aggregation often occurs during active growth in the spring diatom bloom when grazing by heterotrophs is almost non-existent and the environment is nutrient replete (Riebesell, 1988; Kiorboe et al., 1994). It has been estimated that up to 39% of the primary production may be removed via passive sedimentation in the spring (Becquevort and Smith, 2001). Estimates indicating that diatoms represent 40 - 45% of the net primary productivity of the oceans further highlights the scale of this phenomenon (Mann, 1999).

A number of hypotheses have been proposed to explain the apparent paradox of aggregating diatoms. Logan and Alldredge (1989) suggested that the formation of fast-sinking aggregates increases the rate of nutrient uptake by the aggregated algae. While this may be true, it is probably irrelevant because settling aggregates will be quickly removed from the euphotic zone. Smetacek (1985) proposed that aggregation is an adaptive response in nutrient-depleted, grazed populations whereby aggregated cells settle to the sea floor where they revert to a resting stage and return to the euphotic zone when conditions are more favourable. This process would occur in populations close to carrying capacity, but it has been observed that actively growing populations are prone to aggregation and so this requires a different explanation (Riebesell, 1988; Kiorboe et al., 1994). Hansen et al. (1995) described a model of interference-competition between different algal species that suggested that 'sticky' species with smaller cell sizes and higher intrinsic growth rates would benefit from aggregation by removing their competitors

and enhancing their relative abundance in the water column. This model is relevant to actively growing populations and is the model that is discussed here. I extend their analysis from a purely ecological model to an evolutionary game-theoretical approach and relax the assumption of entirely local competition to derive conditions under which 'stickiness' can evolve. Further to this I show that marine snow can be conceptualised as a spiteful trait as well as an interference-competition adaptation (Hamilton, 1970; Gardner et al., 2004; Gardner and West, 2004).

In marine systems there are a number of dinoflagellate species (some non photosynthetic) that commonly bloom after the demise of the silica-depleted diatom bloom and produce neurotoxins, which in sufficient quantity can kill fish, accumulate in filter-feeders and disrupt the trophic structure of the surrounding waters. At the height of the bloom the water can become discoloured producing what is known as a 'red tide'. At present around 200 species of toxic algae have been identified belonging to the dinoflagellates, diatoms, raphidophytes, prymnesiophytes, silicoflagellates, cilliates, and cyanobacteria (Smayda, 1997a; Landsberg, 2002), but with an estimated 75% contributed by the dinoflagellates alone (Smayda, 1997a,b). In freshwater lakes cyanobacteria can produce both neurotoxins and hepatotoxins, which are also lethal for fish. As well as being responsible for massive fish losses to the fishing industry, human consumption of contaminated filter-feeders can result in a range of (sometimes lethal) disorders including paralytic shell-fish poisoning (PSP), and aerosols of toxins can result in respiratory disorders for coastal inhabitants.

Production of toxins by algae presents a paradox. The manufacture and storage of toxins will most likely present a significant energetic cost for algal unicells and the potential benefits will be enjoyed by all of the algal population, both clonal relatives and non-relatives. Assuming the benefits are shared equally among all of the algal lin-

eages, then even with a high relatedness such a behaviour will not increase the relative abundance of the toxin producer's lineage. It has been suggested that toxins may serve merely as an intracellular store of nitrogen, so that any toxic side-effects of toxin released post-cell rupture or consumption are entirely incidental (Steidinger and Baden, 1984; Boyer et al., 1987). This seems wanting as a general explanation, however, as it has been shown that an entire class of toxins, the brevetoxins, are devoid of nitrogen (Baden, 1989).

A more plausible proposition is that the toxins serve as an anti-predator defence and deterrant mechanism (e.g. Wolfe et al., 1997). This idea has received some empirical support in terms of grazers preferring non-toxic algal species when given the choice (Kirk and Gilbert, 1992). One potential problem with this explanation is that during extensive algal blooms the algal density becomes so high (as high as 10^6 cells/ml) that the algae have effectively escaped grazer-control, and yet still produce toxins. I propose another possibility, that toxin production is a response to the depletion of nutrients and degradation of the environment that occur as a result of the algal bloom, with the intention of replenishing nutrients through the rapid death and decay of vertebrate (and/or invertebrate) populations. There is some empirical support for this idea from work showing that toxin production is generally inversely related to the algal growth rate (Proctor et al., 1975; Taroncher-Oldenburg et al., 1999). Additionally, toxin production has been shown to increase when algae are phosphate limited (Hall, 1982; Boyer et al., 1987; Taroncher-Oldenburg et al., 1999). The latter two propositions are explored here in a general model designed to capture the main details whilst remaining tractable.

7.2 Models and analyses

7.2.1 Marine snow

The aggregation of diatoms in the ocean depends on the probability that cells or groups of cells collide and, once they have collided, the probability that they stick. Collision probabilities are governed by the density of cells in the water column and by physical processes such as Brownian motion, shear and differential settlement. Sticking probabilities are governed by the physicochemical properties of the cell surfaces and as such may be modified by the algae themselves. Hence, the probability of both colliding and sticking is simply

$$P[C \cap S] = P[S|C]P[C]. \tag{1}$$

Henceforth, $P[S|C]$ will be replaced with s and $P[C]$ with β. Jackson (1990) described the loss of aggregated diatom dimers from the euphotic zone in a single species as

$$\frac{dC}{dt} = \mu C - 2\beta s C^2 \tag{2}$$

where μ is the per capita growth rate and C is the concentration of the diatoms. The model therefore assumes that the rate at which diatom collisions occur is proportional to the product of their concentrations, C^2, known in chemistry as the law of mass action. It was shown that the dynamics of aggregate formation were dominated by the production of dimers so that more complex aggregate collisions can be neglected (Jackson, 1990; Jackson and Lochmann, 1992).

Hansen et al. (1995) extended this model to a two-species scenario and included density-dependent regulation of the population via an environmental carrying capacity. In a two-species interaction there are two different sticking probabilities and three different collision probabilities, one for each species and a third for the collision between cells of different species. Additionally, they allowed for different per capita growth rates

for the two species. This is the model that is developed here, but referring to lineages as opposed to species. The social arena is assumed to consist of just two lineages, X and Y, and this is the scale over which aggregations can occur. We want to find conditions under which a non-zero stickiness strategy can invade and so assume that lineage X is a minor variant from the average strategy of $s = 0$, and invests δs. The benefit from playing this strategy must be assessed over the growth cycle and not just at an instantaneous point in time, and so explicit solutions for the growth dynamics are required. There is also a small asymmetry in growth rates between the lineages, expressed as $\delta\mu$, which is of the order of δs. The equations describing the density-independent case are

$$\begin{aligned} \frac{dx}{dt} &= (\mu + \delta\mu)x - 2\beta_x \delta s x^2 - 2\beta_{xy}\frac{\delta s}{2}xy \\ \frac{dy}{dt} &= \mu y - 2\beta_{xy}\frac{\delta s}{2}xy \end{aligned} \tag{3}$$

where β_x is the collision probability of cells in lineage X with each other and β_{xy} is the collision probability for cells from lineage X with cells from lineage Y. There is a two-fold cost to being sticky, which arises from the loss of cells through self-sticking and through sticking the other lineage. We can make progress in solving these equations by first finding a solution for $z\ (= x + y)$

$$\frac{dz}{dt} = z(\mu + p\delta\mu) - \delta s z^2 p(\beta_{xy}(1-p) + \beta_x p) \tag{4}$$

where $p = \frac{x}{x+y}$. This equation can be simplified by substituting $p = p_0 + O(\delta s)$, where p_0 is the initial frequency of the X lineage, and collecting terms of $O(\delta s^2)$ this becomes

$$\frac{dz}{dt} = z(\mu + p_0\delta\mu) - \delta s z^2 p_0(\beta_{xy}(1-p_0) + \beta_x p_0) + O(\delta s^2). \tag{5}$$

Separating the variables and integrating both sides

$$t = \int_{z_0}^{z_t} \frac{1}{z(\mu + p_0\delta\mu) - \delta s z^2 p_0(\beta_{xy}(1-p_0) + \beta_x p_0)} dz + O(\delta s^2) \tag{6}$$

we can rearrange to find a solution for z_t that can be expanded using a Taylor series to leading order in δs and $\delta\mu$ to give

$$z_t = \hat{z}_t + \frac{\delta s}{\mu} p_0(1-e^{-\mu t})(p_0(\beta_{xy}-\beta_x)-\beta_{xy})\hat{z}_t^2 + \frac{\delta\mu}{z_0} p_0 e^{-\mu t} t\hat{z}_t^2 + O(\delta s^2) \tag{7}$$

where

$$\hat{z}_t = z_0 e^{\mu t} \tag{8}$$

and may be considered the population size in the neutral case when δs and $\delta\mu$ are equal to zero. Now we wish to find an expression for p_t so that x_t can be expressed as $p_t \times z_t$. To do this it will prove useful to switch to the variable $q = \frac{p}{1-p}$ because the dynamics of q are exponential as opposed to logistic, thereby facilitating an explicit solution. If we note that $q = \frac{x}{y}$ then the quotient rule can be used to differentiate q

$$\frac{dq}{dt} = \frac{y\frac{dx}{dt} - x\frac{dy}{dt}}{y^2} = -\frac{q}{2}(z\delta s(\beta_{xy} + 2\beta_x p - 2\beta_{xy} p) - 2\delta\mu). \tag{9}$$

This expression can be simplified by substituting in $p = p_0 + O(\delta s)$ and $z = \hat{z}_t + O(\delta s)$ and collecting terms of $O(\delta s^2)$

$$\frac{dq}{dt} = -\frac{q}{2}(\hat{z}_t \delta s(\beta_{xy} + 2\beta_x p_0 - 2\beta_{xy} p_0) - 2\delta r) + O(\delta s^2). \tag{10}$$

Separating variables and integrating

$$\int (\hat{z}_t \delta s(\beta_{xy} + 2\beta_x p_0 - 2\beta_{xy} p_0) - 2\delta\mu)\, dt = -2\int_{q_0}^{q_t} \frac{1}{q}\, dq + O(\delta s^2) \tag{11}$$

gives

$$\frac{\delta s}{\mu} z_0 e^{\mu t}(\beta_{xy} + 2\beta_x p_0 - 2\beta_{xy} p_0) + C = -2\ln\left(\frac{q_t}{q_0}\right) + O(\delta s^2). \tag{12}$$

Substituting $t = 0$, and therefore $q_t = q_0$, we can find C

$$C = -\frac{\delta s}{\mu} z_0 (\beta_{xy} + 2\beta_x p_0 - 2\beta_{xy} p_0). \tag{13}$$

Substituting this back in, solving for q_t and expanding the result using a Taylor series

$$q_t = q_0 + q_0 t\delta\mu - \frac{\delta s}{2\mu}(\beta_{xy} + 2\beta_x p_0 - 2\beta_{xy} p_0)(e^{\mu t} - 1)q_0 z_0 + O(\delta s^2). \quad (14)$$

Now we can substitutie $q = \frac{p}{1-p}$ back in, rearrange to find p_t, and expand with a Taylor series

$$p_t = p_0 + \delta\mu(p_0(1-p_0)t) - \frac{\delta s}{2\mu} p_0(1-p_0)z_0(e^{\mu t} - 1)(\beta_{xy} + 2\beta_x p_0 - 2\beta_{xy} p_0) + O(\delta s^2). \quad (15)$$

It can be seen that the $\delta\mu$ term is always positive (so long as $\delta\mu > 0$), but that the δs term can be positive or negative depending on the βs and p_0. Now we can express x_t as $p_t \times z_t$ and expand with a Taylor series to give

$$x_t = \hat{z}_t p_0 + e^{\mu t} p_0 z_0 t p_0 \delta\mu - \frac{\delta s}{2\mu}\hat{z}_t z_0 p_0 e^{\mu t}(e^{\mu t} - 1)(\beta_{xy}(1-p_0) + 2\beta_x p_0) + O(\delta s^2). \quad (16)$$

If, to begin with, we assume entirely local competition (soft selection) then the fitness of lineage X can be expressed as $W = \frac{x_t}{z_t}$, and we can ask when

$$\frac{dW}{d\delta s}|_{\delta\mu=0} > 0 \quad (17)$$

and find this is true when

$$p_0 > \frac{\beta_{xy}}{2(\beta_{xy} - \beta_x)}. \quad (18)$$

Therefore, there is a threshold for the initial kinship (p_0) within the social arena which must be overcome before a sticky variant can invade the population. Thus, the benefit of being sticky displays a positive-frequency dependence in this system. Additionally, it can be seen that an asymmetry in collision probabilities is necessary for invasion of a sticky type, with β_x needing to be less than half of β_{xy} to compensate for the loss of two cells through self-sticking.

It is of interest to interpret this result in terms of Hamilton's rule (Hamilton, 1964), which states that the criterion for the increase of a gene for a social action is $rb - c > 0$

where r is the relatedness between actor and recipient, b is the conferred benefit and c is the cost to the actor. Queller (1994) demonstrated how a negative relatedness can arise in a population where the average relatedness between competitors is zero. In this scenario an average relatedness of zero implies that, for entirely local competition, $p \times 1 + (1-p)r = 0$, where r is the relatedness between the X and Y lineages, and is $r = -\frac{p}{1-p}$. To derive the relatedness in Hamilton's rule we must take into account the probability of interacting with clone mates and competitors, so that we can calculate the relatedness to the average interactant, which for the case of $\delta\mu = 0$ is

$$R = \frac{\beta_x p}{\beta_x p + \beta_{xy}(1-p)} + \frac{\beta_{xy}(1-p)}{\beta_x p + \beta_{xy}(1-p)}\left(\frac{-p}{1-p}\right) = \frac{(\beta_x - \beta_{xy})p}{\beta_x p + \beta_{xy}(1-p)}. \tag{19}$$

In this model the benefit conferred upon the competitor is -1 and the cost to the actor is 1 giving

$$-\frac{(\beta_x - \beta_{xy})p}{\beta_x p + \beta_{xy}(1-p)} - 1 > 0 \tag{20}$$

which is true when

$$p > \frac{\beta_{xy}}{2(\beta_{xy} - \beta_x)} \tag{21}$$

thereby recovering the condition in equation (18). The purpose of deriving this is to demonstrate that the evolution of stickiness can be conceptualised as a spiteful adaptation that involves a negative relatedness (Hamilton, 1970; Gardner et al., 2004; Gardner and West, 2004). The result can be understood in terms of kin discrimination, where the greater the ability to discriminate (the lower β_x is relative to β_{xy}) the greater the benefit from being spiteful, and so the lower the threshold for invasion of a spiteful type. At the extreme of complete kin discrimination, $\beta_x = 0$, $p > \frac{1}{2}$.

It is possible to investigate how small differences in growth rates might influence the invasion of a sticky type by deriving the condition in equation (18) for non-zero $\delta\mu$

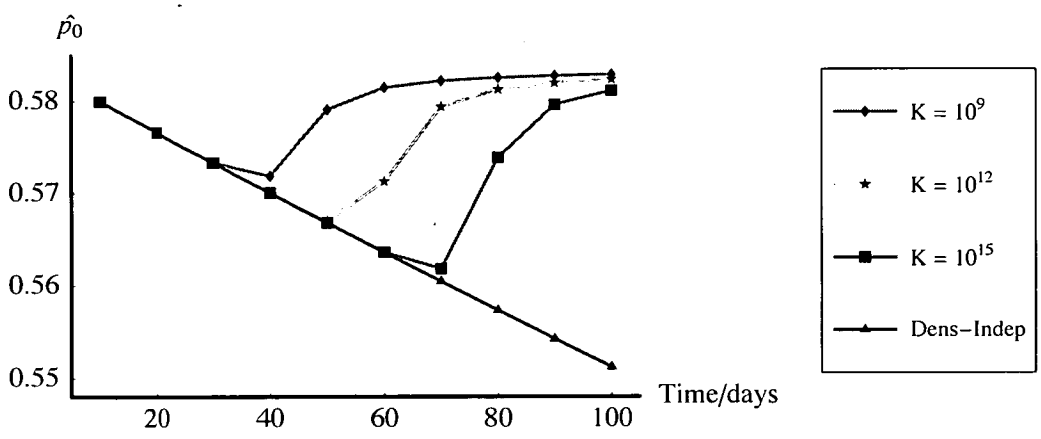


Figure 7.1. A plot showing the frequency invasion threshold for a sticky diatom variant as a function of the time since the start of growth for a range of carrying capacities together with the density-independent case.

giving

$$p_0 > \frac{\beta_{xy}}{2(\beta_{xy} - \beta_x + \beta_{xy} t \delta\mu)}. \tag{22}$$

Positive differences in growth rates will lower the threshold for invasion of a sticky type and this effect scales with the time since the start of growth. With a higher growth rate, a sticky lineage will be able to recover from the cost of its own stickiness and impart a greater negative benefit upon its competitors the longer growth continues for.

So far we have assumed that growth is density-independent, and so is never limited by depletion of the environmental nutrient pool. Relaxing this assumption, equation (3) for the X lineage becomes

$$\frac{dx}{dt} = (\mu + \delta\mu)x\left(1 - \frac{x+y}{K}\right) - 2\beta_x \delta s x^2 - 2\beta_{xy}\frac{\delta s}{2}xy \tag{23}$$

where K is the carrying capacity of the environment. When $\delta\mu = 0$ the condition for invasion remains unchanged, but when $\delta\mu$ is non-zero we have to find the threshold

numerically. Figure 7.1 shows the threshold frequency for the invasion of a sticky type as a function of time since the start of the bloom. Since we have assumed that $\delta\mu$ is very small this is only to illustrate the point that the benefit of being sticky increases with K, the carrying capacity of the environment, as this extends the period of exponential growth.

