Quantitative Traits and Phenotypic Interference

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Daniel Klemmer

aus Bonn

Köln, 2018

Berichterstatter:	Prof. Dr. Michael Lässig
	Prof. Dr. Thomas Wiehe
Tag der mündlichen Prüfung:	25.10.2018

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisor, Professor M. Lässig, for providing me with the opportunity to work on the fascinating research projects culminating in this thesis. Additionally, I would like to thank Christa Stitz for administrative support. I am also thankful for everyone in the groups of M. Lässig and J. Berg for all discussions during and after lunch, and coffee breaks. This research was funded by the DFG through grants SFB 680 and 1310. I would like to extend thanks to the CHEOPS team of the RRZK for technical support and computation time. Finally, I would like to thank my girlfriend and my family for continued support during all the years.

Kurzzusammenfassung

Die Theorie der Evolution quantitativer Merkmale erweckt in den letzten Jahren großes Interesse, da experimentelle Daten deutlich einfacher verfügbar sind. Dadurch, dass experimentelle Methoden immer mehr in der Lage sind systembiologische Datenmengen zu messen, muss die Theorie zur Beschreibung quantitativer Merkmale über die bis jetzt bestehende Theorie zur unabhängigen Evolution quantitativer Merkmale hinausgehen. Diese Thesis untersucht die asexuelle Evolution quantitativer Merkmale, sowohl im kleinen Rahmen einzelner Merkmale, die von externer Adaptation in genetisch verbundenen Teilen des Genoms, als auch im großen, systembiologischen Rahmen des gesamten Organismus mit vielen verbundenen quantitativen Merkmalen. Die Statistik eines quantitativem Merkmals wird durch externe Kopplungseffekte negativ beeinflusst, wobei dieser Effekt durch einen einzigen Parameter beschrieben werden kann, die Neutralitätsschwelle. Für den systembiologischen Anwendungsbereich untersuchen wir den Einfluss der gemeinsamen genomischen Kopplung auf Merkmale mit biophysikalischen Fitnesslandschaften. Das überraschende Ergebnis dieser phänotypischen Interferenz ist ein schneller als linearer Anstieg des genetischen Loads mit der Anzahl der koevolvierenden Phänotypen. Als Konsequenz daraus ergibt sich eine biophysikalisch fundierte Erklärung für die Evolution sexueller Reproduktion. Die negativen Auswirkungen der Kopplung auf Adaptation und Anpassungsfähigkeit verbinden diese Themen mit dem Thema Mendelscher Merkmale in asexuellen Genomen - mit dem Schwerpunkt auf die verminderte Fähigkeit, Informationen in großen, sich schnell anpassenden Genomen zu speichern. Hier finden wir eine neue Klasse von punktuierten Fitnesswellen, bei denen Perioden von Aufrechterhaltung der Funktion durch Anpassungsimpulse unterbrochen werden.

Abstract

The theory of the evolution of quantitative traits is a topic, which has generated a lot of interest in recent years due to experimental data becoming much more easily available. With experimental methods increasingly able to measure systemsbiological amounts of data, theory needs to go beyond the so far established independent evolution of quantitative traits. This thesis investigates the asexual evolution of quantitative traits in two scopes. In the small scope we look at a single quantitative trait being influenced by external adaptation in genetically linked parts of the genome. In the grand, systems-biology scope we look at the whole organism with multiple linked quantitative traits. The statistics of a quantitative trait are negatively influenced by external linkage effects, where the effect can be described by a single parameter, the neutrality threshold. For the systemsbiological scope we study the effect of joint linkage on traits with biophysical fitness landscapes. The surprising result, due to this phenotypic interference, is a faster-than-linear increase of the genetic load with the number of co-evolving phenotypes. As a consequence, this provides us with a biophysically grounded explanation for the evolution of sex. The negative effects of linkage on adaptation and adaptability link these topics together with a chapter on Mendelian traits in asexual genomes. There we focus on the decreased ability to retain information in large, fast adapting genomes. Additionally, we find a new class of punctuated fitness waves, where periods of maintainance of function are interrupted by bursts of adaptation.

Collaborations

The chapters 4 and 5 have been completed in collaboration with colleagues from the institute.

The chapter 4 on the interference of many quantitative traits has been done in collaboration with Torsten Held. The theory has been developed with equal contribution. The chapter 5 on intermittent fitness wave theory has been done in collaboration with Torsten Held, Simone Pompei and Fernanda Pinheiro. I did all the numerical simulations, while the data analysis of the simulations has been done by Torsten Held, Simone Pompei and Fernanda Pinheiro. The analytical theory was developed jointly.

Preprint

The content of chapter 4 is available as a preprint [Held et al., 2018].

1 Introduction

> Albert grunted. "Do you know what happens to lads who ask too many questions?"
> Mort thought for a moment. "No," he said eventually, "what?"
> There was silence.
> Then Albert straightened up and said, "Damned if I know. Probably they get answers, and serve 'em right."
>
> *(Terry Pratchett, Mort)*

Evolutionary theory has moved in recent years from merely trying to explain existing genetic variation and the fossil record, to trying to anticipate and predict future evolutionary changes. The basis for this major paradigm shift lies both in the advancement of theoretical frameworks and the availability of experimental data in a scope that was unimaginable before the availability of whole genome sequencing techniques. The ability to be able to forecast evolution has far reaching implications for human health.

The theoretical frameworks underlying this achievement have come a long way since the beginning of population genetics describing the effect and probability of independent mutations[Fisher, 1930; Wright, 1931; Kimura, 1962]. As the evolution of human pathogens is often asexual, this assumption of independent mutations does not hold in general. Asexual organisms reproduce under complete genetic linkage, which means that mutations occurring in different individuals have to compete with each other for fixation. This introduces long-range interactions on the genome throughout the evolutionary process. Furthermore, pathogens usually do not evolve in a constant environment. The co-evolution of pathogen-host systems

introduces time-dependent selection to these systems through the immune-system adapting to the pathogen. The evolution of drug resistance of bacteria is another case of adaptation to an external pressure [Lukačišinová and Bollenbach, 2017].

Many of the properties needed for the function of a bacterium or a virus are not simple on-or-off type traits described by a single gene. Drug resistance or the binding of a surface protein to attack cells are described by a multitude of genetic sites working in concert. This is called a quantitative trait. The evolution of quantitative traits has come a long way from merely being a phenomenological description tool to explain the adaptation towards an optimum, in the classical geometric model by Fisher[Fisher, 1930]. Recent breakthroughs have been able to describe the evolution of quantitative traits without all the microscopic properties from the sequence level. On the phenotypic level it is now possible to estimate the effect of stabilizing selection [Nour Mohammad et al., 2013] and adaptation [Held et al., 2014] of the trait by measuring a few key trait properties.

These theories have so far been quite distinct from each other: Quantitative traits were studied in isolation and asexual theory was mostly concerned with the competition of mutations.

The first models to predict future short term changes have been applied with impressive accuracy to the evolution of Influenza [Luksza and Lässig, 2014; Morris et al., 2018]. To further improve these predictive models it is important to improve the underlying evolutionary models. This work brings together these long standing, independent research areas within population genetics. It combines the theory of fitness waves [Desai and Fisher, 2007; Rouzine et al., 2008; Hallatschek, 2011; Good et al., 2012; Neher and Hallatschek, 2013; Neher et al., 2013] with a genomic theory, quantitative traits with fast adaptation in a genomic asexual context, and fitness waves with many linked quantitative traits.

We aim to provide answers to the following open questions:

1. How are the statistics of quantitative traits influenced by adaptation in other parts of the genome?

2. What are the consequences of the asexual co-evolution of a large number of genes?

3. More broadly: How does the evolutionary process influence the genome?

What modes of evolution are viable for long-term evolution?

We address these key questions individually in the following chapters using methods of statistical physics to obtain analytical results and verify our model predictions with simulations.

Chapter 2 introduces key concepts of evolutionary biology. Starting with the basic evolutionary forces, we then build up to the evolution of quantitative traits and fitness wave theory needed in the later chapters.

In chapter 3 quantitative traits, which have so far been studied independently of any genomic background, are studied under the influence of selective sweeps and thus in an asexual genomic context. We start by recapitulating the Fokker-Planck equations describing the evolution of a single quantitative trait and add the effect of selective sweeps as an additional force on the trait.

Directly based on this, the next chapter 4 takes this one step further and deals with a minimal model of the joined evolution of many linked quantitative traits. These quantitative traits describe genes that need a folded protein to function, giving rise to a biophysical fitness landscape that is proportional to the folding probability, which is given by a Fermi function. In this chapter we see the well known property in statistical mechanics that the joint behavior of many noisy observables is much easier than studying the interacting behavior of just a few noisy observables. The total fitness behaves quite deterministically according to fitness wave theory, leading to a simple scaling law for the genetic load.

Chapter 5 then links the fitness wave theory, used in chapter 4, to Mendelian genomic loci, i.e. simple loci with an additive fitness landscape only linked through the evolutionary dynamics. There we observe a new evolutionary mode, punctuated fitness waves, giving an anomalous scaling behavior. Also we show, that fitness wave behavior leads to an unstable genome that cannot efficiently maintain information in the long-term.

All chapters are accompanied by simulations to test theory predictions, which are described in detail in the methods of the specific chapters.

2 Introduction to evolutionary theory

> One of the things Ford Prefect had always found hardest to understand about humans was their habit of continually stating and repeating the very very obvious.
>
> *(Douglas Adams, The Hitchhiker's Guide to the Galaxy)*

As one of the most complex systems known in nature, the evolution of species offers a fascinating playground to population geneticists. In this chapter we recapitulate the main concepts of population genetics needed to understand the following chapters. The evolutionary forces can be categorized in stochastic and deterministic forces. Let us start with the most well known aspect of evolution: 'Survival of the fittest', i.e. natural selection.

2.1 Natural selection

The concept of natural selection, which was first formulated by Darwin and Wallace, is by now a well established concept of general knowledge. Its structured mathematical formulation also goes back to the beginning of the last century with Fisher and Wright [Fisher, 1930; Wright, 1931]. Mathematically, natural selection is a deterministic force that can be associated with a potential, the fitness landscape. The fitness landscape can either be a map directly from the genotype, i.e. the most basic description of an organisms information content, or from the phenotype, which is some mapping of the genomic information content to actual biological quantities. The genotypic fitness landscape was first modelled by

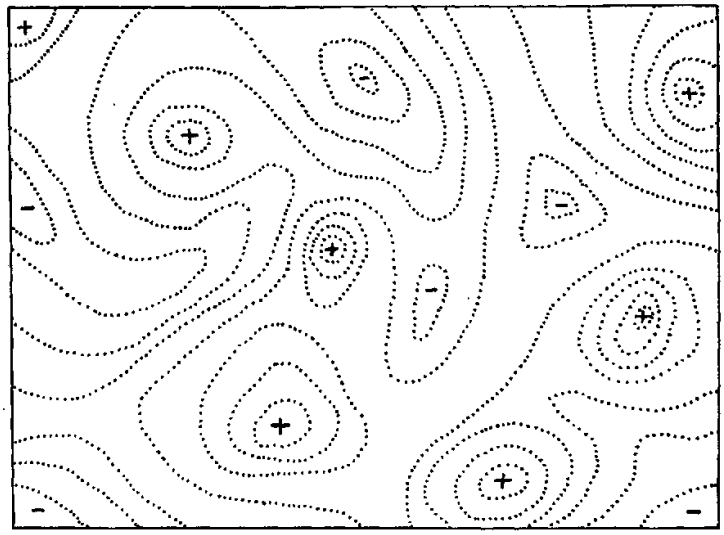

Figure 2.1: **Diagrammatic fitnesslandscape by Wright.** 2-dimensional representation of multidimensional genotype space, where plus and minus signs mark fitness peaks and valleys respectively. Adapted from [Wright, 1932].

Wright [Wright, 1932], while a phenotypic fitness landscape was proposed by Fisher with his geometric model [Fisher, 1930]. In this thesis we are mostly concerned with the latter - Evolution rarely acts directly on the actual genetic basis and is mostly concerned with the functional effect of the genetic information, when it is transcribed into phenotypes.

For a simplified case of two possible types of individuals A and B with associated fitnesses f_A and f_B, the number of individuals $N_i,\ i \in A, B$ of each type changes deterministically according to

$$\frac{d}{dt}N_i(t) = f_i N_i(t),\ i \in A, B. \tag{2.1}$$

The change of the fraction $x_i = N_i/N$, called *frequency*, where N is the total number of individuals, changes accordingly with

$$\begin{aligned} \frac{d}{dt}x_A(t) &= (f_A - f_B)x_A(t)(1 - x_A(t)) \\ &= s_{AB}x_A(t)(1 - x_A(t)). \end{aligned} \tag{2.2}$$

For natural selection to matter, phenotypic diversity is a necessary condition. The best / most fit individuals can only be selected, if there is actually a variety of differences. Furthermore, natural selection actually leads to a decrease of this variation up to the inevitable conclusion of only the fittest remaining. This is in direct conflict with our observation of nature, where we see an abundance of variation. The counteracting force replenishing variation is mutation.

2.2 Mutation

Mutations change the genetic information from one generation to the next due to random copy errors. In spite of this randomness the mutational force from a population genetics perspective is actually also deterministic. Depending on the assumptions we are also able to write down a mutation potential. For the single locus model with two alleles shown as our example in the previous section mutations change the frequency of allele A according to

$$\frac{d}{dt}x_A(t) = \mu(1 - 2x_A(t)). \tag{2.3}$$

Adding to these two deterministic forces, an entropic force associated with random offspring numbers called *genetic drift* acts on the population genetics level.

2.3 Genetic drift

Evolution due to natural selection is not a deterministic process. Randomness enters through the variability in the offspring numbers. This process, which can lead to unfavourable mutations being fixed in a population is called *genetic drift*. For the single locus dynamics, genetic drift leads to a Langevin equation given by

$$\frac{d}{dt}x_A(t) = \eta(t), \tag{2.4}$$

with Gaussian noise $\eta(t)$ describing the genetic drift with $\langle\eta(t)\rangle = 0$ and $\langle\eta(t)\eta(t')\rangle = \frac{1}{N}\delta(t - t')$. Though the name might imply this, it is not a drift in the physical sense, but for population genetics it is associated with a diffusion in probability space. The joint dynamics under these three evolutionary forces leads to a joint

Langevin equation. An equivalent description via a Fokker-Planck equation is given in the next section.

2.4 Fokker-Planck equation of a single locus

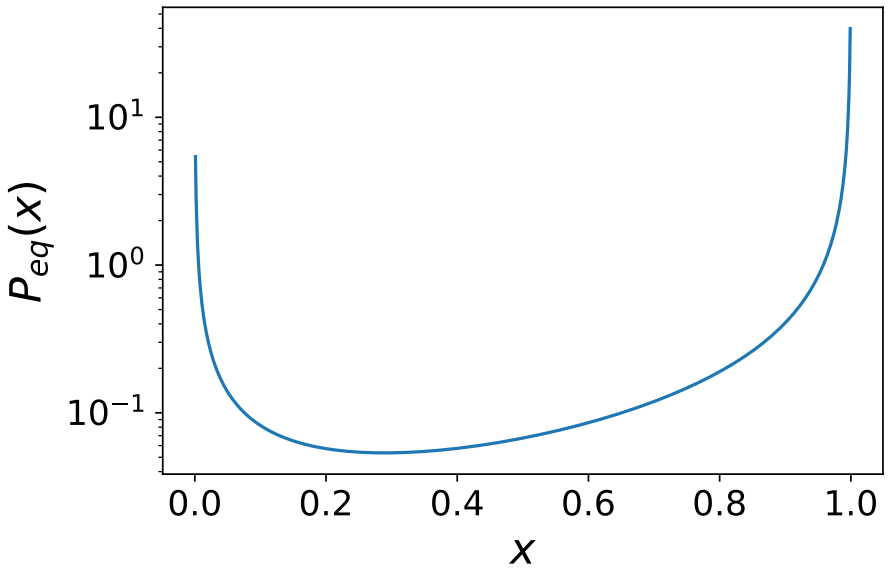

Figure 2.2: **Kimura-Ohta allele frequency spectrum.** Probability distribution of allele frequencies x for a single genetic locus with two alleles.

The evolution of a single genetic locus with two possible values - alleles- can be described by a diffusion approximation[Kimura, 1964]. Here the fraction–called frequency– of the allele with selective advantage s is denoted by x, which leads to a frequency for the deleterious allele of $1 - x$. The population is assumed to have a constant population size N and a uniform mutation rate μ. The distribution of allele frequencies is then given by

$$\partial_t P(x,t) = \frac{1}{2N}\partial_x^2 \left[x(1-x)P(x,t)\right] - \partial_x \left[(\mu(1-2x)+s)P(x,t)\right]. \tag{2.5}$$

The equilibrium distribution of this equation gives the well-known Kimura-Ohta shape for the allele frequencies with directed selection

$$P_{\text{eq}}(x) = \mathcal{N} \times (x(1-x))^{-1+2N\mu} e^{2Nsx}, \tag{2.6}$$

where $\mathcal{N}$ is a normalization factor[Rouzine et al., 2001], see Fig. 2.2.

The probability that a single mutation with selective advantage s becomes present in all individuals is given by the fixation probability[Kimura, 1962]

$$G(s) = \frac{1 - \exp(-2s)}{1 - \exp(-2Ns)}. \tag{2.7}$$

The single locus dynamics is of course only an approximation of anything resembling real organisms. This approximation is valid for low mutation rates or under full recombination, where correlations between sites and mutations can be neglected.

2.5 Asexual evolution

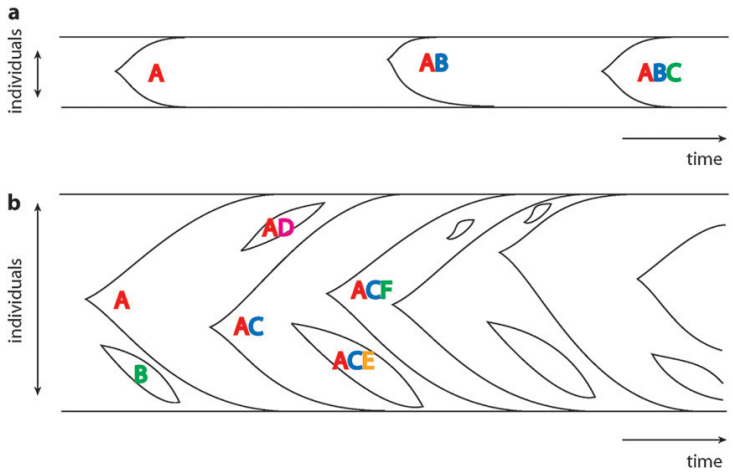

Figure 2.3: **Sequential evolution and competition of mutations.** **a** shows the sequential fixation of mutations without interference. This means single locus theory can be used and the classical Kimura fixation formula is valid, see Eq. 2.7. **b** shows a competition between mutations overlapping in time - here theory needs to take asexual interference effects into account. Adapted from [Desai and Fisher, 2007]

.

In general, for asexual evolution the assumption of neglecting correlations, i.e.

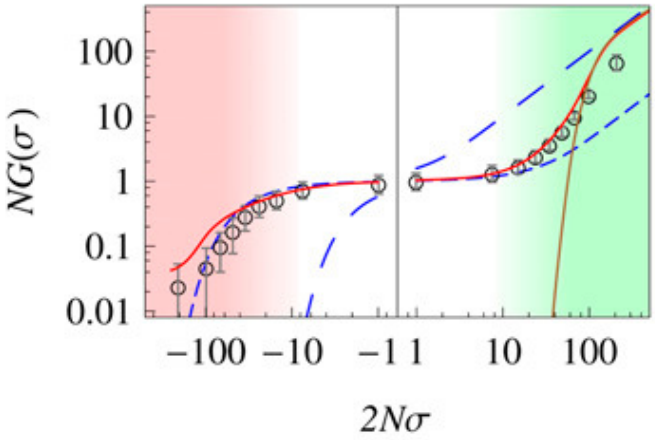

Figure 2.4: **Emergent neutrality.**Fixation probability under clonal interference showing emergent neutrality with respect to Kimura / Haldane fixation probability. Adapted from [Schiffels et al., 2011].

linkage disequilibrium, between mutations and genomic sites is no longer possible. As a result the dynamics depends on the interplay of mutations with each other. A general effect of this is a decrease in the speed of adaptation due to competition between multiple beneficial mutations. Both for fitness wave theory[Good et al., 2012] and under clonal interference[Schiffels et al., 2011] in a genomic context it has been shown that moderately strongly selected mutations act effectively neutral, see Fig. 2.4.

Fixation probability

The fixation probability of a single mutation with selective effect s under the influence of a coalescence rate $\tilde{\sigma}$ is then given by

$$G(s,\tilde{\sigma}) = \begin{cases} \frac{1}{N}e^{s/2\tilde{\sigma}}, & |s| \ll \tilde{\sigma} \text{ or } s < -\tilde{\sigma}, \\ \frac{2s}{1-e^{s/2\tilde{\sigma}}}, & s \gg \tilde{\sigma}. \end{cases} \tag{2.8}$$

This equation was shown for fitness waves [Neher et al., 2013; Good et al., 2014] and in a hierarchical approximation with drivers [Schiffels et al., 2011].

2.6 Multi-locus Fokker-Planck equation

For many linked sites the evolution in frequency space is in a high-dimensional, under-sampled space. The frequencies of the haplotypes $\mathbf{x} = (x_1, \ldots, x_K)$, where K is the number of possible haplotypes, can again be described by a generalized Kimura diffusion equation [Ewens, 2004; Nour Mohammad et al., 2013]

$$\partial_t P(\mathbf{x}, t) = \sum_{\mathbf{a},\mathbf{b}\in\mathcal{A}} \left[\frac{1}{2N} \partial_{x^{\mathbf{a}}} \partial_{x^{\mathbf{b}}} \left(g^{\mathbf{ab}}(\mathbf{x}) P(\mathbf{x}, t) \right) - \partial_{x^{\mathbf{a}}} \left((m^{\mathbf{a}}(\mathbf{x}) + g^{\mathbf{ab}}(\mathbf{x}) s_{\mathbf{b}}(\mathbf{x})) P(\mathbf{x}, t) \right) \right] \tag{2.9}$$

Here $\mathcal{A}$ is the set of possible haplotypes,$g^{\mathbf{ab}}(\mathbf{x})$ is a metric, $m^{\mathbf{a}}(\mathbf{x})$ is associated with mutations, and $s_{\mathbf{b}}(\mathbf{x})$ is the selection coefficient given by the gradient of the fitness landscape $s_{\mathbf{b}}(\mathbf{x}) = \partial_{x^{\mathbf{b}}} F(\mathbf{x})$. This equation is under general conditions hard to solve. For the special case of a quantitative trait the evolution can be mapped onto a stochastic reaction diffusion equation (2.10).

2.7 Quantitative trait evolution

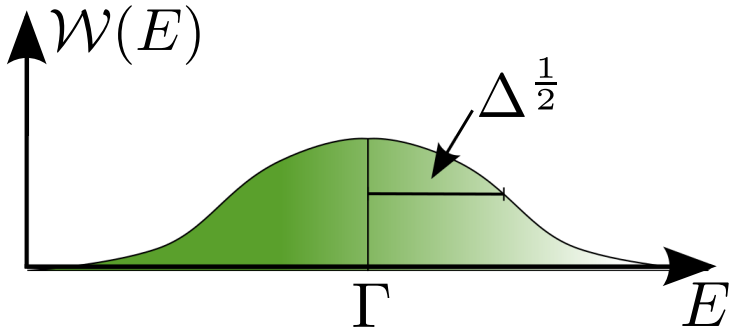

Figure 2.5: **Quantitative trait distribution.** A quantitative trait with a Gaussian distribution $\mathcal{W}(E)$ within a population with mean value Γ and diversity Δ.

