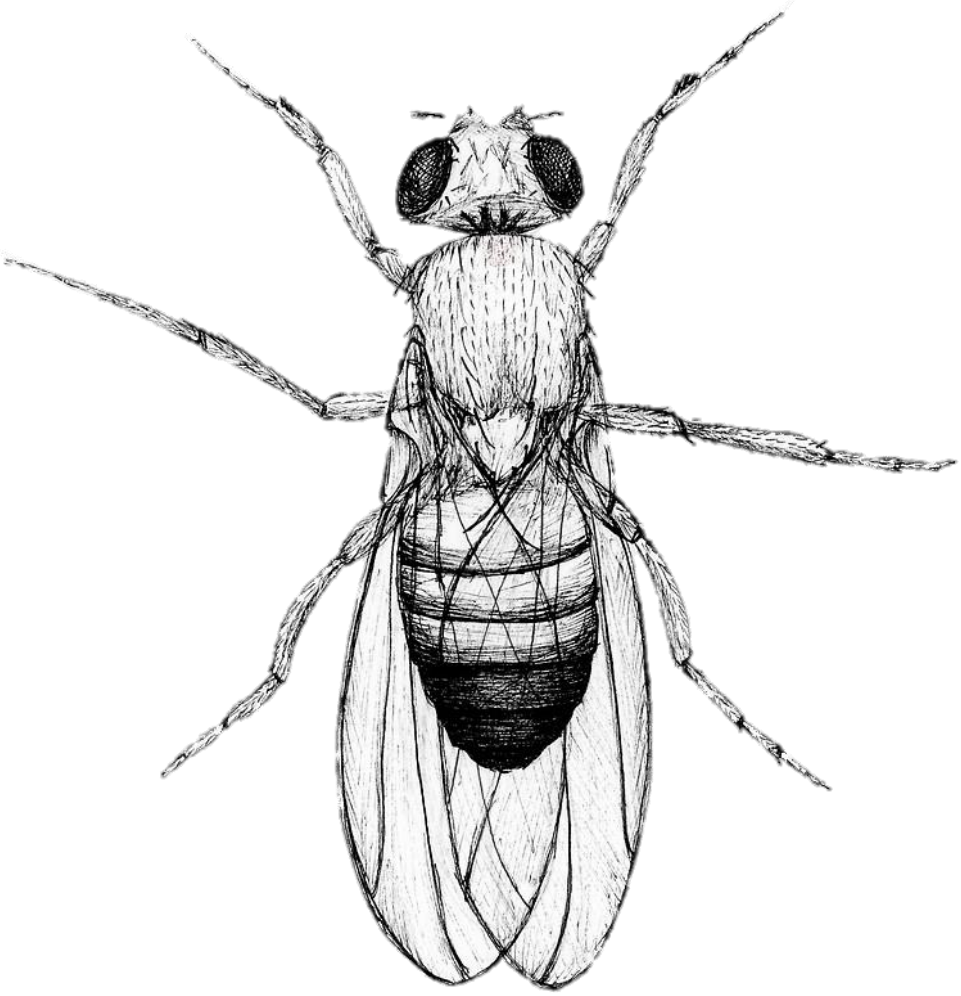
Linking the genetic architecture to divergent evolution of longevity of Drosophila *melanogaster*



Martijn Zoodsma

MSc. Thesis

Supervised by Joost van den Heuvel & Bas Zwaan

March 2020

Wageningen, the Netherlands

WAGENINGEN
UNIVERSITY & RESEARCH

Linking the genetic architecture to divergent evolution of longevity in D. *melanogaster*

Martijn Zoodsma

martijn.zoodsma@wur.nl

960430990090

MSc. Programme Bioinformatics

Wageningen University & Research

Laboratory of Genetics

Course code: GEN-80436

Supervisor

Dr. Joost van den Heuvel

Laboratory of Genetics

Wageningen University & Research

Examiner

Prof. Dr. Bas Zwaan

Laboratory of Genetics

Wageningen University & Research

March 24th, 2020

Wageningen, the Netherlands

Abstract

Understanding the mechanisms and genetic basis of polygenic adaptation is an essential goal of genetic studies. Recently, several evolve and resequence (E&R) studies have sought to investigate the genetic basis of longevity in Drosophila *melanogaster*. A meta-analysis of these studies has revealed a remarkable lack of overlap at the genetic level, yet results appear to converge at a higher biological level. The genetic basis of adaptation appears strikingly different despite successful adaptation across studies. Here, we investigate the consistency of such evolutionary responses and focus on relatedness between populations and the genetic architecture of complex traits.

