Nonlinear effects of recombination on the evolutionary dynamics in Wright-Fisher type models and experiments

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Julian Alexander Klug

aus Bonn

Berichterstatter:
Prof. Dr. Joachim Krug
Prof. Dr. Thomas Wiehe

Tag der mündlichen Prüfung: 4.3.2022

Kurzzusammenfassung

Fitnesslandschaften stellen eine Abbildung zwischen Genotypen und ihrer Fitness dar, die in der Regel ihren Fortpflanzungserfolg widerspiegelt. Mit diesem Ansatz kann Evolution als ein Prozess betrachtet werden, bei dem sich Populationen auf einer Fitnesslandschaft bewegen. Welche Wege in diesem Prozess eingeschlagen werden, hängt dabei entscheidend von deren Struktur ab. Obwohl die Idee schon recht alt ist, hat sie in den letzten Jahren an Bedeutung gewonnen, da es durch Fortschritte in der genetischen Sequenzierung möglich geworden ist, die Struktur von Fitnesslandschaften genauer zu erfassen. Dies wiederum ermöglicht ein umfassenderes und quantitatives Verständnis der Evolution. Es ist jedoch immer noch unklar, wie diese Landschaften in großem Maßstab aussehen. In dieser Dissertation werden daher theoretische sowie empirische Fitnesslandschaften betrachtet und dabei untersucht, wie sich Populationen über diese verteilen. Es wird gezeigt, dass hierbei nicht nur die Struktur der Fitnesslandschaft entscheidend ist, sondern auch, welche evolutionären Kräfte am Werk sind und wie stark. Insbesondere wird auf den Effekt der Rekombination genetischen Materials eingegangen und ein möglicher Vorteil von Rekombination beschrieben, der bisher vergleichsweise wenig Beachtung gefunden hat. Dies ist die Robustheit gegenüber dem Effekt von zufälligen Mutationen, welche in rekombinierenden Populationen größer ist.

Abstract

Fitness landscapes represent a mapping between genotypes and their fitness, which usually reflects their reproductive success. With this approach, evolution can be considered as a process in which populations move on a fitness landscape. The paths taken in this process depend crucially on their structure. Although the idea is quite old, it has gained renewed attention in recent years as advances in genetic sequencing have made it possible to capture the structure of fitness landscapes in greater detail. This in turn facilitates a more comprehensive and quantitative understanding of evolution. However, it is still unclear how these landscapes are structured on a large scale. This dissertation therefore considers theoretical as well as empirical fitness landscapes and investigates how populations are distributed across them. It is shown that not only the structure of the fitness landscape is crucial, but also which evolutionary forces are at work and how strong they are. In particular, the effect of recombination of genetic material is addressed and a possible advantage of recombination is described that has received comparatively little attention so far. This is the robustness against the effect of random mutations, which is greater in recombining populations.

Contents

1 Introduction

Whereas early physics was focused on breaking things down to understand the world at its core, in recent decades the field of complex systems emerged. In complex systems interactions between particles or agents are abundant and if these are neglected, features of the system are lost, also known by the saying: 'the whole is more than the sum of its parts.' The topics of this field range from magnetization as a collective property of many particles to traffic jams as a property of many cars. Because complex systems resist simplification and are multi-layered, they remain a challenge. One can recognize the difficulty of understanding complex systems by the discrepancy that there is a good understanding of individual subatomic particles, but it is still unclear how to properly treat many diseases of complex organisms.

During my physics studies I got especially interested in the dynamics of such complex systems and in their emergent properties. In my opinion, the most intriguing complex systems are currently within the complexity of life. Therefore, in this dissertation certain aspects of evolution are studied.

In evolution, genotypes compete with each other for reproduction. Their carriers are individuals that generate offspring, and at the same time they are their blueprint. These individuals can take on any form, from viruses to bacteria to eukaryotes. Different genotypes can confer different traits to their individuals. These traits in turn can have a positive or negative effect on the reproductive ability, so that certain genotypes gain in frequency and others lose, which is called natural selection. No process is perfect and so the diversity of genotypes is maintained by copying errors during reproduction. Mutation and selection are therefore two evolutionary forces that should emerge in some natural way. However, in addition to these two forces, there is another abundant force that shapes evolution, which is the recombination of genetic material between individuals. For some reason, the blueprints of individuals have favored this additional force, so that it is widespread in nature. This dissertation therefore examines the dynamics of evolving populations and highlights the benefits of recombination.

The following introduction chapter aims to give a brief overview of the main features of this dissertation. For this purpose, a brief overview of the history of evolutionary modelling is presented, followed by an explanation of the evolutionary model used in this work, including its limitations. Afterward, the effect of recombination is discussed and historical arguments for its benefit are outlined. The introduction ends with an explanation of the further structure of the thesis.

1.1 History of modelling evolution in a nutshell

Already Aristotle (4th century BC/1991) philosophized about the diversity of species and brought his ideas to papyrus. However, he thought that all species were static and came into being by spontaneous generation (Capelle, 1955). This static view also did fit well with Christianity later on and was therefore not challenged for a long time. The crucial idea that species could in some form evolve was first formulated as part of a coherent theory by Jean-Baptiste de Lamarck (1809), who coined the term *transmutation of species* (Gould, 2002). This insight was accompanied by paleontology findings that contradicted the image of a static nature and showed that extinction events occurred in the past (Cuvier, 1796). Still, the forces that drive evolution were subject of various theories, e.g. *Lamarckism*, until Charles Darwin's theory of natural selection became well excepted (Darwin, 1859). This led to the notion of a tree of life with a common origin.
One of the first evolutionary experiments that could be statistically evaluated was performed by Gregor Mendel (1865). Through hybridization experiments with different pea plants, he discovered certain rules of heredity, which are today known as laws of Mendelian inheritance. The results showed the existence of discrete traits, which led to the notion of genes. Based on Gregor Mendel's result, Wilhelm Weinberg and G.H. Hardy later separately derived one of the first mathematical descriptions of evolution, known today as Hardy–Weinberg principle (Hardy, 1908; Weinberg, 1908). In an influential next step that linked the concept of natural selection with the notion of genes, Ronald Fisher and J.B.S. Haldane laid the foundation for today's population genetics with a quantitative description of evolution. They both developed mathematical models of evolution to understand how selection shapes the frequency distribution of genes (Fisher, 1919; Haldane, 1924). Their insights also led Haldane to make one of the first predictions about evolutionary dynamics in relation to the color of moths in Manchester, which later proved to be correct (Kettlewell, 1958; Cook & Turner, 2020). Another important figure at that time was Sewall Wright, who was interested in the interaction of genes, called epistasis in this context. He therefore introduced the concept of fitness landscapes, which is essential for this work (Wright, 1932), c.f. Fig. 1. While the number of studies in the field of population genetics exploded in the following years, the notion of fitness landscapes was for a long time not further developed since little was known about their structure. Nowadays, due to the technological advances in genetic sequencing, it becomes feasible to gather more pieces of information about the structure of these landscapes. First empirical fitness landscapes, which consider a small set of mutations and all their combinations, show that genetic interactions are indeed abundant (Weinreich et al., 2013; Szendro et al., 2013). In fitness landscapes, they manifest themselves as deviations from the assumption of independent

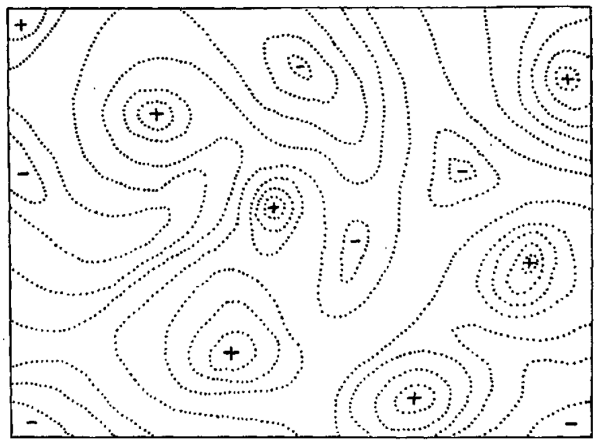

Figure 1: Illustration of a fitness landscape by Sewall Wright (1932)
Fitness landscapes map genotypes to their corresponding fitness, which usually represents their reproductive success. Sewall Wright correctly assumed that there could be many fitness peaks in the vast genotype space due to interacting genes and wondered how the population could move from a lower to a higher fitness peak.

fitness effects of mutations. Their presence makes fitness landscapes interesting in the first place, as they would be otherwise just completely smooth objects containing no significant information. These deviations from independence can take various forms. In the simplest case, only pairs of mutations interact with each other, which is called pairwise epistasis. A distinction can then be made between magnitude and sign epistasis (Weinreich et al., 2005). The former implies that two mutations strengthen or weaken each other in their effect. Sign epistasis, on the other hand, states that there is a change in the sign of the fitness effect of one of the two mutations in the presence of the other mutation. Reciprocal sign epistasis refers to a situation where each of the two mutations changes the sign of the respective other. This last type of interaction is necessary to make fitness landscapes mountainous in the sense that multiple fitness peaks exist (Poelwijk et al., 2011). Besides pairwise epistasis, results on empirical landscape show that also higher-order interactions between several sites are frequent, which can for example be characterized by a Fourier decomposition (Weinberger, 1991; Neidhart et al., 2013; Weinreich et al., 2013; Domingo et al., 2018). Ultimately, all interactions are encoded in the fitness landscape. Its structure determines which paths are likely to be taken by a population and which targets can be reached given a certain initial starting position (De Visser & Krug, 2014). Therefore, fitness landscapes are of great interest for the overarching idea to understand the dynamics of evolution and to make evolution predictable to some degree. In this work, evolution is always studied from the perspective of fitness landscapes.

1.2 Wright-Fisher type models

Besides fitness landscapes, another important ingredient of this thesis are Wright-Fisher type models. These describe variants of the model originally designed by Wright and Fisher to quantitatively describe genotype frequencies in evolving populations (Wright, 1931; Fisher, 1930). The basis of all variants is that populations evolve in discrete generations and that the population size N is kept constant. This means that all individuals in a population reproduce and die at the same time. Furthermore, each individual i with $i = 1, 2, ..., N$ carries a genotype σ_i, which is inherited by its offspring. The genotype determines the reproductive success of an individual, called (Wrightian) fitness $w(\sigma_i) = w_i$ in this context. In the Wright-Fisher type models, the fitness w_i of an individual i describes its average offspring number $w_i = \overline{n_i}$. If a Poisson distribution is assumed for the offspring number,

$$p_i(n_i) = \frac{1}{n_i!} w_i^{n_i} e^{-w_i} \tag{1}$$

a new generation is created by multinomial sampling

$$p(n_1, n_2, ..., n_N | N' = N) = \frac{N!}{n_1!\, n_2! ... n_N!} \prod_{i=1}^{N} \left(\frac{w_i}{N\overline{w}} \right)^{n_i} \tag{2}$$

with

$$N' = \sum_{i=1}^{N} n_i \quad \text{and} \quad \overline{w} = \frac{1}{N} \sum_{i=1}^{N} w_i. \tag{3}$$

Eq. 2 demonstrates that the dynamics are invariant to a multiplication of all fitness values w_i by a common factor. Therefore, if fitness values are considered to be in range $0 \leq w_i \leq 1$, as in this thesis, there is no loss of generality. Furthermore, Eq. 2 demonstrates why the process is often interpreted backwards in time, since then individuals simply select their ancestor i at random with a probability according to $w_i / N\overline{w}$.

In the most basic Wright-Fisher model, it is assumed that there are only two genotypes and that all fitness values are equal. Then the multinomial sampling simplifies to a binomial sampling

$$p(x'|x) = \binom{N}{x} \left(\frac{x}{N} \right)^{x'} \left(\frac{N-x}{N} \right)^{N-x'}, \tag{4}$$

where x represents the number of one of the two genotypes in the population and x' its number in the subsequent generations. $p(x'|x)$ also represents the entries of the transition matrix for the change in genotype numbers of this Markov process. Even though Wright-Fisher type models are generally Markov processes, an analytical analysis for more general cases is often very difficult.

While the dynamics can be considered individual-based as described above, for large populations it is more convenient and numerically efficient to consider the dynamics genotype frequency-based (Zanini & Neher, 2012). With genotype frequencies f_σ, the next generation is then computed through

$$f'_\sigma = \frac{w(\sigma)}{\overline{w}} f_\sigma \quad \text{with} \quad \overline{w} = \sum_\sigma f_\sigma w(\sigma) \tag{5}$$

where the sum is taken over all genotypes. However, in this form, the equation describes deterministic dynamics, referring to the limit $N \to \infty$. To retrieve the stochasticity of finite populations, multinomial sampling needs to be performed in each generation:

$$(n_\sigma, n_\kappa, ...) \sim \text{Multi}(N, (f'_\sigma, f'_\kappa, ...)) \tag{6}$$

Here, $n_\sigma, n_\kappa ...$ denote the integer numbers of individuals carrying a certain genotype. Now, the effect of the individual fitness on the dynamics (Eq. 5) is decoupled from the effect of the population's finiteness (Eq. 6). The latter is referred to as genetic drift in this context. It is another important force that shapes evolution and also plays a central role in some arguments for the benefits of recombination (de Visser & Elena, 2007). The influence of the population's finiteness on the dynamics is demonstrated in chapter 3.
In the Wright-Fisher type model of this dissertation, it is assumed that the fitness values can differ between genotypes. It also takes into account that individuals can acquire mutations each generation and that recombination can take place between individuals. Since the integration of these evolutionary forces will be discussed in more detail in chapters 2 & 3, this will not be explained here.
Instead, the general assumptions of the model are discussed in the following:
On the one hand, the question arises to what extent the assumption of discrete generations is justified. Regarding this issue it can be noted that there are similar models, such as the Moran model, which almost resembles a continuous description. Nevertheless, a comparison of the models shows that the dynamics do not differ except for rescaling (Blythe & McKane, 2007; Wakeley, 2009). Especially in steady states, which we mostly consider in this work, no differences are to be expected. Nevertheless, Wright-Fisher type models have the advantage that numerical simulations run much faster than their continuous counterparts, such that steady states are reached more quickly (Park et al., 2010).
Another strong assumption is certainly that there is no spatial structure between the individuals. In this sense, it is a mean-field theory. However, this assumption can be appropriate if, for example, well-mixed populations in a small volumes are considered. Especially for experiments in test tubes, this assumption should therefore be justifiable,

e.g. microbes in liquid culture.
Another assumption is that the population size is kept constant. Whether this is appropriate certainly depends on the circumstance. Especially if the population size fluctuates strongly and could reach a very small size, bottleneck effects cannot be ignored (Wein & Dagan, 2019). However, in biological experiments the population can also be kept constant by means of special set-ups, e.g. chemostats (Wick et al., 2002).

1.3 Nonlinear dynamics of recombination

One further key aspect of this work is the evolutionary mechanism of recombination. While in the deterministic system that describes infinite populations, the consideration of selection and mutation keeps the dynamics linear, the inclusion of recombination turns it nonlinear. The former two evolutionary forces are linear, since selection is simply determined by the number of offspring of an individual and mutations also arise independently in individuals. Even if the population size in the selection step is kept constant and hence the number of offspring is correlated, this constrain imposes only a normalization that does not change the dynamics with respect to the average genotype frequencies. Contrary, recombination describes a horizontal gene transfer between pairs of individuals and therefore depends quadratically on the population's genotype frequencies. Although certain dynamical systems with recombination can be linearized through a procedure called *Haldane linearization*, it is generally not the case and the procedure itself can be cumbersome (McHale & Ringwood, 1983; Dawson, 2002; Baake & Baake, 2003). The dynamics of nonlinear systems are typically not only less mathematically tractable but also less intuitive. This might be the reason why many different explanations have been proposed for the prevalence of recombination in nature. The most popular explanations that are relevant for this thesis are in the following shortly discussed, along with their limitations.

Weismann effect: The Weismann effect simply states that recombination is beneficial because it increases genetic variation within populations (Weismann, 1889). Although the benefit in this sense remains somewhat vague, it could be understood according to the proverb "don't put all your eggs in one basket", attributed to the author of Don Quixote, Miguel de Cervantes, in the early 16th century. However, it is questionable whether this is always beneficial, and there is also the saying "put all your eggs in one basket, and then watch that basket", which was coined by the US business magnate Andrew Carnegie in the 19th century.
Since proof by proverb is not well accepted in science, the Weismann effect is also interpreted more specifically in terms of an increased fitness variance as a result of the

increased genetic variation (Burt, 2000). In this context, *Fisher's fundamental theorem* specifies that the mean increase in fitness through natural selection is proportional to the fitness variance of the population (Fisher, 1930). However, more recent results have shown that the change of fitness variance through recombination depends on the underlying fitness landscape. For example, on a simple permutation-invariant fitness landscape, where epistasis can be easily defined, results show that recombination reduces fitness variance if epistasis is positive (de Visser & Elena, 2007; Kouyos et al., 2007).
Chapter 4 of this thesis goes on to show that recombination does not always increase genetic variation either.

Fisher-Muller effect: The hypothesis of the *Fisher-Muller effect* is that recombination brings together beneficial mutations that have occurred in different individuals (Fisher, 1930; Muller, 1932). This in turn should lead to a more rapid fitness increase of the population, as there is no need to wait for the beneficial mutations to occur sequentially. In other words, it is argued that beneficial mutations that have occurred in different individuals are not in competition with each other. This is also called clonal interference, which in turn can be lifted by recombination (Gerrish & Lenski, 1998). Numerical results show that this hypothesis holds on non-epistatic fitness landscapes where sexual populations can adapt twice as fast as their asexual counterparts (Park & Krug, 2013). However, on rugged landscape, where sign epistasis is prevalent, numerical results show that the benefit of recombination due to increased fitness gain is only transitory (Nowak et al., 2014). While recombination initially combines beneficial mutations, it can eventually trap the population at suboptimal fitness peaks (Park & Krug, 2011). Still, biological experiments could show signals of the the *Fisher-Muller effect* taking place, leading to faster adaptation e.g. in *Escherichia coli* (Cooper, 2007).

Muller's ratchet: *Muller's ratchet* describes the inevitable accumulation of deleterious mutations in finite populations without recombination under certain conditions (Muller, 1964; Felsenstein, 1974). In the simplest scenario, such a condition is given if all mutations have a fitness disadvantage, are non-interacting and back mutations can be neglected. Without recombination, each individual inherits all the mutations of its ancestor and potentially acquires more, so that after a parameter-dependent number of generations all individuals will likely have at least one mutation. Since back mutations are neglected this process is irreversible referring to a "click" in a ratchet. With time, the minimal number of mutations in individuals constantly increases and the so-called mutational meltdown occurs, which describes the downward spiral that eventually leads to extinction (Gabriel et al., 1993). Such a process could indeed be demonstrated in an evolutionary

experiment with bacteria (Zeyl et al., 2001). Recombination would be beneficial under these conditions because it could stop the ratchet, since descendants have the chance to inherit fewer mutations than their parents if the parental genotypes have mutations at different loci (Bell, 1988).
However, theory predicts that even without recombination, several deviations from the above mentioned strict conditions could either greatly slow down or even stop the ratchet mechanism, e.g. compensatory mutations (Wagner & Gabriel, 1990), beneficial mutations (Rouzine et al., 2008) or synergistic epistasis (Kondrashov, 1994; Jain, 2008). On the other hand, recent results have shown that the effect remains important under a broader range of conditions in spatially structured populations (Park et al., 2018).

Hill-Robertson effect: The *Fisher-Muller effect* and *Muller's ratchet* are conceptually similar in their argument. While the former only consider beneficial mutations, the latter only takes deleterious ones into account. In these scenarios recombination either speeds up the rate of adaptation or slows down the rate of maladaptation. Joe Felsenstein (1974) pointed this out in his work, and argued that they follow essentially the same mechanism. He is also the source of the term *Muller's ratchet* and the term Hill-Robertson effect. With the latter he refers to a paper by Hill and Robertson (1966), who studied the fixation probabilities of beneficial mutations in an additive two-locus model. They demonstrated that in asexual populations both beneficial mutations are interfering with each other, thereby increasing their time to fixation. Recombination lifts such interference and allows the mutations to fix more quickly. Joe Felsenstein describes as the Hill-Robertson effect all situations in finite populations in which selection for one segregating mutation interferes with that for another. The Hill-Robertson effect can arise under various conditions, as in the *Fisher-Muller effect*, in *Muller's ratchet*, or in the case of *background selection*, which is described next. In infinite populations, on the other hand, the effect cannot occur because all genotypes exist instantaneously, so segregating mutations do not compete with each other and only genotypes do.

Background selection: Deleterious mutations are ideally purged by selection quickly. However, when they occur against a background of beneficial mutations, they can hinder the fixation of beneficial mutations on the one hand and the selection against deleterious mutations on the other (Johnson & Barton, 2002). In this case, recombination could be useful, as it creates the possibility that deleterious and beneficial mutations are separated and subsequently selected against/for individually (Peck, 1994; Rice & Chippindale, 2001). Still, recombination could simultaneously break up beneficial mutations or even combine beneficial and deleterious mutations (Moradigaravand & Engelstädter,

2013). Which effect dominates is therefore again situation dependent and not clearly understood. The term *hitchhiking* is related to *background selection* and refers to a case in which a neutral or slightly deleterious mutation reaches high frequency in the population in the background of a beneficial mutation (Smith & Haigh, 1974).

Deterministic mutation hypothesis: While the former arguments consider the rate of fitness change as a potential driver for recombination, the *deterministic mutation hypothesis* considers this to be the mutation load at selection-mutation(-recombination) balance. Equal to *Muller's ratchet* it is assumed that all mutations are deleterious but with the addition that they act synergistically. Under such conditions, the mutation load would be smaller in recombining populations, even for an infinitely large population which could create a benefit (Kondrashov, 1988, 1994).

There are many more hypotheses for the potential benefit of sex but which are less relevant for this thesis. All mentioned arguments have in common, that they impose certain conditions. Yet, recombination is widespread across many lifeforms in nature and the conditions in nature can be very different. A pluralist view could be, that nevertheless one of the conditions is met more often than not, which could generate a net benefit for recombination (West et al., 1999). Another explanation could be, that a simply certain condition is just very abundant in nature. However, this is still an open question.
In this thesis a different potential benefit is explored, which has received much less attention in the literature up to now. While the above arguments all do not take into account the largely neutral nature of many mutations, this is a central point of the potential benefit presented. More specifically, when neutral mutations are abundant, another property of evolving populations is affected by recombination, namely the population's mutational robustness. In this thesis it is shown that mutational robustness strongly increases in the presence of recombination and that this mechanism is quite robust. Furthermore is demonstrated that the effect already occurs at small recombination rates, which could have driven the observed abundance of recombination.

1.4 Structure of the thesis

In chapter 2, mutational robustness is first introduced in more detail and related results in the literature are discussed. The relationship between the population's mutational robustness and the recombination rate is then studied in the context of quasispecies, i.e., infinitely large populations. In this limit, the dynamics are deterministic, which allows the analysis not only to be based on numerical simulations, but also on analytical

results. For a simple two-locus system an explicit relationship between the population's mutational robustness and the recombination rate is derived in certain limits. In addition, the recombination weight is introduced to explain the mechanism. In order to demonstrate that the effect is quite robust, numerical simulations are performed on different multi-locus model landscapes and recombination schemes. In certain limits analytical results are also presented for multi-locus models. Finally, the effect is studied for an empirical landscape. Chapter 3 deals with other aspects of recombination that arise in the context of finite populations which additionally experience genetic drift. This chapter is prefaced with a more detailed justification for the assumption of neutral model landscapes. In the result section, properties related to evolvability and genetic diversity that emerge for finite populations are highlighted. Mutational robustness is also revisited. Results are presented for finite and infinite large landscapes. In addition, different implementations for recombination in the Wright-Fisher model are discussed.
In chapter 4, the results of an experiment that compares asexual to sexual populations are examined. Since the experiment has not been published yet, the design is explained first. This is followed by a statistical analysis of the results. Since the previous chapters make clear that the structure of the fitness landscape determines the dynamics, an attempt is made to infer this from the data. This in turn leads to a better understanding of the results and furthermore shows a signal for increased mutational robustness in the recombining populations.

2 Recombination in quasispecies - 1st manuscript

Alexander Klug, Su-Chan Park and Joachim Krug.
"Recombination and mutational robustness in neutral fitness landscapes."
PLoS computational biology 15.8 (2019): e1006884.

Status: published

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Recombination and mutational robustness in neutral fitness landscapes

Alexander Klug[1], **Su-Chan Park**[2], **Joachim Krug**[1]*

1 Institute for Biological Physics, University of Cologne, Cologne, Germany, **2** Department of Physics, The Catholic University of Korea, Bucheon, Republic of Korea

* jkrug@uni-koeln.de

OPEN ACCESS

Citation: Klug A, Park S-C, Krug J (2019) Recombination and mutational robustness in neutral fitness landscapes. PLoS Comput Biol 15(8): e1006884. https://doi.org/10.1371/journal.pcbi.1006884

Editor: Joshua Payne, Eidgenossische Technische Hochschule Zurich, SWITZERLAND

Received: February 18, 2019

Accepted: July 9, 2019

Published: August 15, 2019

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: AK and JK acknowledge support by Deutsche Forschungsgemeinschaft (DFG, https://www.dfg.de/) through CRC 1310 Predictability in evolution and SPP 1590 Probabilistic structures in evolution. SCP acknowledges support by the Basic Science Research Program through the National Research Foundation of Korea (NRF, https://www.nrf.re.kr/eng/main) funded by the Ministry of Science and ICT (Grant No.

Abstract

Mutational robustness quantifies the effect of random mutations on fitness. When mutational robustness is high, most mutations do not change fitness or have only a minor effect on it. From the point of view of fitness landscapes, robust genotypes form neutral networks of almost equal fitness. Using deterministic population models it has been shown that selection favors genotypes inside such networks, which results in increased mutational robustness. Here we demonstrate that this effect is massively enhanced by recombination. Our results are based on a detailed analysis of mesa-shaped fitness landscapes, where we derive precise expressions for the dependence of the robustness on the landscape parameters for recombining and non-recombining populations. In addition, we carry out numerical simulations on different types of random holey landscapes as well as on an empirical fitness landscape. We show that the mutational robustness of a genotype generally correlates with its recombination weight, a new measure that quantifies the likelihood for the genotype to arise from recombination. We argue that the favorable effect of recombination on mutational robustness is a highly universal feature that may have played an important role in the emergence and maintenance of mechanisms of genetic exchange.

Author summary

Two long-standing and seemingly unrelated puzzles in evolutionary biology concern the ubiquity of sexual reproduction and the robustness of organisms against genetic perturbations. Using a theoretical approach based on the concept of a fitness landscape, in this article we argue that the two phenomena may in fact be closely related. In our setting the hereditary information of an organism is encoded in its genotype, which determines it to be either viable or non-viable, and robustness is defined as the fraction of mutations that maintain viability. Previous work has demonstrated that the purging of non-viable genotypes from the population by natural selection leads to a moderate increase in robustness. Here we show that genetic recombination acting in combination with selection massively enhances this effect, an observation that is largely independent of how genotypes are connected by mutations. This suggests that the increase of robustness may be a major driver underlying the evolution of sexual recombination and other forms of genetic exchange throughout the living world.

2017R1D1A1B03034878). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The reshuffling of genetic material by recombination is a ubiquitous part of the evolutionary process across the entire range of organismal complexity. Starting with viruses as the simplest evolving entities, recombination occurs largely at random during the coinfection of a cell by more than one virus strain [1]. For bacteria the mechanisms involved in recombination are already more elaborate and present themselves in the form of transformation, transduction and conjugation [2, 3]. In eukaryotic organisms, sexual reproduction is a nearly universal feature, and recombination is often a necessary condition for the creation of offspring. Although its prevalence in nature is undeniable, the evolution and maintenance of sex is surprising since compared to an asexual population, only half of a sexual population is able to bear offspring and additionally a suitable partner needs to be found [4, 5]. Whereas the resulting two-fold cost of sex applies only to organisms with differentiated sexes [6], the fact that genetic reshuffling may break up favorable genetic combinations or introduce harmful variants into the genome poses a problem also to recombining microbes that reproduce asexually [7, 8]. Since this dilemma was noticed early on in the development of evolutionary theory, many attempts have been undertaken to identify evolutionary benefits of sex and recombination based on general population genetic principles [9–19].

In this article we approach the evolutionary role of recombination from the perspective of fitness landscapes. The fitness landscape is a mapping from genotype to fitness, which encodes the epistatic interactions between mutations and provides a succinct representation of the possible evolutionary trajectories [20]. Previous computational studies addressing the effect of recombination on populations evolving in epistatic fitness landscapes have revealed a rather complex picture, where evolutionary adaptation can be impeded or facilitated depending on, e.g., the structure of the landscape, the rate of recombination or the time frame of observation [21–26].

Here we focus specifically on the possible benefit of recombination that derives from its ability to enhance the mutational robustness of the population. A living system is said to be robust if it is able to maintain its function in the presence of perturbations [27–31]. In the case of mutational robustness these perturbations are genetic, and the robustness of a genotype is quantified by the number of mutations that it can tolerate without an appreciable change in fitness. Robust genotypes that are connected by mutations therefore form plateaux in the fitness landscape that are commonly referred to as neutral networks [32–35]. Mutational robustness is known to be abundant at various levels of biological organization, but its origins are not well understood. In particular, it is not clear if mutational robustness should be viewed as an evolutionary adaptation, or rather reflects the intrinsic structural constraints of living systems.

Arguments in favor of an adaptive origin of robustness were presented by van Nimwegen *et al.* [32] and by Bornberg-Bauer and Chan [33], who showed that selection tends to concentrate populations in regions of a neutral network where robustness is higher than average. Whereas this result is widely appreciated, the role of recombination for the evolution of robustness has received much less attention. An early contribution that can be mentioned in this context is due to Boerlijst *et al.* [36], who discuss the error threshold in a viral quasi-species model with recombination and point out in a side note that "*in sequence space recombination is always inwards pointing.*" This observation was picked up by Wilke and Adami [37] in a review on the evolution of mutational robustness, where they conjecture that the enhancement of robustness by selection should be further amplified by recombination, because "*recombination alone always creates sequences that are within the boundaries of the current mutant cloud.*" At about the same time, de Visser *et al.* discussed a mechanism based on the spreading of robustness modifier alleles in recombining populations [27] (see also [38]).

In fact indications of a positive effect of recombination on robustness had been reported earlier in computational studies of the evolution of RNA secondary structure [39] and 2D lattice proteins [40] in the presence and absence of recombination. In these systems the native folding structure of a given sequence is determined by its global free energy minimum. Due to the restricted number of attainable folds, most structures are degenerate in the sense that many sequences fold into the same structure. These sequences form neutral networks in sequence space. Xia and Levitt [40] consider two scenarios, in which the evolution of the lattice proteins is dominated by mutation and by recombination, respectively. The results show that in the latter case the concentration of thermodynamically stable protein sequences is enhanced, which is qualitatively explained by the fact that recombination tends to focus the sequences near the center of their respective neutral network. Therefore most often a single mutation does not change the folding structure.

More recently, Azevedo *et al.* [41] used a model of gene regulatory networks to investigate the origin of negative epistasis, which is a requirement for the advantage of recombination according to the mutational deterministic hypothesis [13]. In this study a gene network is encoded by a matrix of interaction coefficients. It is defined to be viable if its dynamics converges to a stable expression pattern and non-viable otherwise. Thus the underlying fitness landscape is again neutral. Based on their simulation results the authors argue that recombination of interaction matrices reduces the recombinational load, which in turn leads to an increase of mutational robustness and induces negative epistasis as a byproduct. In effect, then, recombination selects for conditions that favor its own maintenance. Other studies along similar lines have been reviewed in [42]. Taken together they suggest that the positive effect of recombination on robustness may be largely independent of the precise structure of the space of genotypes or the genotype-phenotype map. Indeed, a related scenario has also been described in the context of computational evolution of linear genetic programs [43].

Finally, in a numerical study that is similar to ours in spirit, Szöllősi and Derényi considered the effect of recombination on the mutational robustness of populations evolving on different types of neutral fitness landscapes [44]. Using neutral networks that were either generated at random or based on RNA secondary structure, they found that recombination generally enhances mutational robustness by a significant amount. Moreover, they showed that this observation holds not only for infinite populations but also for finite populations, as long as these are sufficiently polymorphic.

The goal of this article is to explain these scattered observations in a systematic and quantitative way. For this purpose we begin by a detailed examination of the simplest conceivable setting consisting of a haploid two-locus model with three viable and one lethal genotype [35]. We derive explicit expressions for the robustness as a function of the rates of mutation and recombination that demonstrate the basic phenomenon and guide the exploration of more complex situations. The two-locus results are then generalized to mesa landscapes with L diallelic loci, where genotypes carrying up to k mutations are viable and of equal fitness [45–48]. Using a communal recombination scheme and previous results for multilocus mutation-selection models, we arrive at precise asymptotic results for the mutational robustness for large L and small mutation rates. Subsequently two types of random holey landscape models are considered, including a novel class of sea-cliff landscapes in which the fraction of viable genotypes depends on the distance to a reference sequence. For the isotropic percolation landscape analytic upper and lower bounds on the robustness are derived.

As a first step towards a unified explanation for the effect of recombination on mutational robustness we introduce the concept of the recombination weight, which is a measure for the likelihood of a genotype to arise from a recombination event. In analogy to the classic fitness landscape concept in the context of selection [20], the recombination weight allows one to

identify genotypes that are favored by recombination without referring to any specific evolutionary dynamics. We show that recombination weight correlates with mutational robustness for the landscape structures used in this work, thus providing a mechanistic basis for the enhancement of robustness by recombination. Finally, using an empirical fitness landscape as an example, we quantify the competition between selection and recombination as a function of recombination rate. Throughout we describe the evolutionary dynamics by a deterministic, discrete time model that will be introduced in the next section.

Models and methods

Genotype space

We consider a haploid genome with L loci and the corresponding genotype is represented by a sequence $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_L)$ of length L. The index i labels genetic loci and each locus carries an allele specified by σ_i. Here we rely on binary sequences, which means that there are only two different alleles $\sigma_i \in \{0, 1\}$. This can be either seen as a simplification in the sense that only two alleles are assumed to exist, or in the sense that the genome consisting of all zeros describes the wild type, and the 1's in the sequence display mutations for which no further distinctions are made.

The resulting genotype space is a hypercube of dimension L, where the 2^L genotypes represent vertices, and two genotypes that differ at a single locus and are mutually reachable by a point mutation are connected by an edge. A metric is introduced by the Hamming distance

$$d(\sigma, \kappa) = \sum_i (1 - \delta_{\sigma_i \kappa_i}), \tag{1}$$

which measures the number of point mutations that separate two genotypes σ and κ. Here and in the following the Kronecker symbol is defined as $\delta_{xy} = 1$ if $x = y$ and $\delta_{xy} = 0$ otherwise. The genotype $\overline{\sigma}$ at maximal distance $d(\sigma, \overline{\sigma}) = L$ from a given genotype σ is called its antipodal, and can be defined by $\overline{\sigma}_i = 1 - \sigma_i$. Finally, in order to generate a fitness landscape, a (Wrightian) fitness value w_σ is assigned to each genotype.

Dynamics

The forces that drive evolution are selection, mutation and recombination. To model the dynamics we use a deterministic, discrete-time model with non-overlapping generations, which can be viewed as an infinite population limit of the Wright-Fisher model. Demographic stochasticity or genetic drift is thus neglected. Numerical simulations of evolution on neutral networks have shown that the infinite population dynamics is already observable for moderate population sizes, which justifies this approximation [32, 44]. We will return to this point in the Discussion.

Once the frequency $f_\sigma(t)$ of a genotype σ at generation t is given, the frequency at the next generation is determined in three steps representing selection, mutation, and recombination. After the selection step, the frequency $q_\sigma(t)$ is given as

$$q_\sigma(t) = \frac{w_\sigma}{\overline{w}(t)} f_\sigma(t), \tag{2}$$

where $\overline{w} \equiv \sum_\sigma w_\sigma f_\sigma(t)$ is the mean population fitness at generation t. After the mutation step, the frequency $p_\sigma(t)$ is given as

$$p_\sigma(t) = \sum_\kappa U_{\sigma\kappa} q_\kappa(t), \tag{3}$$

where $U_{\sigma\kappa}$ is the probability that an individual with genotype κ mutates to genotype σ in one generation. Here, we assume that alleles at each locus mutate independently, and the mutation probability μ is the same in both directions ($0 \to 1$ and $1 \to 0$) and across loci. This leads to the symmetric mutation matrix

$$U_{\sigma\kappa} = (1-\mu)^{L-d(\sigma,\kappa)}\mu^{d(\sigma,\kappa)}. \tag{4}$$

In order to incorporate recombination we have to consider the probability that two parents with respective genotypes κ and τ beget a progeny with genotype σ by recombination. This is represented by the following equation:

$$f_\sigma(t+1) = \sum_{\kappa\tau} R_{\sigma|\kappa\tau} p_\kappa(t) p_\tau(t). \tag{5}$$

Descriptively speaking, two genotypes (κ and τ) are taken to recombine with a probability that is equal to their frequency in the population (after selection and mutation). The probability for the offspring genotype σ is then given by $R_{\sigma|\kappa\tau}$. These probabilities depend of course on the parent genotypes κ and τ but also on the recombination scheme. Here we consider a uniform and a one-point crossover scheme; see Fig 1 for a graphical representation. These two represent extremes in a spectrum of possible recombination schemes. Nevertheless we will show that both lead to qualitatively similar results in the regimes of interest. In the case of uniform crossover the recombination probabilities are given by

$$R_{\sigma|\kappa\tau} = \frac{r}{2^L}\left(\prod_i^L (\delta_{\sigma_i\kappa_i} + \delta_{\sigma_i\tau_i})\right) + \frac{1-r}{2}(\delta_{\sigma\kappa} + \delta_{\sigma\tau}) \tag{6}$$

and in the case of one point crossover the probabilities can be written as

$$\begin{aligned} R_{\sigma|\kappa\tau} = \; & \frac{r}{2(L-1)} \sum_{n=1}^{L-1} \left[\left(\prod_{m=1}^{n} \delta_{\sigma_m\kappa_m} \right) \left(\prod_{m=n+1}^{L} \delta_{\sigma_m\tau_m} \right) \right. \\ & \left. + \left(\prod_{m=1}^{n} \delta_{\sigma_m\tau_m} \right) \left(\prod_{m=n+1}^{L} \delta_{\sigma_m\kappa_m} \right) \right] + \frac{1-r}{2}(\delta_{\sigma\kappa} + \delta_{\sigma\tau}). \end{aligned} \tag{7}$$

In both equations a variable $r \in [0, 1]$ appears which describes the recombination rate. For $r = 0$ no recombination occurs and $f_\sigma(t+1)$ is the same as $p_\sigma(t)$. For $r = 1$ recombination is a necessary condition for the creation of offspring (obligate recombination). But also intermediate values of r can be chosen as they occur in nature, e.g., for bacteria and viruses.

In the following we are mostly interested in the equilibrium frequency distribution f_σ^* of a population, which is determined by the stationarity condition

$$f_\sigma(t+1) = f_\sigma(t) = f_\sigma^* \tag{8}$$

for all genotypes σ.

Mutational robustness

From the point of view of fitness landscapes the occurrence of mutational robustness implies that fitness values of neighboring genotypes are degenerate, giving rise to neutral networks in genotype space [29, 32–35]. In order to model this situation we use two-level landscapes that only differentiate between genotypes that are viable ($w_\sigma = 1$) or lethal ($w_\sigma = 0$). Any selective advantage between viable genotypes is assumed to be negligible. The mutational robustness of

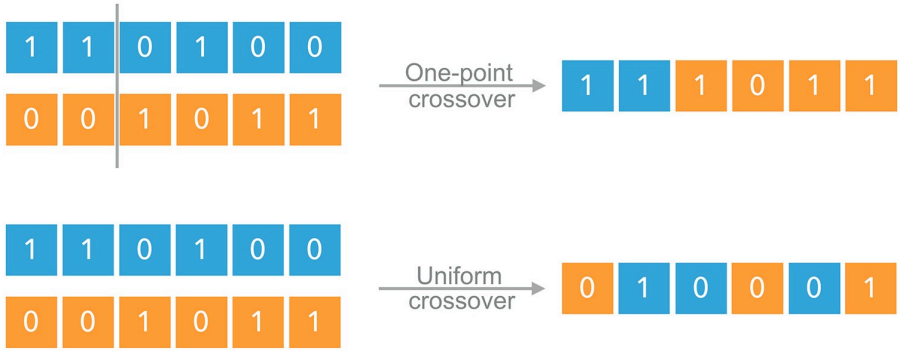

Fig 1. Recombination schemes. In the one-point crossover scheme, the parent genotypes are cut once between two randomly chosen loci and recombined to form the offspring. In the uniform crossover scheme, at each locus of the offspring, an allele present in one of the parents is chosen at random.

https://doi.org/10.1371/journal.pcbi.1006884.g001

a population can then be measured by the average fraction of viable point mutations in an individual, which depends on the population distribution in genotype space [32–34]. It increases if the population mainly adapts to genotypes for which most point mutations are viable. Therefore we define mutational robustness m as the average fraction of viable point mutations of a population,

$$m \equiv \sum_{\sigma \in V} m_\sigma f_\sigma^* \quad \text{with} \quad m_\sigma \equiv \frac{n_\sigma}{L}. \tag{9}$$

Here the sum is over the set V of all viable genotypes and n_σ is the number of viable point mutations of genotype σ. We will refer to m_σ as the mutational robustness of the genotype. The expression is normalized by the total number of loci L, since in an optimal setting the entire population has L viable genotypic neighbors and $m_\sigma = 1$ for all $\sigma \in V$. The value of m is thus constrained to be in the range $[0, 1]$. We weight the genotypes by their stationary frequencies f_σ^*, since we want to determine the mutational robustness of populations that are in equilibrium with their environment.

Recombination weight

In order to elucidate the interplay of recombination and mutational robustness it will prove helpful to introduce a representation of how recombination can transfer genotypes into each other. The number of distinct genotypes that two recombining genotypes are able to create depends on their Hamming distance. In particular, the recombination of two identical genotypes does not create any novelty, whereas a genotype and its antipodal are able to generate all possible genotypes through uniform crossover.

Here we introduce a measure which expresses how many pairs of viable genotypes are able to recombine to a specific genotype. It is complementary to the mutational robustness, in the sense that instead of counting the viable mutation neighbors of a genotype, the size of its recombinational neighborhood of viable recombination pairs is determined. The recombinational neighborhood depends on the recombination scheme and the distribution of viable genotypes in the genotype space. For a given recombination scheme the probability for a genotype σ to be the outcome of recombination of two genotypes κ, τ is given by the recombination

tensor $R_{\sigma|\kappa\tau}$. The *recombination weight* λ_σ is therefore obtained by summing the recombination tensor over all ordered pairs of viable genotypes,

$$\lambda_\sigma = \frac{1}{2^L} \sum_{\kappa \in V, \tau \in V} R_{\sigma|\kappa\tau}. \tag{10}$$

It can be seen from (5) that $\lambda_\sigma = 1$ when all genotypes are viable, and hence the normalization by 2^L ensures that the recombination weight lies in the range [0, 1]. Under this normalization, the recombination weights sum to $\sum_\sigma \lambda_\sigma = |V|^2/2^L$, where $|V|$ stands for the number of viable genotypes. In the following the genotype maximizing λ_σ will be referred to as the *recombination center* of the landscape.

Since neutral landscapes only differentiate between viable (unit fitness) and lethal (zero fitness) genotypes, the recombination weight (10) can alternatively be written as a sum over all ordered pairs of genotypes whereby the recombination tensor is multiplied by the pair's respective fitness,

$$\lambda_\sigma = \frac{1}{2^L} \sum_{\kappa,\tau} R_{\sigma|\kappa\tau} w_\kappa w_\tau. \tag{11}$$

In this way the concept naturally generalizes to arbitrary fitness landscapes. In the absence of recombination ($r = 0$) the recombination weight (11) of a genotype is simply proportional to its fitness, $\lambda_\sigma = \tilde{w} w_\sigma$, where $\tilde{w} = 2^{-L} \sum_\sigma w_\sigma$ is the unweighted average fitness. Within our recombination schemes, the recombination tensor depends linearly on r and, by definition, so does the recombination weight. Accordingly, for general r the recombination weight interpolates linearly between the limiting values at $r = 0$ and $r = 1$. Since λ_σ for $r = 0$ is known, the remaining task will be to find λ_σ for $r = 1$.

Results

In the following sections we investigate how mutational robustness depends on the mutation and recombination rates. In order to test the generality of our results, we use, besides contrasting recombination schemes, also different neutral landscape models such as the mesa [45–48] and the percolation models [35, 49]. Additionally we introduce a more general landscape named sea-cliff model, which combines elements of both the landscape models and contains them as limiting cases. In the end, we discuss mutational robustness and its relation with recombination weight for an empirical landscape.

Two-locus models are commonly used in population genetics to gain a foothold in understanding evolutionary scenarios involving multiple recombining loci [35, 38, 50–57]. Following this tradition, we first discuss a two-locus model and then extend our results to multi-locus models.

Two-locus model

The simplest fitness landscape to study the mutational robustness of a population would be the haploid two-locus model in which all but one genotype are viable [35]; see Fig 2 for a graphical representation of the model. In this setting the population gains mutational robustness if the frequency of the genotype (0,0) for which both point mutations are viable increases relative to the genotypes (0,1) and (1,0). This model has been analyzed previously using a unidirectional mutation scheme where reversions $1 \to 0$ are suppressed [58, 59]. As a consequence, selection cannot contribute to mutational robustness because the genotype (0,0) goes extinct in the

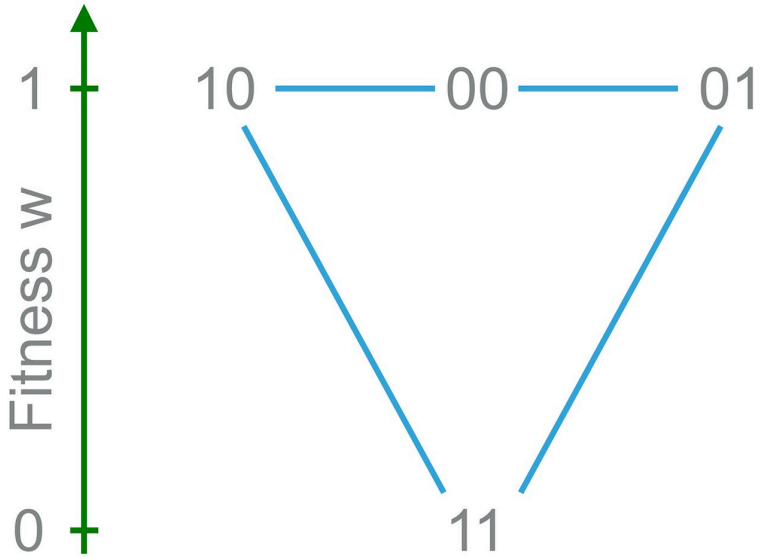

Fig 2. Two-locus model. Genotype (1,1) is lethal while the other three genotypes are viable with the same fitness. Here, genotype (0,0) is most robust since both its single mutants are viable.

https://doi.org/10.1371/journal.pcbi.1006884.g002

absence of recombination. Here we consider the case of bidirectional, symmetric mutations in which both selection and recombination contribute to robustness. A comparison of the two mutation schemes is provided in S1 Appendix.