A quantitative trait is an organism's trait that depends on multiple genomic loci. Thus it is able to have more than simple binary results. Any trait that has a quasi continuous range of possible values is a quantitative trait. As long as the fitness of the trait E does not change with time, the distribution of trait values

$\mathcal{W}(E,t)$within a population is given by a reaction-diffusion equation

$$\begin{aligned}\partial_t \mathcal{W}(E,t) = \left(f(E) - \bar{f}(t)\right)\mathcal{W}(E,t) + \partial_E\left(\mu_1(E)\mathcal{W}(E,t)\right) \\ + \frac{1}{2}\partial_E^2\left(\mu_2(E)\mathcal{W}(E,t)\right) + \eta(E,t)\end{aligned}, \tag{2.10}$$

where $\mu_k(E)$ is the kth moment of the distribution of mutation effects, $f(E)$ is the fitness for trait value E, and $\eta(E,t)$ is a stochastic noise term related to genetic drift. According to [Nour Mohammad et al., 2013] the between population distribution of the mean $\Gamma(t) = \overline{E} = \int dE E\mathcal{W}(E,t)$ and variance within the population $\Delta(t) = \overline{(E-\Gamma)^2} = \int dE(E-\Gamma)^2\mathcal{W}(E,t)$,see Fig. 2.5, can be described by Fokker-Planck equations

$$\frac{\partial Q(\Gamma,t)}{\partial t} = -\frac{\partial}{\partial\Gamma}\left[(m^\Gamma + g^{\Gamma\Gamma}s_\Gamma)Q(\Gamma,t)\right] + \frac{\partial^2}{\partial\Gamma^2}\left[\frac{g^{\Gamma\Gamma}}{2N}Q(\Gamma,t)\right] \tag{2.11}$$

$$\frac{\partial Q(\Delta,t)}{\partial t} = -\frac{\partial}{\partial\Delta}\left[(m^\Delta + g^{\Delta\Delta}s_\Delta)Q(\Delta,t)\right] + \frac{\partial^2}{\partial\Delta^2}\left[\frac{g^{\Delta\Delta}}{2N}Q(\Delta,t)\right]. \tag{2.12}$$

For a quantitative trait under stabilizing selection in a quadratic fitness landscape of the form $f(E) = -c_0(E-E^*)^2$ the two equations have the form

$$\frac{\partial Q(\Gamma,t)}{\partial t} = -\frac{\partial}{\partial\Gamma}\left[(-2\mu(\Gamma - \langle\Gamma\rangle_0) - 2c_0\langle\Delta\rangle(\Gamma - E^*))Q(\Gamma,t)\right] + \frac{\partial^2}{\partial\Gamma^2}\left[\frac{\langle\Delta\rangle}{2N}Q(\Gamma,t)\right] \tag{2.13}$$

$$\frac{\partial Q(\Delta,t)}{\partial t} = -\frac{\partial}{\partial\Delta}\left[(-4\mu(\Delta - E_0^2) - \frac{\Delta}{N} - 2c_0\Delta^2)Q(\Delta,t)\right] + \frac{\partial^2}{\partial\Delta^2}\left[\frac{\Delta^2}{N}Q(\Delta,t)\right], \tag{2.14}$$

where Γ_0 and E_0^2 are the neutrally expected trait mean and variance, respectively.

Mutation-selection-drift equilibrium

The marginal distributions for the mean trait and the trait diversity have well described equilibria given by

$$Q_{\text{eq}}(\Gamma) = \mathcal{N} \times Q_0(\Gamma)\exp(2NF(\Gamma)) \tag{2.15}$$

$$Q_{\text{eq}}(\Delta) = \mathcal{N} \times Q_0(\Delta)\exp(2NF(\Delta)), \tag{2.16}$$

where the neutral equilibria are changed due to selection by a Boltzmann-weight $\exp(2NF)$. The neutral equilibrium distributions are then given by

$$Q_0(\Gamma) = \mathcal{N} \times \exp\left(-\frac{(\Gamma - \Gamma_0)^2}{2E_0^2}\right) \tag{2.17}$$

$$Q_0(\Delta) = \mathcal{N} \times \Delta^{-3-4N\mu} \exp\left(-\frac{4N\mu E_0^2}{\Delta}\right). \tag{2.18}$$

2.8 Fitness wave theory

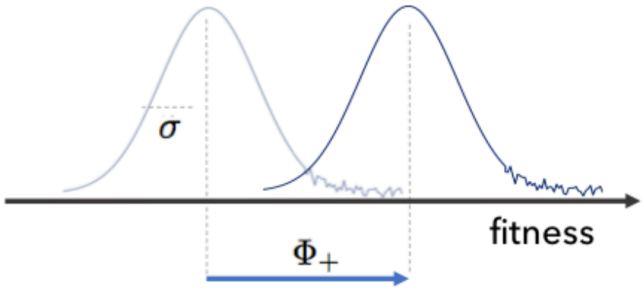

Figure 2.6: **Fitness wave schematic.** Schematic of the evolution according to fitness wave theory. The bulk is gaussian and moves deterministically, while the tip behaves stochastically.

Since fitness depends on more than a single genetic locus, fitness is just another example of a quantitative trait. Fitness wave theory uses this to explain the adaptive evolution in fitness space as a gaussian wave with a stochastic tip[Cohen et al., 2005; Desai and Fisher, 2007; Rouzine et al., 2008; Hallatschek, 2011]. The time evolution equation for the trait value can in this case be simplified to[Good and Desai, 2014]

$$\begin{aligned}\partial_t \mathcal{W}(E,t) = &\left(E - \bar{E}(t)\right)\mathcal{W}(E,t) + \partial_E\left(\mu_1 \mathcal{W}(E,t)\right) \\ &+ \frac{1}{2}\partial_E^2\left(\mu_2 \mathcal{W}(E,t)\right) + \eta(E,t)\end{aligned}. \tag{2.19}$$

A standard assumption of fitness wave theory is a fixed distribution of selection effects, which is in many classical theories just a δ-distribution with a single selection coefficient s. Travelling fitness wave theory is a universal mode of

evolution independent of the actual genomic mechanism; only dependent on the distribution of fitness effects of the mutations.

An important result is that the fitness variance in the population σ^2 is stable and can be related to the inverse coalescence time $\tilde{\sigma}^2$ by

$$\sigma^2 = c\tilde{\sigma}^2, \tag{2.20}$$

where $c = c_0 \log(N\tilde{\sigma})$ is a slowly varying parameter. The actual value of c and the logarithmic dependence on the parameters is dependent on the used model [Neher and Hallatschek, 2013; Neher et al., 2013].

The genealogy associated with fitness wave theory belongs to the universality class of Bolthausen-Sznitman coalescent processes [Brunet et al., 2006; Neher and Hallatschek, 2013]. In this dissertation we will demonstrate the limits and consequences of this mode of evolution to the genomic state of organisms. Further we will show that assuming an underlying genotype to phenotype mapping a standing fitness wave is a generic outcome of the asexual evolutionary process.

In the following chapter we take a closer look at quantitative traits in asexual populations. This will then in later chapters be combined with results from fitness wave theory.

3 Quantitative traits in asexual populations

"No, no! The adventures first, explanations take such a dreadful time."

(Lewis Carroll, Alice's Adventures in Wonderland & Through the Looking-Glass)

QUANTITATIVE traits are ubiquitous in nature. Any measureable biological quantity that is not just defined by a single on-off type binary choice is the result of many genetic loci working in concert to provide the existing variability. The empirical analysis of these phenotypes was first possible with QTL analysis in the 1980s, which is reviewed in[Tanksley, 1993; Mackay, 2001]. Necessarily, many common diseases are associated with quantitative traits [Plomin et al., 2009] and as such the theory of quantitative traits is also of central relevance in the medical field. This is further amplified by the fact, that surface proteins binding to cells for viruses and drug resistance of microbes are also described by quantitative traits [Lukačišinová and Bollenbach, 2017]. With the development of experimental methods and availability of data, a demand for theoretical predictions arose to compare to the observed phenomena. Classical studies of quantitative genetics focused on phenomenological descriptions of trait distributions of quantitative traits [Bulmer, 1972; Barton and Turelli, 1989; Rice, 1990; Hartl and Taubes, 1996; Lynch and Walsh, 1998], often based on considerations first developed by Fisher [Fisher, 1930] with his geometric model, which has since become a standard model for the phenotypic evolution [Tenaillon, 2014]. Even though a direct link to the genomic basis is not part of the framework of Fisher's geometric model, many recent studies have used it to describe epistatic interactions [Martin

et al., 2007; Weinreich and Knies, 2013; Perfeito et al., 2013; Schoustra et al., 2016]. Furthermore, the mapping of a phenotypic landscape to rugged genotypic landscapes has led to a deeper understanding of the connection to a genomic basis [Hwang et al., 2017]. Whereas in the framework of FGM the dimensionality of interest is usually high, a lot of insight on quantitative traits can already be found from one dimensional single peaked landscapes. The evolution of quantitative traits under stabilizing selection can be approximated by a quadratic one-dimensional fitness landscape [Lande, 1976]. Further the evolution was mostly centered near the fitness peak, where sign-epistasis is of major concern. This is due to a generally directionless mutation approximation, which a direct mapping of the genomic evolution to phenotypic evolution as was done in [Nour Mohammad et al., 2013] rectifies. The evolution is then necessarily on the flank of the fitness landscape.

Up until recently evolutionary theory for quantitative traits rested on the assumption that the quantitative trait loci are in or near linkage equilibrium[Barton and de Vladar, 2009; de Vladar and Barton, 2011] to provide analytically solvable models. Recent studies have shown that the evolution of molecular phenotypes can be solved independently from the genetic basis, giving rise to universal behaviour[Nourmohammad et al., 2013] just depending on the first and second moment of the distribution of mutation effects. Even though the evolution of quantitative traits under asexual evolution [Nour Mohammad et al., 2013] and adaptive evolution of the trait itself[Held et al., 2014] has been solved and the influence of selective sweeps within in the trait loci has been studied[Chevin and Hospital, 2008], a missing link remains the influence of asexual evolution on the quantitative trait due to linkage with further parts of the genome. Asexual evolution can give rise to modes of evolution, where the fate of new mutations is strongly dependent on its genomic background(*background selection, fitness wave*)[Desai and Fisher, 2007; Charlesworth, 2012], subsequent helping (*hitchhiking*)[Hill and Robertson, 1966], or competing mutations (*clonal interference*)[Gerrish and Lenski, 1998; Park and Krug, 2007]. A consequence is the emergence of effective neutrality [Schiffels et al., 2011; Good et al., 2012], which randomizes the behaviour of all but the most strongly selected mutations. The regime of rapid adaptation, where clonal interference plays a dominant role in the evolution of a population, has been shown to occur both in evolution experiments[Barroso-Batista et al., 2014; McDonald et al., 2016] and natural populations such as the adaptation of the human

influenza[Strelkowa and Lässig, 2012]. The influence of this strong interference selection on quantitative traits is the subject of this chapter. This is the first statistical theory of a non-neutral quantitative trait under complete linkage. We show that the quantitative trait can be described by the diffusion model introduced by[Nour Mohammad et al., 2013] with a reduced effective population size. Thus for asexual organisms hitchhiking events due to rapid adaptations affect the statistics of quantitative traits under stablizing selection negatively.

3.1 Genetic Draft on Quantitative Traits

We are interested in the influence of adaption in other parts of the genome on a quantitative trait in asexual organisms. For this we study the effects of hitchhiking due to selective sweeps on quantitative traits.

3.1.1 Trait statistics

A quantitative trait is described by its trait value E, which we assume to have an additive map of genotype to phenotype, i.e. $E(\mathbf{a}) = \sum_{i=1}^{L} \epsilon_i a_i$, where each quantitative trait's sequence is given by $\mathbf{a} = (a_1, \ldots, a_L)$ with ϵ_i being the phenotypic effect of site i and $a_i = 0; 1$ described by a binary alphabet. The statistics of the phenotype can be fully described by its trait mean Γ and diversity Δ, if the trait value distribution is Gaussian. For a large number of constituent sites the trait will generally be polymorphic, since $LN\mu \gtrsim 1$. The assumption of a Gaussian distribution of trait values for quantitative traits under stabilizing selection under this condition is due to the large number of trait loci a good discription of the actual distribution.

The evolution of quantitative traits can be described by a diffusion model for the probability distribution between populations of the trait mean Γ and the trait diversity Δ. This description is valid both in the case of a quantitative trait under constant stabilizing selection[Nour Mohammad et al., 2013] and when the quantitative trait itself is undergoing adaptive changes[Held et al., 2014]. The

diffusion equations are given by

$$\frac{\partial Q(\Gamma,t)}{\partial t} = -\frac{\partial}{\partial \Gamma}\left[(m^{\Gamma} + g^{\Gamma\Gamma}s_{\Gamma})Q(\Gamma,t)\right] + \frac{\partial^2}{\partial \Gamma^2}\left[\frac{g^{\Gamma\Gamma}}{2N}Q(\Gamma,t)\right] \tag{3.1}$$

$$\frac{\partial Q(\Delta,t)}{\partial t} = -\frac{\partial}{\partial \Delta}\left[(m^{\Delta} + g^{\Delta\Delta}s_{\Delta})Q(\Delta,t)\right] + \frac{\partial^2}{\partial \Delta^2}\left[\frac{g^{\Delta\Delta}}{2N}Q(\Delta,t)\right], \tag{3.2}$$

with population size N, the metric $g^{\Gamma\Gamma} = \langle\Delta\rangle$, change due to mutations $m^{\Gamma} = -2\mu(\Gamma - \langle\Gamma\rangle_0)$ with mutation rate μ and selection $s_{\Gamma} = f'(\Gamma)$, which is the gradient of the fitness landscape. The neutral dynamics determines the mean value $\langle\Gamma\rangle_0 = \int \Gamma Q_0(\Gamma)d\Gamma$. For Δ the corresponding values are $g^{\Delta\Delta} = 2\Delta^2$, $m^{\Delta} = -4\mu(\Delta - E_0^2)$, and $s_{\Delta} = f'(\Delta)$, where E_0^2 is the expected variance of a random quantitative trait sequence. We expand this model to account for the effect of selective sweeps affecting the quantitative trait.

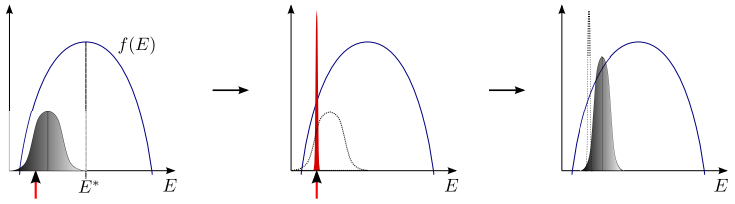

Figure 3.1: **Trait distribution schematic influenced by selective sweeps.** The distribution of trait values within a population is Gaussian. A selective sweep chooses randomly a trait value, which leads to a δ distribution. Afterwards the distribution relaxes again to a Gaussian shape.

3.1.2 The pseudohitchhiking model on a quantitative trait

A minimal model to solve this problem is a simple tightly linked two locus model, where one of these loci is the quantitative trait. We study quantitative traits that are under stabilizing selection. The quantitative trait locus has many different alleles with different trait values E and corresponding allele frequencies x_E. In general the quantitative trait locus will be polymorphic.

The second locus is assumed to be under strong adaptive selection. It incorporates the entire adaptive process of the clonal interference regime by having a rate of selective sweeps ρ with a typical selection coefficient given by the driver mutations

of adaptation process.

This relates the situation to the pseudohitchhiking model proposed by Gillespie [Gillespie, 2000a,b]. We use the basic premises of his model and adapt it to calculate the statistics of a quantitative trait. Just like in his situation we use simplifying assumptions

1. The selection working on the stabilisation of the quantitative trait is negligible with respect to the selection working on the second locus. For the following we assume the quantitative trait to be neutral with respect to the influence by selective sweeps.

2. The fixations at the second locus and thus the selective sweeps form a Poisson process with rate ρ

3. The fixation time at the second locus is small compared to the fixation time at the quantitative trait and the time between substitutions, so that the fixations can be viewed as instantanious.

With these assumptions we can calculate the changes to the mean and the diversity of the quantitative trait. This gives us the wanted statistics of the trait under the influence of much stronger selective sweeps.

The opposite limit of many small effect mutations influencing the quantitative trait gives similar results as we will see from simulations. Furthermore, the trait diversity for neutral traits in this limit is known from fitness wave theory and corroborates our results, when the influence of selection on the trait can be neglected.

Change of the mean trait value Γ

After one selective sweep the population is monomorphic and thus the mean trait value, if a sweep event may have happened is given by

$$\Gamma' = \begin{cases} E_\alpha & \text{with probability} \rho\delta t x_\alpha, \\ \Gamma = \sum_\alpha E_\alpha x_\alpha & \text{with probability} 1 - \rho\delta t \end{cases} \tag{3.3}$$

where x_α is the frequency of trait value E_α in the population before the sweep. The change of the mean is then given by

$$\delta\Gamma = \Gamma' - \Gamma = \begin{cases} E_\alpha - \Gamma & \text{with probability} \rho\delta t x_\alpha, \\ 0 & \text{with probability} 1 - \rho\delta t \end{cases} \tag{3.4}$$

The mean in $\delta\Gamma$ is then

$$\langle \delta\Gamma \rangle|_\Gamma = 0. \tag{3.5}$$

Thus selective sweeps of infinite strength do not directly lead to changes in the between population mean of the mean trait value Γ. The variance of Γ is changed according to

$$\mathbf{Var}(\delta\Gamma')|_\Gamma = \rho\delta t\Delta. \tag{3.6}$$

This means that selective sweeps increase the variance between populations of the mean trait value Γ.

3.2 Effective diffusion model for quantitative traits

The diffusion model for Γ can still be used as an approximation even under the influence of selective sweeps with changed diffusion coefficients, where the added parts are due to assumed independence of the stochastic process of the fitness sweeps and the other processes of the quantitative trait. As the selective sweeps due not change mean of Γ this part of the equation remains unchanged. The changed variance can be additively incorporated in the diffusion constant for Γ.

The changed equation is then given by

$$\begin{aligned}\frac{\partial Q(\Gamma,t)}{\partial t} &= -\frac{\partial}{\partial \Gamma}\left[(-2\mu(\Gamma-\langle\Gamma\rangle_0)-\langle\Delta\rangle f'(\Gamma))Q(\Gamma,t)\right] \\ &+ \frac{\partial^2}{\partial \Gamma^2}\left[\langle\Delta\rangle(\frac{1}{2N}+\rho/2)Q(\Gamma,t)\right] \\ &= -\frac{\partial}{\partial \Gamma}\left[(-2\mu(\Gamma-\langle\Gamma\rangle_0)-\langle\Delta\rangle f'(\Gamma))Q(\Gamma,t)\right] \\ &+ \frac{\partial^2}{\partial \Gamma^2}\left[\langle\Delta\rangle\tilde{\sigma}Q(\Gamma,t)\right]\end{aligned} \tag{3.7}$$

The evolution of Δ is however under the influence of selective sweeps no longer fully described by a diffusive model because jump moments of higher order can no longer be neglected. The most important aspect of the trait variance is its impact on the trait mean due to changed selection. The effect on the trait diversity can be calculated both in response to fitness wave models and under a more idealized saw-tooth model. Here we first show the impact of a simple saw-tooth model,see Fig. 3.2, which is suffient to describe the effect on the trait diversity, as long as the sweeps are the strongest constraint on the diversity. In this simple model the replenishing of diversity is still given by $4\mu E_0^2$, while selective sweeps cut down the diversity to zero leading to a mean trait diversity

$$\langle\Delta\rangle = \frac{4\mu E_0^2}{\rho} = \frac{2\mu E_0^2}{\tilde{\sigma}}. \tag{3.8}$$

Under this assumption $\tilde{\sigma} = \frac{1}{2N} + \rho/2 \approx \rho/2$. The scaling of the mean trait diversity Δ is confirmed by simulations and can be seen for a quadratic landscape in Fig. 3.3. This scaling is further corroborated by an approximate diffusion model of Δ and from a reaction-diffusion model with sweep rates, see chapter 3.4. The approximate diffusion model provides a way to connect the mean diversity to the regime, where selective sweeps do not yet dominate.

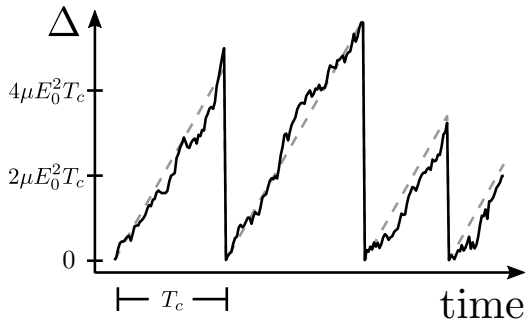

Figure 3.2: **Saw-tooth model for the diversity Δ.** The diversity Δ increases with rate $4\mu E_0^2$ and gets decreased to zero, whenever a sweep happens.

3.3 Moments of the trait average under stabilizing selection

The solution of the diffusion equation is given by

$$Q(\Gamma) = Q_0(\Gamma) \exp\left(\tilde{\sigma}^{-1} F(\Gamma)\right), \tag{3.9}$$

with $Q_0(\Gamma) = Z_0 \exp\left(-(\Gamma - \Gamma_0)^2/(2E_0^2)\right)$ the solution for neutral dynamics. A single peaked quadratic landscape given by

$$f(E) = f^* - c_0(E - E^*)^2 \tag{3.10}$$

is an often used approximation for stabilizing selection[Lande, 1976; Barton and de Vladar, 2009]. It has the added benefit of decoupling the dynamics of the trait mean Γ and diversity Δ [Nour Mohammad et al., 2013]. For traits whose diversity is fully determined by selective sweeps the first two moments, which fully determine the Gaussian distribution of mean trait values, are given by

$$\begin{aligned} \langle \Gamma \rangle &= E^* + \frac{\frac{\mu}{\langle \Delta \rangle}}{\frac{\mu}{\langle \Delta \rangle} + c_0}(\Gamma_0 - E^*) \\ &= E^* + \frac{\tilde{\sigma}}{\tilde{\sigma} + 2c_0 E_0^2}(\Gamma_0 - E^*) \end{aligned} \tag{3.11}$$

and

$$\begin{aligned}\langle \Gamma^2 - \langle \Gamma \rangle^2 \rangle &= \frac{\tilde{\sigma}}{2\frac{\mu}{\langle \Delta \rangle} + 2c_0} \\ &= \frac{\tilde{\sigma}}{\tilde{\sigma} + 2c_0 E_0^2} E_0^2.\end{aligned} \tag{3.12}$$

The squared mean distance to the optimum $\Lambda^2 = (\Gamma - E^*)$ is then given by

$$\langle \Lambda^2 \rangle = \langle (\Gamma - E^*)^2 \rangle = (\langle \Gamma \rangle - E^*)^2 + \langle \Gamma^2 - \langle \Gamma \rangle^2 \rangle. \tag{3.13}$$

3.4 Change of the trait diversity Δ

Calculating the effect of selective sweeps on the trait diversity, we are able to determine the crossover for a diversity that is determined by its own selection to a diversity fully determined by selective sweeps according to the sawtooth model. The trait diversity due to a possible hitchhiking event is given by

$$\Delta' = \begin{cases} 0 & \text{with probability} \rho\delta t, \\ \Delta & \text{with probability} 1 - \rho\delta t \end{cases} \tag{3.14}$$

The change of the diversity is then given by

$$\delta\Delta = \Delta' - \Delta = \begin{cases} -\Delta & \text{with probability} \rho\delta t, \\ 0 & \text{with probability} 1 - \rho\delta t. \end{cases} \tag{3.15}$$

The mean in $\delta\Delta$ is given by

$$\langle \delta\Delta \rangle|_\Delta = -\Delta\rho\delta t. \tag{3.16}$$

The variance is then

$$\begin{aligned}\mathbf{Var}(\delta\Delta)|_\Gamma &= \rho\delta t(1 - \rho\delta t)\Delta^2 \\ &= \rho\delta t\Delta^2 + \mathcal{O}\left((\rho\delta t)^2\right)\end{aligned} \tag{3.17}$$

. This can be incorporated in a reaction-diffusion equation giving

$$\begin{aligned}\frac{\partial Q(\Delta,t)}{\partial t} =& -\frac{\partial}{\partial \Delta}\left[-4\mu(\Delta - E_0^2) - \Delta\frac{1}{N} - 2c_0\Delta^2 Q(\Delta,t)\right] \\ &+ \frac{\partial^2}{\partial \Delta^2}\left[\Delta^2\frac{1}{N}Q(\Delta,t)\right] \\ &- \rho Q(\Delta,t) + \rho\delta(\Delta).\end{aligned} \tag{3.18}$$

3.5 Fokker-Planck equation for Δ

If we for now neglect the higher order jump moments and assume Δ could again be described by a purely diffusive model, we can also write down the Fokker-Planck equation for the trait diversity with the changed diffusion coefficients

$$\begin{aligned}\frac{\partial Q(\Delta,t)}{\partial t} =& -\frac{\partial}{\partial \Delta}\left[(-4\mu(\Delta - E_0^2) - \Delta(\frac{1}{N} + \rho) - 2c_0\Delta^2)Q(\Delta,t)\right] \\ &+ \frac{\partial^2}{\partial \Delta^2}\left[\Delta^2(\frac{1}{N} + \frac{\rho}{2})Q(\Delta,t)\right].\end{aligned} \tag{3.19}$$

This equation can be solved and the stationary solution is given by

$$\begin{aligned}Q(\Delta) = & 2^{\frac{N(\rho+8\mu)}{2N\rho+2}}\Delta^{-3+\frac{8N\mu+N\rho}{2+N\rho}}\exp\left(-\frac{\frac{8N\mu E_0^2}{\Delta} + 4c_0N\Delta}{N\rho+2}\right) \\ &\times\left(\frac{E_0^2\mu}{c_0}\right)^{\frac{N(8\mu+\rho)}{4N\rho+2}+1}\frac{1}{K_{\frac{N(8\mu+\rho)}{2+N\rho}+2}\left(\frac{8\sqrt{2NE_0^2c_0N\mu}}{2+N\rho}\right)}.\end{aligned} \tag{3.20}$$

From the distribution we can obtain the mean of the trait diversity, which is decreasing with increasing sweep rate ρ:

$$\begin{aligned}\langle\Delta\rangle &= \sqrt{\frac{2E_0^2\mu}{c_0}}\frac{K_{\frac{N(8\mu+2\rho)+2}{2+N\rho}}\left(\frac{8\sqrt{2NE_0^2c_0N\mu}}{2+N\rho}\right)}{K_{\frac{N(8\mu+2\rho)+2}{2+N\rho}+1}\left(\frac{8\sqrt{2NE_0^2c_0N\mu}}{2+N\rho}\right)} \\ &\approx \frac{4E_0^2\mu}{4\mu+\frac{1}{N}+\rho}, c \ll \frac{1}{16\mu} \\ &\approx \frac{2E_0^2\mu}{\tilde{\sigma}}, \tilde{\sigma} \gg \mu\end{aligned} \tag{3.21}$$

This approximation assumes small values of Δ, which can be assumed for a high rate of selective sweeps ρ. The result for high sweep rate or small selection is equivalent to the result obtained from the saw-tooth model.