To achieve this, we create multiple populations that are genetically related to varying degrees. Using these populations, we simulate E&R experiments based on a single polygenic, quantitative trait. We then compare the evolutionary response within and between populations. Furthermore, we investigate the influence of several components of the genetic architecture: the number of loci underlying a trait, the allele frequency distribution and the force of selection. Finally, we compare these results to the recent E&R studies.

We find that the consistency of an evolutionary response between two populations depends on the genetic distance between these populations. Genetically differing populations have different genetic architectures, which determine the direction of evolution in this population. We find substantial differences in the genetic architecture between recent E&R studies, which may explain the lack of consistency at the genetic level.

In conclusion, we combine simulations with experimental results to uncover the source of the lack of consistency between E&R studies. Furthermore, we argue whether such a consistent response at the genetic level can be expected.

Keywords: evolve & resequence, Drosophila *melanogaster*, genetic architecture, longevity

Table of Contents

List of figures

List of tables

Introduction

Understanding the mechanisms of polygenic adaptation to environmental pressure is essential in studying evolutionary processes and has been the subject of a great number of studies (Berg & Coop, 2014; Daub et al., 2013; Pritchard et al., 2010; Wellenreuther & Hansson, 2016). Despite tremendous amounts of research, our knowledge concerning the underlying mechanisms remains limited: In what ways does a polygenic trait contribute to an individual's fitness? What genetic variants underlie variation in this trait and how are these genetic variants distributed? How many loci influence a polygenic trait? These questions are long-standing questions in genetic studies and converge on a single subject: the genetic architecture of a trait.

The genetic architecture of a polygenic trait encompasses the characteristics of genetic variation that are responsible for heritable phenotypic variability (Timpson et al., 2018). The genetic architecture comprises, among others, the number of genetic variants underlying this trait, their individual effects on the trait, population frequency, the extent of pleiotropy but also interactions with other loci and the environment (Timpson et al., 2018). Furthermore, there are some lesser obvious components like the effective population size (Hartl & Clark, 1997; Timpson et al., 2018), selection force (Timpson et al., 2018), and linkage disequilibrium between variants (Gazal et al., 2017) that exert influence on the genetic architecture. Research on the genetic architecture has proven challenging because the different components interact with each other, evolve over time (Hansen, 2006) and differ between traits (Mackay, 2001; Timpson et al., 2018). However, increasing our overall understanding of the genetic architecture of polygenic traits is essential, as genome-wide association studies have shown that many traits in the human genome are in fact polygenic traits (Stranger et al., 2011). More research is required if we are to understand the genetic basis of e.g. susceptibility to disease, individual drug adaptation and polygenic adaptation (Mackay, 2001).

An elegant way to research the genetic architecture of polygenic traits are Evolve and Resequence (E&R) studies. These studies attempt to map variants that correlate with a specific trait by relying on (pooled) sequencing of individuals (Kofler & Schlötterer, 2014; Schlötterer et al., 2014). Large populations of model organisms are subjected to selection regimes and their genetic make-up is compared to control populations. In the past, E&R studies in Drosophila *melanogaster* have proven successful in identifying loci correlated to body size (Turner et al., 2011) courtship song (Turner & Miller, 2012), hypoxia tolerance (Zhou et al., 2011) and egg size (Jha et al., 2015), but also adaptation to temperature (Tobler et al., 2014). Recently, several E&R experiments have sought to investigate the genetic basis of longevity in D. *melanogaster* by, for instance, selecting on late reproductive age (Carnes et al., 2015; Fabian et al., 2018; Hoedjes et al., 2019; Remolina et al., 2012). Crucial is that within a study, replicate lines show high degrees of consistency, while between different studies such consistencies are absent (Hoedjes et al., 2019). Moreover, recent research has uncovered significant overlap between these studies at a higher biological level (Heuvel et al., personal communication). It appears that experimental evolution has targeted identical biological pathways across studies, yet the underlying genetic mechanisms remain unclear.