Thus far we have assumed entirely local competition to simplify things, but now we relax this assumption and examine what effects non-zero global competition may have. Global competitors are those outside of the scale of collision interaction in the social arena, but who compete with the focal lineage for global resources. First of all we look at the density-independent case. Fitness is now expressed as

$$W = \frac{x_t}{az_t + (1-a)\bar{z}_t} \tag{24}$$

where a represents the proportion of competition that occurs between individuals within the social arena ($a = 1$ is entirely local competition) and $\bar{z}_t$ is the average population size of a lineage averaged over all of the social arenas. Assuming resources are equally distributed over the n social arenas, that the initial population density is the same between social arenas, and that n is large enough that we can neglect the focal lineage's dynamics from the average, then we would expect $\bar{z}_t$ to be equal to $\hat{z}_t$, the population size in the neutral case. The population dynamics of the metapopulation will be given by

$$\frac{du}{dt} = \mu u \tag{25}$$

where

$$u = \sum_{i=1}^{n} z_i = n\bar{z}. \tag{26}$$

Solving for $\bar{z}_t$ we find

$$\bar{z}_t = z_0 e^{\mu t} \tag{27}$$

which is equal to $\hat{z}_t$ for the density-independent case. Plugging this into the fitness and asking when $\frac{dW}{d\delta s}|_{\delta\mu=0} > 0$ we find this to be true when

$$\frac{2\beta_x - \beta_{xy} - 2a\beta_{xy} + \sqrt{4\beta_x^2 - 4\beta_x\beta_{xy} + \beta_{xy}^2 - 4a\beta_{xy}^2 + 4a^2\beta_{xy}^2}}{4a(\beta_x - \beta_{xy})} < p_0 < \frac{\beta_{xy} - 2\beta_x + 2a\beta_{xy} + \sqrt{4\beta_x^2 - 4\beta_x\beta_{xy} + \beta_{xy}^2 - 4a\beta_{xy}^2 + 4a^2\beta_{xy}^2}}{4a(\beta_{xy} - \beta_x)}. \quad (28)$$

The conditions for increase are now more restrictive, requiring strong local competition and introducing an additional upper limit for p_0. This upper limit reflects the weakened benefit to being sticky for lineages in the majority when there is some non-zero global competition. In the density-dependent case the situation is more complex as local and global competition will enter the growth dynamics via the carrying capacity, but we can solve for the special case, $a = 0$, and find that a sticky variant can never invade showing that some non-zero local competition is essential for evolution of the trait.

7.2.2 Algal toxins

A complete model describing predator-prey interactions would require a description of algal and grazer dynamics concurrently and this would quickly become intractable. Instead we construct a sufficiently general model of algal growth, which includes a constant birth rate and a density-dependent death rate. Hence, the density-dependence of algal death could be caused by nutrient depletion and contamination and degradation of the environment by the algal bloom or by an increased grazing-pressure as the bloom gets denser and grazer population numbers respond in kind. The equations are

$$\begin{aligned} \frac{dx}{dt} &= x(b(1 - c[k + \delta k]) - (d + \delta d)(x + y - n[k + \delta k]x - n[k]y)) \\ \frac{dy}{dt} &= y(b(1 - c[k]) - d(x + y - n[k + \delta k]x - n[k]y)) \end{aligned} \quad (29)$$

where b and d are the per capita birth and death rates respectively, $c[k]$ is the physiological cost of producing toxins as a function of toxin investment, k, and $n[k]$ is the benefit derived from toxin production in terms of an alleviation of the density-dependent death rate and is multiplied by the density of the algae investing in the toxin, so that there is in effect an indiscriminate 'pooling' together of resources (Thornton, 2002b). Cell densities are scaled relative to the carrying capacity.

There is an asymmetry in the death rates between the algal lineages and the model assumes that toxins can be excreted as exotoxins. To solve these equations we must choose specific functions for $c[k]$ and $n[k]$. It seems reasonable to assume that for small investments the physiological cost will be lower than the 'pooled' benefit, but that for higher investments the immediate cost outweighs any potential benefit. To make this mathematically tractable we will use $c[k] = k^2$ and $n[k] = k$. The equation for z is

$$\frac{dz}{dt} = z(z(d + p\delta d)(k + p\delta k - 1) - b(k^2 + 2kp\delta k + p\delta k^2 - 1)). \tag{30}$$

Substituting in $p = p_0 + O(\delta k)$ and solving for z_t we find

$$z_t = \hat{z}_t - \frac{\delta d}{b(1+k)} p_0 (e^{bt(k^2-1}} - 1)\hat{z}_t^2 + \frac{\delta k}{b(1+k)^2 z_0} e^{-bt} p_0 A + O(\delta k^2) \tag{31}$$

where

$$\hat{z}_t = \frac{be^{bt} z_0 (1+k)}{be^{bk^2 t}(1+k) + dz_0(e^{bt} - e^{bk^2 t})}$$

$$A = de^{bt} z_0 - e^{bk^2 t}(dz_0 + 2bk(1+k)t(b + bk - dz_0)). \tag{32}$$

Solving for p_t as before we find

$$p_t = p_0 + \frac{\delta d}{d} p_0 (1 - p_0)(bk^2 t + \ln(b(1+k)) - \ln(be^{bk^2 t}(1+k) + d(e^{bt} - e^{bk^2 t}) z_0)$$
$$-2bk(1 - p_0) p_0 t \delta k + O(\delta k^2). \tag{33}$$

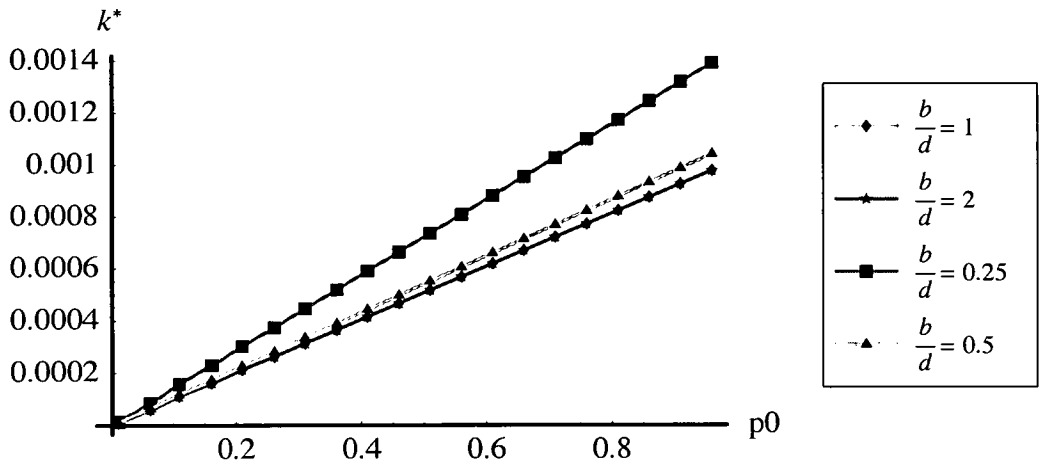


Figure 7.2. Numerical solutions for the ESS level of investment in toxins, k^*, as a function of kinship, p_0, for a range of birth-death rate ratios. The lines for ratios 1 and 2 are almost equal and are indistinguishable. $z_0 = 0.5$, $t = 10$, $\delta d = 0.001$.

Assuming entirely local competition we can ask when $\frac{dW}{d\delta k}|_{k=0} > 0$ and find this is true when $\delta d > 0$ and $p_0 > 0$. There must be an asymmetry in death rates and the focal lineage must have a higher death rate than its competitor. Investment in toxins is an increasing function of the lineage's kinship (Figure 7.2), and is maximised by a low birth rate to death rate ratio. Having a higher death rate means that the relative benefit of investment in toxin production is higher for the X lineage as the lineage stands to lose more if it does not invest in toxins even though their competitors benefit from their toxin. One could imagine a positive feedback occurring whereby toxin production enables the whole population to grow for longer resulting in more density-dependent death and so more toxin poduction.

The growth-reproduction trade-off is a well documented outcome of life-history strategies constrained by juvenile and adult mortality rates and is suggestive of a common asymmetry in natural populations of logistically growing microbes. This is the

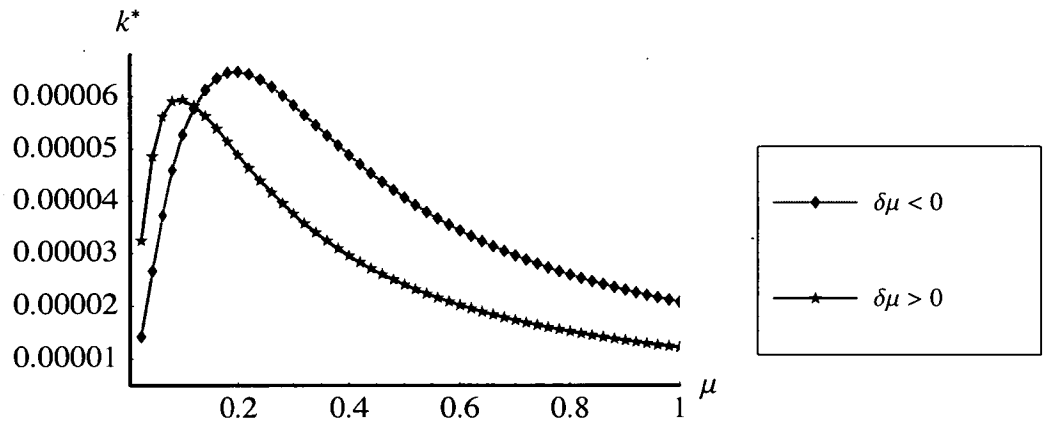


Figure 7.3. A plot showing the ESS investment in toxin production, k^*, as a function of the growth rate, μ, for two different asymmetries in $\delta\mu$. For $\delta\mu = -0.001, f = 1$ and for $\delta\mu = 0.001, f = 3$. In both cases $z_0 = 0.5, t = 10$ and $p_0 = 0.6$.

scenario we turn to now with the inclusion of a density-dependent birth rate. It is difficult to capture this asymmetry in a mathematically tractable fashion using the above model so we modify the model in the following way

$$\begin{aligned}\frac{dx}{dt} &= (\mu+\delta\mu)x(1-f(x+y-(k+\delta k)x-ky))-(k+\delta k)^2x \\ \frac{dy}{dt} &= \mu y(1-f(x+y-(k+\delta k)x-ky))-k^2y \end{aligned} \tag{34}$$

where f^{-1} is the fraction of the total density-limited population size at which overcrowding and environmental degradation and/or grazing pressure results in negative growth rates. Hence, the X-lineage has a larger growth rate prior to the density-induced switch to negative growth after which it has a larger negative growth rate, thereby capturing the growth-reproduction (r-K) trade-off.

Solving as before we once again ask when $\frac{dW}{d\delta k}|_{k=0} > 0$ and find this is true when $p_0 > 0$, $\delta\mu > 0$ and $f > \frac{1}{z_0}$ or when $p_0 > 0$, $\delta\mu < 0$ and $f < \frac{1}{z_0}$. Hence, both sides of the trade-off can invest in toxin production. Slower growers benefit more from toxin

production when the population is still growing and vice versa for faster growers. However, there may be an inherent pressure favouring toxin production in denser populations where the pooling together of resources is more effective and the function for the benefit of toxin production in this model does not properly take this into account.

Figure 7.3 shows some numerical solutions for the ESS toxin investment, k^* as a function of μ for both sides of the r-K trade-off. Solutions were found by setting $\frac{dW}{d\delta k}|_{\delta k=0,k=k^*} = 0$ and solving for k^*. The second derivative is always negative so long as $0 < p_0 < 1$. Toxin production is favoured at lower growth rates for the lineage with $\delta\mu > 0$ reflecting an increased capacity to invest when growth rates are not too negative, but both lineages show highest investment for lower growth rates probably because $\delta\mu$ comprises a larger fraction of their growth making the costs of not investing larger. The lineage with $\delta\mu < 0$ invests more overall and this reflects the reduced relative cost of toxin investment when the population is actively growing (i.e. for this case $f < \frac{1}{z_0}$). These solutions are of the order $\delta\mu$.

So far we have assumed entirely local competition, but it is of interest to ask how a non-zero global competition affects toxin production. It is difficult to derive solvable equations in the general case, but it is possible to get an answer for the special case of $a = 0$ and it is found that toxin production will always be favoured simply when $p_0 > 0$ in the case of entirely global competition. This makes sense as in this model a global competitor is one who cannot benefit from your toxin production, but who you compete with for other global resources, meaning that the toxin producing lineage is the only lineage to benefit from its own toxin. Therefore, any non-zero global competition will enhance the evolution of toxin producing variants.

7.3 Discussion

Evolutionary models of two important natural phenomena associated with phytoplankton have been developed and analysed and in both cases it has been shown that asymmetries in life-history parameters, such as growth rates, enable the evolution of traits which are costly to the individual by biasing the relative benefit of the trait in favour of the actor.

In the case of marine snow it was shown that an ability to kin discriminate, expressible through β_x, coupled with higher growth rates enable sticky variants to compensate for the costs of being sticky. These results corroborate the numerical analysis of Hansen et al. (1995) that suggested that small cell size (low β_x) and fast growth rates enable sticky lineages to outcompete non-sticky competitors and dominate the water column. The benefit of being sticky was also shown to increase as the carrying capacity of the environment increases when there are asymmetries in growth rates, consistent with the results of Hansen et al. (1995) suggesting that eutrophic environments will favour more investment in stickiness. Additionally, it has been shown that the evolution of stickiness involves a negative relatedness and is an increasing function of the scale of competition, indicating that it may be considered both a spiteful trait as well as an interference-competition adaptation (Hamilton, 1970; Gardner et al., 2004; Gardner and West, 2004).

Kiorboe et al. (1994) investigated marine snow in a mixed diatom population during a spring diatom bloom in a shallow Danish fjord. They found that there was aggregation occurring whilst the environment was not nutrient limiting, providing support for aggregation as a competitive trait during active diatom growth. Tallberg and Heiskanen (1998) found that the flux of settling phytoplankton in a Finnish bay was highest during the spring. In a laboratory test of diatom aggregate formation, Kiorboe et al. (1990) investigated the potential effects of nutrient limitation and growth status on the

stickiness of three different diatom species *in vitro*. One species appeared to increase in stickiness as the nutrients became depleted, a pattern that is consistent with the proposal of Smetacek (1985). Another species, however, was sticky even in nutrient replete conditions and, interestingly, appeared to become maximally sticky in the transition between exponential and stationary growth. In another study Kiorboe and Hansen (1993) examined the stickiness of diatom species in batch cultures and found for one species that stickiness decreased with the age of the batch culture, but did not find any pattern with nutrient or bacterial concentrations.

It seems likely that the stickiness of diatoms may serve more than one function in the diatom life-cycle and life-history strategies, and more complex models may be required to ascertain the relative contribution of different putative functions during the growth-cycle. The situation is more complicated than just the sticking of cells that have physically collided. Diatom species are also known to excrete transparent exopolymer particles (TEP), a complex carbohydrate mucilage, which can enhance aggregate formation (Kiorboe and Hansen, 1993; Corzo et al., 2000; Engel, 2000; Passow, 2002). Malej and Harris (1993) have suggested that the production of TEP is an adaptation to inhibit copepod grazing in a 'pooling' of resources manner. The relationship between TEP production and cell surface stickiness is likely to be complex, but warrants further investigation.

In the case of algal toxins it was shown that very slight asymmetries in growth rates or death rates can enable the evolution of toxin production by tipping the benefit of producing toxin in favour of the actor as they stand to lose more by not investing. The results suggested that both r-selected and K-selected algal types may benefit from toxin production during periods of population decline and exponential growth respectively. The validity of these results, however, may rest on the choice of function describing the

benefit accrued from toxin production and on the assumption that toxin is excreted as an exotoxin. Studies suggest that the toxin accumulates in the food chain (Yasumoto, 2005). In natural populations the pooling together of toxin resources may be most effective when the population is dense and close to the environmental carrying capacity. Also, if the toxin is only effective upon cell death and rupture or after consumption by a grazer then toxin production would be favoured when death rates are inflated by density-dependent effects or when grazer populations are increased in response to algal density.

Empirical studies to date have suggested that toxin production is increased by nutrient limitation and is inversely related to the algal growth rate (Proctor et al., 1975; Hall, 1982; Boyer et al., 1987; Taroncher-Oldenburg et al., 1999). It has been proposed that toxins are used as an anti-predator defence mechanism because, in the majority of toxic algae, in partcular the dinoflagellates, there is a limited nutrient affinity, which results in small growth rates making the algae vulnerable to grazing (Smayda, 1997a). In the present study toxin production appeared to be greater when growth rates were small, although this is attributed to the asymmetry forming a relatively larger component of the lineage's fitness, and in turn benefiting toxin production more.

From this analysis it is proposed that in the case of entirely local competition toxin production is viewed as a competitive strategy based upon a lineage's particular life-history circumstances as opposed to a cooperative group behaviour. There is an altruistic element to toxin production though, as producing the toxin is costly and is an increasing function of a lineage's kinship in the social arena. The production of toxins, however, does have a group effect, prolonging the bloom length and feeding back positively to induce further toxin production. These are perhaps the only times when toxin production comes to the attention of humans, and so it is of interest to know how non-blooming

algae regulate their toxin production.

It is suggested that these results may have some relevance for other 'public goods games' such as siderophore production in parasitic bacteria (West and Buckling, 2003; Griffin et al., 2004). It was shown by West and Buckling (2003) and Griffin et al. (2004) that siderophore production requires non-zero global competition if competitor growth rates are symmetrical. From this study it can be seen that small asymmetries between competitors can enable toxin production even when competition is entirely local, and will also increase toxin production when there is non-zero global and local competition.

For both marine snow and algal toxin production it is likely that facultative behaviours will be adaptive since in both cases the optimal allocation of resources into the behaviours is dependent on the density of the population (and their clone mates). Indeed it has been shown that algae can both facultatively alter their cell stickiness (Kiorboe and Hansen, 1993; Jackson, 1995) and facultatively invest in toxin production (Proctor et al., 1975; Hall, 1982; Boyer et al., 1987; Taroncher-Oldenburg et al., 1999). For this reason it is proposed that an understanding of density-dependent processes in phyoplankton may yield further insight into the important behaviours that they exhibit.

8 Summary

8.1 Experimental work

In Chapter 3 we described the selection of *C. reinhardtii* lines to divergent environments (pH 6 and pH 8) and detailed the subsequent measures of RL from crosses between the evolved lines. The measures of RL were significantly positive for lines selected in pH 6 and in pH 8, indicating the existence of positive epistasis between loci in the parents. In contrast there was no epistasis found for lines selected in the intermediate environment, pH 6.6. Together with evidence of adaptation for the lines selected at pH 6 and pH 8, but not at pH 6.6, this suggests that groups of co-adapted alleles were selected, and that recombination, in the short-term at least, will act to retard this epistatic selection.

It is desirable to trace the evolution of this RL through time so that the role of standing variation versus *de novo* mutation may be compared. Further to this, it is of great interest to directly compare the dynamics of asexual adaptation with that of populations undergoing regular bouts of recombination during selection, and to then compare the subsequent RLs between populations. Such a comparison may shed light on whether recombination can select for alleles with large average additive effects so that they are good across a range of backgrounds. It will also allow for a comparison of the short-term and long-term effects of recombination during adaptation, which may be expected to switch from negative to positive as selection progresses.

The experiment also showed that the extent of the RL between populations is not influenced by the difference between their selective environments. This result indicates that the build-up of incompatibilities between isolated populations is not necessarily enhanced by differences in selective environments and supports previous work that highlights that different populations in uniform environments will fix different mutations

and hence adaptations (Cohan and Hoffmann, 1986; Hoffmann and Cohan, 1987; Cohan et al., 1989).