3.6 Fitness load on the quantitative trait

The fitness load on the quantitative trait is the difference between the optimal value f^* and the mean population fitness of the quantitative trait $\bar{f}$

$$\mathcal{L} = f^* - \bar{f}. \tag{3.22}$$

Under the influence of selective sweeps this value can be calculated . The between population average of the fitness load in adaptively evolving populations is given by

$$\langle \mathcal{L} \rangle = c_0 \langle (\Gamma - E^*)^2 \rangle + c_0 \langle \Delta \rangle \tag{3.23}$$

$$= c_0 \frac{\tilde{\sigma}}{2\frac{\mu}{\langle \Delta \rangle} + 2c_0} + c_0 \left(\frac{\frac{\mu}{\langle \Delta \rangle}}{\frac{\mu}{\langle \Delta \rangle} + c_0} (\Gamma_0 - E^*) \right)^2 + c_0 \langle \Delta \rangle \tag{3.24}$$

$$= c_0 \frac{\tilde{\sigma}}{\tilde{\sigma} + 2c_0 E_0^2} E_0^2 \tag{3.25}$$

$$+ c_0 \left(\frac{\tilde{\sigma}}{\tilde{\sigma} + 2c_0 E_0^2} \right)^2 (\Gamma_0 - E^*)^2 \tag{3.26}$$

$$+ c_0 \frac{2E_0^2 \mu}{\tilde{\sigma}} \tag{3.27}$$

$$\approx \frac{\tilde{\sigma}}{2} + c_0 \frac{2E_0^2 \mu}{\tilde{\sigma}} \tag{3.28}$$

$$\approx \frac{\tilde{\sigma}}{2}. \tag{3.29}$$

The approximations are under the assumption of strongly selected quantitative traits. The change from the expected value of the load under strong selection $\langle \mathcal{L}_{free} \rangle = \frac{\sigma}{2} = \frac{1}{4N}$ is given by the rate of selective sweeps $\langle \mathcal{L} - \mathcal{L}_{free} \rangle = \frac{\rho}{4}$. For small selection all three terms scale with the strength of selection of the quantitative trait c_0. The dominant contribution is then given by $c_0(\Gamma_0 - E^*)^2$.

The fitness diversity Δ_f, which is the variance of fitness within the population is given by

$$\begin{aligned} \langle \Delta_f \rangle &\approx \langle \Delta (f'(\Gamma))^2 \rangle \\ &\approx \langle \Delta \rangle \langle (f'(\Gamma))^2 \rangle \\ &= 4c_0 \langle \Delta \rangle \langle \Lambda^2 \rangle. \end{aligned} \tag{3.30}$$

We see a nice agreement of the theory expectations for the fitness observables with numerical results obtained from simulations, see Fig. 3.3.

3.7 Numerical verification for finite sweeps

If we use a population, where the infinite sweep assumption is violated, we can obtain a neutrality threshold $\tilde{\sigma}$ from neutral trait diversity to predict the statistics

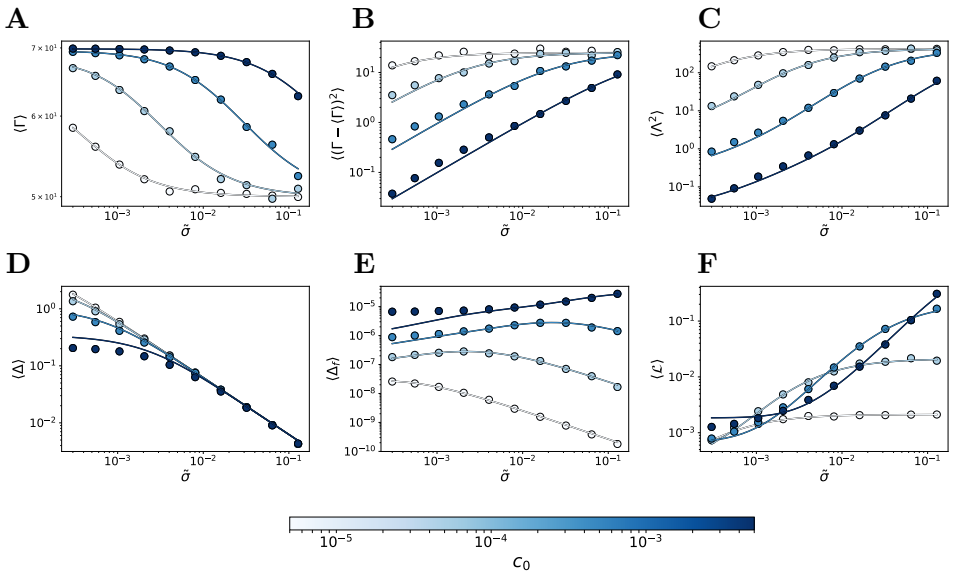

Figure 3.3: **Trait statistics under influence of selective sweeps for a quadratic landscape.** Results from a simulation with a quadratic landscape with sweeps of infinite strength with parameters $N = 10000$, $N\mu = 0.125$, c_0 are shown in the figure as points as a function of the sweeprate $\tilde{\sigma}$. The lines are theory expectations. **A** shows the between population mean of the mean trait value $\langle\Gamma\rangle$, **B** the variance of the mean trait value $\langle(\Gamma - \langle\Gamma\rangle)^2\rangle$, **C** the squared mean distance to the optimum $\langle\Lambda^2\rangle = \langle(\Gamma - E^*)^2\rangle$, **D** the mean population diversity $\langle\Delta\rangle$, **E** the mean fitness diversity $\langle\Delta_f\rangle$, and **F** the genetic load $\langle\mathcal{L}\rangle$.

of a quantitative trait under stabilizing selection on the same genome. For a quantitative trait under neutral dynamics, i.e. without selection, the trait diversity is determined by Eq. 4.12, as long as the mutation rate is not too big $N\mu \ll 1$. This gives an inferred neutrality threshold of

$$\tilde{\sigma} = \frac{2\mu E_0^2}{\langle \Delta \rangle_0}. \tag{3.31}$$

Using this we verify the quantitative trait statistics as before with simulations. The validity of this for a quantitative traits without an additional genome influencing its statistics, is given by [Nour Mohammad et al., 2013].

As can be seen in the figures for constant(Fig. 3.4) and fluctuating selection(Fig. 3.5), the theory expectation fits the data up to a small factor of order 1. Thus, we are able to infer the neutrality threshold from the diversity of neutral sites and can use the inferred neutrality threshold to estimate the statistics of the mean of other linked quantitative traits.

3.8 Discussion

We have shown that the influence on quantitative traits due to hitchhiking can be significant and as such that quantitative traits need to be evaluated in their whole genomic context. For the trait selective sweeps in different parts of the genome can quickly become the strongest influence constraining the diversity. In turn the mean trait value can still be described by a diffusion theory, where the effective population size is given by the neutrality threshold $\tilde{\sigma}$. As such, not only can the coalescence time be inferred from neutral sequence diversity as shown by[Rice et al., 2015; Good et al., 2014], but also from moderately selected quantitative traits. The decreased efficacy of selection on the quantitative traits, leads to much less adapted traits in general, which means that in adaptively evolving asexual populations quantitative traits are much more likely to be at the flank of a stabilizing fitness landscape than at the peak.

We apply this result for the statistics of a quantitative trait under the influence of external selective sweeps in the following chapter 4. There we consider the joint evolution of many quantitative traits influencing and at the same time being influenced by the effective neutrality threshold $\tilde{\sigma}$ together with a focus on a

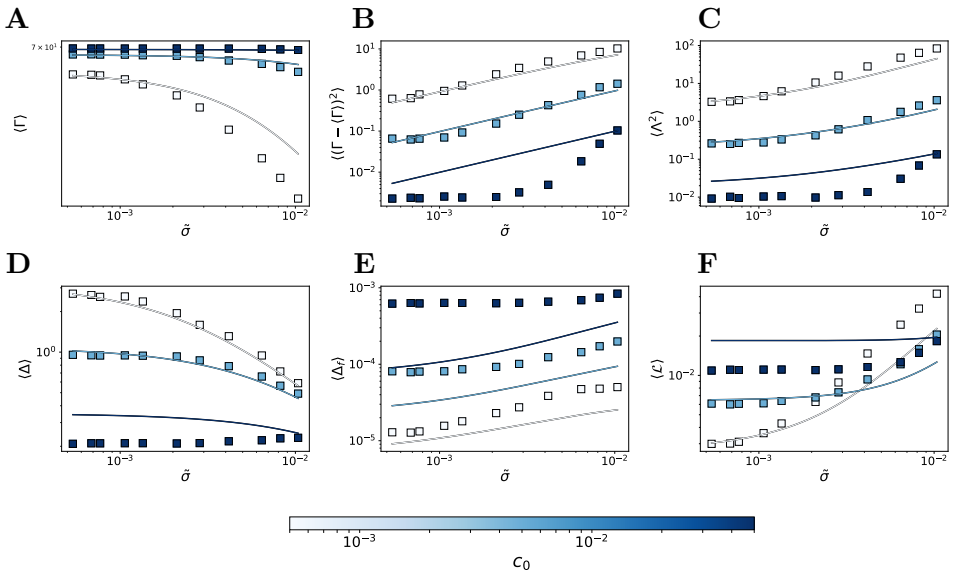

Figure 3.4: **Trait statistics under influence of external sites with constant selection for a quadratic landscape.** Results from a simulation with a quadratic landscape with a varying number of additional loci without selection flips($\gamma = 0$) (see Methods). Parameters are given by $N = 1000$, $N\mu = 0.125$, c_0 as shown in the figure as a function of the neutrality threshold $\tilde{\sigma}$. This threshold is obtained from a neutral trait via Eq. 3.31. The lines are theory expectations. **A** shows the between population mean of the mean trait value $\langle\Gamma\rangle$, **B** the variance of the mean trait value $\langle(\Gamma - \langle\Gamma\rangle)^2\rangle$, **C** the squared mean distance to the optimum $\langle\Lambda^2\rangle = \langle(\Gamma - E^*)^2\rangle$, **D** the mean population diversity $\langle\Delta\rangle$, **E** the mean fitness diversity $\langle\Delta_f\rangle$, and **F** the genetic load $\langle\mathcal{L}\rangle$.

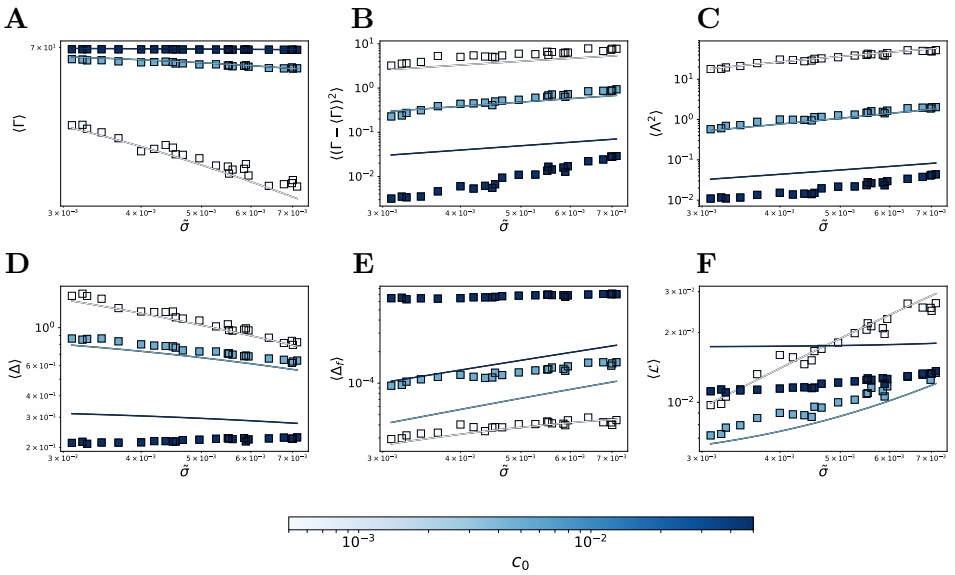

Figure 3.5: **Trait statistics under influence of external sites with selection flips for a quadratic landscape.** Results from a simulation with a quadratic landscape and additional loci with varying rates of selection flips (see Methods) with parameters $N = 1000$, $N\mu = 0.125$, c_0 as shown in the figure as a function of the neutrality threshold $\tilde{\sigma}$. This threshold is obtained from a neutral trait via Eq. 3.31. The lines are theory expectations. **A** shows the between population mean of the mean trait value $\langle\Gamma\rangle$, **B** the variance of the mean trait value $\langle(\Gamma-\langle\Gamma\rangle)^2\rangle$, **C** the squared mean distance to the optimum $\langle\Lambda^2\rangle = \langle(\Gamma - E^*)^2\rangle$, **D** the mean population diversity $\langle\Delta\rangle$, **E** the mean fitness diversity $\langle\Delta_f\rangle$, and **F** the genetic load $\langle\mathcal{L}\rangle$.

biophysical fitness landscape given by a Fermi function.

3.9 Methods

3.9.1 Simulations

The theoretical predictions have been verified using Fisher-Wright simulations. A population consists of N individuals with genomes $\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(N)}$. There are two types of simulations:

1. For the first kind we use simulations, where we directly implement the pseudohitchhiking model. A genotype $\mathbf{a} = (a_1, \ldots, a_L)$ consists of a quantitative trait with L binary alleles $a_k = 0, 1$ $(k = 1, \ldots, L)$. Each sequence defines a trait $E(\mathbf{a}) = \sum_{k=1}^{\ell} \mathcal{E}_k a_k + E_0$, where E_0 is the expectation value of the trait under neutral evolution. The resulting effect distribution of point mutations has as a second moment $\epsilon_G^2 = \sum_{k=1}^{\ell} \mathcal{E}_k^2/\ell$ and a first moment $\kappa_0 \epsilon_G = \sum_{k=1}^{\ell} \mathcal{E}_k(1 - 2\langle a_k \rangle)/\ell$, where $\langle a_k \rangle$ is the state-dependent probability of a mutation at site k being beneficial and brackets $\langle . \rangle$ denote averaging across parallel simulations or time. The genomic fitness is $f(\mathbf{a}) = f(E)$ with $f(E)$ given by a quadratic or exponential function. In each generation, the sequences undergo point mutations with probability $\mu\tau_0$ for each site, where τ_0 is the generation time, and the sequences of the next generation are drawn by multinomial sampling with a probabilities proportional to $1 + \tau_0 f(\mathbf{a})$. Additionally there is a probability of a sweep occuring each generation with probability $\rho\tau_0$; the next generation is then populated by identical sequences, where the particular sequence is chosen by drawing with uniform probability from among the current generation.

2. For the second kind of simulation, the interference is generated by an additional sequence part, which does not encode a quantitative trait, but is additive in fitness. Here a genotype $\mathbf{a} = (a_1, \ldots, a_{L_{\text{phen}}}, a_{L_{\text{phen}}+1}, \ldots, a_{L_{\text{phen}}+L_{\text{single}}})$ consists of a quantitative trait with $L = L_{\text{phen}} + L_{\text{single}}$ binary alleles $a_k = 0, 1$ $(k = 1, \ldots, L)$. Each sequence again defines a trait $E(\mathbf{a}) = \sum_{k=1}^{L_{\text{phen}}} \mathcal{E}_k a_k + E_0$. The genomic fitness is $f(\mathbf{a}) = f(E) + \sum_{i=L_{\text{phen}}+1}^{L} s_i \eta_i a_i$ with $f(E)$ given by a quadratic or exponential function and s_i being a site specific additive

fitness effect and $\eta_i = \pm 1$ defining the direction of selection. Just like for the first kind, in each generation, the sequences undergo point mutations with probability $\mu\tau_0$ for each site, where τ_0 is the generation time, and the sequences of the next generation are drawn by multinomial sampling with a probabilities proportional to $1 + \tau_0 f(\mathbf{a})$. Additionally, with probability $\gamma\tau_0$ the selection of a given site flips $\eta_i \to -\eta_i$.

Appendix

3.10 Time-dependent buildup of the fitness load

Starting in equilibrium and introducing selective sweeps the time dependent load can be calculated under the assumption that the timescale for Δ is much smaller than the timescale for Γ to reach a stationary state. Under this assumption we have a simple relaxation of $Q(\Gamma)$ to its new equilibrium value under the influence of selective sweeps, leading to

$$\begin{aligned}
\langle L(t)\rangle = {} & c_0\left(\frac{1}{4N}+\frac{\rho}{4}\right)\frac{1}{c_0+\frac{\mu}{\langle\Delta\rangle_\rho}} \\
& + \left(c_0\frac{1}{4N}\frac{1}{c_0+\frac{\mu}{\langle\Delta\rangle_{\rho=0}}} \right. \\
& \quad \left. -c_0\left(\frac{1}{4N}+\frac{\rho}{4}\right)\frac{1}{c_0+\frac{\mu}{\langle\Delta\rangle_\rho}} \right) e^{-4(\mu+c_0\langle\Delta\rangle_\rho)t} \\
& + c_0\left[\frac{\mu}{\mu+c_0\langle\Delta\rangle_\rho}(\Gamma_0-E^*)+\right. \\
& \left.\left(\frac{\frac{\mu}{\langle\Delta\rangle_{\rho=0}}\Gamma_0+c_0E^*}{\frac{\mu}{\langle\Delta\rangle_{\rho=0}}+c_0}-\frac{\frac{\mu}{\langle\Delta\rangle_\rho}\Gamma_0+c_0E^*}{\frac{\mu}{\langle\Delta\rangle_\rho}+c_0}\right)\right. \\
& \left. e^{-2(\mu+c_0\langle\Delta\rangle_\rho)t}\right]^2 \\
& + c_0\langle\Delta\rangle_\rho
\end{aligned} \tag{3.32}$$

4 Phenotypic Interference and fitness waves

> "As with all things, layer settles upon layer, and in time the deepest, darkest ones become forgotten - yet they have shaped all that lies above."
>
> *(Steven Erikson, The Bonehunters)*

Asexually reproducing populations evolve under complete genetic linkage. Hence, selection on an allele at one genomic locus can interfere with the evolution of simultaneously present alleles throughout the genome. Linkage-induced interference interactions between loci include background selection (the spread of a beneficial allele is impeded by linked deleterious alleles), hitchhiking or genetic draft (a neutral or deleterious allele is driven to fixation by a linked beneficial allele), and clonal interference between beneficial alleles originating in disjoint genetic clades (only one of which can reach fixation). These interactions and their consequences for genome evolution have been studied extensively in laboratory experiments [Wiser et al., 2013; Barroso-Batista et al., 2014], natural populations [Betancourt et al., 2009; Strelkowa and Lässig, 2012], and theory [Gerrish and Lenski, 1998; Desai and Fisher, 2007; Rouzine et al., 2008; Hallatschek, 2011; Schiffels et al., 2011; Good et al., 2012; Neher and Hallatschek, 2013; Neher et al., 2013]. The most prominent global effect of interference is to reduce the speed of evolution, which has been observed in laboratory evolution experiments [de Visser et al., 1999; Perfeito et al., 2007; McDonald et al., 2016]. The fitness cost of interference, which has also been measured [Cooper, 2007; Couce et al., 2017], is the center piece of classic arguments for the evolutionary advantage of sex [Fisher, 1930; Muller, 1932; Eigen, 1971; Felsenstein, 1974; Kondrashov, 1993]. Much less clear is how

interference affects the system-wide evolution of molecular phenotypes, such as protein stabilities and affinities governing gene regulation and cellular metabolism. This is the topic of the present chapter, which looks at the systems-biological consequences of interference evolution. We establish that interference generates a long-term degradation of an organism's molecular functions by the accumulation of deleterious mutations. This effect is strongly dependent on genome size: it becomes an evolutionary force constraining organismic complexity and driving the evolution of recombination.

Our analysis is based on simple biophysical models of molecular evolution [Gerland and Hwa, 2002; Berg et al., 2004; Chen and Shakhnovich, 2009; Goldstein, 2011; Serohijos and Shakhnovich, 2014; Manhart and Morozov, 2015; Chi and Liberles, 2016]. In a minimal model, an individual organism consists of g genes and each gene carries a single quantitative trait G, the stability of its protein. The trait is encoded in multiple sites of the gene sequence and is affected by mutations at these sites, most of which will make the protein less stable. Selection on a gene is described by a standard thermodynamic fitness landscape $f(G)$, which is a sigmoid function with a high-fitness plateau corresponding to stable proteins and a low-fitness plateau corresponding to unfolded proteins (Fig. 4.1A). We also discuss an stability-affinity protein model with a two-dimensional fitness landscape $f(G, E)$; this model includes enzymatic or regulatory functions of genes, specifically the protein binding affinity E to a molecular target. The genome-wide mutation-selection balance in these fitness landscapes describes populations maintaining the functionality of their molecular traits; we refer to this state as *housekeeping evolution*. We analyze its long-term evolutionary forces on genome architecture that arise independently of short-term adaptive processes, such as the evolution of resistance.

Over a wide range of model parameters, we find that housekeeping evolution takes place in an evolutionary mode of *phenotypic interference*. In this mode, genetic and phenotypic variants in multiple genes generate standing fitness variation under complete genetic linkage, a so-called traveling fitness wave [Desai and Fisher, 2007; Rouzine et al., 2008; Hallatschek, 2011; Good et al., 2012; Neher and Hallatschek, 2013; Neher et al., 2013]. We show that phenotypic interference is a system-wide collective dynamics with a universal feedback between the global fitness wave and selection on individual phenotypic variants. This feedback generates a

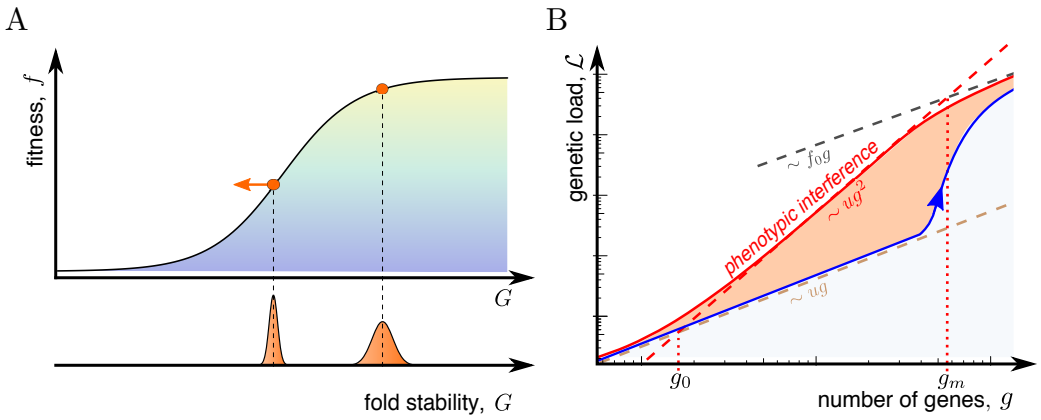

Figure 4.1: **Fitness landscape and fitness cost of phenotypic interference.** A: Minimal biophysical fitness model. The fitness of an individual gene, f, is a sigmoid function of its fold stability, G. This function has a high-fitness region of stable, functional proteins (yellow), an inflection point at intermediate fitness marking marginally functional proteins, and a low-fitness region of dysfunctional proteins (blue). The mutation-selection dynamics on this landscape generates high-fitness equilibria ($\tilde{\sigma} \ll f_0$, red dot) and unstable states at lower fitness ($\tilde{\sigma} \gtrsim f_0$, red dot with arrow), depending on the fitness difference f_0 between functional and dysfunctional proteins and the coalescence rate $\tilde{\sigma}$; see also Fig. 4.4B. The fold stability distribution of a population in these states is shown below. B: The total genetic load $\mathcal{L}$ in a genome is shown as a function of the number of genes, g, for different models of genome evolution. Red line: Asexual evolution in the minimal biophysical model has an evolutionary regime of *phenotypic interference* where $\mathcal{L}$ increases quadratically with g; see Eq. (4.2) and simulation data shown in Fig. 4.4A. This regime arises from the competition of phenotypic variants within a population. The nonlinear scaling of $\mathcal{L}$ sets in at a small gene number g_0 and ends at a much larger value g_m, which marks the crossover to genomes with a large fraction of dysfunctional genes (grey line). Blue line: Under asexual evolution in a model with discrete gene fitness effects, the onset of load nonlinearity and interference occurs at $g \sim g_m$ and is associated with the onset of Muller's ratchet [Muller, 1964; Gordo and Charlesworth, 2000; Rouzine et al., 2008]. Brown line: Sexual evolution reduces $\mathcal{L}$ to a linear function of g, if the recombination rate is above the transition point R^* given by Eq. (4.8).

fitness cost, defined as the difference between the mean population fitness and the fitness maximum of fully functional genes, that increases quadratically with g (Fig. 4.1B). The fitness cost of interference quantifies its systems-biological effects: the maintenance of each gene degrades stability and function of all other genes by increasing the accumulation of deleterious mutations. The cost nonlinearity sets in already at a small number of genes, g_0, and generates strong selection against genome complexity in viable, asexually reproducing organisms. This distinguishes our biophysical models from classical models of mutational load, which predict a linear fitness cost up to a much larger error threshold g_m associated with mutational meltdown [Muller, 1964; Eigen, 1971; Gordo and Charlesworth, 2000; Rouzine et al., 2008] (Fig. 4.1B).