Before embarking on an expensive and labour-intensive E&R experiment, scientists face numerous choices concerning the setup of the experiment. Factors to consider are, among others, population size, number of biological replicates, length of the artificial selection process, selection of flies for the founder populations and determining when to sequence the population. These choices determine the power of the experiment and thus its capability to identify low-effect variants (Kofler & Schlötterer, 2014). Software to simulate E&R experiments is readily available and allows scientists to test different strategies and examine their performance (Haller & Messer, 2017; Neuenschwander et al.,

2008; Vlachos & Kofler, 2018; Zanini & Neher, 2012). However, these models fail to consider arguably essential biological factors. Often, these models operate under the assumption that a locus' contribution to a trait is purely additive. Dominance interactions between alleles and epistatic interactions between loci are assumed to have no significant contribution to the individual's phenotype, or on the course of evolution. While some models do allow for epistatic interactions (Vlachos & Kofler, 2018), fitness values must be entered manually for each interaction. This is impractical for any simulations that involve more than 10 loci. Furthermore, the impact of the physical structure of chromosomes is underestimated. Linkage disequilibrium (LD) between loci is not always calculated and reported, while its importance is undebated (Hill & Robertson, 1968; Pritchard & Przeworski, 2001; Reich et al., 2001). We argue that these models fail to incorporate important biological components and are therefore incomplete. Below, we highlight the influence of these components on an evolutionary process and finally describe the aim of this study.

Chromosomal structure creates physical linkage between neighbouring loci and therefore causes non-random segregation of loci. Loci that co-segregate more than would be expected at random are said to be in LD (Hill & Robertson, 1968). Recombination events separate alleles from each other and break down LD. Because recombination events are more likely to happen as the distance between two loci increases, a decrease in LD over distance is expected. This suggests that loci situated near a locus under active selection can co-segregate and increase in population frequency – the so-called hitchhiking effect (Barton, 2000). This short-range LD around strongly selected positions generates a lot of false positives in artificial selection experiments (Nuzhdin & Turner, 2013). When an individual experiences selective pressure at e.g. 100 loci, thousands of loci may co-segregate and produce the illusion of a genome-wide response to selection. In the era of low-cost sequencing and an abundance of information, the real challenge for scientists is to separate these false positives from true results. Some *in silico* evolutionary models fail to calculate LD (Haller & Messer, 2017; Vlachos & Kofler, 2018), while we believe that is essential in shaping the genetic composition of a population experiencing selective pressure.

Alleles may interact with each other, both within a locus (dominance) and across loci (epistasis). In the simplest form, an allele may mask or enhance the effect of another allele, depending on the sign of the interaction. The effect of a mutation is then dependent on the genetic background it appears in (Weinreich et al., 2005). The role and influence of epistasis on evolution have been extensively discussed in literature (Hansen, 2013; Templeton, 2000; Whitlock et al., 1996; Wolf et al., 2000), yet no compelling answer has been presented. Recently, Barton argued that epistasis does not significantly alter the evolutionary process and that the evolution of complex traits can very well be described by an infinitesimal model (Barton, 2017). On the other hand, some believe that epistasis has been ignored in evolutionary studies for too long (Carlborg & Haley, 2004). It has been proven that first-order interaction components are converted to additive genetic variance. In turn, the higher-order interaction components contribute to the first-order interaction component (Barton & Turelli, 2004; Hill, 2017). While the interaction components themselves may not contribute much to the phenotype, their sum is a substantial fraction of the additive genetic variance (Hill, 2017). Furthermore, studies have shown that epistatic interactions play an important role in the evolutionary process because they alter the fitness landscape (Szendro et al., 2013; Whitlock et al., 1996). In short, populations may travel different genetic paths depending on their relative starting position, arising mutations and their timing. In conclusion, dominance and epistatic interactions have the potential to influence a locus' contribution to a quantitative trait. Furthermore, these interactions determine the ways in which a population can respond to selective pressure. Nevertheless, dominance and epistatic interactions are often not recognized by currently available *in silico* evolutionary models (Haller & Messer, 2017; Kessner & Novembre, 2014; Vlachos & Kofler, 2018)