We proceed to solve the equilibrium condition Eq (8). Since the equilibrium genotype frequencies f^*_{01} and f^*_{10} are the same due to the symmetry of the landscape and the mutation scheme, the recombination step at stationarity reads

$$\begin{aligned} f^*_{00} &= p_{00} - \rho(p_{00}p_{11} - p_{10}p_{01}) \quad \Leftrightarrow \quad f_0 = p_0 - \rho D, \\ f^*_{10/01} &= p_{10/01} + \rho(p_{00}p_{11} - p_{10}p_{01}) \quad \Leftrightarrow \quad f_1 = p_1 + 2\rho D, \\ f^*_{11} &= p_{11} - \rho(p_{00}p_{11} - p_{10}p_{01}) \quad \Leftrightarrow \quad f_2 = p_2 - \rho D, \end{aligned} \tag{12}$$

where p_σ is the (equilibrium) frequency of genotype σ after the mutation step, f_i and p_i are the corresponding lumped frequencies [60] of all genotypes with i 1's, and $D \equiv p_{00}p_{11} - p_{10}p_{01} = p_0p_2 - p_1^2/4$ is the linkage disequilibrium after the mutation step. Notice that the one-point and uniform crossover schemes give the same equation form except that the parameter ρ is given by $\rho = r$ in the case of one-point crossover and $\rho = r/2$ for uniform crossover. However, we would like to emphasize that this is a mere coincidence of the two-locus model which disappears as soon as L is larger than 2.

The lumped frequencies q_i of all genotypes with i 1's after the selection step are given by

$$q_0 = \frac{f_0}{1 - f_2}, \quad q_1 = \frac{f_1}{1 - f_2}, \quad q_2 = 0. \tag{13}$$

Applying the mutation step we obtain

$$\begin{aligned} p_0 &= q_0(1-\mu)^2 + \mu(1-\mu)q_1 = \mu(1-\mu) + (1-\mu)(1-2\mu)q_0, \\ p_1 &= q_1[(1-\mu)^2 + \mu^2] + 2\mu(1-\mu)q_0 = 1 - 2\mu + 2\mu^2 - (1-2\mu)^2 q_0, \\ p_2 &= \mu(1-\mu)q_1 + \mu^2 q_0 = \mu(1-\mu) - \mu(1-2\mu)q_0, \\ D &= p_0 p_2 - p_1^2/4 = -\frac{1}{4}(1-2\mu)^2(1-q_0)^2, \end{aligned} \tag{14}$$

where we have used the normalization $q_0 + q_1 = 1$ to express the right hand sides in terms of q_0. Putting everything together, the problem is reduced to solving the following third order polynomial equation for q_0,

$$\begin{aligned} 0 &= q_0(1-f_2) - f_0 = q_0(1 - p_2 + \rho D) - p_0 + \rho D \\ &= \frac{\rho}{4}(1-2\mu)^2(1-q_0)^2(1+q_0) + \mu\left[1 - 2q_0 - q_0^2 - \mu(1-2q_0)(1+q_0)\right], \end{aligned} \tag{15}$$

from which we can in principle find exact analytic expressions for f_σ^*. However, it is difficult to extract useful information from the exact solution. In the following we will therefore provide approximate solutions.

If we neglect recombination ($\rho = 0$), we obtain the following equilibrium genotype frequency distribution:

$$\begin{aligned} f_{00}^*(\rho=0) &= \frac{1-\mu}{2}\sqrt{8 - 16\mu + 9\mu^2} - \frac{1}{2}(2 - 5\mu + 3\mu^2) \\ &\approx (\sqrt{2} - 1) + \left(\frac{5}{2} - 2\sqrt{2}\right)\mu + O(\mu^2), \\ f_{01/10}^*(\rho=0) &= \frac{1}{4}(4 - 9\mu + 6\mu^2) - \frac{1-2\mu}{4}\sqrt{8 - 16\mu + 9\mu^2} \\ &\approx \left(1 - \frac{1}{\sqrt{2}}\right) + \left(\frac{3}{\sqrt{2}} - \frac{9}{4}\right)\mu + O(\mu^2), \\ f_{11}^*(\rho=0) &= \frac{\mu}{2}\left(4 - 3\mu - \sqrt{8 - 16\mu + 9\mu^2}\right) \approx \left(2 - \sqrt{2}\right)\mu + O(\mu^2). \end{aligned} \tag{16}$$

When $\rho = 1$, which corresponds to the one-point crossover scheme with $r = 1$, linkage equilibrium ($f_{00}f_{11} = f_{10}f_{01}$) is restored after one generation [55]. Accordingly, we can treat each locus independently and get rather simple expressions for f_σ^* as

$$\begin{aligned} f_{00}^*(\rho=1) &= \frac{1}{4}(2 + \mu - \sqrt{\mu^2 + 4\mu})^2 \approx 1 - 2\sqrt{\mu} + 2\mu + O(\mu^{3/2}), \\ f_{01/10}^*(\rho=1) &= \frac{1}{4}\left(2 + \mu - \sqrt{\mu^2 + 4\mu}\right)\left(\sqrt{\mu^2 + 4\mu} - \mu\right) \approx \sqrt{\mu} - \frac{3\mu}{2} + O(\mu^{3/2}), \\ f_{11}^*(\rho=1) &= \frac{1}{4}(\sqrt{\mu^2 + 4\mu} - \mu)^2 \approx \mu - O(\mu^{3/2}). \end{aligned} \tag{17}$$

We depict the equilibrium solutions for the above two cases in Fig 3.

Now, the mutational robustness

$$m = \frac{1}{2}(2f_{00}^* + f_{10}^* + f_{01}^*) = f_{00}^* + f_{10}^* = f_0 + \frac{1}{2}f_1 \tag{18}$$

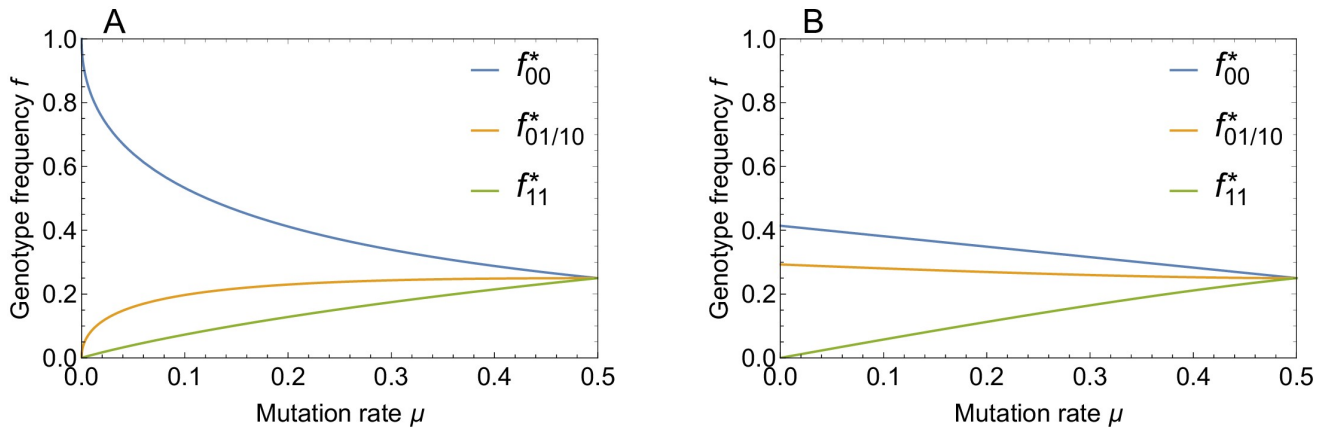

Fig 3. Equilibrium genotype frequencies in the two locus model. Genotype frequencies in the stationary state are shown as a function of mutation rate for (A) strong recombination ($\rho = 1$) and (B) no recombination ($\rho = 0$).

https://doi.org/10.1371/journal.pcbi.1006884.g003

for the above two cases is obtained as

$$m(\mu, \rho = 0) = \frac{1}{4}(\mu + \sqrt{8 - 16\mu + 9\mu^2}) \approx \frac{1}{\sqrt{2}} - \left(\frac{1}{\sqrt{2}} - \frac{1}{4}\right)\mu + O(\mu^2), \tag{19}$$

$$m(\mu, \rho = 1) = \frac{1}{2}(2 + \mu - \sqrt{\mu^2 + 4\mu}) \approx 1 - \sqrt{\mu} + \frac{\mu}{2} + O(\mu^{3/2}), \tag{20}$$

which is depicted in Fig 4. These results encapsulate in a simple form the main topic of this paper. Selection alone ($\rho = 0$) leads to a moderate increase of robustness from the baseline value $m = \frac{1}{2}$ corresponding to a random distribution over genotypes, which is attained at

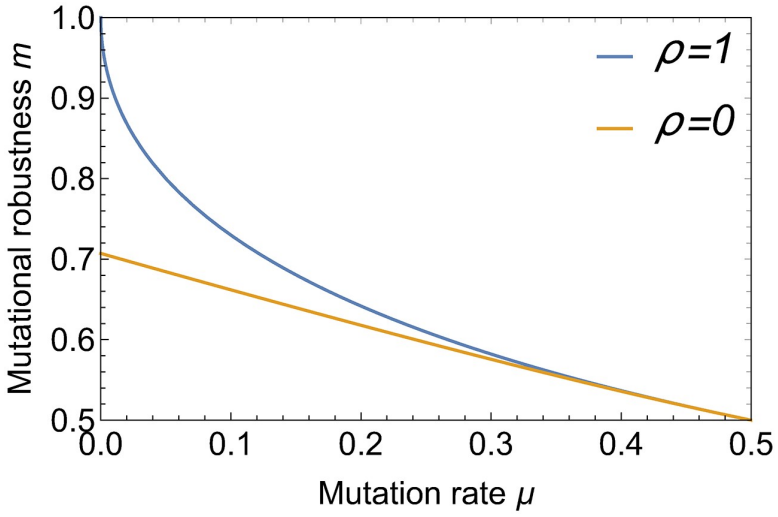

Fig 4. Mutational robustness as a function of mutation rate. The figure shows the robustness in the two-locus model at $\rho = 0$ and $\rho = 1$. Recombination leads to a massive enhancement of robustness for small mutation rates.

https://doi.org/10.1371/journal.pcbi.1006884.g004

$\mu = \frac{1}{2}$, to $m = \frac{1}{\sqrt{2}}$ for $\mu \to 0$. In contrast, for recombining populations ($\rho = 1$) robustness is massively enhanced at small mutation rates due to the strong frequency increase of the most robust genotype (0,0) and reaches the maximal value $m = 1$ at $\mu = 0$. The underlying mechanism is analogous to Kondrashov's deterministic mutation hypothesis, which posits that recombination makes selection against deleterious mutations more effective when these interact synergistically [13]. In the present case recombination increases the frequency of the double mutant genotype (1, 1), which is subsequently purged by selection, and thereby effectively drives the frequency of the allele 1 at both loci to zero. The enhancement of the frequency of the genotype (0,0) by recombination is also reflected in the recombination weights, which take on the values

$$\lambda_{00} = \frac{3}{4} + \frac{\rho}{4}, \quad \lambda_{01} = \lambda_{10} = \frac{3}{4} - \frac{\rho}{4}, \quad \lambda_{11} = \frac{\rho}{4}. \tag{21}$$

Thus the genotype (0,0) is the recombination center of the two-locus landscape.

Next we investigate how mutational robustness varies with μ for intermediate recombination rates, assuming that μ is small. As can be seen from Eq (15), the asymptotic behavior of the solution for small ρ and μ depends on which of the two parameters is smaller. We first consider the case $\rho \ll \mu \ll 1$. Defining $l = \rho/(4\mu) \ll 1$, Eq (15) is approximated by

$$0 = l(1 - q_0)^2(1 + q_0) + 1 - 2q_0 - q_0^2 - \mu(1 - 2q_0)(1 + q_0), \tag{22}$$

where we kept terms up to $O(\mu)$, since we have not determined whether l is smaller than μ or not. Since $q_0 = \sqrt{2} - 1$ is the solution of Eq (22) for $l = \mu = 0$, we set $q_0 = \sqrt{2} - 1 + al + b\mu$ and solve the equation to leading order, which gives

$$q_0 \approx \sqrt{2} - 1 + (3 - 2\sqrt{2})l - \left(\frac{3}{2} - \sqrt{2}\right)\mu. \tag{23}$$

The mutational robustness then follows as

$$m = f_0 + \frac{f_1}{2} = \frac{1}{2} + \frac{p_0 - p_2}{2} = \frac{1}{2} + (1 - 2\mu)\frac{q_0}{2} \approx \frac{1}{\sqrt{2}} + \frac{3 - 2\sqrt{2}}{2}l - \left(\frac{1}{\sqrt{2}} - \frac{1}{4}\right)\mu, \tag{24}$$

which is consistent with our previous result for $\rho = 0$; see Eq (19). We note that in this regime it is sufficient for the recombination rate to be of order $O(\mu^2)$ to compensate the negative effect of mutations on mutational robustness, as the two effects cancel when $\rho = \rho_c$ with

$$\rho_c = 2(5 + 4\sqrt{2})\mu^2 \approx 21.3 \times \mu^2. \tag{25}$$

In the regime $\rho \gg \mu$, Eq (15) is approximated as

$$(1 - 4\mu)(1 - q_0)^2(1 + q_0) + s(1 - 2q_0 - q_0^2) = 0, \tag{26}$$

with $s = 4\mu/\rho$. Again we have kept terms up to $O(\mu)$ because μ and s are of the same order if $\rho = O(1)$. Since the solution of Eq (26) for $\mu = s = 0$ is $q_0 = 1$, we set $q_0 = 1 - \alpha$ with $\alpha \ll 1$. Inserting this into Eq (26), we get $\alpha \approx \sqrt{s}$. Since $\alpha \gg \mu$, $q_0 = 1 - \sqrt{s}$ is the approximate solution to leading order. Hence

$$m = \frac{1}{2} + (1 - 2\mu)\frac{q_0}{2} \approx 1 - \sqrt{\frac{s}{4}} = 1 - \sqrt{\frac{\mu}{\rho}}, \tag{27}$$

which is again consistent with our previous result for $\rho = 1$ in Eq (20). The square root

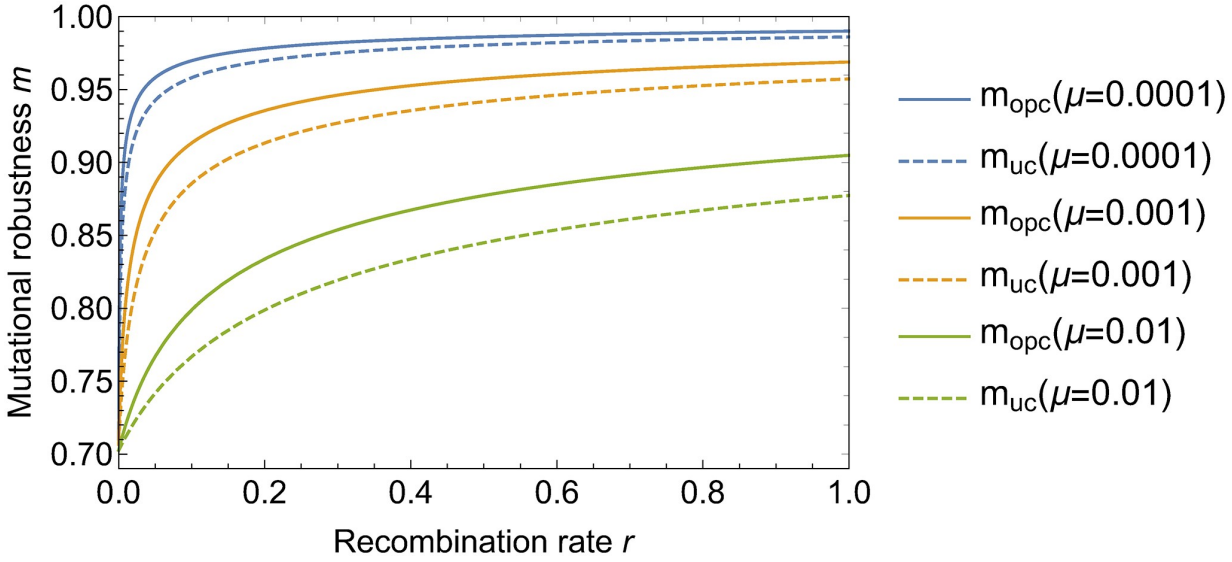

Fig 5. Mutational robustness as a function of recombination rate. The figure shows the mutational robustness for one-point crossover (m_{opc}) and uniform crossover (m_{uc}) and three different values of the mutation rate μ. When mutations are rare, a small amount of recombination is sufficient to significantly increase mutational robustness.

https://doi.org/10.1371/journal.pcbi.1006884.g005

dependence on μ/ρ derives from the corresponding behavior of the genotype frequency f^*_{00} and has been noticed previously in the model with unidirectional mutations [58, 59].

For arbitrary ρ and μ, we have to use the full Eq (15). Fig 5 illustrates the behaviour of mutational robustness as a function of the recombination rate for different mutation rates and both recombination schemes. For small μ, a low rate of recombination suffices to bring the robustness close to its maximal value $m = 1$. More precisely, according to Eq (27), a robustness $m > 1 - \epsilon$ is reached for recombination rates $\rho > \mu/\epsilon^2$.

To summarize, we have seen that analytic results for the two-locus model are easily attainable. For multi-locus models it is much more challenging to derive analytical results, particularly in the presence of recombination. By way of contrast the dynamics induced only by mutation and selection are easier to understand: While mutations increase the genotype diversity in the population, fitter ones grow in frequency through selection, which reduces diversity. Although one might expect that recombination would increase diversity, a number of studies have shown that recombination is more likely to impede the divergence of populations. Recombining populations tend to cluster on single genotypes or in a limited region of a genotype space and furthermore the waiting times for peak shifts in multipeaked fitness landscapes diverge at a critical recombination rate [22, 26, 54–56, 61]. The results for the two-locus model presented above are consistent with this behaviour, as the genotype heterogeneity of the population decreases with increasing recombination rate (S1 Fig).

In the following we will investigate how the focusing effect of recombination enhances the mutational robustness of the population in three different multi-locus models.

Mesa landscape

In the mesa landscape it is assumed that up to a certain number k of mutations all genotypes are functional and have unit fitness, whereas genotypes with more than k mutations are lethal

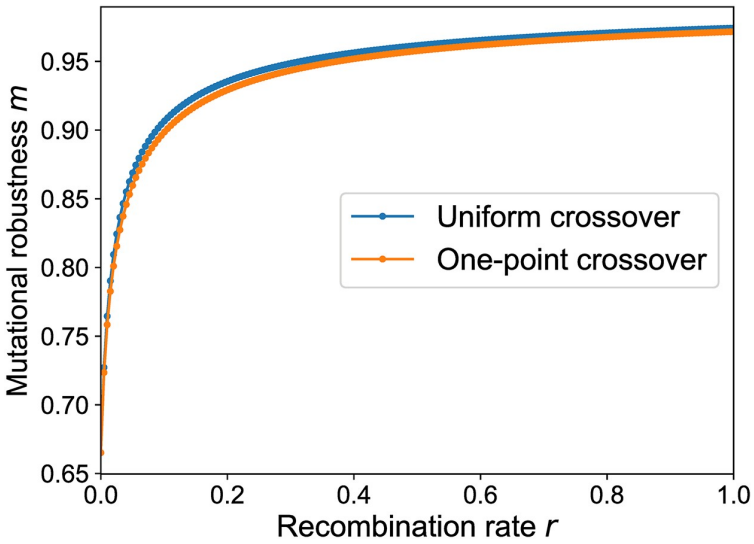

Fig 6. Mutational robustness in a mesa landscape as a function of recombination rate. Data points are obtained by numerically iterating the selection-mutation-recombination dynamics until the equilibrium state is reached. The parameters of the mesa landscape are $L = 6$, $k = 2$ and the mutation rate is $\mu = 0.001$.

https://doi.org/10.1371/journal.pcbi.1006884.g006

and have fitness zero [48]. Hence the fitness landscape is defined as

$$w_\sigma = \begin{cases} 1, & \text{if } d_\sigma \leq k, \\ 0, & \text{otherwise,} \end{cases} \tag{28}$$

where d_σ is the Hamming distance to the wild-type sequence (0, 0, . . ., 0) or, equivalently, the number of loci with allele 1. We will refer to k as the mesa width or as the critical Hamming distance.

Such a scenario can for example be observed in the evolution of regulatory motifs, where the fitness depends on the binding affinity of the regulatory proteins and d_σ corresponds to the number of mismatches compared to the original binding motif [45, 47]. The two-locus model discussed in the preceding section corresponds to the mesa landscape with critical Hamming distance $k = 1$ and sequence length $L = 2$. Here we ask to what extent the behavior observed for the two-locus model generalizes to longer sequences and variable k. Numerical simulations suggest that the strong increase of mutational robustness with recombination rate indeed persists in the general setting, and the particular recombination scheme seems to have only a minor influence; see Fig 6.

Whereas an analytical treatment for general L, k and intermediate recombination rates appears to be out of reach, accurate approximations are available in the limiting case of strong recombination or of no recombination, assuming mutation rate is small. The full derivations for both cases can be found in S1 Appendix. In the following we summarize the main results.

Strong recombination. In the limit of strong recombination we demand linkage equilibrium after each recombination step. This is satisfied if we use the so-called communal recombination scheme [62]. In this scheme an individual is not the offspring of a pair of parents. Rather, its genotype is aggregated by choosing the allele at each locus from a randomly selected parent. Hence the probability of occurrence of an allele at each locus in the offspring genotype after recombination is given by the corresponding allele frequency of the whole population,

which is precisely the definition of linkage equilibrium. In order to obtain an approximation for the mutational robustness we further assume that the mutation rate μ is small, which in turn implies a low frequency of mutant alleles. Following the derivation in S1 Appendix this leads us to the expression

$$m_{\mathrm{cr}} \approx 1 - \binom{L-1}{k}^{1/(k+1)} \mu^{k/(k+1)} + \frac{k}{k+1}\mu, \tag{29}$$

which can be approximated as

$$m_{\mathrm{cr}} \approx 1 - U^{k/(k+1)}(k!)^{-1/(k+1)} + \frac{k}{k+1}\mu \tag{30}$$

for $L \gg k$, where $U = L\mu$ is the genome-wide mutation rate and the subscript signifies the communal recombination scheme. Using Eq (29) and setting $L = 2$ and $k = 1$ we retrieve the result (20) for the two-locus model. Furthermore comparing Eqs (29) and (30) to numerical simulations of communal recombination illustrates their validity for large L (S2 Fig). If we use uniform crossover and one-point crossover instead of communal recombination, the numerical simulations suggest that the leading behaviour of $1 - m$ is still a function of $U = L\mu$ with the same exponent $k/(k + 1)$, which supports the universality of our findings with respect to the recombination scheme; see S3 Fig.

No recombination. In order to obtain analytical results in the absence of recombination we assume that the mutation rate is small enough that only a single point mutation occurs in one generation. This condition is fulfilled if $U = L\mu \ll 1$. Interestingly, we observe that in this regime the equilibrium frequencies after selection are independent of U. Therefore also the mutational robustness after selection, denoted by M_{nr}, is independent of U. The relation between mutational robustness after selection (M_{nr}) and after mutation (m_{nr}) is given by

$$m_{\mathrm{nr}} = M_{\mathrm{nr}}(1 - U) + M_{\mathrm{nr}}^2 U, \tag{31}$$

which makes it suffice to find M_{nr}.

Assuming $k/L \ll 1$ it is possible to link the set of stationarity conditions to the Hermite polynomials $H_n(x)$. This yields an approximation for the mutational robustness after selection as

$$M_{\mathrm{nr}} = \sqrt{\frac{y_k}{L}} + o(L^{-1/2}), \tag{32}$$

where $\sqrt{y_k/2}$ is the largest zero of $H_{k+1}(x)$. Correspondingly, the mutational robustness after mutation is

$$m_{\mathrm{nr}} = \sqrt{\frac{y_k}{L}}(1 - U) + \frac{y_k}{L}U. \tag{33}$$

A comparison to the exact solutions for M_{nr}, which have been obtained up to $k = 4$, confirms this approximation. If we further assume that $1 \ll k \ll L$, we find $y_k \sim 4k$, which leads to

$$m_{\mathrm{nr}} = 2\sqrt{\frac{k}{L}}(1 - U) + 4\frac{k}{L}U. \tag{34}$$

Results for the joint limit $k, L \to \infty$ at fixed ratio $x = k/L$ can be obtained from the analysis of Ref. [48], which yields

$$M_{\rm nr} = \begin{cases} 2\sqrt{x(1-x)}, & \text{if } x < 1/2, \\ 1, & \text{if } x \geq 1/2 \end{cases} \tag{35}$$

and therefore

$$m_{\rm nr} = \begin{cases} 2\sqrt{x(1-x)}(1-U) + 4x(1-x)U, & \text{if } x < 1/2, \\ 1, & \text{if } x \geq 1/2. \end{cases} \tag{36}$$

The leading behaviour for small x coincides with Eq (34). A comparison of the approximations to numerical solutions is given in S4 Fig.

Comparison of the two cases. It is instructive to compare the results obtained above to the mutational robustness m_0 of a uniform population distribution. For the latter we assume that all viable genotypes have the same frequency and all lethal genotypes have frequency zero. For the mesa model this yields

$$m_0(L,k) = \frac{1}{\sum_{i=0}^{k}\binom{L}{i}}\left[\binom{L}{k}\frac{k}{L} + \sum_{i=0}^{k-1}\binom{L}{i}\right] \approx \min[2k/L, 1], \tag{37}$$

where the last approximation is valid for $L \to \infty$. In S5 Fig the behavior of m_0, $m_{\rm nr}$ and $m_{\rm cr}$ is depicted as a function of various model parameters. Similar to the results obtained for the two-locus model, we see that selection gives rise to a moderate increase of robustness (from $2k/L$ to $2\sqrt{k/L}$ for $1 \ll k \ll L$), but recombination has a much stronger effect and leads to values close to the maximal robustness $m = 1$ for a broad range of conditions.

To elucidate the underlying mechanism, it is helpful to consider the shape of the equilibrium frequency distributions in genotype space (Fig 7). The combinatorial increase of the number of genotypes with increasing d_σ generates a strong entropic force that selection alone cannot efficiently counteract. As a consequence, the non-recombining population distribution is localized near the brink of the mesa at $d_\sigma = k$ [48]. In contrast, the contracting property of recombination [44] allows it to localize the population in the interior of the fitness plateau where most genotypes are surrounded by viable mutants.

S6 Fig shows the corresponding recombination weight profile. Similar to the genotype frequencies in Fig 7(B) the recombination weight decays rapidly with increasing Hamming distance for $r > 0$, but the decay appears to be faster than exponential. Interestingly, at $d = k$ the recombination weight decreases with increasing r [see also Eq (21)]. The method used to compute λ_σ for large mesa landscapes is explained in S1 Appendix.

Percolation landscapes

In the percolation landscape genotypes are randomly chosen to be viable ($w_\sigma = 1$) with probability p and lethal ($w_\sigma = 0$) with probability $1 - p$. An interesting property of the percolation model is the emergence of two different landscape regimes [49, 63–65]. Above the percolation threshold p_c, viable genotypes connected by single mutational steps form a cluster that extends over the whole landscape, whereas below p_c only isolated small clusters appear. Since the percolation threshold depends inversely on the sequence length, $p_c \approx \frac{1}{L}$, for large L a small fraction of viable genotypes suffices to create large neutral networks. This allows a population to evolve to distant genotypes without going through lethal regions, and correspondingly the percolation model is often used to study speciation [35, 49]. A network representation of the

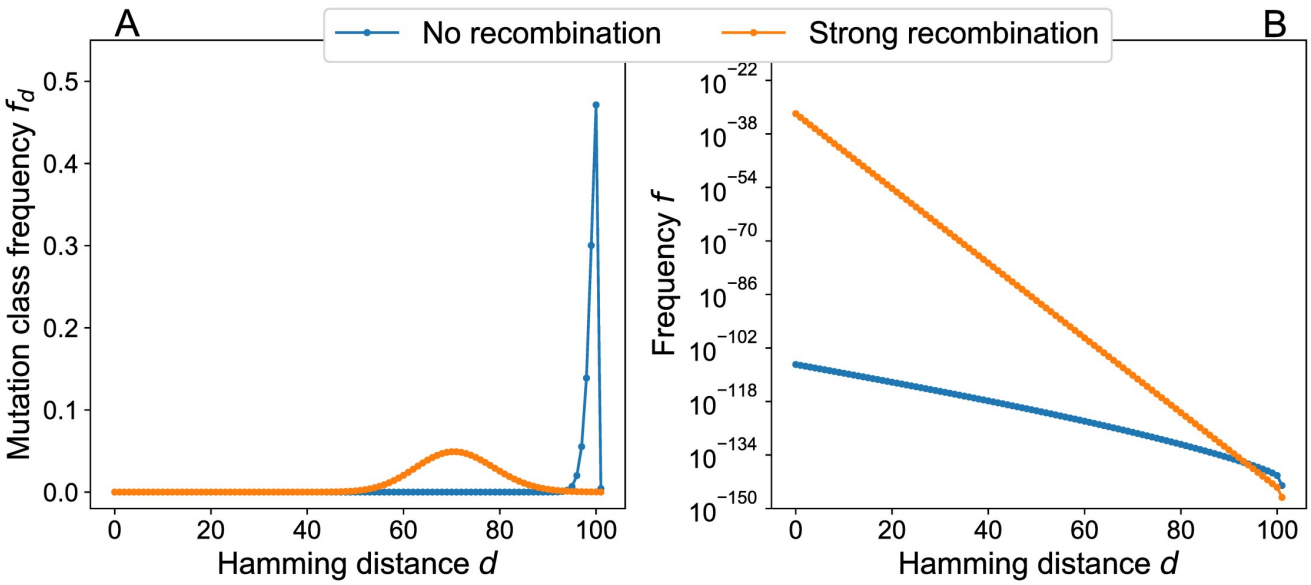

Fig 7. Equilibrium genotype distributions in a mesa landscape for strongly and non-recombining populations. Stationary states for populations with communal recombination and no recombination have been computed by assuming that only single point mutations occur with $U = 0.01$. Landscape parameters are $L = 1000$ and $k = 100$. The resulting mutational robustness is $m_{nr} \approx 0.572$ for the non-recombining population and $m_{cr} \approx 1.000$ for communal recombination. (A) Lumped mutation class frequencies on linear scales. In the absence of recombination the majority of the population is located at the critical Hamming distance $d = k$, whereas in the case of strong recombination the distribution is broader and shifted away from the brink of the mesa. (B) Genotype frequencies on semi-logarithmic scales. In both cases the genotype frequencies decrease exponentially with the Hamming distance to the wild type, but the distribution has much more weight at small distances in the case of recombination.

https://doi.org/10.1371/journal.pcbi.1006884.g007

percolation model is shown in Fig 8. The algorithm used to generate this visual representation is explained in S1 Appendix.

Fig 9 shows three exemplary stationary genotype frequency distributions on the landscape depicted in Fig 8. In the absence of recombination the equilibrium frequency distribution is unique, but in the presence of recombination the non-linearity of the dynamics implies that multiple stationary states may exist [54, 55, 61]. Fig 9 displays two stationary distributions for $r = 1$ which are accessed from different initial conditions. It is visually apparent that the recombining populations are concentrated on a small number of highly connected genotypes, leading to a significant increase of mutational robustness.

To quantify this effect, the average mutational robustness $\overline{m}$ is calculated as a function of the recombination rate according to the following numerical protocol:

- A percolation landscape for given L and p is generated and the initial population is distributed uniformly among all genotypes.
- The population is evolved in the absence of recombination ($r = 0$) until the unique equilibrium frequency distribution is reached, for which the mutational robustness m is calculated.
- Next the recombination rate is increased by predefined increments. After increasing r, the population is again evolved using the stationary state obtained before the increment of r as the initial condition, until it reaches a new stationary state for which the mutational robustness is measured.
- When the recombination rate has reached $r = 1$, a new percolation landscape is generated and the process starts all over again. This is done for an adjustable number of runs over which the average is taken.

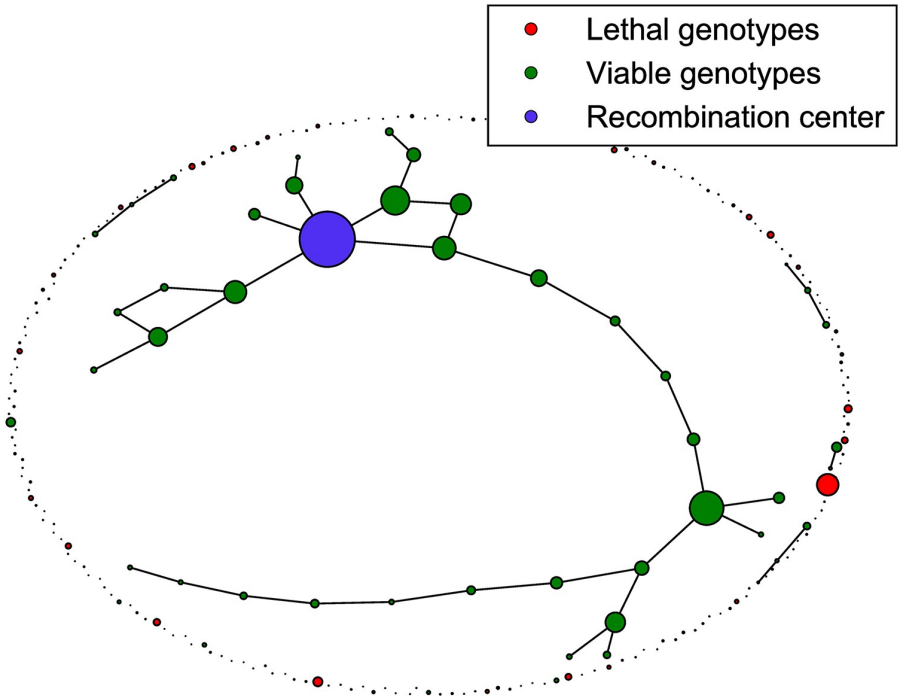

Fig 8. Network representation of a percolation landscape. The figure shows a percolation landscape with $L = 8$ loci and a fraction $p = 0.2$ of viable genotypes. Viable genotypes at Hamming distance $d = 1$ are connected by edges, and the node area of a genotype σ is proportional to λ_σ^6, where the recombination weight λ_σ is defined in Eq (10). The recombination center is the genotype with the largest recombination weight.

https://doi.org/10.1371/journal.pcbi.1006884.g008

The results of such a computation are shown in Fig 10. Similar to the mesa landscapes, a strong increase of mutational robustness is observed already for small rates of recombination, and the effect is largely independent of the recombination scheme. However, in contrast to the mesa landscape the robustness does not reach its maximal value $m = 1$ for $r = 1$ and small μ. This reflects the fact that maximally connected genotypes with $m_\sigma = 1$ are very rare at this particular value of p.

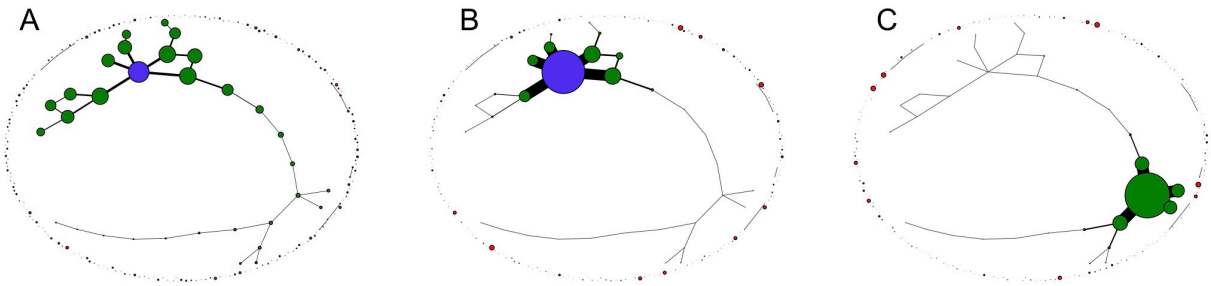

Fig 9. Stationary states in a percolation landscape. The figure shows three different stationary population distributions in the percolation landscape depicted in Fig 8. Node areas are proportional to the stationary frequency of the respective genotype in the population, and the edge width $e_{\sigma,\tau}$ between neighboring genotypes is proportional to the frequency of the more populated one, $e_{\sigma,\tau} \propto \max[f_\sigma^*, f_\tau^*]$. (A) Unique stationary state of a non-recombining population. (B,C) Stationary states for recombining populations undergoing uniform crossover with $r = 1$. The recombination center (purple) is the most populated genotype in (A,B), but not in (C). In all cases the mutation rate is $\mu = 0.01$.

https://doi.org/10.1371/journal.pcbi.1006884.g009

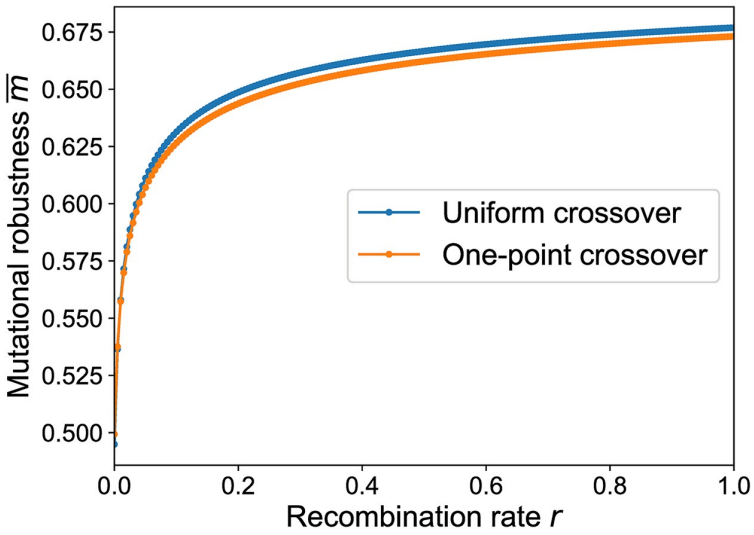

Fig 10. Average mutational robustness in the percolation landscape as a function of recombination rate. Mutational robustness is computed for 250 randomly generated percolation landscapes with $L = 6$ and $p = 0.4$, and the results are averaged to obtain $\overline{m}(r)$. The mutation rate is $\mu = 0.001$.

https://doi.org/10.1371/journal.pcbi.1006884.g010

For the purpose of comparison we also determined the average mutational robustness $\overline{m}_0$ of a uniform population distribution for the percolation model. Conditioned on the number $\nu = |V|$ of viable genotypes and assuming that $\nu \geq 1$, we have $m_0(\nu, L) = n(\nu, L)/L$, where $n(\nu, L)$ is the average number of viable neighbors of a viable genotype. The latter is given by the expression

$$n(L, \nu) = \frac{(\nu - 1)L}{2^L - 1}, \tag{38}$$

since for a given viable genotype there are $\nu - 1$ remaining genotypes, each of which has the probability $L/(2^L - 1)$ to be a neighboring one. Taking into account that the number of viable genotypes is binomially distributed with parameter p and that the empty hypercube ($\nu = 0$) should yield $m_0 = 0$ we obtain

$$\overline{m}_0 = \sum_{\nu=1}^{2^L} \frac{(\nu - 1)}{2^L - 1} \binom{2^L}{\nu} p^\nu (1 - p)^{2^L - \nu} = \frac{2^L p - 1 + (1 - p)^{2^L}}{2^L - 1}, \tag{39}$$

which simplifies to $\overline{m}_0 = p$ when $2^L p \gg 1$. Note that the condition $2^L p \gg 1$ is naturally satisfied beyond the percolation threshold.

Fig 11 illustrates that the dynamics induced by mutation and selection already increase mutational robustness compared to $\overline{m}_0$ and that the addition of recombination even further increases mutational robustness for all values of p. The figure also displays the expected maximum number of viable neighbors of any genotype in the landscape, $\overline{m}_{\max}$, which provides an upper bound on the robustness. The fact that the numerically determined robustness remains below this bound for all p shows that the ability of recombination to locate the most connected genotype is limited. In S1 Appendix it is shown that $\lim_{L\to\infty} \overline{m}_{\max} = 1$ for $p > \frac{1}{2}$.

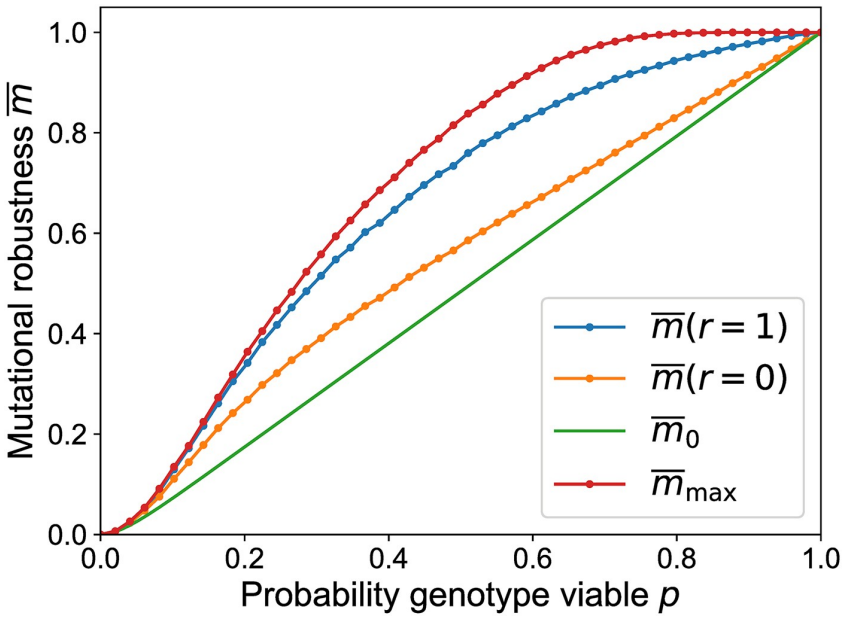

Fig 11. Mutational robustness in the percolation landscape as a function of the fraction of viable genotypes. The robustness for recombining ($\overline{m}(r=1)$) and non-recombining ($\overline{m}(r=0)$) populations is obtained by averaging over 6800 randomly generated landscapes with $L = 6$ and $\mu = 0.001$. In the same way the average maximal robustness $\overline{m}_{\max}$ is estimated. The full line shows the analytic expression (39) for the robustness of a uniformly distributed population.

https://doi.org/10.1371/journal.pcbi.1006884.g011

As outlined above, the algorithm used to generate Figs 10 and 11 computes the mutational robustness of a particular stationary frequency distribution of the recombining population which is smoothly connected to the unique non-recombining stationary state. Although one expects this state to be representative in the sense of being reachable from many initial conditions, for large enough r there can be multiple stationary states that will generally display different robustness (see Fig 9). To illustrate this point, S7 Fig shows the results of a simulation of the percolation model where all stationary states were identified using localized initial conditions, and the mutational robustness was computed separately for each state. Whereas on average the mutational robustness is always enhanced by recombination, there are rare instances when recombination reduces the robustness compared to the non-recombining case. This may happen, for example, if recombination traps the population on a small island of viable genotypes [22, 26, 55, 56].

Sea-cliff landscapes

In this section we introduce a novel class of fitness-landscape models (to be called sea-cliff landscapes) that interpolates between the mesa and percolation landscapes. Similar to the mesa landscape, the fitness values of the sea-cliff model are determined by the distance to a reference genotype κ^*. The model differs from the mesa landscape in that it is not assumed that all genotypes have zero fitness beyond a certain number of mutations. Instead, the likelihood for a mutation to be lethal (to "fall off the cliff") is taken to increase with the Hamming distance from the reference genotype. This is mathematically realized by a Heaviside step function $\theta(x)$ that contains an uncorrelated random contribution η_σ and the distance measure

$d(\sigma, \kappa^*)$,

$$w_\sigma = \theta[\eta_\sigma - d(\sigma, \kappa^*)] = \begin{cases} 1, & \text{if } \eta_\sigma > d(\sigma, \kappa^*), \\ 0, & \text{if } \eta_\sigma < d(\sigma, \kappa^*). \end{cases} \tag{40}$$

This construction is similar in spirit to the definition of the Rough-Mount-Fuji model [66, 67].

The average shape of the landscape can be tuned by the mean c and the standard deviation s of the distribution of the random variables η_σ, which we assume to be Gaussian in the following. The average fitness at distance d from the reference sequence is then given by

$$\overline{w}(d) = \text{Prob}(w_\sigma = 1) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{d-c}{s\sqrt{2}}\right)\right], \tag{41}$$

where $\text{erf}(x)$ is the error function. Note that the mesa landscape is reproduced if we take the limit $s \to 0$ for fixed c in the range $k < c < k+1$ and the percolation landscape is reproduced if we take a joint limit $s, |c| \to \infty$ with c/s fixed.

To fix c and s we introduce two distances $d_<$ and and $d_>$ such that $\overline{w}(d_<) = 0.99$ and $\overline{w}(d_>) = 0.01$, which leads to the relations

$$c = \frac{1}{2}(d_< + d_>) \quad \text{and} \quad s \approx 0.215(d_> - d_<). \tag{42}$$

The model can be generalized to include several predefined reference sequences,

$$w(\sigma) = \theta\left\{\sum_{\kappa^*} \theta[\eta_{\sigma,\kappa^*} - d(\sigma, \kappa^*)]\right\}, \tag{43}$$

which allows to create a genotype space with several highly connected clusters. Depending on the Hamming distance between the reference sequences and the variables c and s, clusters can be isolated or connected by viable mutations.

Fig 12 shows stationary states in the absence and presence of recombination for two different sea-cliff landscapes with one and two reference genotypes, respectively. Similar to the other landscape models, mutational robustness increases strongly with recombination, due to a population concentration within a neutral cluster. In the presence of two reference genotypes the recombining population should be concentrated within a single cluster. Otherwise lethal genotypes would be predominantly created through recombination of genotypes on different clusters. This observation can also be interpreted in the context of speciation due to genetic incompatibilities [49, 61]. Without recombination genotypes on both clusters have a nonvanishing frequency, but still the larger cluster is more populated. In contrast to the percolation landscape, robustness reaches a value close to unity for large r, because highly connected genotypes are abundant close to the reference sequence (S8 Fig).