Remarkably, the genome-wide steady state of evolution affords an analytic solution in our minimal model. We develop this solution in the following section; then we turn to model extensions and biological consequences on genome complexity under asexual evolution. Housekeeping evolution in these models also provides a biophysically grounded rationale for the evolution of sex. We show that long-term selective pressure on the recombination rate induces a first-order phase transition to a mode of sexual evolution without genome-wide interference, and we obtain a simple estimate of the transition recombination rate R^* that can be directly compared to data.

4.1 Theory of phenotypic interference

The solution of the minimal model has two parts that will be discussed in order. First, the mean fitness variance of a single quantitative trait at evolutionary equilibrium depends in a simple way on a global evolutionary parameter, the coalescence rate $\tilde{\sigma}$. Second, for the steady state of housekeeping evolution, the fitness variances of all traits combine to the total standing fitness variation, which in turn sets the coalescence rate and leads to a closure of the derivation.

Evolution of a quantitative trait under interference selection

The stability G of a protein is the free energy difference between the unfolded and the folded state (Methods). This trait gains heritable variation Δ_G by new mutations

at a speed $u\epsilon_G^2$, where u is the total mutation rate and ϵ_G^2 is the mean square stability effect of its sequence sites. The trait loses variation by coalescence at a rate $\tilde{\sigma}$. These processes determine an equilibrium stability variation $\Delta_G = u\,\epsilon_G^2/(2\tilde{\sigma})$. This type of relation is well known for neutral sequence variation in models of genetic draft [Gillespie, 2000b] and of traveling fitness waves [Good et al., 2014; Rice et al., 2015]. It can be derived more generally from a diffusion theory for quantitative traits under selection; see 4.4 and chapter 3. Next, we consider the mutation-selection equilibrium of a gene on the flank of the fitness landscape $f(G)$. We equate the rate of stability increase by selection, $\Delta_G\, f'(G)$, with the rate of trait degradation by mutations, $u\epsilon_G$, using that most mutations in a functional trait are deleterious (Methods). This relates the mean square selection coefficient at trait sites, $s^2 = \epsilon_G^2 f'^2(G)$, and the fitness variance $\Delta_f \approx \Delta_G f'^2(G)$ to the coalescence rate,

$$s^2 = 4\tilde{\sigma}^2, \qquad \Delta_f = 2u\tilde{\sigma}. \tag{4.1}$$

These relations are universal; that is, they do not depend on details of the fitness landscape and the trait effect distribution of sequence sites. Remarkably, trait fluctuations by genetic drift and genetic draft also leave their form invariant (Section 4.5 and Fig 4.5).

Eqs. (4.1) express a salient feature of selection on quantitative traits: the strength of selection on genetic variants is not fixed *a priori*, but is an emergent property of the global evolutionary process. A faster pace of evolution, i.e., an increase in coalescence rate $\tilde{\sigma}$, reduces the efficacy of selection [Schiffels et al., 2011; Good et al., 2012; Rice et al., 2015]. In a downward curved part of the fitness landscape, this drives the population to an equilibrium point of lower fitness and higher fitness gradients. The resulting equilibrium tunes typical selection coefficients to marginal relevance, where mean fixation times $1/s$ are of the order of the coalescence time $1/\tilde{\sigma}$. This point marks the crossover between effective neutrality ($s \ll \tilde{\sigma}$) and strong selection ($s \gg \tilde{\sigma}$); consistently, most but not all trait sites carry their beneficial allele [Schiffels et al., 2011].

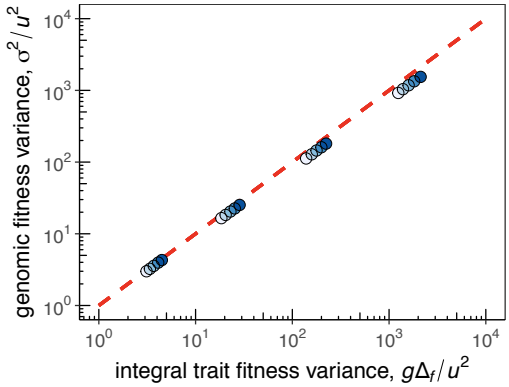

Figure 4.2: **Additivity of the genomic fitness variance.** For housekeeping evolution in the minimal biophysical model, we plot the total fitness variance, σ^2, against the additive part $\Delta_{f,1} + \cdots + \Delta_{f,g}$. The additivity is used in the closure of the evolutionary dynamics, Eqs. (4.2)– (4.4).

Housekeeping evolution of multiple traits

The non-adaptive stationary scenario of housekeeping evolution[1] builds on the assumption that over long time scales, selection acts primarily to repair the deleterious effects of mutations, because these processes are continuous and affect the entire genome. In contrast, short-term adaptive processes are often environment-dependent, transient, and affect only specific genes. Here we discuss a closed solution of the phenotypic interference dynamics for housekeeping evolution. In section 4.6, we extend this approach to scenarios of adaptive evolution and show that these do not affect the conclusions of the chapter.

In a housekeeping equilibrium, the total fitness variation σ^2 is simply the sum of the fitness variances of individual genes, $\sigma^2 = g\Delta_f$ (4.2). Moreover, traveling wave theory shows that σ^2 and the coalescence rate $\tilde{\sigma}$ are simply related, $\sigma^2/\tilde{\sigma}^2 = c_0 \log(N\sigma)$, where N is the population size and $c_0 \sim 10^2$ [Neher et al., 2013; Neher and Hallatschek, 2013]. Together with Eq. (4.1), we obtain the global fitness wave

$$\sigma^2 = 4\frac{u^2 g^2}{c}, \qquad \tilde{\sigma} = 2\frac{ug}{c}, \tag{4.2}$$

[1] Our terminology of housekeeping evolution must not be confused with housekeeping genes for cell metabolism. We denote by this the non-adaptive evolutionary steady state.

as well as corresponding characteristics of individual traits,

$$\frac{\Delta_G}{\epsilon_G^2} = \frac{c}{4g}, \qquad s^2 = 4\tilde{\sigma}^2 = 16\frac{u^2 g^2}{c^2}, \tag{4.3}$$

in terms of the slowly varying parameter

$$c = \frac{\sigma^2}{\tilde{\sigma}^2} \approx c_0 \log(Nug). \tag{4.4}$$

As shown in Methods, this parameter has a simple interpretation: it estimates the complexity of the fitness wave, that is, the average number of genes with simultaneously segregating beneficial genetic variants destined for fixation. Fisher-Wright simulations of the minimal model confirm Eqs. (4.2)–(4.4); they reproduce the joint pattern of σ^2, $\tilde{\sigma}^2$, Δ_G, and s^2 and infer the wave complexity c (Fig. 4.3).

These relations are the centerpiece of phenotypic interference theory. They show that the collective evolution of molecular quantitative traits under genetic linkage depends strongly on the number of genes that encode these traits. The dependence is generated by a feedback between the global fitness variation, σ^2, and mean square local selection coefficients at genomic sites, s^2. In section 4.4, we show that this feedback also tunes the evolutionary process to the crossover point between independently evolving genomic sites and strongly correlated fitness waves composed of multiple small-effect mutations.

4.1.1 Biological implications of phenotypic interference

Interference selection against complexity

The feedback of phenotypic interference has an immediate consequence for the genetic load, which is determined by the average position of genes on the fitness landscape. We first consider stable and functional genes located in the concave part of the minimal model landscape $f(G)$ (Fig. 4.1A). This part can be approximated by its exponential tail, where the load is proportional to the slope $f'(G)$. Eq. (4.3) then predicts a load $sk_BT/\epsilon_G \approx 2\tilde{\sigma}$ per gene, where we have used that typical reduced effect sizes ϵ_G/k_BT are of order 1 (Methods). This implies a superlinear

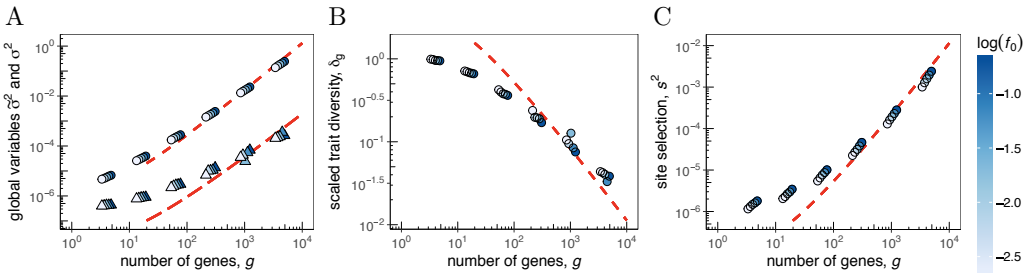

Figure 4.3: **Global and local scaling under phenotypic interference.** A: Average total fitness variance, σ^2 (circles) and coalescence rate $\tilde{\sigma}^2$ (triangles) versus number of genes, g, for asexual evolution. Simulation data for different average gene selection coefficients f_0 (indicated by color) are compared to model results, $\sigma^2 \sim g^2/c$ (short-dashed line) and $\tilde{\sigma}^2 \sim g^2/c^2$ (long-dashed line) for $g > g_0 \sim 10^2$; Eqs. (4.2) and (4.4). B: Average scaled trait diversity, $\delta_G = \Delta_G/\epsilon_G^2$, versus g. Simulation data (circles); model results, $\delta_G \sim c/g$ (dashed line; Eq. (4.3)). Values $\delta_G < 1$ indicate that individual proteins are in the low-mutation regime. C: The mean square selection coefficient at sequence sites, s^2, versus g. Simulation results (circles); model results, $s^2 = 4\tilde{\sigma}^2 \sim g^2/c^2$ (short-dashed line as in A; Eq. (4.4)). The scaling $s^2 \sim \tilde{\sigma}^2$ is independent of f_0, signalling that site selection coefficients emerge from a feedback between global and local selection (see text). Other simulation parameters: $N = 1000, u = 1.25 \times 10^{-3}$, $\epsilon_G/k_BT = 1$; see section 4.3.4 for simulation details.

scaling of the total equilibrium genetic load,

$$\mathcal{L}_{\text{int}}(g) \approx 2g\tilde{\sigma} = 4\frac{ug^2}{c}, \tag{4.5}$$

which sets on at a small gene number g_0 given by the condition $g_0 \approx c/4$ (Fig. 4.1B, numerical simulations are shown in Fig. 4.4A). The superlinearity of the genetic load is the most important biological effect of phenotypic interference. As detailed in section 4.6 and Fig. 4.6, this scaling holds more generally for a sufficient number of quantitative traits evolving under genetic linkage; it does not depend on details of the fitness landscape and of the underlying biophysical processes. For example, active protein degradation, a ubiquitous process that drives the thermodynamics of folding out of equilibrium [Hochstrasser, 1996], does not affect

our conclusions. Another example is the stability-affinity model, which has two quantitative traits per gene that evolve in a two-dimensional sigmoid fitness landscape $f(G, E)$ [Manhart and Morozov, 2015; Chéron et al., 2016]. We show that under reasonable biophysical assumptions, evolution in a stability-infinity model produces a 2-fold higher interference load than the minimal model, $\mathcal{L}_{\rm int}(g) \approx 8ug^2/c$. Alternative models with a quadratic single-peak fitness landscape generate an even stronger nonlinearity of the load, $\mathcal{L}_{\rm int}(g) \sim g^3$. In contrast, a discrete model with a fitness effect f_0 of each gene shows a linear load up to a characteristic gene number $g_m = (f_0/u)\log(Nf_0)$ associated with the onset of mutational meltdown by Muller's ratchet [Muller, 1964; Gordo and Charlesworth, 2000; Rouzine et al., 2008].

The interference load builds up with a time lag given by the relaxation time to equilibrium,

$$\tau = \frac{1}{u} = 2\frac{g}{c}\frac{1}{\tilde{\sigma}}. \tag{4.6}$$

Deleterious mutations in an organism's genes build up on a time scale τ, which exceeds the coalescence time $\tilde{\sigma}^{-1}$. Therefore, the load $\mathcal{L}_{\rm int}$ affects the long-term fitness of a population against competing lineages. Specifically, it generates strong long-term selection against genome complexity: the fitness cost for each additional gene, $\mathcal{L}'_{\rm int}(g)$, can take sizeable values even at moderate genome size. For example, in a "standard" microbe of the complexity of *E. coli*, a 10 % increase in gene number may incur an additional load $\Delta\mathcal{L} \approx 3 \times 10^{-2}$ under the stability-affinity model (with parameters $g = 5000, u = 10^{-6}, N = 10^8$). In comparison, the discrete model leads to a much smaller value $\Delta\mathcal{L} = 5 \times 10^{-4}$ for the same parameters.

It is instructive to compare the interference load of an extra gene with its physiological fitness cost $\mathcal{L}'_{\rm phys}(g)$, which is generated primarily by the synthesis of additional proteins (and is part of the overall fitness effect f_0). For a gene with an average expression level, $\mathcal{L}'_{\rm phys}(g) = \lambda/g$ with a constant $\lambda \sim 1$ reflecting the (re-)allocation of metabolic resources in the cell; see refs. [Scott et al., 2010; Lynch and Marinov, 2015]. This cost acts as a selective force on changes of genome size, which take place within a coalescence interval $\tilde{\sigma}^{-1}$. Importantly, $\mathcal{L}'_{\rm phys}$ is much smaller than $\mathcal{L}'_{\rm int}$ for a standard microbe, suggesting a two-scale evolution of genome sizes. On short time scales, the dynamics of gene numbers is permissive and allows the rapid acquisition of adaptive genes; these changes are neutral with

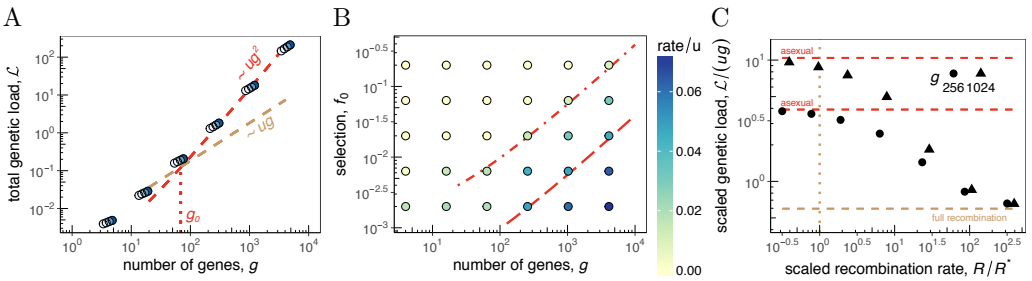

Figure 4.4: **Genetic load, gene loss, and transition to sexual evolution.** A: Total genetic load $\mathcal{L}$ versus the number of genes g for asexual evolution. Simulation results (circles) for different values of f_0 (indicated by color, as in Fig. 4.3); model results: interference load $\mathcal{L}_{\text{int}} \sim ug^2/c$ (red line; Eq. (4.5)) for $g > g_0$ (dotted line) and null model $\mathcal{L} = ug$ (brown line) as in Fig. 4.1B. The superlinear behavior of $\mathcal{L}$ indicates strong selection against genome complexity. B: Rate of gene loss (indicated by color, in units of u) as a function of the gene selection coefficient, f_0, and the number of co-evolving genes g. Genes with $f_0 \sim \tilde{\sigma}$ (long-dashed line, cf. Fig. 4.3A) have appreciable loss rates; genes with $f_0 \gtrsim 10\tilde{\sigma}$ (dashed-dotted line) have negligible loss rates, i.e., are conserved under phenotypic interference. C: Scaled genetic load, $\mathcal{L}/(ug)$, versus scaled recombination rate, R/R^*, for different genome sizes. The observed load rapidly drops from the superlinear scaling of phenotypic interference, $\mathcal{L} = 4ug^2/c$ (asymptotic data: red lines), to the linear scaling of unlinked genes, $\mathcal{L} \sim ug$ (brown line). This signals a (fluctuation-rounded) transition to sexual evolution at the threshold recombination rate $R^* = 2ug/c$ (dotted line, see Eq. (4.8)). Other simulation parameters as in Fig. 4.3; see section 4.3.4 for simulation details.

respect to $\mathcal{L}'_{\text{int}}$. On longer time scales (of order τ), marginally relevant genes are pruned in a more stringent way, for example, by invasion of strains with more compact genomes.

Interference drives gene loss

The near-neutral dynamics of genome size extends to gene losses, which become likely when a gene gets close to the inflection point of the sigmoid fitness landscape (Fig. 4.1A). The relevant threshold gene fitness, f_0^c, is set by the coalescence rate,

which leads to

$$f_0^c \sim 2\tilde{\sigma} = \frac{4ug}{c} \tag{4.7}$$

in the minimal model. Strongly selected genes ($f_0 \gg 2\tilde{\sigma}$) have equilibrium trait values firmly on the concave part of the landscape, resulting in small loss rates of order $u\exp(-f_0/2\tilde{\sigma})$; these genes can be maintained over extended evolutionary periods. Marginally selected genes ($f_0 \lesssim 2\tilde{\sigma}$) have near-neutral loss rates of order u [Schiffels et al., 2011], generating a continuous turnover of genes. According to Eq. (4.7), the threshold f_0^c for gene loss increases with genome size, which expresses again the evolutionary constraint on genome complexity. The dependence of the gene loss rate on f_0 and $\tilde{\sigma}$ is confirmed by simulations (Fig. 4.4B). The housekeeping coalescence rate $\tilde{\sigma} = 2ug/c$ sets a lower bound for the fitness threshold f_0^c, adaptive evolution can lead to much larger values of $\tilde{\sigma}$ and f_0^c.

The transition to sexual evolution

Recombination breaks up genetic linkage at a rate R per genome and per generation (R is also called the genetic map length). Evolutionary models show that recombination generates linkage blocks that are units of selection; a block contains an average number ξ of genes, such that there is one recombination event per block and per coalescence time, $R\xi/(g\tilde{\sigma}(\xi)) = 1$ [Weissman and Barton, 2012; Neher, 2013; Neher et al., 2013]. Depending on the recombination rate, these models predict a regime of asexual evolution, where selection acts on entire genotypes ($\xi \sim g$), and a distinct regime of sexual evolution with selection acting on individual alleles ($\xi \ll g$). Here we focus on the evolution of the recombination rate itself and establish a selective avenue for the transition between asexual and sexual evolution. With the phenotypic interference scaling $\tilde{\sigma}(\xi) = 2u\xi/c$ for $\xi \gtrsim c$, as given by Eq. (4.2), our minimal model produces an instability at a threshold recombination rate

$$R^* = \frac{2ug}{c}. \tag{4.8}$$

This signals a first-order phase transition to sexual evolution with the genetic load as order parameter (Fig. 4.4C). For $R < R^*$, the population is in the asexual mode of evolution ($\xi \sim g$), where interference produces a superlinear load $\mathcal{L}_{\text{int}} = 2ug^2/c$. For $R > R^*$, efficient sexual evolution generates much smaller block sizes ($\xi \sim c$). In

this regime, the mutational load drops to the linear form $\mathcal{L} = ug \ll \mathcal{L}_{\rm int}$, providing a net long-term fitness gain $\Delta\mathcal{L} \simeq \mathcal{L}_{\rm int}$. However, the process of recombination itself entails a direct short-term cost $\mathcal{L}_{\rm rec}$ [Kondrashov, 1993]. If we assume that cost to be of order 1 per event, we obtain $\mathcal{L}_{\rm rec} \sim R^* = \tilde{\sigma}$ close to the transition. This cost is much smaller than the gain $\Delta\mathcal{L}$ and remains marginal (i.e., $\mathcal{L}_{\rm rec}/\tilde{\sigma} \sim 1$).

Together, our theory of phenotypic interference suggests a specific two-step scenario for the evolution of sex. Recombination at a rate of order R^* is near-neutral at short time scales, so a recombining variant of rate R^* arising in an asexual background population can fix by genetic drift and draft. Recombining strains acquire a long-term benefit $\Delta\mathcal{L} \sim gR^* = g\tilde{\sigma}$, so they can outcompete asexual strains in the same ecological niche. The threshold rate R^* is of the order of the genome-wide mutation rate ug, so even rare facultative recombination can induce the transition. This scenario builds on the basic biophysics of molecular traits but does not require *ad hoc* assumptions on adaptive pressure, on rate and effects of beneficial and deleterious mutations, or on genome-wide epistasis [Kondrashov, 1993]. It is at least consistent with observed recombination rates in different parts of the tree of life: genome average values are always well above R^*; a high-resolution recombination map of the *Drosophila* genome shows low-recombining regions with values above but of order R^* [Comeron et al., 2012; Schiffels et al., 2017] (4.1).

4.2 Discussion

Here we have developed the evolutionary genetics of multiple quantitative traits in non-recombining populations. We find a specific evolutionary mode of phenotypic interference, which is characterized by a feedback between global fitness variation and local selection coefficients at genomic sites. This feedback generates highly universal features, which include the complexity of the evolutionary process and the scaling of coalescence rate and genetic load with gene number, as given by Eqs. (4.2)–(4.4).

Phenotypic interference produces strong selection against genome complexity in asexual populations, which implies selection in favor of recombination above a threshold rate R^* given by Eq. (4.8). The underlying genetic load originates from the micro-evolutionary interference of phenotypic variants within a population and unfolds with a time delay beyond the coalescence time, as given by Eq. (4.6).

	Saccharomyces cerevisiae	*Drosophila melanogaster*	*Arabidopsis thaliana*
μ	$3{\cdot}10^{-8}$ [Lynch et al., 2008]	$3{\cdot}10^{-9}$ [Keightley et al., 2014]	$7{\cdot}10^{-9}$[Ossowski et al., 2009]
ℓ	1401 [Harrison et al., 2003]	1500 [Harrison et al., 2003]	2232 [Derelle et al., 2006]
g	6563 [Derelle et al., 2006]	14332 [Harrison et al., 2003]	26990 [Feng et al., 2009]
R	$3{\cdot}10^{-2}$ [Ruderfer et al., 2006]	$2{\cdot}10^{-3} - 1{\cdot}10^{-0}$ [Fiston-Lavier et al., 2010; Comeron et al., 2012]	$2{\cdot}10^{-0}$ [Salome et al., 2012]
R^*	$6{\cdot}10^{-5}$	$1{\cdot}10^{-3}$	$9{\cdot}10^{-3}$

Table 4.1: **Genome data and estimates of threshold recombination rates.** Point mutation rate μ, average gene length ℓ (in bp), gene number g and recombination rate R per genome (map length) are shown for three recombining species. The parameter range for *D. melanogaster* describes local recombination rates in different parts of the chromosomes (in the same units) [Comeron et al., 2012]. An upper bond of the threshold recombination rate R^* marking the transition to sexual evolution is obtained from Eq. (4.8) (with $ug = \mu\ell g$ and $c \approx c_0 \approx 100$).

Therefore, the interference load is a macro-evolutionary selective force that impacts the long-term fitness and survival of a population in its ecological niche.

Molecular complexity, the broad target of phenotypic interference, can be regarded as a key systems-biological observable. In our simple biophysical models, we measure complexity by number of stability and binding affinity traits in a proteome. More generally, we can define complexity as the number of (approximately) independent molecular quantitative traits, which includes contributions from an organism's regulatory, signalling, and metabolic networks that scale in a nonlinear way with genome size. Interference selection affects the complexity and architecture of all of these networks, establishing new links between evolutionary and systems biology to be explored in future work.

We have seen that the co-evolution of quantitative traits in asexual populations provides a framework, where fitness wave theory can be applied without an artificial infinite sites model. We now turn in the next chapter to Mendelian traits in a genomic model to investigate, under which conditions fitness waves occur in this context.