8.2 Theory

In Chapter 4, a two-locus model of the evolution of genetic robustness was developed and analysed. The results showed that recombination facilitates the evolution of a costly robustness modifier that decreases the selective detriment of a deleterious allele. The effect of recombination is to decouple the long-term cost of the modifier's association with the deleterious allele from its short-term benefit of suffering a weaker deleterious effect in the context of the deleterious allele. It was suggested that synergistic epistasis might be the evolutionary outcome of sex and recombination since recombination facilitates the evolution of robustness, and synergistic epistasis is a form of genetic robustness. Therefore, it is of interest to develop models that allow for modifiers of epistasis between loci with deleterious mutations.

Chapter 5 discusses the potential role of epistatic selection for generating patterns of gene clustering within chromosomes. It was concluded that in general highly conserved genes that are constitutively expressed will tend to be clustered together into transcriptionally active chromosomal domains and that epistatic clustering will result in clustering patterns only for special cases, the sex chromosomes being the prime example.

Chapters 6 and 7 involve more general non-linear models. In Chapter 6, a sexual selection model is analysed where the development of conspicuous male traits are influenced by their maternal investment. In this way females can choose males that indicate the investment levels of their mothers, which is genetically determined. The model shows that maternal effects may be important components of mate choice. In Chapter

7, models of marine snow and algal toxin production are analysed. Both of the models show that small asymmetries in life history traits can bias the relative benefit of the trait for the actor and enable its evolution.

References

Acampora, D., D'Esposito, M., Faiella, A., Pannese, M., Migliaccio, E., Morelli, F., Stornaiuolo, A., Nigro, V., Simeone, A. and Boncinelli, E., 1989. The human Hox gene family. *Nucleic Acids Res* 17, 10385–10402.

Alldredge, A. L. and Silver, M. W., 1988. Characteristics, dynamics and significance of marine snow. *Progress in Oceanography* 20, 41–82.

Allen, J. A. and Clarke, B., 1968. Evidence for apostatic selection by wild passerines. *Nature* 220, 501.

Ancel, L. W. and Fontana, W., 2000. Plasticity, evolvability and modularity in RNA. *J Exp Zool* 288, 242–283.

Andersson, M., 1986. Evolution of condition-dependent sex ornaments and mating preferences: sexual selection based on viability differences. *Evolution* 40, 804–816.

Andersson, M., 1994. *Sexual Selection*. Princeton Univ Press.

Armes, N. and Fried, M., 1996. Surfeit locus homologs are widely distributed in invertebrate genomes. *Mol Cell Biol* 16, 5591–5596.

Awadalla, P. and Charlesworth, D., 1999. Recombination and selection at brassica self-incompatibility loci. *Genetics* 152, 413–425.

Baden, D. G., 1989. Brevetoxins: unique polyether dinoflagellate toxins. *The FASEB J* 3, 1807–1817.

Bagheri, H. and Wagner, G. P., 2004. Evolution of dominance in metabolic pathways. *Genetics* 168, 1713–1735.

Baker, H. G., 1965. *The genetics of colonizing species*, Academic Press, inc., chap. Characteristics and modes of origin of weeds. pp. 147–168.

Barton, N. H., 1995. A general model for the evolution of recombination. *Genet Res* 65, 123–144.

Barton, N. H. and Charlesworth, B., 1984. Genetic revolution, founder effects, and speciation. *Ann Rev Ecol Syst* 15, 133–164.

Barton, N. H. and Turelli, M., 1991. Natural and sexual selection on many loci. *Genetics* 127, 229–255.

Bateson, W., 1909. *Mendel's Principles of Heredity*. Cambridge Univ Press.

Becquevort, S. and Smith, W. O., 2001. Aggregation, sedimentation and biodegradability of phytoplankton-derived material during spring in the Ross Sea, Antarctica. *Deep Sea Res* 48, 4155–4178.

Behre, K. and Schwabe, G. H., 1970. Auf Surtsey-Island in sommer 1968 nachgewiesene nicht marine algen. *Schr Naturw Ver Schleswig-Holstein* , 31–100.

Bell, G., 1990a. The ecology and genetics of fitness in *Chlamydomonas*. I. Genotype-by-environment interactions among pure strains. *Proc Roy Soc Lon B* 240, 295–321.

Bell, G., 1990b. The ecology and genetics of fitness in *Chlamydomonas*. II. The properties of mixtures of strains. *Proc Roy Soc Lon B* 240, 323–350.

Bell, G., 1991a. The ecology and genetics of fitness in *Chlamydomonas*. II. Genotype-by-environment interaction within strains. *Evolution* 45, 668–679.

Bell, G., 1991b. The ecology and genetics of fitness in *Chlamydomonas*. IV. The properties of mixtures of gentoypes of the same species. *Evolution* 45, 1036–1046.

Bell, G., 2005. Experimental sexual selection in *Chlamydomonas*. *J Evol Biol* 18, 722–734.

Bell, G. A. C., 1997. Experimental evolution in *Chlamydomonas*. I. Short-term selection in uniform and diverse environments. *Heredity* 78, 490–497.

Bjorklund, D. F., 2006. Mother knows best: epigenetic inheritance, maternal effects and the evolution of human intelligence. *Dev Rev* 26, 213–242.

Black, J. G., 1996. *Microbiology. Principles and applications*. Prentice-Hall, inc., third edn.

Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W. L., Duke, K., Kiraly, M. and Kim, S. K., 2002. A global analysis of *Caenorhabditis elegans* operons. *Nature* 20, 797–798.

Bold, H. C., 1942. The cultivation of algae. *Bot Rev* 8, 69–138.

Bold, H. C. and Wynne, M. J., 1985. *Introduction to the algae*. Prentice-Hall, inc., second edn.

Bordenstein, S. R. and Drapeau, M. D., 2001. Genotype-by-environment interaction and the Dobzhansky-Muller model of postzygotic isolation. *J Evol Biol* 14, 490–501.

Bortoluzzi, S., Rampoldi, L., Simionati, B., Zimbello, R., Barbon, A., d'Alessi, F., Tiso, N., Pallavicini, A., Toppo, S., Cannata, N., Valle, G., Lanfranchi, G. and Danieli, G. A., 1998. A comprehensive, high-resolution genomic transcript map of human skeletal muscle. *Genome Res* 8, 817–825.

Bourque, G., Zdobnov, E. M., Bork, P., Pevzner, P. A. and Tesler, G., 2005. Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. *Genome Res* 15, 98–110.

Boutanaev, A. M., Kalmykova, A. I., Shevelyov, Y. Y. and Nurminsky, D. I., 2002. Large clusters of co-expressed genes in the *Drosophila* genome. *Nature* 420, 666–669.

Boyer, G. L., Sullivan, J. J., Anderson, R. J., Harrison, P. J. and Taylor, F. J. R., 1987. Effects of nutrient limitation on toxin production and composition in the marine dinoflagellate *Protogonyaulax tamarensis*. *Mar Biol* 96, 123–128.

Boyes, D. C., Nasrallah, M. E., Vrebalov, J. and Nasrallah, J. B., 1997. The self-incompatibility (S) haplotypes of brassica contain highly divergent and rearranged sequences of ancient origin. *Plant Cell* 9, 237–247.

Bristol, B. M., 1920. On the alga-flora of some dessicated English soils: an important factor in soil biology. *Ann Botany* 34, 35–80.

Bristol-Roach, B. M., 1928. The present position of our knowledge of the distribution and function of algae in the soil. *Proc 1st Inter Cong Soil Sci* 30, 30–38.

Brooks, L. D., 1988. *The evolution of sex*, Sinauer Press, chap. The evolution of recombination rates. pp. 87–105.

Brown, R. M. J., Larson, D. A. and Bold, H. C., 1964. Airborne algae; their abundance and heterogeneity. *Science* 143, 583–585.

Burch, C. L. and Chao, L., 2004. Epistasis and its relationship to canalisation in the rna virus phi-6. *Genetics* 167, 559–567.

Buss, L. W. and Grosberg, R. K., 1990. Morphogenetic basis for phenotypic differences in hydroid competitive behaviour. *Nature* 343, 63–66.

Butcher, A. C., 1969. Non-allelic interactions and genetic isolation in wild populations of *Aspergillus nidulans*. *Heredity* 24, 621–631.

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M., van Asperen, R., Boon, K., Voute, P. A., Heisterkamp, S., van Kampen A. and Versteeg, R., 2001. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* 291, 1289–1292.

Cavalier-Smith, T., 1981. *Molecular and Cellular aspects of microbial evolution*, Cambridge Univ Press, chap. The origin and early evolution of the eukaryote cell. pp. 33–84.

Cavalier-Smith, T., 1993. *The eukaryotic genome: Organisation and regulation*, Cambridge Univ Press, chap. Evolution of the eukaryotic genome. pp. 333–385.

Cavalli, L. and Maccacaro, G., 1952. Polygenic inheritance of drug resistance in the bacterium *E. coli*. *Heredity* 6, 311–331.

Chandler, M. G. and Pritchard, R. H., 1975. The effect of gene concentration and relative gene dosage on gene output in *Escherichia coli*. *Mol Gen Genet* 138, 127–141.

Charlesworth, B., 1974. Inversion polymorphism in a two-locus genetic system. *Genet Res* 23, 259–280.

Charlesworth, B., 1990. Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet Res* 55, 199–221.

Charlesworth, B., 1993. Directional selection and the evolution of sex and recombination. *Genet Res* 61, 205–224.

Charlesworth, B. and Barton, N. H., 1996. Recombination load associated with selection for increased recombination. *Genet Res* 67, 27–41.

Charlesworth, B. and Charlesworth, D., 1973. Selection of new inversions in multi-locus genetic systems. *Genet Res* 21, 167–183.

Charlesworth, B. and Charlesworth, D., 1975. An experiment on recombination load in *Drosophila melanogaster*. *Genetical Res* 25, 267–274.

Charlesworth, B., Coyne, J. A. and Barton, N. H., 1987. The relative rates of evolution of sex chromosomes and autosomes. *Am Nat* 130, 113–146.

Charlesworth, D., Bartolome, C., Schierup, M. H. and Mable, B. K., 2003. Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. *Mol Biol Evol* 20, 1741–1753.

Charlson, R. J., Lovelock, J. E., Andreae, M. O. and Warren, S. G., 1987. Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature* 326, 655–661.

Charnov, E. L. and Schaffer, W. M., 1973. Life-history consequences of natural selection: Cole's result revisited. *Am Nat* 107, 791–793.

Chen, F. C. and Li, W. H., 2001. Genomic divergences between humans and other hominids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet* 68, 444–456.

Cheverud, J. M. and Routman, E. J., 1995. Epistasis and its contribution to genetic variance components. *Genetics* 139, 1455–1461.

Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J. and Davis, R. W., 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell* 2, 65–73.

Christiansen, F. B., 1991. On conditions for evolutionary stability for a continuously varying character. *Am Nat* 138, 37–50.

Clarke, B., 1969. Evidence for apostatic selection. *Heredity* 240, 347.

Cockerham, C. C., 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* 39, 859–882.

Cody, M. L., 1966. A general theory of clutch size. *Evolution* 20, 174–184.

Cohan, F. M. and Hoffmann, A. A., 1986. Genetic divergence under uniform selection. II. Different responses to selection for knockdown resistance to ethanol among *Drosophila melanogaster* populations and their replicate lines. *Genetics* 114, 145–163.

Cohan, F. M. and Hoffmann, A. A., 1989. Uniform selection as a diversifying force in evolution: evidence from *Drosophila*. *Am Nat* 134, 613–637.

Cohan, F. M., Hoffmann, A. A. and Gayley, T. W., 1989. A test of the role of epistasis in divergence under uniform selection. *Evolution* 43, 766–774.

Colegrave, N., 2002. Sex releases the speed limit on evolution. *Nature* 420, 664–666.

Colegrave, N., Kaltz, O. and Bell, G., 2002. The ecology and genetics of fitness in

Chlamydomonas. VIII. The dynamics of adaptation to novel environments after a single episode of sex. *Evolution* 56, 14–21.

Coleman, A. W., 1983. *Survival strategies of the algae*, Cambridge Univ. Press, chap. The roles of resting spores and akinetes in chlorophyte survival. pp. 1–21.

Collins, S. and Bell, G., 2004. Phenotypic consequences of 1,000 generations of selection at elevated CO_2 in a green alga. *Nature* 431, 566–569.

Corzo, A., Morillo, J. A. and Rodriguez, S., 2000. Production of transparent expopolymer particles (TEP) in cultures of Chaetoceros calcitrans under nitrogen limitation. *Aq Microb Ecol* 23, 63–72.

Cotton, S., Fowler, K. and Pomiankowski, A., 2004. Condition dependence of sexual ornament size and variation in the stalk-eyed fly *Cyrtodiopsis dalmanni*. *Evolution* 58, 1038–1046.

Couturier, E. and Rocha, E. P., 2006. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol Microbiol* 59, 1506–1518.

Coyne, J., Barton, N. and Turelli, M., 1997. A critique of Sewall Wright's shirfting balance theory of evolution. *Evolution* 51, 643–671.

Coyne, J., Barton, N. and Turelli, M., 2000. Is Wright's shifting balance process important in evolution? *Evolution* 54, 306–317.

Cremer, T. and Cremer, C., 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2, 292–301.

Crook, D. J., Hopper, J. A., Ramaswamy, S. B. and Higgins, R. A., 2004. Courtship behaviour of the soybean stem borer *Dectes texanus texanus* (Coleoptera: Cerambycidae): evidence for a female contact sex pheromone. *Ann Entomol Soc Amer* 97, 600–604.

Crow, J., 1956. Genetics of DDT resistance in *Drosophila*. *Proc Intern Genet Symp* , 408–409.

Crow, J. and Kimura, M., 1970. *An introduction to population genetics theory*. Burgess Publishing Company.

Crow, J. F. and Kimura, M., 1965. Evolution in sexual and asexual populations. *Am Nat* 99, 439–450.

Danforth, W. F., 1962. *Physiology and biochemistry of algae*, Academic Press, inc., chap. Substrate assimilation and heterotrophy. pp. 99–123.

Darwin, C., 1859. *On the Origin of Species by Means of Natural Selection*. Murray.

Dawson, K. J., 1999. Evolutionarily stable mutation rates. *J Theor Biol* 194, 143–157.

de Grouchy, J., Turleàu, C., Roubin, M. and Colin, F. C., 1973. *Chromosome Identification. Proc. 23rd Nobel Symp.*, Academic Press, inc., chap. Chromosomal evolution of man and primates (*Pan troglodytes, Gorilla gorilla, Pongo pygmaeus*). pp. 124–131.

de Visser, J. A. G. M., Hermisson, J., Wagner, G. P., Meyers, L. A., Bagheri-Chaichian, H., Blanchard, J., Chao, L., M., C. J., Elena, S. F., Fontana, W., Gibson, G., Hansen, T. F., Krakauer, D., Lewontin, R. C., Ofria, C., Rice, S. H., von Dassow, G., Wagner, A. and Whitlock, M. C., 2003. Evolution and detection of genetic robustness. *Evolution* 57, 1959–1972.

de Visser, J. A. G. M. and Hoekstra, R. F., 1998. Synergistic epistasis between loci affecting fitness: evidence in plants and fungi. *Genet Res* 71, 39–49.

de Visser, J. A. G. M., Hoekstra, R. F. and van den Ende, H., 1996. The effect of sex and deleterious mutations on fitness in *Chlamydomonas*. *Proc Roy Soc Lon B* 263, 193–200.

de Visser, J. A. G. M., Hoekstra, R. F. and van den Ende, H., 1997. Test of interaction between genetic markers that affect fitness in *Aspergillus niger*. *Evolution* 51, 1499–1505.

Dempsey, A. A., Pabalan, N., Tang, H. and Liew, C.-C., 2001. Organization of human cardiovascular-expressed genes on chromosomes 21 and 22. *J Mol Cell Cardiol* 33, 587–591.

Dobzhansky, T., 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* 21, 113–135.

Dobzhansky, T., 1970. *Genetics of the evolutionary process*. Columbia Univ Press.

Dobzhansky, T., Levene, H., Spassky, B. and Spassky, N., 1959. Release of genetic variability through recombination. III. *Drosophila prosaltans*. *Genetics* 44, 75–92.

Duhig, T., Ruhrberg, C., More, O. and Fried, M., 1998. The human surfeit locus. *Genomics* 52, 72–78.

Eberharter, A. and Becker, P. B., 2002. Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep* 3, 224–229.

Edelman, G. M. and Gally, J. A., 2001. Degeneracy and complexity in biological systems. *Proc Natl Acad Sci USA* 98, 13763–13768.

Edwards, J. S. and Palsson, B. O., 1999. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J Biol Chem* 274, 17410–17416.

Edwards, J. S. and Palsson, B. O., 2000a. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* 97, 5528–5533.

Edwards, J. S. and Palsson, B. O., 2000b. Robustness analysis of the *E. coli* metabolic network. *Biotech Progress* 16, 927–939.

Eigen, M. and Schuster, P., 1979. *The Hypercycle: A principle of Natural Self-Organisation.* Springer-Verlag.

Elena, S. F., 1999. Little evidence for synergism among deleterious mutations in a non-segmented RNA virus. *J Mol Evol* 49, 703–707.

Elena, S. F. and Lenski, R. E., 1997. Test of interaction between deleterious mutations in *E. coli*. *Nature* 390, 395–398.

Emlen, S. T. and Oring, L. W., 1977. Ecology, sexual selection, and the evolution of mating systems. *Science* 197, 215–223.

Engel, A., 2000. The role of transparent exopolymer particles (TEP) in the increase in apparent particle stickiness (α) during the decline of a diatom bloom. *J Plankton Res* 22, 485–497.

Eshel, I., 1983. Evolutionary and continuous stability. *J Theor Biol* 103, 99–111.

Eshel, I. and Motro, U., 1981. Kin selection and strong evolutionary stability of mutual help. *Theor Pop Biol* 19, 420–433.

Ettl, H., 1976. Die gattung *Chlamydomonas* Ehrenberg. *Beih Nova Hedwigia* 49, 1–1122.

Eversole, R. A., 1956. Biochemical mutants of *Chlamydomonas reinhardi*. *Am J of Botany* 43, 404–407.

Fares, M. A., Ruiz-Gonzalez, M. X., Moya, A., Elena, S. F. and Barrio, E., 2002. GroEL buffers against deleterious mutations. *Nature* 417, 398.

Feder, J. L., Roethele, J. B., Filchak, K., Niedbalski, J. and Romero-Sverson, J., 2003. Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* 163, 939–953.

Feldman, M. W., Christiansen, F. B. and Brooks, L. D., 1980. Evolution of recombination in a constant environment. *Proc Natl Acad Sci USA* 77, 4838–4841.

Ferris, P. J., Armbrust, E. V. and Goodenough, U. W., 2002. Genetic structure of the mating type locus of *Chlamydomonas reinhardtii*. *Genetics* 160, 181–200.

Ferris, P. J. and Goodenough, U. W., 1994. The mating type locus of *Chlamydomonas reinhardtii* contains highly rearranged DNA sequences. *Cell* 76, 1135–1145.