Mutational robustness and recombination weight

Comparing Figs 6 and 10 and S8 Fig, the dependence of mutational robustness on the recombination rate is seen to be strikingly similar. Despite the very different landscape topographies, in all cases a small amount of recombination gives rise to a massive increase in robustness compared to the non-recombining baseline. For the mesa landscape this effect can be plausibly attributed to the focusing property of recombination, which counteracts the entropic spreading towards the fitness brink and localizes the population inside the plateau of viable genotypes. In the case of the holey landscapes, however, it is not evident that focusing the

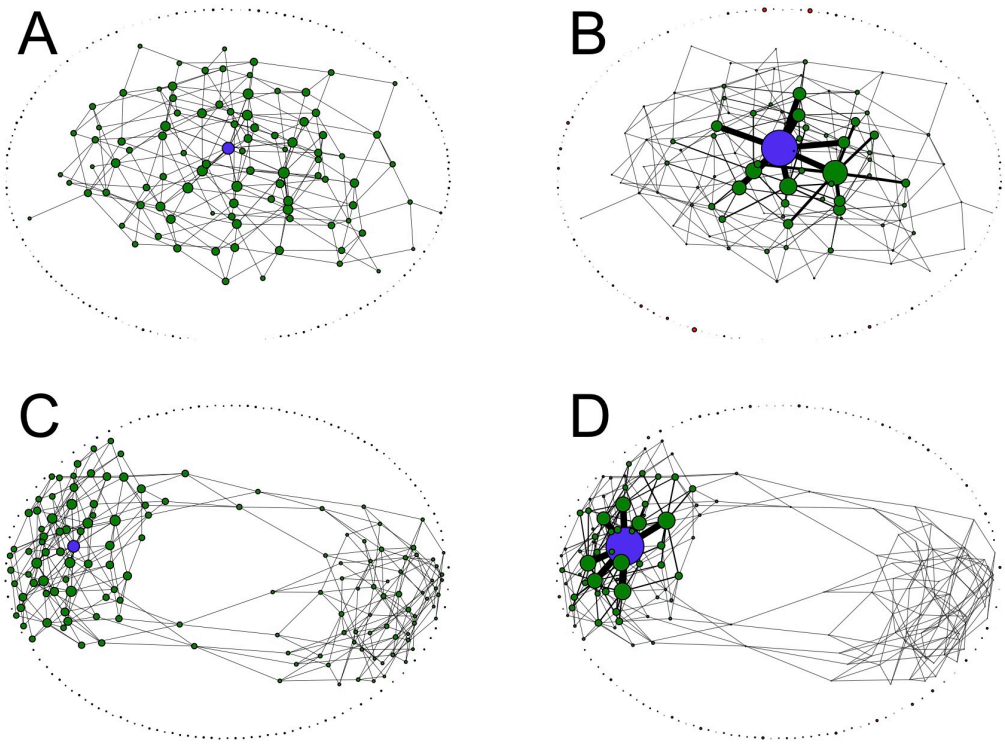

Fig 12. Stationary states in two different sea-cliff landscapes with and without recombination. (A,B) A single reference genotype with landscape parameters $L = 8$, $d_< = 1$ and $d_> = 6$. (C,D) Two reference genotypes which are antipodal to each other with landscape parameters $L = 8$, $d_< = 2$ and $d_> = 4.2$. (A,C) Stationary frequency distribution in the absence of recombination. (B,D) Stationary frequency distribution with uniform crossover and $r = 1$. In all cases node areas are proportional to genotype frequencies, and the recombination center is marked in blue. The edge width between neighboring genotypes is proportional to the frequency of the more populated one. The mutation rate is $\mu = 0.01$.

https://doi.org/10.1371/journal.pcbi.1006884.g012

population towards the center of its genotypic range will on average increase robustness, since viable and lethal genotypes are randomly interspersed.

To establish the relation between recombination and mutational robustness on the level of individual genotypes, in Fig 13 we plot the recombination weight of each genotype against its robustness m_σ. A clear positive correlation between the two quantities is observed both for percolation and sea-cliff landscapes. Additionally we differentiate between viable and lethal genotypes. In the percolation landscape viable genotypes are uniformly distributed in the genotype space, which implies that lethal and viable genotypes have on average the same number of viable point-mutations. Nevertheless the recombination weight of viable genotypes is larger. The fitness of a genotype influences its own recombination weight, because the genotype itself is a possible parental genotype in the recombination event.

In non-neutral fitness landscapes the redistribution of the population through recombination competes with selection responding to fitness differences, and the generalized definition (11) of the recombination weight captures this interplay. To exemplify the relation between recombination weight and mutational robustness in this broader context, we use an empirical fitness landscape for the filamentary fungus *Aspergillus niger* originally obtained in [68]. In a nutshell, two strains of *A. niger* (N411 and N890) were fused to a diploid which is unstable and creates two haploids by random chromosome arrangement. Both strains are isogenic to each

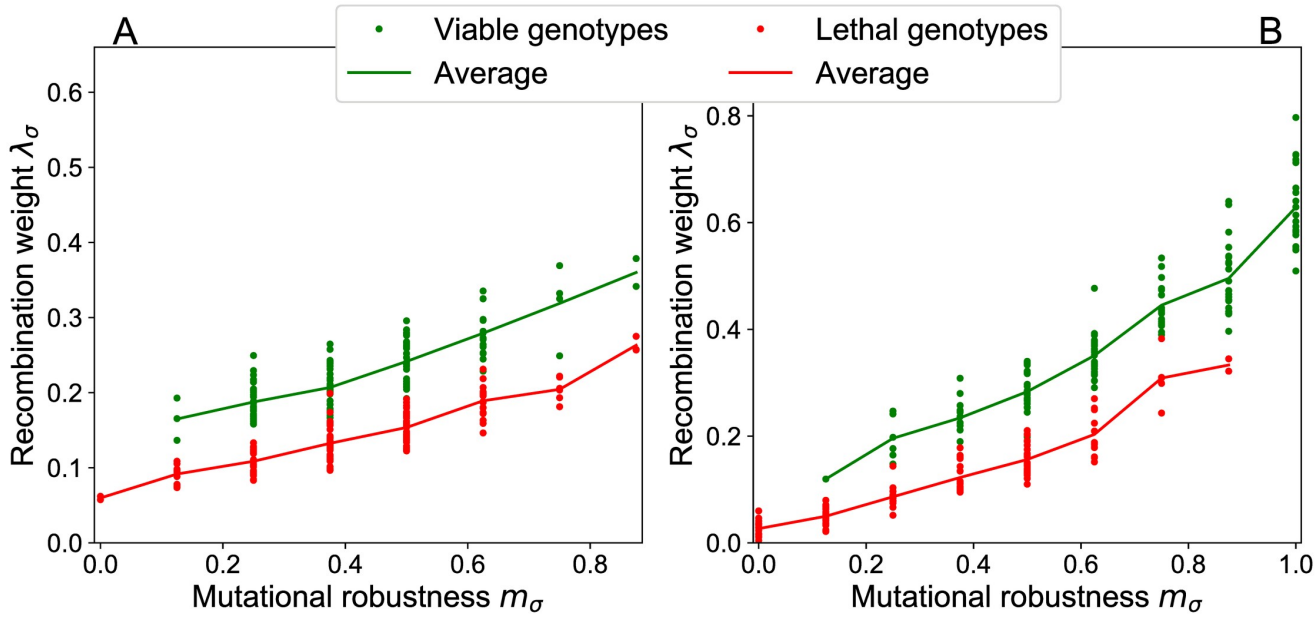

Fig 13. Mutational robustness correlates with recombination weight. The recombination weight of genotypes is plotted against their mutational robustness for (A) a percolation landscape with parameters $L = 8$, $p = 0.4$ and (B) a sea-cliff landscape with parameters $L = 8$, $d_< = 2$, $d_> = 6$. For the evaluation of the recombination weight (10), uniform crossover at rate $r = 1$ is assumed.

https://doi.org/10.1371/journal.pcbi.1006884.g013

other, except that N890 has 8 marker mutations on different chromosomes, which were induced by low UV-radiation. Through this process $2^8 = 256$ haploid segregants can theoretically be created of which 186 were isolated in the experiment. As a result of a statistical analysis it was concluded that the missing 70 haploids have zero fitness [69].

In order to illustrate the fitness landscape, a network representation is employed where genotypes are arranged in a plane according to their fitness and their Hamming distance to the wild type, which in this case is the genotype of maximal fitness. In Fig 14A and 14B node sizes are adjusted to the recombination weights and mutational robustness of genotypes, respectively, in order to display the distribution of these quantities. In accordance with the analyses for neutral fitness landscapes, a clear correlation between the recombination weights and mutational robustness is shown in Fig 14C. Since fitness values are not binary we further consider the correlation between the recombination weights and fitness values (Fig 14D). The recombination center is one of the maximally robust genotypes with $m_\sigma = 1$, but it is not the fittest within this group. The wild type has maximal fitness but, by comparison, lower robustness ($m_\sigma = 7/8$).

Fig 15 highlights how the recombination weights change as a function of the recombination rate and how this affects the stationary state of a population. For small recombination rates the recombination weight of each genotype mainly depends on its own fitness, and therefore the wild type coincides with the recombination center. With increasing recombination rate the connectivity of the surrounding genotype network becomes more important and the recombination center switches to a genotype at Hamming distance $d = 2$. In contrast to the numerical protocol described previously, in the simulations used to generate Fig 15D–15F the population is reset to a uniform distribution before the recombination rate is increased. Otherwise the population would continue to adapt to the wild type, which has the highest fitness and from

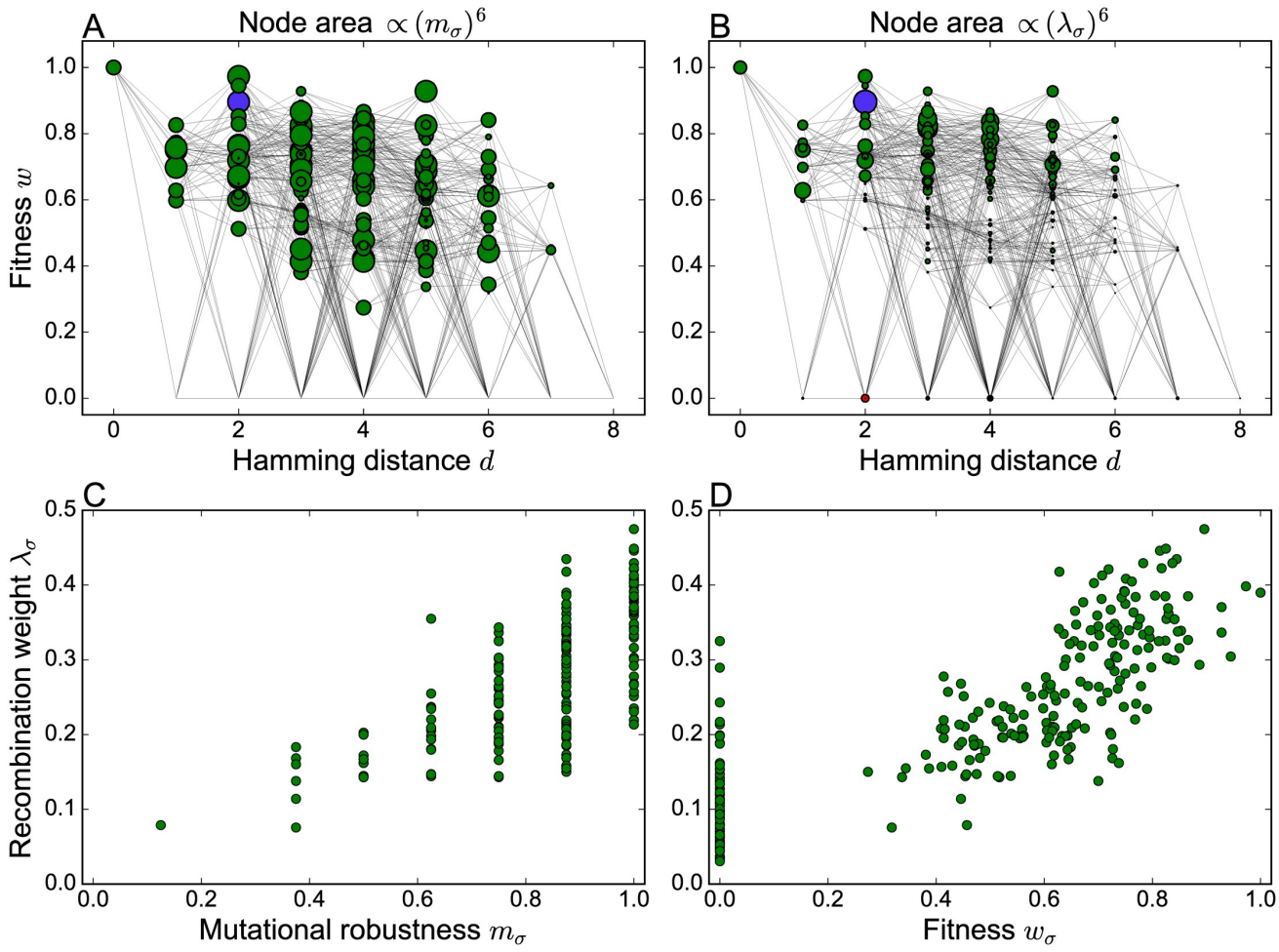

Fig 14. The empirical *A. niger* fitness landscape. (A,B) Two-dimensional network representation of the fitness landscape with node sizes determined by the mutational robustness m_σ and the recombination weight λ_σ, respectively. In order to make the differences between genotypes more conspicuous, the node area is chosen proportional to the sixth power of these quantities. The recombination weight is evaluated for uniform crossover with $r = 1$, and the recombination center is highlighted in purple. (C,D) Recombination weight plotted against mutational robustness and genotype fitness, respectively. Lethal genotypes with $w_\sigma = 0$ appear only in panel D.

https://doi.org/10.1371/journal.pcbi.1006884.g014

which it cannot escape because of peak trapping [22, 26]. Starting from an initially uniform distribution the population will adapt to one of three possible final genotypes which depend on the recombination rate. For small and large recombination rates the most abundant genotype coincides with the recombination center (Fig 15D and 15F), whereas for intermediate recombination rates the population chooses another genotype that is also located at Hamming distance $d = 2$ but has higher fitness (Fig 15E). The recombination center ultimately dominates the population, not only because it is maximally connected ($m_\sigma = 1$), but also because the genotypes that it is connected to have high fitness. In this sense the sequence of transitions in the most abundant genotype that occur with increasing recombination rate is akin to the scenario described previously in non-recombining populations as the "survival of the flattest" [48, 70]. Along this sequence mutational robustness increases monotonically whereas the average fitness of the population actually declines (S9 Fig).

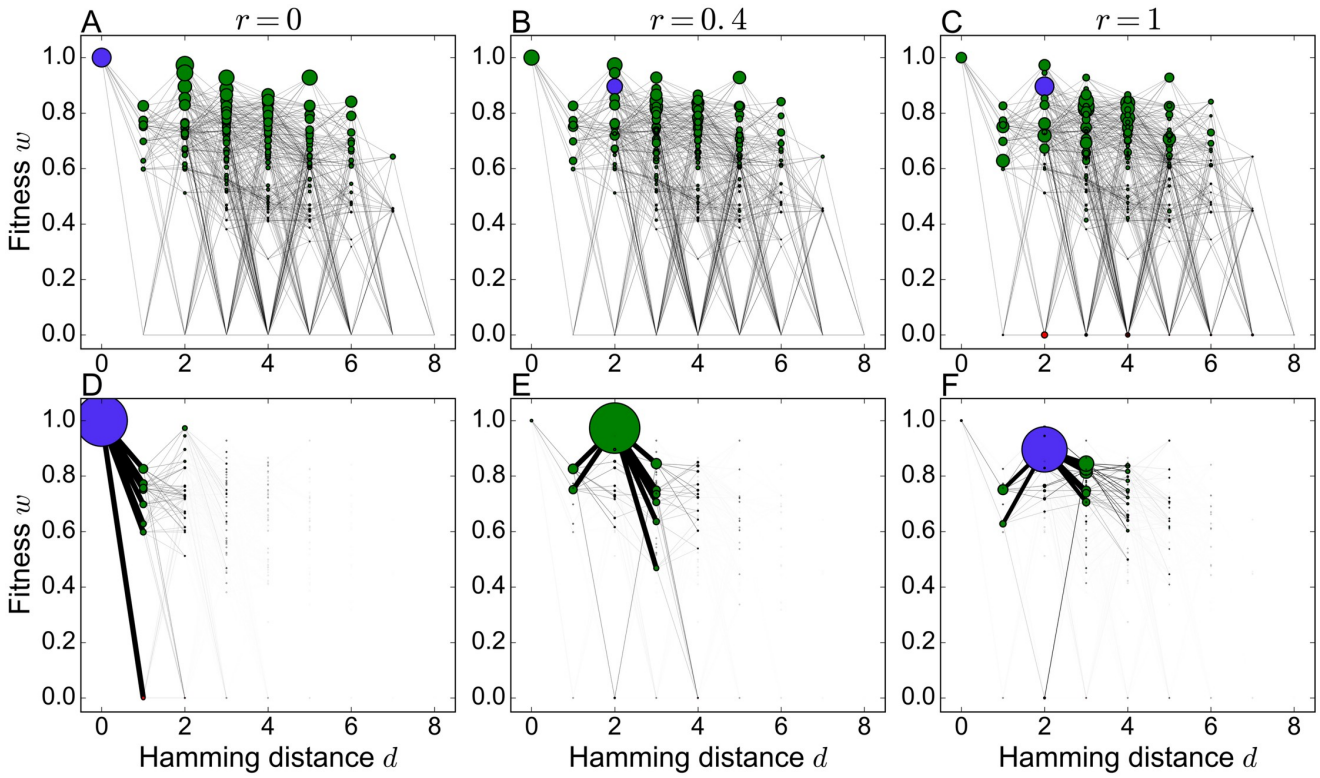

Fig 15. Recombination weights and stationary states at different recombination rates. (A-C) Two-dimensional network representation of the *A. niger* fitness landscape with node areas proportional to the sixth power of the recombination weight for recombination rates $r = 0$, $r = 0.4$ and $r = 1$, respectively. (D-F) Two-dimensional network representation of the *A. niger* fitness landscape with node areas proportional to the stationary genotype frequency at the same recombination rates and mutation rate $\mu = 0.005$. The edge width between neighboring genotypes is proportional to the frequency of the more populated one.

https://doi.org/10.1371/journal.pcbi.1006884.g015

Discussion

Despite a century of research into the evolutionary bases of recombination, a general mechanism explaining the ubiquity of genetic exchange throughout the domains of life has not been found [17, 18]. Even within the idealized scenario of a population evolving in a fixed environment, whether or not recombination speeds up adaptation and leads to higher fitness levels depends in a complicated way on the structure of the fitness landscape and the parameters of the evolutionary dynamics [21–26].

The most important finding of the present work is that, by comparison, the effect of recombination on mutational robustness is much simpler and highly universal. Irrespective of the number of loci, the structure of the fitness landscape or the recombination scheme, recombination leads to a significant increase of robustness that is usually much stronger than the previously identified effect of selection [32–34]. This suggests that the evolution of recombination may be closely linked to the evolution of robustness, and that similar selective benefits are involved in the two cases. Although the relation of robustness to evolutionary fitness is subtle and not fully understood [27], it has been convincingly argued that robustness enhances evolvability and hence becomes adaptive in changing environments [29, 31, 71, 72]. A common perspective on recombination, robustness and evolvability can help to develop novel hypotheses about the evolutionary origins of these phenomena that can be tested in future computational or empirical studies.

On a quantitative level, we have shown that robustness generally depends on the ratio of recombination to mutation rates, and that the robustness-enhancing effect saturates when $r \gg \mu$. This observation highlights the importance of r/μ as an evolutionary parameter. Interestingly, even in bacteria and archaea, which have traditionally been regarded as essentially non-recombining, the majority of species displays values of r/μ that are significantly larger than one [73–75]. Similarly, a recent study of the evolution of *Siphoviridae* phages revealed a ratio of recombination events to mutational substitutions of about 24 [76]. In eukaryotes this ratio is expected to be considerably higher [40]. This indicates that most organisms maintain a rate of recombination that is sufficient to reap its evolutionary benefits in terms of increased robustness.

In order to clarify the mechanism through which recombination enhances robustness, we have introduced the concept of the recombination weight, which is a measure for the likelihood of a genotype to arise from the recombination of two viable parental genotypes. The recombination weight defines a "recombination landscape" over the space of genotypes which is similar in spirit to, but distinct from, previous mathematical approaches to conceptualizing the way in which recombining populations navigate a fitness landscape [77]. It is complementary to the more commonly used notion of a recombination load, which refers to the likelihood for a viable genotype to recombine to a lethal one [41, 42]. In many cases the maximum of the recombination weight correctly predicts the most populated genotype in a recombining population at low mutation rate. Moreover, the concept generalizes to non-neutral landscapes and thus permits to address situations where selection and recombination compete.

Provided recombination weight is correlated with mutational robustness for the individual genotypes, this explains the positive effect of recombination on the population-level robustness. Whether or not such a correlation exists will generally depend on the structure of the fitness landscapes. For simple neutral landscapes such as the mesa landscape it is an immediate consequence of the focusing property of recombination, but for more complex neutral networks the relationship between the two quantities is nontrivial and needs to be studied on a case-by-case basis. Although a positive correlation was observed numerically both for the holey landscapes and the empirical landscape considered in this work, it is not difficult to construct landscapes where the genotypes with high recombination weight are not highly robust. As a simple but instructive example, in S10 Fig we show results for an 'atoll' landscape where a ring of viable genotypes surrounds a central hole of lethals.

Throughout this work the effects of genetic drift have been neglected. We expect that our results will be applicable to finite populations as long as the population is sufficiently diverse rather than being monomorphic. This requires the population-wide mutation rate $N\mu L$ to be much larger than unity [32, 44]. If $N\mu L \ll 1$ the population is almost always monomorphic and recombination has no effect. In this regime the population explores the fitness landscape as a random walker and the observed mutational robustness is the uniform robustness m_0. In S11 Fig we present the results of finite population simulations on a mesa landscape, which show a sharp transition from the random walk regime to the behavior predicted by the deterministic theory when $N\mu L \sim 1$.

Future work should be directed towards extending the present investigation to more realistic genotype-phenotype maps arising, for example, from the secondary structures of biopolymers such as RNA or proteins [39, 40, 44], or from simple genetic, metabolic or logical networks [29, 41, 43, 78]. There is ample evidence from numerical studies that a favorable effect of recombination on mutational robustness is present also in these more complex systems, but a detailed analysis of the underlying mechanism has not been carried out. This would entail, in particular, the generalization to genotype spaces composed of sequences

carrying more than two alleles per site. We expect that at least part of the analysis for the mesa landscapes carries over to this setting, and in fact some results for the non-recombining case have already been obtained [48]. More importantly, the role of the topology of the corresponding neutral networks in shaping the correlation between recombination weight and robustness needs to be explored systematically. Research along these lines will help to corroborate the relationship between recombination and robustness that we have sketched, and to further elucidate the origins of these two pervasive features of biological evolution.

Supporting information

S1 Appendix. This appendix contains detailed derivations of analytic results presented in the main text.
(PDF)

S1 Fig. Population heterogeneity decreases with increasing recombination rate. The figure shows the entropy of the genotype frequency distribution in the two-locus model defined as $S = -\sum_\sigma f_\sigma^* \ln(f_\sigma^*)$. For small mutation rates the strongly recombining population primarily consists of a single genotype, which implies that $S \to 0$.
(PDF)

S2 Fig. Mutational robustness for the mesa landscape with communal recombination. The figure compares the analytic approximations in Eqs (29) and (30) to the numerical solution of the stationary genotype frequency distribution for the communal recombination scheme. The two panels show the mutational robustness as a function of the genome-wide mutation rate in linear (A) and double-logarithmic (B) scales, respectively. The parameters of the mesa landscape are $L = 30$ and $k = 3$.
(PDF)

S3 Fig. Mutational robustness in a mesa landscape with different recombination schemes. The figure compares the analytic results for communal recombination (m_{cr}) with numerical data obtained using uniform crossover (m_{uc}) and one-point crossover (m_{opc}) at $r = 1$. The landscape parameters are $L = 5$, $k = 2$ and robustness is plotted as a function of the genome-wide mutation rate $L\mu$. (A) Mutational robustness on linear scales. (B) Double-logarithmic plot of $1 - m$ vs. $L\mu$, illustrating the power-law behavior $1 - m \sim (L\mu)^b$ with the exponent $b = k/(k + 1) = 2/3$ predicted by the analysis of the communal recombination model.
(PDF)

S4 Fig. Mutational robustness for the mesa landscape in the absence of recombination. The figure compares the analytic predictions in Eqs (35) and (36) to the numerical solution for the genotype frequency distribution in the absence of recombination. The two panels show the mutational robustness (A) after selection and (B) after mutation as a function of the scaled mesa width $x_0 = k/L$ for $L = 1000$ and $U = 0.01$.
(PDF)

S5 Fig. Mutational robustness in mesa landscapes with and without recombination. Numerical results for communal recombination (m_{cr}) and no recombination (m_{nr}) are shown as dots. The mutational robustness m_0 of a uniformly distributed population, given by Eq (37), as well as the analytic expressions Eqs (30) and (36) are depicted as lines. (A) Robustness as a function of mutation rate $U = L\mu$ for a landscape with $L = 1000$ and $k = 10$. (B) Robustness as a function of mesa width k at fixed $L = 1000$ and $U = L\mu = 0.01$. (C) Robustness as a function of

genome length L at fixed $k = 10$ and $U = 0.01$. (D) Robustness as a function of genome length L at fixed $k = 10$ and $\mu = 0.001$.
(PDF)

S6 Fig. Recombination weight in a mesa landscape. The parameters of the mesa landscape are $L = 100$ and $k = 10$. For $r = 0$ the recombination weight is directly proportional to the fitness and hence equal for all viable genotypes. Already small rates of recombination are sufficient to redistribute the recombination weight such that the weight of genotypes with small Hamming distance is strongly enhanced. Beyond $d = 20$ the recombination weight is identically zero, since the recombinant of two viable genotypes cannot carry more than $2k$ mutations.
(PDF)

S7 Fig. Mutational robustness for different stationary states within a percolation landscape. The figure compares the mutational robustness of non-recombining ($r = 0$) and recombining ($r = 1$) populations on individual realizations of the percolation model with $L = 6$ and three values of p. In order to obtain different stationary states we used localized initial population distributions of the form $f_\tau(0) = \delta_{\tau\sigma}$ for all genotypes with mutational robustness $m_\sigma \neq 0$ and propagated them until stationarity. Since the stationary populations are usually highly concentrated for large r and small μ, this is a natural choice in order to access all stationary states. Each data point represents the robustness of the recombining population $m(r = 1)$ for a particular stationary state. Data points within the same landscape are plotted above the corresponding unique robustness of the non-recombining population $m(r = 0)$ and connected by a vertical line. The orange crosses show the average over all initial conditions.
(PDF)

S8 Fig. Average mutational robustness in the sea-cliff landscape as a function of recombination rate. Mutational robustness is computed for 200 randomly generated sea-cliff landscapes with parameters $L = 6$, $d_< = 1$ and $d_> = 5$, and the results are averaged to obtain $\overline{m}(r)$. The mutation rate is $\mu = 0.001$.
(PDF)

S9 Fig. Mutational robustness and average fitness in the empirical *A. niger* fitness landscape. The mutational robustness and the population-averaged fitness in the stationary state are computed as a function of recombination rate by evolving the population from a uniform initial genotype distribution at mutation rate $\mu = 0.005$. Jumps mark changes in the most populated genotype.
(PDF)

S10 Fig. Recombination on an atoll landscape. This landscape is similar to the mesa landscape but includes an inner critical radius within which genotypes are lethal. In this example the inner radius is chosen to be 1 such that only the wild type is lethal. The outer radius is 2 and the sequence length is $L = 7$. The recombination rate is $r = 1$ and the mutation rate is $\mu = 0.001$. The frequencies f_n of the stationary state at the same Hamming distance n are lumped together. The population is concentrated at distance 1 which is most robust since only one point mutation is lethal, but the recombination center coincides with the lethal wild type. This example shows that the correlation between recombination weight and mutational robustness depends on the topology of the neutral network.
(PDF)

S11 Fig. Finite population size effects. The figure shows the mutational robustness in a mesa landscape with parameter $L = 6$, $k = 2$ as a function of mutation rate. The finite population

results were obtained using Wright-Fisher dynamics for $N = 1000$ individuals. For small mutation rates such that $N\mu L \ll 1$ the monomorphic population performs a random walk among viable genotypes, which leads to the uniform mutational robustness m_0 given by Eq (37) (green dashed line). In this regime recombination cannot have any effect. For $N\mu L > 1$ the robustness rises sharply to the value predicted by the infinite population approach. At the maximal mutation rate $\mu = 0.5$ the population is uniformly distributed among all (lethal or viable) genotypes after the mutation step and recombination has again no effect.
(PDF)

Acknowledgments

JK acknowledges the kind hospitality of the Higgs Centre for Theoretical Physics at the University of Edinburgh during the completion of the manuscript.

Author Contributions

Conceptualization: Alexander Klug, Joachim Krug.

Formal analysis: Alexander Klug, Su-Chan Park, Joachim Krug.

Funding acquisition: Su-Chan Park, Joachim Krug.

Investigation: Alexander Klug, Su-Chan Park, Joachim Krug.

Methodology: Alexander Klug, Su-Chan Park, Joachim Krug.

Project administration: Joachim Krug.

Software: Alexander Klug.

Supervision: Joachim Krug.

Visualization: Alexander Klug.

Writing – original draft: Alexander Klug, Su-Chan Park, Joachim Krug.

Writing – review & editing: Alexander Klug, Su-Chan Park, Joachim Krug.

References

1. Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? Nat Rev Microbiol. 2011; 9:617–626. https://doi.org/10.1038/nrmicro2614 PMID: 21725337
2. Redfield RJ. Do bacteria have sex? Nat Rev Genet. 2001; 2:634–639. https://doi.org/10.1038/35084593 PMID: 11483988
3. Feil EJ, Spratt BG. Recombination and the Population Structures of Bacterial Pathogens. Annu Rev Microbiol. 2001; 55:561–590. https://doi.org/10.1146/annurev.micro.55.1.561 PMID: 11544367
4. Maynard Smith J. The Evolution of Sex. Cambridge, UK: Cambridge University Press; 1978.
5. Michod RE, Levin BR, editors. The Evolution of sex: an examination of current ideas. Sunderland, Massachusetts: Sinauer Associates Inc.; 1988.
6. Holmes CM, Nemenman I, Weissman DB. Increased adaptability to sudden environmental change can more than make up for the two-fold cost of males. Europhys Lett. 2018; 123:58001. https://doi.org/10.1209/0295-5075/123/58001
7. D J P Engelmoer ID, Rozen DE. Conservative Sex and the Benefits of Transformation in Streptococcus pneumoniae. PLOS Pathogens. 2013; 9:e1003758. https://doi.org/10.1371/journal.ppat.1003758
8. Moradigaravand D, Engelstädter J. The Evolution of Natural Competence: Disentangling Costs and Benefits of Sex in Bacteria. Amer Nat. 2013; 182:E112–E126. https://doi.org/10.1086/671909
9. Weismann A. Essays Upon Heredity and Kindred Biological Problems. Oxford University Press; 1889.
10. Muller HJ. Some genetic aspects of sex. Amer Nat. 1932; 66:118–138. https://doi.org/10.1086/280418

11. Muller HJ. The relation of recombination to mutational advance. Mutat Res. 1964; 1:2–9. https://doi.org/10.1016/0027-5107(64)90047-8
12. Felsenstein J. The evolutionary advantage of recombination. Genetics. 1974; 78:737–756. PMID: 4448362
13. Kondrashov AS. Deleterious mutations and the evolution of sexual reproduction. Nature. 1988; 336:435–440. https://doi.org/10.1038/336435a0 PMID: 3057385
14. Kondrashov AS. Classification of hypotheses on the advantage of amphimixis. J Hered. 1993; 84:372–387. https://doi.org/10.1093/oxfordjournals.jhered.a111358 PMID: 8409359
15. Feldman MW, Otto SP, Christiansen FB. Population genetic perspectives on the evolution of recombination. Annu Rev Genet. 1997; 30:261–295. https://doi.org/10.1146/annurev.genet.30.1.261
16. Burt A. Perspective: Sex, recombination, and the efficacy of selection—was Weismann right? Evolution. 2000; 54:337–351. https://doi.org/10.1111/j.0014-3820.2000.tb00038.x PMID: 10937212
17. Otto SP, Lenormand T. Resolving the paradox of sex and recombination. Nat Rev Genet. 2002; 3:252–261. https://doi.org/10.1038/nrg761 PMID: 11967550
18. de Visser JAGM, Elena SF. The evolution of sex: empirical insights into the roles of epistasis and drift. Nat Rev Genet. 2007; 8:139–149. https://doi.org/10.1038/nrg1985 PMID: 17230200
19. Otto SP. The evolutionary enigma of sex. Amer Nat. 2009; 174:S1–S14. https://doi.org/10.1086/599084
20. de Visser JAGM, Krug J. Empirical fitness landscapes and the predictability of evolution. Nat Rev Genet. 2014; 15:480–490. https://doi.org/10.1038/nrg3744 PMID: 24913663
21. Kondrashov FA, Kondrashov AS. Multidimensional epistasis and the disadvantage of sex. Proc Nat Acad Sci USA. 2001; 98:12089–12092. https://doi.org/10.1073/pnas.211214298 PMID: 11593020
22. de Visser JAGM, Park SC, Krug J. Exploring the Effect of Sex on Empirical Fitness Landscapes. Amer Nat. 2009; 174:S15–S30. https://doi.org/10.1086/599081
23. Misevic D, Kouyos RD, Bonhoeffer S. Predicting the Evolution of Sex on Complex Fitness Landscapes. PLOS Comp Biol. 2009; 5:e1000510. https://doi.org/10.1371/journal.pcbi.1000510
24. Moradigaravand D, Engelstädter J. The Effect of Bacterial Recombination on Adaptation on Fitness Landscapes with Limited Peak Accessibility. PLOS Comp Biol. 2012; 8:e1002735. https://doi.org/10.1371/journal.pcbi.1002735
25. Moradigaravand D, Kouyos R, Hinkley T, Haddad M, Petropoulos CJ, Engelstädter J, et al. Recombination Accelerates Adaptation on a Large-Scale Empirical Fitness Landscape in HIV-1. PLOS Genetics. 2014; 10:e1004439. https://doi.org/10.1371/journal.pgen.1004439 PMID: 24967626
26. Nowak S, Neidhart J, Szendro IG, Krug J. Multidimensional Epistasis and the Transitory Advantage of Sex. PLOS Comp Biol. 2014; 10:e1003836. https://doi.org/10.1371/journal.pcbi.1003836
27. de Visser JAGM, Hermisson J, Wagner GP, Meyers LA, Bagheri-Chaichian H, Blanchard JL, et al. Perspective: Evolution and detection of genetic robustness. Evolution. 2003; 57:1959–1972. https://doi.org/10.1554/02-750R PMID: 14575319
28. Kitano H. Biological robustness. Nat Rev Genet. 2004; 5:826–837. https://doi.org/10.1038/nrg1471 PMID: 15520792
29. Wagner A. Robustness and Evolvability in Living Systems. Princeton: Princeton University Press; 2005.
30. Lenski RE, Barrick JE, Ofria C. Balancing robustness and evolvability. PLOS Biol. 2006; 4:e428. https://doi.org/10.1371/journal.pbio.0040428 PMID: 17238277
31. Masel J, Trotter MV. Robustness and evolvability. Trends Genet. 2010; 26:406–414. https://doi.org/10.1016/j.tig.2010.06.002 PMID: 20598394
32. van Nimwegen E, Crutchfield JP, Huynen M. Neutral evolution of mutational robustness. Proc Nat Acad Sci USA. 1999; 96:9716–9720. https://doi.org/10.1073/pnas.96.17.9716 PMID: 10449760
33. Bornberg-Bauer E, Chan HS. Modeling evolutionary landscapes: Mutational stability, topology, and superfunnels in sequence space. Proc Nat Acad Sci USA. 1999; 96:10689–10694. https://doi.org/10.1073/pnas.96.19.10689 PMID: 10485887
34. Wilke CO. Adaptive evolution on neutral networks. Bull Math Biol. 2001; 63:715–730. https://doi.org/10.1006/bulm.2001.0244 PMID: 11497165
35. Gavrilets S. Fitness Landscapes and the Origin of Species. Princeton: Princeton University Press; 2004.
36. Boerlijst MC, Bonhoeffer S, Nowak MA. Viral Quasi-Species and Recombination. Proc R Soc Lond B. 1996; 263:1577–1584. https://doi.org/10.1098/rspb.1996.0231

37. Wilke CO, Adami C. Evolution of mutational robustness. Mutat Res. 2003; 522:3–11. https://doi.org/10.1016/s0027-5107(02)00307-x PMID: 12517406

38. Gardner A, Kalinka AT. Recombination and the evolution of mutational robustness. J Theor Biol. 2006; 241:707–715. https://doi.org/10.1016/j.jtbi.2006.01.011 PMID: 16487979

39. Huynen MA, Hogeweg P. Pattern Generation in Molecular Evolution: Exploitation of the Variation in RNA Landscapes. J Mol Evol. 1994; 39:71–79. https://doi.org/10.1007/BF00178251 PMID: 7520506

40. Xia Y, Levitt M. Roles of mutation and recombination in the evolution of protein thermodynamics. Proc Nat Acad Sci USA. 2002; 99:10382–10387. https://doi.org/10.1073/pnas.162097799 PMID: 12149452

41. Azevedo RBR, Lohaus R, Srinivasan S, Dang KK, Burch CL. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. Nature. 2006; 440:87–90. https://doi.org/10.1038/nature04488 PMID: 16511495

42. Singhal S, Gomez SM, Burch CL. Recombination drives the evolution of mutational robustness. Curr Opin Syst Biol. 2019; 13:142–149. https://doi.org/10.1016/j.coisb.2018.12.003

43. Hu T, Banzhaf W, Moore JH. The effects of recombination on phenotypic exploration and robustness in evolution. Artificial Life. 2014; 20:457–470. https://doi.org/10.1162/ARTL_a_00145 PMID: 25148550

44. Szöllősi GJ, Derényi I. The effect of recombination on the neutral evolution of genetic robustness. Math Biosci. 2008; 214:58–62. https://doi.org/10.1016/j.mbs.2008.03.010 PMID: 18490032

45. Gerland U, Hwa T. On the Selection and Evolution of Regulatory DNA Motifs. J Mol Evol. 2002; 55:386–400. https://doi.org/10.1007/s00239-002-2335-z PMID: 12355260

46. Peliti L. Quasispecies evolution in general mean-field landscapes. Europhys Lett. 2002; 57:745–751. https://doi.org/10.1209/epl/i2002-00526-5

47. Berg J, Willmann S, Lässig M. Adaptive evolution of transcription factor binding sites. BMC Evol Biol. 2004; 4:42. https://doi.org/10.1186/1471-2148-4-42 PMID: 15511291

48. Wolff A, Krug J. Robustness and epistasis in mutation-selection models. Phys Biol. 2009; 6:036007. https://doi.org/10.1088/1478-3975/6/3/036007 PMID: 19411737

49. Gavrilets S. Evolution and speciation on holey adaptive landscapes. Trends Ecol Evol. 1997; 12:307–312. https://doi.org/10.1016/S0169-5347(97)01098-7 PMID: 21238086

50. Bürger R. The Mathematical Theory of Selection, Recombination and Mutation. Chichester, UK: John Wiley & Sons; 2000.

51. Crow JF, Kimura M. Evolution in sexual and asexual populations. Amer Nat. 1965; 99:439–450. https://doi.org/10.1086/282389

52. Eshel I, Feldman MW. On the evolutionary effect of recombination. Theor Pop Biol. 1970; 1:88–100. https://doi.org/10.1016/0040-5809(70)90043-2

53. Kimura M. The role of compensatory neutral mutations in molecular evolution. J Genet. 1985; 64:7–19. https://doi.org/10.1007/BF02923549

54. Higgs PG. Compensatory neutral mutations and the evolution of RNA. Genetica. 1998; 102/103:91–101. https://doi.org/10.1023/A:1017059530664

55. Park SC, Krug J. Bistability in two-locus models with selection, mutation, and recombination. J Math Biol. 2011; 62:763–788. https://doi.org/10.1007/s00285-010-0352-x PMID: 20617437

56. Altland A, Fischer A, Krug J, Szendro IG. Rare Events in Population Genetics: Stochastic Tunneling in a Two-Locus Model with Recombination. Phys Rev Lett. 2011; 106:088101. https://doi.org/10.1103/PhysRevLett.106.088101 PMID: 21405603

57. Weinreich DM, Sindi S, Watson RA. Finding the boundary between evolutionary basins of attraction, and implications for Wright's fitness landscape analogy. J Stat Mech. 2013; 2013:P01001. https://doi.org/10.1088/1742-5468/2013/01/P01001

58. Nowak MA, Boerlijst MC, Cooke J, Maynard Smith J. Evolution of genetic redundancy. Nature. 1997; 388:167–171. https://doi.org/10.1038/40618 PMID: 9217155

59. Phillips PC, Johnson NA. The Population Genetics of Synthetic Lethals. Genetics. 1998; 150:449–458. PMID: 9725860

60. Baake E, Georgii HO. Mutation, selection, and ancestry in branching models: a variational approach. J Math Biol. 2007; 54:257–303. https://doi.org/10.1007/s00285-006-0039-5 PMID: 17075709

61. Paixão T, Bassler KE, Azevedo RBR. Emergent speciation by multiple Dobzhansky-Muller incompatibilities. bioRxiv. 2015;Available from: https://www.biorxiv.org/content/early/2015/07/07/008268.

62. Neher RA, Shraiman BI, Fisher DS. Rate of Adaptation in Large Sexual Populations. Genetics. 2010; 184:467–481. https://doi.org/10.1534/genetics.109.109009 PMID: 19948891

63. Gavrilets S, Gravner J. Percolation on the Fitness Hypercube and the Evolution of Reproductive Isolation. J Theor Biol. 1997; 184:51–64. https://doi.org/10.1006/jtbi.1996.0242 PMID: 9039400

64. Reidys CM. Random Induced Subgraphs of Generalized *n*-Cubes. Adv Appl Math. 1997; 19:360–377. https://doi.org/10.1006/aama.1997.0553

65. Reidys CM. Large components in random induced subgraphs of *n*-Cubes. Discrete Mathematics. 2009; 309:3113–3124.

66. Aita T, Uchiyama H, Inaoka T, Nakajima M, Kokubo T, Husimi Y. Analysis of a Local Fitness Landscape with a Model of the Rough Mt. Fuji-Type Landscape: Application to Prolyl Endopeptidase and Thermolysin. Biopoly. 2000; 54:64–79. https://doi.org/10.1002/(SICI)1097-0282(200007)54:1%3C64::AID-BIP70%3E3.0.CO;2-R

67. Neidhart J, Szendro IG, Krug J. Adaptation in Tunably Rugged Fitness Landscapes: The Rough Mount Fuji Model. Genetics. 2014; 198:699–721. https://doi.org/10.1534/genetics.114.167668 PMID: 25123507

68. de Visser JAGM, Hoekstra RF, van den Ende H. Test of interaction between genetic markers that affect fitness in *Aspergillus niger*. Evolution. 1997; 51:1499–1505. https://doi.org/10.1111/j.1558-5646.1997.tb01473.x PMID: 28568630

69. Franke J, Klözer A, de Visser JAGM, Krug J. Evolutionary accessibility of mutational pathways. PLOS Comp Biol. 2011; 7:e1002134. https://doi.org/10.1371/journal.pcbi.1002134

70. Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature. 2001; 412:331–333. https://doi.org/10.1038/35085569 PMID: 11460163

71. Draghi JA, Parsons TL, Wagner GP, Plotkin JB. Mutational robustness can facilitate adaptation. Nature. 2010; 463:353–355. https://doi.org/10.1038/nature08694 PMID: 20090752

72. Payne JL, Wagner A. The causes of evolvability and their evolution. Nat Rev Genet. 2019; 20:24–38. https://doi.org/10.1038/s41576-018-0069-z PMID: 30385867

73. Guttman DS, Dykhuizen DE. Clonal divergence in Escherichia-coli as a result of recombination, not mutation. Science. 1994; 266:1380–1383.

74. Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. ISME J. 2009; 3:199–208. https://doi.org/10.1038/ismej.2008.93 PMID: 18830278

75. Didelot X, Maiden MCJ. Impact of recombination on bacterial evolution. Trends Microbiol. 2010; 18:315–322. https://doi.org/10.1016/j.tim.2010.04.002 PMID: 20452218

76. Kupczok A, Neve H, Huang KD, Hoeppner MP, Heller KJ, Franz CMAP, et al. Rates of Mutation and Recombination in *Siphoviridae* Phage Genome Evolution over Three Decades. Mol Biol Evol. 2018; 35:1147–1159. https://doi.org/10.1093/molbev/msy027 PMID: 29688542

77. Stadler PF, Wagner G. Algebraic Theory of Recombination Spaces. Evol Comp. 1997; 5:241–275. https://doi.org/10.1162/evco.1997.5.3.241

78. García-Martín JA, Catalán P, Manrubia S, Cuesta JA. Statistical theory of phenotype abundance distributions: A test through exact enumeration of genotype spaces. Europhys Lett. 2018; 123:28001. https://doi.org/10.1209/0295-5075/123/28001

Recombination and mutational robustness in neutral fitness landscapes: Supplementary appendix

Alexander Klug,[1] Su-Chan Park,[2] and Joachim Krug[1]
[1]Institute for Biological Physics, University of Cologne, Cologne, Germany
[2]Department of Physics, The Catholic University of Korea, Bucheon, Republic of Korea

I. VISUALIZATION OF FITNESS LANDSCAPES AS NETWORKS

In order to visualize random neutral fitness landscapes with more than two loci we make use of a network representation, where genotypes that differ by a single mutation are connected by an edge. Nodes of the network then represent genotypes, which are arranged according to a spring layout that is based on a Fruchterman-Reingold force-directed algorithm [1]. To describe this algorithm briefly, nodes are made to repel each other, which is counteracted by edges that function as springs. This leads to a process of spring-force relaxation that arrives at an equilibrium state which in turn is used for the node positions. The equilibrium state is characterized by clustering of highly connected regions of nodes. Therefore this algorithm is only useful if not all nodes have the same number of edges. Hence edges attached to lethal genotypes are deleted. This leads to a network in which only viable genotypes that differ by a single mutation are connected. Lethal genotypes are off the grid and create a ring of repelled nodes.

II. TWO-LOCUS MODEL WITH UNIDIRECTIONAL MUTATION

Following Nowak *et al.* [2], we consider the two-locus model with unidirectional mutations from allele 0 to allele 1 at rate μ and one-point crossover at rate r. Based on the relation

$$q_0 = \frac{r}{4\tilde{\mu}} q_1^2, \quad \tilde{\mu} = \frac{\mu}{1-\mu} \tag{A1}$$

between the lumped genotype frequencies after selection, the expression

$$M = q_0 + \frac{1}{2} q_1 = 1 - \frac{\tilde{\mu}}{r}\left(\sqrt{1 + \frac{r}{\tilde{\mu}}} - 1\right) \tag{A2}$$

can be derived for the mutational robustness after selection. For $r \to 0$ this reduces to $M = \frac{1}{2}$ independent of μ, which is smaller than the value $M = \frac{2}{3}$ expected for a random distribution over the viable genotypes ($q_0 = \frac{1}{3}, q_1 = \frac{2}{3}$). In the absence of recombination, the unidirectional mutations drive the entire population into the least robust genotypes (0,1) and (1,0), such that $q_0 = 0$ and $q_1 = 1$. On the other hand, for $r = 1$ Eq (A2) becomes $M = (1 + \sqrt{\mu})^{-1}$, which can be compared to the corresponding expression

$$M = \frac{m}{1 - f_2} = \frac{2}{2 - \mu + \sqrt{\mu^2 + 4\mu}} \tag{A3}$$

obtained from Eq (17) of the main text. The two expressions coincide for $\mu \to 0$, but for larger μ the bidirectional model has higher robustness, because both selection and recombination contribute to focusing the population onto the robust genotype (0,0) (Fig A1).