4.3 Methods

4.3.1 Biophysical fitness models

In thermodynamic equilibrium at temperature T, a protein is folded with probability $p_+(G) = 1/[1+\exp(-G/k_BT)]$, where G is the Gibbs free energy difference between the unfolded and the folded state and k_B is Boltzmann's constant. A minimal biophysical fitness model for proteins takes the form

$$f(G) = f_0\, p_+(G) = \frac{f_0}{1+\exp(-G/k_BT)} \tag{4.9}$$

with a single selection coefficient capturing functional benefits of folded proteins and metabolic costs of misfolding [Chen and Shakhnovich, 2009; Goldstein, 2011; Serohijos and Shakhnovich, 2014]. Similar fitness models based on binding affinity have been derived for transcriptional regulation [Gerland and Hwa, 2002; Berg et al., 2004; Mustonen et al., 2008]; the rationale of biophysical fitness models has been reviewed in refs. [Lässig, 2007; Chi and Liberles, 2016]. In section 4.6, we introduce alternative fitness landscapes for proteins and show that our results depend only on broad characteristics of these landscapes. The minimal global fitness landscape for a system of g genes with traits $G_1, \ldots, G_g$ and selection coefficients $f_{0,1}, \ldots, f_{0,g}$ is taken to be additive, i.e., without epistasis between genes,

$$f(G_1, \ldots, G_g) = \sum_{i=1}^{g} \frac{f_{0,i}}{[1+\exp(-G_i/k_BT)]}. \tag{4.10}$$

4.3.2 Evolutionary model

We characterize the population genetics of an individual trait G by its population mean Γ and its expected variance Δ_G. The trait mean follows the stochastic evolution equation

$$\dot{\Gamma} = -u\kappa\epsilon_G + \Delta_G f'(\Gamma) + \eta(t) \tag{4.11}$$

with white noise $\eta(t)$ of mean $\langle\eta(t)\rangle = 0$ and variance $\langle\eta(t)\eta(t')\rangle = \tilde{\sigma}\Delta_G\,\delta(t-t')$. This dynamics is determined by the rate u, the mean effect $(-\kappa)\epsilon_G$, and the mean square effect ϵ_G^2 of trait-changing mutations, which determine the diversity $\Delta_G = u\epsilon_G^2/(2\tilde{\sigma})$ [Nour Mohammad et al., 2013]. We use effects $\epsilon_G \approx 1$–$3\,k_BT$,

which have been measured for fold stability [Tokuriki et al., 2007; Zeldovich et al., 2007] and for molecular binding traits [Gerland and Hwa, 2002; Kinney et al., 2010; Tuğrul et al., 2015], and a mutational bias $\kappa = 1$, which is consistent with the observation that most mutations affecting a functional trait are deleterious.

4.3.3 Housekeeping equilibrium

The deterministic equilibrium solution ($\dot{\Gamma} = 0, \eta = 0$) of Eq. (4.11) determines the dependence of Δ_G and the associated fitness variance $\Delta_f = \Delta_G f'^2(G)$ on $\tilde{\sigma}$, as given by Eq. (4.1); the same scaling follows from the full stochastic equation (section 4.5). The derivation of the global housekeeping equilibrium, Eqs. (4.2)–(4.4), uses two additional inputs: the additivity of the fitness variance, $\sigma^2 = g\Delta_f$, which is confirmed by our simulations (4.2), and the universal relation $\sigma^2/\tilde{\sigma}^2 = c_0 \log(N\sigma)$ [Neher et al., 2013; Neher and Hallatschek, 2013] in a travelling fitness wave, where the coalescence rate $\tilde{\sigma}$ is generated predominantly by genetic draft. Eqs. (4.2)–(4.4) determine further important characteristics of phenotypic interference:

(a) The complexity of the fitness wave, defined as the average number of beneficial substitutions per coalescence time, is $(v_+ g)/\tilde{\sigma} \sim ug/2\tilde{\sigma} = c/4$, using that trait-changing mutations are marginally selected, Eq. (4.1), and have nearly neutral fixation rates $v_+ \sim u/2$ per gene.

(b) The evolutionary equilibria of stable genes ($f_0 \gg \tilde{\sigma}$) are located in the high-fitness part of the minimal fitness landscape, $f \simeq f_0[1 - \exp(-G/k_B T)]$. These genes have an average fitness slope $f' = (\Delta_f/\Delta_G)^{1/2} = 2\tilde{\sigma}/\epsilon_G$, an average trait $\Gamma = -k_B T \log(2\tilde{\sigma} k_B T/f_0 \epsilon_G)$, and an average load $\mathcal{L}_{\text{int}}(g)$ given by Eq. (4.5).

(c) The scaling regime of Eqs. (4.2)–(4.4) sets in at a gene number g_0 given by the condition $g_0 = c/4$; this point also marks the crossover from the linear load $\mathcal{L}_0(g) = ug$ to the nonlinear form $\mathcal{L}_{\text{int}}(g)$.

4.3.4 Simulations

In-silico evolution of stability traits. We use a Wright-Fisher process to simulate the evolution of stability traits in a population. A population consists of N

individuals with genomes $\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(N)}$. A genotype $\mathbf{a} = (\mathbf{a}_1, \ldots, \mathbf{a}_g)$ consists of g segments; each segment is a subsequence $\mathbf{a}_i = (a_{i,1}, \ldots, a_{i,\ell})$ with binary alleles $a_{j,k} = 0, 1$ $(i = 1, \ldots, g;\ k = 1, \ldots, \ell)$. A segment $\mathbf{a}$ defines a stability trait $G(\mathbf{a}) = \sum_{k=1}^{\ell} \mathcal{E}_k a_k + G_0$, where G_0 is the expectation value of the trait under neutral evolution. The resulting effect distribution of point mutations has as a second moment $\epsilon_G^2 = \sum_{\imath=1}^{\ell} \mathcal{E}_k^2/\ell$ and a first moment $\kappa_0 \epsilon_G = \sum_{\imath=1}^{\ell} \mathcal{E}_k(1 - 2\langle a_k \rangle)/\ell$, where $\langle a_k \rangle$ is the state-dependent probability of a mutation at site k being beneficial and brackets $\langle . \rangle$ denote averaging across parallel simulations or time. The genomic fitness is $f(\mathbf{a}) = \sum_{i=1}^{g} f(G(\mathbf{a}_i); f_{0,i})$ with $f(G)$ given by Eq. (4.9) and gene-specific amplitudes $f_{0,i}$. In each generation, the sequences undergo point mutations with probability $\mu\tau_0$ for each site, where τ_0 is the generation time, and the sequences of the next generation are drawn by multinomial sampling with a probabilities proportional to $1 + \tau_0 f(\mathbf{a})$.

Simulations are performed with parameters $N = 1000$, $N\mu = 0.0125$, each trait with genomic base of size $\ell = 100$, and each site with equal effect $\mathcal{E}_k = 1$. The quantitative trait dynamics is insensitive to the form of the effect distribution [Nour Mohammad et al., 2013; Held et al., 2014]. To increase the performance of the simulations, we do not keep track of the full genome. We only store the number of deleterious alleles $n_i = \sum_{k=1}^{\ell} a_{i,k}$ for each trait, we draw mutations with rate $u = \mu\ell$, and we assign to each mutation a beneficial change $\mathcal{E}$ with probability n_i/ℓ and a deleterious change $-\mathcal{E}$ otherwise. This procedure produces the correct genome statistics for bi-allelic sites with uniform trait effects $\mathcal{E}_i = \mathcal{E}$. Simulation data are shown with theory curves for $\kappa = 1$, which provide a good fit to all amplitudes; the input κ_0 is different by a factor of order 1 which includes fluctuation effects (Section 4.5).

Housekeeping evolution. For the simulations in Figs. 4.3 and 4.4A, where we are not explicitly interested in the loss of genes, we use an exponential approximation of the stable regime of the stability fitness landscape. The reason is a limited accessible parameter range in simulations constraining the values of f_0 and $\tilde{\sigma}$ due to finite N. We checked that the exponential approximation gives the same results as the full model in the regime $f_0/\tilde{\sigma} \gg 1$, where the gene loss rate in the biophysical landscape is negligible.

Loss rate measurements. In the biophysical landscape used in Fig 4.4B, a long-term stationary population is maintained by evolving 70% of the traits in a biophysical fitness landscape with selection f_0; the remaining 30% of the traits are modeled to be essential with selection $10f_0$. Gene loss is defined by the condition $G < -3.5k_BT$. To maintain a constant number of genes, lost genes are replaced immediately with an input trait value $G > 0$.

Recombination. For simulations with recombination (Fig. 4.4C), we draw recombination events with rate NR for the whole population from a Poisson distribution. Each recombination event is implemented as one crossover between the genomes of two individuals at a random, uniformly distributed position of the genomes.

Appendix

4.4 Trait diversity and cross-over scaling of the fitness wave

Equilibrium of trait diversity. Consider a quantitative trait G evolving by mutations, coalescence caused by genetic drift and genetic draft, and stabilizing selection in a fitness landscape $f(G)$. Mutations and coalescence alone generate an equilibrium of the trait diversity Δ,

$$\Delta_G = \frac{u\epsilon_G^2}{2\tilde{\sigma}}, \tag{4.12}$$

as derived in chapter 3 and ref. [Nour Mohammad et al., 2013]. This expression is valid if stabilizing selection on the trait diversity can be neglected, i.e., if [Nour Mohammad et al., 2013]

$$\frac{\mathcal{L}_\Delta}{\tilde{\sigma}} \equiv \frac{\Delta_G |f''(\Gamma)|}{\tilde{\sigma}} \lesssim 1. \tag{4.13}$$

Here we show that this condition is self-consistently fulfilled throughout the phenotypic interference regime. Evaluating the expected fitness curvature in the high-fitness part of the minimal fitness landscape, Eq. (4.9), where $f''(\Gamma) = -f'(\Gamma)/k_BT$, and in the mutation-coalescence equilibrium given by Eq. (4.12), we obtain $f'' = -2\tilde{\sigma}/(\epsilon_g k_B T)$. By Eq. (4.3), the condition (4.13) then reduces to

$$\frac{\Delta_G}{\epsilon_G^2} = \frac{c}{4g} \lesssim 1, \tag{4.14}$$

which is identical to the condition for phenotypic interference. This relation expresses an important scaling property of the phenotypic interference regime: individual traits evolve in the low-mutation regime and are monomorphic at most times. In contrast, the the global trait diversity defines a polymorphic fitness wave,

$$\frac{4g\Delta_G}{\epsilon_G^2} = \frac{\sigma^2}{\tilde{\sigma}^2} = c \gtrsim g_0. \tag{4.15}$$

Cross-over scaling of the fitness wave. A travelling fitness wave maintained by mutations at genomic sites with a fixed selection coefficient s has two distinct scaling regimes [Neher et al., 2013; Neher, 2013],

$$\sigma^2 = \begin{cases} sug, & (g \lesssim g_c) \\ c\tilde{\sigma}^2 = (\frac{c}{4})^{1/3}(s^2 ug)^{2/3} & (g \gtrsim g_c), \end{cases} \tag{4.16}$$

which correspond to independently evolving sites and to an asymptotic fitness wave with strong interference selection, respectively. At the crossover point $g_c = cs/(4u)$, the relation

$$\tilde{\sigma}(g_c) = \frac{2ug_c}{c} = \frac{1}{2}s \tag{4.17}$$

is valid. Comparing this relation with the *generic* scaling under phenotypic interference, $\tilde{\sigma} = 2ug/c = s/2$ as given by Eqs. (4.2) and (4.3), we conclude that the phenotypic fitness wave is locked in the crossover region of marginal interference. This feature reflects the feedback between global and local selection in a phenotypic fitness landscape, which tunes selection coefficents to the g-dependent value $s = 4ug/c$. Consistently, the phenotypic fitness wave has a fitness variance $\sigma^2 \sim g^2$, compared to the scaling $\sigma^2 \sim g^{4/3}$ of the asymptotic regime at fixed selection coefficients (up to log corrections).

4.5 Stochastic theory of phenotypic interference

In section 4.1 we derive the scaling relations of phenotypic interference, Eqs. (4.1) – (4.5), using the evolution equation for quantitative traits, Eq. (4.11), in its deterministic limit ($\eta = 0$). Here we show that the full evolution equation generates the same scaling. We convert Eq. (4.11) into an equivalent diffusion equation [Nour Mohammad et al., 2013] and chapter 3 for the probability density $Q(\Gamma, t)$,

$$\frac{\partial}{\partial t} Q(\Gamma, t) = \left[\tilde{\sigma} \Delta_G \frac{\partial^2}{\partial \Gamma^2} + \frac{\partial}{\partial \Gamma} \left(\kappa \epsilon_G u - \Delta_G f'(\Gamma) \right) \right] Q(\Gamma, t) \tag{4.18}$$

with the average trait diversity Δ_G given by Eq. (4.12). The equilibrium probability distribution $Q_{\mathrm{eq}}(\Gamma)$ describes the stationary fluctuations of the population mean trait $\Gamma(t)$ of a stable gene around its long-term average $\langle \Gamma \rangle = \int \Gamma \, Q_{\mathrm{eq}}(\Gamma) \, d\Gamma$; these

fluctuations are generated by genetic drift and (predominantly) genetic draft. In the biophysical fitness landscape, Eq. (4.9), the equilibrium distribution can be evaluated analytically,

$$Q_{\rm eq}(\Gamma) = \frac{\left(\frac{f_0}{\tilde\sigma}\right)^{2\kappa\frac{k_BT}{\epsilon_G}} \exp\left(-\frac{2\kappa\Gamma}{\epsilon_G} - \frac{f_0}{\tilde\sigma}{\rm e}^{-\Gamma/k_BT}\right)}{k_BT\,\mathrm{Gamma}\left(2\kappa\frac{k_BT}{\epsilon_G}\right)}, \tag{4.19}$$

where Gamma and PolyGamma are standard transcendental functions. This function is plotted in Fig. A in 4.5. The resulting average,

$$\frac{\langle\Gamma\rangle}{k_BT} = -\log\left(\frac{\tilde\sigma}{f_0}\right) - \mathrm{PolyGamma}\left(2\kappa\frac{k_BT}{\epsilon_G}\right), \tag{4.20}$$

shows that genes are slightly more stable than estimated from the deterministic average, $\Gamma/k_BT = -\log(2\kappa\tilde\sigma k_BT/f_0\epsilon_G)$. Through the nonlinearity of the fitness landscape, the fluctuations of the mean trait Γ induce fluctuations of the conditional average fitness variance, $\langle\Delta_f\rangle(\Gamma) = \Delta_G f'^2(\Gamma)$. We obtain the equilibrium distribution

$$Q_{\rm eq}(\Delta_f) = \mathrm{Gamma}_{\rm gen}\left(\Delta_f;\ 2\kappa\frac{k_BT}{\epsilon_G}, \frac{1}{2}u\tilde\sigma\frac{\epsilon_G^2}{(k_BT)^2}, \frac{1}{2}, 0\right), \tag{4.21}$$

with $\mathrm{Gamma}_{\rm gen}$ denoting the generalized gamma distribution (Fig. B in 4.5). The average fitness variance

$$\langle\Delta_f\rangle = 2\tilde\sigma u\kappa^2\left(1 + \frac{\epsilon_G}{2\kappa k_BT}\right) \tag{4.22}$$

differs from its deterministic counterpart, Eq. (4.1) by a prefactor of order 1. Similarly, the Γ fluctuations induce fluctuations of the interference load of individual genes,

$$Q_{\rm eq}(\mathcal{L}_{\rm gene}) = \mathrm{Gamma}_{\rm dist}\left(\mathcal{L}_{\rm gene};\ 2\kappa\frac{k_BT}{\epsilon_G}, \tilde\sigma\right) \tag{4.23}$$

(Fig. C in 4.5). The resulting dependence

$$\langle\mathcal{L}_{\rm gene}\rangle = 2\kappa\frac{k_BT}{\epsilon_G}\tilde\sigma \tag{4.24}$$

is identical to the deterministic case; the fluctuation effect on $\langle\Gamma\rangle$, Eq. (4.20), is offset by the fluctuation load in a downward-curved fitness landscape.

4.6 Model extensions

In this section, we develop alternative evolutionary models of quantitative traits under genetic linkage. The mode of phenotypic interference, which is characterized by a superlinear scaling of the genetic load with genome complexity, occurs in all cases, suggesting it is a generic property of this class of models. Specifically, we discuss housekeeping dynamics in extended models of protein evolution and we extend our analysis to adaptive processes.

Active protein degradation. This non-equilibrium process affects a wide range of proteins, for example through the ubiquitin-proteasome pathway[Hochstrasser, 1996]. It ensures that regulatory proteins are rapidly cleared once their function ends (at a particular point of the cell cycle). Consider a simple model, which has a constant rate K^- of active degradation and a rate $K_G^+ = K_G^0 \mathrm{e}^{G/k_BT}$ for the folding process. Here we do not model details of the pathways of protein synthesis from and degradation into amino acid constituents, which would only affect the total protein concentration but not their state probabilities. In the steady state, proteins are folded with probability

$$\tilde{p}_+(G) = \frac{1}{1+\nu_G \mathrm{e}^{-G/k_BT}}, \tag{4.25}$$

where $\nu_G = K^-/K_G^0$. Hence, this model retains the sigmoid form of the fitness landscape given in Eq. (4.9) and shown in Fig. 4.1 , and our evolutionary conclusions remain invariant.

Stability-affinity model. This model extends the minimal protein model by explicitly including protein function, which is assumed to be mediated through binding to a molecular target. Proteins can be in three thermodynamic states: functional, i.e., folded and target-bound $(++)$, folded and unbound $(+-)$, and unfolded $(--)$. We assume that unfolded proteins cannot bind their target, which

implies that the fourth state of unfolded proteins localized to their target $(-+)$ is suppressed by the entropy loss of localization.

We consider two different thermodynamic ensembles of these proteins. In *thermodynamic equilibrium,* the statistics of this ensemble is governed by two quantitative traits, which are defined as free energy differences: the fold stability $G \equiv G_{--} - G_{+-}$ and the reduced binding affinity $E \equiv G_{+-} - G_{++}$, which includes the entropy loss of localization and depends on the ligand concentration. The equilibrium state probabilities p_{++}, p_{--}, and p_{--} are given by Boltzmann statistics depending on the traits G and E; in particular,

$$p_{++}(G,E) = \frac{1}{1 + \mathrm{e}^{-E/k_BT} + \mathrm{e}^{-(E+G)/k_BT}}. \tag{4.26}$$

Equilibrium models of this kind are well known in protein biophysics [Phillips et al., 2013; Monod et al., 1965], and have been used to build fitness landscapes [Manhart and Morozov, 2015; Chéron et al., 2016]. *Active degradation* is again a ubiquitous process that drives the thermodynamics out of equilibrium; this process is particularly relevant for target-bound proteins that would have a long lifetime at thermodynamic equilibrium. Here we assume a single degradation rate K^- for the processes $(++) \to (--)$ and $(+-) \to (--)$, a rate $K_G^+ = K_G^0 \mathrm{e}^{G/k_BT}$ for the folding process $(--) \to (+-)$, and a rate $K_E^+ = K_E^0 \mathrm{e}^{E/k_BT}$ for the binding process $(+-) \to (++)$. In this model, the folding and binding processes decouple, and we obtain the non-equilibrium steady-state probability

$$p_{++}(G,E) = \frac{1}{\left(1 + \nu_G \mathrm{e}^{-G/k_BT}\right)\left(1 + (1+\nu_E)\mathrm{e}^{-E/k_BT}\right)} \tag{4.27}$$

with $\nu_G = K^-/K_G^0$ and $\nu_E = K^-/K_E^0$. From these ensembles, we build thermodynamic fitness landscapes

$$f(G,E) = f_0\, p_{++}(G,E) \tag{4.28}$$

analogous to Eq. (4.9); these landscapes are plotted in Fig. 4.6A and B.

The population genetics of the two-trait system is described by the population mean values Γ_G and Γ_E, the diversities Δ_{GG} and Δ_{EE}, and the covariance Δ_{GE}. Under mutations, coalescence, and selection given by the fitness landscape $f(G,E)$,

the mean traits follow a stochastic evolution equation analogous to Eq. (4.11),

$$\begin{pmatrix}\dot{\Gamma}_G\\ \dot{\Gamma}_E\end{pmatrix} = -\begin{pmatrix}u_G\kappa_G\epsilon_G\\ u_E\kappa_E\epsilon_E\end{pmatrix} + \begin{pmatrix}\Delta_{GG} & \Delta_{GE}\\ \Delta_{GE} & \Delta_{EE}\end{pmatrix}\begin{pmatrix}\partial_G f(E,G)\\ \partial_E f(E,G)\end{pmatrix} + \begin{pmatrix}\eta_G\\ \eta_E\end{pmatrix} \tag{4.29}$$

with white noise of mean and variance

$$\begin{pmatrix}\langle\eta_G\rangle\\ \langle\eta_E\rangle\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}, \qquad \begin{pmatrix}\langle\eta_G(t)\eta_G(t')\rangle & \langle\eta_G(t)\eta_E(t')\rangle\\ \langle\eta_G(t)\eta_E(t')\rangle & \langle\eta_E(t)\eta_E(t')\rangle\end{pmatrix} = \tilde{\sigma}\delta(t-t')\begin{pmatrix}\Delta_{GG} & \Delta_{GE}\\ \Delta_{GE} & \Delta_{EE}\end{pmatrix}. \tag{4.30}$$

Here we discuss the simplest stationary states of housekeeping evolution in this model, using the deterministic limit of the evolution equation ($\eta_G = \eta_E = 0$). The trait diversities Δ_{GG} and Δ_{EE} are given as in Eq. (4.12), and we assume that pleiotropic sites have uncorrelated effects on both traits, i.e., $\Delta_{GE} = 0$; this has recently been observed in [Otwinowski, 2018]. We again set $\kappa_G = \kappa_E = 1$, which says that most random mutations reduce stability and affinity.

In equilibrium, the high-fitness part of the fitness landscape takes the asymptotic form $f(G,E) \simeq f_0[1-\mathrm{e}^{-E/k_BT}(1+\mathrm{e}^{-G/k_BT}+\mathrm{e}^{-E/k_BT})]+O((e^{-E/k_BT}, e^{-G/k_BT})^3)$. The mutation-selection equilibrium leads to mean trait values

$$\begin{pmatrix}\Gamma_G\\ \Gamma_E\end{pmatrix} \approx \begin{pmatrix}k_BT\log\left(\frac{\epsilon_G}{\epsilon_E}-1\right)\\ -k_BT\log\left(2\frac{\tilde{\sigma}}{f_0}\left(\frac{k_BT}{\epsilon_E}-\frac{k_BT}{\epsilon_G}\right)\right)\end{pmatrix}, \tag{4.31}$$

where only the E-component depends on the coalescence rate $\tilde{\sigma}$. Comparison with the minimal model, Eq. (4.1), shows that in the stable part of the fitness landscape, the equilibrium stability-affinity model becomes an essentially one-dimensional problem for the affinity trait E [Manhart and Morozov, 2015]. The total fitness variance per gene, $\Delta_f = 2(u_G+u_E)\tilde{\sigma}$, is of the universal form (4.1) with an effective mutation rate

$$u = u_G + u_E. \tag{4.32}$$

We conclude that housekeeping evolution in this model follows the same scaling as in the minimal model, Eqs. (4.1) – (4.8), with the parameter u given by Eq. (4.32). However, the equilibrium model lacks evolutionary stability, because lack of folding stability ($G > 0$) can be compensated by a stronger binding affinity.

With active degradation, the high-fitness part of the fitness landscape takes the

asymptotic form $f(G,E) \simeq f_0[1-(1+\nu_E)\mathrm{e}^{-E/k_BT}-\nu_G\mathrm{e}^{-G/k_BT}]+O((e^{-E/k_BT}, e^{-G/k_BT})^2)$. Hence, for stable genes ($f_0 \gg \tilde{\sigma}$), the evolutionary dynamics of the traits G and E becomes approximately independent. The traits of each gene are at a mutation-selection equilibrium of the universal form (4.1), generating a combined fitness variance $\Delta_f = 2(u_G + u_E)\tilde{\sigma}$. Therefore, housekeeping evolution in this model also follows the same scaling as in the minimal model, Eqs. (4.1) – (4.8), with a total mutation rate per gene given by Eq. (4.32) and an effective value of g that is twice the number of genes,

$$g_{\mathrm{eff}} = 2g. \tag{4.33}$$

In particular, the system-wide interference load is about twice the value of the minimal model, $\mathcal{L}_{\mathrm{int}} \approx 8ug^2/c$. This estimate disregards the additional contribution from the enhanced total mutation rate, Eq. (4.32), which takes into account that $u_E \ll u_G$ for many binding domains.

The form invariance of housekeeping evolution in these models shows the robustness of the phenotypic interference mode. It also suggests that in more general contexts, we can define genomic complexity as the number of quantitative traits that evolve (approximately) independently; see the Discussion of this chapter.

Single-peak fitness model. A minimal model of stabilizing selection is a quadratic landscape [Lande, 1976; de Vladar and Barton, 2014; Nourmohammad et al., 2013],

$$f(E) = -f_0(E - E^*)^2 \tag{4.34}$$

(Fig. C in 4.6). This model penalizes deviations from an optimal trait value E^*. In contrast to the biophysical landscape, there is no gene loss in a quadratic landscape, because there are no constraints on its slope. As long as mutations generate trait equilibria predominantly on one flank of the landscape, the basic scaling of phenotypic interference, Eqs. (4.1) – (4.4), is universal and, hence, the same as in the minimal model. The genetic load for a single gene, $\mathcal{L}_{\mathrm{gene}} = -f(\Gamma) = u^2\epsilon_E^2/4\Delta_E^2 f_0$, has been derived in [Nour Mohammad et al., 2013]. With Δ_E given by Eq. (4.12), we find a system-wide interference load

$$\mathcal{L}_{\mathrm{int}} = g\mathcal{L}_{\mathrm{gene}} = \frac{4u^2g^3}{c^2\epsilon_E^2 f_0}. \tag{4.35}$$

Hence, the single-peak model has an even stronger load nonlinearity than the biophysical fitness landscapes.