In summary, we argue that existing evolutionary models do not sufficiently cover all biological aspects. In this study, we construct a new evolutionary model and investigate the lack of consistency at the genetic level between recent E&R experiments: why do populations evolve to identical phenotypic values by different genetic changes? To study the consistency of evolution between experiments, we compare the genetic makeup between control and selection lines between these experiments. Due to time restrictions, we limit ourselves to additive genetic variance and leave dominance and genetic variance for future research. We start by investigating whether the consistency of evolution depends on the genetic distance between populations. Subsequently, we turn to the genetic architecture and examine several components: the number of underlying loci, the nucleotide diversity in the population (e.g. the shape of the allele frequency distribution), and finally the force of selection. In the process, we highlight the striking influence of the chromosomal structure and the importance of non-random segregation of loci. Subsequently, we prove its influence on the evolutionary process by performing simulations with increasing recombination rates. Finally, we investigate differences in the genetic architecture in recent E&R studies. In conclusion, we combine simulated and experimental results to connect the genetic architecture to divergent evolution of longevity in D. *melanogaster*.

Methods

In this study, we investigate the consistency of an evolutionary response within and between populations. Specifically, we focus on the impact of genetic distance between populations and the genetic architecture of a trait: the number of loci underlying a trait, the shape of the allele frequency distribution and the force of selection applied to the population.

Genetic determination of the phenotypes

The genetic determination of an individual's phenotype consists of only additive genetic variance. For simplicity, we denote the average effect of allele substitution α (Falconer & Mackay, 1996) as the genotypic effect of a locus. We normalize the additive effect per locus so that the additive contribution for a heterozygous locus equals zero (Table 1). Furthermore, we scale the total additive genetic variance in the population to be 1.0 to allow for comparisons between different populations. Subsequently, we add environmental genetic variance to the individual phenotypes. Here, we use h^2 = 0.25, signifying that the variance of the underlying normal distribution equals 4.0. Finally, we standardize all phenotypes by subtracting the population mean phenotype and dividing by the standard deviation of the phenotypic values in the population. The optimum that individuals evolve towards is then expressed as units of standard deviation (*z*-units). The fitness of an individual with a specific phenotype is then calculated as:

$$e^{0-(^{phenotype-optimum}/_{width})^2} \quad (1)$$

where *optimum* is the number of standard deviation units the population mean is separated from the optimum, and *width* denotes the width of the fitness curve.

Recombination

The number of recombination events that occurs per mating event is taken from a Poisson distribution with mean λ = 1. This means that, on average, every chromosome has had a single recombination event between the two parental gametes. The invading strand is picked at random, and recombination events are placed randomly across the chromosome. Furthermore, multiple recombination events on the same chromosome are located independent of each other.

Genetically differing populations

Populations with different genetic backgrounds were simulated in QMSim (Sargolzaei & Schenkel, 2009). The history of the simulated populations was set up to reflect the natural population structures of the population of D. *melanogaster* as used by Hoedjes et al (2019) in their E&R study. The initial population (N=20.000) individuals can randomly mate for 100.000 generations to establish mutation-drift equilibrium, Subsequently, the population is split repeatedly over time to create genetically different populations (Fig. 1). Genetic distance between populations is calculated as Nei's genetic distance (Nei, 1972). We create control and selection lines by cloning the population. In the selection lines, reproductive success depends on the fitness values of both parents, while random mating is allowed in the control lines.

Genetic architecture

We consider three different aspects of the genetic architecture of a complex trait: the number of loci underlying a trait, the shape of the allele frequency distribution and the force of selection applied to the population. In the simulations performed here, individuals possessed a total of 4 diploid chromosomes. Simulations with differing numbers of loci per chromosome were run with identical number of chromosomes but differing number of loci per chromosome.

Table 1 **Contribution to the phenotype for different genotypes.** For each genotype, the contribution to the phenotype is normalized with respect to the heterozygous individuals. Only the additive genetic variance is considered.

Genotype	Aa	Aa / aA	AA
Genotypic effect size	$-\alpha$	*0*	α

We created populations with different allele frequencies using QMSim (Sargolzaei & Schenkel, 2009). For any population at equilibrium, the shape of the allele frequency distribution depends on $\vartheta = 4 * Ne * \mu$ (Davoudi et al., 2018). By altering the mutation rate μ, three different populations were created with ϑ = 0.02, 0.2 and 0.8. Lower values of ϑ indicate a more pronounced U-shaped allele frequency distribution.