Ferris, P. J. and Goodenough, U. W., 1997. Mating type in *Chlamydomonas* is specified by *mid*, the minus-dominance gene. *Genetics* 146, 859–869.

Ferris, P. J., Woessner, J. P. and Goodenough, U. W., 1996. A sex recognition glycoprotein is encoded by the plus mating type gene *fus1* of *Chlamydomonas reinhardtii*. *Mol Biol Cell* 7, 1235–1248.

Field, C. B., Behrenfeld, M. J., Randerson, J. T. and Falkowski, P., 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237–240.

Fischer, G., Rocha, E. P. C., Brunet, F., Vergassola, M. and Dujon, B., 2006. Highly variable rates of genome rearrangements between hemiascomycetous yeast lineages. *PLoS* 2, 0253–0261.

Fisher, R. A., 1915. The evolution of sexual preference. *Eugen Rev* 7, 184–192.

Fisher, R. A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans Royal Soc Edinburgh* 52, 399–433.

Fisher, R. A., 1922. On the dominance ratio. *Proc Roy Soc Edin* 42, 321–341.

Fisher, R. A., 1925. *Statistical methods for research workers*. Hafner.

Fisher, R. A., 1928. The possible modification of the response of the wild type to recurrent mutations. *Am Nat* 62, 115–126.

Fisher, R. A., 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Fogg, G. E., 1975. *Algal cultures and phytoplankton ecology*. Wisconsin Univ Press, second edn.

Frank, S. A., 2003. Genetic variation of polygenic characters and the evolution of genetic degeneracy. *J Evol Biol* 16, 138–142.

Franklin, I. and Lewontin, R. C., 1970. Is the gene the unit of selection? *Genetics* 65, 707–734.

Frey, K. J. and Maldonado, U., 1967. Relative productivity of homogeneous and heterogeneous oat cultivars in optimum and suboptimum environments. *Crop Sci* 7, 532–535.

Gabrielsson, B. L., Carlsson, B. and Carlsson, L. M. S., 2000. Partial genome scale analysis of gene expression in human adipose tissue using DNA array. *Obes Res* 8, 374–384.

Gadgil, M. and Solbrig, O. T., 1972. The concept of r- and K-selection: evidence fromwild flowers and some theoretical considerations. *Am Nat* 106, 14–31.

Gardner, A. and West, S. A., 2004. Spite and the scale of competition. *J Evol Biol* 17, 1195–1203.

Gardner, A., West, S. A. and Buckling, A., 2004. Bacteriocins, spite and virulence. *Proc Roy Soc Lon B* 271, 1529–1535.

Garson, K., Duhig, T., Armes, N., Colombo, P. and Fried, M., 1995. A gene in the tightly clustered mouse surfeit locus is highly conserved and transcribed divergently from the rpL7A (surf3) gene. *Genomics* 30, 163–170.

Gavrilets, S., 1997. Evolution and speciation on holey adaptive landscapes. *TREE* 12, 307–312.

Gessler, D. D. and Xu, S., 1999. On the evolution of recombination and meiosis. *Genet Res* 73, 119–131.

Gillham, N. W., 1969. Uniparental inheritance in *Chlamydomonas reinhardtii*. *Am Nat* 103, 355–388.

Ginzburg, L. R. and Taneyhill, D. E., 1994. Population-cycles of forest lepidoptera - a maternal effect hypothesis. *J Anim Ecol* 63, 79–92.

Goho, S. and Bell, G., 2000. The ecology and genetics of fitness in *Chlamydomonas*. IX. The rate of accumulation of variation of fitness under selection. *Evolution* 54, 416–424.

Goodnight, C. J., 1987. On the effect of founder events on the additive genetic variance. *Evolution* 41, 171–179.

Goodnight, C. J., 1988. Epistasis and the effect of founder events on the additive genetic variance. *Evolution* 42, 441–454.

Goodnight, C. J., 1995. Epistasis and the increase in additive genetic variance: implications for phase 1 of Wright's shifting-balance process. *Evolution* 49, 502–511.

Goodnight, C. J. and Wade, M. J., 2000. The ongoing synthesis: a reply to coyne, barton and turelli. *Evolution* 54, 317–324.

Grafen, A., 1990a. Sexual selection unhandicapped by the Fisher process. *J Theor Biol* 144, 473–516.

Grafen, A., 1990b. Biological signals as handicaps. *J Theor Biol* 144, 517–546.

Griffin, A. S., West, S. A. and Buckling, A., 2004. Cooperation and competition in pathogenic bacteria. *Nature* 430, 1024–1027.

Haldane, J. B. S., 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. *J Genet* 8, 299–309.

Haldane, J. B. S., 1931. A mathematical theory of natural selection. Part VIII. Metastable population. *Proc Camb Phil Soc* 27, 137–142.

Haldane, J. B. S., 1937. The effect of variation on fitness. *Am Nat* 71, 337–349.

Haldane, J. B. S., 1956. The theory of selection for malanism in *Lepidoptera. Proc Roy Soc Lon B* 145, 303–306.

Hall, S., 1982. *Toxins and toxicity of* Protogonyaulax *from the northeast Pacific*. Ph.D. thesis, University of Alaska.

Hamilton, W. D., 1964. The genetical evolution of social behaviour. I. *J Theor Biol* 7, 1–16.

Hamilton, W. D., 1970. Selfish and spiteful behaviour in an evolutionary model. *Nature* 228, 1218–1220.

Hamilton, W. D. and Lenton, T. M., 1998. Spora and Gaia: how microbes fly with their clouds. *Ethol Ecol Evol* 10, 1–16.

Hamilton, W. D. and Zuk, M., 1982. Heritable true fitness and bright birds: a role for parasites? *Science* 218, 384–387.

Hanks, L. M., 1999. Influence of the larval host plant on reproductive strategies of cerambycis beetles. *Ann Rev Entomol* 44, 483–505.

Hansen, J. L. S., Timm, U. and Kiorboe, T., 1995. Adaptive significance of phytoplankton stickiness with emphasis on the diatom *Skeletonema costatum. Mar Biol* 123, 667–676.

Harris, E. H., 1989. *The* Chlamydomonas *sourcebook: a comprehensive guide to biology and laboratory use*. Academic Press, inc.

Hayes, B. J., Visscher, P. M., McPartlan, H. C. and Goddard, M. E., 2003. Novel multilocus measure of linkage disequilibrium to esimate past effective population size. *Genome Res* 13, 635–643.

Hebets, E. A., 2003. Subadult experience influences adult mate choice in an arthropod: exposed female wolf spiders prefer males of a familiar phenotype. *Proc Natl Acad Sci USA* 100, 13390–13395.

Heisler, I. L., 1984. A quantitative genetic model for the origin of mating preferences. *Evolution* 38, 1283–1295.

Hermisson, J., Redner, O. and Wagner, H., 2002. Mutation-selection balance: ancestry, load and maximum principle. *Theor Pop Biol* 62, 9–46.

Hoffmann, A. A. and Cohan, F. M., 1987. Genetic divergence under uniform selection. III. Selection for knockdown resistance to ethanol among *Drosophila pseudoobscura* populations and their replicate lines. *Heredity* 58, 425–434.

Huang, S. L., Singh, M. and Kojima, K., 1971. A study of frequency-dependent selection observed in the esterase-6 locus of *Drosophila melanogaster* using a conditioned media method. *Genetics* 68, 97–104.

Hudson, R. R., 1992. Gene trees, species trees and the segregation of ancestral alleles. *Genetics* 131, 509–512.

Hughes, A. L. and Hughes, M. K., 1985. Female choice of mates in a polygynous insect, the whitespotted sawyer *Monochamus scutellatus*. *Behav Ecol Sociobiol* 17, 385–387.

Hughes, A. L. and Yeager, M., 1998. Natural selection at major histocompatibility complex loci of vertebrates. *Ann Rev Genet* 32, 415–435.

Hurst, L. D., Pal, C. and Lercher, M. J., 2004. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5, 299–310.

Hutchinson, G. E., 1967. *A Treatise on limnology: an introduction to lake biology and the limnoplankton*, vol. 2. Wiley.

Huxley, C. and Fried, M., 1990. The mouse surfeit locus contains a cluster of six genes associated with four CpG-rich islands in 32 kilobases of genomic DNA. *Mol Cell Biol* 10, 605–614.

Hwang, C. J., Monk, B. C. and Goodenough, U. W., 1981. Linkage of mutations affecting minus flagellar membrane agglutinability to the mt-minus mating type locus of *Chlamydomonas reinhardtii*. *Genetics* 99, 41–48.

Ingolia, N. T., 2004. Topology and robustness in the *Drosophila* segment polarity network. *PLoS* 2, 805–815.

Jackson, G. A., 1990. A model of the formation of marine algal flocs by physical coagulation processes. *Deep Sea Res* 37, 1197–1211.

Jackson, G. A., 1995. Coagulation of marine algae. *Aquatic Chemistry* 244, 203–217.

Jackson, G. A. and Lochmann, S. E., 1992. Effect of coagulation on nutrient and light limitation of an algal bloom. *Limnol Oceanog* 37, 77–89.

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

Jacobs, P. A., 1977. *Population cytogenetics: studies in humans*, Academic Press, inc., chap. Structural rearrangements in the chromosomes of man. pp. 81–97.

Jakubiec, H., 1984. Studies on the Polish Chlamydomonadaceae. *Nova Hedwigia* 39, 269–292.

Jinks, J. L., Caten, C. E., Simchen, G. and Croft, J. H., 1966. Heterokaryon incompatibility and variation in wild populations of *Aspergillus nidulans*. *Heredity* 21, 227–239.

Kacser, H. and Burns, J. A., 1981. The molecular basis of dominance. *Genetics* 97, 639–666.

Kaltz, O. and Bell, G., 2002. The ecology and genetics of fitness in *Chlamydomonas*. XII. Repeated sexual episodes increases rates of adaptation to novel environments. *Evolution* 56, 1743–1753.

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D. P., Zipperlen, P. and Ahringer, J., 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421, 231–237.

Karlin, S. and Feldman, M., 1969. Linkage and selection: New equilibrium properties of the two locus symmetric viability model. *Proc Natl Acad Sci USA* 62, 70–74.

Kauffman, S. and Levin, S., 1987. Towards a general theory of adaptive walks on rugged landscapes. *J Theor Biol* 128, 11–45.

Kawecki, T. J., 2000. The evolution of genetic canalization under fluctuating selection. *Evolution* 54, 1–12.

Kempthorne, O., 1954. The correlations between relatives in a random mating population. *Proc Roy Soc Lon B* 143, 103–113.

Kimura, M., 1956. A model of a genetic system which leads to closer linkage by natural selection. *Evolution* 10, 278–287.

Kimura, M. and Maruyama, T., 1966. Mutational load with epistatic gene interactions in fitness. *Genetics* 54, 1337–1351.

Kiorboe, T., Andersen, K. P. and Dam, H. G., 1990. Coagulation efficiency and aggregate formation in marine phytoplankton. *Mar Biol* 107, 235–245.

Kiorboe, T. and Hansen, J. L. S., 1993. Phytoplankton aggregate formation - observations of patterns and mechanisms of cell sticking and the significance of exopolymeric material. *J Plankton Res* 15, 993–1018.

Kiorboe, T., Lundsgaard, C., Olesen, M. and Hansen, J. L. S., 1994. Aggregation and sedimentation processes during a spring phytoplankton bloom: a field experiment to test coagulation theory. *J Mar Res* 52, 297–323.

Kirk, K. L. and Gilbert, J. J., 1992. Variation in herbivore response to chemical defenses - zooplankton foraging on toxic cyanobacteria. *Ecology* 73, 2208–2217.

Kirkpatrick, M., 1982. Sexual selection and the evolution of female choice. *Evolution* 36, 1–12.

Kirkpatrick, M., 1987. Sexual selection by female choice in polygynous animals. *Ann Rev Ecol Syst* 18, 43–70.

Kirkpatrick, M. and Barton, N., 2006. Chromosome inversions, local adaptation and speciation. *Genetics* 173.

Kirkpatrick, M., Johnson, T. and Barton, N. H., 2002. General models of multilocus evolution. *Genetics* 161, 1727–1750.

Ko, M. S., Threat, T. A., Wang, X., Horton, J. H., Cui, Y., Wang, X., Pryor, E., Paris, J., Wells-Smith, J., Kitchen, J. R., Rowe, L. B., Eppig, J., Satoh, T., Brant, L., Fujiwara, H., Yotsuomoto, S. and Nakashima, H., 1998. Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome. *Hum Mol Genet* 7, 1967–1978.

Kojima, K., 1959a. Role of epistasis and overdominance in stability of equilibria with selection. *Proc Natl Acad Sci USA* 45, 984–989.

Kojima, K., 1959b. Stable equilibria for the optimum model. *Proc Natl Acad Sci USA* 45, 989–993.

Kojima, K. and Tobari, Y. N., 1969. The pattern of viability changes associated with genotype frequency at the alcohol dehydrogenase locus in a population of *Drosophila melanogaster*. *Genetics* 61, 201–209.

Kojima, K. and Yarbrough, K. M., 1967. Frequency-dependent selection at the esterase-6 locus in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 57, 645–649.

Kondrashov, A. S., 1984. Deleterious mutations as an evolutionary factor. I. The advantage of recombination. *Genet Res* 44, 199–218.

Kondrashov, A. S. and Crow, J. F., 1991. Haploidy or diploidy: which is better? *Nature* 351, 314–315.

Kumar, S., Filipski, A., Swarna, V., Walker, A. and Hedges, S. B., 2005. Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proc Natl Acad Sci USA* 102, 18842–18847.

Lande, R., 1979. Effective deme sizes during long-term evolution estimated from rates of chromosomal rearrangement. *Evolution* 33, 234–251.

Lande, R., 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution* 34, 292–307.

Lande, R., 1981. Models of speciation by sexual selection on polygenic traits. *Proc Natl Acad Sci USA* 78, 3721–3725.

Lande, R., 1984. The expected fixation rate of chromosomal inversions. *Evolution* 38, 743–752.

Landsberg, J. H., 2002. The effects of harmful algal blooms on aquatic organisms. *Rev Fisheries Sci* 10, 113–390.

Lau, K. F. and Dill, K. A., 1990. Theory for protein mutability and biogenesis. *Proc Natl Acad Sci USA* 87, 638–642.

Lee, J. M. and Sonnhammer, E. L. L., 2003. Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res* 13, 875–882.

Lennard, A. C. and Fried, M., 1991. The bidirectional promoter of the divergently transcribed mouse surf-1 and surf-2 genes. *Mol Cell Biol* 11, 1281–1294.

Lenski, R. E., 1988. Experimental studies of pleiotropy and epistasis in *Escherichia coli*. I. Variation in competitive fitness among mutants resistant to virus T4. *Evolution* 42, 425–433.

Lenski, R. E., Rose, M. R., Simpson, S. C. and Tadler, S. C., 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2000 generations. *Am Nat* 138, 1315–1341.

Lercher, M. J., Blumenthal, T. and Hurst, L. D., 2003. Co-expression of neighbouring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. *Genome Res* 13, 238–243.

Lercher, M. J., Urrutia, A. O. and Hurst, L. D., 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature Genetics* 31, 180–183.

Lewin, J. C., 1963. *Symposium on marine microbiology*, Springfield, chap. Heterotrophy in marine diatoms. pp. 229–235.

Lewis, L. A. and McCourt, R. M., 2004. Green algae and the origin of land plants. *Am J of Botany* 91, 1535–1556.

Lewontin, R. C., 1964a. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 49, 49–67.

Lewontin, R. C., 1964b. The interaction of selection and linkage. II. Optimum models. *Genetics* 50, 757–782.

Lewontin, R. C., 1971. The effect of genetic linkage on the mean fitness of a population. *Proc Natl Acad Sci USA* 68, 984–986.

Lewontin, R. C. and Kojima, K., 1960. The evolutionary dynamics of complex polymorphisms. *Evolution* 14, 458–472.

Li, L. J., Z.and Luo, Mei, H. W., Wang, D. L., Shu, Q. Y., Tabien, R., Zhong, D. B., Ying, C. S., Stansel, G. S., J. W.and Khush and Paterson, A. H., 2001. Overdominant epistatc loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics* 158, 1737–1753.

Logan, B. E. and Alldredge, A. L., 1989. Potential for increased nutrient uptake by flocculating diatoms. *Mar Biol* 101, 443–450.

Longhurst, A., Sathyendranath, S., Platt, T. and Caverhill, C., 1995. An estimate of global primary production in the ocean from satellite radiometer data. *J Plankton Res* 17, 1245–1271.

Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., Mandel, G. and Rosenfeld, M. G., 2002. Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* 298, 1747–1752.

Lynch, M. and Walsh, B., 1998. *Genetics and analysis of quantitative traits*. Sinauer Associates, Inc.

Lyttle, T. W., 1991. Segregation distorters. *Ann Rev Genet* 25, 511–557.

MacArthur, R. H., 1962. Some generalized theorems of natural selection. *Proc Natl Acad Sci USA* 48, 1893–1897.

MacArthur, R. H. and Wilson, E. O., 1967. *The theory of island biogeography*. Princeton Univ Press.

Maguire, J., B., 1963. The passive dispersal of small aquatic organisms and their colonisation of isolated bodies of water. *Ecol Mongr* 33, 161–185.

Malej, A. and Harris, R. P., 1993. Inhibition of copepod grazing by diatom exudates: a factor in the development of mucus aggregates. *Mar Ecol Prog Ser* 96, 33–42.

Malmberg, R. L., 1977. The evolution of epistasis and the advantage of recombination in populations of bacteriophage T4. *Genetics* 86, 607–621.

Mann, D. G., 1999. The species concept in diatoms. *Phycologia* 38, 437–495.

Mather, K., 1943. Polygenic inheritance and natural selection. *Biol Rev* 18, 32–64.

Matynard Smith, J. and Price, G., 1973. The logic of animal conflict. *Nature* 246, 15–18.

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

Maynard Smith, J., 1978. *The evolution of sex*. Cambridge Univ Press.

Maynard Smith, J., 1980. Selection for recombination in a polygenic model. *Genet Res* 33, 121–128.

Maynard Smith, J., 1982. *Evolution and the Theory of Games*. Cambridge Univ Press.

Maynard Smith, J., 1988. Selection for recombination in a polygenic model: the mechanism. *Genet Res* 51, 59–63.

Mayo, O. and Burger, R., 1997. The evolution of dominance: a theory whose time has passed? *Biol Rev* 72, 97–110.

Mayr, E., 1963. *Animal species and evolution*. Belknap.

Meir, E., von Dassow, G., Munro, E. and Odell, G. M., 2002. Robustness, flexibility and the role of lateral inhibition in the neurogenic network. *Curr Biol* 12, 778–786.

Metting, B., 1981. The systematics and ecology of soil algae. *The Botanical Review* 47, 195–312.