Phenotypic interference in adaptive evolution. Here we show that the phenotypic interference scaling extends to simple models of adaptive evolution. In the minimal biophysical model, we assume that protein stabilities are still at an evolutionary equilibrium of the universal form (4.1), generating a combined fitness variance $g\Delta_f = 2gu\tilde{\sigma}$. However, the global fitness variance acquires an additional contribution from adaptive evolution of other system functions,

$$\sigma^2 = c\tilde{\sigma}^2 = 2\tilde{\sigma}ug + \phi, \tag{4.36}$$

where ϕ is the *fitness flux* or rate of adaptive fitness gain [Mustonen and Lässig, 2010]. Mathematically, this term quantifies the deviations of the adaptive evolutionary process from equilibrium (defined by detailed balance). Closure of the modified dynamics leads to an increased coalescence rate $\tilde{\sigma}$,

$$\tilde{\sigma} = \frac{2ug}{\mathcal{C}} + \frac{\phi}{2ug} + O\left(\frac{\phi^2}{(ug)^3}\right). \tag{4.37}$$

However, the adaptive term remains subleading to the housekeeping term for large g; this is true even if we assume that ϕ is proportional to g. Hence, the total interference load, $\mathcal{L}_{\text{int}} = g\tilde{\sigma} = 2ug^2/c + \phi/u + \ldots$, retains the leading nonlinearity generated by housekeeping evolution, as given by Eq. (4.5). Only for very high fitness flux ($\phi \gg u^2g^2/c$), coalescence becomes dominated by adaptation, leading to a substantial decrease in the efficacy of selection.

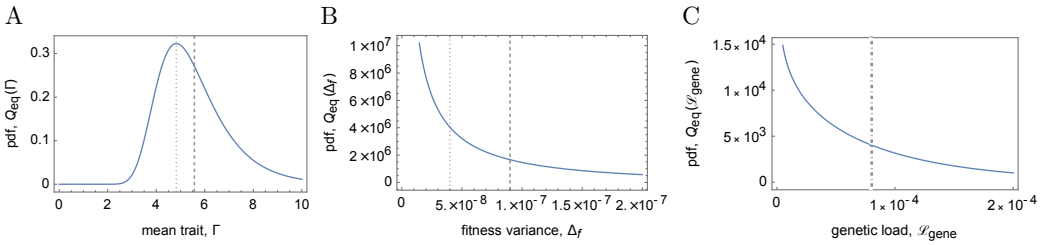

Figure 4.5: **Equilibrium distributions under stochastic evolution.** The figure shows the probability density functions (A) of the mean population trait, $Q_{\mathrm{eq}}(\Gamma)$, (B) of the conditional expected fitness variance, $Q_{\mathrm{eq}}(\Delta_f)$, and (C) of the genetic load per gene, $Q_{\mathrm{eq}}(\mathcal{L}_{\mathrm{gene}})$; see Eqs. (4.19) – (4.23). These distributions measure deviations from long-term averages (dashed lines), which are generated by genetic drift and draft. The corresponding deterministic solutions are marked by dotted lines; both lines coincide in (C). All pdfs are shown for $\tilde{\sigma} = f_0/100 = 10^{-4}$; other parameters as in Fig. 2. See section 4.5.

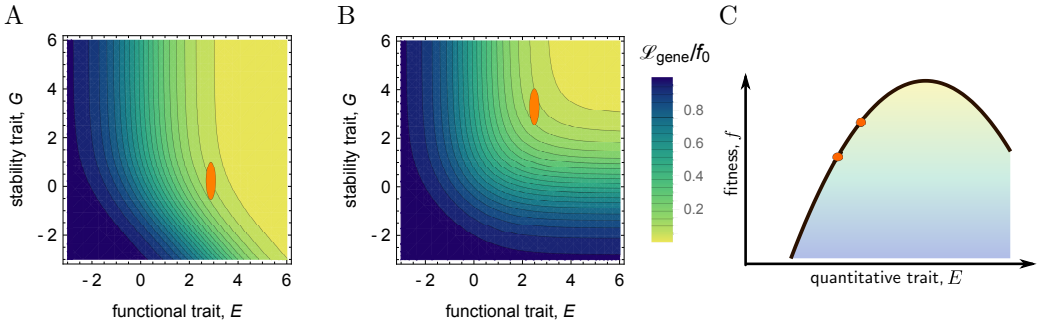

Figure 4.6: **Extended fitness landscapes.** (A, B) Thermodynamic fitness landscapes $f(G, E)$ of the stability-affinity model, Eqs. (4.26) – (4.28), are shown as functions of the stability G and the affinity E. Stable populations, characterized by stationary mean values (Γ_G, Γ_E) and variances (Δ_G, Δ_E), are marked by red ellipsoids. (A) Thermodynamic equilibrium. (B) Non-equilibrium driven by active degradation of folded proteins. In the high-fitness part, this landscape becomes approximately additive in G and E. (C) Quadratic fitness landscape $f(E)$, Eq. (4.34), as a minimal model for stabilizing selection on a quantitative trait E. Stable population states on a flank of the landscape are marked by red dots. See section 4.6.

5 Punctuated genomic fitness waves

"There is no magic. There is only knowledge, more or less hidden."

(Gene Wolfe, Shadow & Claw)

The debate about the mode and tempo of evolution has been a long running question. Asexual evolution in particular is strongly characterized by interference interactions due to genomic linkage. These linkage interactions, which include the competition of beneficial mutations (clonal interference), hitchhiking of neutral and deleterious mutations, and background selection, where a mutation's fate is mainly determined by its linked alleles, have been studied theoretically [Gerrish and Lenski, 1998; Desai and Fisher, 2007; Rouzine et al., 2008; Hallatschek, 2011; Schiffels et al., 2011; Good et al., 2012; Neher and Hallatschek, 2013; Neher et al., 2013] and experimentally [Wiser et al., 2013; Barroso-Batista et al., 2014] to understand this particular mode of evolution. For high mutation rates and large population sizes a strongly background determined, collective type of behavior – the travelling fitness wave – has been of major interest in the last 10 years due its analytical framework, which predicts the decrease of adaptation measured in laboratory evolution experiments[de Visser et al., 1999; Perfeito et al., 2007; McDonald et al., 2016]. Moreover, the coalescent process of the travelling fitness wave has been shown to belong to the universality class of the Bolthausen-Sznitman coalescent[Brunet et al., 2007].

However a recent experimental result for *E.coli* has led to questioning of the so far accepted theories for long-term evolutionary behavior. An increase to the overall fitness does not prevent genome degradation[Couce et al., 2017]. To close the gap between the observed long-term consequence of genome degradation with the existing theoretical framework is the goal of this chapter. Using a genomic model together with fitness wave theory, we are able to show that fast adaptation

not only entails genome degradation, but even facilitates it. Furthermore, we find a mode of punctuated fitness waves, where due to a separation of selection and timescales maintenance of function is interrupted by adaptive bursts.

We classify stationary adaptation in a genomic framework in four different phases as is shown in Fig. 5.3: The evolution of a small number of sites according to independent substitutions is compatible with a viable genome, but is of no particular interest for the mode asexual evolution considered here. Another viable stationary evolutionary mode, we call *standing fitness wave*, which does still exhibit a linear scaling with genome size, but already shows a stable fitness distribution. A viable stationary evolutionary mode with adaptation,*punctuated fitness waves*, which shows a clear separation of time- and selection scales giving rise to intermittent adaptation. And finally, fitness waves associated with the Bolthausen-Sznitman coalescent, which in the context of a simple genomic model cannot maintain sites at their preferred alleles for long-term evolution, giving rise to an unstable, *molten* genome. The pattern of substitutions for sites with different selection coefficients are shown in Fig. 5.1. Increasing genome size, increases the neutrality threshold $\tilde{\sigma}$ and in turn enables substitutions for sites with higher selection coefficients. Alternatively, in adaptive scenarios neofunctionalization events enable adaptive substitutions even for sites, where substitutions would otherwise be exponentially suppressed.

As a consequence adaptation has to occur intermittently, in a turbulent way, or can only be transient.

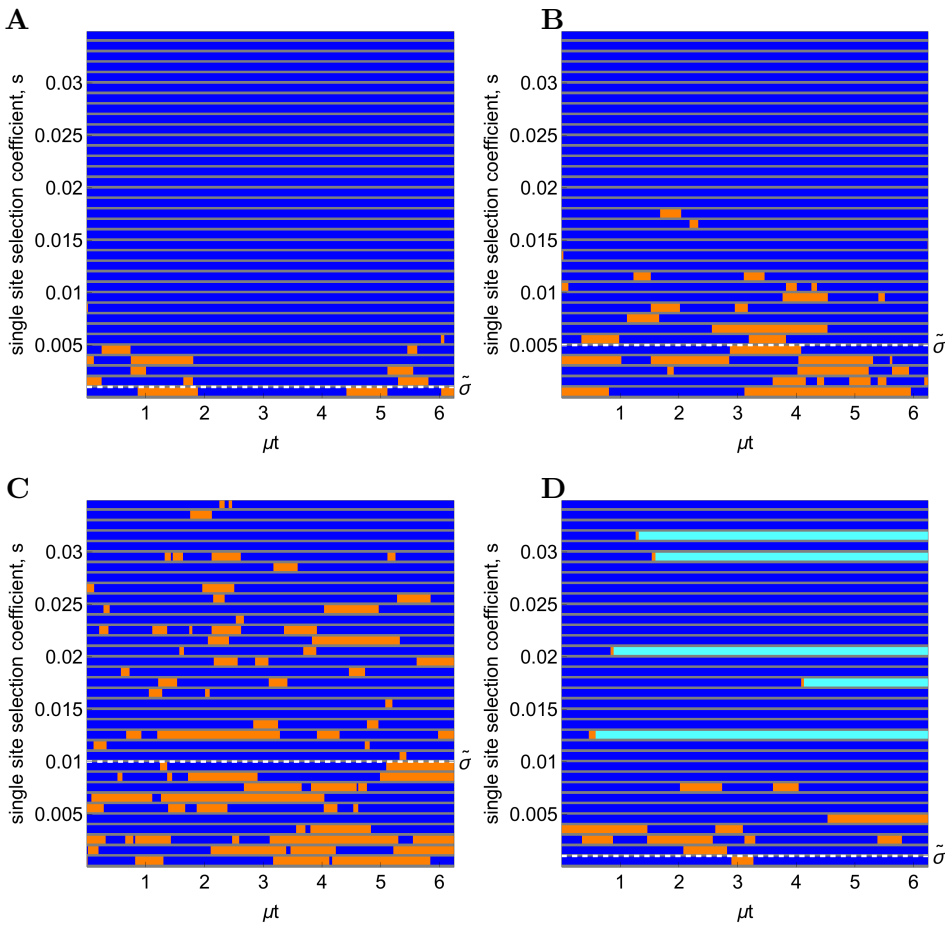

Figure 5.1: **Substitution pattern for stationary genomic asexual adaptation.** Different loci showing the state of the fixed allele for different numbers of loci L (**A,D**: $L = 100$, **B**: $L = 500$, **C**: $L = 5000$) and rate of neofunctionalization events γ (**A,B,C**: $\gamma = 0$, **D**: $\gamma > 0$) shown as a function of evolutionary time μt. The beneficial state is shown in blue and the deleterious state in orange. For the panel **D**, which is the only panel with a neofunctionalization rate $\gamma > 0$, the light blue color identifies the beneficial state after a neofunctionalization event. We see the pattern of substitutions mainly constrained in the regime of sites with selection values around and below the neutrality threshold $\tilde{\sigma}$. Increasing the number of loci L, increases the neutrality threshold, and increases the percentage of sites which are not fixed at the beneficial allele.

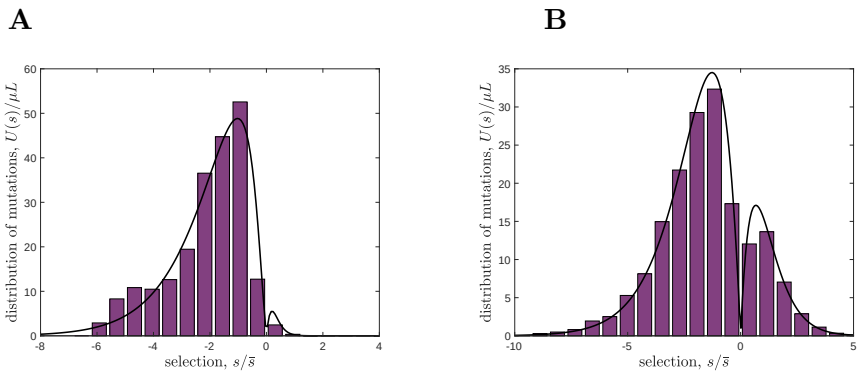

Figure 5.2: **Mutation distribution.** Distribution of mutations $U(s)$ for $L = 100$(**A**) and $L = 5000$(**B**). The distribution of mutations in our genomic model is an outcome of the evolutionary dynamics, since the genomic state determines the direction of the new mutation, see Eq. 5.1. This leads to a bi-modal distribution of mutation effects as beneficial sites can only have deleterious mutations and deleterious sites have beneficial mutations. For large L many more sites are in the deleterious state giving rise to more beneficial mutations.

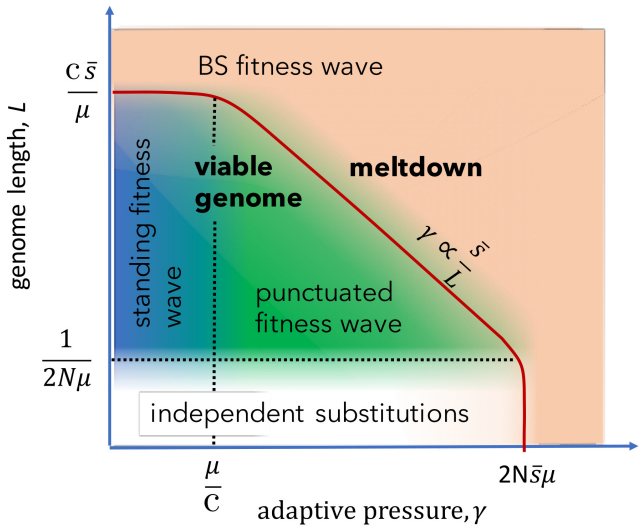

Figure 5.3: **Phase diagram for stationary genomic asexual adaptation.** Different evolutionary modes as a function of the number of loci L and the rate of neofunctionalization events γ. In the limit of very low γ we recover the equilibrium phases as function of the genomic mutation rate: Small values of $\mu L < \frac{1}{2N}$, characterize the regime of independent substitutions, where $\sigma^2 \sim \mu L s$ and $\tilde{\sigma} = 1/2N$. For intermediate values of the genomic mutation rate $\frac{1}{2N} < \mu L < c\bar{s}$ we identified the new regime of standing fitness waves. This region of parameters is associated with a viable genome. By increasing the genomic mutation rate, the variance of fitness crosses over to the regime of infinitesimal fitness waves, associated with the Bolthausen-Sznitman coalescent. Since here $\tilde{\sigma} \gtrsim \bar{s}$, a substantial fraction of the genome cannot be maintained and is then molten. At intermediate values of the genomic mutation rate and increasing values of the adaptive pressure γ, we found the new regime of punctuated fitness waves that is compatible with a viable genome. For high values of γ we cross-over into smooth Bolthausen-Sznitman waves, which are again associated with a molten genome.

5.1 Housekeeping Evolution

Of central importance to the evolutionary success of an organism is its ability to transmit information to its offspring. This can only be achieved, if selection acting on its genes is strong enough to counteract genetic drift or more importantly for asexual organisms genetic draft. The probability of being in the beneficial allele for a given genetic locus is tightly linked to its substitution probability towards the deleterious allele state. For a minimal genomic evolutionary model we consider *housekeeping evolution*, where the selective pressure remains constant for maintaining each genes function and we consider no adaptive pressure. In this minimal genomic model each locus has two states and a selection coefficient s, which describes the fitness difference between the beneficial and deleterious state of this locus. Necessarily, the distribution of mutations $U(s)$ as seen in Fig. 5.2 is an outcome of the evolutionary process. It is determined by the distribution of selection effects of sites $\chi(s)$ and the genomic state $\lambda(s)$

$$U(s) = \mu L \chi(s) \lambda(s). \tag{5.1}$$

Each locus will be in a marginal evolutionary equilibrium, satisfying detailed balance determined by the fixation probability of mutations at the locus.

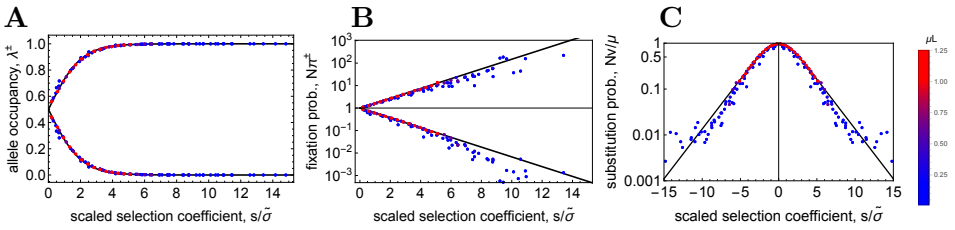

Figure 5.4: **Housekeeping evolution (I): Single site statistics in the standing fitness wave regime.** Collapse plot of **A** Site occupancy of beneficial and deleterious allele $\lambda^{\pm}$, **B** fixation probability $\pi^{\pm}$, and **C** fixation probability $v^{\pm}$ as a function of the rescaled selection coefficient $s/\tilde{\sigma}$ in the regime $\mu NL > 1$. Results from the simulations are compared with the theoretical expectation (black line) of Eq. 5.3, 5.2, and 5.4. Simulation parameters are $N = 1000, 2N\bar{s} = 30, 2N\mu = 0.25$.

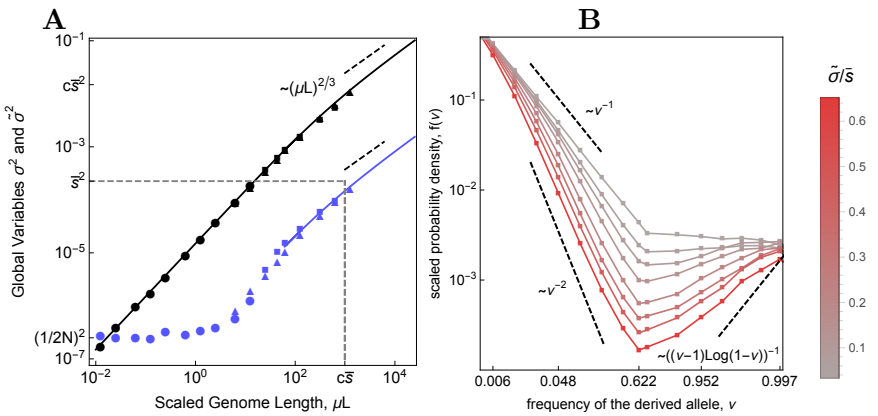

Figure 5.5: **Housekeeping evolution (II):Fitness variance, coalescence rate and frequency spectrum. A** Variance of fitness σ^2 (black) and inverse coalescent time $\tilde{\sigma}$ (blue) as a function of the mutational input/scaled genome length μL. The points are results of simulations (see Methods) and the dashed lines show theoretical expectations of the two known limits: for small μL the variance of fitness has a linear dependence on the mutational input, whereas for large values it shows the 2/3rd power dependence obtained on the universal limit of infinitesimal fitness waves. **B** Frequency-spectrum of derived alleles at neutral sites. It illustrates the crossover from Kingman behavior, with the characteristic inverse frequency decay $(1/\nu)$, for small values of μL to Bolthausen-Sznitman behavior with the characteristic inverse squared frequency decay $(1/\nu^2)$ at small frequencies, and the inverse logarithmic increase $(1/((\nu-1)\log(1-\nu)))$ at large frequencies. Simulation parameters as in Fig. 5.4.

$$\pi_{\pm}(s,\tilde{\sigma}) = \begin{cases} \frac{1}{N} e^{\pm s/2\tilde{\sigma}}, & (\mu NL \gg 1 \,\&\, |s| \ll \tilde{\sigma}) \text{ or } s < -\tilde{\sigma}, \\ \frac{\pm 2s}{1-e^{\pm s/2\tilde{\sigma}}}, & |s| \gg \tilde{\sigma}. \end{cases} \tag{5.2}$$

This fixation probability was shown for mutations in waves [Neher et al., 2013; Good et al., 2014] and in a hierarchical approximation with drivers [Schiffels et al., 2011], i.e. different regimes. For our model, it was reproduced from simulations in [Rice et al., 2015], who also observed an interpolation in the middle regime. The equation defines the efficacy of selection through selected mutations with $s \lesssim \tilde{\sigma}$ acting effectively neutral, hence the neutrality threshold $\tilde{\sigma}$. The threshold has been associated with the coalescence time in previous publications[Neher et al., 2013;

Rice et al., 2015].

For our 2-state model for each locus the probabilities of the beneficial and deleterious allele λ_+ and λ_- immediately follow from the fixation probability

$$\Rightarrow \lambda_\pm(s, \tilde{\sigma}) = \frac{1}{1 + e^{\mp s/\tilde{\sigma}}}. \tag{5.3}$$

This has been shown in [Rice et al., 2015]. We verify that both equations, (5.2) and (5.3) are valid in our system by simulations in figure 5.4.

Together, they define the substitution rates of beneficial and deleterious alleles, $v_\pm(s, \tilde{\sigma}) = \mu\ \lambda_\mp(s, \tilde{\sigma})\Pi_\pm(s, \tilde{\sigma})$,

$$v_\pm(s, \tilde{\sigma}) = \frac{\mu}{2\cosh(s/2\tilde{\sigma})} \begin{cases} \approx \mu & (s \lesssim \tilde{\sigma}) \\ \ll \mu & (s \gtrsim \tilde{\sigma}), \end{cases} \tag{5.4}$$

that clearly satisfy detailed balance. This expression contains important lessons about this system. While substitutions at loci with selection coefficient $s \lesssim \tilde{\sigma}$ occur approximately at a neutral rate μ, substitutions at loci with $s \gtrsim \tilde{\sigma}$ are exponentially rare. The important result following this analysis is that since the dynamics of substitutions at loci with large selection coefficients is essentially frozen, fitness fluctuations are of the order of $\tilde{\sigma}$ or smaller. The resulting substitution patterns stressing this behavior in housekeeping evolution can be seen in Fig. 5.1.