We simulated varying intensities of selection by adjusting the width of the bell-shaped fitness curve. Increasing the width of this curve results in less severe fitness drop-off for individuals that are farther away from the genotypic optimum. Inversely, decreasing the width increases the selection force.

Model design

By default, we simulate 4 control populations and 4 selection populations. In each population, the following steps are performed:

- Initial statistical calculations
    - Normalization of additive genetic variance. The total additive genetic variance is calculated as $V_a = 2 * p * q * \alpha^2$ summed over all loci. We multiplied the genotypic score per locus a with $1 / sqrt(a)$, thereby ensuring a total additive genetic variance in the population of 1.0.
    - Environmental variance is added to the phenotypes. Because the heritability is set to 0.25, environmental variance is taken from a normal distribution with standard deviation 3.0.
    - Phenotype standardization. We subtract the mean phenotype from all phenotypes and divide by the standard deviation of the phenotypic values in the population.
    - Fitness value are calculated for as indicated in formula 1.
- Mating of two individuals. In the selection lines, the mating chance of two individuals is proportional to the product of their fitness values. In control lines, random mating is allowed. The following steps are repeated until the next generation has been created:
    - Pick random parents until a match is found where fitness values are high enough.
    - Gamete formation for both parents: Recombination is applied between homologous chromosomes.
    - Mutations are introduced at a pre-specified rate
    - Offspring is formed by combining gametes from both parents and added to the new generation.

- Statistical calculations for the new generation: allele frequencies, the mean phenotype in the population and possibly linkage disequilibrium between loci are calculated for the new population. Environmental variance is added to all phenotypes. Subsequently, phenotypic scores are standardized.

Once the total number of generations has been reached, output files are generated, and the simulation is ended. Output files include genotypic effects of loci used in the simulations, allele frequencies per generation and mean phenotype in the population per generation. LD between loci may also be calculated and written to files for further analysis. Finally, genotypes of the last generation may be saved and can be used as input for further simulations. To save disk space, genotypes are converted to the genotypic code: {0 2 3 4 5}, which corresponds to: 0 - homozygous for allele 1, 2 - homozygous for allele 2, 3 - heterozygous, 4 - heterozygous, 5 - unknown. The chromosomal structure is not explicitly stored but can be retrieved by calculating the number of loci per chromosome and splitting the chromosomes accordingly.

Measure of consistency of evolution

We measure the consistency of an evolutionary process both within and between experiments. At the end of every simulation, we calculate the difference in allele frequency per locus between every control and its matched selection line. To calculate the consistency of the evolutionary process within an experiment, we take the mean allele frequency difference between two control-selection comparisons and compare this to the mean allele frequency difference in the remaining two control-selection comparisons. To calculate the consistency of evolution between experiments, we take the mean difference in allele frequency across the whole experiment (4 control-selection comparisons) and calculate the correlation to the mean difference in allele frequency in another experiment. We denote the correlation between two separate experiments as $\rho(\Delta AF)$.

Implementation & visualization

The model used in this study is implemented in Python 3 and is dependent on the external libraries NumPy, Joblib and Scikit-learn which are part of the SciPy ecosystem (Virtanen et al., 2020). Visualization was performed using the statistical language R (R Core Team, 2019) with the package ggplot2 (Wickham, 2016).