MHC-Consortium, 2001. Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401, 921–923.

Michalakis, Y. and Roze, D., 2004. Epistasis in RNA viruses. *Curr Biol* 306, 1492–1493.

Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, W., K.and Vergeer, Yannoutsos, N., Grosveld, F. and Fraser, P., 1996. Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* 87, 105–114.

Moczek, A. P., 2002. Allometric plasticiy in a polyphenic beetle. *Ecol Entomol* 27, 58–67.

Mousseau, T. A. and Dingle, H., 1991. Maternal effects in insect life histories. *Ann Rev Entomol* 36, 511–534.

Mousseau, T. A. and Fox, C. W., 1998a. The adaptive significance of maternal effects. *TREE* 13, 403–407.

Mousseau, T. A. and Fox, C. W., 1998b. *Maternal effects as adaptations*. Oxford Univ Press.

Mukai, T. and Yamaguchi, O., 1974. The genetic structure of natural populations of *D. melanogaster*. XI. Genetic variability in a local population. *Genetics* 76, 339–366.

Muller, H. J., 1932. Some genetic aspects of sex. *Am Nat* 8, 118–138.

Muller, H. J., 1939. Reversibility in evolution considered from the standpoint of genetics. *Biol Rev Camb Philos Soc* 14, 261–280.

Muller, H. J., 1940. *The new systematics*, Clarendon Press, chap. Bearing of the *Drosophila* work on systematics. pp. 185–268.

Murphy, W. J., Larkin, D. M., Everts-van der Wind, A., Bourque, G. and Tesler, G., 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* 309, 613–617.

Nasrallah, J. B. and Nasrallah, M. E., 1993. Pollen-stigma signalling in the sporophytic self-incompatibility response. *Plant Cell* 5, 1325–1335.

Nei, M., 1967. Modification of linkage intensity by natural selection. *Genetics* 57, 625–641.

Nei, M., 1968. Evolutionary change of linkage intensity. *Nature* 218, 1160–1161.

Nei, M., 1987. *Molecular evolutionary genetics*. Columbia Univ Press.

Nei, M., 2003. Genome evolution: let's stick together. *Heredity* 90, 411–412.

Nei, M., Kojima, K. and Schaffer, H. E., 1967. Frequency changes of new inversions in populations under mutation-selection equilibria. *Genetics* 57, 714–750.

Nei, M., Maruyama, T. and Wu, C.-I., 1983. Models of evolution of reproductive isolation. *Genetics* 103, 557–579.

Newman, E. I. and Rovira, A. D., 1975. Allelopathy among some British grassland species. *J Ecol* 63, 727–737.

O'Donald, P., 1980. *Genetic Models of Sexual Selection*. Cambridge Univ Press.

Ohta, T. and Kojima, K., 1968. Survival probabilities of new inversions in large populations. *Biometrics* 240, 501–516.

Orians, G. H., 1969. On the evolution of mating systems in birds and mammals. *Am Nat* 103, 589–603.

Orr, H. A., 1991. A test of Fisher's theory of dominance. *Proc Natl Acad Sci USA* 88, 11413–11415.

Orr, H. A., 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* 139, 1805–1813.

Otto, S. P. and Bourguet, D., 1999. Balanced polymorphisms and the evolution of dominance. *Am Nat* 153, 561–574.

Pal, C. and Hurst, L. D., 2003. Evidence for co-evolution of gene order and recombination rate. *Nature Genetics* 33, 392–395.

Passow, U., 2002. Transparent expopolymer particles (TEP) in aqautic environments. *Progress in Oceanography* 55, 287–333.

Pen, I., 2000. Reproductive effort in viscous populations. *Evolution* 54, 293–297.

Pepper, J. W., 2003. The evolution of evolvability in genetic linkage patterns. *BioSystems* 69, 115–126.

Perrot, V., Richerd, S. and Valero, M., 1991. Transition from haploidy to diploidy. *Nature* 351, 315–317.

Pevzner, P. and Tesler, G., 2003. Genome rearrangements in mammalian evolution: Lessons from human and mouse genomes. *Genome Res* 13, 37–45.

Phillips, P. C., 1998. The language of gene interaction. *Genetics* 149, 1167–1171.

Pianka, E. R., 1970. On r and K selection. *Am Nat* 104, 592–597.

Pomiankowski, A., 1987. Sexual selection: the handicap principle does work - sometimes. *Proc Roy Soc Lon B* 231, 123–145.

Pomiankowski, A., 1988. The evolution of female mate preferences for male genetic quality. *Oxf Surv Evol Biol* 5, 136–184.

Proctor, N. H., Chan, S. L. and Trevor, A. J., 1975. Production of saxitoxin by cultures of *Gonyaulax catenella*. *Toxicon* 13, 1–9.

Proulx, S. R. and Phillips, P. C., 2005. The opportunity for canalization and the evolution of genetic networks. *Am Nat* 165, 147–162.

Provine, W. B., 1971. *The Origins of Theoretical Population Genetics*. Univ Chicago Press.

Queitsch, C., Sangster, T. A. and Lindquist, S., 2002. Hsp90 as a capacitor of phenotypic variation. *Nature* 417, 618–624.

Queller, D. C., 1994. Genetic relatedness in viscous populations. *Evol Ecol* 8, 70–73.

Riebesell, U., 1988. Particle aggregation during a diatom bloom. II. Biological aspects. *Mar Ecol Prog Ser* 69, 281–291.

Riley, M. A., Goldstone, C. M., Wertz, J. E. and Gordon, D., 2003. A phylogenetic approach to assessing the targets of microbial warfare. *J Evol Biol* 16, 690–697.

Riley, M. A. and Wertz, J. E., 2002. Bacteriocins: evolution, ecology and application. *Ann Rev Microbiol* 56, 117–137.

Robyr, D., Suka, Y., Xenarios, I., Kurdistani, S. K., Wang, A., Suka, N. and Grunstein, M., 2002. Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell* 17, 437–446.

Roman, C. S. and Robrow, M., 1973. The sites of radiation induced-breakage in human lymphocyte chromosomes, determined by quinacrine fluorescence. *Mutat Res* 18, 325–331.

Roughgarden, J., 1971. Density dependent natural selection. *Ecology* 52, 453–468.

Round, F. E., 1981. *The ecology of algae*. Cambridge Univ. Press.

Roy, P. J., Stuart, J. M., Lund, J. and Kim, S. K., 2002. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* 418, 975–979.

Ruddle, F. H., Bartels, J. L., Bentley, K. L., Kappen, C., Murtha, M. T. and Pendleton, J. W., 1994. Evolution of Hox genes. *Ann Rev Genet* 28, 423–442.

Rutherford, S. L. and Lindquist, S., 1998. Hsp90 as a capacitor for morphological evolution. *Nature* 396, 336–342.

Ruvolo, M., 1997. Molecular phylogeny of the hominids: Inferences from multiple independent DNA sequence data sets. *Mol Biol Evol* 14, 248–265.

Sager, R. and Granick, S., 1953. Nutritional studies with *Chlamydomonas reinhardi*. *Ann N Y Acad Sci* 56, 831–838.

Saino, N., Calza, S. and Moller, A. P., 1998. Effects of a dipteran ectoparasite on immune response and growth trade-offs in barn swallow, *Hirundo rustica*, nestlings. *Oikos* 81, 217–228.

Schaeffer, S. W., Goetting-Minesky, M. P., Kovacevic, M., Peoples, J. R., Graybill, J. L., Miller, J. M., Kim, K., Nelson, J. G. and Anderson, W. W., 2003. Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. *Proc Natl Acad Sci USA* 100, 8319–8324.

Schmalhausen, I. I., 1949. *Factors of Evolution. The Theory of Stabilising Selection.* The Blakiston Co., Philadelphia.

Schuster, P., Fontana, W., Stadler, P. F. and Hofacker, I., 1994. From sequences to shapes and back: a case study in RNA secondary structures. *Proc Roy Soc Lon B* 255, 279–284.

Seabright, M., 1973. High resolution studies on the pattern of induced exchanges in the human karyotype. *Chromosoma* 40, 333–346.

Seger, J., 1985. Unifying genetic models for the evolution of female choice. *Evolution* 39, 1185–1193.

Seger, J. and Trivers, R., 1986. Asymmetry in the evolution of female mating preferences. *Nature* 319, 771–773.

Sen, N. and Fogg, G. E., 1966. Effects of glycollate on the growth of a planktonic *Chlorella*. *J Exp Botany* 17, 417–425.

Sharp, P., Shields, D. C., Wolfe, K. H. and Li, W. H., 1989. Chromosomal location and evolutionary rate variation in enterobacterial genes. *Science* 246, 808–810.

Shen-Orr, S. S., Milo, R., Mangan, S. and Alon, U., 2002. Network motifs in the transcriptional regulation of *E. coli*. *Nat Genet* 31, 64–68.

Shorter, R. and Frey, K. J., 1979. Relative yields of mixtures and monocultures of oat genotypes. *Crop Sci* 19, 548–553.

Smayda, T. J., 1997a. Harmful algal blooms: Their ecophysiology and general relevance to phytoplankton blooms in the sea. *Limnol Oceanogr* 42, 1137–1153.

Smayda, T. J., 1997b. What is a bloom? A commentary. *Limnol Oceanog* 42, 1132–1136.

Smetacek, V., 1985. Role of sinking in diatoms life history cycles: ecological, evolutionary and geological significance. *Mar Biol* 84, 239–251.

Soane, I. D. and Clarke, B., 1973. Evidence for apostatic selection using olfactory cues. *Nature* 241, 62–63.

Soler, J. J., de Neve, L. and Perez-Contreras, T., 2003. Trade-off between immunocompetence and growth in magpies: an experimental study. *Proc Roy Soc Lon B* 270, 241–248.

Sorokin, C. and Krauss, R. W., 1958. The effects of light intensity on the growth rates of green algae. *Plant Phys* 33, 109–113.

Spassky, B., Spassky, N., Levene, H. and Dobzhansky, T., 1958. Release of genetic variability through recombination. I. *Drosophila pseudoobscura*. *Genetics* 43, 844–865.

Spiess, E. B., 1958a. The effect of genetic recombination on viability in *Drosophila*. *Cold Spr Harb Symp Quant Biol* 23, 239–250.

Spiess, E. B., 1958b. Release of genetic variability through recombination. II. *Drosophila persimilis*. *Genetics* 44, 43–58.

Stearns, S. C., 2002. Progress on canalization. *Proc Natl Acad Sci USA* 99, 10229–10230.

Steidinger, K. A. and Baden, D. G., 1984. *Dinoflagellates*, Academic Press, inc., chap. Toxic marine dinoflagellates. pp. 201–261.

Sturtevant, A. H. and Mather, K., 1938. The interrelations of inversions, heterosis and recombination. *Am Nat* 72, 447–452.

Sueoka, N., 1960. Mitotic replication of deoxyribonucleic acid in *Chlamydomonas reinhardi*. *Proc Natl Acad Sci USA* 46, 83–91.

Syrett, P. J., 1962. *Physiology and biochemistry of algae*, Academic Press, inc., chap. Nitrogen Assimilation. pp. 171–188.

Szathmary, E., 1999. *Levels of Selection in Evolution*, Princeton Univ Press, chap. The first replicators. pp. 31–52.

Tallberg, P. and Heiskanen, A. S., 1998. Species-specific phytoplankton sedimentation in relation to primary production along an inshore-offshore gradient in the Baltic Sea. *J Plankton Res* 20, 2053–2070.

Tanabe, H., Muller, S., Neusser, J., M.and von Hase, Calcagno, E., Cremer, M., Solovei, I., Cremer, C. and Cremer, T., 2002. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci USA* 99, 4424–4429.

Taroncher-Oldenburg, G., Kulis, D. M. and Anderson, D. M., 1999. Coupling of saxitoxin biosynthesis to the G(1) phase of the cell cycle in the dinoflagellate *Alexandrin fundyenese*: Temperature and nutrient effects. *Nat Toxins* 7, 207–219.

Taylor, P. D., 1996. Inclusive fitness arguments in genetic models of behaviour. *J Math Biol* 34, 654–674.

Teichmann, S. A. and Veitia, R. A., 2004. Genes encoding subunits of stable complexes are clustered on the yeast chromosomes. *Genetics* 167, 2121–2125.

Templeton, A. R., 1981. Mechanisms of speciation - a population genetics approach. *Ann Rev Ecol Syst* 12, 23–48.

Thacker, A. and Syrett, P. J., 1972. The assimilation of nitrate and ammonium by *Chlamydomonas reinhardi*. *New Phytol* 71, 423–433.

Thornton, D. C. O., 2002a. Diatom aggregation in the sea: mechanisms and ecological implications. *Eur J Phycol* 37, 149–161.

Thornton, D. C. O., 2002b. Individuals, clones or groups? Phytoplankton behaviour and units of selection. *Ethol Ecol Evol* 14, 165–173.

Tomkins, J. L., 1999. Environmental and genetic determinants of the male forceps length dimorphism in the european earwig *Forficula auricularia* L. *Behav Ecol Sociobiol* 47, 1–8.

Tomkins, J. L. and Brown, G. S., 2004. Population density drives the local evolution of a threshold dimorphism. *Nature* 431, 1099–1103.

Tomkins, J. L. and Simmons, L. W., 1998. Female choice and manipulations of forceps size and symetry in the earwig *Forficula auricularia* L. *Anim Behav* 56, 347–356.

Trivers, R. L., 1972. *Sexual Selection and the Descent of Man*, Heinemann, chap. Parental investment and sexual selection. pp. 136–179.

Trivers, R. L., 1985. *Social Evolution*. Benjamin/Cummings.

Trivers, R. L. and Willard, D. E., 1973. Natural selection of parental ability to vary the sex ratio of offspring. *Science* 179, 90–92.

Tucic, N., Ayala, F. J. and Marinkovic, D., 1981. Correlation between recombination frequency and fitness in *Drosophila melanogaster*. *Genetica* 56, 61–69.

Turner, J. R. G., 1967. Why does the genotype not congeal? *Evolution* 21, 645–656.

Uzzell, T. and Pilbeam, D., 1971. Phyletic divergence dates of hominid primates: a comparison of fossil and molecular data. *Evolution* 25, 615–635.

van Den Hoek, C., Mann, D. G. and Jahns, H. M., 1995. *Algae: an introduction to phycology*. Cambridge Univ. Press.

van Nimwegen, E., Crutchfield, J. P. and Huynen, M., 1999. Neutral evolution of mutational robustness. *Proc Natl Acad Sci USA* 96, 9716–9720.

van Valen, L. and Levins, R., 1968. The origins of inversion polymorphisms. *Am Nat* 102, 5–24.

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

Waddington, C. H., 1940. *Organisers and Genes*. Cambridge Univ. Press, Cambridge.

Waddington, C. H., 1942. Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565.

Wade, M. J., Johnson, N. A. and Toquenaqa, Y., 1999. Temperature effects and genotype-by-environment interaction in hybrids: Haldane's rule in flour beetles. *Evolution* 53, 855–865.

Wagner, A., 2000. Robustness against mutations in genetic networks of yeast. *Nat Genet* 24, 355–361.

Wagner, A., 2001. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol Biol Evol* 18, 1283–1292.

Wagner, A., 2005. Distributed robustness versus redundancy as causes of mutational robustness. *BioEssays* 27, 176–188.

Wagner, G. P., Booth, G. and Bagheri-Chaichian, H., 1997. A population genetic theory of canalization. *Evolution* 51, 329–347.

Wagner, G. P., Laubichler, M. D. and Bagheri-Chaichian, H., 1998. Genetic measurement theory of epistatic effects. *Genetica* 102/3, 569–580.

Wang, J., 2005. Estimation of effective population sizes from data on genetic markers. *Phil Trans Roy Soc B* 360, 1395–1409.

Weismann, A., 1889. *Essays on heredity and kindred biological subjects*. Oxford Univ Press.

West, S. A. and Buckling, A., 2003. Cooperation, virulence and siderophore production in bacterial parasites. *Proc Roy Soc Lon B* 270, 37–44.

Whitlock, M. C., Phillips, P. C., Moore, F. B.-G. and Tonsor, S. J., 1995. Multiple fitness peaks and epistasis. *Ann Rev Ecol Syst* 26, 601–629.

Wilke, C. O., 2001. Selection for fitness versus selection for robustness in RNA secondary structure folding. *Evolution* 55, 2412–2420.

Wilke, C. O. and Adami, C., 2003. Evolution of mutational robustness. *Mut Res* 522, 3–11.

Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. and Adami, C., 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* 412, 331–333.

Wilkins, A. S., 1997. Canalization: a molecular genetic perspective. *BioEssays* 19, 257–261.

Wilkinson, G. S. and Taper, M., 1999. Evolution of genetic variation for condition-dependent traits in stalk-eyed flies. *Proc Roy Soc Lon B* 266, 1685–1690.

Williams, E. J. B. and Bowles, D. J., 2004. Co-expression of neighbouring genes in the genome of *Arabidopsis thaliana*. *Genome Res* 14, 1060–1067.

Williams, G. C., 1966. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton Univ Press.

Williams, T. and Fried, M., 1986. A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3' ends. *Nature* 322, 275–279.

Wolf, J. B., Brodie III, E. D. and Moore, A. J., 1999. The role of maternal and paternal effects in the evolution of parental quality by sexual selection. *J Evol Biol* 12, 1157–1167.

Wolf, J. B., Moore, A. J. and Brodie III, E. D., 1997. The evolution of indicator traits for parental quality: the role of maternal and paternal effects. *Am Nat* 150, 639–649.

Wolfe, G. V., Steinke, M. and Kirst, G. O., 1997. Grazing-activated chemical defence in a unicellular marine alga. *Nature* 387, 894–897.

Wright, S., 1929. Evolution in a Mendelian population. *Anat Rec* 44, 287–333.

Wright, S., 1931. Evolution in Mendelian populations. *Genetics* 10, 97–159.

Wright, S., 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc 6th Intern Cong Genet* 1, 356–366.

Wright, S., 1934. Physiological and evolutionary theories of dominance. *Am Nat* 68, 25–53.

Wright, S., 1963. *Methodology in Mammalian Genetics*, Holden-Day, chap. Genic Interaction. pp. 159–192.

Wright, S., 1977. *Evolution and the Genetics of Populations. Vol III.* Univ Chicago Press.

Wu, C. I., 1991. Inferences of species phylogeny in relation to segregation of ancient polymorphisms. *Genetics* 127, 429–435.

Yang, Y. S., Song, H. D., Shi, W. J., Hu, R., Han, Z. G. and Chen, J. L., 2002. Chromosome localization analysis of genes strongly expressed in human visceral adipose tissue. *Endocrine* 18, 57–66.

Yang, Z. H., 2002. Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. *Genetics* 162, 1811–1823.

Yarbrough, K. and Kojima, K., 1967. The mode of selection at the polymorphic esterase-6 locus in cage populations of *Drosophila melanogaster*. *Genetics* 57, 677–686.