5.1.1 Melting fitness waves

The fitness variance within a population is of central interest in population genetics, since it determines the speed of adaptation. For small μL, because substitutions are independent, the variance of fitness depends linearly on the genomic mutation rate. When μL is large, interference between substitutions render the variance of fitness sublinear in the mutational input. Fisher's fundamental theorem relates the fitness variance to the increase in fitness. For housekeeping evolution the fitness is in equilibrium and the increase in fitness due to the variance σ^2 is balanced by the decrease in fitness due to deleterious mutations ϕ_-. For our genomic model this is

given by

$$\sigma^2 = -\phi_- = \mu\ell \int \mathrm{d}s\, \chi(s) s(\lambda_+ - \lambda_-) = \mu\ell \int \mathrm{d}s\, \chi(s) s \tanh(s/2\tilde{\sigma}). \tag{5.5}$$

The distinction between essentially frozen sites with $s > \tilde{\sigma}$, where substitutions are exponentially suppressed, and molten sites with $s < \tilde{\sigma}$, where substitutions are effectively neutral, lets us split the variance into contributions from stable and unstable sites.

$$\begin{aligned} \sigma^2 &= \mu\ell \int_0^{\tilde{\sigma}} \mathrm{d}s\, \chi(s) s \tanh(s/2\tilde{\sigma}) + \mu\ell \int_{\tilde{\sigma}}^{\infty} \mathrm{d}s\, \chi(s) s \tanh(s/2\tilde{\sigma}) \\ &\approx \mu\ell \int_0^{\tilde{\sigma}} \mathrm{d}s\, \chi(s) \frac{s^2}{2\tilde{\sigma}} + \mu\ell \int_{\tilde{\sigma}}^{\infty} \mathrm{d}s\, \chi(s) s \\ &= \sigma^2_{\text{unstable}} + \sigma^2_{\text{stable}} \end{aligned} \tag{5.6}$$

For a large number of mutations the fitness distribution is approximately Gaussian

$$\mathcal{W}(f) = Z^{-1} \exp\left(\frac{-(f-\overline{f})^2}{2\sigma^2}\right). \tag{5.7}$$

After a time given by the coalescence timescale $\tau_c = \tilde{\sigma}^{-1}$ all present individuals will be descendents of a single individual before this time leading to a self-consistency condition given by

$$\int_{\overline{f}+\frac{d\overline{f}}{dt}t_c}^{\infty} \mathcal{W}(f) df = \frac{1}{N}. \tag{5.8}$$

The speed of the fitness increase is given by Fisher's fundamental theorem $\frac{d\overline{f}}{dt} = \sigma^2$, which leads to

$$\frac{\sqrt{\pi}}{2} \mathrm{erfc}\left(\frac{1}{\sqrt{2}} \frac{\sigma}{\tilde{\sigma}}\right) = \frac{1}{N}. \tag{5.9}$$

Approximating the error function $\mathrm{erfc}(x) \approx \frac{1}{\sqrt{\pi}} \frac{e^{-x^2}}{x}$, we get the fitness variance proportional to the coalescence rate with logarithmic corrections

$$\sigma^2 \sim 2\tilde{\sigma}^2 \log(N\sigma). \tag{5.10}$$

This type of proportionality of the form

$$\sigma^2 = c\tilde{\sigma}^2, \tag{5.11}$$

with a proportionality factor $c = c_0 \log(N\tilde{\sigma})$ is well known from fitness wave theory[Neher et al., 2013; Neher and Hallatschek, 2013] with a model dependent prefactor. It leads to

$$\sigma^2_{\text{molten}} + \sigma^2_{\text{stable}} = c\tilde{\sigma}^2. \tag{5.12}$$

Extending the well known formula for the fitness wave to the prewave regime, we get

$$\tilde{\sigma} = \begin{cases} 1/2N, & \mu LN < 1 \\ \sigma/\sqrt{c}, & \mu LN > 1. \end{cases} \tag{5.13}$$

This extension to the prewave regime is validated by numerical results as the theory obtained from the closure describes the numerical data far into this regime, see Fig. 5.5. In contrast to the classical fitness wave theories, here not all mutations are of small effect. Closing solves all the non-adaptive regimes

$$\frac{\overline{s^2}_{\text{molten}}}{\tilde{\sigma}}(1-\alpha) + \overline{s}_{\text{stable}}\alpha = c\frac{\tilde{\sigma}^2}{\mu L} \tag{5.14}$$

where $\alpha := \int_{\tilde{\sigma}}^{\infty} \mathrm{d}s\, \chi(s)$ is the stable fraction of the genome, $\overline{s^2}_{\text{molten}} := \int_0^{\tilde{\sigma}} \mathrm{d}s\, \chi(s) s^2$, $\overline{s}_{\text{stable}} := \int_{\tilde{\sigma}}^{\infty} \mathrm{d}s\, \chi(s) s$, and $\overline{s^2}_{\text{molten}} \ll \overline{s^2}$ in the pre-wave regime. c sets the melting from below and the transition to the wave at $c \approx 1/2$, where all s-statistics start coinciding with $\tilde{\sigma}$, $s \sim \tilde{\sigma}$. We can now discuss the results in these regimes reproducing the wave [Neher and Hallatschek, 2013].

$$\sigma^2 = \begin{cases} \mu L \bar{s}, & L \lesssim c\bar{s}/\mu \\ c^{1/3}(\mu L \overline{s^2})^{2/3}, & L \gtrsim c\bar{s}/\mu, \end{cases} \tag{5.15}$$

with the factor $c = c_0 \log(N(\mu L \overline{s^2}))$, $c_0 \sim 100$ having a weak dependence on population size. The two regimes correspond to the variance expected for independent sites and the infinitesimal wave regime [Neher and Hallatschek, 2013]. Consistently the frequency spectrum for the derived alleles at neutral loci in the independent substitution regime is given by the scaling known for Kingman coalescence $\frac{1}{\nu}$ and

for the fitness wave by the Bolthausen-Sznitman coalescent scaling $\frac{1}{\nu^2}$ [Neher and Shraiman, 2011; Neher and Hallatschek, 2013], see Fig 5.5B. The crossover between the two regimes occurs for $\mu L \sim \bar{s}$, in agreement with the intuitive picture that interference sets in when the timescale of selection is of the order of the timescale of mutational input.

As shown in Fig. 5.5A, this is also verified in the genomic model. We identify three regimes which differ in the behavior of σ and $\tilde{\sigma}$ as a function of the mutational input. For small values of μL, the genomic model recovers the coalescent time of the single-locus theory and the variance of fitness of the regime of independent substitutions (cf. Eq. 5.15). For very large μL, the variance of fitness depends on the mutational input with the characteristic power dependence found in the asymptotic infinitesimal limit of travelling fitness waves (see Eq. 5.15), and which also sets the coalescent time as given in Eq. 5.13. However, for intermediate values of μL, we find a new regime, where the coalescent time is set by selection, but the variance of fitness still behaves as in the regime of independent substitutions. This regime is due to the melting of frozen sites from below. We obtain an increasing fraction of unstable sites, but the fitness variance is still made up almost entirely by small fluctuations of the stable sites. So in this regime the total fitness variance is still determined by polymorphisms from stable sites with selection $s \gg \tilde{\sigma}$. Due to the coalescence rate in this regime being smaller than the mean site selection coefficient $\tilde{\sigma} < \bar{s}$, the average effect of positive mutations will also be smaller than $\bar{s}$.

5.1.2 Emergent neutrality and genome melting

The crossover between the regime of standing fitness waves and the Bolthausen-Sznitman limit occurs for the same values of μL where $\tilde{\sigma} \gtrsim \bar{s}$ (see Fig. 5.5A and B) This has remarkable consequences in the context of a genome. Indeed, as Eq. 5.3 and Fig. 5.4 show, for $s/\tilde{\sigma} \lesssim 1$, and cf. Eq. 5.4, substitutions occur essentially at the neutral rate in such loci. Thus, as a function of the mutational input, the regime where $\tilde{\sigma} > \bar{s}$ is the regime where the typical locus in this genome is neutral with respect to its substitution rate, and moreover, is not strongly fixed in the beneficial allele. This is clearly visible in Fig. 5.6, where we see a strong reduction in the average fixed beneficial allele. Because this analysis shows this genome to

be unfit for storage of heritable information, it implies that in the framework of a minimal genomic model under housekeeping evolution, the limit of Bolthausen-Sznitman waves is not compatible with a viable genome in housekeeping evolution. This however does not exclude possible scenarios of adaptive evolution, which we consider in the next section.

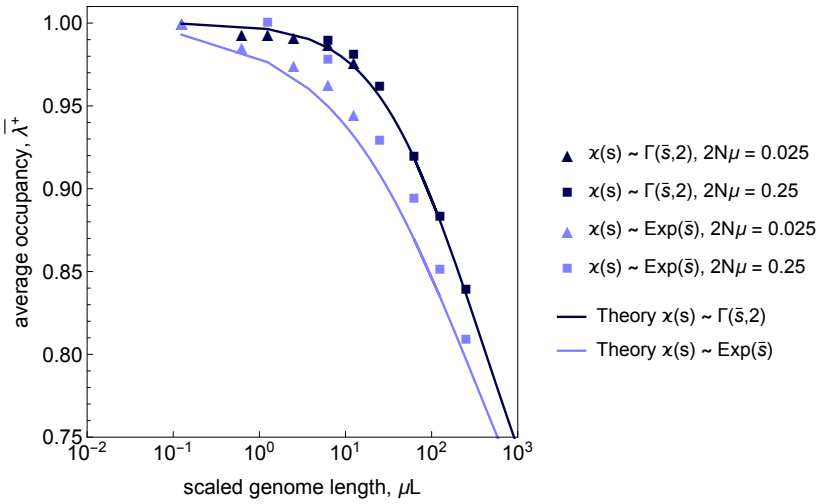

Figure 5.6: **Housekeeping evolution (III):Meltdown in equilibrium.** Mean probability of beneficial allele $\bar{\lambda}^+$ as a function of mutational input μL. The points are simulation results, lines are theory expectations for different site distributions. Simulation parameters as in Fig. 5.4.

5.2 Adaptive evolution and punctuated fitness waves

In nature a surplus of beneficial substitutions leads to adaptive scenarios. Extending the minimal genomic model to adaptive scenarios, we move away from housekeeping evolution. The selection acting on the genome is then time-dependent, which can be quantified by a non-zero fitness flux ϕ[Mustonen and Lässig, 2010].

5.2.1 Minimal model of adaptation and punctuated fitness waves

A simple way of incorporating time-dependent selection in the genomic model is by allowing the direction of selection to flip at individual loci [Mustonen and Lässig,

2007]. We call this change of selection a neofunctionalization event, which occurs at a rate γ. Thus, upon occurrence, the deleterious state of the affected locus becomes beneficial, and the beneficial state becomes deleterious, creating pressure for adaptation. The adaptation can then be described by the cumulative fitness flux $\Phi(\tau) = \int_0^\tau (\mathrm{var}_f(t) + \phi^-(t))\,\mathrm{d}t$, where $\phi^-(t)$ is the negative change of fitness due to mutations.

We start by considering the regime of standing fitness waves, where the genome is still viable, and we add moderate adaptive pressure. As shown in Fig. 5.8, the behavior of the fitness flux (scaled in units of the mean selection coefficient $\bar{s}$) as a function of scaled time (total number of neofunctionalization events $\gamma L \Delta t$) has a clear separation of time and selective scales. It consists of long regions of quasi-equilibrium, where fitness flux has no average increase or decrease, with fluctuations of the order of $\tilde{\sigma}$. These are interrupted by sudden adaptive sweeps, where fluctuations of fitness flux are much larger than $\bar{s}$, yielding a substantial fitness flux increase. The advance in fitness flux is now *punctuated*, and not diffusive as in the limit of infinitesimal fitness waves.

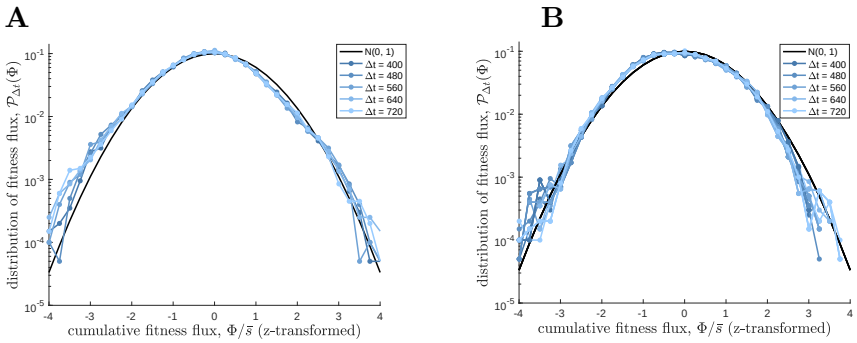

Figure 5.7: **Equilibrium tail.** **A** Equilibrium distribution of fitness flux Φ in units of $\bar{s}$ z-transformed for $L = 100$. **B** Equilibrium distribution of fitness increase z-transformed for $L = 5000$. The line is in both cases the Gaussian expectation. Simulation parameters as in Fig. 5.4.

5.2.2 New universality class: Turbulent fitness waves

The punctuated character of the genomic fitness waves is a direct consequence of the adaptive sweeps. Neofunctionalization events can affect any locus in this genome,

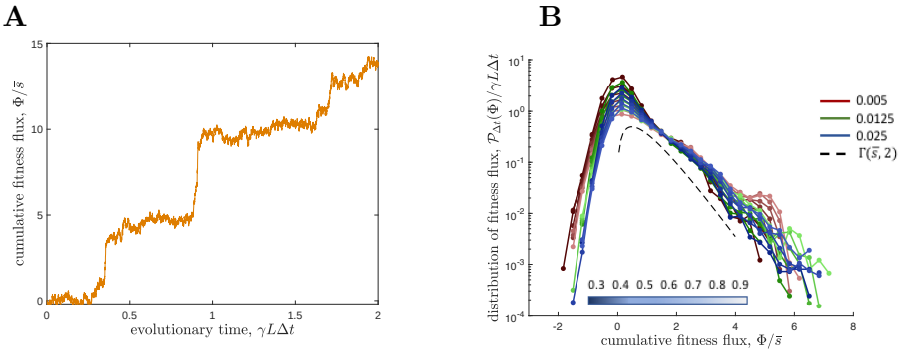

Figure 5.8: **Adaptive evolution (I): New universality class of a punctuated fitness wave. A** Total cumulative fitness flux $\Phi(\tau)$ as a function of evolutionary time τ in adaptive evolution with neofunctionalization rate γ. **B** Rescaled distribution of fitness flux $\mathcal{P}_{\Delta t}(\Phi)$ showing the tail of the distribution for various values of γ and genome size L. Simulation parameters as in Fig. 5.4.

thereby allowing for substitutions at the loci of large effect ($s \gtrsim \bar{s}$) otherwise frozen in equilibrium. Thus, in a time interval Δt larger than the average sweep time $1/\bar{s}$, and smaller than the mean waiting time of a neofuctionalization event $1/\gamma L$, the advance in fitness flux Φ is generated by no-sweep and single sweep configurations, and admits a simple description by a first order expansion in the sweep rate

$$\mathcal{P}_{\Delta t} = (1 - \gamma L\Delta t)\mathcal{P}^{\text{bulk}}_{\Delta t}(\Phi) + (\gamma L\Delta t)\chi(\Phi) + \mathcal{O}((\gamma L\Delta t)^2) \;\; (1/\bar{s} \ll \Delta t \ll 1/\gamma L). \tag{5.16}$$

It depends, to first order, only on the number of neofunctionalization events $\gamma L\Delta t$ and on the input distribution of selection coefficients of the loci χ. This can be directly verified by comparing the probability distribution scaled with $\gamma L\Delta t$ at different times and for systems of different number of loci. As we show in Fig. 5.8, the result is a collapse of curves that recover χ.

This distribution of advance in fitness flux has centered moments with anomalous scaling

$$\langle \Phi^k \rangle^c_{\Delta t} \sim (\gamma L\Delta t)\overline{\Phi^k} + \ldots \;\; (k \geq 1). \tag{5.17}$$

Here all the moments scale linearly with time, illustrating once more the departure from the diffusive behavior, where $\langle \Phi^k \rangle^c \sim (\Delta t)^{k/2}$ for k even and zero otherwise. Remarkably, Eq. 5.17 has the exact same properties of the distribution of

longitudinal velocity differences of processes described by the stochastic Burgers equation in 1D [Bouchaud et al., 1995; Lässig, 2000] . The one-dimensional stochastic Burgers equation is one of the simplest non-linear systems out of equilibrium. It has a multitude of applications in the statistical physics of vehicular dynamics [Debashish Chowdhury, 2000], surface growth [Kardar et al., 1986], directed polymers [Kardar and Zhang, 1987; Bouchaud et al., 1995], and also cosmology [Zel'dovich, 1970; Arnold et al., 1982; Gurbatov and Saichev, 1984; Vergassola et al., 1994]. A key feature of this turbulent theory is the occurence of singular events with bigger scale, which are built up dynamically by the system.

Thus, adaptive genome evolution, by punctuated fitness waves, generates a new evolutionary regime that is different from the waves of Bolthausen-Sznitman process and belong to the same universality class of Burger's turbulence processes.

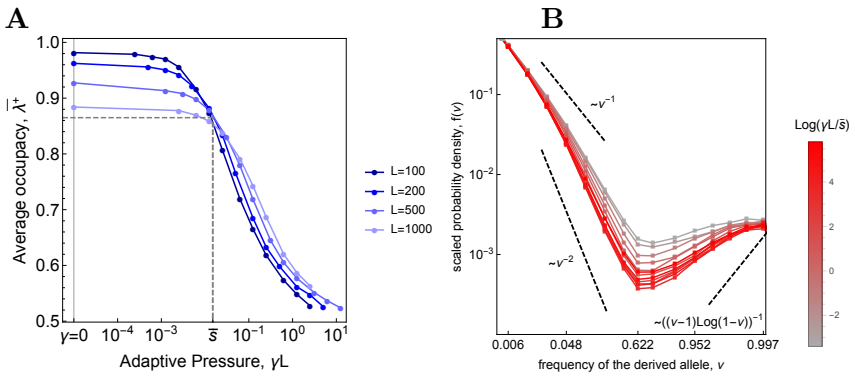

Figure 5.9: **Adaptive evolution (II). A** Mean genomic state of the beneficial allele λ^+ as a function of the adaptive pressure given by the flip / neofunctionalization rate γ. **B** Single site frequency spectrum of the derived allele for neutral sites for different values of γ. We see a crossover from the Kingman towards the Bolthausen-Sznitman frequency spectrum. Other parameters : $2N\mu = 0.25$, $2N\bar{s} = 30$, selection coefficients gamma-distributed ($k = 2$). Simulation parameters as in Fig. 5.4.

5.2.3 Meltdown in non-equilibrium

Once the average time between neofunctionalization events is on the order of the timescale of a single event, interference between these events starts to appear. The clonal interference according to [Schiffels et al., 2011] leads to a neutrality

threshold given by the sweeprate $\alpha\gamma L$ determined by

$$\tilde{\sigma} = \tilde{\sigma}_{\text{eq}} + \alpha\gamma L = \begin{cases} \alpha\gamma L, & \text{in adaptive wave} \\ \tilde{\sigma}_{\text{eq}} \propto \sqrt{\mu L s} & \text{in housekeeping evolution.} \end{cases} \tag{5.18}$$

leading eventually again to a fitness wave behavior, see Fig. 5.9B. This is a consequence of the establishment time of adaptive mutations being smaller than their fixation time[Desai and Fisher, 2007] and several beneficial mutations exist simultaneously leading to this crossover. Studying the neutrality threshold $\tilde{\sigma}$ we see that in this case most individual alleles are again unstable $s \leq \tilde{\sigma}$ and a fitness wave in non-equilibrium is again not a viable long-term mode of evolution, cf. Fig. 5.9A. The actual dynamics in our simulations however does not fully reach the frequencyspectrum associated with fitness waves. This is due to another effect related with high flip rates γ. For flip rates, which are faster than the single site substitution rate at a single locus, we observe a transition towards a neutral site due to the site not being able to react to changes in the environment, called micro-evolution[Takahata et al., 1975; Mustonen and Lässig, 2008]. The probability to be in the beneficial allele shows a transition at $\gamma = 2Ns\mu$ given by

$$\lambda_+ \approx \frac{2Ns\mu + \gamma}{2Ns\mu + 2\gamma}. \tag{5.19}$$

As we are only concerned with the effect on the genomic state and not the actual method of reduced stability of the beneficial allele λ_+, we note that for micro-evolution according to Eq. (5.19) and for adaptive evolution according to Eq. (5.18) for different distributions leads to only small changes for critical meltdown values of λ_+ and fraction of stable sites α, see Fig 5.10. In particular, if an coalescence time of a few hundred generations is observed in experiments [Kryazhimskiy et al., 2014; Barroso-Batista et al., 2014], only alleles with selection coefficients larger than 1/few hundred generations can be maintained on long-term; the rest of the genome gets molten.

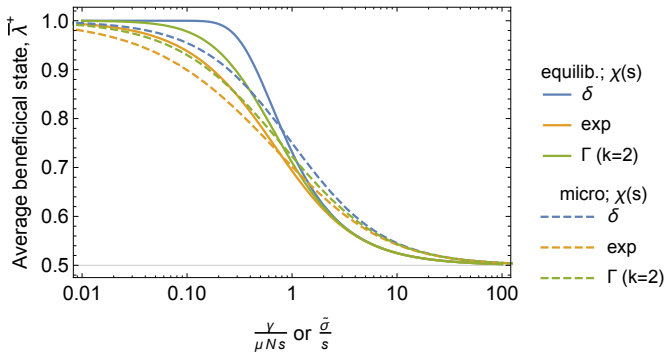

Figure 5.10: **Adaptive evolution (III):Micro-evolution and adaptation.** Average beneficial allele $\lambda^+(\gamma)$ (dashed lines) and $\lambda^+(\tilde{\sigma})$ (solid lines) for all our distributions of s.

5.3 Transient adaptation

In the context of this minimal genomic model, fixations of a large number of deleterious mutations in the regime of fast adaptation makes this mode of evolution incompatible with a viable genome. But clonal interference and coalescent times of a few hundred generations - and therefore fast adaptation, are observed in experiments with asexual populations [Kryazhimskiy et al., 2014; Barroso-Batista et al., 2014]. So what can we learn about evolution by considering our findings in a broader sense?

Our analysis shows that a steady-state regime of fast adaptation is possible to occur over short periods of time. However, this comes with the cost of genetic draft.

The time-scale and effect of genome degradation can be estimated if we consider a hypothetical starting point of a viable genome, where most of the sites are fixed in the beneficial allele. Our housekeeping evolution analysis has shown that sites with selection coefficient $s \leq \tilde{\sigma}$ are essentially neutral, and therefore substitutions occur at a rate μ. Thus, the rate of deleterious substitutions can be written as

$$v_- = (1-\alpha)L\mu, \tag{5.20}$$

where $(1-\alpha)$ is the fraction of sites with $s \leq \tilde{\sigma}$, with the corresponding fitness

decline

$$\dot{F} = -(1-\alpha)L\mu\bar{s}. \tag{5.21}$$

This decline in fitness would occur until the steady state associated with the minimal model of housekeeping evolution would be reached. For *E. Coli*, where $L \sim 10^6$ and $\mu \sim 10^{-10}$ per bp per replication, and assuming the order of 10 generations per day, this yields the loss of approximaterly one locus per year and fitness decline, per year, of the order of the mean selection coefficient of the genome. While such effects wouldn't appear to be strong even in very long-term experiments in the lab, evolution in the wild would have enough time to degrade the genomes. This leads to an important implication: since fast adaptation as observed in laboratory populations is only sustainable for short periods of time, this mode of asexual evolution must be transient.

5.4 Discussion

We have shown that adaptive steady states have to be characterized by punctuated adaptation. This has far reaching implications: First of all, this means that for long-term processes in non-recombining populations with Mendelian traits the classical fitness wave theory cannot be applied as this would require an unstable genome. However, fast adaptation according to fitness waves has been observed in laboratory experiments. This is perfectly compatible with our results as long as this mode of evolution is only transient. Short-term fitness wave evolution is permissible, but comes with the cost of degradation according to Eq. 5.21. This degradation has been observed in experiments.

Introducing broad epistasis in the form of quantitative traits into the system introduces a natural separation of scales between the selective effect of individual mutations and the selection on the whole phenotype. Under these circumstances quantitative trait loci can become effectively neutral, while the whole trait is still under efficient selection. The negative effect of fast adaptation is a direct result of the non-recombining evolutionary process. Different modes of recombination, i.e. horizontal gene transfer, are able to break the linkage between loci and can then reduce the negative effects of fast adaptation. The machinery to break linkage comes with its own costs, which have to be offset by the gain of possible fast

long-term adaptation.

In the final chapter we recapitulate the main results of this thesis and answer the guiding questions raised in the introduction.

5.5 Methods

5.5.1 Simulations

We implement our minimal model of genome evolution as a Fisher-Wright simulation with fixed population size N and individuals with a sequence $\mathbf{a} = (a_1, \ldots, a_L)$ of size L with a binary alphabet, so that $a_i \in 0, 1$ for $i \in [1; L]$. The fitness of a sequence is given by $f(\mathbf{a}) = \sum_{i=1}^{L} s_i a_i$. In each generation, the sequences undergo point mutations with probability $\mu\tau_0$ for each site, where τ_0 is the generation time, and the sequences of the next generation are drawn by multinomial sampling with probabilities proportional to $1 + \tau_0 f(\mathbf{a})$. Further each generation each site has a probability $\tau_0\gamma$ to flip the preferred allele at each site, effectively changing s_i to $-s_i$.

6 Discussion

"M is for magic. All the letters are, if you put them together properly. You can make magic with them, and dreams, and, I hope, even a few surprises..."

(Neil Gaiman, M Is for Magic)

We have provided multiple ways, to improve the current status quo of evolutionary theory by joining together previously disjoint fields within population genetics.

In chapter 3 we have expanded the theory of quantitative traits to include the impact of fast adaptation in its genomic context. For the most this can be achieved through a single parameter, the neutrality threshold as seen in Eq. (3.7). Using the diversity of neutral sites we are able to infer the neutrality threshold as shown in Eq. (3.31). This confirms previous results for the neutral sequence diversity from fitness wave theories mapped onto a quantitative trait [Good et al., 2014; Rice et al., 2015].

In chapter 4 we have investigated a minimal realistic model of an asexual organism with many linked genes with biophysical fitness landscapes and shown that the fitness load scales quadratically with the number of genes, Eq. (4.5). The faster-than-linear scaling does not depend on the particular form of the fitness landscape and only takes into account the number of independent biological functions, giving a measure for the complexity of an organism. The correlation due to linkage of different sites can be absorbed in an effective independent evolution model with a given neutrality threshold $\tilde{\sigma}$ as can be seen in Fig. 4.2. Surprisingly, even a small recombination rate on the order of the mutation rate, Eq. (4.8), is already enough for a phase transition from phenotypic interference to the limit of independent gene evolution under full recombination, suggesting an avenue for the evolution of sex according to the difference in genetic load over the long-term. In our analysis

we have seen that fitness waves naturally occur in a minimal housekeeping scenario for phenotypic interference. In this context there is a separation between the selective scale of the individual mutations, which is given by the coalescence rate $\tilde{\sigma}$, Eq. (4.1), and the bigger selection of the whole trait. Thus quantitative traits can have near neutral single mutations and still describe stable genes as long as the selection on the whole gene is strong enough .

For Mendelian traits no separation of scales is possible in the selection of the traits. In chapter 5 we have shown that fast adaptation in a genomic context leads to a loss of genes and thus a corresponding fitness decline, Eq. (5.21). Thus the mode of fast adaptation as seen in evolutionary experiments can only be transient. An adaptive mode that does not degrade the genome over the long-term, is the punctuated fitness wave that is characterized by phases of maintenance of function interrupted by short adaptive bursts. This mode of evolution is further distinguished by anomalous scaling of the moments of the fitness flux as known from Burger's turbulence, Eq. (5.17).