III. MUTATIONAL ROBUSTNESS ON THE MESA LANDSCAPE WITH COMMUNAL RECOMBINATION

In this section, we calculate the mutational robustness in equilibrium for the mesa landscape, using the communal recombination scheme [3]. Since fitness depends only on the Hamming distance from the wild type, the equilibrium allele-frequency distribution at each locus is the same after mutation. In the following we denote the (equilibrium) frequency of allele 0 (1) after the mutation step by π_0 ($\pi_1 = 1 - \pi_0$). Then the equilibrium frequency f^*_σ of a genotype σ after recombination becomes

$$f^*_\sigma = \pi_0^{L-n} \pi_1^n, \tag{A4}$$

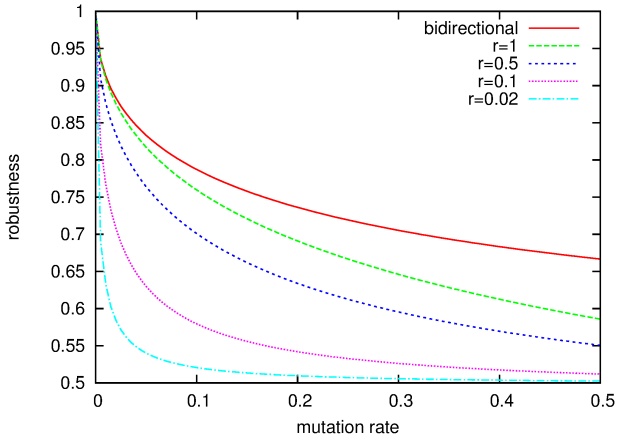

Fig A1. **Mutational robustness in the two-locus model with unidirectional mutations**. The figure shows the mutational robustness after selection obtained for the unidirectional mutation scheme, Eq (A2), as function of μ for different r. For comparison the corresponding result Eq (A3) for the bidirectional mutation scheme with $r = 1$ is also depicted.

where n is the Hamming distance from the wild type. The lumped frequency of all genotypes in the class with n mutations is then given by

$$f_n = \binom{L}{n} \pi_0^{L-n} \pi_1^n. \tag{A5}$$

Denoting the corresponding lumped frequency after selection by q_n and using the mesa landscape defined in Eq (28) of the main text, we get

$$q_n = \begin{cases} f_n \bar{w}^{-1}, & n \leq k, \\ 0, & n > k, \end{cases} \tag{A6}$$

where $\bar{w} = \sum_{n=0}^{k} f_n$ is the mean fitness. The lumped frequency p_n after mutation then satisfies

$$p_d = \sum_{n=0}^{L} \mu(d|n) q_n, \tag{A7}$$

where $\mu(d|n)g_d$ is the probability that a mutation changes the Hamming distance from n to d. The p_d in turn determine the allele frequency after mutation through

$$\pi_1 = \frac{1}{L} \sum_{d=0}^{L} d p_d = \frac{1}{L} \sum_d d \sum_n \mu(d|n) q_n = \frac{1}{L} \sum_n h(n) q_n, \tag{A8}$$

where $h(n) = \sum_d \mu(d|n) d$ is the average Hamming distance of a mutant generated from a genotype with Hamming distance n. One can easily calculate $h(n)$ for the mutation scheme Eq (4) of the main text, which yields

$$h(n) = n(1-\mu) + (L-n)\mu = L\mu + (1-2\mu)n. \tag{A9}$$

This expression has a simple interpretation: On average a fraction $1-\mu$ of the n mutated sites is not mutated, and a fraction μ of the $L-n$ non-mutated sites aquires a new mutation. Inserting Eq (A9) into Eq (A8), we finally obtain

$$\begin{aligned} \pi_1 &= \frac{1}{L} \sum_{n=0}^{k} \left[L\mu + (1-2\mu)n \right] q_n = \mu + \frac{1}{L} (1-2\mu) \sum_{n=0}^{k} n q_n = \mu + \frac{1}{L\bar{w}} (1-2\mu) \left(L\pi_1 - \sum_{n=k+1}^{L} n f_n \right) \\ &= \mu + \frac{\pi_1}{\bar{w}} (1-2\mu) - \frac{1}{L\bar{w}} (1-2\mu) \sum_{n=k+1}^{L} n f_n, \end{aligned} \tag{A10}$$

where we have used that

$$\sum_{n=0}^{L} n f_n = \sum_{n=0}^{L} n p_n = L\pi_1, \tag{A11}$$

because the allele frequency is not changed by recombination.

Up to now, everything is exact. It is a formidable, if not impossible, task to find an exact solution of Eq (A10), so we will solve the problem approximately for small μ. Since μ is small, it is plausible to assume that $\pi_1 \ll 1$ as well. Under this assumption, we can find an approximate expression for $\bar{w}$ as follows:

$$\begin{aligned}
\bar{w} &= \sum_{n=0}^{k} f_n = 1 - \sum_{n=k+1}^{L} f_n \approx 1 - \binom{L}{k+1}\pi_1^{k+1}(1-\pi_1)^{L-k-1} - \binom{L}{k+2}\pi_1^{k+2}(1-\pi_1)^{L-k-2} \\
&\approx 1 - \binom{L}{k+1}\pi_1^{k+1} + (L-k-1)\binom{L}{k+1}\pi_1^{k+2} - \binom{L}{k+2}\pi_1^{k+2} \\
&= 1 - \binom{L}{k+1}\pi_1^{k+1} + (k+1)\binom{L}{k+2}\pi_1^{k+2} \equiv 1 - C_1\pi_1^{k+1} + (k+1)C_2\pi_1^{k+2},
\end{aligned} \tag{A12}$$

where we have kept terms up to order π_1^{k+2}, $C_i = \binom{L}{k+i}$ $(i = 1, 2)$, and $1/j!$ should be interpreted as 0 if j is a negative integer. Note that the above formula is actually exact for $k \geq L - 2$.

Now we approximate Eq (A10) term by term. First, we get

$$\frac{\pi_1}{\bar{w}}(1-2\mu) \approx \pi_1\left[1 + C_1\pi_1^{k+1} - (k+1)C_2\pi_1^{k+2}\right](1-2\mu) \approx \pi_1 - 2\mu\pi_1 + C_1\pi_1^{k+2}, \tag{A13}$$

where we have kept terms up to π_1^{k+2} and $\mu\pi_1$. Second, we get

$$\begin{aligned}
\frac{1-2\mu}{\bar{w}}\sum_{n=k+1}^{L} n f_n &\approx \left[1 + C_1\pi_1^{k+1} - (k+1)C_2\pi_1^{k+2}\right](1-2\mu)\left[(k+1)C_1\pi_1^{k+1}(1-(L-k-1)\pi_1) + (k+2)C_2\pi_1^{k+2}\right] \\
&\approx (k+1)C_1\pi_1^{k+1} - k\frac{L!}{(k+1)!(L-k-2)!}\pi_1^{k+2}.
\end{aligned} \tag{A14}$$

Accordingly, we arrive at

$$\begin{aligned}
\pi_1 &\approx \mu + \pi_1 - 2\mu\pi_1 + C_1\pi_1^{k+2} - \frac{k+1}{L}C_1\pi_1^{k+1} + k\frac{(L-1)!}{(k+1)!(L-k-2)!}\pi_1^{k+2} \\
&= \pi_1 + \mu - 2\mu\pi_1 - \binom{L-1}{k}\pi_1^{k+1} + (L-k)\binom{L-1}{k}\pi_1^{k+2},
\end{aligned} \tag{A15}$$

that is,

$$\mu \approx B^{-(k+1)}\pi_1^{k+1} + 2\mu\pi_1 - (L-k)B^{-(k+1)}\pi_1^{k+2}, \tag{A16}$$

where $B = [k!(L-k-1)!/(L-1)!]^{1/(k+1)}$. Since the leading behavior of π_1 is $B\mu^{1/(k+1)}$, we set

$$\pi_1 = B\mu^{1/(k+1)}(1+g), \tag{A17}$$

where $g = o(1)$. Inserting Eq (A17) into Eq (A16) and expanding up to the leading order in g, we obtain

$$\begin{aligned}
\mu &\approx \mu(1+g)^{k+1} + 2B\mu^{(k+2)/(k+1)} - (L-k)B\mu^{(k+2)/(k+1)} \\
&\approx \mu + \mu(k+1)g + (2+k-L)B\mu\mu^{1/(k+1)},
\end{aligned} \tag{A18}$$

which yields

$$g \approx \frac{L-k-2}{k+1}B\mu^{1/(k+1)}. \tag{A19}$$

Therefore the mutational robustness becomes

$$\begin{aligned} m &= \sum_{n=0}^{k-1} f_n + \frac{k}{L} f_k = 1 - \sum_{n=k+2}^{L} f_n - \frac{L-k}{L} f_k - f_{k+1} \approx 1 - \frac{L-k}{L}\binom{L}{k}\left[\pi_1^k - (L-k)\pi_1^{k+1}\right] - \binom{L}{k+1}\pi_1^{k+1} \\ &= 1 - \frac{(L-1)!}{k!(L-k-1)!}\pi_1^k + \frac{(L-1)!}{(k+1)!(L-k-1)!}(kL-k^2-k)\pi_1^{k+1} \\ &= 1 - B^{-(k+1)}\pi_1^k + B^{-(k+1)}\pi_1^{k+1}\frac{kL-k^2-k}{k+1} \approx 1 + \mu\frac{kL-k^2-k}{k+1} - B^{-(k+1)}\pi_1^{k+1}\pi_1^{-1} \\ &\approx 1 + \mu\frac{kL-k^2-k}{k+1} - \mu[1+(k+1)g](1-g)\mu^{-1/(k+1)}B^{-1} \approx 1 + \mu\frac{kL-k^2-k}{k+1} - \mu^{k/(1+k)}(1+kg)B^{-1} \\ &= 1 + \mu\frac{kL-k^2-k}{k+1} - \mu^{k/(1+k)}B^{-1} - \mu^{k/(1+k)}kgB^{-1} \approx 1 - \mu^{k/(1+k)}B^{-1} + \mu\frac{kL-k^2-k}{k+1} - \mu\frac{k(L-k-2)}{k+1} \\ &= 1 - \binom{L-1}{k}^{1/(k+1)}\mu^{k/(k+1)} + \mu\frac{k}{k+1}. \end{aligned} \tag{A20}$$

If $L \gg k$, m can be approximated as

$$m \approx 1 - (L\mu)^{k/(k+1)}(k!)^{-1/(k+1)} + \mu\frac{k}{k+1}. \tag{A21}$$

IV. MUTATIONAL ROBUSTNESS ON THE MESA LANDSCAPE IN THE ABSENCE OF RECOMBINATION

Here we calculate the mutational robustness for the mesa landscape in the absence of recombination and under the assumption that the mutation rate is small. Here this is taken to imply that the genome-wide mutation rate $U \equiv L\mu \ll 1$, which implies that multiple mutations are negligible in the mutation step. Using the same notation as before, the lumped equilibrium frequencies after mutation f_n and after selection q_n then satisfy the relations

$$\bar{w} = \sum_{n=0}^{k} f_n, \quad q_n = \frac{f_n}{\bar{w}}, \quad f_n = (1-U)q_n + U\frac{L-n+1}{L}q_{n-1} + U\frac{n+1}{L}q_{n+1}, \tag{A22}$$

where $q_n = 0$ for $n > k$ and $q_{-1} = 0$. Since $f_n = 0$ for $n > k+1$, we have

$$\bar{w} = 1 - f_{k+1} = 1 - U\frac{L-k}{L}q_k. \tag{A23}$$

This yields a closed set of equations for the q_n, which reads

$$q_n\left[1 - U\left(1 - \frac{k}{L}\right)q_k\right] = (1-U)q_n + U\frac{L-n+1}{L}q_{n-1} + U\frac{n+1}{L}q_{n+1} \tag{A24}$$

or

$$\frac{n+1}{L}q_{n+1} = M_k q_n - \frac{L-n+1}{L}q_{n-1}, \tag{A25}$$

with

$$M_k = 1 - \frac{L-k}{L}q_k = \sum_{n=0}^{k-1} q_n + \frac{k}{L}q_k = 1 - \left(1 - \frac{k}{L}\right)q_k. \tag{A26}$$

Note that M_k can be interpreted as mutational robustness measured before mutation and after selection. Interestingly, q_n's do not depend on U if no multiple mutations are allowed. Since mutational robustness after mutation is given by

$$\begin{aligned} m =& \sum_{n=0}^{k-1} f_n + \frac{k}{L}f_k = 1 - f_{k+1} - \frac{L-k}{L}f_k = \bar{w}\left(1 - \frac{L-k}{K}q_k\right) = M_k\bar{w} \\ =& M_k - UM_k(1-M_k) = M_k(1-U) + UM_k^2, \end{aligned} \tag{A27}$$

it is sufficient to find M_k.

Defining $\xi_n \equiv (2L)^{n/2}\binom{L}{n}^{-1} q_n/q_0$ and $y \equiv M_k\sqrt{L/2}$, we obtain from (A25)

$$\left(1-\frac{n}{L}\right)\xi_{n+1} = 2y\xi_n - 2n\xi_{n-1}. \tag{A28}$$

We write down the first few terms for later purposes,

$$\xi_0 = 1, \quad \xi_1 = 2y, \quad \xi_2 = (4y^2-2)\frac{L}{L-1}. \tag{A29}$$

If $n/L \ll 1$, Eq (A28) is approximated as

$$\xi_{n+1} = 2y\xi_n - 2n\xi_{n-1}, \tag{A30}$$

which is the recursion relation of the Hermite polynomials $H_n(y)$. Since $\xi_0 = H_0$ and $\xi_1 = H_1$ for any L, we find the approximate solution for ξ_n as $\xi_n = H_n(y)$ for $n \ll L$. If $k/L \ll 1$, the Hermite polynomial becomes an accurate solution for all n. Since $\xi_{k+1} = 0$ by definition and $\xi_n > 0$ for $n \le k$, y should be the largest solution of the equation

$$H_{k+1}(y) = 0. \tag{A31}$$

If we denote the largest zero of Eq (A31) by $\sqrt{y_k/2}$, we thus conclude

$$M_k = \sqrt{\frac{y_k}{L}} + o(L^{-1/2}). \tag{A32}$$

The first few zeros are given by

$$y_1 = 1, \quad y_2 = 3, \quad y_3 = 3+\sqrt{6}, \quad y_4 = 5+\sqrt{10}. \tag{A33}$$

The approximation can be compared to the exact solutions for M_k which have been obtained up to $k = 4$ by solving Eq (A22),

$$\begin{aligned}
M_1 &= \frac{1}{\sqrt{L}}, \quad M_2 = \frac{\sqrt{3L-2}}{L} = \sqrt{\frac{3}{L}} + O(L^{-3/2}),\\
M_3 &= \frac{\sqrt{3L-4+\sqrt{6L^2-3L+16}}}{L} = \left(\frac{3+\sqrt{6}}{L}\right)^{1/2} + O(L^{-3/2}),\\
M_4 &= \frac{\sqrt{5L-10+\sqrt{10L^2-5L+76}}}{L} = \left(\frac{5+\sqrt{10}}{L}\right)^{1/2} + O(L^{-3/2}),
\end{aligned}$$

which are indeed consistent with Eq (A32) and the first four y_k's in Eq (A33). Using Eq (A27) the robustness after mutation is then given by

$$m \approx \sqrt{\frac{y_k}{L}}(1-U) + U\frac{y_k}{L}. \tag{A34}$$

Now we consider the case of large k. If we still assume $1 \ll k \ll L$, the above approximation is valid. Since the asymptotic behavior of the largest zero of $H_n(x)$ is $\sim\sqrt{2n+1}$ [4, p. 132], we find $y_k \sim 4k$, which gives

$$m \approx 2\sqrt{\frac{k}{L}}(1-U). \tag{A35}$$

The approximation leading to Eq (A32) is however not valid if k/L remains finite as $L \to \infty$. To treat this problem, we may refer to previous work on the mesa landscape [5] that makes use of a maximum principle for permutation-invariant fitness landscapes [6]. This principle states that the stationary population mean fitness $\bar{w}$ is given by

$$\bar{w} = \max_{x\in[0,1]}\left\{\omega(x) - U\left[1-2\sqrt{x(1-x)}\right]\right\}, \tag{A36}$$

where $\omega(x) = \lim_{L\to\infty} w_{xL}$ is the limiting value of the fitness of a genotype with $n = xL$ mutations. To account for the fact that genotypes with more than k mutation are lethal, the fitness function has to be taken to be $\omega(x) = 1$ if $x \leq x_0 \equiv k/L$ and $\omega(x) = -\infty$ if $x > x_0$, which is slighlty different from the setting of Ref. [5]. Nevertheless the result for the stationary fitness is the same,

$$\bar{w} = \begin{cases} 1 - U\left[1 - 2\sqrt{x_0(1-x_0)}\right], & \text{if } x_0 < 1/2, \\ 1, & \text{if } x_0 \geq 1/2. \end{cases} \tag{A37}$$

Combining Eqs (A23) and (A26) we see that $M_k = 1 - U^{-1}(1 - \bar{w})$, and therefore

$$M_k = \begin{cases} 2\sqrt{x_0(1-x_0)}, & \text{if } x_0 < 1/2, \\ 1, & \text{if } x_0 \geq 1/2. \end{cases} \tag{A38}$$

Note that the leading behavior of M_k for small x_0 is the same as the Hermite polynomial solution Eq (A35).

V. RECOMBINATION WEIGHT ON THE MESA LANDSCAPE WITH UNIFORM CROSSOVER

In order to efficiently compute the recombination weight for uniform crossover on the mesa landscape, one has to exploit the permutation invariance of the landscape. In the following we denote the recombination weight λ_σ of genotype σ as $\lambda(L, a, k, r)$, since it is fully defined by the sequence length L, the mesa width k, the Hamming distance $a \equiv d_\sigma$ to the wild type and the recombination rate r. To start with we first note that the Hamming distances between an offspring genotype σ and its parent genotypes κ, τ also determine the Hamming distance between both parent genotypes through the relation [7]

$$d(\sigma, \kappa) + d(\sigma, \tau) = d(\kappa, \tau). \tag{A39}$$

For the following it is convenient to introduce the variables i and j which represent the Hamming distance $d(\sigma, \kappa)$ and $d(\sigma, \tau)$, respectively. Eq (A39) is useful since the Hamming distance $i + j$ between the parent genotypes determines their number of possible distinct offspring genotypes through recombination. Hence the probability that the offspring genotype σ is generated by two genotypes at distance i and j is given by

$$\frac{1}{2^{i+j}} r + \frac{1-r}{2} \left(\delta_{i0} + \delta_{j0}\right), \tag{A40}$$

where the second term includes the possibility of no recombination for which at least one of the parent genotypes needs to be the same as the offspring genotype, see also Eq (6) of the main text. Next we consider the number of genotypes at Hamming distance i and j as well as their respective fitness. The number of potential parent genotypes at Hamming distance i is given by $\binom{L}{i}$ which can be rewritten as

$$\binom{L}{i} = \sum_{x=0}^{i} \binom{a}{x}\binom{L-a}{i-x} = \sum_{x=\max(0,i+a-L)}^{\min(i,a)} \binom{a}{x}\binom{L-a}{i-x}. \tag{A41}$$

We make use of the fact that in order to create a genotype at distance i, we can mutate x out of a 1-alleles and $i-x$ out of $L-a$ 0-alleles from the offspring genotype for which the number of arrangements is given by a binomial coefficient. Since the sum might contain zero terms we can restrict the summation range further. Through this expression it is possible to relate to each genotype its fitness which is given by

$$w(k, (a-x) + (i-x)) = \theta(k - (a-x) - (i-x)), \tag{A42}$$

where $(a-x) + (i-x)$ denotes the number of 1-alleles in the parent genotype and θ is the Heaviside step function with $\theta(0) = 1$. After choosing a parent genotype at distance i the remaining number of suitable parent genotypes at Hamming distance j is thus given by

$$\sum_{y=0}^{j} \binom{a-x}{y}\binom{L-a-(i-x)}{j-y} = \sum_{y=\max(0,j+a-L+i-x)}^{\min(j,a-x)} \binom{a-x}{y}\binom{L-a-(i-x)}{j-y}, \tag{A43}$$

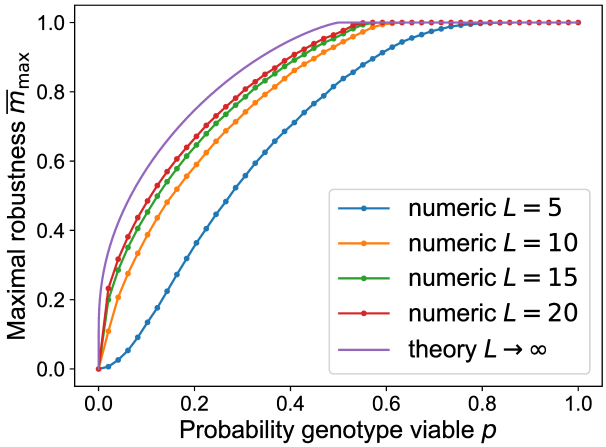

Fig A2. **Maximal degree of the viable network in the percolation landscape.** The figure shows numerical results for the expected maximal degree of a viable genotype in percolation landscapes of different size L. For $L \to \infty$ the results converge to the solution z^* of the equation $s_p(z^*) = \ln 2$, where $s_p(z)$ is given in Eq (A47).

with fitness

$$w(k,(a-y)+(j-y)) = \theta(k-(a-y)-(j-y)). \tag{A44}$$

Since the allele of at least one parent genotype needs to coincide with the allele of the offspring genotype, the number of 1-alleles that one can mutate is reduced by x. The same logic applies to the number of 0-alleles one can mutate, which is reduced by $i-x$. Finally in order to compute the recombination weight we have to sum over all possible combinations of distances (i,j) which are restricted due to Eq (A39) to be in the range $0 \le i+j \le L$. For efficient computation one should avoid double counting of ordered pairs (i,j) and (j,i) which yield the same contribution to the recombination weight. Combining these considerations leads to a more efficient expression for the recombination weight on the mesa landscape,

$$\begin{aligned}\lambda(L,k,a,r) =& \frac{1}{2^L}\sum_{i=0}^{\lfloor L/2\rfloor}\sum_{j=i}^{L-i}\sum_{x=\max(0,i+a-L)}^{\min(i,a)}\binom{a}{x}\binom{L-x}{i-x}\theta(k+2x-a-i)\times \\ &\sum_{y=\max(0,j+a-L+i-x)}^{\min(j,a-x)}\binom{a-x}{y}\binom{L+x-a-i}{j-y}\theta(k+2y-a-j)\left[\frac{r}{2^{i+j}}(2-\delta_{ij})+(1-r)\delta_{i0}\right],\end{aligned} \tag{A45}$$

where $\lfloor z \rfloor$ stands for the greatest integer that is less than or equal to z. As explained in the main text $\lambda(L,k,a,r)$ depends linearly on the recombination rate r. We use Eq (A45) for numerical calculations.

VI. MAXIMAL ROBUSTNESS IN THE PERCOLATION LANDSCAPE

To estimate the number of viable neighbors of a genotype in the percolation landscape in the limit of large L, we start from the observation that the expected number of genotypes with k viable neighbors is

$$\mathbb{E}(n_k) = 2^L\binom{L}{k}p^k(1-p)^{L-k} \sim \exp[L(\ln 2 - s_p(k/L))], \tag{A46}$$

where

$$s_p(z) = -z\ln(p) - (1-z)\ln(1-p) + z\ln(z) + (1-z)\ln(1-z) \tag{A47}$$

is the large deviation function of the binomial distribution [8]. For a given p, there is thus a value $z^*(p)$ defined by $s_p(z^*) = \ln 2$ such that, for $L \to \infty$, $\mathbb{E}(n_k) \to \infty$ if $k < z^*L$ and $\mathbb{E}(n_k) \to 0$ if $k > z^*L$. Using standard probabilistic

arguments this can be shown to imply that genotypes with k neighbors are present (absent) with probability 1 if $k < z^*L$ ($k > z^*L$), respectively. Thus the expected maximal robustness is $\overline{m}_{\max} = z^*$. Since $s_p(1) = \ln(1/p)$, $z^* = 1$ for $p \geq \frac{1}{2}$. Fig A2 compares the asymptotic behavior of $\overline{m}_{\max}$ for $L \to \infty$ to simulation results at finite L.

[1] URL `http://networkx.readthedocs.io/en/networkx-1.11/`.
[2] M. A. Nowak, M. C. Boerlijst, J. Cooke, and J. Maynard Smith, Nature **388**, 167 (1997).
[3] R. A. Neher, B. I. Shraiman, and D. S. Fisher, Genetics **184**, 467 (2010).
[4] G. Szegö, *Orthogonal polynomials* (American Mathematical Society, Providence, Rhode Island, 1975), 4th ed.
[5] A. Wolff and J. Krug, Phys. Biol. **6**, 036007 (2009).
[6] J. Hermisson, O. Redner, H. Wagner, and E. Baake, Theor. Pop. Biol. **62**, 9 (2002).
[7] M. C. Boerlijst, S. Bonhoeffer, and M. A. Nowak, Proc. Biol. Sci. **263**, 1577 (1996).
[8] D. Sornette, *Critical Phenomena in Natural Sciences* (Springer, Berlin, 2000).

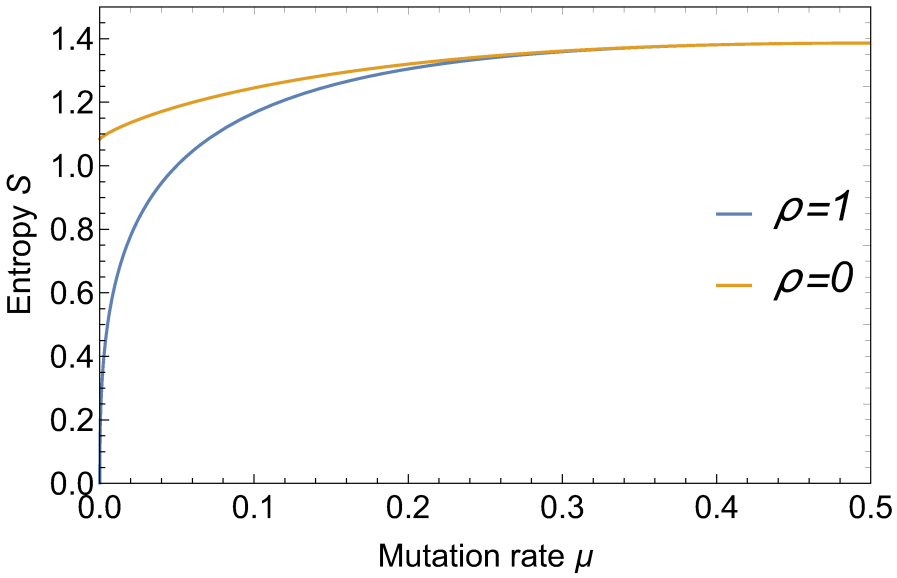

FIG. S1. **Population heterogeneity decreases with increasing recombination rate.** The figure shows the entropy of the genotype frequency distribution in the two-locus model defined as $S = -\sum_\sigma f_\sigma^* \ln(f_\sigma^*)$. For small mutation rates the strongly recombining population primarily consists of a single genotype, which implies that $S \to 0$.

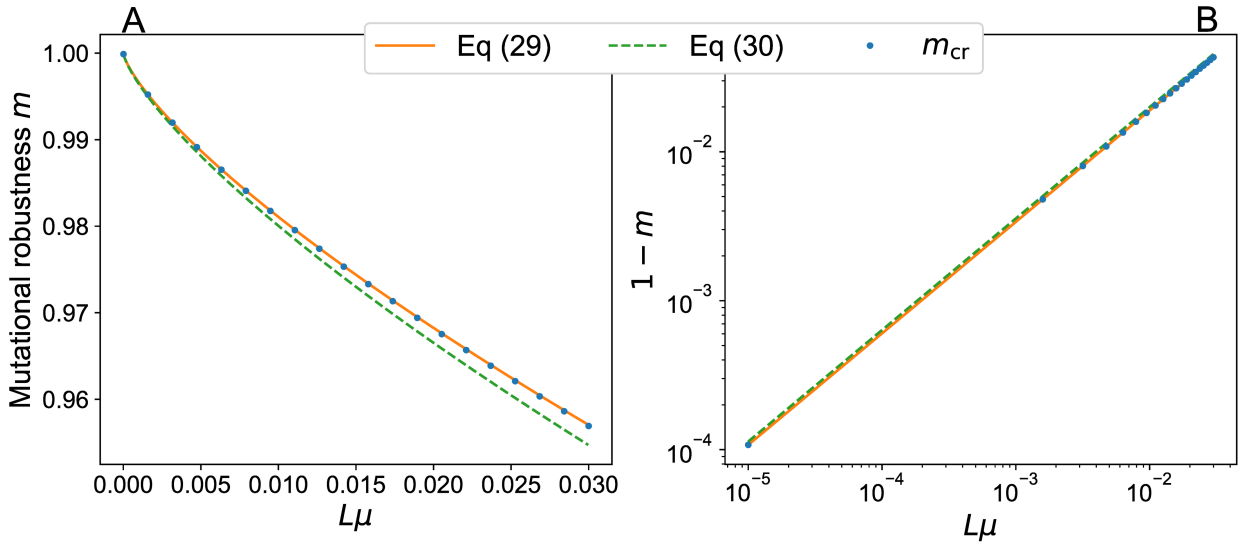

FIG. S2. **Mutational robustness for the mesa landscape with communal recombination.** The figure compares the analytic approximations in Eqs (29) and (30) to the numerical solution of the stationary genotype frequency distribution for the communal recombination scheme. The two panels show the mutational robustness as a function of the genome-wide mutation rate in linear (A) and double-logarithmic (B) scales, respectively. The parameters of the mesa landscape are $L = 30$ and $k = 3$.

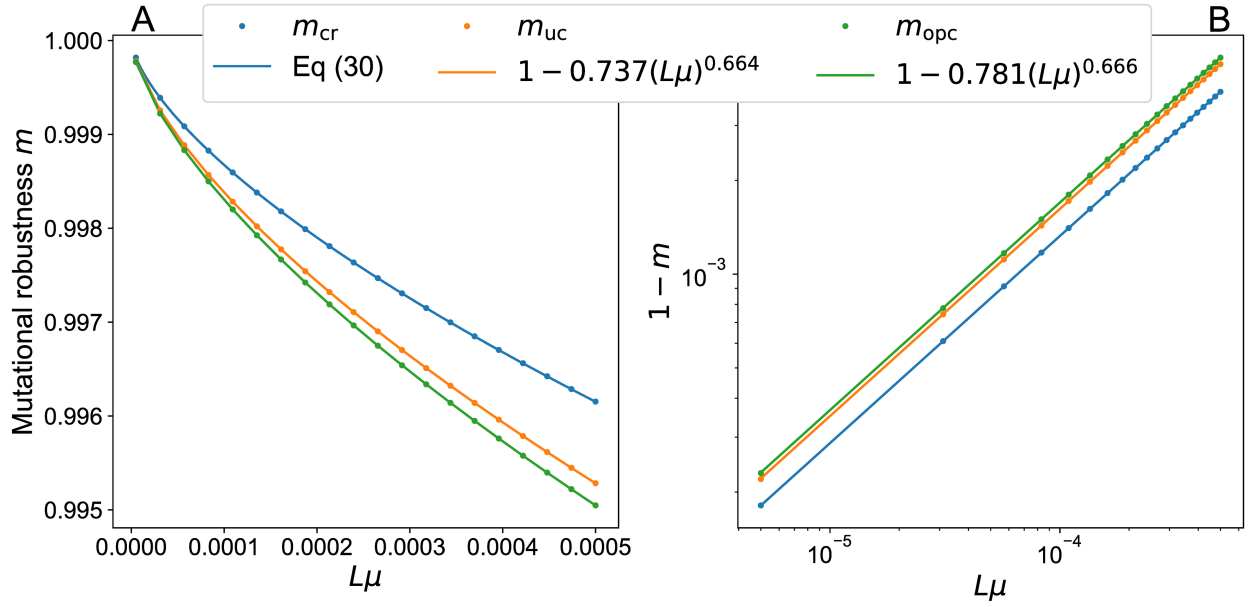

FIG. S3. **Mutational robustness in a mesa landscape with different recombination schemes.** The figure compares the analytic results for communal recombination ($m_{\rm cr}$) with numerical data obtained using uniform crossover ($m_{\rm uc}$) and one-point crossover ($m_{\rm opc}$) at $r = 1$. The landscape parameters are $L = 5$, $k = 2$ and robustness is plotted as a function of the genome-wide mutation rate $L\mu$. (A) Mutational robustness on linear scales. (B) Double-logarithmic plot of $1 - m$ vs. $L\mu$, illustrating the power-law behavior $1 - m \sim (L\mu)^b$ with the exponent $b = k/(k + 1) = 2/3$ predicted by the analysis of the communal recombination model.

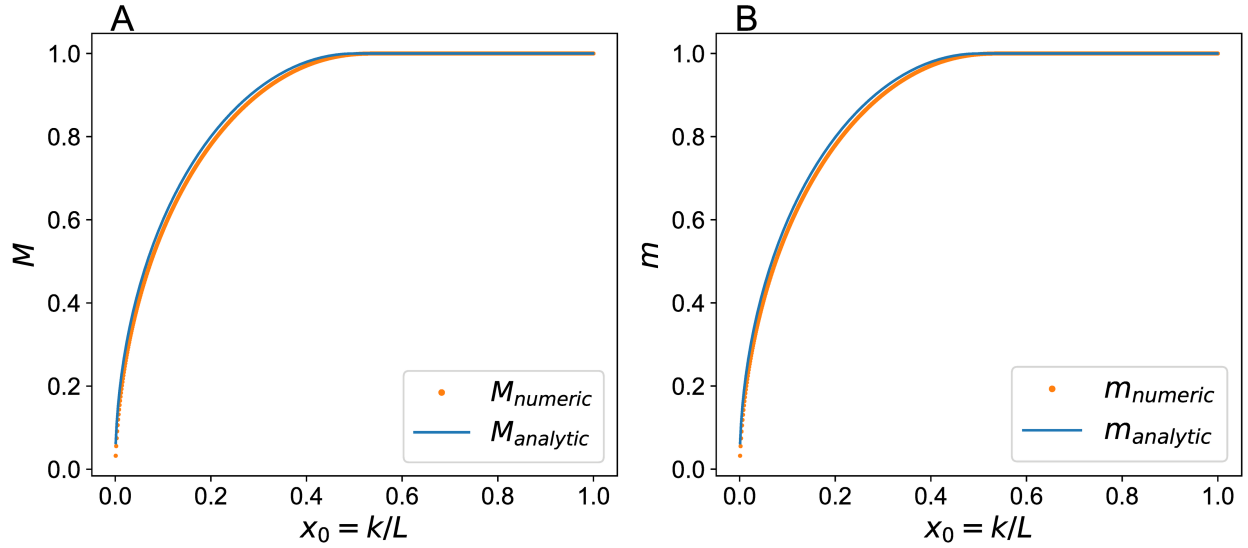

FIG. S4. **Mutational robustness for the mesa landscape in the absence of recombination.** The figure compares the analytic predictions in Eqs (35) and (36) to the numerical solution for the genotype frequency distribution in the absence of recombination. The two panels show the mutational robustness (A) after selection and (B) after mutation as a function of the scaled mesa width $x_0 = k/L$ for $L = 1000$ and $U = 0.01$.

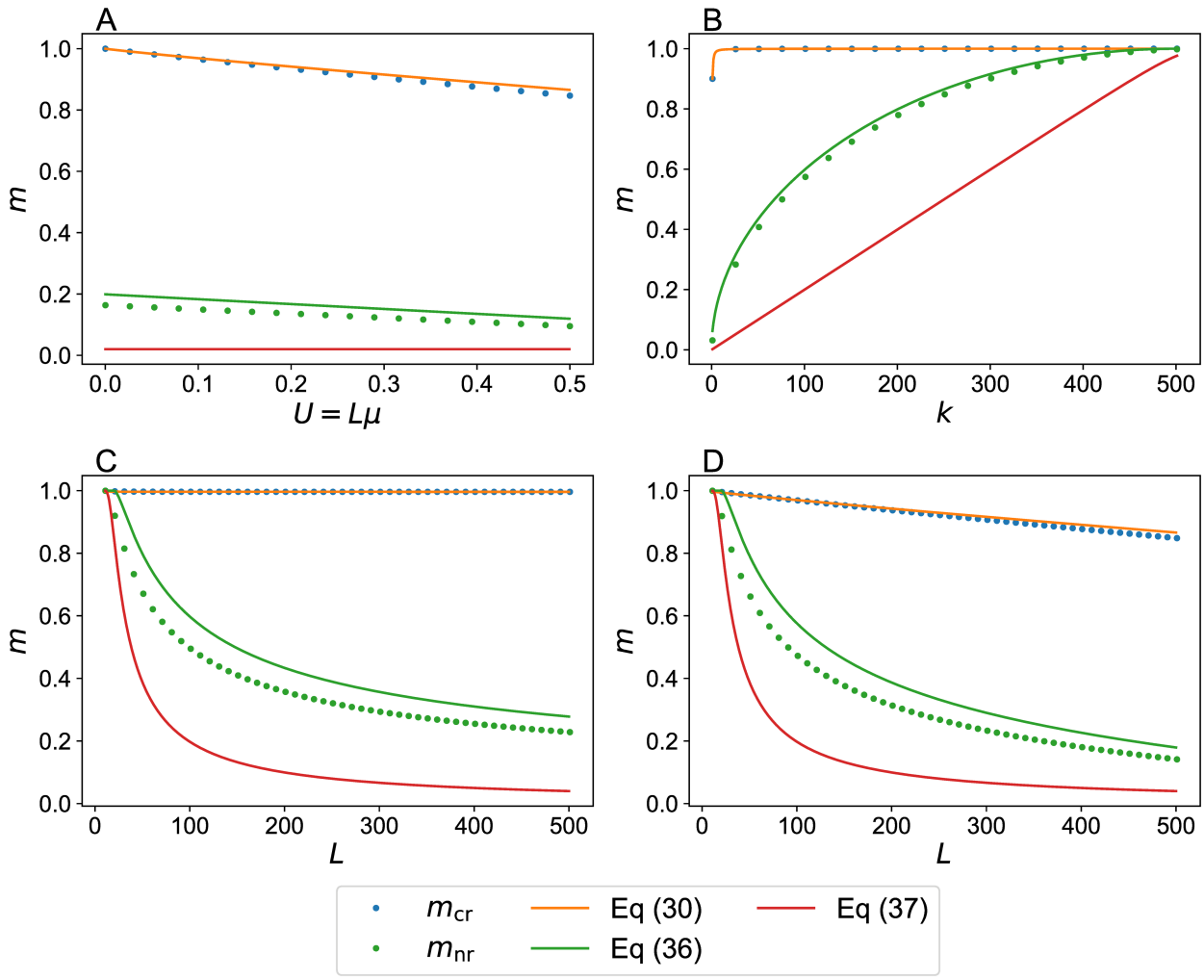

FIG. S5. **Mutational robustness in mesa landscapes with and without recombination**. Numerical results for communal recombination (m_{cr}) and no recombination (m_{nr}) are shown as dots. The mutational robustness m_0 of a uniformly distributed population, given by Eq (37), as well as the analytic expressions Eqs (30) and (36) are depicted as lines. (A) Robustness as a function of mutation rate $U = L\mu$ for a landscape with $L = 1000$ and $k = 10$. (B) Robustness as a function of mesa width k at fixed $L = 1000$ and $U = L\mu = 0.01$. (C) Robustness as a function of genome length L at fixed $k = 10$ and $U = 0.01$. (D) Robustness as a function of genome length L at fixed $k = 10$ and $\mu = 0.001$.

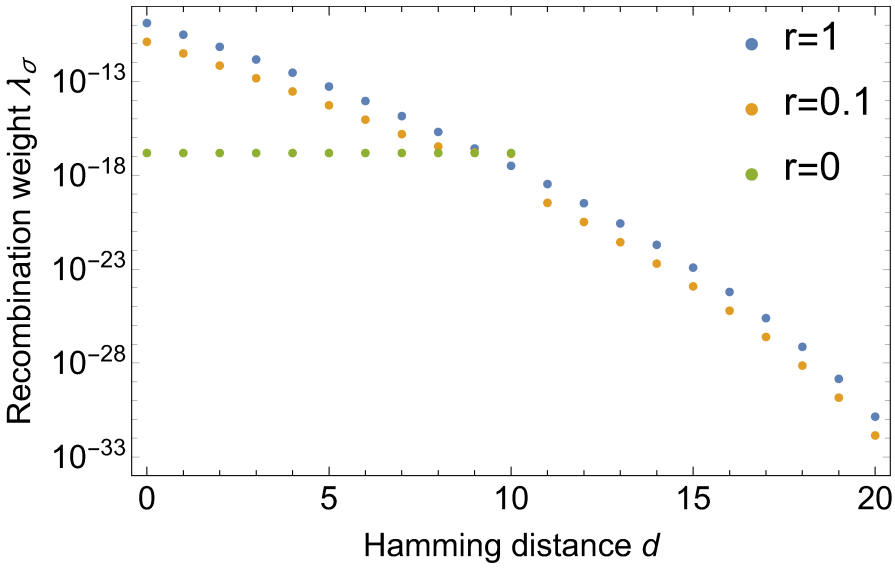

FIG. S6. **Recombination weight in a mesa landscape.** The parameters of the mesa landscape are $L = 100$ and $k = 10$. For $r = 0$ the recombination weight is directly proportional to the fitness and hence equal for all viable genotypes. Already small rates of recombination are sufficient to redistribute the recombination weight such that the weight of genotypes with small Hamming distance is strongly enhanced. Beyond $d = 20$ the recombination weight is identically zero, since the recombinant of two viable genotypes cannot carry more than $2k$ mutations.

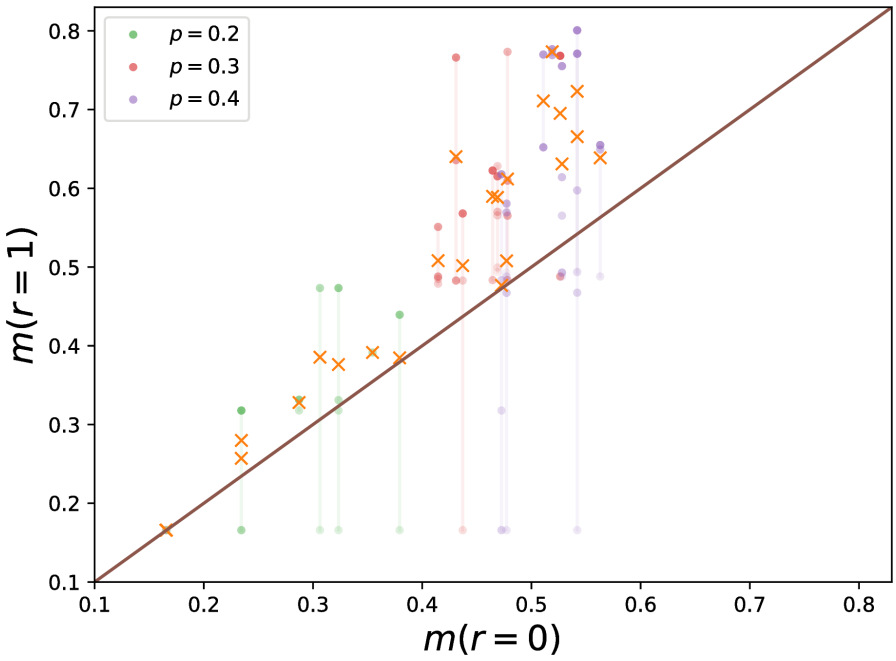

FIG. S7. **Mutational robustness for different stationary states within a percolation landscape**. The figure compares the mutational robustness of non-recombining ($r = 0$) and recombining ($r = 1$) populations on individual realizations of the percolation model with $L = 6$ and three values of p. In order to obtain different stationary states we used localized initial population distributions of the form $f_\tau(0) = \delta_{\tau\sigma}$ for all genotypes with mutational robustness $m_\sigma \neq 0$ and propagated them until stationarity. Since the stationary populations are usually highly concentrated for large r and small μ, this is a natural choice in order to access all stationary states. Each data point represents the robustness of the recombining population $m(r = 1)$ for a particular stationary state. Data points within the same landscape are plotted above the corresponding unique robustness of the non-recombining population $m(r = 0)$ and connected by a vertical line. The orange crosses show the average over all initial conditions.

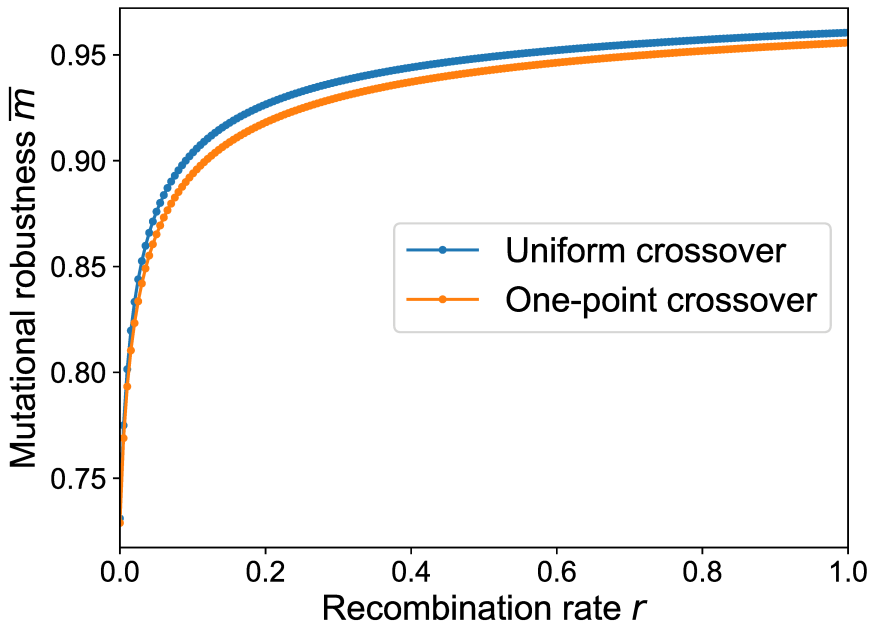

FIG. S8. **Average mutational robustness in the sea-cliff landscape as a function of recombination rate.** Mutational robustness is computed for 200 randomly generated sea-cliff landscapes with parameters $L = 6$, $d_< = 1$ and $d_> = 5$, and the results are averaged to obtain $\overline{m}(r)$. The mutation rate is $\mu = 0.001$.