Yasumoto, T., 2005. Chemistry, etiology, and food chain dynamics of marine toxins. *Proc Japan Acad B* 81, 43–51.

Yeager, M. and Hughes, A. L., 1999. Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunol Rev* 167, 45–58.

Yu, S. B., Li, J. X., Xu, C. G., Tan, Y. F., Gao, Y. J., Li, X. H., Zhang, Q. and Maroof, M. A. S., 1997. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA* 94, 9226–9231.

Yuhki, N., Beck, T., Stephens, R. M., Nishigaki, Y., Newmann, K. and O'Brien, S. J., 2003. Comparative genomeorganization of human, murine, and feline MHC class II region. *Genome Res* 13, 1169–1179.

Zahavi, A., 1975. Mate selection - a selection for a handicap. *J Theor Biol* 53, 205–214.

Zahavi, A., 1977. The cost of honesty (further remarks on the handicap principle). *J Theor Biol* 67, 603–605.

Zhang, J. and Nei, M., 1996. Evolution of antennapedia-class homeobox genes. *Genetics* 142, 295–303.

Zhivotovsky, L. A., Feldman, M. W. and Christiansen, F. B., 1994. Evolution of recombination among multiple selected loci: a generalized reduction principle. *Proc Natl Acad Sci USA* 91, 1079–1083.

A Appendix

Algal Media			
	Medium Type, g/l		
	Bolds	Modified Eversole	Sueoka
$NaNO_3$	0.25	-	-
NH_4Cl	-	0.5	0.5
$MgSO_4.7H_2O$	0.05	0.5	0.02
K_2HPO_4	0.1	0.02	1.44
KH_2PO_4	0.15	-	0.72
$CaCl_2$	0.0332	0.01	0.01
NaCl	0.025	-	-
$FeCl_3.6H_2O$	-	0.01	-
1M Tris-HCl (pH 6 or pH 8)	-	50 ml/l	-
Trace Elements			
EDTA	0.01	-	0.01
KOH	0.031	-	0.031
$FeSO_4.7H_2O$	0.005	-	0.005
H_3BO_3	0.0114	0.1	0.0114
Na_2MoO_4	0.00108	0.02	0.00108
$ZnSO_4$	0.0157	0.1	0.0157
$Co(NO_3)_2.6H_2O$	0.0005	-	0.0005
$MnCl_2.4H_2O$	0.00226	0.04	0.00226
$CuSO_4$	0.00158	0.004	0.00158
pH	6.6	6 or 8	6.8

Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Theoretical Biology 241 (2006) 707–715
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Recombination and the evolution of mutational robustness

Andy Gardner[a,b,c,*], Alex T. Kalinka[a]

[a]*Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK*
[b]*Department of Mathematics & Statistics, Queen's University, Kingston, Ont., Canada K7L 3N6*
[c]*Department of Biology, Queen's University, Kingston, Ont., Canada K7L 3N6*

Received 25 July 2005; received in revised form 8 December 2005; accepted 5 January 2006
Available online 20 February 2006

Abstract

Mutational robustness is the degree to which a phenotype, such as fitness, is resistant to mutational perturbations. Since most of these perturbations will tend to reduce fitness, robustness provides an immediate benefit for the mutated individual. However, robust systems decay due to the accumulation of deleterious mutations that would otherwise have been cleared by selection. This decay has received very little theoretical attention. At equilibrium, a population or asexual lineage is expected to have a mutation load that is invariant with respect to the selection coefficient of deleterious alleles, so the benefit of robustness (at the level of the population or asexual lineage) is temporary. However, previous work has shown that robustness can be favoured when robustness loci segregate independently of the mutating loci they act upon. We examine a simple two-locus model that allows for intermediate rates of recombination and inbreeding to show that increasing the effective recombination rate allows for the evolution of greater mutational robustness.

Keywords: Canalization; Epistasis; Linkage disequilibrium; Multilocus methodology; Mutation–selection balance

1. Introduction

It has long been observed that many developmental traits display a high degree of phenotypic robustness, that is, the phenotype is remarkably immune to environmental and genetic perturbations (Waddington, 1940; Schmalhausen, 1949). Waddington (1942) described the phenomenon as 'canalization', and proposed an adaptive explanation. He reasoned that traits under stabilising selection towards some intermediate optimum should benefit from any mechanism that prevents deviation from that optimum due to either genetic or environmental perturbations. Within the class of genetic perturbations are those that are due to deleterious mutations. In recent years, mutational robustness has attracted renewed interest, on both theoretical and empirical fronts (Wagner et al., 1997; Wilkins, 1997; Rutherford and Lindquist, 1998; van Nimwegen et al., 1999; Kawecki, 2000; Wagner, 2000; Wilke, 2001; Wilke et al., 2001; Queitsch et al., 2002; Wilke and Adami, 2003; de Visser et al., 2003; Proulx and Phillips, 2005). Most attention has been given to adaptive explanations, although some researchers have speculated that mutational robustness is a by-product of adaptation against environmental perturbations (Wagner et al., 1997; Burch and Chao, 2004) or simply an emergent property of genetic systems (Kacser and Burns, 1981; von Dassow et al., 2000; Meir et al., 2002; Shen-Orr et al., 2002).

All genetic models for the evolution of robustness require some form of gene interaction or epistasis between the loci involved. In this respect, a distinction can be drawn between the two different models of robustness that are commonly discussed; whether the epistasis is exclusively between the loci involved in the trait or between the trait loci and an unrelated locus (a 'modifier'). There is evidence for the former in RNA and protein folding where there is often extensive degeneracy between the primary sequence and the secondary or tertiary structure (Maynard Smith, 1970; Lau and Dill, 1990; Schuster et al., 1994;

*Corresponding author. Departments of Biology and Mathematics & Statistics, Queen's University, Kingston, Ont., Canada K7L 3N6.
Tel.: +44 131 650 5508; fax: +44 131 650 6564.
E-mail addresses: andyg@mast.queensu.ca (A. Gardner), a.t.kalinka@sms.ed.ac.uk (A.T. Kalinka).

0022-5193/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2006.01.011

an Nimwegen et al., 1999; Wilke, 2001) and in metabolic nd developmental pathways in prokaryotes and eukarotes, where distributed network architectures facilitate obustness via internal pathway degeneracy (Edwards and alsson, 1999, 2000a, b; von Dassow et al., 2000; Meir t al., 2002; Shen-Orr et al., 2002; Ingolia, 2004; Wagner, 005). Wagner (2005) refers to this form of robustness as distributed robustness' and argues, based on empirical vidence, for the primacy of its role in mutational obustness in favour of gene redundancy arising from gene uplicates (Wagner 2000, 2001, 2005). Conversely, there is vidence that heat shock proteins, such as *Hsp*90 in *Drosophila* (Rutherford and Lindquist, 1998) and *Arabiopsis* (Queitsch et al., 2002), and *GroEL* in *Escherischia oli* (Fares et al., 2002), behave as modifiers of mutational obustness. For the remainder of this article, we will restrict ttention to the modifier view of robustness.

The evolution of mutational robustness is conceptually milar to the adaptive evolution of dominance proposed y Fisher (1928). In both cases it is the heritable deviation rom the wild type that is being buffered, and the selective dvantage of the modifier is of the order of the mutation ate (Wright, 1929). Fisher believed that although the elective advantage is weak, in a large population with a umber of recessive mutations the accumulated selective ressure would drive the evolution of dominance. Wright ook the view that dominance emerged as an intrinsic roperty of metabolic pathways and proposed an alterative 'physiological' theory of dominance (Wright, 1934). Kacser and Burns (1981) provided considerable support for Wright's argument with a model of a multienzyme system hat showed that the flux of the enzyme pathway is nsensitive to concentration changes in the enzymes nvolved, suggesting that dominance is an inevitable roperty of such systems. More recently, however, Bagheri nd Wagner (2004) have shown that this may only be the ase when one neglects nonlinear enzyme interactions. Currently, empirical evidence appears not to support Fisher's adaptationist hypothesis (Orr, 1991), although it nay be relevant in situations involving strong selection Haldane, 1956; Mayo and Burger, 1997). The debate ontinues. Another related phenomenon that has received nuch attention is the evolutionary transition from aploidy to diploidy. A benefit may be afforded by an xtended diploid phase due to the masking of recessive r partially recessive deleterious mutations (Crow and Kimura, 1965). Here, the adaptationist view appears to ave a plausible theoretical foundation (Kondrashov and Crow, 1991; Perrot et al., 1991) although, interestingly, it is ncompatible with Fisher's view of dominance since it equires that newly arisen deleterious mutations are always at least partially) recessive (Perrot et al., 1991). Together vith the evolution of mutational robustness, these scenaros involve evolutionary modification of the genetic system self driven by the immediate benefit of alleviating the ffects of deleterious mutations, which are of course a biquitous evolutionary phenomenon.

A classic result that motivates the present study is that at equilibrium the mutation load (L^*) of the population is invariant with respect to the fitness consequences of deleterious alleles. Assuming fitnesses combine multiplicatively across loci, an allele which arises by recurrent irreversible mutation at rate μ and incurs a fitness decrement s will equilibrate at frequency μ/s in a haploid population (mutation–selection balance). Hence the average fitness contributed by this locus is $(1-\mu/s)\times 1+(\mu/s)\times(1-s)=1-\mu$; the mutation load at this locus is then μ, and not a function of s (Haldane, 1937). The result has been generalized for all loci in the genome, giving a population load of mutations $L*=1-e^{-U}$ (Kimura and Maruyama, 1966; Kondrashov, 1988), where U is the per genome per generation mutation rate, and hence the decrement to fitness due to individual mutations is again irrelevant. The reason for this is intuitive: if mutations are more harmful they are more readily removed from the population by selection. Those mutations with large deleterious effects are held at low frequency at mutation–selection balance, and thus cause the same decrement to the mean fitness of the population as less harmful, and hence more frequently encountered, mutations.

The action of mutational robustness is to reduce the magnitude of a mutation's fitness effect. Whilst it may be temporarily advantageous to reduce the selection coefficient associated with the deleterious mutation, this leads to the accumulation of mutations that would otherwise have been cleared by selection, and so a closed population (i.e. no flow of genetic material between populations) with enhanced robustness does not improve its equilibrium mutation load. Thus there is no long-term benefit for being robust, at the level of the closed population. This mutational decay of robust systems has received only limited attention (Frank, 2003). If robustness has an intrinsic cost, such as the energetic cost of synthesizing the robustness gene product, then in the long term it will cause a net disadvantage for the population. Therefore, in an asexual population, we predict eventual loss of robust lineages. However, robustness might be favoured in a sexual population. Since the benefit of robustness (a reduced impact of mutations) accrues to the robust lineage, yet the cost (an increased equilibrium frequency of mutations) is shared by the whole population, robust lineages may have a relative advantage. It seems that this will increasingly be the case as the rate of recombination (in particular, between robustness genes and those genes undergoing mutational perturbation) is increased. This has received some attention, and the hypothesis is supported by contrasting the predictions of models of complete linkage, in which costly robustness is never favoured (Hermisson et al., 2002, p. 26), with those which assume free recombination, in which costly robustness can evolve (Wagner et al., 1997; Dawson, 1999). However, results for robustness evolution with intermediate recombination rates are lacking (de Visser et al., 2003, p. 1962).

We examine a simple model that captures the essence of this problem. The dynamics of the system are described

using a multilocus methodology (developed by Barton and Turelli, 1991; Kirkpatrick et al., 2002) that highlights allele frequencies and linkage disequilibria, which is more natural than following genotype frequencies. Also, it provides a general notation that neatly partitions the various causes of evolutionary change, and allows for arbitrary complexity so that the model is readily extensible within this single framework. Specifically, we develop exact analytical recursions describing the dynamics of a costly robustness modifier and its association with a mutating locus, and from this we generate an invasion condition to determine when this modifier will increase in frequency when vanishingly rare. We then make an assumption of minor robustness variants to examine how the robustness phenotype evolves in the longer term, moving from multilocus population genetics to an evolutionary game theoretic analysis.

2. Model and analysis

2.1. Two-locus model

We consider a simple model which captures all the important features of this problem—a large population of sexual haploids, with a life cycle which involves (i) selection, followed by (ii) mutation, and finally (iii) mating to form diploid zygotes, which undergo meiosis to form the next generation of haploid individuals. All notation used in this article are summarized and defined in Table 1. A locus i suffers recurrent, irreversible mutation, from the wild-type allele with value $X_i = 0$ to mutant allele with value $X_i = 1$, at a rate μ. The fitness of the wild type is 1, and the fitness of the mutant is $1-s$ in the absence of robustness. The frequency of the mutant is denoted p_i, and thus the frequency of the wild type is $q_i = 1 - p_i$. A second locus j controls the expression of the deleterious mutant, when it occurs at the first locus: with robustness k, the fitness contributed by the first locus is $1-(1-k)s$. Two alleles, with varying robustness effect, are present. The 'resident' allele has value $X_j = 0$ and robustness effect k_x, and the 'variant' allele has value $X_j = 1$ and robustness effect k_y. The robustness locus also incurs a direct (intrinsic) cost, with the resident contributing $1-c_x$, and the variant $1-c_y$, to an individual's fitness. The frequency of the variant is denoted p_j and the frequency of the resident is $q_j = 1 - p_j$. We will assume that the direct effects of the loci multiply to give genotype fitness. The four genotype fitnesses are summarized in Table 2. The effects of linkage and inbreeding are described by an effective rate of recombination parameter, r_e.

Following the above model, an individual's fitness may be written in the form:

$$\begin{aligned} w &= (1-X_i)(1-X_j)w_{00} + X_i(1-X_j)w_{10} + (1-X_i)X_jw_{01} \\ &\quad + X_iX_jw_{11} = (1-X_i)(1-X_j)(1-c_x) \\ &\quad + X_i(1-X_j)(1-(1-k_x)s)(1-c_x) \\ &\quad + (1-X_i)X_j(1-c_y) + X_iX_j(1-(1-k_y)s)(1-c_y), \end{aligned} \tag{1}$$

Table 1
Summary of notation used in this article

Notation	Definition
μ	Deleterious mutation rate at a single locus
s	Selection coefficient associated with deleterious mutation
U	Per genome per generation mutation rate
L^*	Equilibrium mutation load
i	Locus under recurrent mutation
j	Locus controlling robustness
i, j	A generic gene position
A, B	A generic set of gene positions
W	The set of all gene positions contributing to fitness
$X_\mathbf{i}$	Allelic value for gene position **i** (0 or 1)
$p_\mathbf{i}$	Frequency of the $X_\mathbf{i} = 1$ allele
$q_\mathbf{i} = 1 - p_\mathbf{i}$	Frequency of the $X_\mathbf{i} = 0$ allele
$\zeta_\mathbf{i} = X_\mathbf{i} - p_\mathbf{i}$	Allelic deviation for gene position **i**
$\zeta_\mathbf{A} = \prod_{\mathbf{i}\in\mathbf{A}}\zeta_\mathbf{i}$	Allelic deviation for a set **A** of gene positions
$D_\mathbf{A} = E[\zeta_\mathbf{A}]$	Association for set **A** of gene positions
w	Fitness of an individual
$\bar{w}$	Population mean fitness
$a_\mathbf{A}$	Multilocus selection coefficient for set **A** of gene positions
z	A generic robustness strategy
x	Resident robustness strategy
$y = x + \delta x$	Variant robustness strategy
x^*	Equilibrium robustness strategy
k_z, $k[z]$	Robustness effect associated with strategy z
c_z, $c[z]$	Cost of robustness associated with strategy z
r_e	Effective rate of recombination
$\lambda = 1 + \delta\lambda$	Invasion fitness of robustness variant; its asymptotic rate of increase

Table 2
Genotype fitness (w) as a function of allelic value ($X = 0$, 1) at the mutating locus (i) and the robustness locus (j).

		X_j	
		0	1
X_i	0	$1-c_x$	$1-c_y$
	1	$(1-(1-k_x)s)(1-c_x)$	$(1-(1-k_y)s)(1-c_y)$

where $w_{X_iX_j}$ is the fitness of the (X_i, X_j) genotype (see Table 2). This fitness function is analogous to Eq. (7) in Barton and Turelli (1991).

2.2. Multilocus population statistics

The multilocus framework of Kirkpatrick et al. (2002) describes individuals and populations according to deviations from average values. An allelic deviation ($\zeta_\mathbf{i} = X_\mathbf{i} - p_\mathbf{i}$) is defined for a generic gene position **i**, and describes the deviation of the allelic value ($X_\mathbf{i} = 0$ or 1) from the population average ($p_\mathbf{i}$) at that position. Thus, the population average allelic deviation for a single gene position is zero. A corresponding deviation term ($\zeta_\mathbf{A} = \prod_{\mathbf{i}\in\mathbf{A}}\zeta_\mathbf{i}$) may be assigned to a set **A** of gene positions, and is the product of the allelic deviations for all the positions in

hat set. Note that the average deviation for a set of two gene positions (**i** and **j**) is equal to the allelic covariance between these positions ($E[\zeta_{\mathbf{ij}}] = E[(X_{\mathbf{i}} - p_{\mathbf{i}})(X_{\mathbf{j}} - p_{\mathbf{j}})] = Cov[X_{\mathbf{i}}, X_{\mathbf{j}}]$), and thus is equivalent to the linkage disequilibrium ($D_{\mathbf{ij}}$) between these gene positions. In general, the population average deviation for a set **A** of gene positions will be denoted $D_{\mathbf{A}}$. Thus, the population composition with respect to a set of gene positions **B** may be fully described by the set of allele frequencies ($p_{\mathbf{i}}$, $\mathbf{i} \in \mathbf{B}$) at these positions, and the statistical associations ($D_{\mathbf{A}}$, $\mathbf{A} \subseteq \mathbf{B}$) between these positions. If an association term corresponds to a set of gene positions in which a particular position features several times, for example $D_{\mathbf{iiA}}$, then a reduction formula may be applied to re-express this as $p_{\mathbf{i}}q_{\mathbf{i}}D_{\mathbf{A}} + (1-2p_{\mathbf{i}})D_{\mathbf{iA}}$, as outlined by Kirkpatrick et al. (2002).