To conclude we have shown how the drastic effects of asexual linkage incorporated in the neutrality threshold link our different chapters together. As a main result fitness wave theory is a valuable tool to describe the asexual evolution of short term laboratory evolution experiments, but fails when it comes to long-term processes in nature that do not incorporate quantitative traits. The combination of fitness wave theories together with phenotypic and Mendelian genomic models can lead to fascinating new insights to evolutionary modes – showing surprising new features like phenotypic interference and the melting of genomic sites from below.

Bibliography

V. I. Arnold, S. F. Shandarin, and Y. B. Zeldovich. The large scale structure of the universe i. general properties. one-and two-dimensional models. *Geophysical & Astrophysical Fluid Dynamics*, 20(1-2):111–130, 1982. doi: 10.1080/03091928208209001.

J. Barroso-Batista, A. Sousa, M. Lourenço, M.-L. Bergman, D. Sobral, J. Demengeot, K. B. Xavier, and I. Gordo. The first steps of adaptation of escherichia coli to the gut are dominated by soft sweeps. *PLos Genet*, 10(3):1–12, 03 2014. doi: 10.1371/journal.pgen.1004182.

N. H. Barton and H. P. de Vladar. Statistical mechanics and the evolution of polygenic quantitative traits. *Genetics*, 181(3):997–1011, 2009.

N. H. Barton and M. Turelli. Evolutionary quantitative genetics: How little do we know? *Annu Rev Genet*, 23(1):337–370, dec 1989. doi: 10.1146/annurev.ge.23.120189.002005.

J. Berg, S. Willmann, and M. Lässig. Adaptive evolution of transcription factor binding sites. *BMC Evol Biol*, 4(1):42, 2004. ISSN 1471-2148. doi: 10.1186/1471-2148-4-42.

A. J. Betancourt, J. J. Welch, and B. Charlesworth. Reduced effectiveness of selection caused by a lack of recombination. *Current Biology*, 19(8):655–660, Apr. 2009. doi: 10.1016/j.cub.2009.02.039.

J. P. Bouchaud, M. Mézard, and G. Parisi. Scaling and intermittency in burgers turbulence. *Phys Rev E*, 52:3656–3674, Oct 1995. doi: 10.1103/PhysRevE.52.3656.

E. Brunet, B. Derrida, A. H. Mueller, and S. Munier. Noisy traveling waves: Effect of selection on genealogies. *EPL (Europhysics Letters)*, 76(1):1, 2006.

É. Brunet, B. Derrida, A. H. Mueller, and S. Munier. Effect of selection on ancestry: An exactly soluble case and its phenomenological generalization. *Physical Review E*, 76(4), oct 2007. doi: 10.1103/physreve.76.041104.

M. G. Bulmer. The genetic variability of polygenic characters under optimizing selection, mutation and drift. *Genet Res*, 19(01):17, feb 1972. doi: 10.1017/s0016672300014221.

B. Charlesworth. The effects of deleterious mutations on evolution at linked sites. *Genetics*, 190(1):5–22, 2012. ISSN 0016-6731. doi: 10.1534/genetics.111.134288.

P. Chen and E. I. Shakhnovich. Lethal mutagenesis in viruses and bacteria. *Genetics*, 183(2):639–650, 2009. doi: 10.1534/genetics.109.106492.

N. Chéron, A. W. R. Serohijos, J.-M. Choi, and E. I. Shakhnovich. Evolutionary dynamics of viral escape under antibodies stress: A biophysical model. *Protein Sci*, 25(7):1332–1340, 2016. ISSN 1469-896X. doi: 10.1002/pro.2915.

L.-M. Chevin and F. Hospital. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics*, 180(3):1645–1660, 2008. doi: 10.1534/genetics.108.093351.

P. B. Chi and D. A. Liberles. Selection on protein structure, interaction, and sequence. *Protein Sci*, 25(7):1168–1178, 2016. ISSN 1469-896X. doi: 10.1002/pro.2886.

E. Cohen, D. A. Kessler, and H. Levine. Front propagation up a reaction rate gradient. *Phys. Rev. E*, 72:066126, Dec 2005. doi: 10.1103/PhysRevE.72.066126.

J. M. Comeron, R. Ratnappan, and S. Bailin. The many landscapes of recombination in *drosophila melanogaster*. *PLos Genet*, 8(10):1–21, 10 2012. doi: 10.1371/journal.pgen.1002905.

T. F. Cooper. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *escherichia coli*. *PLoS Biol*, 5(9):1–7, 08 2007. doi: 10.1371/journal.pbio.0050225.

A. Couce, L. V. Caudwell, C. Feinauer, T. Hindré, J.-P. Feugeas, M. Weigt, R. E. Lenski, D. Schneider, and O. Tenaillon. Mutator genomes decay, despite

sustained fitness gains, in a long-term experiment with bacteria. *Proc Natl Acad Sci*, 114(43):E9026–E9035, 2017. doi: 10.1073/pnas.1705887114.

A. G. J. M. de Visser, C. W. Zeyl, P. J. Gerrish, J. L. Blanchard, and R. E. Lenski. Diminishing returns from mutation supply rate in asexual populations. *Science*, 283(5400):404–406, 1999. ISSN 0036-8075. doi: 10.1126/science.283.5400.404.

H. P. de Vladar and N. Barton. Stability and response of polygenic traits to stabilizing selection and mutation. *Genetics*, 197(2):749–767, 2014. ISSN 0016-6731. doi: 10.1534/genetics.113.159111.

H. P. de Vladar and N. H. Barton. The statistical mechanics of a polygenic character under stabilizing selection, mutation and drift. *J R Soc Interface*, 8 (58):720–739, Mar. 2011.

A. S. Debashish Chowdhury, Ludger Santen. Statistical physics of vehicular traffic and some related systems. *Physics Reports*, 329(4-6):199–329, may 2000. doi: 10.1016/s0370-1573(99)00117-9.

E. Derelle, C. Ferraz, S. Rombauts, P. Rouzé, A. Z. Worden, S. Robbens, F. Partensky, S. Degroeve, S. Echeynié, R. Cooke, Y. Saeys, J. Wuyts, K. Jabbari, C. Bowler, O. Panaud, B. Piégu, S. G. Ball, J.-P. Ral, F.-Y. Bouget, G. Piganeau, B. De Baets, A. Picard, M. Delseny, J. Demaille, Y. Van de Peer, and H. Moreau. Genome analysis of the smallest free-living eukaryote ostreococcus tauri unveils many unique features. *Proc Natl Acad Sci U S A*, 103(31): 11647–11652, Jan. 2006. ISSN 1091-6490.

M. M. Desai and D. S. Fisher. Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics*, 176(3):1759–1798, 2007. doi: 10.1534/genetics.106.067678.

M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, 1971. ISSN 1432-1904. doi: 10.1007/BF00623322.

W. Ewens. *Mathematical Population Genetics: I. Theoretical Introduction*. Interdisciplinary Applied Mathematics. Springer, 2004. ISBN 9780387201917.

J. Felsenstein. The evolutionary advantage of recombination. *Genetics*, 78(2): 737–756, Mar. 1974. ISSN 0016-6731.

J. Feng, D. Q. Naiman, and B. Cooper. Coding dna repeated throughout intergenic regions of the arabidopsis thaliana genome: evolutionary footprints of rna silencing. *Mol Biosyst*, 5:1679–1687, 2009. doi: 10.1039/B903031J.

R. A. Fisher. *The genetical theory of natural selection.* Oxford Clarendon Press, 1930. doi: 10.5962/bhl.title.27468.

A.-S. Fiston-Lavier, N. D. Singh, M. Lipatov, and D. A. Petrov. Drosophila melanogaster recombination rate calculator. *Gene*, 463(1–2):18 – 20, 2010. ISSN 0378-1119. doi: http://dx.doi.org/10.1016/j.gene.2010.04.015.

U. Gerland and T. Hwa. On the selection and evolution of regulatory DNA motifs. *J Mol Evol*, 55(4):386–400, 2002.

P. J. Gerrish and R. E. Lenski. The fate of competing beneficial mutations in an asexual population. *Genetica*, 102:127–144, Mar 1998. ISSN 1573-6857. doi: 10.1023/A:1017067816551.

J. H. Gillespie. The neutral theory in an infinite population. *Gene*, 261(1):11–18, dec 2000a. doi: 10.1016/s0378-1119(00)00485-6.

J. H. Gillespie. Genetic drift in an infinite population: The pseudohitchhiking model. *Genetics*, 155(2):909–919, 2000b. ISSN 0016-6731.

R. A. Goldstein. The evolution and evolutionary consequences of marginal thermostability in proteins. *Proteins: Struct , Funct , Bioinf*, 79(5):1396–1407, 2011. ISSN 1097-0134. doi: 10.1002/prot.22964.

B. H. Good and M. M. Desai. Deleterious passengers in adapting populations. *Genetics*, 198(3):1183–1208, 2014. doi: 10.1534/genetics.114.170233.

B. H. Good, I. M. Rouzine, D. J. Balick, O. Hallatschek, and M. M. Desai. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proceedings of the National Academy of Sciences*, 109(13):4950–4955, 2012. doi: 10.1073/pnas.1119910109.

B. H. Good, A. M. Walczak, R. A. Neher, and M. M. Desai. Genetic diversity in the interference selection limit. *PLos Genet*, 10(3):e1004222, 03 2014. doi: 10.1371/journal.pgen.1004222.

I. Gordo and B. Charlesworth. The degeneration of asexual haploid populations and the speed of Muller's ratchet. *Genetics*, 154(3):1379–1387, 03 2000.

S. N. Gurbatov and A. I. Saichev. Probability distribution and spectra of potential hydrodynamic turbulence. *Radiophysics and Quantum Electronics*, 27(4):303–313, Apr 1984. ISSN 1573-9120. doi: 10.1007/BF01036611.

O. Hallatschek. The noisy edge of traveling waves. *Proc Natl Acad Sci*, 108(5): 1783–1787, 2011. doi: 10.1073/pnas.1013529108.

P. M. Harrison, D. Milburn, Z. Zhang, P. Bertone, and M. Gerstein. Identification of pseudogenes in the drosophila melanogaster genome. *Nucleic Acids Res*, 31 (3):1033–1037, 2003. doi: 10.1093/nar/gkg169.

D. L. Hartl and C. H. Taubes. Compensatory nearly neutral mutations: Selection without adaptation. *J Theor Biol*, 182(3):303–309, oct 1996. doi: 10.1006/jtbi.1996.0168.

T. Held, A. Nourmohammad, and M. Lässig. Adaptive evolution of molecular phenotypes. *J Stat Mech: Theory Exp*, 2014(9):P09029, 2014.

T. Held, D. Klemmer, and M. Lässig. Survival of the simplest: the cost of complexity in microbial evolution. *bioRxiv*, 2018. doi: 10.1101/287599. Equal contribution by TH and DK.

W. G. Hill and A. Robertson. The effect of linkage on limits to artificial selection. *Genetics Research*, 8(03):269–294, 1966. doi: 10.1017/S0016672300010156.

M. Hochstrasser. Ubiquitin-dependent protein degradation. *Annu Rev Genet*, 30: 405–439, 1996. ISSN 0066-4197. doi: 10.1146/annurev.genet.30.1.405.

S. Hwang, S.-C. Park, and J. Krug. Genotypic complexity of fisher's geometric model. *Genetics*, 206(2):1049–1079, 2017. ISSN 0016-6731. doi: 10.1534/genetics.116.199497.

M. Kardar and Y.-C. Zhang. Scaling of directed polymers in random media. *Phys Rev Lett*, 58:2087–2090, May 1987. doi: 10.1103/PhysRevLett.58.2087.

M. Kardar, G. Parisi, and Y.-C. Zhang. Dynamic scaling of growing interfaces. *Phys Rev Lett*, 56:889–892, Mar 1986. doi: 10.1103/PhysRevLett.56.889.

P. D. Keightley, R. W. Ness, D. L. Halligan, and P. R. Haddrill. Estimation of the spontaneous mutation rate per nucleotide site in a drosophila melanogaster full-sib family. *Genetics*, 196(1):313–320, 2014. ISSN 0016-6731. doi: 10.1534/genetics.113.158758.

M. Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47(6):713–719, 1962.

M. Kimura. Diffusion models in population genetics. *J Appl Probab*, 1:177, 1964.

J. B. Kinney, A. Murugan, C. G. Callan, and E. C. Cox. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc Natl Acad Sci*, 107(20):9158–9163, 2010. doi: 10.1073/pnas.1004290107.

A. S. Kondrashov. Classification of hypotheses on the advantage of amphimixis. *J Hered*, 84(5):372–387, 1993. doi: 10.1093/oxfordjournals.jhered.a111358.

S. Kryazhimskiy, D. P. Rice, E. R. Jerison, and M. M. Desai. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*, 344 (6191):1519–1522, 2014. ISSN 0036-8075. doi: 10.1126/science.1250939.

R. Lande. Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 30(2):314–334, 1976.

M. Lässig. Dynamical anomalies and intermittency in burgers turbulence. *Physical Review Letters*, 84(12):2618–2621, mar 2000. doi: 10.1103/physrevlett.84.2618.

M. Lässig. From biophysics to evolutionary genetics: statistical aspects of gene regulation. *BMC Bioinformatics*, 8(Suppl 6):S7, Sept. 2007. doi: 10.1186/1471-2105-8-s6-s7.

M. Lukačišinová and T. Bollenbach. Toward a quantitative understanding of antibiotic resistance evolution. *Curr Opin Biotechnol*, 46:90–97, aug 2017. doi: 10.1016/j.copbio.2017.02.013.

M. Luksza and M. Lässig. A predictive fitness model for influenza. *Nature*, 507 (7490):57–61, feb 2014. doi: 10.1038/nature13087.

M. Lynch and G. K. Marinov. The bioenergetic costs of a gene. *Proc Natl Acad Sci*, 112(51):15690–15695, 2015. doi: 10.1073/pnas.1514974112.

M. Lynch and B. Walsh. *Genetics and Analysis of Quantitative Traits*. Oxford University Press, 1998. ISBN 0878934812.

M. Lynch, W. Sung, K. Morris, N. Coffey, C. R. Landry, E. B. Dopman, W. J. Dickinson, K. Okamoto, S. Kulkarni, D. L. Hartl, and W. K. Thomas. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci*, 105(27):9272–9277, 2008. doi: 10.1073/pnas.0803466105.

T. F. C. Mackay. Quantitative trait loci in drosophila. *Nat Rev Genet*, 2(1):11–20, jan 2001. doi: 10.1038/35047544.

M. Manhart and A. V. Morozov. Protein folding and binding can emerge as evolutionary spandrels through structural coupling. *Proc Natl Acad Sci*, 112(6): 1797–1802, 2015. doi: 10.1073/pnas.1415895112.

G. Martin, S. F. Elena, and T. Lenormand. Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat Genet*, 39(4):555–560, mar 2007. doi: 10.1038/ng1998.

M. J. McDonald, D. P. Rice, and M. M. Desai. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature*, 531:233–236, 02 2016. doi: 10.1038/nature17143.

J. Monod, J. Wyman, and J.-P. Changeux. On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*, 12(1):88–118, 1965. ISSN 0022-2836. doi: 10.1016/s0022-2836(65)80285-6.

D. H. Morris, K. M. Gostic, S. Pompei, T. Bedford, M. Luksza, R. A. Neher, B. T. Grenfell, M. Lässig, and J. W. McCauley. Predictive modeling of influenza shows the promise of applied evolutionary biology. *Trends in Microbiology*, 26 (2):102–118, feb 2018. doi: 10.1016/j.tim.2017.09.004.

H. J. Muller. Some genetic aspects of sex. *Am Nat*, 66(703):118–138, 1932. ISSN 00030147, 15375323.

H. J. Muller. The relation of recombination to mutational advance. *Mutat Res*, 106:2–9, 1964.

V. Mustonen and M. Lässig. Adaptations to fluctuating selection in drosophila. *Proceedings of the National Academy of Sciences*, 104(7):2277–2282, 2007.

V. Mustonen and M. Lässig. Molecular evolution under fitness fluctuations. *Phys Rev Lett*, 100(10), mar 2008. doi: 10.1103/physrevlett.100.108101.

V. Mustonen and M. Lässig. Fitness flux and ubiquity of adaptive evolution. *Proc Natl Acad Sci*, 107:4248–53, 2010.

V. Mustonen, J. Kinney, C. G. J. Callan, and M. Lässig. Energy-dependent fitness: A quantitative model for the evolution of yeast transcription factor binding sites. *Proc. Natl. Acad. Sci. USA*, 105(34):12376–12381, 2008. doi: 10.1073/pnas.0805909105.

R. A. Neher. Genetic draft, selective interference, and population genetics of rapid adaptation. *Annu Rev Ecol Evol Syst*, 44(1):195–215, 2013. doi: 10.1146/annurev-ecolsys-110512-135920.

R. A. Neher and O. Hallatschek. Genealogies of rapidly adapting populations. *Proceedings of the National Academy of Sciences*, 110(2):437–442, 2013. doi: 10.1073/pnas.1213113110.

R. A. Neher and B. I. Shraiman. Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics*, 188(4):975–996, 2011. ISSN 0016-6731. doi: 10.1534/genetics.111.128876.

R. A. Neher, T. A. Kessinger, and B. I. Shraiman. Coalescence and genetic diversity in sexual populations under selection. *Proceedings of the National Academy of Sciences*, 110(39):15836–15841, 2013. doi: 10.1073/pnas.1309697110.

A. Nour Mohammad, S. Schiffels, and M. Lässig. Evolution of molecular phenotypes under stabilizing selection. *J Stat Mech Theor Exp*, 2013(01):P01012, jan 2013. doi: 10.1088/1742-5468/2013/01/P01012.

A. Nourmohammad, T. Held, and M. Laessig. Universality and predictability in molecular quantitative genetics. *Current opinion in genetics & development*, 23(6):684–693, 2013.

S. Ossowski, K. Schneeberger, J. I. Lucas-Lledó, N. Warthmann, R. M. Clark, R. G. Shaw, D. Weigel, and M. Lynch. The rate and molecular spectrum of spontaneous mutations in arabidopsis thaliana. *Science*, 327(5961):92–94, 2009. ISSN 0036-8075. doi: 10.1126/science.1180677.

J. Otwinowski. Inferring protein stability and function from a high-throughput assay. *arXiv preprint*, 1802.08744, 2018.

S.-C. Park and J. Krug. Clonal interference in large populations. *Proceedings of the National Academy of Sciences*, 104(46):18135–18140, 2007. doi: 10.1073/pnas.0705778104.

L. Perfeito, L. Fernandes, C. Mota, and I. Gordo. Adaptive mutations in bacteria: High rate and small effects. *Science*, 317(5839):813–815, 2007. ISSN 0036-8075. doi: 10.1126/science.1142284.

L. Perfeito, A. Sousa, T. Bataillon, and I. Gordo. Rates of fitness decline and rebound suggest pervasive epistasis. *Evolution*, 68(1):150–162, sep 2013. doi: 10.1111/evo.12234.

R. Phillips, J. Kondev, J. Theriot, and N. Orme. *Physical Biology of the Cell*. Garland Science, 2013. ISBN 9780815344506.

R. Plomin, C. M. A. Haworth, and O. S. P. Davis. Common disorders are quantitative traits. *Nat Rev Genet*, 10(12):872–878, oct 2009. doi: 10.1038/nrg2670.

D. P. Rice, B. H. Good, and M. M. Desai. The evolutionarily stable distribution of fitness effects. *Genetics*, 200(1):321–329, 2015. doi: 10.1534/genetics.114.173815.

S. H. Rice. A geometric model for the evolution of development. *J Theor Biol*, 143 (3):319–342, apr 1990. doi: 10.1016/s0022-5193(05)80033-5.

I. M. Rouzine, A. Rodrigo, and J. M. Coffin. Transition between stochastic evolution and deterministic evolution in the presence of selection: General theory

and application to virology. *Microbiology and Molecular Biology Reviews*, 65(1): 151–185, mar 2001. doi: 10.1128/mmbr.65.1.151-185.2001.

I. M. Rouzine, É. Brunet, and C. O. Wilke. The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. *Theor Popul Biol*, 73(1):24 – 46, 2008. ISSN 0040-5809. doi: http://dx.doi.org/10.1016/j.tpb.2007.10.004.

D. M. Ruderfer, S. C. Pratt, H. S. Seidel, and L. Kruglyak. Population genomic analysis of outcrossing and recombination in yeast. *Nat Genet*, 38(9):1077–1081, Sept. 2006. ISSN 1061-4036.

P. A. Salome, K. Bomblies, J. Fitz, R. A. E. Laitinen, N. Warthmann, L. Yant, and D. Weigel. The recombination landscape in arabidopsis thaliana f2 populations. *Heredity*, 108(4):447–455, Apr. 2012. ISSN 0018-067X.

S. Schiffels, G. J. Szöllősi, V. Mustonen, and M. Lässig. Emergent neutrality in adaptive asexual evolution. *Genetics*, 189(4):1361–1375, 2011. doi: 10.1534/genetics.111.132027.

S. Schiffels, V. Mustonen, and M. Lässig. The asexual genome of Drosophila. *ArXiv e-prints*, (1711.10849), Nov. 2017.

S. Schoustra, S. Hwang, J. Krug, and J. A. G. M. de Visser. Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus. *Proc Biol Sci*, 283(1837):20161376, aug 2016. doi: 10.1098/rspb.2016.1376.

M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330(6007):1099–1102, 2010. ISSN 0036-8075. doi: 10.1126/science.1192588.

A. W. Serohijos and E. I. Shakhnovich. Merging molecular mechanism and evolution: theory and computation at the interface of biophysics and evolutionary population genetics. *Curr Opin Struct Biol*, 26(0):84 – 91, 2014. ISSN 0959-440X. doi: http://dx.doi.org/10.1016/j.sbi.2014.05.005. New constructs and expression of proteins / Sequences and topology.

N. Strelkowa and M. Lässig. Clonal interference in the evolution of influenza. *Genetics*, 192(2):671–682, 2012. ISSN 0016-6731. doi: 10.1534/genetics.112.143396.

N. Takahata, K. Ishii, and H. Matsuda. Effect of temporal fluctuation of selection coefficient on gene frequency in a population. *Proc Natl Acad Sci*, 72(11): 4541–4545, nov 1975. doi: 10.1073/pnas.72.11.4541.

S. D. Tanksley. Mapping polygenes. *Annu Rev Genet*, 27(1):205–233, dec 1993. doi: 10.1146/annurev.ge.27.120193.001225.

O. Tenaillon. The utility of fisher's geometric model in evolutionary genetics. *Annual Review of Ecology, Evolution, and Systematics*, 45:179–201, 11 2014.

N. Tokuriki, F. Stricher, J. Schymkowitz, L. Serrano, and D. S. Tawfik. The stability effects of protein mutations appear to be universally distributed. *J Mol Biol*, 369(5):1318–1332, 2007. ISSN 0022-2836. doi: http://dx.doi.org/10.1016/j.jmb.2007.03.069.

M. Tuğrul, T. Paixão, N. H. Barton, and G. Tkačik. Dynamics of transcription factor binding site evolution. *PLoS Genet*, 11(11):1–28, 11 2015. doi: 10.1371/journal.pgen.1005639.

M. Vergassola, B. Dubrulle, U. Frisch, and A. Noullez. Burgers' equation, Devil's staircases and the mass distribution for large-scale structures. *Astron Astrophys*, 289:325–356, Sept. 1994.

D. M. Weinreich and J. L. Knies. Fisher's geometric model of adaptation meets the functional synthesis: Data on pairwise epistasis for fitness yields insights into the shape and size of phenotype space. *Evolution*, pages n/a–n/a, jun 2013. ISSN 0014-3820. doi: 10.1111/evo.12156.

D. B. Weissman and N. H. Barton. Limits to the rate of adaptive substitution in sexual populations. *PLos Genet*, 8(6):1–18, 06 2012. doi: 10.1371/journal.pgen.1002740.

M. J. Wiser, N. Ribeck, and R. E. Lenski. Long-term dynamics of adaptation in asexual populations. *Science*, 342(6164):1364–1367, 2013. ISSN 0036-8075. doi: 10.1126/science.1243357.

S. Wright. Evolution in mendelian populations. *Genetics*, 16(2):97–159, 03 1931.

S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proceedings of the Sixth International Congress of Genetics*, volume 1, pages 356–66, 1932.

K. B. Zeldovich, P. Chen, and E. I. Shakhnovich. Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proc Natl Acad Sci*, 104(41):16152–16157, 2007. doi: 10.1073/pnas.0705366104.

Y. B. Zel'dovich. Gravitational instability: An approximate theory for large density perturbations. *Astron Astrophys*, 5:84–89, Mar. 1970.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Michael Lässig betreut worden.

Es liegen keine Teilpublikationen vor.

Promotionsbeginn: 02/2013
Promotionsende: 10/2018

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Veränderungen, dem Dekanat unverzüglich mitzuteilen.

.....

Datum Unterschrift