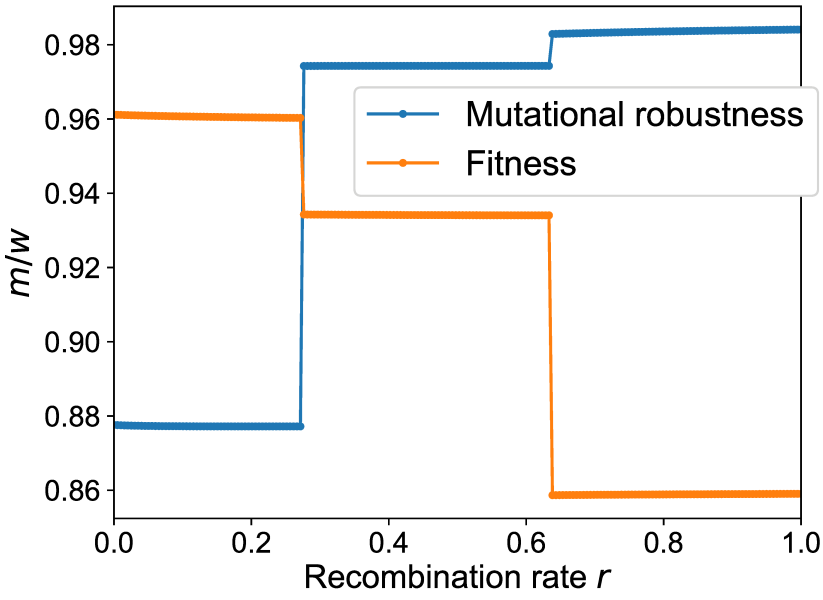

FIG. S9. **Mutational robustness and average fitness in the empirical *A. niger* fitness landscape.**. The mutational robustness and the population-averaged fitness in the stationary state were computed as a function of recombination rate by evolving the population from a uniform initial genotype distribution at mutation rate $\mu = 0.005$. Jumps mark changes in the most populated genotype.

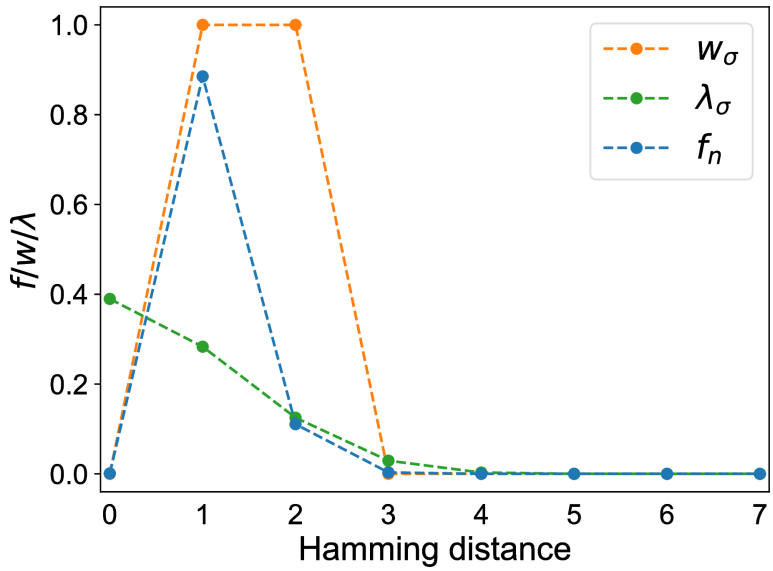

FIG. S10. **Recombination on an atoll landscape**. This landscape is similar to the mesa landscape but includes an inner critical radius within which genotypes are lethal. In this example the inner radius is chosen to be 1 such that only the wild type is lethal. The outer radius is 2 and the sequence length is $L = 7$. The recombination rate is $r = 1$ and the mutation rate is $\mu = 0.001$. The frequencies f_n of the stationary state at the same Hamming distance n are lumped together. The population is concentrated at distance 1 which is most robust since only one point mutation is lethal, but the recombination center coincides with the lethal wild type. This example shows that the correlation between recombination weight and mutational robustness depends on the topology of the neutral network.

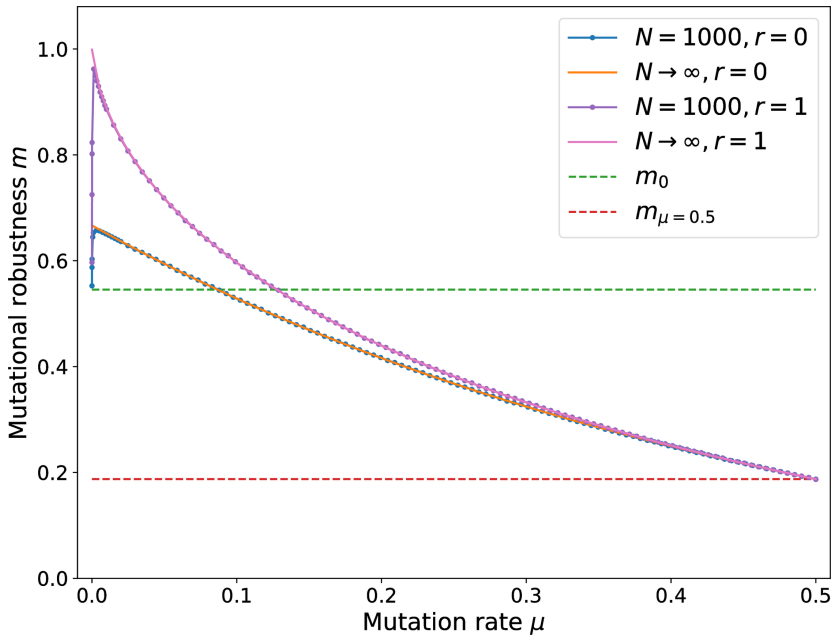

FIG. S11. **Finite population size effects**. The figure shows the mutational robustness in a mesa landscape with parameter $L = 6, k = 2$ as a function of mutation rate. The finite population results were obtained using Wright-Fisher dynamics for $N = 1000$ individuals. For small mutation rates such that $N\mu L \ll 1$ the monomorphic population performs a random walk among viable genotypes, which leads to the uniform mutational robustness m_0 given by Eq (37) (green dashed line). In this regime recombination cannot have any effect. For $N\mu L > 1$ the robustness rises sharply to the value predicted by the infinite population approach. At the maximal mutation rate $\mu = 0.5$ the population is uniformly distributed among all (lethal or viable) genotypes after the mutation step and recombination has again no effect.

3 Recombination in finite populations - 2nd manuscript

Alexander Klug and Joachim Krug.
"Effect of recombination on the evolvability, genetic diversity and mutational robustness of neutrally evolving finite populations."

Status: in preparation

Effects of recombination on the evolvability, genetic diversity and mutational robustness of neutrally evolving populations

Alexander Klug and Joachim Krug[1]
Institute for Biological Physics, University of Cologne, Cologne, Germany

ABSTRACT

Many effects attributed to recombination have been invoked to explain the advantage of sex. The most prominent arguments focus on either evolvability, genetic diversity, or mutational robustness to justify why the benefit of recombination overcomes its costs, with partially contradicting results. As a consequence, understanding which aspects of recombination are most important in a given situation remains an open problem for theoretical and experimental research. In this study, we focus on finite populations evolving on neutral networks, which already display remarkably complex behavior. We aim to provide a comprehensive overview of the effects of recombination by jointly considering different measures of evolvability, genetic diversity, and mutational robustness over a broad parameter range, such that many evolutionary regimes are covered. We find that several of these measures vary non-monotonically with the rates of mutation and recombination. Moreover, the presence of lethal genotypes that introduce inhomogeneities in the network of viable states qualitatively alters the effects of recombination. We conclude that conflicting trends induced by recombination can be explained by an emerging trade-off between evolvability and genetic diversity on the one hand, and mutational robustness and fitness on the other. Finally, we discuss how different implementations of the recombination scheme in theoretical models can affect the observed dependence on recombination rate through a coupling between recombination and genetic drift.

The neutral theory of evolution assumes that mutations have either no selective effect or are highly deleterious. This approximation of the distribution of fitness effects was suggested by Kimura based on observations of surprisingly high substitution rates in the amino acid sequence of certain proteins, although their function remained essentially unchanged (Kimura 1968, 1983). Today it is understood that the abundant neutrality in molecular evolution can arise through a wide range of mechanisms. Large portions of the genome are non-coding, allowing mutations to accumulate freely (Nobrega *et al.* 2004). However, also in the coding regions, neutrality is prevalent due to degeneracies in the genotype–phenotype mapping at multiple levels between the blueprint, the DNA, and the final functional structure, which can be a protein, a cell, or an entire organism (Manrubia *et al.* 2021). For example, on the scale of proteins, the degeneracies arise through synonymous mutations and through many different amino acid chains that fold to the same structure (Guo *et al.* 2004; Bloom *et al.* 2005). On the scale of cells, neutrality is observed in regulatory gene networks (Azevedo *et al.* 2006; Ciliberti *et al.* 2007) and metabolic reaction networks (Rodrigues and Wagner 2009). Moreover, recent microbial evolution experiments (Johnson *et al.* 2019) and theory (Reddy and Desai 2021) indicate that populations consistently adapt to regions of the genotype space where diminishing-returns and increasing-costs epistasis are common, which implies that the beneficial effects of mutations are almost neutral, whereas deleterious mutations typically have large negative selection coefficients. Apart from truly neutral mutations, small effect mutations can be effectively neutral if the absolute magnitude of the selection coefficient is smaller than the reciprocal of the population size (Ohta 2002). Therefore neutrality is particularly important for small populations.

[1]Corresponding author: Institute for Biological Physics, University of Cologne, Zülpicher Str. 77, D-50937 Köln, Germany. E-mail: jkrug@uni-koeln.de

The assumption of a binary distribution of fitness effects, where mutations are either selectively neutral or highly deleterious, can be conceptualized as a flat fitness landscape with holes (Gavrilets 1997, 2004) or as a neutral network with varying node degrees (van Nimwegen *et al.* 1999; Wilke 2001). With potentially many loci at which mutations can occur, of which at least a few are selectively neutral, large clusters of viable genotypes connected by point mutations form. These clusters may span the entire sequence space and thus populations can evolve continuously without being trapped at a fitness peak (Gavrilets 2004). In this way large neutral networks are argued to increase evolvability, since populations are able to explore large parts of genotype space leading to ever fitter genotypes (Wagner 2005).

However, for a complete description of evolution on neutral fitness landscapes, also the population dynamics has to be specified. A commonly used simplification is to consider the population as a point on the fitness landscape, implying that only a single genotype is present at a time. Neutral evolution then proceeds as a simple random walk on the neutral network through a sequence of fixation events (Maynard Smith 1970; Gavrilets 1997). This scenario applies in the weak mutation regime where the mutation supply is low (de Visser and Krug 2014). At the opposite end of the spectrum of evolutionary dynamics, quasispecies theory considers populations as continuously distributed clouds of genotypes in sequence space (Jain and Krug 2007; Domingo and Schuster 2016). An important difference compared to the weak mutation regime is that while in a simple random walk, all viable connected genotypes have the same probability of being currently occupied by the population (Hughes 1996), in the quasispecies regime mutationally robust genotypes, i.e. genotypes with an above-average number of viable point mutations, are preferentially occupied (van Nimwegen *et al.* 1999; Bornberg-Bauer and Chan 1999). This effect is strongly enhanced in recombining populations (Szöllősi and Derényi 2008; Klug *et al.* 2019; Singhal *et al.* 2019).

Quasispecies theory is deterministic and, strictly speaking, only applies to infinitely large populations (Wilke 2005). Therefore genetic drift is absent, and all genotypes have a frequency greater than zero by definition. Moreover, in this limit, the population reaches a stationary state determined by a selection-mutation(-recombination) balance, where the frequencies of all genotypes become constant in time. Since the number of genotypes grows exponentially with the number of loci, a shortcoming of this approximation is that it quickly becomes unrealistic for large but finite populations and is only applicable for short sequences, where the population can cover all genotypes. In this case quasispecies theory can approximate finite populations quite well (van Nimwegen *et al.* 1999; Szöllősi and Derényi 2008).

The purpose of this article is to describe and understand neutrally evolving finite populations in large sequence spaces for which the deterministic quasispecies limit does not apply. Within this setting, we explore a broad parameter range, such that all possible evolutionary regimes are covered. In particular, we include recombination and study its effect across a wide range of recombination rates. We believe that such a comprehensive study, which to the best of our knowledge has not been performed previously, is essential for elucidating the conditions under which recombination carries a selective advantage (Weismann 1891; Muller 1932, 1964; Felsenstein 1974; Kondrashov 1988; Feldman *et al.* 1996; Burt 2000; de Visser and Elena 2007; Otto 2009). In previous work we argued that the universal and somewhat underappreciated effect of recombination on mutational robustness may play an important role in this context (Klug *et al.* 2019). Working in the deterministic limit of an infinitely large populations, we showed that mutational robustness increases monotonically with the recombination rate r, independent of model details, and that for low mutation rate μ and small r, mutational robustness grows linearly with r/μ. The deterministic limit allowed us to find precise analytical results, but many relevant questions cannot be addressed in this framework. Here we consider finite populations in large sequence spaces. We are interested in the evolvability of the population and ask how quickly new genotypes are discovered, and how many generations it takes to discover all viable genotypes. The discovery rate of new genotypes can be crucial, e.g., if through environmental perturbations like an immune response certain genotypes become fitter over time or if the majority of the neutral network loses fitness and an escape mutation needs to be found. While the discovery of new genotypes is therefore essential for long-term survival, the accumulation of lethal mutations entails the risk of extinction. This induces an evolutionary trend towards increasing mutational robustness, another measure we investigate.

In order explain how the discovery rate and the mutational robustness of the population change with the parameters, we consider different measures of genotype diversity, such as the mean Hamming distance, the number of segregating mutations and the number of distinct genotypes. Certain properties like the number of segregating mutations and mean Hamming distance are independent of recombination if all genotypes are viable, but become recombination dependent when some genotypes are lethal. Other properties like the number of distinct genotypes and the discovery rate grow monotonically with r if all genotypes are viable, while in the presence lethal genotypes, the dependence becomes non-monotonic. We also find that with recombination, the discovery rate can become non-monotonic in the mutation rate, such that higher mutation rates may lead to reduced evolvability. Furthermore, we discuss different implementations of recombination in the Wright-Fisher model. Depending on the model details, recombination can act as an additional source of genetic drift which matters in small populations or large sequence spaces.

Outline. In the first section Models and methods, we define the structure of the genotype space and the fitness landscape. We consider both finite and infinite-sites settings. We further define the population dynamics and the implementation of recombination. Next we introduce the relevant measures of diversity, robustness and evolvability and describe the visualization of our results. In the second section Results and analysis, we first give an overview of the evolutionary regimes on neutral networks. We then explain our results in the limit of infinite sequence spaces (infinite-sites model) and continue with the results for finite sequence spaces. At the end of this section we discuss aspects of the results that show a non-robust dependence on the implementation of recombination. The results are summarized and conclusions are presented in the last section Discussion. Due to the complexity of the problem, our work relies primarily on extensive numerical simulations, but analytic results are also presented when available.

Models and methods

Genotype space

We consider haploid genomes with L diallelic loci, which can be expressed as sequences

$$\sigma = (\sigma_1, \sigma_2, ..., \sigma_L) \tag{1}$$

of symbols drawn from a binary alphabet $\sigma_i = \{-1, 1\}$. This translates to a genotype space that has the properties of a hypercube $H_2^L = \{-1, 1\}^L$ of dimension L, where each of the 2^L vertices represents a genotype. Genotypes of vertices connected by an edge differ at a single locus and are therefore mutually reachable by a point mutation. The natural metric in this genotype space is the Hamming distance

$$d(\sigma^i, \sigma^j) = \sum_{k=1}^{L} (1 - \delta_{\sigma_k^i, \sigma_k^j}), \tag{2}$$

which quantifies the number of point mutations that separate two genotypes σ^i and σ^j. For our analyses we consider both the so-called infinite-sites model (*ism*) corresponding to the limit $L \to \infty$, and the finite-sites model (*fsm*) with finite L. The *ism* originally introduced by Kimura (1969) is easier to handle analytically, since back mutations do not occur and all mutations are novel. However, certain quantities of interest such as the mutational robustness, for which the number of viable point mutations needs to be computed, and the time until full discovery of all viable genotypes cannot be defined within the *ism*. We therefore consider both models and compare results.

Fitness landscape

In our simulations, we either assume that all genotypes are viable, or that a fraction $1 - p$ of genotypes is lethal. We show that the addition of lethal genotypes strongly alters the structure of the genotype cloud and in particular the effect of recombination. To be maximally agnostic about the distribution of lethal genotypes in sequence space, we assume that each genotype is viable with probability p and otherwise lethal. This kind of fitness landscape is known as a percolation landscape (Gavrilets and Gravner 1997).

In the case of the *fsm* we add the constraint that the resulting network of viable genotypes on the hypercube is connected, i.e. that between any two viable genotypes there is a path of viable point mutations. In our simulations this is achieved by discarding all percolation landscape realizations that do not satisfy this condition; Fig. S1 shows how the fraction of connected landscapes varies with p. The constraint is added in order to avoid situations in which the initial population is trapped in a disconnected cluster of viable genotypes. In the case of the *ism* there is no additional constraint and the fitness of a novel genotype is generated once it has been discovered by the population.

Besides containing only minimal assumptions, we also chose this landscape model because its random nature makes it rich in possible structures, in the sense that it can contain regions with many viable point mutations and regions where genotypes are more often lethal and populations must evolve along a narrow fitness ridge. Furthermore, the choice of this landscape has the benefit of only adding one more parameter p to our analysis. From the point of view of the neutral network, the parameter p determines the degree distribution.

Dynamics

To model the evolutionary forces of selection, mutation and recombination, we use individual-based Wright-Fisher models with discrete, non-overlapping generations and a constant population size N. For the implementation of selection and recombination, we found different computational schemes in the literature. Initial simulations showed that, whereas for large populations in the *fsm* the models become indistinguishable, the model details become apparent for small populations in the *fsm* and at arbitrary population sizes in the *ism*. In the following, the different schemes are explained. For the main part of the article, we show results for only one of the models, but mention important differences when they exist, and discuss the differences in detail in the subsection Recombination-induced genetic drift.

Figure 1 illustrates the course of one generation for three different selection-recombination schemes. In the main text, we use the model that we refer to as *concurrent recombination*. In this model, selection and recombination occur in a single step, whereas in the other two models referred to as *successive recombination* schemes these processes require two separate steps. Despite these differences, the models also share similarities, which we explain first. All models have in common that an individual j has two ancestors in the previous generation with a probability equal to the recombination rate r. If recombination occurs we employ a uniform crossover scheme, which means that at each locus, the allele from one of the two ancestors k,l is chosen with equal probability,

$$R: \ \sigma_i^j \mapsto \begin{cases} \sigma_i^k & \text{with prob. } 1/2, \\ \sigma_i^l & \text{with prob. } 1/2, \end{cases} \quad \forall\ i. \tag{3}$$

Furthermore, the mutation step always occurs last, separate from the two other processes. During the mutation step, each locus of each individual mutates with probability μ to the opposite allele,

$$M: \ \sigma_i^j \mapsto \begin{cases} \sigma_i^j & \text{with prob. } 1 - \mu \\ -\sigma_i^j & \text{with prob. } \mu, \end{cases} \quad \forall\ i, j. \tag{4}$$

In the *ism* we take the joint limits $L \to \infty$ and $\mu \to 0$ at finite genome-wide mutation rate $U = L\mu$. In this limit the number of mutations per individual and generation is Poisson distributed with mean U. Importantly, each mutation is then novel and back mutations cannot occur. This can be expressed by characterising each individual's genotype by the set of acquired novel mutations, e.g.

$$M: \ \sigma^j = \{\tau, \zeta\} \to \begin{cases} \{\tau, \zeta\} \\ \{\tau, \zeta, \lambda\} \\ \{\tau, \zeta, \lambda, \gamma\} \\ ... \end{cases} \quad \forall\ j, \tag{5}$$

denoted by Greek letters. It is necessary to track the mutations carried by each invididual also in the *ism* in order to be able to implement recombination, which combines mutations and breaks them apart. However, once a mutation is fixed in the population, i.e., it is present in all individuals, it can be omitted due to the lack of back mutations, thereby keeping the list of stored mutation finite. Also, the stored list of fitness values of discovered genotypes can be purged of those genotypes whose mutation set does not contain newly fixed mutations, as they cannot be reached anymore.

While the features discussed so far are the same in all models, the differences are the following.

Concurrent recombination: In this case all individuals that are not the product of a recombination event select one ancestor with a probability proportional to the ancestor's fitness. For our neutral landscapes, this implies that viable ancestors are drawn uniformly with replacement. Simultaneously those individuals that are the product of a recombination event select two different ancestors with a probability proportional to their fitness. For a neutral landscape, this implies that each descendant takes a random sample of size two without replacement from the pool of viable ancestors.

Successive recombination: In both successive recombination models, selection occurs first, where all individuals choose one ancestor, again in our case uniformly with replacement among the viable ones. Next, recombination takes place independent of fitness. In the case of the simple successive recombination model, individuals that are a product of recombination choose two different ancestors that survived selection. In the case of successive recombination with mating pairs, all individuals that survived selection and happen to recombine are pooled in groups of two, which then create two offspring individuals. These two offspring individuals are complementary in their recombined material, that is, if one offspring has the allele of the first parent, the other offspring will have the allele of the second parent at the corresponding locus.

Importantly, the three recombination schemes differ in the way in which recombination couples to genetic drift. In the concurrent recombination and the successive recombination with mating pairs model, genetic drift is independent of the recombination rate r, but this is not the case for the simple successive recombination model (see Appendix). Recombination-dependent drift is a confounding factor that needs to be taken into account in the interpretation of the results of the latter model. Since the concurrent and the successive recombination with mating pairs models are implemented in two commonly used open software packages (Haller and Messer 2016; Zanini and Neher 2012), we stick to a non-recombination dependent genetic drift model in the main text and only occasionally refer to differences that would otherwise appear. The recombination-dependent genetic drift model has been used by Nowak *et al.* (2014). Being aware about these differences might be important for the design of experiments, e.g., in the context of *in vitro* recombination (Pesce *et al.* 2016). Of the two non-recombination dependent genetic drift models, we choose the concurrent recombination model because it is somewhat simpler. In particular, in this model the number of recombining individuals does not have to be an even number.

Measures of evolvability, diversity and robustness

To quantify evolvability in the *ism* we consider the discovery rate r_{dis} of novel genotypes and the fixation rate r_{fix} of mutations. The discovery rate r_{dis} is the average number of novel viable genotypes that are discovered in each generation, either through mutation or recombination. The fixation rate r_{fix} of mutations measures the average number of segregating mutations that become fixed in each generation.

In the *fsm* we monitor evolvability through the time t_{fdis} and the number of mutation events N_{mut} until full discovery. Starting from a monomorphic population carrying a single randomly selected viable genotype, we say that full discovery is reached when all viable genotypes have been present in at least one individual in at least one generation. The time is measured

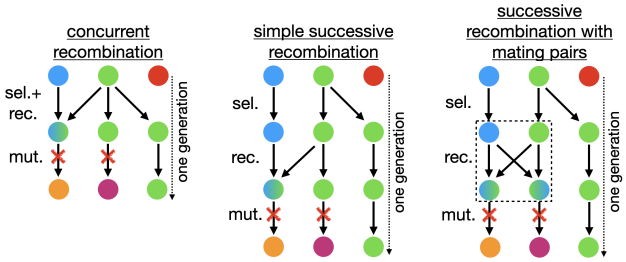

Figure 1 One generation of evolution in the three selection-recombination-mutation schemes discussed in the text. Nodes represent individuals and colors genotypes, with green and blue nodes being fit and red nodes unfit. The arrows show the lineages from the ancestors to the descendants. Individuals with two incoming arrows are the product of uniform recombination. In successive recombination with mating pairs the recombining individuals are grouped in pairs and each pair creates two descendants, which is indicated by the dashed box. Mutations are indicated by red crosses.

in generations and a mutation event occurs if an individual acquires one or multiple mutations during reproduction.

Genetic diversity is characterized through several well-known measures of population genetics, for which the time average is descriptive for the randomly drifting genotype cloud (note that in the regimes of interest here, the population cannot attain an equilibrium state, because the genotype space is larger than the population). Such measures are the pairwise mean Hamming distance between two individuals in the population

$$d_{pw} = \frac{1}{N(N-1)} \sum_{\substack{i,j \\ i \neq j}} d(\sigma^i, \sigma^j), \tag{6}$$

the number of viable distinct genotypes

$$Y = |\{\sigma^i | i \in 1, 2, ..., N \wedge \sigma^i \text{is viable}\}|, \tag{7}$$

and the number of segregating mutations

$$S = \sum_i^L \left(|\{\sigma_i^j\}_{j \in 1,2,..,N}| - 1 \right), \tag{8}$$

i.e., the number of loci at which both alleles are present in the population. These measures are used for the *ism* as well as for the *fsm*.

Additionally, for the *fsm* we consider the mutational robustness m of the population. The robustness m_{σ^i} of an individual i is equal to the fraction of viable point mutations of its genotype σ^i if it is itself viable, and equal to 0 otherwise. The mutational robustness of the population is the average robustness of all individuals,

$$m = \frac{1}{N} \sum_i^N m_{\sigma^i}. \tag{9}$$

This quantity depends on the population distribution in genotype space, and increases if the population moves to genotypes for which most point mutations are viable. Apart from these measures for evolvability, genetic diversity and mutational robustness we also measure the mean fitness defined by

$$\overline{w} = \frac{1}{N} \sum_i^N w_i, \tag{10}$$

as well as the average viable recombination fraction, which represents the fraction of recombination events per generation that generate a viable genotype. The latter can be considered a measure of recombination robustness.

Except for the measures for evolvability in the *fsm*, our numerical results always show the averages in steady state. For the *ism*, we evolve and evaluate the population for $t = \sqrt{10^{10}/U}$ generations for each data point. For the *fsm*, we evolve the population until all genotypes have been discovered once and, except for the measures of evolvability, more than 10^5 mutation events have occurred. This is done for 10^4 landscape realizations and for each data point. We measure all quantities at the end of a generation, i.e., after the mutation step, and denote the averages of measured quantities by overbars.

Illustration of results

3D wireframe plots. In order to represent the numerical results comprehensively, we mostly use 3D wireframe plots. In these plots, the wireframe lines run along either constant recombination rate or constant mutation rate to guide the eye. The color of the wireframe lines depends on their height. Additionally a contour plot is shown below the wireframe. The viewing angles vary and are selected such that the results can be seen in the best possible way.

Graph representation. To visualize the population distribution in sequence space, we use a graph representation. In this graph each genotype in the population is represented as a node, and nodes whose genotypes differ by a single point mutation are connected by an edge. Therefore, the resulting graph only contains information about nearest neighbor relationships. However, as long as the genotype cloud is not distributed too broadly in sequence space, we expect most nodes to have at least one edge, thereby forming clusters of connected components in the high dimensional sequence space. To arrange the nodes in two dimensions we use a forced-based algorithm called *ForceAtlas2* (Jacomy *et al.* 2014). This leads to a configuration in which nodes that share many edges form visual clusters. The frequencies of the genotypes are represented by the node sizes.

Data availability

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article. All numerical calculations including simulations described in this work were implemented in Python. All relevant source codes are available upon request.

Results and analysis

Evolutionary regimes

To organize the discussion of the results, we recall here the distinct evolutionary regimes that can be realized on neutral networks. In the *monomorphic* or *weak mutation* regime $NU \ll 1$, mutations are rare and either fix with probability $1/N$ or go extinct through genetic drift before another mutation originates. In this regime, the population consists of a single genotype most of the time and can be described by a "blind ant" random walker (Hughes 1996; van Nimwegen *et al.* 1999). Without lethal genotypes, a step to one of the current genotype's mutational neighbors is taken with equal probability at rate U independent of N. With lethal genotypes, a step is discarded when the randomly chosen point mutation is lethal. On a connected network this implies that the population occupies on average each genotype, irrespective of its degree, with equal probability (Hughes 1996). In this regime, the effect of recombination is minimal since recombination is fueled by combining segregating mutations, which do not exist most of the time. Furthermore, as long as there are not more than two genotypes in the population within one generation, the population is by definition in linkage equilibrium. Nonetheless, in subsection Recombination-induced genetic drift we show that recombination can have an effect in this regime if it couples to the genetic drift.

In the *polymorphic* regime ($NU \geq 1$), where the population is a cloud of competing genotypes, two subregimes can be distinguished. If the population size is large compared to the number of genotypes ($N \gg 2^L$), all viable genotypes can become occupied and an equilibrium state is reached. In this case genetic drift is irrelevant and the equilibrium distribution can be described by assuming deterministic dynamics of quasispecies type. In the absence of lethal genotypes, all genotypes then have the same frequency in the population, similar to the monomorphic regime. Importantly, though, with lethal genotypes, the population distribution becomes non-uniform, in that robust viable genotypes that have less lethal point mutations exhibit higher frequency (van Nimwegen *et al.* 1999). This imbalance increases dramatically with increasing recombination rate (Klug *et al.* 2019).

Of particular interest in the context of the present work is the second polymorphic subregime, where the population size is smaller than the number of genotypes ($N \ll 2^L$) or even smaller than the number of loci ($N \ll L$), and the population clearly cannot attain an equilibrium state. Instead, it will diffuse as a cloud of genotypes on the neutral network. This subregime has, to the best of our knowledge, not been fully covered in the literature, in particular in the presence of recombination. In the following sections, we first study the population structure in the *ism*, where $N \ll L$ is guaranteed by definition. Subsequently we consider a finite number of sites and study the *fsm* for $N \ll 2^L$.

Infinite-sites model

In this section, we present numerical results that show the impact of recombination and lethal genotypes on the population structure in the *ism*. We keep the population size fixed at $N = 100$ and discuss the dependence on the recombination rate r, the mutation rate U and the fraction of viable genotypes p.

Discovery rate. Figure 2 displays the discovery rate of formerly unexplored viable genotypes for $p = 1$ and $p = 0.5$. Without recombination, the discovery rate is given by

$$r_{dis} = pN(1 - e^{-U}) \overset{U \ll 1}{\approx} pNU \tag{11}$$

since each mutation generates a new genotype which is viable with probability p. The numerical results show that the effect of recombination on the discovery rate depends on the mutation rate U, the product NU and the fraction of viable genotypes p. If $NU \ll 1$, the effect of recombination is minimal, since there are very few segregating mutations that can be recombined. However, for $NU \geq 1$ we notice a rather complex dependence. In the absence of lethal genotypes ($p = 1$), recombination increases the discovery rate monotonically. The relative increase is maximal if $U \ll 1$ but $NU \gg 1$. If U is of order 1, the effect of recombination becomes smaller since the maximum discovery rate is capped by the population size N and almost exhausted through mutations.

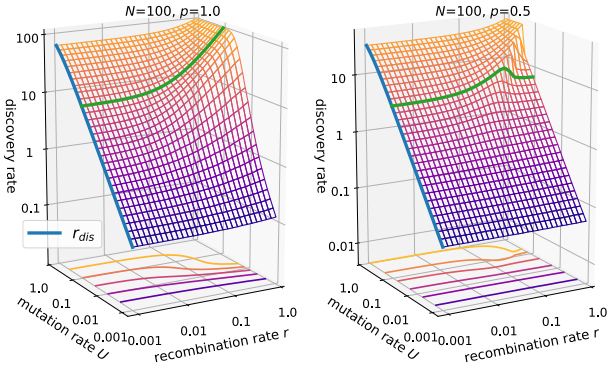

Figure 2 Discovery rate in the *ism* for population size $N = 100$ and $p = 1.0$ (left panel) vs. $p = 0.5$ (right panel). The green line at $U = 0.1$ highlights the different effects of recombination in the absence ($p = 1$) and presence ($p = 0.5$) of lethal genotypes. The blue lines in both panels show Eq. 11.

In the presence of lethals ($p = 0.5$) the behavior is more surprising. For small U, recombination has almost no effect, even if $NU \geq 1$. As U increases, an intermediate recombination rate becomes optimal for the discovery rate. This intermediate peak shifts to larger r with increasing U and the drop-off becomes sharper until U is of order 1, where the intermediate peak vanishes and the behavior is again similar to $p = 1$. These results indicate three regimes for the effect of recombination in the *ism*: (i) $NU \ll 1$, (ii) $NU \geq 1$ and $U \ll 1$, and (iii) $U \approx 1$.

Fixation rate. Next, we consider the fixation rate of segregating mutations. In the absence of recombination mutations are expected to fix at rate U, since in each generation NU mutations occur, a fraction $1/N$ of which goes to fixation. The results in Fig. 3 for $p = 1$ confirm this expectation and show that concurrent recombination has no effect if all genotypes are viable. However, for $p = 0.5$ the fixation rate is seen to dramatically decline at large recombination rates when $NU \geq 1$ and $U \ll 1$. This disruption of fixation is released only when the mutation rate becomes of order 1. We further notice, that at very large mutation rates, the fixation rate exceeds U.

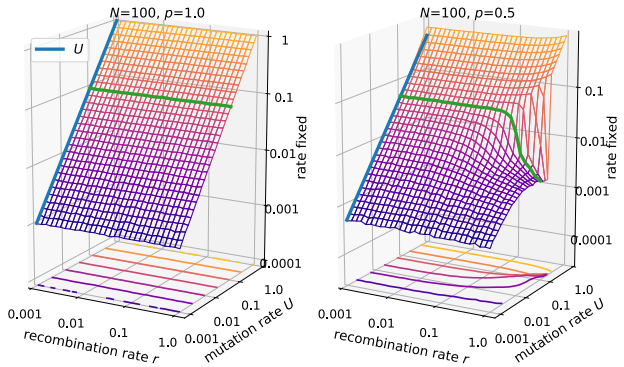

Figure 3 Fixation rate of segregating mutations in the *ism* with $N = 100$ and $p = 1.0$ (left panel) vs. $p = 0.5$ (right panel). The green line is drawn at $U = 0.1$.

Number of distinct genotypes. To understand the results for evolvability, we next consider measures of genetic diversity. The number of distinct genotypes Y is naturally closely related to the discovery rate, since with more novelty discovered in each generation, more distinct genotypes should accumulate, cf. Fig. 4. An analytical expression for the case of no recombination and

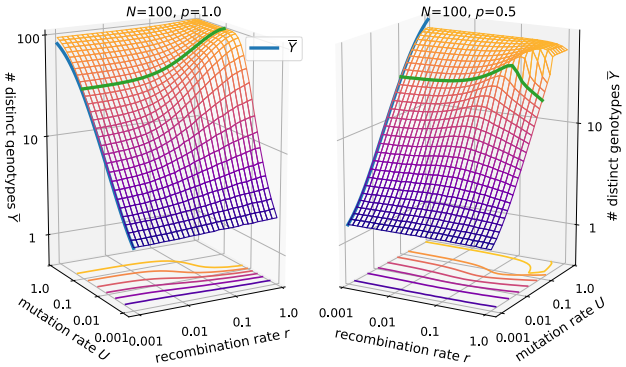

Figure 4 Number of distinct genotypes in the *ism* with $N = 100$ and $p = 1.0$ (left panel) vs. $p = 0.5$ (right panel). The blue lines shows Eq. 12, where θ is replaced by θ^* for $p < 1$. The green line is drawn at $U = 0.1$.

no lethal genotypes was derived in Ewens (2012):

$$\overline{Y}(r=0) = \sum_{i=0}^{N-1} \frac{\theta}{\theta+i} \quad \text{with} \quad \theta = 2NU. \tag{12}$$

Our numerical results fit this expression and furthermore show that in the absence of recombination, the formula can be extended to include lethal genotypes by replacing θ with

$$\theta^* = 2pNU. \tag{13}$$

The reasoning behind this replacement is that the parent population size is effectively reduced by a factor p in each generation. In general, the effect of recombination on the number of distinct genotypes is similar to that on the discovery rate. A noticeable difference in the case of $p = 0.5$ is that $\overline{Y}$ is not maximal at $U \approx 1$ but at a slightly smaller value. The reason is that at such high mutation rates a significant fraction $(1-p)U$ of individuals does not survive each generation, such that the population repeatedly goes through a bottleneck, in which distinct genotypes are lost. This leads to the observed increased fixation rate at $U \approx 1$ in Fig. 3.

Number of segregating mutations. Next we consider the number of segregating mutations S (Fig. 5). These span an effective sequence space of size 2^S in which recombination can create novel genotypes. An analytical expression for the number of segregating mutations in the *ism* was developed by Watterson (1975), again assuming no recombination and no lethal genotypes. It can be derived from the expectation of the length of the genealogical tree to the most recent common ancestor, which is given by (Wakeley 2009)

$$\overline{T}_{MRCA} = 2N \sum_{i=1}^{N-1} \frac{1}{i} \tag{14}$$

for the Wright-Fisher model. The total tree length is equal to the total time, and multiplying by the mutation rate U and

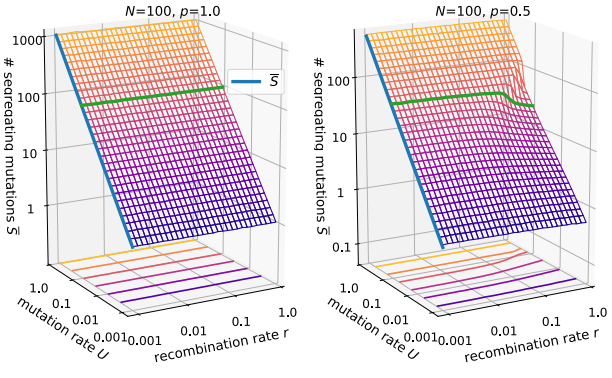

Figure 5 Number of segregating mutations in the *ism* with $N = 100$ and $p = 1.0$ (left panel) vs. $p = 0.5$ (right panel). The blue lines shows Eq. 15. The green line is drawn at $U = 0.1$.

the fraction of viable genotypes yields the average number of segregating mutations

$$\overline{S} = \theta^* \sum_{i=1}^{N-1} \frac{1}{i}. \tag{15}$$

The results also show that for $p = 1$ and the concurrent recombination model used here, the number of segregating mutations is independent of r. This is not generally true but depends on the implementation of recombination (see Recombination-induced genetic drift). For $p = 0.5$ and $NU \geq 1$, $U \ll 1$, recombination generally decreases the number of segregating mutations. For example, at $U = 0.1$ the number decreases from around 50 to about 20, which still yields an enormous effective sequence space compared to the population size. This alone cannot explain the decrease in evolvability seen in Figs. 2, 4.

Mean Hamming distance. Since recombination occurs between pairs of individuals we now consider the pairwise mean Hamming distance $\overline{d}_{pw}$ (Fig. 6). From Eq. 14 we conclude that the

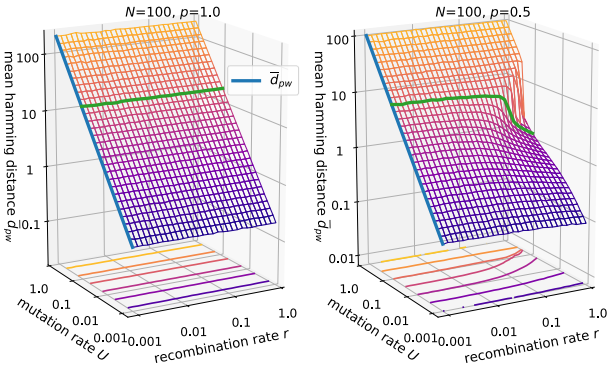

Figure 6 Mean Hamming distance in the *ism* with $N = 100$ and $p = 1.0$ (left panel) vs. $p = 0.5$ (right panel). The blue lines shows Eq. 16. The green line is drawn at $U = 0.1$.

mean length of the genealogical tree to the most recent common ancestor for two random individuals is given by $2N$, which directly leads to the expression

$$\overline{d}_{pw} = \theta^* \tag{16}$$

for the mean Hamming distance in non-recombining populations. The results for $p = 0.5$ show that recombination contracts the population cloud in sequence space in the presence of lethal genotypes. The functional relationship is similar to the number of segregating mutations but the relative change is more dramatic, e.g., for $U = 0.1$ the distance drops from 10 to about 1.

Cross section of the population cloud. To further understand the contraction in sequence space, we took a cross section of the population cloud by measuring the Hamming distance of each individual to the ancestral genotype that contains only fixed mutations, averaged over many generations (Fig. 7). This quantity is equal to the number of segregating mutations in an individual. For $r = 0$ the Hamming distance distribution is seen to follow a hypoexponential distribution, which converts to a Poisson distribution for large r. The hypoexponential distribution follows from the correspondence between the Hamming distance and the time to the most-recent-common-ancestor T_{MRCA}, the distribution of which is well known (Wakeley 2009). With a mutation rate U and a fraction p of viable genotypes, this yields the distribution

$$f_{h_{MRCA}}(d) = \frac{2}{\theta^*} \sum_{i=2}^{N} \binom{i}{2} e^{-\binom{i}{2}\frac{2d}{\theta^*}} \prod_{j=2, j\neq i}^{N} \frac{\binom{j}{2}}{\binom{j}{2} - \binom{i}{2}}, \tag{17}$$

with mean θ^*. At high recombination rates, segregating mutations become well mixed among all individuals, and the number of segregating mutations acquired by an individual follows a Poisson distribution. For $p = 1$ the mean is independent of r and equal to θ^*, but for $p = 0.5$ we observe a strong contraction of the distance distribution at $r = 1$. This shows that the focal genotype, around which the contraction occurs, is the ancestral genotype that contains only fixed mutations.

Mean fitness and recombination load. The contraction of the genotype distribution described in the preceding paragraphs must be beneficial for the population in the sense that the mean fitness increases. For our binary fitness distribution, the mean fitness is determined by the mutation and recombination load, i.e. the fraction of mutation and recombination events per generation that are lethal. The rate of lethal mutations is always $U(1 - p)$ and cannot be optimized in the *ism*. Contrary to that, the outcome of recombination events depends on the genotype composition of the population. If the population is contracted around a focal genotype and most individuals are closely related to each other, the effective sequence space for recombination is much smaller than 2^S. Therefore it becomes likely that a recombination event will not create a novel genotype but a genotype that already exists in the current genotype cloud and that, more importantly, is viable, which benefits the mean fitness.

This connection is shown in Fig. 8. If the population is sparsely distributed which happens at high mutation rates, most recombination events create novel genotypes, which leads to a viable recombination fraction equal to p. Contrary to that in a monomorphic population ($NU \ll 1$) no novelty results from recombination. In the regime $NU \geq 1, U \ll 1$, e.g. at $U = 0.1$, we see that with increasing r the viable fraction initially decreases, as the population becomes more diverse, leading to a decrease in fitness. At large recombination rates this trend reverses as the population becomes concentrated around a focal genotype, which then also leads to a fitness increase. Thus the contraction of the population is an adaptive response to the fitness decline caused by an increased recombination load.

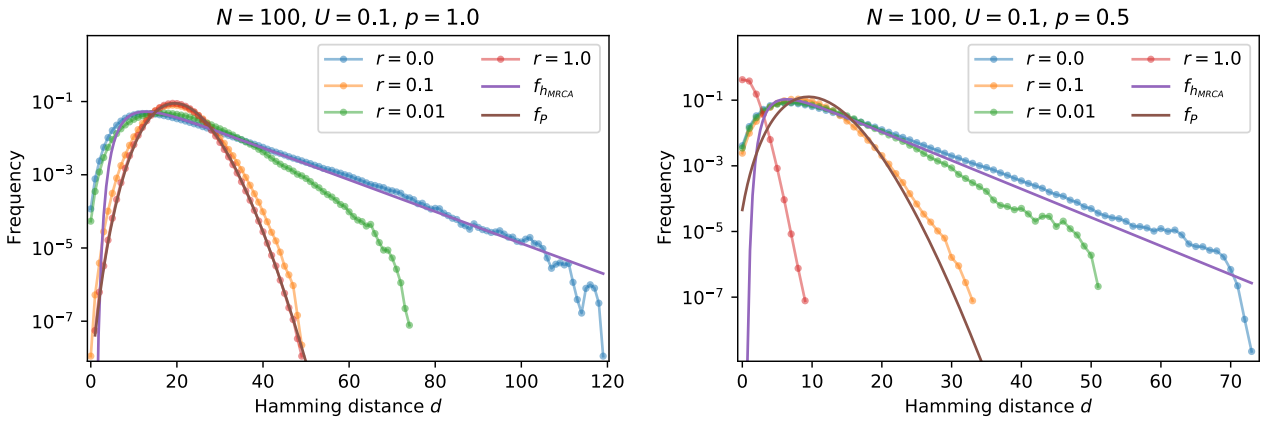

Figure 7 Cross section of the population cloud in the *ism* with $N = 100$, $U = 0.1$ and $p = 1.0$ (left panel) vs. $p = 0.5$ (right panel). The figures show the distribution of the Hamming distance to the genotype containing only fixed mutations. The hypoexponential distribution $f_{h_{MRCA}}$ is given by Eq. 17 and f_P is a Poisson distributions with mean θ^*. The data were accumulated over 10^6 generations.

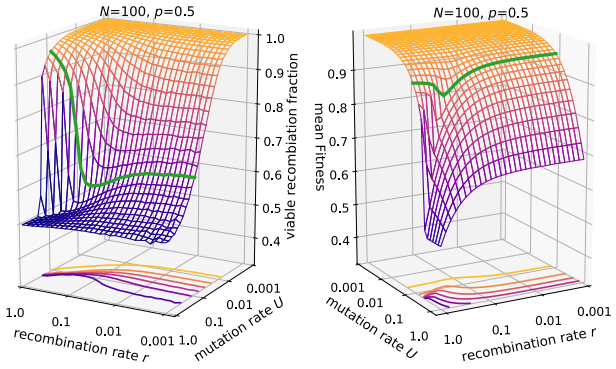

Figure 8 Viable recombination fraction and mean fitness in the *ism* for $N = 100$, $p = 0.5$. The green line is drawn at $U = 0.1$.

Dependence on the fraction of viable genotypes. The results presented so far were obtained for the two values $p = 1$ and $p = 0.5$ of the fraction of viable genotypes. Figure 9 shows cross sections of the previously shown 3D plots at either fixed recombination rate $r = 1$ (left column) or fixed mutation rate $U = 0.1$ (right column) and four different values of p. For fixed $r = 1$, the lethal genotypes strongly reduce evolvability by contracting the genotype cloud when the mutation rate is low, but the contraction is released once the mutation rate is strong enough. At this point the population evolves independent of fitness and therefore the measures coincide with the results for $p = 1$. However, the mean fitness is strongly reduced and equal to p. With smaller p this transition happens at larger U and strikingly at small enough p the numerical results display a discontinuity as a function of U.

The dependence on mutation rate resembles the error threshold phenomenon of quasispecies theory, in which the population delocalizes from a fitness peak in a finite-dimensional sequence space when the mutation rate is increased above a critical value (Jain and Krug 2007; Domingo and Schuster 2016). In the quasispecies context it has been shown that error thresholds do not occur in neutral landscapes with lethal genotypes, at least not in the absence of recombination (Wilke 2005; Wagner and Krall 1993). Moreover, for non-recombining populations the mean population fitness is generally continuous at the error threshold, whereas recombination can induce discontinuous fitness changes and bistability (Boerlijst *et al.* 1996). Although the transfer of results from the infinite population quasispecies model with finite sequence length L to the finite population *ism* is not straightforward, our results are generally consistent with previous work in that there is no discontinuity in the absence of recombination, while at sufficiently large recombination rates we find evidence for a discontinuous error threshold in the percolation landscape with lethal genotypes. From the perspective of quasispecies theory, the shift of the transition to larger U for decreasing p may tentatively be explained as an effect of increased selection pressure in landscapes with a larger fraction of lethal genotypes.

The results for fixed mutation rate $U = 0.1$ displayed in the right column of Fig. 9 show that the contraction of the population only occurs if the recombination rate is sufficiently large, whereas otherwise recombination increases genetic diversity and evolvability. Importantly, with increasing fraction of lethal genotypes the contraction occurs at smaller recombination rates.