We may now describe how sets of gene positions impact upon an individual's fitness. Making the substitution $X_{\mathbf{i}} = \zeta_{\mathbf{i}} + p_{\mathbf{i}}$ into the fitness function (1), this may be rearranged into the form

$$\frac{w}{\bar{w}} = 1 + a_i(\zeta_i - D_i) + a_j(\zeta_j - D_j) + a_{ij}(\zeta_{ij} - D_{ij}), \tag{2}$$

where $a_{\mathbf{A}}$ is the contribution of the deviation for a set of gene positions **A** to relative fitness, $w/\bar{w}$. This is analogous to Eq. (6) of Barton and Turelli (1991) and Eq. (7) of Kirkpatrick et al., (2002). The $a_{\mathbf{A}}$ terms provide selection coefficients for a multilocus analysis (Kirkpatrick et al., 2002). For the present model, we have

$$\begin{aligned} a_i &= -s(1 - ((1 - p_j)(k_x + c_x(1 - k_x)) \\ &\quad + p_j(k_y + c_y(1 - k_y))))/\bar{w}, \\ a_j &= (sp_i((k_y + c_y(1 - k_y)) - (k_x + c_x(1 - k_x))) \\ &\quad - (c_y - c_x))/\bar{w}, \\ a_{ij} &= s((k_y + c_y(1 - k_y)) - (k_x + c_x(1 - k_x)))/\bar{w}. \end{aligned} \tag{3}$$

Mean fitness is found by taking an average of w over the population:

$$\begin{aligned} \bar{w} &= 1 - sp_i(1 - k_x)(1 - c_x) - (p_jc_y + (1 - p_j)c_x) \\ &\quad - s((1 - k_y)(1 - c_y) - (1 - k_x)(1 - c_x))(p_ip_j + D_{ij}). \end{aligned} \tag{4}$$

Having extracted multilocus population fitness statistics from the model, we can now use them to make some remarkably elegant statements about how selection moulds the allele frequencies and linkage disequilibrium of this system. After selection we will consider mutation and then transmission.

2.3. Selection

The multilocus methodology provides simple recursion expressions for the change in allele frequencies and genetic associations due to selection. The basic equation is

$$\Delta_S D_{\mathbf{A}} = \sum_{\mathbf{B} \subseteq \mathbf{W}} a_{\mathbf{B}}(D_{\mathbf{AB}} - D_{\mathbf{A}}D_{\mathbf{B}}), \tag{5}$$

where **W** is the set of all gene positions contributing to fitness. For the case of a single gene position ($\mathbf{A} = \mathbf{i}$), we may use expression (5) to describe the change in allele frequency ($p_{\mathbf{i}}$) due to selection:

$$\Delta_S p_{\mathbf{i}} = \sum_{\mathbf{B} \subseteq \mathbf{W}} a_{\mathbf{B}} D_{\mathbf{iB}}. \tag{6}$$

A complication arises in that the association after selection ($D'_{\mathbf{A}} = D_{\mathbf{A}} + \Delta_S D_{\mathbf{A}}$) is described with respect to allele frequencies before selection. It will usually be helpful to correct for this, and the procedure is described in Kirkpatrick et al. (2002). No correction is necessary for the expressions describing allele frequency change. In the context of the present model, the change in the frequency of the deleterious mutation that is due to selection is described by

$$\begin{aligned} p'_i &= p_i + a_iD_{ii} + a_jD_{ij} + a_{ij}D_{iij} = p_i + a_ip_iq_i + a_jD_{ij} \\ &\quad + a_{ij}(1 - 2p_i)D_{ij}. \end{aligned} \tag{7}$$

This notational framework makes clear the causes of evolutionary change: here we see that the response to selection ($\Delta_S p_i = p'_i - p_i$) is given by the product of the strength of selection operating directly on the focal locus (a_i) and the variation at that locus (p_iq_i), plus the product of selection operating directly on the other locus (a_j) and the association between the two loci (D_{ij}), plus the product of selection due to the epistatic interaction between the two loci (a_{ij}) and the appropriate association ($(1-2p_i)D_{ij}$). Similarly, the change in allele frequency, due to selection, at the robustness locus is given by

$$\begin{aligned} p'_j &= p_j + a_iD_{ij} + a_jD_{jj} + a_{ij}D_{ijj} = p_j + a_iD_{ij} \\ &\quad + a_jp_jq_j + a_{ij}(1 - 2p_j)D_{ij}. \end{aligned} \tag{8}$$

From expression (5), the change in the association between the loci is described by

$$\begin{aligned} D'_{ij} &= D_{ij} + a_iD_{iij} + a_jD_{ijj} + a_{ij}(D_{iijj} - D_{ij}^2) - (p'_i - p_i)(p'_j - p_j) \\ &= D_{ij} + a_i(1 - 2p_i)D_{ij} + a_j(1 - 2p_j)D_{ij} \\ &\quad + a_{ij}(p_iq_ip_jq_j + (1 - 2p_i)(1 - 2p_j)D_{ij} - D_{ij}^2) \\ &\quad - (p'_i - p_i)(p'_j - p_j), \end{aligned} \tag{9}$$

where the trailing term corrects. for change in allele frequency (Kirkpatrick et al., 2002).

2.4. Mutation

The change in frequency of the deleterious allele after mutation is described by

$$p''_i = p'_i + \mu(1 - p'_i). \tag{10}$$

Since the j locus does not undergo mutation, $p''_j = p'_j$. From Kirkpatrick et al. (2002), the change in the linkage disequilibrium due to mutation is given by

$$D''_{ij} = (1 - \mu)D'_{ij}. \tag{11}$$

2.5. Transmission

Transmission—the union of gametes, crossing over, and fair meiosis—does not alter the allele frequencies in this model (so $p_i''' = p_i''$ and $p_j''' = p_j''$, where triple primes denote the variable is measured after transmission), but it does impact on the linkage disequilibrium. This is reduced by a fraction equal to the effective rate of recombination (Crow and Kimura, 1970), and so we have

$$D_{ij}''' = (1 - r_e)D_{ij}''. \tag{12}$$

2.6. Invasion analysis

We have obtained recursions describing the change in the frequencies of the deleterious mutation (p_i) and robustness modifier (p_j) and the linkage disequilibrium (D_{ij}) over a single generation incorporating selection, mutation and transmission. We now consider that the variant robustness allele is vanishingly rare ($p_j \to 0$, and hence $D_{ij} \to 0$), and examine the conditions under which this rare allele will increase in frequency (invasion). We will assume that μ is sufficiently small for us not to have to worry about fixation of the deleterious mutation i.e. $\mu < (1 - k_x)s$. We will assume that the deleterious mutation is initially at its equilibrium point, $p_i^* = \mu/((1 - k_x)s)$. While the variant is rare, the evolutionary dynamics at the j locus has vanishing impact on dynamics at the i locus, so in any generation we may express the frequency of the deleterious mutation as $p_i = p_i^* + \delta p_i$, where $\delta p_i \to 0$. Making this substitution, and summarising the changes in the allele frequency at the robustness locus and linkage disequilibrium due to selection, mutation and recombination, obtains $p_j''' = \alpha_1 p_j + \alpha_2 D_{ij} + O(\delta p_i^2, p_j^2, D_{ij}^2)$ and $D_{ij}''' = \alpha_3 p_j + \alpha_4 D_{ij} + O(\delta p_i^2, p_j^2, D_{ij}^2)$, where

$$\alpha_1 = \frac{(1 - c_y)(1 - k_x - \mu(1 - k_y))}{(1 - \mu)(1 - c_x)(1 - k_x)},$$

$$\alpha_2 = -\frac{s(1 - c_y)(1 - k_y)}{(1 - \mu)(1 - c_x)},$$

$$\alpha_3 = \frac{(1 - r_e)(k_y - k_x)(1 - c_y)\mu((1 - k_x)s - \mu)}{(1 - c_x)(1 - k_x)^2 s(1 - \mu)}, \text{ and}$$

$$\alpha_4 = \frac{(1 - r_e)(1 - c_y)((1 - k_x)(1 - (1 - k_y)s) - \mu(k_y - k_x))}{(1 - \mu)(1 - c_x)(1 - k_x)}. \tag{13}$$

Neither of these recursions are functions of δp_i, therefore we need not explicitly follow the frequency of the deleterious mutation, so long as we assume it is close to its equilibrium. The asymptotic rate of increase of the rare variant, its 'invasion fitness', is given by the leading eigenvalue for the above system. This is the solution λ to the characteristic equation $(\alpha_1 - \lambda)(\alpha_4 - \lambda) - \alpha_2\alpha_3 = 0$ that has the largest magnitude. The condition for invasion of the robustness variant is $\lambda > 1$.

2.7. Evolution of robustness

We have obtained a condition for the invasion of a given resident population by a given variant robustness allele. We now ask the following questions: (1) is there a resident allele that cannot be invaded by any robustness variant? (2) Will the population converge on this evolutionarily stable state? In other words, we are interested in identifying the endpoint of robustness evolution in the longer term. To address this, we will now consider a continuum of robustness strategies (z), from zero robustness ($z = 0$) to full robustness ($z = 1$), each encoded by an allele at the j locus. The cost and effect of robustness parameters from the previous sections are now considered as functions of the robustness strategy ($c[z]$ and $k[z]$; where $c[0] = k[0] = 0$ and dc/dz, $dk/dz > 0$ for all z). The resident allele encodes the robustness strategy $z = x$, and the variant encodes $z = y$. Thus, $c_x = c[x]$, $k_x = k[x]$, $c_y = c[y]$ and $k_y = k[y]$. For ease of analysis, we will consider only local stability, restricting our attention to $y = x + \delta x$ where $\delta x \to 0$. Since this represents near-neutrality, the invasion fitness of the variant will be of the form $\lambda = 1 + \delta\lambda$, where $\delta\lambda \to 0$. Upon this assumption, we may solve the characteristic equation from earlier to obtain $\delta\lambda \to ((1 - \alpha_1)(1 - \alpha_4) - \alpha_2\alpha_3)/(\alpha_1 + \alpha_4 - 2)$ as $\delta x \to 0$, or

$$\delta\lambda \approx \frac{\mu r_e(1 - c[x])k'[x] - (1 - k[x])(r_e + (1 - r_e)(1 - k[x])s - \mu)c'[x]}{(1 - c[x])(1 - k[x])(r_e + (1 - r_e)(1 - k[x])s - \mu)}\delta x, \tag{14}$$

where the primes denote derivatives evaluated at the resident robustness strategy, i.e. $c'[x] = dc[z]/dz|_{z=x}$ and $k'[x] = dk[z]/dz|_{z=x}$. Marginal invasion fitness is given by $\partial\lambda/\partial y|_{y=x} = \delta\lambda/\delta x$. Setting $r_e = 0$, marginal invasion fitness reduces to $\partial\lambda/\partial y|_{y=x} = -c'[x]/(1 - c[x])$, which is negative for all x: over the whole range of resident strategies, selection favours variants with reduced robustness. Hence, the only equilibrium point in the absence of recombination is at $x^* = 0$. This means that when the effective rate of recombination is zero, costly robustness cannot evolve (Hermisson et al., 2002). We now ask, for $r_e > 0$, what is the end point of robustness evolution? We are therefore looking for a strategy that is both evolutionarily stable once attained (an ESS; Maynard Smith and Price, 1973; Maynard Smith, 1982), and is also attainable (i.e. convergence stable, so that when x is close to x^*, y closer to x^* will invade; Eshel and Motro, 1981; Taylor, 1996). A strategy that is both an ESS and is convergence stable is termed a 'continuously stable strategy' (CSS; Eshel, 1983; Christiansen, 1991). If a strategy x^* is evolutionarily stable, then it must satisfy $\partial\lambda/\partial y|_{y=x=x^*} = 0$. Thus,

$$\frac{d}{dr_e}\left[\frac{\partial\lambda}{\partial y}\bigg|_{y=x=x^*}\right] = \frac{\partial}{\partial r_e}\left[\frac{\partial\lambda}{\partial y}\bigg|_{y=x=x^*}\right] + \frac{\partial}{\partial x^*}\left[\frac{\partial\lambda}{\partial y}\bigg|_{y=x=x^*}\right]\frac{dx^*}{dr_e} = 0, \tag{15}$$

which can be re-arranged to give

$$\frac{dx^*}{dr_e} = -\frac{\partial/\partial r_e[\partial\lambda/\partial y|_{y=x=x^*}]}{\partial/\partial x^*[\partial\lambda/\partial y|_{y=x=x^*}]}. \tag{16}$$

Noting that convergence stability implies $\partial[\partial\lambda/\partial y|_{y=x=x^*}]/\partial x^* < 0$ (Taylor, 1996), the CSS satisfies

$$\operatorname{sgn}\left[\frac{\mathrm{d}x^*}{\mathrm{d}r_e}\right] = \operatorname{sgn}\left[\frac{\partial}{\partial r_e}\left[\left.\frac{\partial\lambda}{\partial y}\right|_{y=x=x^*}\right]\right] \tag{17}$$

(Pen, 2000), where the function sgn returns the sign, or sense, of its real argument, i.e. positive or negative or zero. The partial derivative on the RHS is

$$\frac{\partial}{\partial r_e}\left[\left.\frac{\partial\lambda}{\partial y}\right|_{y=x=x^*}\right] = \frac{\mu((1-k[x^*])s-\mu)k'[x^*]}{(1-k[x^*])(r_e+(1-r_e)(1-k[x^*])s-\mu)}. \tag{18}$$

Since RHS of (18) >0, it follows from (16) that the CSS x^* (when it exists) is a monotonically increasing function of r_e. Thus, we expect the endpoint of evolution to be a greater degree of robustness the higher the effective rate of recombination. Applying the same procedure to the selection coefficient (s) and mutation rate (μ) obtains

$$\frac{\partial}{\partial s}\left[\left.\frac{\partial\lambda}{\partial y}\right|_{y=x=x^*}\right] = -\frac{\mu r_e(1-r_e)k'[x^*]}{(r_e+(1-r_e)(1-k[x^*])s-\mu)^2} < 0 \tag{19}$$

and

$$\frac{\partial}{\partial \mu}\left[\left.\frac{\partial\lambda}{\partial y}\right|_{y=x=x^*}\right] = \frac{r_e(r_e+(1-r_e)(1-k[x^*])s)k'[x^*]}{(1-k[x^*])(r_e+(1-r_e)(1-k[x^*])s-\mu)^2} > 0 \tag{20}$$

i.e. the CSS x^* is a decreasing function of s and an increasing function of μ, so we expect the endpoint of evolution to be a greater degree of robustness as we decrease the magnitude of the deleterious effect of mutations and as we increase the mutation rate. From (14), the exact value of the CSS x^* can be found by solving the equation

$$\mu r_e(1-c[x])k'[x] - (1-k[x])(r_e+(1-r_e)(1-k[x])s-\mu)c'[x] = 0. \tag{21}$$

Some representative numerical examples are given in Figs. 1A and 2A. The assumption of vanishing variation is somewhat artificial, and so we have used simulations to test the predictions using a similar two-locus model that allows for continuum alleles which are simultaneously extant (simulation results are presented in Figs. 1B and 2B). We find that numerical solutions to the analytical prediction given by (21) and the results of the simulations are generally in good agreement. Depending on the choice of parameters and robustness functions, there may be: (1) a single internal equilibrium, which is a CSS (Fig. 1A); (2) an unstable equilibrium in addition to the CSS (Fig. 2A, e.g. lines for $r_e = 0.1$, 0.05, 0.01); or (3) no internal equilibria (Fig. 2A, e.g. $r_e = 0.5$). The simulations confirm that a population initialized at close to zero robustness will ultimately find itself trapped at the CSS, where this exists

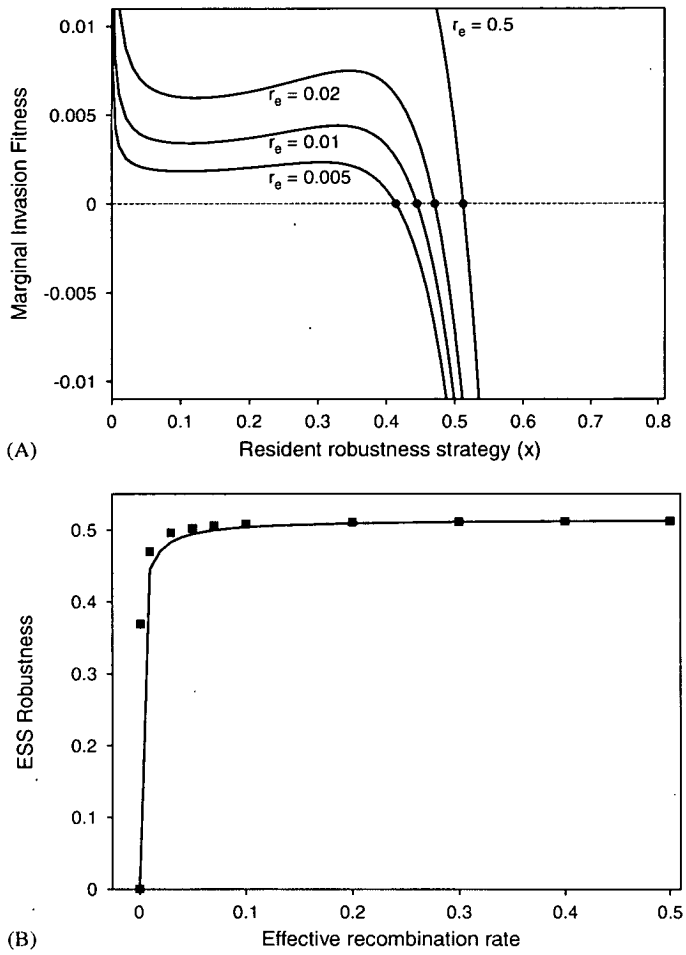


Fig. 1. (A) Marginal invasion fitness $\partial\lambda/\partial y|_{y=x}$ as a function of the resident robustness strategy (x) and the effective recombination rate (r_e), assuming $\mu = 0.01, s = 0.1, c[z] = z^{10}, k[z] = z^{1/2}$. The sign of marginal invasion fitness determines the direction of selection; if it is positive then variant strategies increasing robustness are favoured, and if it is negative then variant strategies reducing robustness are favoured. The convergence stable robustness strategy (marked by a filled circle) is an increasing function of the effective rate of recombination. (B) For the same model, simulation results (squares) confirm the analytical prediction (line) that robustness increases with the effective rate of recombination.

(Figs. 1B and 2B). Due to the assumption that mutation–selection balance holds the deleterious mutation at intermediate frequency (i.e. $\mu < (1-k_x)s$, which may also be written as $k_x < 1-\mu/s$) the present analysis does not allow for examination of the evolution of almost complete robustness ($k_x = 1$).