Large populations. For large population sizes the number of segregating mutations grows rapidly, such that storing the part of the fitness landscape that could be revisited through recombination becomes computationally challenging. This limits the range of population sizes that can be explored. Nevertheless, the results for the discovery rate for $N = 1000$ displayed in Fig. S2 suggest that, for large populations, the interesting regime with a non-monotonous behavior in r appears in an even larger range of mutation rates than expected from the conditions $U \ll 1$ and $NU \geq 1$.

Summary of ism results. As expected, for $NU \ll 1$ or $U \approx 1$, recombination has almost no effect since the population is either monomorphic or dominated by mutations. In contrast, for $NU \geq 1$ and $U \ll 1$ the behavior is rather complex. While

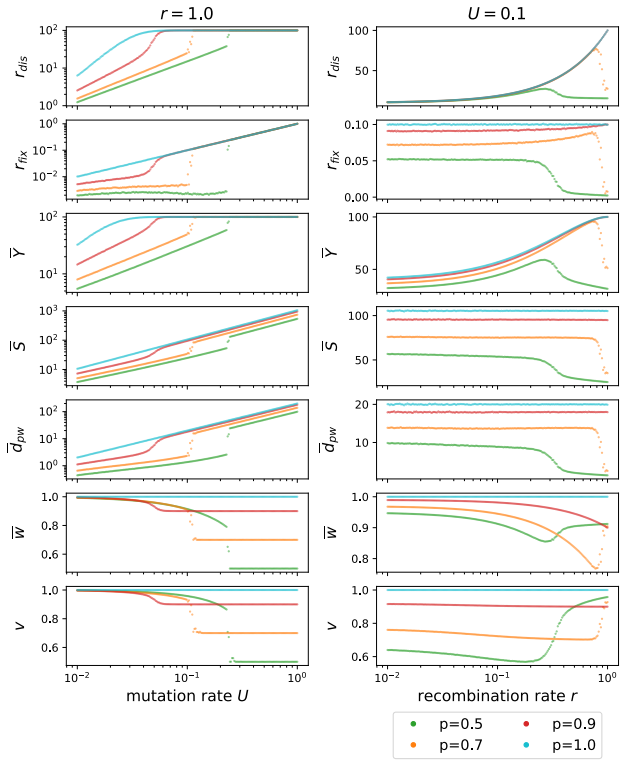

Figure 9 Cross sections of 3D plots of various measures characterizing the *ism* for four different values of the fraction p of viable genotypes. The left column shows the dependence on the mutation rate U for fixed $r = 1.0$, and the right column shows the dependence on the recombination rate r at fixed $U = 0.1$. The last row shows the fraction v of viable genotypes created by recombination events. The population size is $N = 100$.

low recombination rates generally diversify the population and increase evolvability, the population can dramatically change its genotype composition at high recombination rates, such that most genotypes are tightly clustered around a focal genotype. In the percolation landscape, the onset of this structural change depends on the fraction of lethal genotypes, whereas for general fitness landscapes we expect it to be determined by the degree distribution of the neutral network. The focal genotype of a tightly clustered population contains no segregating mutations. As a consequence the clustering decreases the discovery rates as well as the fixation rate, but the mean fitness increases. We conclude that recombination does not generally lead to increased evolvability, but may instead reduce diversity in order to increase fitness.

Finite-sites model

In this section, we study the effect of recombination in the finite-sites model. In contrast to the *ism* back mutations are possible, and the number of viable point mutations varies between genotypes. Therefore the population can optimize its mutational robustness to increase fitness. We keep the population size and sequence length fixed at $N = 100$ and $L = 10$, respectively, and investigate the interplay between the mutation rate per site μ and the recombination rate r.

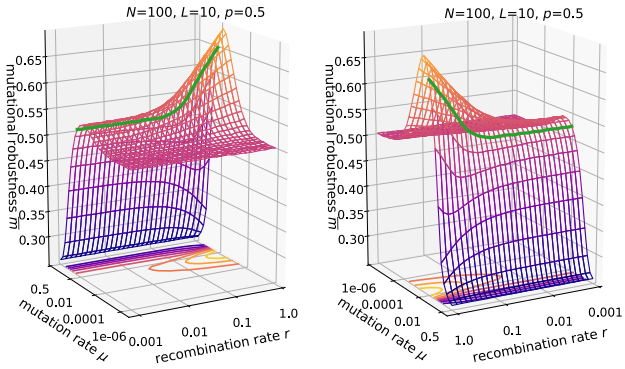

Figure 10 Mutational robustness in the *fsm* for $N = 100$, $L = 10$, $p = 0.5$. The two panels show the same data from two different viewing angles. The green line is drawn at $L\mu = 0.1$.

Mutational robustness. The results for the mutational robustness displayed in Fig. 10 show a complex dependence on r and μ, which reflects the different evolutionary regimes described above. Similar to the *ism*, for $NL\mu \ll 1$ the population is essentially monomorphic. In this regime it behaves like a random walker and all genotypes have equal occupation probability, such that the mutational robustness is equal to the average network degree $p = 0.5$. For a monomorphic population, recombination has no effect. With increasing mutation rate ($NL\mu \geq 1, \mu \ll 0.5$), the population becomes polymorphic but also more mutationally robust. Strikingly, with recombination, this effect is strongly amplified, as was observed previously in the quasispecies regime (Klug *et al.* 2019; Szöllősi and Derényi 2008). Similar to the results presented for the *ism* (Fig. 8), the increase in mutational robustness is accompanied by an increase in mean fitness (Fig. S3). In Klug *et al.* (2019) we showed that mutationally robust genotypes are more likely to be the outcome of recombination events, because they have a larger share of potentially viable parents. Therefore increased mutational robustness is a universal feature of recombining populations.

However, even higher mutation rates ($\mu \approx 0.5$) are detrimental to robustness, and recombination then also has a slightly negative effect. In this regime, the population is not concentrated anymore on a focal genotype but becomes highly delocalized and almost independent of the previous generation. Because of this, recombination events will produce random genotypes; note that for $\mu = 0.5$, all genotypes have the same probability after mutation, independent of viability. Thus, similar to the *ism* we can define three regimes with qualitatively different effects of recombination: (i) $NL\mu \ll 1$, (ii) $NL\mu \geq 1$ and $\mu \ll 0.5$, and (iii) $\mu \approx 0.5$.

Time to full discovery. We quantify evolvability in the *fsm* in terms of the time until all genotypes have been discovered, which can be interpreted as the inverse of the average discovery rate. For ease of comparison with the results for the *ism* shown in Fig. 2, in Fig. 11 we display the inverse of the time to full discovery in the *fsm* for $p = 1$ and $p = 0.5$. For $p = 1$, the overall behavior is similar to the *ism*, but the dependence on the recombination rate is significantly weaker. For $p = 0.5$, we see a decrease in the discovery rate for large r that is much more pronounced than in the *ism*. Whereas in the *ism* the discovery rate never

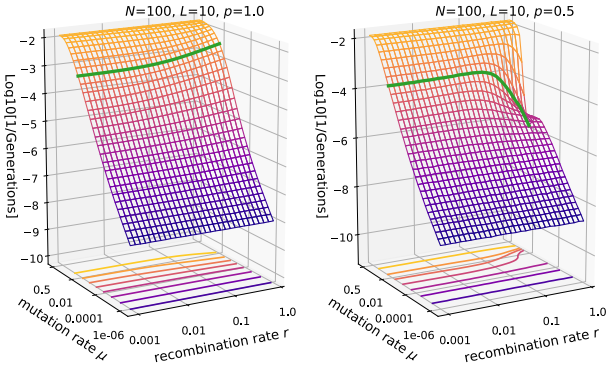

Figure 11 Reciprocal of the time to full discovery in the *fsm* with $N = 100$, $L = 10$ and $p = 1$ (left panel) vs. $p = 0.5$ (right panel). The green line is drawn at $L\mu = 0.1$.

drops below its value in the absence of recombination ($r = 0$), here the time to full discovery *diverges* at large recombination rates when $NL\mu \geq 0$ and $\mu \ll 1$. Furthermore, the dependence on mutation rate becomes non-monotonic at large r, which does not happen in the *ism*. The increase in the time to full discovery coincides with increased mutational robustness (Fig. 10) and a large viable recombination fraction (Fig. S3). Therefore this divergence occurs because the population is focused and entrenched in the highly robust regions of the fitness landscape. In this way a fitness ridge surrounded by many lethal genotypes can become almost impassable for strongly recombining populations.

In Figure 12 this phenomenon is explored over a wider range of parameters. When the fraction of viable genotypes is closer to $p = 1$, the rate of discovery displays an intermediate maximum as a function of r (panel A), and this behavior becomes more pronounced for longer sequences (panel B). Figure 12 C shows that strongly recombining populations are consistently slower in discovering the remaining undiscovered genotypes, which are presumably those that exhibit low mutational robustness.

Number of mutation events until full discovery. As a second measure of evolvability we consider the total number of mutation events N_{mut} until all viable genotypes have been discovered (Fig. S4). In the random walk regime $NL\mu \ll 1$, N_{mut} is independent of μ and r, since each mutation has the probability $1/N$ to go to fixation. As the mutation rate increases and the population spreads over the genotype space, fewer mutation events are necessary for full discovery. Similar to the time to full discovery t_{fdis}, depending on the fraction of viable genotypes, recombination can be either beneficial or detrimental for evolvability. In fact the two measures are related by

$$N_{mut} = t_{fdis} NL\mu \tag{18}$$

as long double mutations, which we count as a single mutation event, are sufficiently rare ($L\mu \ll 1$).

Genetic diversity. Figure S5 summarizes results for the genetic diversity in the *fsm*. Overall the impact of recombination on the genetic diversity is similar to, but less pronounced than the results for the *ism*. In particular, no discontinuities are observed in the variation of diversity measures with mutation rate (compare to Fig. 9). Because the number of segregating mutations and the mean Hamming distance are bounded in the *fsm*, the capacity

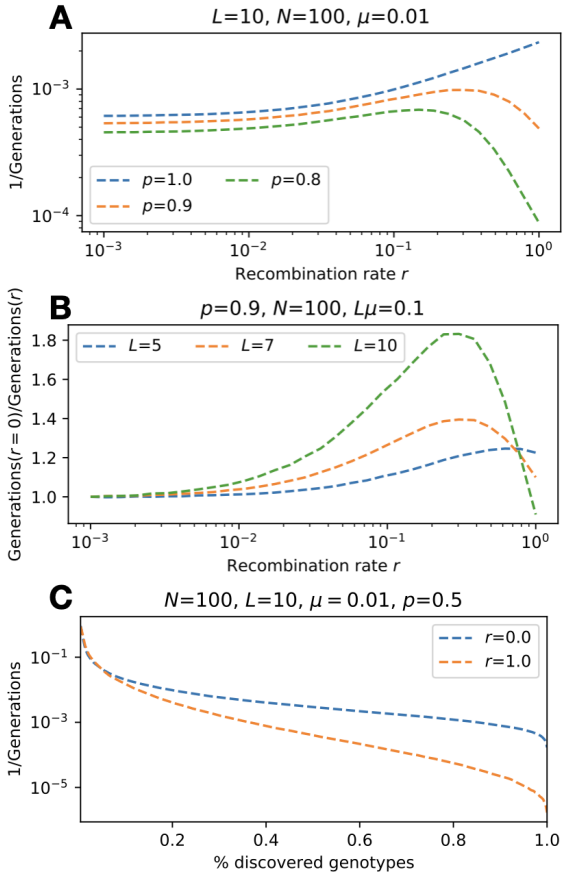

Figure 12 **A**: Reciprocal of the time to full discovery is shown as a function of recombination rate for different values of the fraction p of viable genotypes. **B**: Relative change in the reciprocal of the time to full discovery in obligately recombining ($r = 1$) vs. non-recombining ($r = 0$) populations for different values of the sequence length L and fixed genome-wide mutation rate $U = L\mu = 0.1$. **C**: Reciprocal of the time to discover a given fraction of all viable genotypes for a non-recombining and recombining population. In panel **C** the parameters are the same as in Fig. 11.

of recombination for creating diversity is limited. For example, while in the *ism* there are around 100 segregating mutations for $U = 0.1$, in the *fsm* with $L\mu = 0.1$ this number is limited by the sequence length $L = 10$ (middle row of S5). This suggests that the non-monotonic effect of recombination on evolvability in the *fsm* for $p \leq 1$ is mostly caused by an increase in mutational robustness. For $p = 1$ an analytical expression for the mean Hamming distance is derived in the Appendix. For the concurrent recombination scheme the result given in Eq. 34 is independent of the recombination rate, but this property is model dependent (see Recombination-induced genetic drift for further discussion).

Graph representation. To further illustrate the genotype composition of the population, a graph representation is employed in Fig. 13. These graphs show snapshots of genotype clouds that have evolved for sufficiently many generations to be independent of the initial condition. The population parameters are chosen to be in the regime $NL\mu \geq 1$, $\mu \ll 0.5$. The graphs of the recombining populations consist of significantly more edges representing mutational neighbors, which implies that they form a densely connected component in sequence space. By comparison, the graphs of the non-recombining populations have fewer edges, which implies that the populations are more dispersed. This is consistent with the results for the cross section of the genotype cloud in the *ism* in Fig. 7, which show a narrower distribution for recombining populations.

Figure S6 shows genotype clouds for larger values of L and N, but within the same evolutionary regime. The increased population size allows us to study the frequency distribution of genotypes in the population sorted by their rank. Remarkably, in non-recombining populations ($r = 0$) the distribution is exponential whereas for $r = 1$ we observe a heavy-tailed power law distribution, a feature that appears to be independent of p. The histograms in Fig. S6 also highlight the fact that, depending on the fraction p of viable genotypes, recombination can either increase or decrease the genetic diversity. In terms of mutational robustness, the graph representation shows that genotypes with high frequency exhibit an above-average robustness, thereby increasing the mutational robustness of the population.

Time evolution. So far we have studied the impact of recombination on stationary populations, where the effects of mutation, selection and drift balance on average. However, such a stationary state is generally not reached within a few generations, and in particular in the context of evolution experiments it is also important to understand the transient behavior. Since experiments usually consider a predefined small set of loci and track their evolution, we consider the temporal evolution in the *fsm*.

As an example, Fig. 14 shows the time evolution of mutational robustness m for obligately recombining and non-recombining populations at different mutation rates. The population is initially monomorphic and starts on a random viable genotype. To account for the variability between the trajectories observed in different realizations of the evolutionary process, the shading around the lines showing the average robustness represents the standard deviation of m. Analogous results for measures of evolvability and diversity are shown in Figs. S7, S8, S9 and S10.

As expected, the time scale for the establishment of the stationary regime is determined primarily by the mutation rate, since mutations create the diversity on which selection and recombination can act. At the lowest mutation rate recombin-

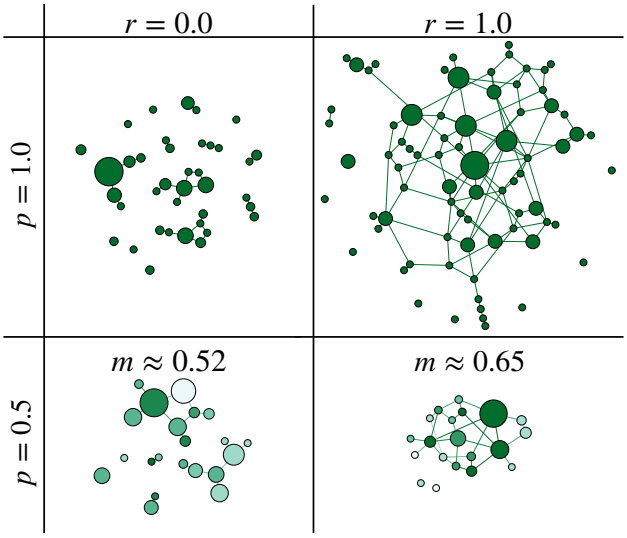

Figure 13 Graph representation of genotype clouds in the *fsm* with $N = 100$, $L = 10$, $\mu = 0.01$ after 10^6 generations of evolution starting from a random viable genotype. Links connect genotypes that differ by a single mutation, and node sizes represent the frequency of the corresponding genotype in the population; see Illustration of results for details. The networks on the left show non-recombining populations ($r = 0$) and on the right obligately recombining populations ($r = 1$). In the top panels all genotypes are viable ($p = 1$), whereas in the bottom panels half of the genotypes are lethal ($p = 0.5$). In the bottom row the mutational robustness of genotypes is shown by color coding with dark green representing high robustness and pale green low robustness, and the average robustness m is also indicated.

ing and non-recombining populations behave in the same way, and with increasing mutation rate the evolutionary regimes described above are traversed. This implies in particular that the ordering between the lines representing $r = 1$ and $r = 0$ may change as a function of μ (Figs. 14, S7 and S10). The distinction between recombining and non-recombining populations is often most pronounced at intermediate values of the mutation rate (Figs. S9 and S10).

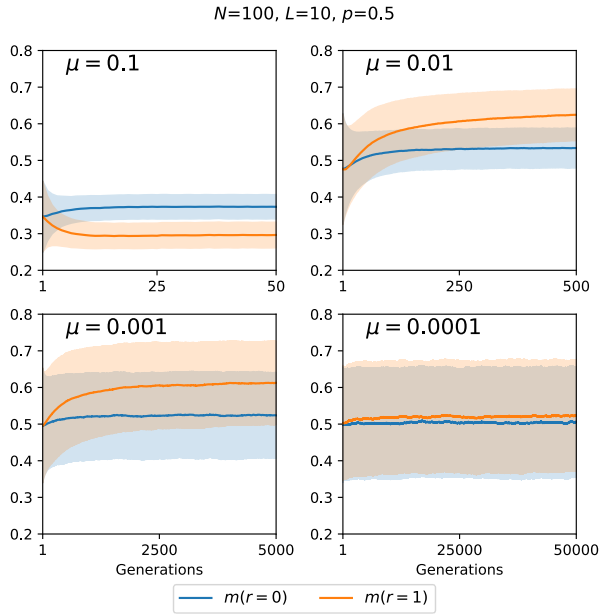

Figure 14 Time evolution of mutational robustness in the *fsm* with $N = 100$, $L = 10$, $p = 0.5$ for different values of the mutation rate μ. Each panel compares obligately recombining ($r = 1$) and non-recombining ($r = 0$) populations. Thick lines represent the mean over 5000 landscape realizations and the shaded areas the corresponding standard deviation.

Recombination-induced genetic drift

If recombination does not occur concurrently with selection but successively, it can act as an additional source of genetic drift. This recombination-induced drift is a confounding factor that needs to be accounted for when interpreting the results obtained with successive recombination models. As an example, Fig. 15 shows results for the numbers of distinct genotypes and segregating mutations for the *ism* with $p = 1$ using the simple successive recombination scheme. While in the concurrent recombination model the number of distinct genotypes is strictly increasing with r at $p = 1$ (Fig. 4), the effect of recombination in the simple successive model is mutation rate dependent and can be non-monotonic even at $p = 1$. This is due to a decrease in the number of segregating mutations and the mean Hamming distance with increasing r which occurs through the additional genetic drift (see Appendix). In the *fsm* the effect is similar and can result in an intermediate peak in the mutational robustness as a function of r when $NL\mu \approx 1$ (Fig. S11). While the effect in the *ism* occurs at all population sizes, in the *fsm* it is only significant for relatively small populations, because the number of segregating mutations is capped at L.

If the design of a simulation model or an *in vitro* experiment requires that recombination and selection occur independently, but recombination should not be an additional source of genetic drift, then successive recombination with mating pairs, as illustrated in Fig. 1, might be an option. Alternatively, one can mitigate the additional genetic drift in the simple successive recombination model by performing the selection and recombination step in each generation within a large population from which a small sample of size N is subsequently drawn.

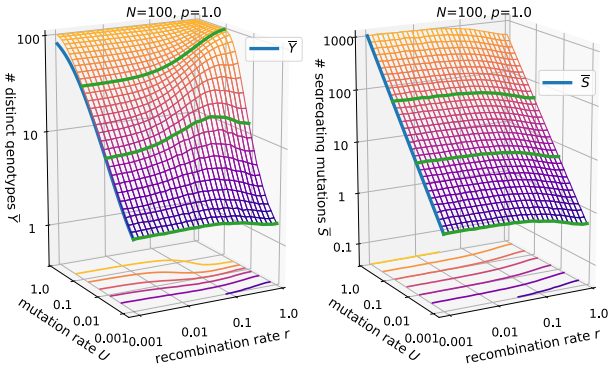

Figure 15 Number of distinct genotypes and segregating mutations in the simple successive recombination model for $p = 1$ in the *ism*. Compare to Figs. 4, 5. Green lines show constant $U = 0.1$, $U = 0.01$ and $U = 0.001$.

Discussion

The most apparent effect of recombination is to shuffle alleles among individuals. Therefore, one might think that recombination always increases genetic diversity and, through that, evolvability, as formulated in Weismann's hypothesis (Weismann 1891). In this study, we have shown that while this is true on a neutral network in the absence of lethal genotypes, the effect of recombination in the presence of lethal or highly deleterious mutations is much more complex.

More precisely, if the fraction of lethal genotypes in the fitness landscape is large enough, we find the emergence of two different regimes for the effect of recombination. While small recombination rates increase diversity in accordance with Weismann's hypothesis, at sufficiently high recombination rates we observe a strong contraction of the genotype cloud and reduced evolvability. This contraction regime is characterized by a clustering of the population in sequence space around a focal genotype, which we have shown to be the most recent common ancestor. Therefore most genotypes have only a few segregating mutations as well as a small pairwise mean Hamming distance, which leads to a reduced number of distinct genotypes. This benefits mean fitness, as recombination events more often lead to viable genotypes. Results for the finite-sites model further reveal, that polymorphic genotype clouds are most dense around mutationally robust genotypes and that a contraction through recombination therefore greatly increases mutational robustness. The trade-off is a reduced evolvability. In the infinite-sites model this is manifested through a reduced discovery rate. Furthermore, the lower frequency of segregating mutations leads to a reduced fixation rate of segregating mutations.

Therefore, even in an infinite-sites setting, recombination can, in some sense, entrench the population. However, as the

number of potential mutation sites is unbounded, the discovery rate never falls below the discovery rate in the absence of recombination. In contrast, in the finite-sites model strongly recombining populations can become trapped in mutationally robust regions, such that the time to full discovery diverges. The recombination-induced trapping of the population at fitness peaks is well known from studies on non-neutral fitness landscapes (Eshel and Feldman 1970; Weinreich and Chao 2005; de Visser *et al.* 2009; Jain 2010; Park and Krug 2011; Altland *et al.* 2011), but our results show that a similar phenomenon occurs in neutral landscapes with fitness plateaus and ridges. In terms of genetic diversity the results in the finite and infinite-sites settings are similar but more gradual in the latter case, as the number of segregating mutations and the mean Hamming distance are capped.

Overall, our numerical simulations show a very consistent increase in mutational robustness with recombination rate in polymorphic populations. Related to this is the observation that recombination leads to a heavy-tailed frequency distribution of genotype abundance, and that the most frequent genotypes have an above-average robustness, thereby increasing the robustness of the whole population. As discussed in previous work, the most frequent genotypes are more likely those that have an above-average fraction of possible viable parent combinations (Klug *et al.* 2019).

The broad scope of our investigation demonstrates that the effects of recombination vary widely across parameter combinations and evolutionary regimes, and this has to be accounted for when interpreting apparent contradictions between different experiments. Furthermore, it is important to distinguish long-term effects of recombination from short-term effects. In this study, we mainly considered long-term effects in stationary populations. Short-term effects can be different, in particular when evolution proceeds in a changing environment (Becks and Agrawal 2012; Nowak *et al.* 2014).

Literature Cited

Altland, A., A. Fischer, J. Krug, and I. G. Szendro, 2011 Rare events in population genetics: stochastic tunneling in a two-locus model with recombination. Physical review letters **106**: 088101.

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

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

Bloom, J. D., J. J. Silberg, C. O. Wilke, D. A. Drummond, C. Adami, *et al.*, 2005 Thermodynamic prediction of protein neutrality. Proceedings of the National Academy of Sciences **102**: 606–611.

Boerlijst, M. C., S. Bonhoeffer, and M. A. Nowak, 1996 Viral quasi-species and recombination. Proc. R. Soc. Lond. B **263**: 1577–1584.

Bornberg-Bauer, E. and H. S. Chan, 1999 Modeling evolutionary landscapes: Mutational stability, topology, and superfunnels in sequence space. Proceedings of the National Academy of Sciences **96**: 10689–10694.

Burt, A., 2000 Perspective: sex, recombination, and the efficacy of selection—was weismann right? Evolution **54**: 337–351.

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

de Visser, J. A. G. and S. F. Elena, 2007 The evolution of sex: empirical insights into the roles of epistasis and drift. Nature Reviews Genetics **8**: 139–149.

de Visser, J. A. G. and J. Krug, 2014 Empirical fitness landscapes and the predictability of evolution. Nature Reviews Genetics **15**: 480–490.

de Visser, J. A. G., S.-C. Park, and J. Krug, 2009 Exploring the effect of sex on empirical fitness landscapes. The American Naturalist **174**: S15–S30.

Domingo, E. and P. Schuster, editors, 2016 *Quasispecies: From Theory to Experimental Systems*. Springer Science & Business Media.

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

Ewens, W. J., 2012 *Mathematical population genetics 1: theoretical introduction*. Springer Science & Business Media.

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

Felsenstein, J., 1974 The evolutionary advantage of recombination. Genetics **78**: 737–756.

Gavrilets, S., 1997 Evolution and speciation on holey adaptive landscapes. Trends in ecology & evolution **12**: 307–312.

Gavrilets, S., 2004 *Fitness landscapes and the origin of species*. Princeton University Press.

Gavrilets, S. and J. Gravner, 1997 Percolation on the fitness hypercube and the evolution of reproductive isolation. Journal of theoretical biology **184**: 51–64.

Guo, H. H., J. Choe, and L. A. Loeb, 2004 Protein tolerance to random amino acid change. Proceedings of the National Academy of Sciences **101**: 9205–9210.

Haller, B. C. and P. W. Messer, 2016 SLiM 2: Flexible, Interactive Forward Genetic Simulations. Molecular Biology and Evolution **34**: 230–240.

Hughes, B., 1996 *Random Walks and Random Environments*. Clarendon Press, Oxford.

Jacomy, M., T. Venturini, S. Heymann, and M. Bastian, 2014 Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. PloS one **9**: e98679.

Jain, K., 2010 Time to fixation in the presence of recombination. Theoretical Population Biology **77**: 23–31.

Jain, K. and J. Krug, 2007 Adaptation in simple and complex fitness landscapes. In *Structural Approaches to Sequence Evolution*, edited by H. E. R. Ugo Bastolla, Markus Porto and M. Vendruscolo, Springer.

Johnson, M. S., A. Martsul, S. Kryazhimskiy, and M. M. Desai, 2019 Higher-fitness yeast genotypes are less robust to deleterious mutations. Science **366**: 490–493.

Kimura, M., 1968 Evolutionary rate at the molecular level. Nature **217**: 624–626.

Kimura, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics **61**: 893.

Kimura, M., 1983 *The neutral theory of molecular evolution*. Cambridge University Press.

Klug, A., S.-C. Park, and J. Krug, 2019 Recombination and mutational robustness in neutral fitness landscapes. PLoS computational biology **15**: e1006884.

Kondrashov, A. S., 1988 Deleterious mutations and the evolution of sexual reproduction. Nature **336**: 435–440.
Manrubia, S., J. A. Cuesta, J. Aguirre, S. E. Ahnert, L. Altenberg, *et al.*, 2021 From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. Physics of Life Reviews **38**: 55–106.
Maynard Smith, J., 1970 Natural selection and the concept of a protein space. Nature **225**: 563–564.
Muller, H. J., 1932 Some genetic aspects of sex. The American Naturalist **66**: 118–138.
Muller, H. J., 1964 The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **1**: 2–9.
Nobrega, M. A., Y. Zhu, I. Plajzer-Frick, V. Afzal, and E. M. Rubin, 2004 Megabase deletions of gene deserts result in viable mice. Nature **431**: 988–993.
Nowak, S., J. Neidhart, I. G. Szendro, and J. Krug, 2014 Multidimensional epistasis and the transitory advantage of sex. PLoS Computational Biology **10**: e1003836.
Ohta, T., 2002 Near-neutrality in evolution of genes and gene regulation. Proceedings of the National Academy of Sciences **99**: 16134–16137.
Otto, S. P., 2009 The evolutionary enigma of sex. The American Naturalist **174**: S1–S14.
Park, S.-C. and J. Krug, 2011 Bistability in two-locus models with selection, mutation, and recombination. Journal of Mathematical Biology **62**: 763–788.
Pesce, D., N. Lehman, and J. A. G. M. de Visser, 2016 Sex in a test tube: testing the benefits of in vitro recombination. Phil. Trans. R. Soc. B **371**: 20150529.
Reddy, G. and M. M. Desai, 2021 Global epistasis emerges from a generic model of a complex trait. eLife **10**: e64740.
Rodrigues, J. F. M. and A. Wagner, 2009 Evolutionary plasticity and innovations in complex metabolic reaction networks. PLoS Comput Biol **5**: e1000613.
Serva, M. and L. Peliti, 1991 A statistical model of an evolving population with sexual reproduction. Journal of Physics A: Mathematical and General **24**: L705.
Singhal, S., S. M. Gomez, and C. L. Burch, 2019 Recombination drives the evolution of mutational robustness. Current Opinion in Systems Biology **13**: 142–149.
Szöllősi, G. J. and I. Derényi, 2008 The effect of recombination on the neutral evolution of genetic robustness. Mathematical biosciences **214**: 58–62.
van Nimwegen, E., J. P. Crutchfield, and M. Huynen, 1999 Neutral evolution of mutational robustness. Proceedings of the National Academy of Sciences **96**: 9716–9720.
Wagner, A., 2005 Robustness, evolvability, and neutrality. FEBS letters **579**: 1772–1778.
Wagner, G. and P. Krall, 1993 What is the difference between models of error thresholds and Muller's ratchet? Journal of Mathematical Biology **32**: 33–44.
Wakeley, J., 2009 *Coalescent theory: An introduction*. Roberts & Co. Publishers.
Watterson, G., 1975 On the number of segregating sites in genetical models without recombination. Theoretical population biology **7**: 256–276.
Weinreich, D. M. and L. Chao, 2005 Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. Evolution **59**: 1175–1182.
Weismann, A., 1891 *Essays upon heredity and kindred biological problems*, volume 1. Clarendon press.
Wilke, C. O., 2001 Adaptive evolution on neutral networks. Bulletin of Mathematical Biology **63**: 715–730.
Wilke, C. O., 2005 Quasispecies theory in the context of population genetics. BMC evolutionary biology **5**: 1–8.
Zanini, F. and R. A. Neher, 2012 FFPopSim: an efficient forward simulation package for the evolution of large populations. Bioinformatics **28**: 3332–3333.

Appendix

In the following, analytical expressions for the mean Hamming distance $\overline{d}_{pw}$ are derived for all three recombination models illustrated in Fig. 1. Throughout this appendix these models are abbreviated as *cr* (concurrent recombination), *ssr* (simple successive recombination) and *srmp* (successive recombination with mating pairs), respectively. The derivation is based on the approach of Serva and Peliti (1991), which is generalized to take into account recombination. It is assumed that there are no lethal genotypes ($p = 1$). The results apply to the *fsm* and the *ism* and show that the recombination rate influences $\overline{d}_{pw}$ only in the *ssr* model. The r-dependence decreases with N in the *fsm* but remains independent of N in the *ism*.

To start, in the selection step each individual α picks a parent α' at random from the previous generation. During the mutation step, a mutation occurs at each locus i with probability μ, changing its state from -1 to 1 or vice versa,

$$\sigma_i^{\alpha}(t+1) = \begin{cases} \sigma_i^{\alpha'}(t), & \text{with prob. } 1-\mu \\ -\sigma_i^{\alpha'}(t), & \text{with prob. } \mu. \end{cases} \tag{19}$$

This can also be written as

$$\sigma_i^{\alpha}(t+1) = \epsilon_i^{\alpha}(t)\sigma_i^{\alpha'}(t) \tag{20}$$

with

$$\epsilon_i^{\alpha}(t) = \begin{cases} +1, & \text{with prob. } 1-\mu \\ -1, & \text{with prob. } \mu. \end{cases} \tag{21}$$

During the recombination step, individuals recombine at rate r. Altogether this leads to

$$\begin{aligned} \sigma_i^{\alpha}(t+1) =& \epsilon_i^{\alpha}(t)\Big[\kappa^{\alpha}(t)\left(\xi_i^{\alpha}(t)\sigma_i^{\alpha'}(t) + (1-\xi_i^{\alpha}(t))\sigma_i^{\alpha''}(t)\right) \\ &+ (1-\kappa^{\alpha}(t))\sigma_i^{\alpha'''}(t+1)\Big] \end{aligned} \tag{22}$$

where α', α'' are the parental genotypes in case of recombination and α''' is the parent in case of no recombination. The random variable κ^{α} determines whether a recombination event occurs and is given by

$$\kappa^{\alpha}(t) = \begin{cases} 1 & \text{with prob. } r, \\ 0 & \text{with prob. } 1-r. \end{cases} \tag{23}$$

The random variable ξ_i^{α} determines from which parent the allele is taken in case of a recombination event:

$$\xi_i^{\alpha}(t) = \begin{cases} 1 & \text{with prob. } 1/2, \\ 0 & \text{with prob. } 1/2. \end{cases} \tag{24}$$

Next the relatedness Q of the population is computed, which is defined by

$$Q = \binom{N}{2}^{-1} \sum_{(\alpha,\beta)} \frac{1}{L} \sum_{i=1}^{L} \sigma_i^{\alpha}\sigma_i^{\beta} \tag{25}$$

where the sum runs over all different pairs of individuals (α, β). We are interested in the average relatedness $\overline{Q}$ in the stationary state, and therefore make use of Eq. 22 by evaluating

$$\begin{aligned}
&\overline{\sigma_i^\alpha(t+1)\sigma_i^\beta(t+1)} \\
&= \underbrace{\epsilon_i^\alpha \epsilon_i^\beta}_{(1-2\mu)^2} \Bigg[\underbrace{\kappa^\alpha \kappa^\beta}_{r^2} \underbrace{\left(\xi_i^\alpha \sigma_i^{\alpha'} + (1-\xi_i^\alpha)\sigma_i^{\alpha''}\right)\left(\xi_i^\beta \sigma_i^{\beta'} + (1-\xi_i^\beta)\sigma_i^{\beta''}\right)}_{(*)} \\
&\overline{+ \underbrace{(1-\kappa^\alpha)(1-\kappa^\beta)}_{(1-r)^2} \sigma_i^{\alpha'''}\sigma_i^{\beta'''}} \\
&\overline{+ \underbrace{\kappa^\alpha(1-\kappa^\beta)}_{r(1-r)} \underbrace{\left(\xi_i^\alpha \sigma_i^{\alpha'} + (1-\xi_i^\alpha)\sigma_i^{\alpha''}\right)\sigma_i^{\beta'''}}_{\overline{\sigma_i^{\alpha'}\sigma_i^{\beta'''}}}} \\
&\overline{+ \underbrace{(1-\kappa^\alpha)\kappa^\beta}_{r(1-r)} \underbrace{\sigma_i^{\alpha'''}\left(\xi_i^\beta \sigma_i^{\beta'} + (1-\xi_i^\beta)\sigma_i^{\beta''}\right)}_{\overline{\sigma_i^{\beta'}\sigma_i^{\alpha'''}}}} \Bigg].
\end{aligned} \tag{26}$$

To simplify notation, the t-dependence is suppressed on the right hand side. Equation 26 holds for all three models, but differences emerge in the probability that the individuals α, β share a parent as well as in the term marked by $(*)$.

The probability that two non-recombining individuals have the same parent $\alpha''' = \beta'''$ through selection is given by $1/N$ in all models. This leads to

$$\overline{\sigma_i^{\beta'''}(t)\sigma_i^{\alpha'''}(t)} = \frac{1}{N} + \left(1 - \frac{1}{N}\right)\overline{Q}(t). \tag{27}$$

With *cr* and *srwm*, a recombining and non-recombining individual share parents $\alpha' = \beta'''$ with probability $1/N$, leading again to the right hand side of Eq. 27. However, for *ssr*, the probability that they share a common parent $\alpha' = \beta'''$ is $1/N + (1-1/N)1/N \approx 2/N$, as this can occur either during the recombination or selection step. This yields

$$\overline{\sigma_i^{\beta'''}(t)\sigma_i^{\alpha'}(t)} = \overline{\sigma_i^{\beta'}(t)\sigma_i^{\alpha'''}(t)} \approx \frac{2}{N} + \left(1 - \frac{2}{N}\right)\overline{Q}(t). \tag{28}$$

We next turn to the evaluation of the term in Eq. 26 marked by $(*)$. In the *cr* and *ssr* model the random variables ξ_i are not correlated between individuals. Therefore in both models $(*)$ simplifies to

$$(*) = \overline{\sigma_i^{\alpha'}(t)\sigma_i^{\beta'}(t)}. \tag{29}$$

For *cr* this leads to the right hand side of Eq. 27, while for *ssr* we get Eq. 28 using the same argument as before. Similar to *ssr*, for *srwm*, recombining individuals have an increased chance of sharing a parent $\alpha' = \beta'$ since they either can belong to the same mating pair with probability $2/(rN)$ or share a parent during selection with probability $1/N$. However, this is exactly balanced by the constraint that mating pairs are complementary in their recombined material, which is reflected in a correlation of the random variables ξ_i. Simply put, the increased chance that two individuals share the same parent is offset by the constraint that they always inherit the allele of the respective other parental genotype. Therefore $(*)$ again leads to Eq. 27 for *srwm*.

Summarizing, we have

$$\begin{aligned}
\text{suc. rec. pairs:}\quad & \overline{Q(t+1)} = (1-2\mu)^2\left[\frac{1}{N} + (1-\frac{1}{N})\overline{Q}(t)\right]. \\
\text{conc. rec.:}\quad & \overline{Q(t+1)} = (1-2\mu)^2\left[\frac{1}{N} + (1-\frac{1}{N})\overline{Q}(t)\right]. \\
\text{simp. suc. rec.:}\quad & \overline{Q(t+1)} \approx (1-2\mu)^2\Big[r^2(\frac{2}{N} + (1-\frac{2}{N})\overline{Q}(t)) \\
& \qquad + (1-r)^2(\frac{1}{N} + (1-\frac{1}{N})\overline{Q}(t)) \\
& \qquad + 2r(1-r)(\frac{2}{N} + (1-\frac{2}{N})\overline{Q}(t))\Big].
\end{aligned} \tag{30}$$

Note that the r dependence has vanished with *cr* & *srmp*. Next we compute the stationary relatedness by setting $\overline{Q(t+1)} = \overline{Q(t)}$. For *cr* & *srmp* this yields

$$\overline{Q} = \frac{(1-2\mu)^2}{4(1-\mu)\mu(N-1)+1}, \tag{31}$$

while for *ssr* we get

$$\overline{Q} = \frac{(1-2\mu)^2[1+r(2-r)]}{4(1-\mu)\mu(N-1) - (1-2\mu)^2 r^2 + 2(1-2\mu)^2 r + 1}. \tag{32}$$

The relatedness is connected to the mean Hamming distance through

$$\overline{d}_{pw} = \frac{L(1-\overline{Q})}{2}, \tag{33}$$

which leads in the cases of *cr* & *srmp* to

$$\overline{d}_{pw} = \frac{2(1-\mu)\mu LN}{4(1-\mu)\mu(N-1)+1} \tag{34}$$

and in the case of *ssr* to

$$\overline{d}_{pw} = \frac{2(1-\mu)\mu LN}{4(1-\mu)\mu(N-1) - (1-2\mu)^2 r^2 + 2(1-2\mu)^2 r + 1}. \tag{35}$$

In Figures S5 and S12 we compare the expressions in Eqs. 26 and 35 to numerical simulations and find excellent agreement.

In the deterministic limit $N \to \infty$ with finite L the mean Hamming distance is the same for all recombination models:

$$\overline{d}_{pw} = \frac{L}{2}. \tag{36}$$

However, in the *ism* limit ($L \to \infty, \mu \to 0, L\mu = U$) the result for *cr* & *srmp* leads to

$$\overline{d}_{pw} = 2NU = \theta \tag{37}$$

while the result for *ssr* reads

$$\overline{d}_{pw} = \frac{\theta}{1 + r(2-r)}. \tag{38}$$

Hence in this limit the dependence on the recombination rate persists independent of the population size.

To obtain an analytical expression for $\overline{S}$ in the *ism* under *ssr* we adopt a relation between the mean Hamming distance and the number of segregating sites $\overline{S}$ that has been derived for the non-recombining case (Wakeley 2009). This leads us to

$$\overline{S} \approx \overline{d}_{pw} \sum_{i=1}^{N-1} \frac{1}{i} = \frac{\theta}{1 + r(2-r)} \sum_{i=1}^{N-1} \frac{1}{i}. \tag{39}$$

Figure S12 shows that, at least for the parameter regime of interest here, Eq. 39 provides an accurate approximation of the simulation results.

Supplementary figures

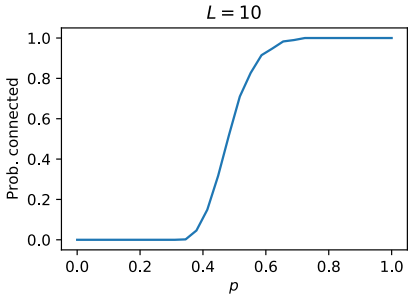

Figure S1 Probability that the network of viable genotypes is connected in a percolation landscape, where genotypes are viable or lethal independently with probability p. The sequence length is $L = 10$.

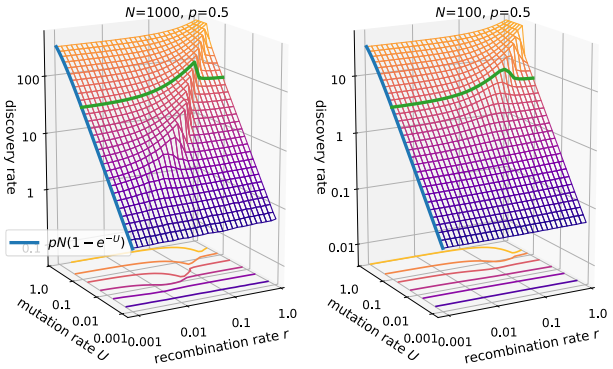

Figure S2 Discovery rate in the *ism* for viable fraction $p = 0.5$ and population size $N = 1000$ (left panel) vs. $N = 100$ (right panel). The green line is drawn at $U = 0.1$ and the blue lines in both panels show Eq. 11. The right panel is identical to that in Fig. 2.

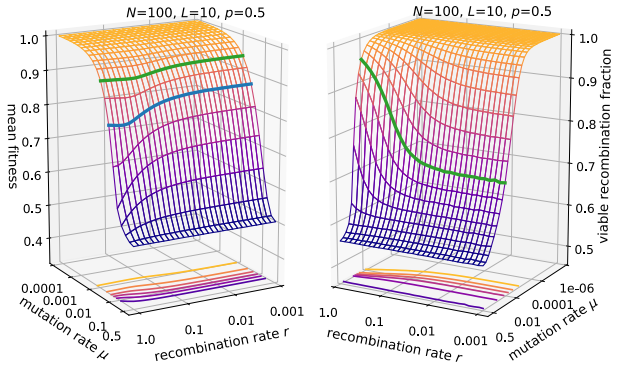

Figure S3 Mean fitness and viable recombination fraction in the *fsm* with concurrent recombination. The green line is drawn at $\mu = 0.01$ in both panels. Similar to the results for the *ism* (Fig. 8), the fitness displays an intermediate minimum at the point where the population structure changes. This is best visible at $\mu = 0.03$ (blue line). Compared to the *ism* the variation in mean fitness and viable recombination fraction is less pronounced.

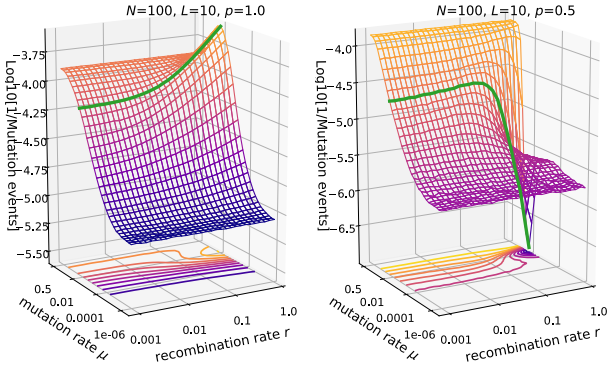

Figure S4 Reciprocal of the total number of mutation events until full discovery in the *fsm* with $N = 100$, $L = 10$ and $p = 1.0$ (left panel) vs. $p = 0.5$ (right panel). The green line is drawn at $L\mu = 0.1$.

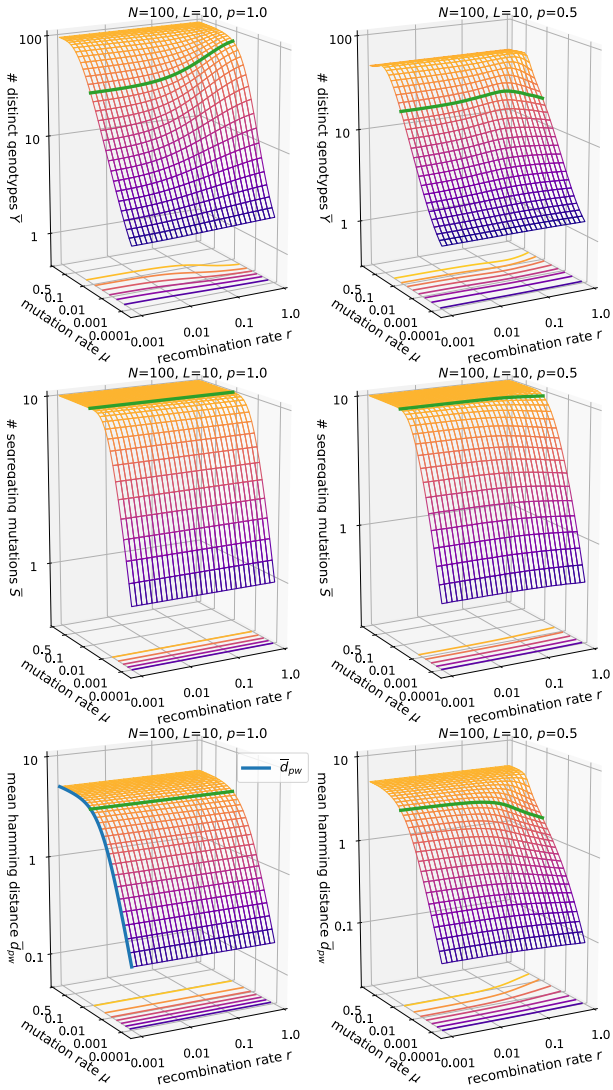

Figure S5 Genetic diversity in the *fsm* with $N = 100$, $L = 10$ and $p = 1.0$ (left column) vs. $p = 0.5$ (right column). The green line is drawn at $L\mu = 0.1$. The blue line in the bottom left panel shows the expression in Eq. 34 for the mean Hamming distance at $p = 1$.