Although the simulations and analytical predictions have a very good fit, they are not perfect. In particular, the simulations tend to give an end-point of robustness evolution that is higher than predicted from the game theoretic approach. There is reason to suspect that the invasion analysis underestimates the benefits of robustness. For example consider a variant robustness allele with associated cost such that its asymptotic rate of increase is exactly 1. Initially, this allele will increase in frequency, and will eventually settle at a neutral equilibrium. To some

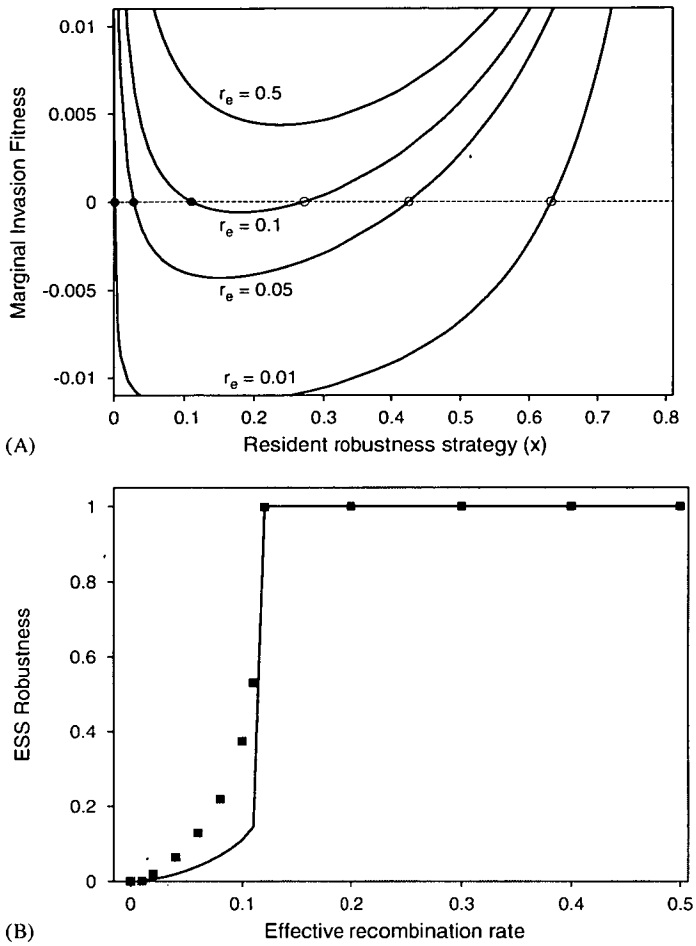


Fig. 2. (A) Marginal invasion fitness $\partial\lambda/\partial y|_{y=x}$ as a function of the resident robustness strategy (x) and the effective recombination rate (r_e), assuming $\mu = 0.01, s = 0.1, c[z] = 0.15z, k[z] = z^{1/2}$. The sign of marginal invasion fitness determines the direction of selection; if it is positive then variant strategies increasing robustness are favoured, and if it is negative then variant strategies reducing robustness are favoured. The convergence stable robustness strategy (where it exists, e.g. for $r_e = 0.01$, 0.05, 0.1, marked by a filled circle) is an increasing function of the effective rate of recombination; there is also sometimes an unstable equilibrium (marked by an empty circle). (B) For the same model, simulation results (squares) confirm the analytical prediction (line) that robustness increases with the effective rate of recombination, and that there is no internal stable end point for $r_e > 0.11$.

extent then, this allele has been favoured, although technically it does not invade. More generally, deviations from the analytical predictions will occur due to slowness in attaining the evolutionary endpoint, as selection acts weakly upon robustness and only a finite number of generations are simulated.

3. Discussion

We have examined the evolution of costly mutational robustness in a simple two-locus model for when recombination (r_e) between the two loci is intermediate. Previously, only the extremes of zero recombination (Hermisson et al., 2002) and freely recombining loci (Wagner et al., 1997; Dawson, 1999) have been considered. A multilocus methodology has been employed to obtain recursions for allele frequencies at the robustness locus and the association between this locus and the locus that is under recurrent mutation. The result is an analytical condition for when the robustness variant invades a population. Restricting attention to minor variants, we have used this condition to determine how the end point of robustness evolution varies with the effective rate of recombination (r_e), the intrinsic deleterious effect of the mutation (s) and the mutation rate (μ). Consistent with previous theory, we find that costly robustness cannot be favoured when the effective rate of recombination is zero. In addition, we show that, where one exists, the internal stable endpoint of robustness evolution is an increasing function of effective recombination rate and the mutation rate, and is a decreasing function of the intrinsic deleterious effect of the mutation. Although the analysis assumes vanishing robustness variation in the population at any time, simulations that relax this assumption reveal the analytical treatment is robust.

Why do we predict enhanced robustness with increasing effective rate of recombination? Recombination favours robustness in two ways: (1) by allowing a robust lineage to discard the excess of deleterious mutations it has accumulated, and (2) these deleterious mutations are inflicted upon non-robust lineages where they cause enhanced damage to fitness, thus increasing the relative fitness of the robust lineages. Put another way, by breaking down the association between the robustness gene and the target of the robustness effect, recombination decouples the immediate benefit of robustness (enhanced fitness in the context of a mutant genotype) from the long-term cost (increased frequency of mutations at mutation–selection equilibrium); the former accruing only to robust individuals, and the latter being paid by the population as a whole. This benefit for robustness is mirrored in the Perrot et al. (1991) model for the evolution of diploidy, which features gene flow between haploids and diploids. With this in mind, the model predicts increased maladaptation in the genomes of sexual, outbred populations, whereas the genomes of asexual or inbred populations should be less afflicted with the mutationally decayed remains of robust networks.

The endpoint of robustness evolution is predicted to be a decreasing function of the selection coefficient associated with deleterious mutations. This is because as the strength of selection upon the mutating locus increases, so does the epistasis between the robustness and mutating loci, which results in a greater association between these. Since it is the build-up of this linkage disequilibrium which acts to disfavour the evolution of robustness, weaker selection against the deleterious mutant favours enhanced robustness. An analogous result emerged from Perrot et al.'s (1991) model for the evolution of diploidy. Interestingly, this is responsible for the run-away selection for robustness observed in some of the simulations (Fig. 2B), because as the population becomes more robust to deleterious

mutations there is reduced selection acting upon the mutating locus, and thus a lower build-up of linkage disequilibrium between the two loci. Intuitively, it would seem that mutational robustness should be increasingly favoured as deleterious mutations become more, rather than less, harmful. Yet it is the mutation load and not the mutation effect that is crucial (Proulx and Phillips, 2005), so the endpoint of robustness evolution is an increasing function of the mutation rate and not the deleterious mutational effect. This is consistent with Wright's (1929) view that selection for mutational robustness will be of the order of mutation rate.

It is of interest to compare the present results with previous models for the evolution of mutational robustness. Wagner et al. (1997) investigated the evolution of a modifier of mutational robustness impacting upon a number of loci underlying a quantitative trait under Gaussian stabilising selection. Individual-based simulations showed that the selection coefficient acting on the modifier tends to increase with the intensity of the stabilising selection, a result that was verified in deterministic simulations by Kawecki (2000). This is in contrast to the result we report here where the modifier invades most easily at lower strengths of selection against the deleterious allele because it is here that the modifier experiences weakest linkage disequilibrium with the deleterious allele. Wagner et al. (1997) attribute their result to stronger stabilising selection enabling stronger selection for canalization. Their result crucially depends on the possibility of back mutation, plus high mutation rates, so that back mutation is strong relative to selection. The present analysis assumes no back mutation, and this appears to be the reason for the disparity. The neglecting of back mutation seems reasonable if there are many alleles that give rise to defective gene products and only a few that code for a correctly functioning protein. However, further work is needed to clarify the impact of back mutation on the evolution of robustness.

We note some possibilities for the evolution of synergistic epistasis, where an individual's fitness declines more rapidly with increasing numbers of deleterious mutations than predicted by a multiplicative fitness scheme. In many models, mutational robustness is synonymous with synergistic epistasis (de Visser et al., 2003; Michalakis and Roze, 2004). For example the classic 'neutral network' (van Nimwegen et al., 1999) models of robustness—involving individuals with less than some threshold number of mutations having wild-type fitness, and individuals exceeding that threshold being inviable—presents an extreme form of synergistic epistasis. A substantial amount of theory has been devoted to the evolution of sex and recombination given synergistic epistasis between deleterious mutations (Kimura and Maruyama, 1966; Kondrashov, 1988; Charlesworth, 1990). Inferring from the present analysis, we suggest that synergistic epistasis can be an evolutionary outcome of sex and recombination, insofar as the latter processes promote the evolution of robustness, and synergistic epistasis emerges as a consequence. This hypothesis has some empirical support—there is a general trend towards weak synergistic epistasis between deleterious mutations among eukaryotes, but no trend in prokaryotes (de Visser et al., 1997; Elena and Lenski, 1997; de Visser and Hoekstra, 1998; Elena, 1999; Burch and Chao, 2004). This is beyond the scope of the present analysis, which has maximally one deleterious mutation in each individual, though it presents an interesting problem for the future.

Currently, no convincing empirical evidence has been published that demonstrates that genetic robustness exists as an adaptation. One reason for this is that, while it is possible to demonstrate that heritable variation is buffered in particular organisms, it is not easy to determine whether genetic robustness is the primary function, merely a side-effect of evolution for environmental robustness (Rutherford and Lindquist, 1998; Ancel and Fontana, 2000; Queitsch et al., 2002; Burch and Chao, 2004), or perhaps simply an emergent property of gene networks (Kacser and Burns, 1981; von Dassow et al., 2000; Edelman and Gally, 2001; Meir et al., 2002; Shen-Orr et al., 2002). A closely related problem is that the selection coefficient for a modifier of genetic robustness will be very weak, typically of the order of the mutation rate itself. However, the evolution of genetic robustness as a primary function may be plausible if there is migration between subpopulations in a heterogeneous environment (Mayr, 1963; Otto and Bourguet, 1999; Stearns, 2002). Migration rates can be much higher than mutation rates and therefore provide a stronger selective pressure for the buffering of (locally) maladapted alleles. Additionally, genetic robustness may evolve when selection fluctuates over time (Kawecki, 2000) and when selective sweeps take a population out of equilibrium (Mayo and Burger, 1997). It is with a view to extending the analysis to more complicated multilocus models that we have employed the methodology of Kirkpatrick et al. (2002), which permits arbitrary complexity within a single notational framework.

Acknowledgements

We thank J.-B. Andre, H. Bagheri, N. Barton, A. de Cara, T. Johnson, M. Kirkpatrick, J. Taylor, S. West and an anonymous reviewer for discussion and comments, plus BBSRC and NSERC for funding.

References

Ancel, L.W., Fontana, W., 2000. Plasticity, evolvability and modularity in RNA. J. Exp. Zool. 288, 242–283.

Bagheri, H.C., Wagner, G.P., 2004. Evolution of dominance in metabolic pathways. Genetics 168, 1713–1735.

Barton, N.H., Turelli, M., 1991. Natural and sexual selection on many loci. Genetics 127, 229–255.

Burch, C.L., Chao, L., 2004. Epistasis and its relationship to canalization in RNA virus phi 6. Genetics 167, 559–567.

Charlesworth, B., 1990. Mutation–selection balance and the evolutionary advantage of sex and recombination. Genet Res. 55, 199–221.

Christiansen, F.B., 1991. On conditions for evolutionary stability for a continuously varying character. Am. Nat. 138, 37–50.

Crow, J.F., Kimura, M., 1965. Evolution in sexual and asexual populations. Am. Nat. 99, 439–450.

Crow, J.F., Kimura, M., 1970. An Introduction to population genetics theory. Burgess Publishing Company, Minneapolis, Minnesota.

Dawson, K.J., 1999. The dynamics of infinitesimally rare alleles applied to the evolution of mutation rates and the expression of deleterious mutations. Theor. Popul. Biol. 55, 1–22.

de Visser, J.A.G.M, Hoekstra, R.F., 1998. Synergistic epistasis between loci affecting fitness: evidence in plants and fungi. Genet. Res. 71, 39–49.

de Visser, J.A.G.M., Hoekstra, R.F., van den Ende, H., 1997. Test of interaction between genetic markers that affect fitness in *Aspergillus niger*. Evolution 51, 1499–1505.

de Visser, J.A.G.M., Hermisson, J., Wagner, G.P., Ancel-Meyers, L., Bagheri-Chaichian, H., Blanchard, J.L., Chao, L., Cheverud, J.M., Elena, S.F., Fontana, W., Gibson, G., Hansen, T.F., Krakauer, D., Lewontin, R.C., Ofria, C., Rice, S.H., von Dassow, G., Wagner, A., Whitlock, M.C., 2003. Perspective: evolution and detection of genetic robustness. Evolution 57, 1959–1972.

Edelman, G.M., Gally, J.A., 2001. Degeneracy and complexity in biological systems. Proc. Natl Acad. Sci. USA 98, 13763–13768.

Edwards, J.S., Palsson, B.O., 1999. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. J. Biol. Chem. 274, 17410–17416.

Edwards, J.S., Palsson, B.O., 2000a. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc. Natl Acad. Sci. USA 97, 5528–5533.

Edwards, J.S., Palsson, B.O., 2000b. Robustness analysis of the *E. coli* metabolic network. Biotech. Progress 16, 927–939.

Elena, S.F., 1999. Little evidence for synergism among deleterious mutations in a nonsegmented RNA virus. J. Mol. Evol. 49, 703–707.

Elena, S.F., Lenski, R.E., 1997. Test of interaction between deleterious mutations in *E. coli*. Nature 390, 395–398.

Eshel, I., 1983. Evolutionary and continuous stability. J. Theor. Biol. 103, 99–111.

Eshel, I., Motro, U., 1981. Kin selection and strong evolutionary stability of mutual help. Theor. Popul. Biol. 19, 420–433.

Fares, M.A., Ruiz-Gonzales, M.X., Moya, A., Elena, S.F., Barrio, E., 2002. GroEL buffers against deleterious mutations. Nature 417, 398.

Fisher, R.A., 1928. The possible modification of the response of the wild type to recurrent mutations. Am. Nat. 62, 115–126.

Frank, S.A., 2003. Genetic variation of polygenic characters and the evolution of genetic degeneracy. J. Evol. Biol. 16, 138–142.

Haldane, J.B.S., 1937. The effect of variation on fitness. Am. Nat. 71, 337–349.

Haldane, J.B.S., 1956. The theory of selection for melanism in *Lepidoptera*. Proc. R. Soc. London B 145, 303–306.

Hermisson, J., Redner, O., Wagner, H., Baake, E., 2002. Mutation–selection balance: ancestry, load, and maximum principle. Theor. Popul. Biol. 62, 9–46.

Ingolia, N.T., 2004. Topology and robustness in the *Drosophila* segment polarity network. PLoS Biol. 2, 805–815.

Kacser, H., Burns, J.A., 1981. The molecular basis of dominance. Genetics 97, 639–666.

Kawecki, T.J., 2000. The evolution of genetic canalization under fluctuating selection. Evolution 54, 1–12.

Kimura, M., Maruyama, T., 1966. The mutational load with epistatic gene interactions in fitness. Genetics 54, 1337–1351.

Kirkpatrick, M., Johnson, T., Barton, N.H., 2002. General models of multilocus evolution. Genetics 161, 1727–1750.

Kondrashov, A.S., 1988. Deleterious mutations and the evolution of sexual reproduction. Nature 336, 435–440.

Kondrashov, A.S., Crow, J.F., 1991. Haploidy or diploidy: which is better? Nature 351, 314–315.

Lau, K.F., Dill, K.A., 1990. Theory for protein mutability and biogenesis. Proc. Natl Acad. Sci. USA 87, 638–642.

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

Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge.

Maynard Smith, J., Price, G., 1973. The logic of animal conflict. Nature 246, 15–18.

Mayo, O., Burger, R., 1997. The evolution of dominance: a theory whose time has passed? Biol. Rev. 72, 97–110.

Mayr, E., 1963. Animal Species and Evolution. Harvard University Press, Cambridge, MA.

Meir, E., von Dassow, G., Munro, E., Odell, G.M., 2002. Robustness, flexibility and the role of lateral inhibition in the neurogenic network. Curr. Biol. 12, 778–786.

Michalakis, Y., Roze, D., 2004. Epistasis in RNA viruses. Science 306, 1492–1493.

Orr, H.A., 1991. A test of Fisher's theory of dominance. Proc. Natl Acad. Sci. USA 88, 11413–11415.

Otto, S.P., Bourguet, D., 1999. Balanced polymorphisms and the evolution of dominance. Am. Nat. 153, 561–574.

Pen, I., 2000. Reproductive effort in viscous populations. Evolution 54, 293–297.

Perrot, V., Richerd, S., Valero, M., 1991. Transition from haploidy to diploidy. Nature 351, 315–317.

Proulx, S.R., Phillips, P.C., 2005. The opportunity for canalization and the evolution of genetic networks. Am. Nat. 165, 147–162.

Queitsch, C., Sangster, T.A., Lindquist, S., 2002. Hsp90 as a capacitor of phenotypic variation. Nature 417, 618–624.

Rutherford, S.L., Lindquist, S., 1998. Hsp90 as a capacitor for morphological evolution. Nature 396, 336–342.

Schmalhausen, I.I., 1949. Factors of Evolution: The Theory of Stabilising Selection. Blakiston, Philadelphia, PA, USA.

Schuster, P., Fontana, W., Stadler, P.F., Hofacker, I., 1994. From sequences to shapes and back: A case study in RNA secondary structures. Proc. R. Soc. London B 255, 279–284.

Shen-Orr, S.S., Milo, R., Mangan, S., Alon, U., 2002. Network motifs in the transcriptional regulation of *E. colii*. Nat. Genet. 31, 64–68.

Stearns, S.C., 2002. Progress on canalization. Proc. Natl Acad. Sci. USA 99, 10229–10230.

Taylor, P.D., 1996. Inclusive fitness arguments in genetic models of behaviour. J. Math Biol. 34, 654–674.

van Nimwegen, E., Crutchfield, J.P., Huynen, M., 1999. Neutral evolution of mutational robustness. Proc. Natl Acad. Sci. USA 96, 9716–9720.

von Dassow, G.E., Meir, E., Munrow, E.M., Odell, G.M., 2000. The segment polarity network is a robust developmental module. Nature 406, 188–192.

Waddington, C.H., 1940. Organisers and Genes. Cambridge University Press, Cambridge.

Waddington, C.H., 1942. Canalization of development and the inheritance of acquired characters. Nature 150, 563–565.

Wagner, A., 2000. Robustness against mutations in genetic networks of yeast. Nat. Genet. 24, 355–361.

Wagner, A., 2001. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. Mol. Biol. Evol. 18, 1283–1292.

Wagner, A., 2005. Distributed robustness versus redundancy as causes of mutational robustness. BioEssays 27, 176–188.

Wagner, G.P., Booth, G., Bagheri-Chaichian, H., 1997. A population genetic theory of canalization. Evolution 51, 329–347.

Wilke, C.O., 2001. Selection for fitness versus selection for robustness in RNA secondary structure folding. Evolution 55, 2412–2420.

Wilke, C.O., Adami, C., 2003. Evolution of mutational robustness. Mut. Res. 522, 3–11.

Wilke, C.O., Wang, J.L., Ofria, C., Lenski, R.E., Adami, C., 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature 412, 331–333.

Wilkins, A.S., 1997. Canalization: a molecular genetic perspective. BioEssays 19, 257–261.

Wright, S., 1929. Fisher's theory of dominance. Am. Nat. 68, 24–53.

Wright, S., 1934. Physiological and evolutionary theories of dominance. Am. Nat. 68, 25–53.