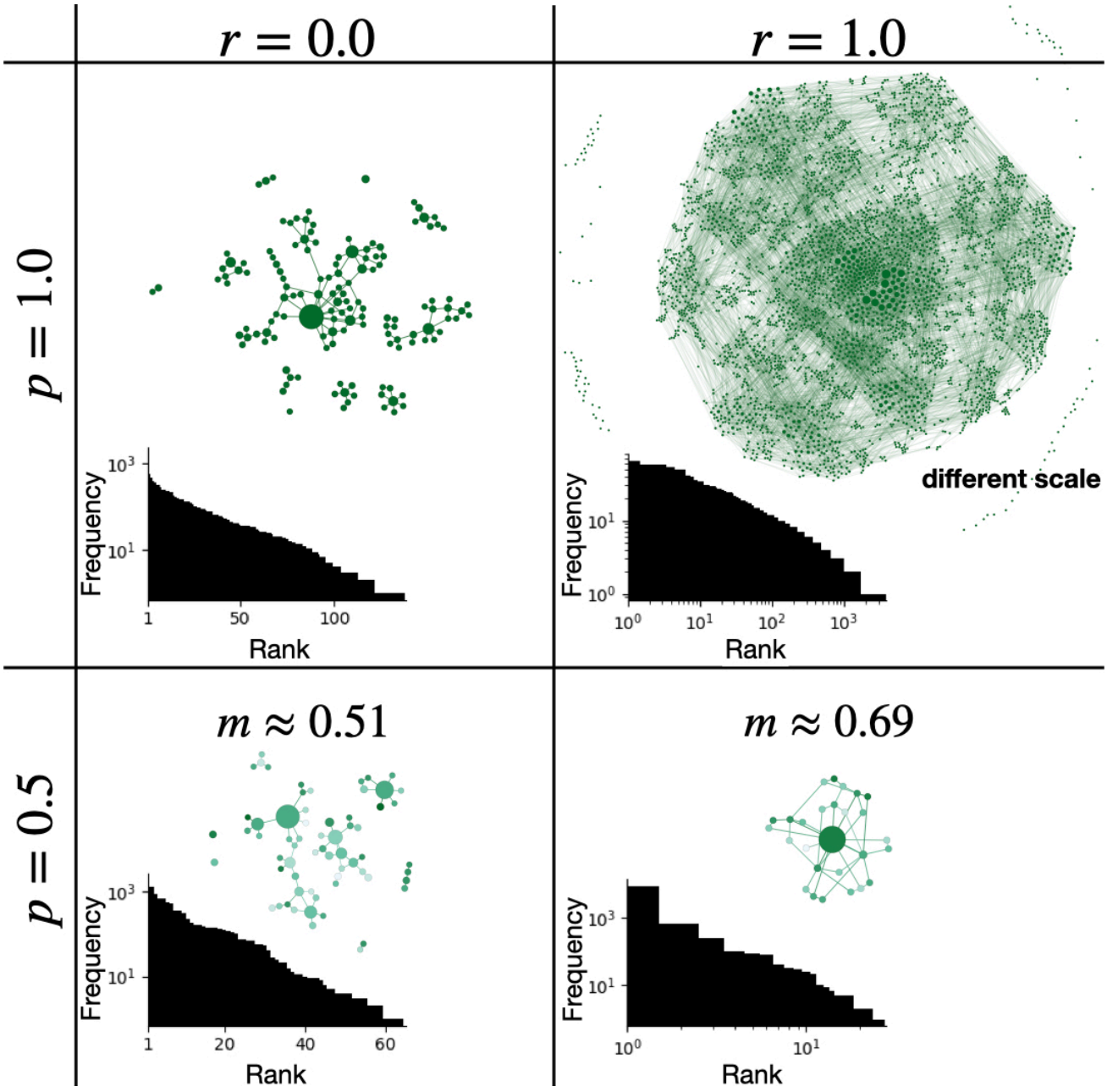

Figure S6 Same as Fig. 13 but with $N = 10000$, $L = 14$, $\mu = 0.0001$ after 10^7 generations. Inset histograms show the frequency distribution sorted by rank. The histograms are in semi-log scale for $r = 0$ and in log-log scale for $r - 1$. Through the number of ranks, the histograms also display the number of existing distinct genotypes. Note that for both values of p recombination makes the frequency distribution heavy-tailed, but it may either increase or decrease the number of distinct genotypes.

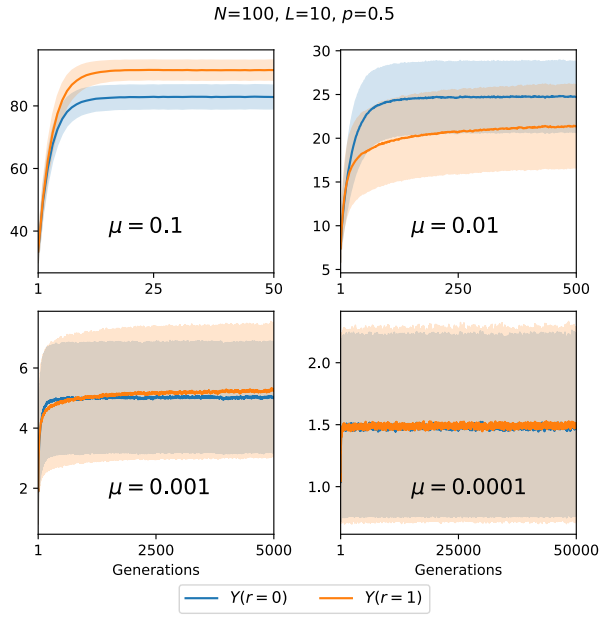

Figure S7 Time evolution of of the number of distinct genotypes in the *fsm* with $N = 100$, $L = 10$, $p = 0.5$ for different values of the mutation rate μ. Each panel compares obligately recombining ($r = 1$) and non-recombining ($r = 0$) populations. Thick lines represent the mean over 5000 landscape realizations and the shaded areas the corresponding standard deviation.

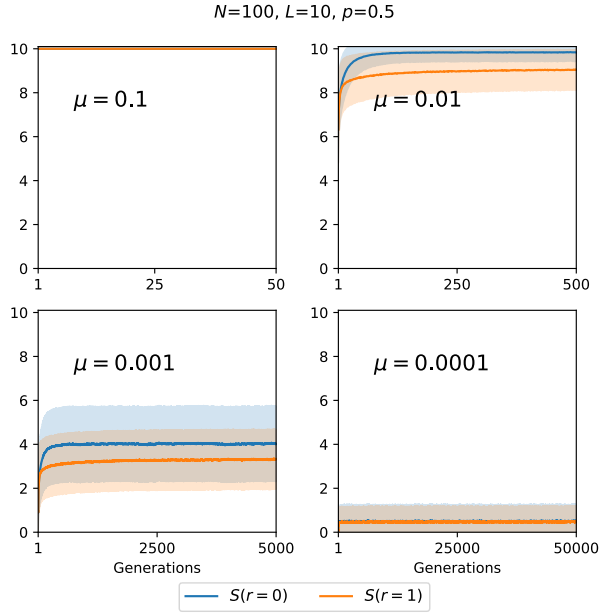

Figure S8 Time evolution of the number of segregating mutations in the *fsm* with $N = 100$, $L = 10$, $p = 0.5$ for different values of the mutation rate μ. Each panel compares obligately recombining ($r = 1$) and non-recombining ($r = 0$) populations. Thick lines represent the mean over 5000 landscape realizations and the shaded areas the corresponding standard deviation.

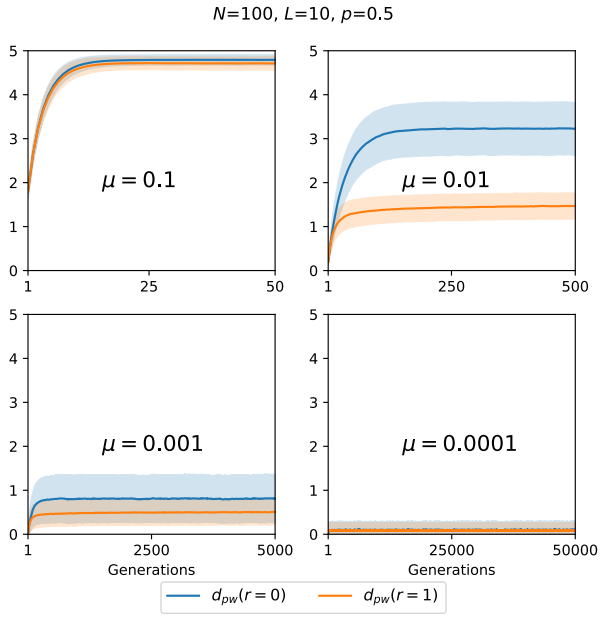

Figure S9 Time evolution of the pairwise mean Hamming distance in the *fsm* with $N = 100$, $L = 10$, $p = 0.5$ for different values of the mutation rate μ. Each panel compares obligately recombining ($r = 1$) and non-recombining ($r = 0$) populations. Thick lines represent the mean over 5000 landscape realizations and the shaded areas the corresponding standard deviation.

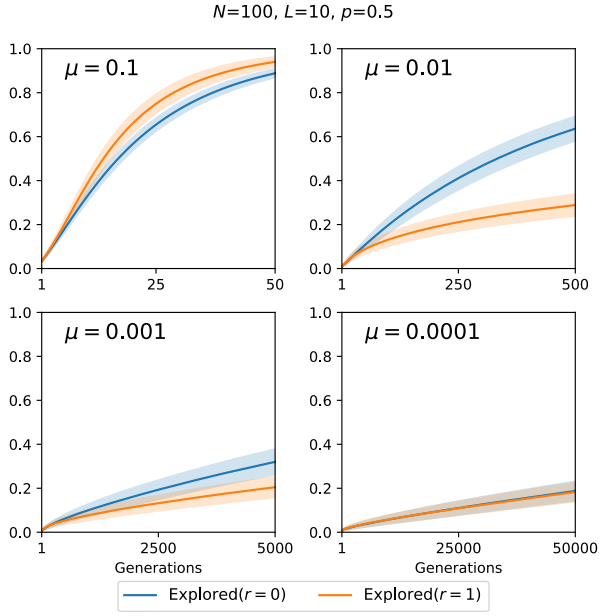

Figure S10 Time evolution of the fraction of explored viable genotypes in the *fsm* with $N = 100$, $L = 10$, $p = 0.5$ for different values of the mutation rate μ. Each panel compares obligately recombining ($r = 1$) and non-recombining ($r = 0$) populations. Thick lines represent the mean over 5000 landscape realizations and the shaded areas the corresponding standard deviation.

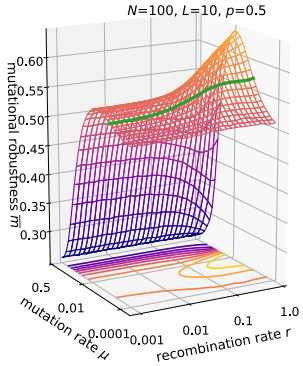

Figure S11 Mutational robustness in the *fsm* with $N = 100$, $L = 10$, $p = 0.5$ and simple successive recombination dynamics. The green line at $\mu = 0.001$ ($NL\mu = 1$) shows a non-monotonic variation with recombination rate, which is caused by recombination-dependent genetic drift.

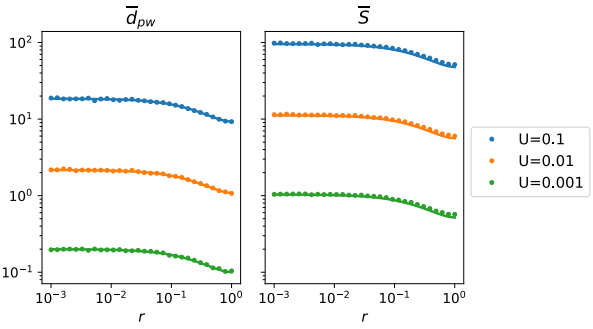

Figure S12 Comparison of numerical results (represented by dots) for the mean Hamming distance $\overline{d}_{pw}$ and the mean number of segregating sites $\overline{S}$ to the analytical expressions in Eqs. 35 and 39 (represented by lines), for the simple successive recombination model. Simulations were carried out using the *ism* with population size $N = 100$ and three different mutation rates. While for $\overline{d}_{pw}$ the fit is perfect, for $\overline{S}$ some deviations are discernible at large r.

4 Recombination in a directed evolution experiment

This section discusses the results of a directed evolution experiment on the effects of recombination, conducted by Diego Pesce in the laboratory of Arjan de Visser, which I have analyzed further. The experimental results are currently unpublished and in the following cited as Pesce and de Visser (n.d.). Therefore, the motivation and the experimental design are explained first. Their general motivation for conducting experimental evolution with recombination is also described in Pesce et al. (2016).

4.1 Motivation

Antibiotic resistance is of great interest, especially in clinical research. β-lactamase alleles, present in *Escherichia coli* and several other bacteria play a vital role for such resistance by creating enzymes that can break down the antibiotic's structure (Cooksey et al., 1990). These enzymes are highly adaptable and can increase their resistance-conferring ability by a factor of several magnitudes through mutations (Salverda et al., 2010). To date, more than 100 variants with different levels of resistance have already been described in the literature (Jacoby & Munoz-Price, 2005; Salverda et al., 2010). One of the most common variant is the β-lactamase allele TEM-1 (Cooksey et al., 1990), which is efficient against several β-lactam antibiotics. However, it only shows low resistance against the antibiotic *cefotaxime* (Schenk et al., 2012). Interestingly, a specific set of five mutations can increase its resistance $\sim$ 100.000 fold against *cefotaxime* (Stemmer, 1994; Hall, 2002). In one of the hallmark papers of evolutionary biology that motivated the field in recent times, it has been shown that the evolutionary path from TEM-1 to the highly resistant allele is tightly constrained (Weinreich et al., 2006). It is constrained in that if evolution is viewed as an adaptive walk in which only beneficial mutations are sequentially fixed, most pathways are inaccessible (108 out of $5! = 120$ pathways are inaccessible). This could imply that evolution is much more predictable than previously thought. While adaptive walks only consider selection and mutation, the motivation of this experiment was to investigate whether recombination is beneficial in the search for a highly *cefotaxime* resistant variant and if so, how. Furthermore, the question of whether the mutation rate and the starting position in sequence space have an influence was investigated. However, the experiment did not attempt to mimic long-term evolution, in which a balance between selection, recombination, and mutation occurs, but was rather a directed evolutionary experiment in which the fittest genotypes were amplified.

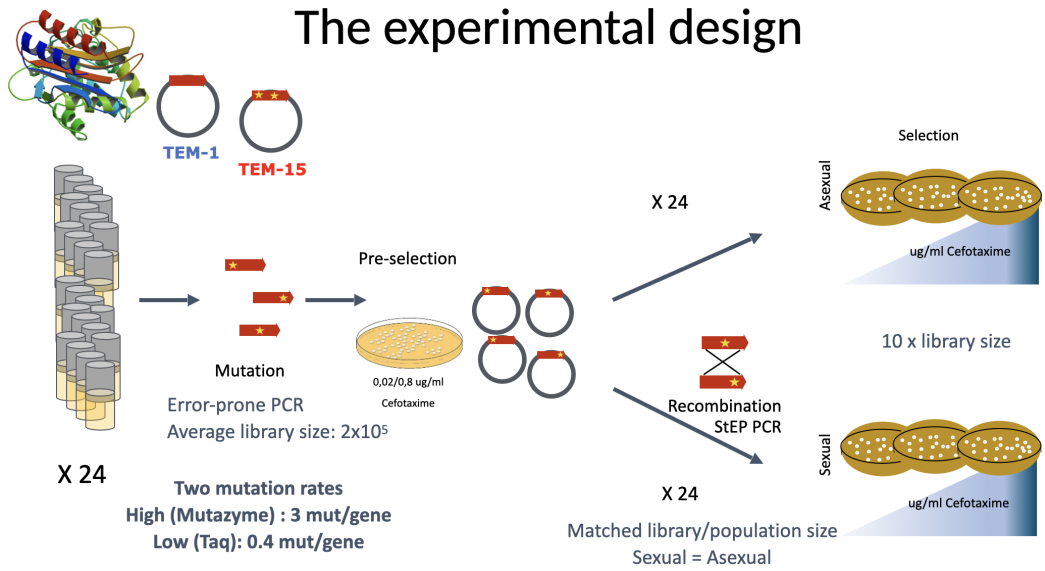

Figure 2: Illustration of the experimental design.
Source: Diego Pesce.

4.2 Experimental design

The experimental design consists of several steps which are sequentially performed (Fig. 2). These steps are explained in the following.

Initial genotype: Each experimental line starts with one of two variants, referred to as TEM-1 and TEM-15. From the point of view of a fitness landscape these are two different initial positions in sequence space that could potentially open different pathways. In the following, we regard TEM-1 as the wild type, having no mutations. From that point of view, the TEM-15 variant has two mutations which are G238S and E104K. These two mutations already increase the resistance ≈64-fold to *cefotaxime*. Their impact on the structure of TEM-1 is explained in more detail in Salverda et al. (2010). In former experiments, TEM-15 often evolved in populations that have been exposed to antibiotic concentrations (Salverda et al., 2011; Schenk et al., 2015). Therefore using these two different initial genotypes can also be interpreted as a comparison of the effect of recombination in early vs. late adaptation.

Mutation: After the initial variant is picked, the evolutionary process starts with a mutation step in which random mutations are introduced in the TEM-1 and TEM-15 variant, respectively, by using error-prone PCR. The mutation rate can be adjusted and is chosen to be either $U_{low} = 0.4\,\text{mut/gene}$ or $U_{high} = 3\,\text{mut/gene}$ for each line. In total, there are four possible configurations, which are also abbreviated as TEM-1-Low, TEM-1-

High, TEM-15-Low, TEM-15-High in the following. This mutation step results in a large library of unknown variants.

Preselection: The mutation step is followed by a preselection step. For this purpose the variants are amplified and introduced to *E. Coli* through a plasmid on which the variant is living on. These plasmids are taken up by *E. Coli*, which allows the bacteria to create TEM variant enzymes. For preselection, the *E. Coli* bacteria are put on a Petri dish with a small *cefotaxime* concentration in order to select for functional TEM variants ($c_1 = 0.02$ ug/ml for initial TEM-1 variant, and $c_2 = 0.8$ ug/ml for initial TEM-15 variant).

Treatment: The experiment then branches off into two different arms. In one of them a recombination step follows and in the other one it does not for comparison. Therefore each line leads to two paired variants, which are the result of the same mutant library, but with two different subsequent treatments (asex/sex). In the case of recombination, the TEM variants which survived preselection are pool-wise recombined. Recombination is introduced in vitro by first extracting the plasmid from the surviving *E. Coli*. The TEM allele is then fragmented by enzymatic digestion. To induce in vitro recombination, incidental template switching during PCR is exploited. The probability for incidental template switching can be increased by abbreviating the polymerase-catalyzed extension. This leads to not fully extended fragments which might bind to other templates during the next annealing phase. After several PCR cycles, the fragments form a full sequence that can have taken up mutations from different TEM variants. This method is called *Staggered Extension Process (StEP)* (Zhao et al., 1998) and a modified version is used for this experiment. Through this protocol, three to five cross-overs per gene are introduced (Pesce & de Visser, n.d.). After the alleles are reassembled to full length, they are again introduced to *E. Coli* through a plasmid.

Greedy selection: In the final step, the fittest variant is selected. For this purpose, in both branches, the bacteria are placed on a series of plates with increasing *cefotaxime* concentrations to select the fittest variant. The initial density of bacteria on the Petri dishes is small enough, such that we can assume that each visible forming colony belongs to one variant. The variant of the largest visible colony on the plate with the highest *cefotaxime* concentration is then picked as the final variant of the respective experimental line. It is then sequenced and its MIC is measured through microtitre plates with a square-root-two-fold increase in *cefotaxime*. In addition to the MIC, the growth rate in the absence of *cefotaxime* was also evaluated compared to TEM-1 and TEM-15 with TEM-15 as initial genotype and only to TEM-1 with TEM-1 as initial genotype.

In total 196 final variants have been measured, as each combination of initial genotype (TEM-1/TEM-15), mutation rate (low/high) and treatment (asex/sex) is performed in 24 experimental lines.

4.3 Results

In the following, the results are analyzed from different angles. In the preliminary note the expectation from simulations for the effect of recombination in the experimental setup is discussed. Afterward, the results are first looked at statistically and then from the point of view of a fitness landscape.

4.3.1 Preliminary note

The experiment employs one iteration of mutation, preselection, and recombination with a final greedy selection step. Therefore, the results for the effect of recombination of the previous chapter cannot be directly transferred since these considered mostly populations in their stationary state. However, the insights from the previous chapters are helpful in interpreting the dynamics of recombination in this setting. Some aspects of the experiment are nevertheless simulated to verify whether the overall behavior is consistent with the expectations and the experimental results.

Simulation model: For this purpose, the Wright-Fisher model is employed including mutation, preselection and recombination with L diallelic loci and N individuals (c.f. chapter 3). The order of the evolutionary forces and the initial condition is adjusted to the experiment. This entails that the population is initially monomorphic, which changes after the mutation step. For the mutation step, each locus mutates with probability μ, such that $U = \mu L$ total mutations per gene occur on average. Subsequently, a preselection step is performed, distinguishing only between functional or non-functional genotypes. In this sense, the population evolves for one generation on an underlying holey neutral landscape. After mutation and preselection, the resulting population is duplicated to create two lines. Only in one of the two lines recombination with a uniform crossover scheme takes place. In the following final greedy selection step, the fittest/most resistant genotype that exists within the population is selected. Both selected genotypes are considered to be the result of one paired line.
This model is quite robust with respect to the exact fitness values because fitness differences between functional genotypes are irrelevant during preselection. Only the structure of the underlying neutral network is essential for the effect of recombination, which is also

discussed in the next paragraph. Moreover, during greedy selection, the only information that matters is which genotype in the population has the highest fitness.

Hamming distance distribution: Even though the experiment illuminates parts of the fitness landscape, most of it remains hidden. Still, certain assumptions or hypotheses can be tested using the described simulation model above. To get a better sense of the impact of recombination under the experimental design, Fig. 3 compares how asexual and sexual populations are distributed around the initial genotype after one generation but before greedy selection takes place. For this purpose, the Hamming distance of all individuals to the wild type is measured in different scenarios, which are described in the figure caption. The scenarios are rather artificial and should only give a general insight.
In the first scenario preselection is left out. In this case recombination would have no effect at all. This is due to the fact that neither genetic drift nor epistasis through selection is at work, so that linkage disequilibrium (LD) cannot arise. Without LD there are no non-random associations and recombination has no effect.
With preselection, but in a scenario in which all genotypes are functional, recombination would not change the mean values of the Hamming distance distribution. However, it would decrease the frequency variability, illustrated by the error bars.
Furthermore, two different scenarios with lethal genotypes are illustrated. In these, recombination changes the shape of the Hamming distance distribution by filling the "frequency holes" torn open by preselection. For "$d > 1$ lethal" genotypes at distance two are created at the cost of genotypes at distance one. Contrary, for "$d = 1$ lethal" genotypes at distance one are created at the cost of of the wild type and larger Hamming distances.
The latter two scenarios demonstrate, that recombination can not only increase the most distant genotypes but can also reduce their frequency. In general, which genotypes gain or lose in frequency through recombination is therefore a question of the underlying neutral network. In this sense, recombination might be favorable or not to create resistant genotypes, e.g. if they are at large Hamming distances. Important to note is that in all cases the population's average Hamming distance to the initial genotype is the same between the sexual and asexual lines, since recombination is only able to reshuffle mutations among individuals.
As a common pattern, recombination does increase the genetic diversity in all scenarios in terms of distinct genotypes within the population, which is shown in the figure legends. This in turn should always be beneficial in the search of the most resistant genotype.

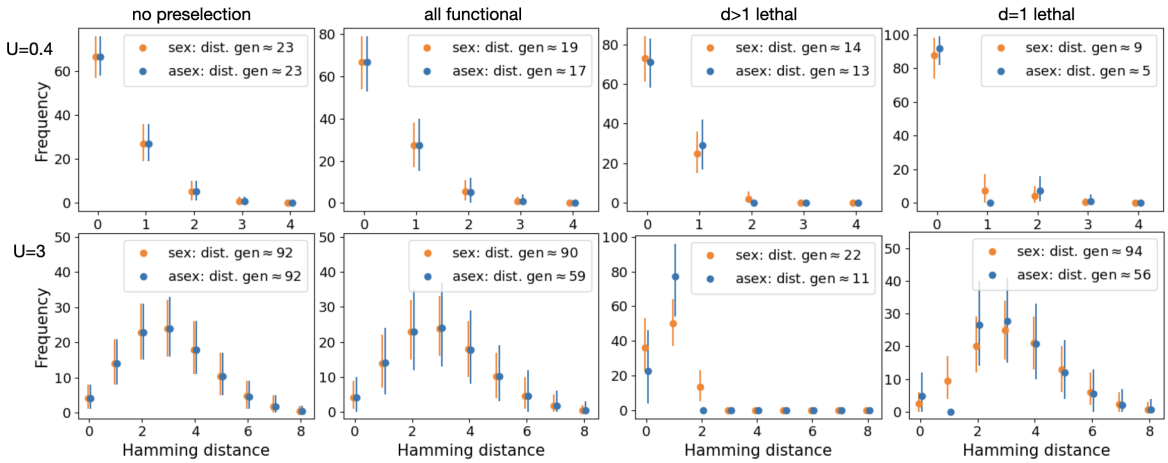

Figure 3: Simulation results for the Hamming distance distribution of the population in different scenarios.
Two different mutation rates, $U = 0.4$ (top row) & $U = 3$ (bottom row) and four different configurations (columns) are considered. The population size is $N = 100$ and the sequence length is $L = 250$. Dots represent the mean of 1000 experiments and bars represent the 2.5% to the 97.5% quantile. In the first column preselection is left out. In the second column all genotypes are functional. In the third column all genotypes above Hamming distance one are lethal. In that sense it is a mesa landscape with critical distance one. In the last column only genotypes at distance one are lethal, which could be illustrated as a ring of lethal genotypes around the wild type. The legends contain as additional information the average number of distinct genotypes for each treatment.

4.3.2 Statistical analysis

The experimental data paints a complex picture for the effect of recombination. Results for the MIC of the selected variants and their number of acquired mutations are discussed in the following.

MIC: In the case of TEM-1, recombination consistently produces a greater increase in final MIC resistance (Tab. 1, Fig. 4). A Wilcoxon signed-rank test further shows that the paired results between asexual and sexual lines significantly differ (Pesce & de Visser, n.d.). Contrary, in the background of TEM-15, the effect of recombination is ambiguous in terms of MIC increase and whether there is a general statistical difference in the samples according to the Wilcoxon signed-rank test. At first glance, this might suggest that the pool of beneficial mutations is already exhausted. If this is the case, recombination would only rarely be able to combine beneficial mutations, such that the difference between both treatments becomes marginal. But other metrics argue against this conclusion.
First, choosing a higher mutation rate for TEM-15 leads to significantly higher MIC values, indicating that many combinations of mutations exist which further increase resistance

(Tab. 1). Second, comparing the MIC fold increases of TEM-15 to those of TEM-1 reveals that evolution does not slow down, c.f. Fig. 4. It is therefore surprising that recombination does not further increase resistance in TEM-15 as it does in TEM-1. However, this contradiction might be resolved if epistatic interactions exist for which the underlying fitness landscape must be considered. This aspect is further investigated in section 4.3.3.

Initial genotype	Mutation rate	MIC [µg /mL]		Same distribution?
		Asex	Sex	
TEM-1	Low	0.36 ± 0.20	0.89 ± 0.38	$p<0.05$
TEM-1	High	0.58 ± 0.32	2.27 ± 1.56.	$p<0.05$
TEM-15	Low	28.64 ± 17.93	25.62 ± 11.64	$p\approx0.30$
TEM-15	High	101.42 ± 99.53	130.66 ± 118.40	$p\approx0.22$

Table 1: Summary for average MIC values.
Shown are the arithmetic mean MIC values with standard deviation for each configuration. The p-value is determined using Wilcoxon signed-rank test.

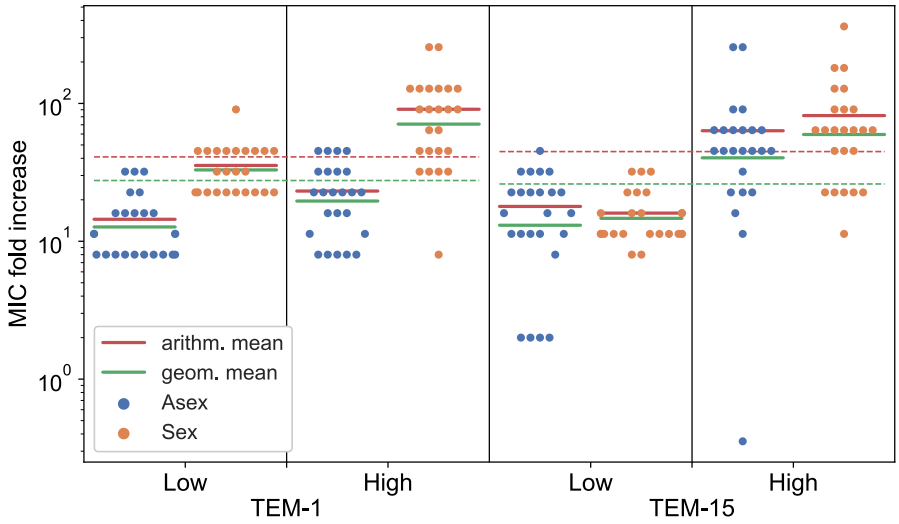

Figure 4: Swarm plots of MIC fold increase.
The illustration shows all data points for each configuration. Besides the arithmetic mean, also the geometric mean is included for each configuration since the MIC has been measured with a multiplicative increase. Depending on the choice of mean, the recombination effect is either marginally positive or negative for TEM-15-Low. To compare the dependency to the initial genotype, the dashed lines represent the mean across both mutation rates and treatments for TEM-1 and TEM-15, respectively.

Number of mutations: Studying the number of acquired mutations of the selected variants, the results show that high mutation rates indeed translate into an increased number of mutations (Tab. 2, Fig. 5). This is to be expected as a higher mutation rate pushes the population further out in sequence space. Still, it demonstrates that resistant variants composed of multiple mutations truly exist around both initial genotypes which not necessarily might be given.

The difference between the asexual and sexual lines is small with a tendency for more mutations in sexual lines. This tendency is stronger for TEM-1 and at high mutation rates. An interesting aspect is that for both treatments fewer mutations occur for TEM-15. Especially at low mutation rates, the small number of mutations in asexual and sexual lines is striking. Possible reasons are discussed in 4.3.3.

Initial genotype	Mutation rate	Mutations		Same distribution?
		Asex	Sex	
TEM-1	Low	1.92±0.91	2.13±0.33	p≈0.23
TEM-1	High	2.79±1.19	3.33±1.21	p≈0.17
TEM-15	Low	1.58±0.64	1.58±0.64	p≈0.97
TEM-15	High	2.96±0.84	3.29±1.24	p≈0.34

Table 2: Summary for the number of acquired mutations.
Shown are the arithmetic mean number with standard deviation of the final variants. The p-value is determined using Wilcoxon signed-rank test.

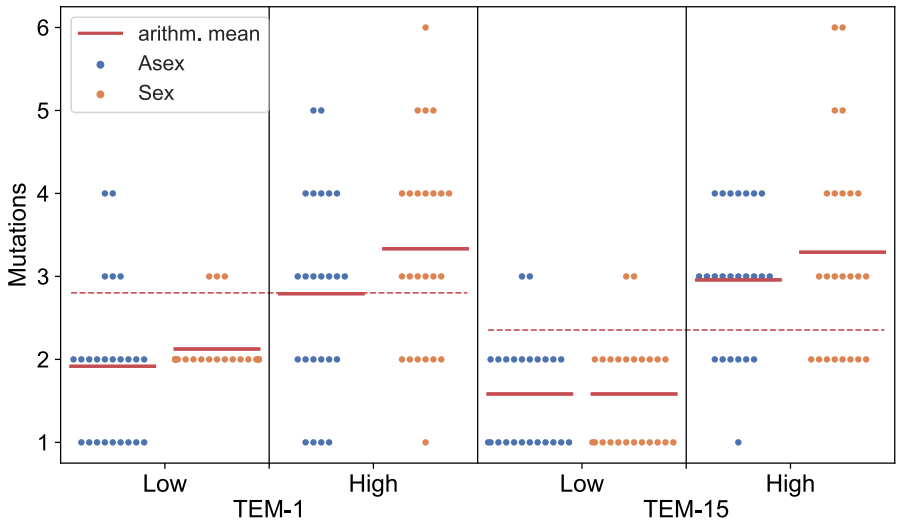

Figure 5: Swarm plots for the number of mutations.
The illustration shows besides all data points the arithmetic mean for each configuration. The dashed lines illustrate the arithmetic mean for TEM-1 and TEM-15, respectively, across both mutation rates and treatments.

Classical recombination effects: Since sexual and asexual lines are paired, it is also possible to evaluate these pairs individually for the effect of recombination (Pesce & de Visser, n.d.). By relating the difference in the MIC between the sexual and asexual line of each pair to the difference of their acquired mutations, one can distinguish between different classical arguments for the effect of recombination. It is then counted how often each effect has occurred among all pairs.
For this purpose, the Fisher-Muller effect, background selection and recombination load are considered. Fisher-Muller effect and background selection occur if the sexual line has a higher MIC. In the former, the number of mutations is larger in the sexual line, indicating that recombination combined beneficial mutations. Contrary, in the latter, the number of mutations is smaller in sexual lines, indicating that recombination broke beneficial and deleterious mutations apart. If the asexual line reached higher MIC, the pair is counted as recombination load. Besides these effects, coupled lines are counted as neutral if both treatments reached the same MIC. If the sexual lines reached higher MIC, but the number of mutations is equal, these are counted as mixed. Results indicate that for TEM-1, the Fisher-Muller effect was most important (Tab. 3). However, for TEM-15 no recombination effect played a dominant role.

Initial	Mutation	Δ MIC>0			Δ MIC=0	Δ MIC<0
genotype	rate	FM	BS	Mixed	Neutral	RL
TEM-1	Low	10	3	9	1	1
TEM-1	High	11	6	6	0	1
TEM-15	Low	1	1	6	7	9
TEM-15	High	9	2	2	4	7

Table 3: Counts for the observed classical recombination effects.
Compared are the Fisher-Muller effect (FM), background selection (BS), mixed, neutral and recombination load (RL) for each configuration.

Growth rate: Besides the MIC also the growth rates relative to the initial genotype have been measured. Results are shown in Fig. 6. Unlike the MIC results, the growth rates do not show any significant increase to the initial genotype. At high mutation rates, they rather show a slight decrease. Between asexual and sexual lines, the results indicate a minor trend for smaller growth rates in sexual lines. Peculiar is that for TEM-15-Low, the growth rates are mostly positive, independent of treatment. This will be discussed in section 4.3.3 in more detail.

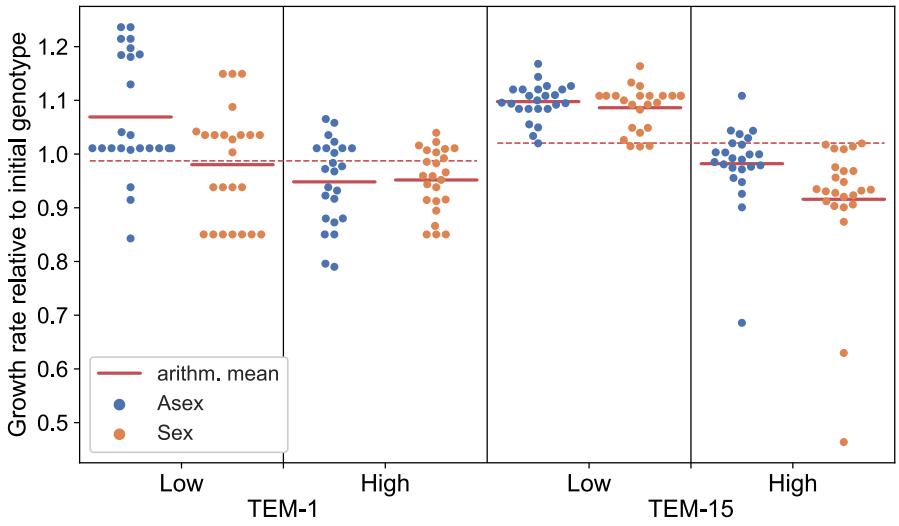

Figure 6: Swarm plots for the growth rate increase relative to the initial genotype.
Besides all data points, the arithmetic mean for each configuration is shown. The dashed lines illustrate the arithmetic mean for TEM-1 and TEM-15, respectively, across both mutation rates and treatments.

Genotypic diversity/Repeatability: High repeatability implies low genetic diversity and vice versa. Therefore, these two measures are linked. Since each final variant is sequenced, the distribution of all variants in sequence space can be analyzed. For this purpose, different measures are considered. These are the number of distinct genotypes, the number of distinct mutations, the entropy and mean Hamming distance. The entropy H is based on the frequencies f_σ of genotypes in the population for a given configuration.

$$H = -\sum_{\sigma} f_\sigma \log(f_\sigma) \tag{7}$$

Results are shown in Table 4. All measures draw a similar picture that evolution is generally more repeatable

1. at low mutation rates as expected,
2. in TEM-1,
3. in sexual lines at low mutation rates.

The second point is a signal that certain pathways have a particularly high probability in TEM-1. Already former experiments have shown that the TEM-15 variant often appears in antibiotic concentrations starting at TEM-1 (Salverda et al., 2011; Schenk et al., 2015). This experiment is no exception, as will also be illustrated in section 4.3.3. The third point is interesting, as it fits the results of chapter 3 on neutral evolution with lethal genotypes:

While at low mutation rates, the mean Hamming distance decreases with recombination, this effect is not observable at high mutation rates.

Initial genotype	Mutation rate	Entropy		#dist. geno.		#dist. mut.		d_{pw}	
		Asex	Sex	Asex	Sex	Asex	Sex	Asex	Sex
TEM-1	Low	2.24	1.79	14	8	22	9	1.99	1.75
TEM-1	High	2.89	3.04	20	22	34	33	3.75	3.54
TEM-15	Low	2.60	2.29	16	15	20	17	2.74	2.10
TEM-15	High	3.17	3.17	24	24	41	50	5.29	5.99

Table 4: Different measures for the distribution of the final variants in sequence space.
In total 127 distinct genotypes have been the final variant (including TEM-15) and 140 distinct mutations have been sequenced across all variants. The entropy is based on the frequency distribution of genotypes.

Combined consideration of MIC fold increase in relation to Hamming distance: Certain conclusions about the underlying fitness landscape can be drawn from the data. For this purpose, in Fig. 7 the MIC fold increase of all distinct genotypes is plotted as a function of their Hamming distance to the initial genotype to obtain a one-dimensional description of the observed fitness landscape. The representation reveals that particularly fit genotypes only exist at larger Hamming distances. Moreover, based on the mean values, there is hardly any reduction in the MIC increase with increasing Hamming distance for both initial genotypes, showing the potential for adaptation. Interestingly, the figure also indicates why the average number of mutations at low mutation rates is higher in the TEM-1 background than in TEM-15 (Fig. 5). This might be due to the fact that in the TEM-1 background only one genotype has been selected for at distance one and it cannot outcompete any other genotype. Contrary, for TEM-15 several mutations have been selected for at distance one of which the most resistant is at least as resistant as a significant fraction (31/72) of distinct genotypes at greater Hamming distances.

Combined consideration of growth rate in relation to Hamming distance: In Fig. 8 the growth rate of distinct genotypes is shown as a function of the Hamming distance to both initial genotypes. In contrast to the results for the MIC (Fig. 7), there is no clear relationship to the number of mutations. However, it is striking that in the case of TEM-15, almost all single mutations increase the growth rate. This is likely due to the fact that TEM-15 shows a significantly below-average growth rate. This in turn explains, why for the configuration TEM-15-Low the growth rate is peculiar above average (Fig. 6). Still, several mutations combined decrease the growth rate again. The kind of mutations that restore growth are discussed in the next section 4.3.3.

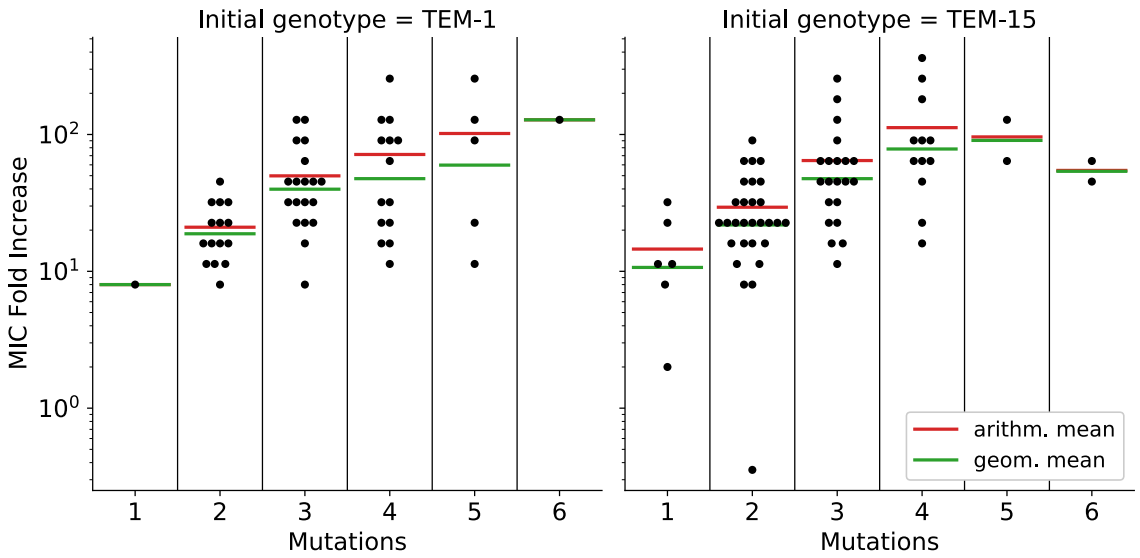

Figure 7: Swarm plots for the MIC fold increase among all distinct genotypes as a function of the Hamming distance.
The Hamming distance is measured relative to the corresponding initial genotype. For TEM-1, 55 distinct genotypes have been measured and 72 for TEM-15.

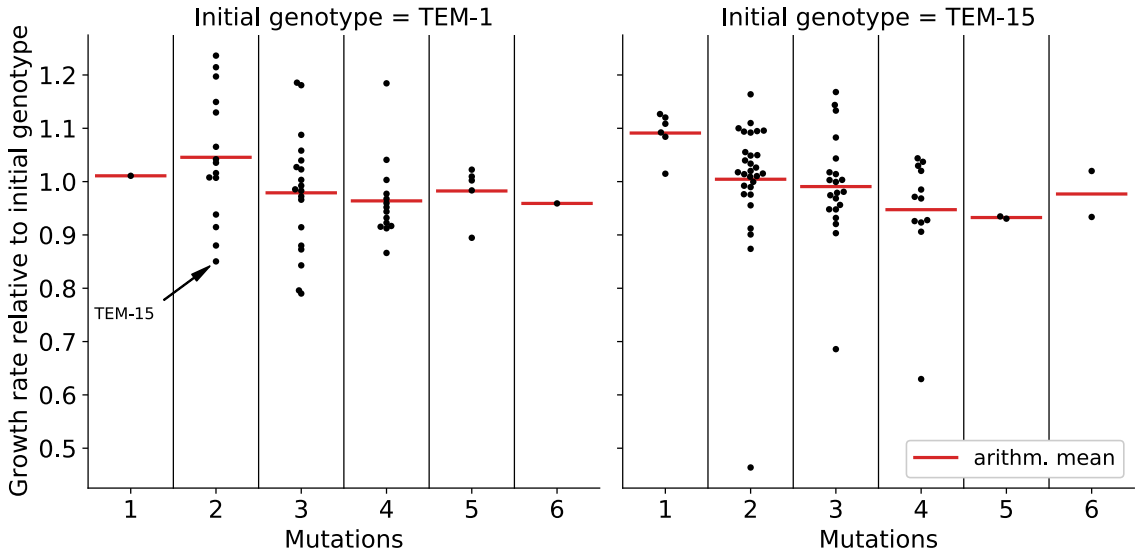

Figure 8: Swarm plots for the growth rate increase among all distinct genotypes as a function of the Hamming distance.
The Hamming distance is measured relative to the corresponding initial genotype. Notably, TEM-15, marked by an arrow, has one of the lowest growth rates compared to TEM-1.

4.3.3 Fitness landscape of β-lactamese

As discussed in the previous sections, details of the underlying fitness landscape are crucial to understand the particular observed effect of recombination. To get a grasp of the high-dimensional structure, in the following a graph representation is employed to create a two-dimensional visual impression of the observed fitness landscape. For this purpose, the measured MIC values and growth rates of all distinct genotypes are used. It should be noted that only the most resistant genotypes are measured, such that there are potentially many more genotypes that are functional but which have not been measured. Nevertheless, it serves a better understanding, as will be shown. In the following, the graph is explained first and subsequently different findings are discussed.

Graph representation: In the graph, each node represents a genotype that has been discovered at least once across all configurations. The node's color represents either the genotype's MIC value or growth rate. The frequency of a genotype for a certain configuration or across all configurations is encoded through the node size. Furthermore, a pie chart around each genotype illustrates the ratio between its frequency in asexual to sexual lines.
Edges connect nodes that differ by a point mutation. The arrow of each edge points to the genotype containing the mutation indicated on the edge label. The genotype marked with the arrow also contains all other mutations of the genotype from which the arrow originates. Therefore, in order to know all mutations of a particular genotype, it is necessary to follow the arrows starting from TEM-1. This also implies that arrows generally point to the genotypes that have a greater Hamming distance to TEM-1. Dashed edges are between genotypes that only differ at the same locus. Thus they have the same Hamming distance to TEM-1 and no arrow is shown. Edges of mutations that lead to a 10 fold or stronger MIC increase or to a at least 0.2 change in growth rate are drawn in orange to illustrate high impact mutations. Important to note is that the arrows should not signal any time information in terms of mutations being sequentially selected.
The nodes are first arranged according to a force-based algorithm as a starting point. This leads to a representation in which genotypes with many edges form visual clusters. Genotypes without edges drift to the periphery and are randomly distributed there. Subsequently, the nodes' positions have been manually adjusted for better visibility of the edge labels. After that, visible clusters are highlighted with a background color. The color is determined by the central genotype of the cluster, which is defined to be the one with the most edges. For each cluster the mean Log[MIC] and its standard deviation are determined. Finally, genotypes that do not have any edges but are at least at Ham-

ming distance two to a central genotype of these clusters are placed manually around the plateaus. Next to them are written the two mutations in which they differ from the central genotype.
The frequency of genotypes across all configurations with either the MIC values or growth rates are shown in Figs. 9, 10. The frequency of genotypes for the individual configurations are shown in Figs. 11, 12, 13, 14. The network structure is kept the same across all figures and only the node sizes change according to the illustrated configuration. Edges are drawn in all cases to provide orientation. Since TEM-1 never occurs, it is illustrated by a star.

Connectivity: Through the procedure described above, most genotypes can be arranged in a meaningful way with respect to each other (103 of 128 distinct genotypes). The remaining genotypes (25) are randomly distributed on the periphery. Of those meaningful arranged nodes, the majority form one large connected component (73 out of 103). The remaining nodes (30) are at Hamming distance two from the central node of one of the clusters. The large connectivity of the resulting graph is a property arising through the high-dimensional sequence space. With a different alleles per loci, the number of genotypes grows with a^L with sequence length L, while the number of edges between genotypes at Hamming distance one grows faster with $(a-1)La^{L-1}$. Therefore, as long as the observed genotypes are not distributed too broadly in the sequence space or their number is too small, it is to be expected that large components will form.

Cluster: While most genotypes appear only once, there are a few genotypes that occur comparatively often and in this sense seem to have a large basin of attraction. These coincide with the central nodes of the highlighted clusters. Since the nodes within most highlighted clusters have similar MIC, this is a strong indication that the high-frequency nodes really contain the mutations that drive evolution. The high frequency genotypes are G238S, G238S-E104K, and G238S-E104K-X, where X stands for a signal peptide mutation. One exception with regard to MIC similarity is the cluster consisting of the mutations G238S-E104K-A11*, which shows increased variability. Besides the small G238S-E104K-L12* cluster, it is the only signal peptide cluster with a synonymous mutation (Fig. 9).

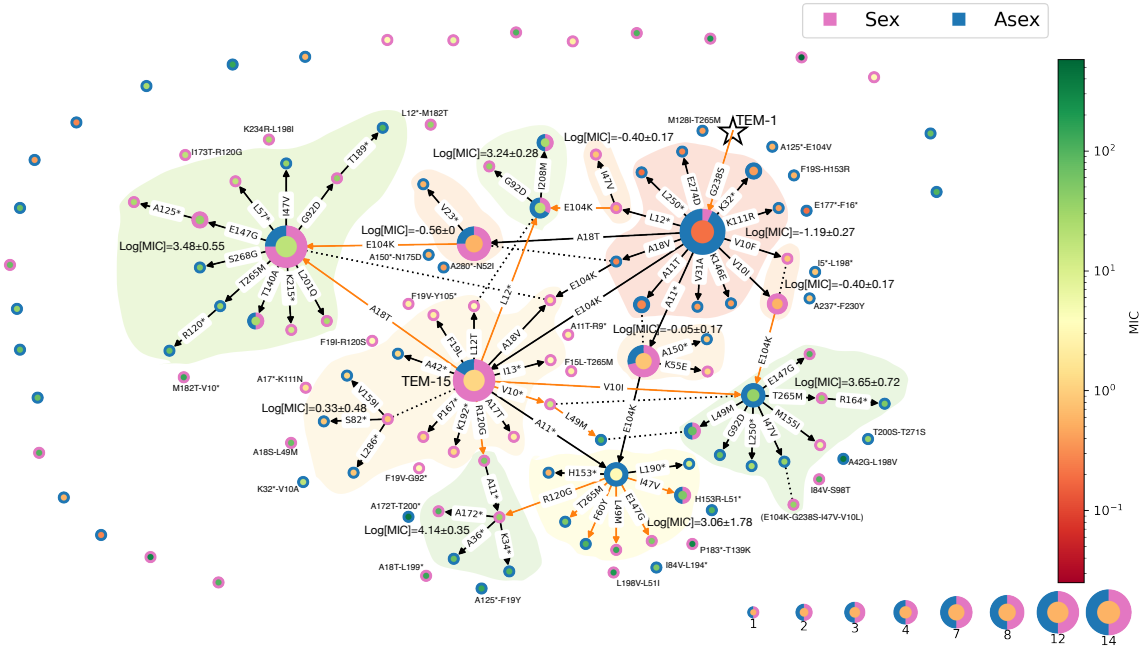

Figure 9: Frequencies of genotypes across all configurations.
The colors represent absolute MIC values.

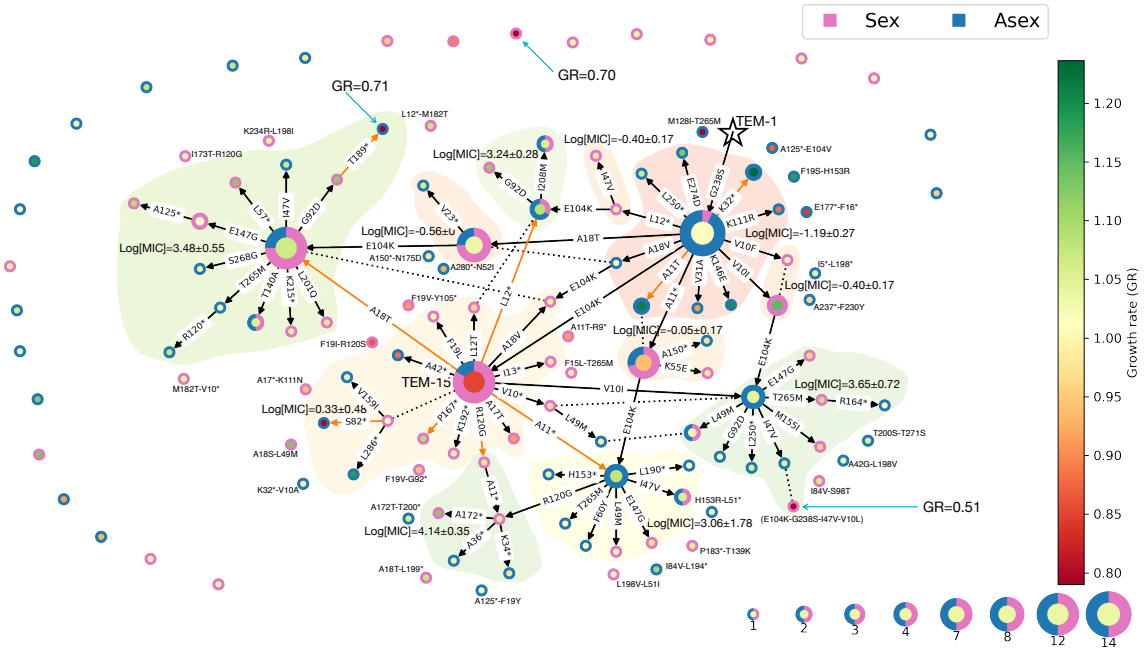

Figure 10: Frequencies of genotypes across all configurations.
The colors represent growth rates relative to TEM-1. Three outliers are off the scale and marked with a blue arrow.

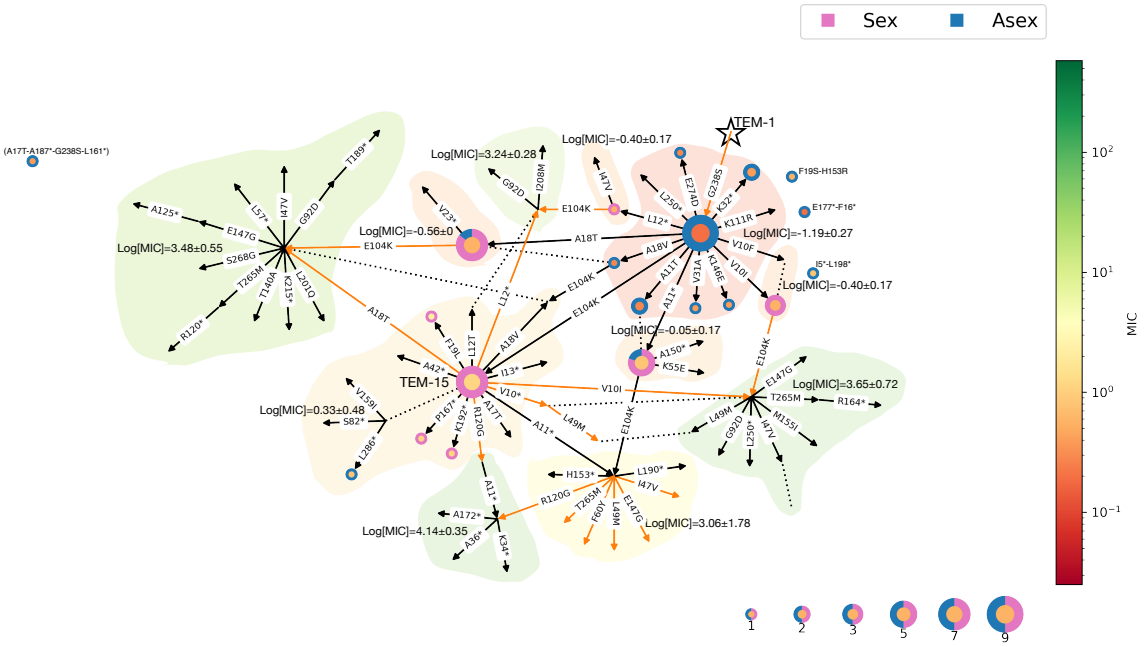

Figure 11: Frequencies of genotypes for the configuration TEM-1-Low.

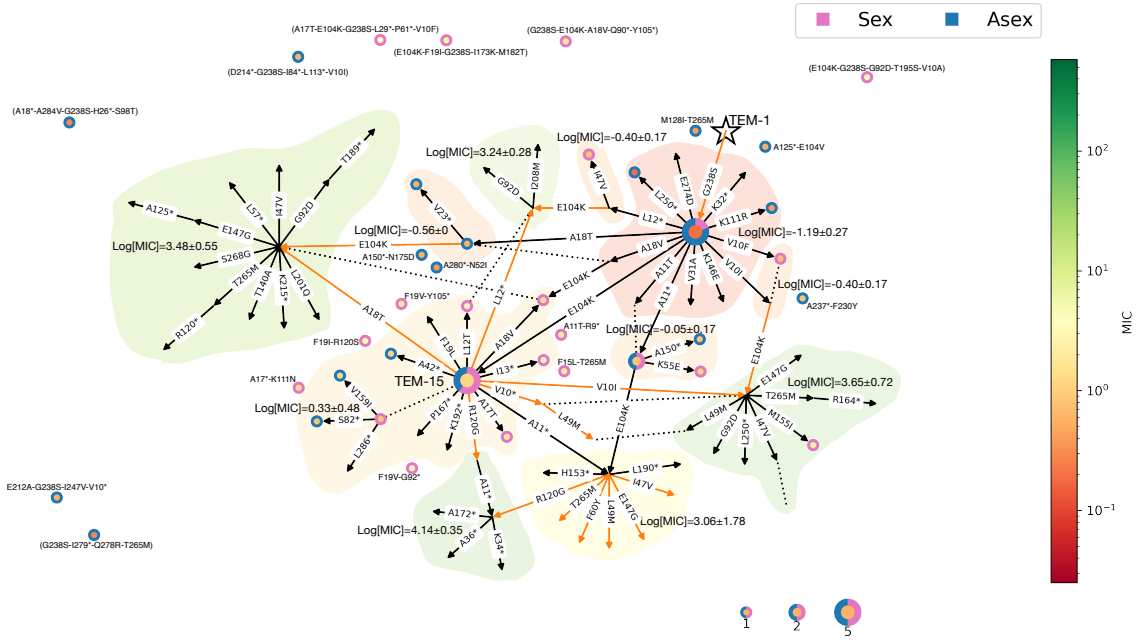

Figure 12: Frequencies of genotypes for the configuration TEM-1-High.

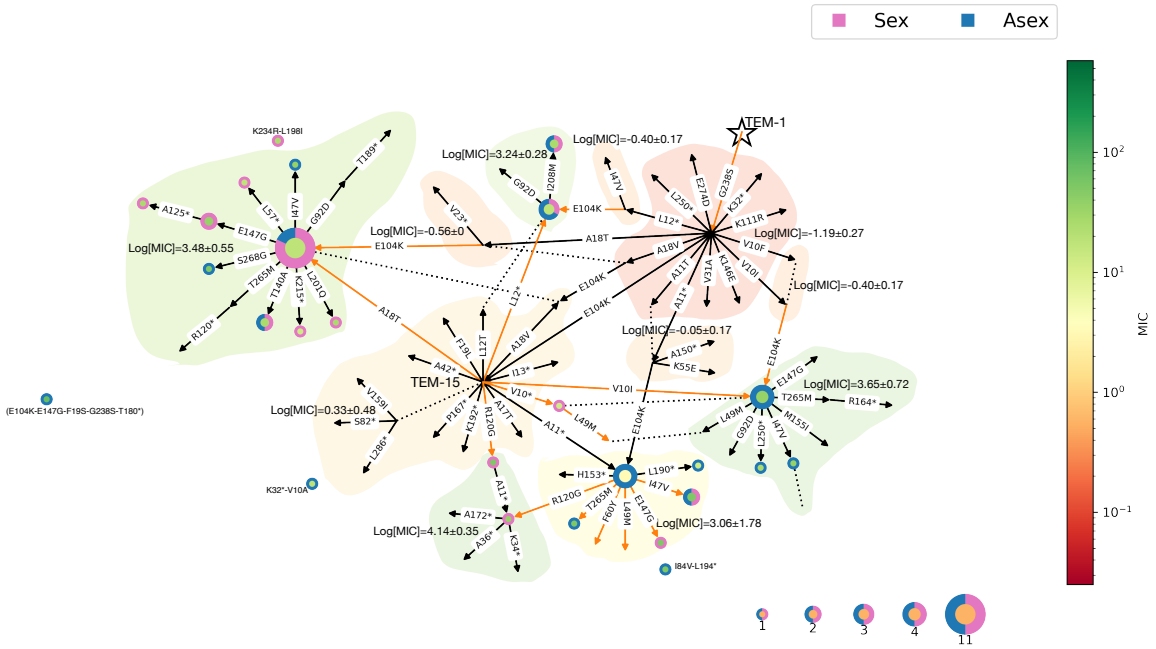

Figure 13: Frequencies of genotypes for the configuration TEM-15-Low.

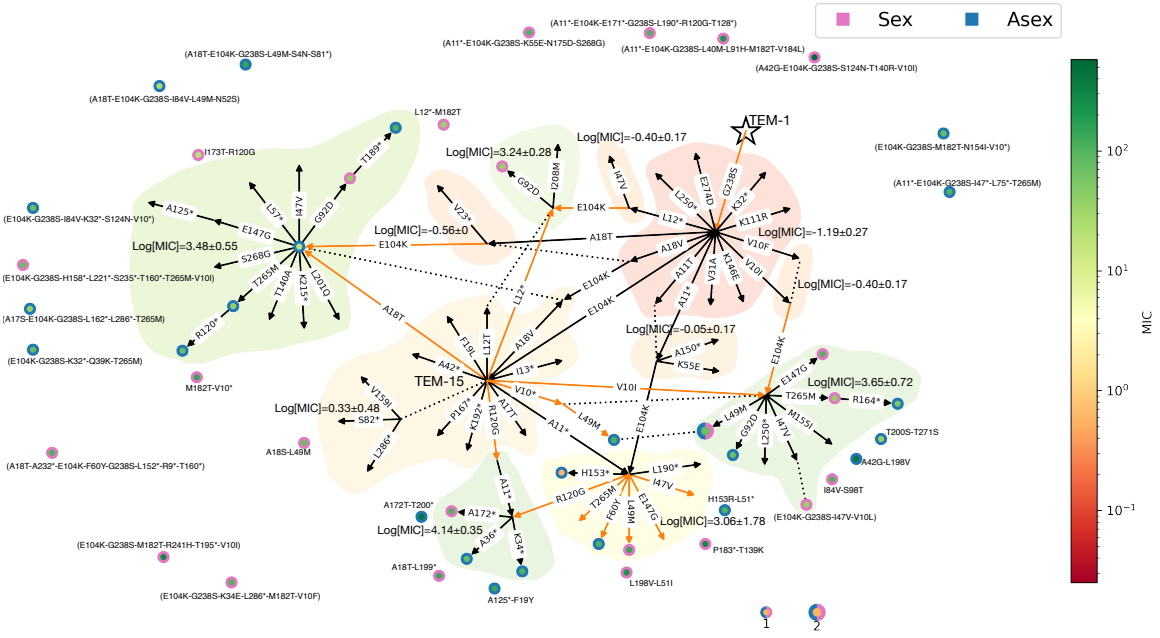

Figure 14: Frequencies of genotypes for the configuration TEM-15-High.

Signal peptide mutations: The graph illustrates that signal peptide mutations are of high significance. Especially the mutations A18T, V10I, L12* have a particularly strong effect on MIC as indicated by the orange arrows (Fig. 9). Interestingly, their effect seems to be enhanced in the presence of the E104K-G238S mutations, which can be noticed by comparing arrows of those signal peptide mutations outgoing from TEM-15 to those outgoing from the G238S variant. Furthermore, the graph shows that in the background of TEM-15, all single point mutations except for R120G, K192*, and P167* are signal peptide mutations. Moreover, all genotypes at distance two around the TEM-15 cluster also have a signal peptide mutation. This indicates that they are essential for further adaptation in the background of TEM-15. Results on the growth rate indicate why this might be (Fig. 10). While TEM-15 has a significantly below-average growth rate, all signal peptide mutations are able to restore the growth rate to around average levels. This indicates that selection for growth may also have taken place in the background of TEM-15. It might also explain why the mutation A11*, which has one of the strongest effects on the growth rate, occurs several times in the background of TEM-15, although the MIC increase is rather small compared to e.g., A18T, V10I, L12* (Fig. 9, 10). Moreover, it might explain why the A18T mutation is the most common mutation in the TEM-15 background, as it combines one of the largest increases in MIC with one of the strongest increases in growth rate. However, it does not rank first in either measure. More observations concerning signal peptide mutations are listed in the following, indicating that they are especially important for further adaptation in the background of TEM-15.

- Signal peptide mutations occur in 146/192=0.76 lines.
- 52/96=0.54 variants in the TEM-1 background acquire a signal peptide mutation.
- 16/19=0.84 variants in the TEM-1 background that have acquired the E104K-G238S mutations plus other mutations have a signal peptide mutation amongst them. Exceptions are E104K-G238S-P167* & E104K-G238S-K192* & E104K-G238S-A42*.
- 94/96=0.98 variants in TEM-15 background acquire a signal peptide mutation.
- The V10I mutation exhibits the strongest increase in MIC as a single point mutation in the TEM-15 background. It is also present in the most resistant variant E104K-G238S-V10I-A42G-S124N-T140R.
- In descending order, the following signal peptide mutations have the greatest positive influence on the MIC in the background of TEM-15:
  1.V10I 2.A18T & L12* 3.V10*.

- In descending order, the following signal peptide mutations have the greatest positive influence on the growth rate in the background of TEM-15.:
  1.L12* 2.A11* 3.A18T.

Dark Matter around TEM-15-A18T?: While a combined consideration of growth rate and MIC provides an idea of why the mutation A18T is most frequently selected in the background of TEM-15, the question remains open as to why it is disproportionately selected in the sexual lines at low mutation rates (Fig. 13). This is surprising because it is a single point mutation and recombination thus does not produce a benefit in the sense that mutations are combined. One explanation we would like to point out for this is that genotypes consisting of the TEM-15-A18T mutations among potentially other mutations could be more often functional than, for example, genotypes consisting of the TEM-15-V10I mutations, which is more resistant but could therefore have a disadvantage if recombination occurs. This difference in robustness might not be observed, since genotypes with TEM-15-A18T plus additional mutations might be functional, but could have a smaller MIC compared to TEM-15-A18T. In that sense, it would be dark matter, which cannot be observed within the experimental design. Translated onto a neutral landscape, this would imply that the TEM-15-A18T genotype has significantly more functional point mutation neighbors than the TEM-15-V10I genotype and is thus more robust to mutations. If this were the case, double mutations containing A18T could survive the preselection step and then recombine with TEM-15, resulting in a higher probability of observing the A18T single point mutation genotype. With respect to the recombination weight explained in chapter 1, it can also be argued that due to the higher robustness of TEM-15-A18T, its recombination weight increases and, importantly, is greater than that of TEM-15-V10I. This in turn increases its frequency in the sexual lines.
A reduced model landscape for this hypothesis is shown and explained in Fig. 15 along with the results of the simulation. For the simulation the protocol explained in 4.3.1 is used. The results show that such an observation could indeed be made. However, only with a relatively small population, since with a sufficiently large population, the fittest genotype would be discovered every time. Moreover, the frequency of the fittest genotype would in any case exceed that of the more robust genotype. A possible explanation why this is not the case in the experiment could be that the growth rate of TEM-15-A18T is higher. This is not taken into account in this simplified model. However, within the experimental design the growth rate could have an effect on the distribution for example during preselection. This could be addressed by accepting a deviation from the neutral landscape assumption during preselection, but this has not yet been verified.

Fisher-Muller effect in the case of TEM-1: The results for TEM-1 as initial genotype are shown in Figs. 11, 12. They illustrate why the number of observed mutations and the MIC is greater in the sexual lines, leading to a strong signal for the Fisher-Muller effect. Because whereas asexual lines mostly carry only the G238S mutation and sporadically another mutation that does not have a major effect, in sexual lines, the G238S mutation are often combined with E104K. These two mutations together have the largest fitness effect of all genotypes found at Hamming distance two. The effect is particularly visible at low mutation rates. At high mutation rates, G238S-E104K is also observed in some asexual lines. Therefore, the Fisher-Muller effect really seems to be at work in this case. Moreover, in our simulations, we can reproduce a similar effect. For this, we assume the simplified model landscape shown in Fig. 16.

4.4 Discussion

The results demonstrate that the effect of recombination clearly depends on the initial TEM variant. In general, as discussed, the particular effect of recombination hinges on the underlying fitness landscape. Therefore, an attempt was made to characterize the fitness landscape using the observed sequenced genotypes. The results strongly indicate that the local landscape structure around both initial variants differs, which could explain the different recombination effects. While for TEM-1, only the G238S mutation at distance one seems to offer a clear fitness advantage, for TEM-15, there are numerous mutations at distance one that are selected. The results also suggest that for TEM-15 not only the MIC but also the growth rate plays a role. In a future work the performed simulation might be improved if these are also taken into account. However, further complications could arise at this point since the growth rates are only measured in the absence of *cefotaxime*, while these usually are concentration-dependent (Ruelens & de Visser, 2021). In terms of evolvability, the results suggest that evolutionary pathways tend to be narrow in early evolution (TEM-1), but become broader once the TEM-15 mutations and a signal peptide mutation are acquired. Especially for TEM-15-High there seems to be no basin of attraction anymore (Fig. 14).

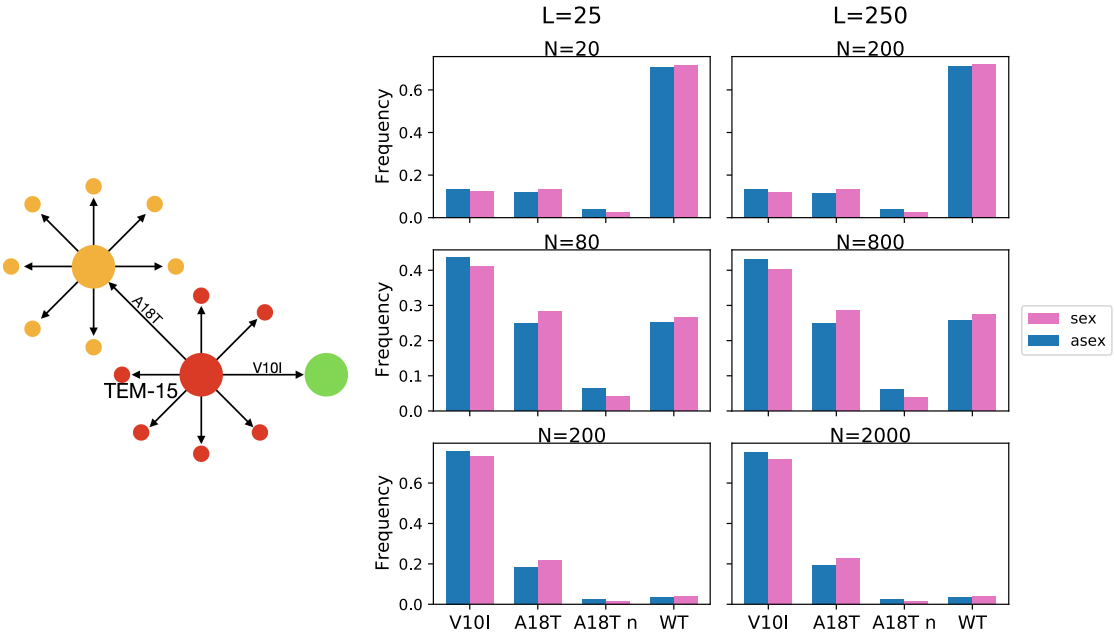

Figure 15: TEM-15 model landscape and simulation results.
The left figure is a sketch of the model fitness landscape for $L = 8$. In the model landscape, it is assumed that all L point mutations of TEM-15 are functional. Two point mutations are special. One of these is A18T, whose own point mutation neighbors are also all functional. The other special mutation is V10I, which has no additional functional neighbors but has the highest fitness. This is crucial for the greedy selection step. Thus, V10I is selected in the experiment if this genotype is present in the population after one generation. The second highest fitness belongs to A18T, followed by its point mutation neighbors. If these genotypes are also not present, TEM-15 is drawn. MIC is the proxy for fitness in this case and growth rate was not included. The results of numerical simulations are shown in the figure on the right for different population sizes N and sequence lengths L. "WT" stands for TEM-15 and "A18T n" for any point mutation neighbor of A18T that is not TEM-15. The mutation rate is $U = 0.4$ (low mutation rate) in all cases. The results reveal that the observed frequency of A18T is higher in recombining populations and that the results depend only on the ratio between N and L in this scenario.

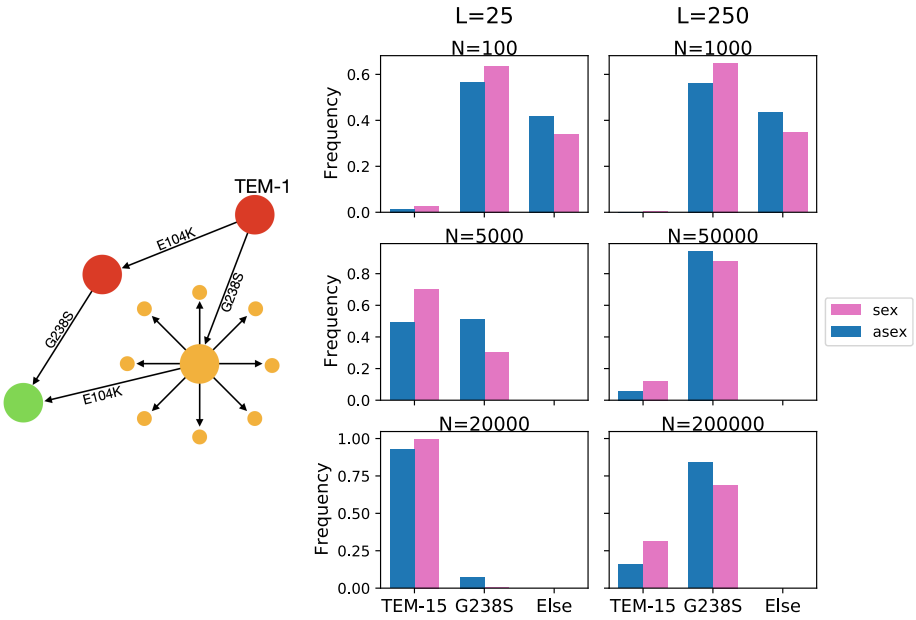

Figure 16: TEM-1 model landscape and results.
The left figure is a sketch of the model fitness landscape for $L = 10$. Only two point mutations of TEM-1 are functional, namely E104K and G238S. The latter has higher fitness than TEM-1 and E104K. Moreover, all L point mutations of G238S are functional. The combination E104K and G238S leads to another functional genotype which is also the fittest one. Results are illustrated on the right for different population sizes N and sequence length L. The mutation rate is $U = 0.4$ (low mutation rate). "Else" represents either a E104K or TEM-1 observation.

5 Summary and outlook

In this thesis, evolution on fitness landscapes was studied under selection, mutation, and recombination dynamics.

In chapter 2, quasispecies were used to address in particular the effect of recombination on mutational robustness. The results demonstrate that recombination increases mutational robustness in a broad parameter range. Explicit equations were derived for a two-locus model. These show that in this model, mutational robustness depends on the r/μ ratio in certain limits. This can be interpreted to reflect that higher mutation rates μ distribute the population more evenly in sequence space while increased recombination rates r return the population to particularly robust genotypes. Additionally, three different multi-locus models were employed: the percolation, mesa, and sea-cliff model. While the former two already exist in the literature, the latter one was introduced as an intermediate model containing the former two models as limiting cases. For all multi-locus models, robustness was shown to increase with r in a fashion similar to the two-locus model. Even the choice of recombination scheme does not fundamentally change this behavior, as demonstrated. Moreover, additional analytical results could be derived for the mesa model in limiting cases, displaying good agreement with the numerical simulations. These illustrate that selection alone moderately increases robustness, but that recombination has a much greater effect, leading to values near the maximum robustness for a wide range of conditions. For the percolation model, we discussed that several distinct stationary states can exist in the presence of recombination. These are usually characterized by a highly concentrated population in a well-connected region in sequence space, leading to increased mutational robustness. If there are several such well-connected regions, the recombining population moves into one of them. In which one depends on the initial placement of the population in the sequence space. Furthermore, the recombination mechanism is explained under the notion of fitness landscapes by introducing the recombination weight as a new measure. It is shown that this measure predicts well which genotypes might reach high frequency at stationarity and that there is a positive correlation with the robustness of genotypes. Finally, we consider an empirical landscape in which increased selection for mutational robustness by recombination is likewise observed, but at the cost of the population's average fitness.

In chapter 3, finite population sizes N were considered in a percolation model and in addition to mutational robustness, the population's evolvability and genetic diversity were taken into account. While the quasispecies approximation of chapter 2 is appropriate if $N\mu L \gg 1$ and $N \gg 2^L$, here we deliberately chose not to meet these criteria. We therefore considered, on the one hand, the limit of infinitely long sequences ($L \to \infty$) with finite

$U = \mu L$ and, on the other hand, finite genotype spaces which are larger than the chosen population size ($N \ll 2^L$). At first, different evolutionary regimes that arise due to the finiteness of the population were discussed. Depending on the product NU, the population can be either considered as a random walker ($NU \ll 1$) or as a diffusive genotype cloud ($NU \geq 1$). Subsequently, the effect of recombination was studied in a wide parameter range, such that all evolutionary regimes are covered. In general, the results show that recombination is particularly significant for the parameter range $NU \geq 1$ and $U \ll 1$, as the population is otherwise either too monomorphic or dominated by the entropic force of mutation. Results for the infinite-sites model demonstrate that in the absence of lethal genotypes, evolvability in terms of the discovery rate and genetic diversity in terms of the number of distinct genotypes only increases with the recombination rate. Other metrics like the fixation rate, the number of segregating mutations, and mean Hamming distance remain independent of recombination. However, it was shown that this changes once lethal genotypes are present. Then several non-monotonicites arise. Most remarkably are those that are a function of the recombination rate. While for small recombination rates, the discovery rate and the number of distinct genotypes increase as in the absence of lethal genotype, they quickly drop again at large recombination rates. A more detailed analysis shows, that at large recombination rates, the genotype composition of the population changes, such that recombination events more often generate viable genotypes. This is achieved if the genotype cloud is highly concentrated on a focal genotype, from which mean fitness benefits. But evolvability and genetic diversity decline across all metrics, showing a clear trade-off. Results for the finite-sites model draw a similar picture. Here, we could also study mutational robustness, which shows very similar behavior to the quasispecies regime for $NU \geq 1$ and $U \ll 1$. This demonstrates once again that the increase of mutational robustness due to recombination exists under many conditions. In the end, the difference between three model implementations of recombination was discussed. While the model details would not matter for quasispecies with finite sequence length, the choice can become important otherwise. Depending on the implementation, recombination could act as an additional source of genetic drift besides selection, which gives rise to further non-monotonicities.

In chapter 4, the results of an experiment on the evolution of the antibiotic resistance enzyme TEM-1 β-lactamase in *E. Coli* was studied. The experiment is interesting in the context of this thesis, as it compares evolution with and without in-vitro recombination. A statistical analysis of the measured MICs, growth rates, and number of mutations initially leaves open questions about the particular effect of recombination. While for TEM-1 as a starting point the Fisher-Muller effect seems to play a significant role, the effect for TEM-15 is unclear. Therefore, the sequenced mutations were utilized to reconstruct the

observed underlying fitness landscape, which was then illustrated as a directed graph. In this graph, genotypes are represented as nodes, and although many intermediate nodes are missing, the majority of the measured ones could be arranged relative to each other in a meaningful way. Further information such as the frequency of each measured genotype and the ratio between asexual and sexual lines were encoded in the graph, such that almost all experimental results can be read from it. On the basis of this graph, we recognized certain patterns, which were subsequently discussed in more detail. Simulations were also employed, showing that for TEM-1 the Fisher-Muller effect was dominant because of the underlying fitness landscape structure. Contrary, for TEM-15, the results indicate that recombination selects for mutational robustness. However, simulations could not clearly confirm this. We suspect that the growth rate must also have played a role at this point, but we did not include it in the simulation.

For further studies, it would be interesting to investigate whether stronger signals for mutational robustness can be detected in evolutionary experiments. At the moment, in vitro recombination is still a complex task, and therefore most evolutionary experiments only consider mutation and selection. However, this might change, and there are already a few limited studies on the effect of recombination in evolutionary experiments that probe theoretical hypotheses (Cooper, 2007; McDonald et al., 2016). Still, the recombination rate was not modified in these experiments and only the fitness increase was considered. On the theoretical side, it would be interesting to incorporate deviations from the assumption that fitness is equal among all viable genotypes. Depending on the origin of the neutrality, one could consider, on the one hand, fitness landscapes with rugged fitness plateaus or, on the other hand, several flat fitness plateaus of different heights that are distributed in sequence space. The former could, for example, represent the fact that even synonymous mutations show fitness differences, while the latter scenario could represent the degeneracy of different phenotypes in sequence space (Zwart et al., 2018; Manrubia et al., 2021). These scenarios could then also be combined. Since environmental conditions change frequently, it would also be interesting to study the population dynamics with plateaus that have time-dependent heights. In this context, one could consider the effect of recombination on the mutational robustness and average fitness increase simultaneously. Furthermore, in this thesis, constant populations were assumed. It would therefore be interesting to investigate what effects bottleneck situations could have on neutral landscapes with recombination. Moreover, only haploid populations with diallelic loci were considered. A generalization would be interesting to verify whether the effect of recombination on mutational robustness persists.

A References

Aristotle. (4th century BC/1991). *History of animals* (Vol. 3). Harvard University Press Cambridge, MA.

Baake, M., & Baake, E. (2003). An exactly solved model for mutation, recombination and selection. *Canadian Journal of Mathematics*, *55*(1), 3–41.

Bell, G. (1988). Recombination and the immortality of the germ line. *Journal of Evolutionary Biology*, *1*(1), 67–82.

Blythe, R. A., & McKane, A. J. (2007). Stochastic models of evolution in genetics, ecology and linguistics. *Journal of Statistical Mechanics: Theory and Experiment*, *2007*(07), P07018.

Burt, A. (2000). Perspective: sex, recombination, and the efficacy of selection—was Weismann right? *Evolution*, *54*(2), 337–351.

Capelle, W. (1955). Das Problem der Urzeugung bei Aristoteles und Theophrast und in der Folgezeit. *Rheinisches Museum für Philologie*, *98*(2. H), 150–180.

Cook, L. M., & Turner, J. R. (2020). Fifty per cent and all that: what Haldane actually said. *Biological Journal of the Linnean Society*, *129*(3), 765–771.

Cooksey, R., Swenson, J., Clark, N., Gay, E., & Thornsberry, C. (1990). Patterns and mechanisms of beta-lactam resistance among isolates of Escherichia coli from hospitals in the United States. *Antimicrobial agents and chemotherapy*, *34*(5), 739–745.

Cooper, T. F. (2007). Recombination speeds adaptation by reducing competition between beneficial mutations in populations of Escherichia coli. *PLoS biology*, *5*(9), e225.

Cuvier, G. (1796). Note on the skeleton of a very large species of quadruped, hitherto unknown, found in Paraguay and deposited in the Cabinet of Natural History in Madrid. *Magasin Encyclopédique*.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. Murray.

Dawson, K. J. (2002). The evolution of a population under recombination: How to linearise the dynamics. *Linear algebra and its applications*, *348*(1-3), 115–137.

de Lamarck, J.-B. d. M. (1809). *Philosophie zoologique, ou Exposition des considérations relatives à l'histoire naturelle des animaux...* (Vol. 1). Musée d'Histoire Naturelle.

de Visser, J. A. G., & Elena, S. F. (2007). The evolution of sex: empirical insights into the roles of epistasis and drift. *Nature Reviews Genetics*, *8*(2), 139–149.

De Visser, J. A. G., & Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics*, *15*(7), 480–490.

Domingo, J., Diss, G., & Lehner, B. (2018). Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature*, *558*(7708), 117–121.

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics*, *78*(2), 737–756.

Fisher, R. A. (1919). The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, *52*(2), 399–433.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford Clarendon Press.

Gabriel, W., Lynch, M., & Bürger, R. (1993). Muller's ratchet and mutational meltdowns. *Evolution*, *47*(6), 1744–1757.

Gerrish, P. J., & Lenski, R. E. (1998). The fate of competing beneficial mutations in an asexual population. *Genetica*, *102*, 127–144.

Gould, S. J. (2002). *The structure of evolutionary theory*. Harvard University Press.

Haldane, J. (1924). A mathematical theory of natural and artificial selection–I. *Bulletin of mathematical biology*, *52*(1-2), 209–40.

Hall, B. G. (2002). Predicting evolution by in vitro evolution requires determining evolutionary pathways. *Antimicrobial agents and chemotherapy*, *46*(9), 3035–3038.

Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science*, *28*(706), 49–50.

Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research*, *8*(3), 269–294.

Jacoby, G. A., & Munoz-Price, L. S. (2005). The new β-lactamases. *New England Journal of Medicine*, *352*(4), 380–391.

Jain, K. (2008). Loss of least-loaded class in asexual populations due to drift and epistasis. *Genetics*, *179*(4), 2125–2134.

Johnson, T., & Barton, N. H. (2002). The effect of deleterious alleles on adaptation in asexual populations. *Genetics*, *162*(1), 395–411.

Kettlewell, H. D. (1958). A survey of the frequencies of Biston betularia (L.)(Lep.) and its melanic forms in Great Britain. *Heredity*, *12*(1), 51–72.

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

Kondrashov, A. S. (1994). Muller's ratchet under epistatic selection. *Genetics*, *136*(4), 1469–1473.

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

Manrubia, S., Cuesta, J. A., Aguirre, J., Ahnert, S. E., Altenberg, L., Cano, A. V., ... others (2021). From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Physics of Life Reviews*.

McDonald, M. J., Rice, D. P., & Desai, M. M. (2016). Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature*, *531*(7593), 233–236.

McHale, D., & Ringwood, G. (1983). Haldane linearisation of baric algebras. *Journal of the London Mathematical Society*, *2*(1), 17–26.

Mendel, G. (1865). Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn*, *4*, 3–47.

Moradigaravand, D., & Engelstädter, J. (2013). The evolution of natural competence: disentangling costs and benefits of sex in bacteria. *The American Naturalist*, *182*(4), E112–E126.

Muller, H. J. (1932). Some genetic aspects of sex. *The American Naturalist*, *66*(703), 118–138.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *1*(1), 2–9.

Neidhart, J., Szendro, I. G., & Krug, J. (2013). Exact results for amplitude spectra of fitness landscapes. *Journal of theoretical biology*, *332*, 218–227.

Nowak, S., Neidhart, J., Szendro, I. G., & Krug, J. (2014). Multidimensional epistasis and the transitory advantage of sex. *PLoS computational biology*, *10*(9), e1003836.

Park, S.-C., Klatt, P., & Krug, J. (2018). Rare beneficial mutations cannot halt Muller's ratchet in spatial populations (a). *EPL (Europhysics Letters)*, *123*(4), 48001.

Park, S.-C., & Krug, J. (2011). Bistability in two-locus models with selection, mutation, and recombination. *Journal of mathematical biology*, *62*(5), 763–788.

Park, S.-C., & Krug, J. (2013). Rate of adaptation in sexuals and asexuals: a solvable model of the Fisher–Muller effect. *Genetics*, *195*(3), 941–955.

Park, S.-C., Simon, D., & Krug, J. (2010). The speed of evolution in large asexual populations. *Journal of Statistical Physics*, *138*(1), 381–410.

Peck, J. R. (1994). A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics*, *137*(2), 597–606.

Pesce, D., & de Visser, A. (n.d.). *Recombination benefits depend on initial position in the fitness landscape.* (unpublished)

Pesce, D., Lehman, N., & de Visser, J. A. G. (2016). Sex in a test tube: testing the benefits of in vitro recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1706), 20150529.

Poelwijk, F. J., Tănase-Nicola, S., Kiviet, D. J., & Tans, S. J. (2011). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of theoretical biology*, *272*(1), 141–144.

Rice, W. R., & Chippindale, A. K. (2001). Sexual recombination and the power of natural selection. *Science*, *294*(5542), 555–559.

Rouzine, I. M., Brunet, É., & Wilke, C. O. (2008). The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. *Theoretical population biology*, *73*(1), 24–46.

Ruelens, P., & de Visser, J. A. G. (2021). Choice of β-lactam resistance pathway depends critically on initial antibiotic concentration. *Antimicrobial Agents and Chemotherapy*, AAC–00471.

Salverda, M. L., Dellus, E., Gorter, F. A., Debets, A. J., Van Der Oost, J., Hoekstra, R. F., ... de Visser, J. A. G. (2011). Initial mutations direct alternative pathways of protein evolution. *PLoS genetics*, *7*(3), e1001321.

Salverda, M. L., De Visser, J. A. G., & Barlow, M. (2010). Natural evolution of TEM-1 β-lactamase: experimental reconstruction and clinical relevance. *FEMS microbiology reviews*, *34*(6), 1015–1036.

Schenk, M. F., Szendro, I. G., Krug, J., & de Visser, J. A. G. (2012). Quantifying the adaptive potential of an antibiotic resistance enzyme. *PLoS genetics*, *8*(6), e1002783.

Schenk, M. F., Witte, S., Salverda, M. L., Koopmanschap, B., Krug, J., & de Visser, J. A. G. (2015). Role of pleiotropy during adaptation of TEM-1 β-lactamase to two novel antibiotics. *Evolutionary applications*, *8*(3), 248–260.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, *23*(1), 23–35.

Stemmer, W. P. (1994). Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, *370*(6488), 389–391.

Szendro, I. G., Schenk, M. F., Franke, J., Krug, J., & De Visser, J. A. G. (2013). Quantitative analyses of empirical fitness landscapes. *Journal of Statistical Mechanics: Theory and Experiment*, *2013*(01), P01005.

Wagner, G. P., & Gabriel, W. (1990). Quantitative variation in finite parthenogenetic populations: what stops Muller's ratchet in the absence of recombination? *Evolution*, *44*(3), 715–731.

Wakeley, J. (2009). *Coalescent theory: an introduction* (No. 575: 519.2 WAK).

Wein, T., & Dagan, T. (2019). The effect of population bottleneck size and selective regime on genetic diversity and evolvability in bacteria. *Genome biology and evolution*, *11*(11), 3283–3290.

Weinberg, W. (1908). Über Vererbungsgesetze beim Menschen. *Zeitschrift für induktive Abstammungs-und Vererbungslehre*, *1*(1), 440–460.

Weinberger, E. D. (1991). Fourier and Taylor series on fitness landscapes. *Biological cybernetics*, *65*(5), 321–330.

Weinreich, D. M., Delaney, N. F., DePristo, M. A., & Hartl, D. L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *science*, *312*(5770), 111–114.

Weinreich, D. M., Lan, Y., Wylie, C. S., & Heckendorn, R. B. (2013). Should evolutionary geneticists worry about higher-order epistasis? *Current opinion in genetics & development*, *23*(6), 700–707.

Weinreich, D. M., Watson, R. A., & Chao, L. (2005). Perspective: sign epistasis and genetic costraint on evolutionary trajectories. *Evolution*, *59*(6), 1165–1174.

Weismann, A. (1889). *Essays on heredity and kindred biological subjects.* Oxford Univ. Press, Oxford, UK.

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

Wick, L. M., Weilenmann, H., & Egli, T. (2002). The apparent clock-like evolution of Escherichia coli in glucose-limited chemostats is reproducible at large but not at small population sizes and can be explained with Monod kinetics. *Microbiology*, *148*(9), 2889–2902.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proc. 6th Int. Congress Genet*, *1*, 356–366.

Zanini, F., & Neher, R. A. (2012). FFPopSim: an efficient forward simulation package for the evolution of large populations. *Bioinformatics*, *28*(24), 3332–3333.

Zeyl, C., Mizesko, M., & De Visser, J. A. G. (2001). Mutational meltdown in laboratory yeast populations. *Evolution*, *55*(5), 909–917.

Zhao, H., Giver, L., Shao, Z., Affholter, J. A., & Arnold, F. H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature biotechnology*, *16*(3), 258–261.

Zwart, M. P., Schenk, M. F., Hwang, S., Koopmanschap, B., de Lange, N., van de Pol, L., ... de Visser, J. A. G. (2018). Unraveling the causes of adaptive benefits of synonymous mutations in TEM-1 β-lactamase. *Heredity*, *121*(5), 406–421.

B Declaration of individual contributions

1st manuscript

Alexander Klug, Su-Chan Park and Joachim Krug.
"Recombination and mutational robustness in neutral fitness landscapes."
PLoS computational biology 15.8 (2019): e1006884.

Author contribution: The individual contributions are published along with the article, c.f. page 28 of the article. The first author wrote the first draft of the article, which was subsequently improved by the other authors. Su-Chan Park joined the project later and strengthened the analytical results. The first author developed the idea of this project during his master's program. Therefore, parts of the publication can already be found in a much less elaborated form within his master's thesis. This entails the definition of mutational robustness, the notion of the recombination weight, and the considered model landscapes.

2nd manuscript

Alexander Klug and Joachim Krug.
"Effect of recombination on the evolvability, genetic diversity and mutational robustness of neutrally evolving finite populations."

Author contribution: The first author conceptualized the question of the project and performed the simulations. Furthermore, the analysis is the work of the first author and the first draft. The draft was then subsequently improved by the second author.

C Acknowledgments

Finally, it is time to thank everyone who has supported me in one way or another during my doctoral studies.

First and foremost, I would like to thank my supervisor Joachim Krug for giving me the chance and freedom to explore this fascinating research field of evolutionary biology. I am very grateful for all the guidance along this journey and for the provided opportunities to attend conferences, international summer schools, and a short trip to Boston during my PhD. I consider all the experiences during this time as very valuable for myself and my education.

I would also like to thank Thomas Wiehe for reviewing this dissertation and Berenike Maier for taking over the chair of the thesis committee.

Moreover, many thanks to all the current and former lab members for interesting scientific and non-scientific discussions, which have made research even more enjoyable.

Furthermore, I have to say thank you to my fellow students Jonas, Hannes, Marc, and Philipp, with whom studying physics in Cologne was much more fun. Special thanks to Alva for all the refreshing walks we shared, where I was able to gather new thoughts.

I also owe a huge thank you to my wonderful girlfriend Jennifer for always being supportive.

Zuletzt danke ich meiner Familie, die mich immer unterstützt hat und Dank der ich natürlich erst soweit gekommen bin.

D Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Teilpublikation:
Klug, A., Park, S. C., & Krug, J. (2019). Recombination and mutational robustness in neutral fitness landscapes. PLoS computational biology, 15(8), e1006884.

Julian Alexander Klug

Köln, den 21.12.2021