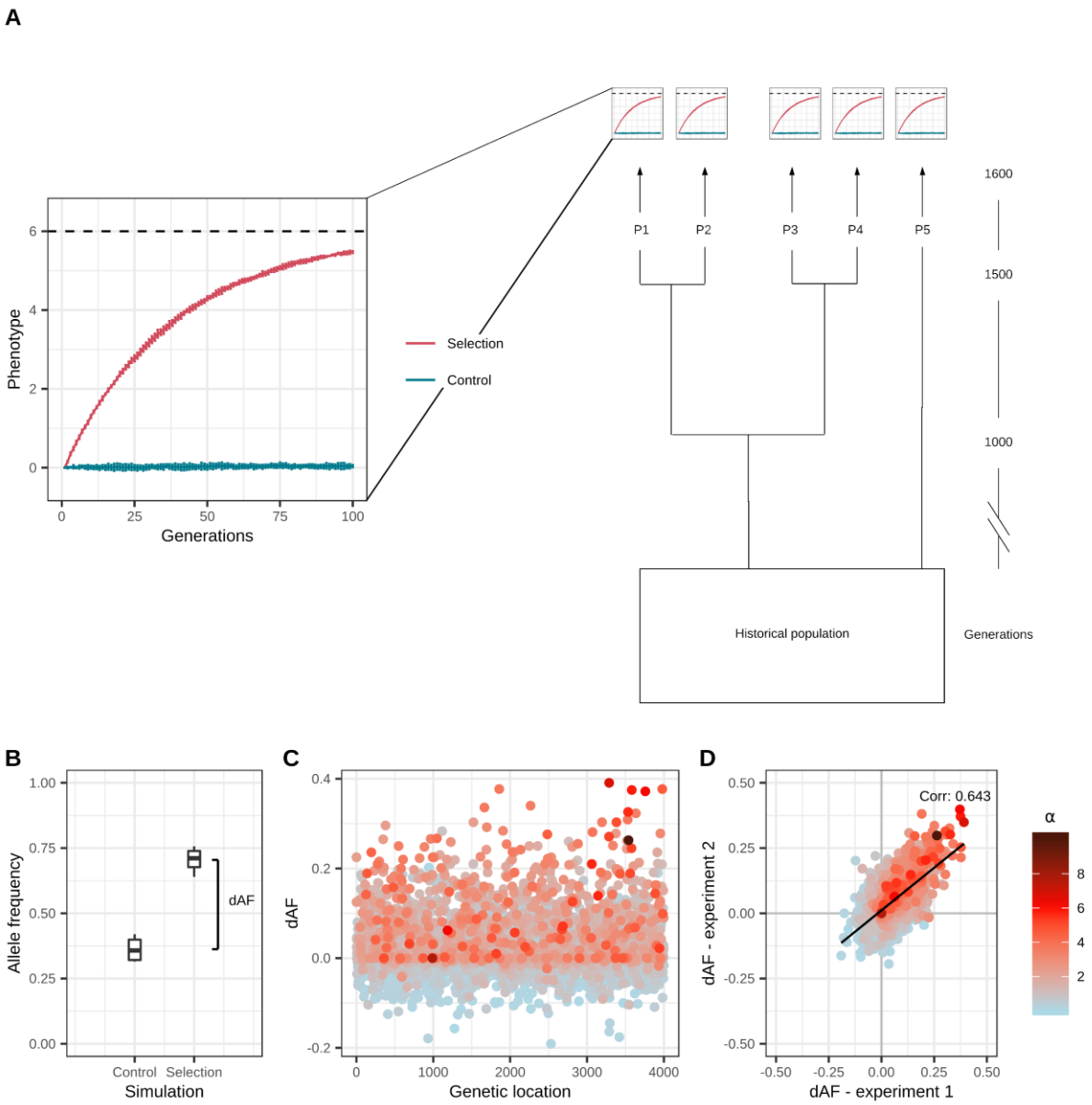


Figure 1 **Population history and comparison between results. A)** The founder population is split repeatedly over time to create varying degrees of genetic relatedness between populations. With each of these genetically different populations, we perform E&R experiments. **B)** The difference in allele frequency between control and selection lines for a single locus, denoted as ΔAF. **C)** The difference in allele frequency between control and selection lines for all loci combined, relative to their genetic location. Points are coloured for their genotypic effect, with red representing higher genotypic effect sizes. **D)** The comparison between two separate E&R experiments. For each point, the x-coordinate is given by its ΔAF in the first experiment, while the y-coordinate is given by the locus' ΔAF in the second experiment. Points are coloured according to a locus' genotypic effect size, with red representing higher genotypic effect sizes.

Results

We investigated the lack of consistency at the genetic level between recent E&R studies that sought to uncover the genetic basis of longevity. To do so, we created an initial founder population and repeatedly split this population over time to create genetically different populations. From these populations, 4 control and 4 selection lines are made. Using these populations, we simulated E&R experiments and compared their genetic makeup (Fig. 1).

The populations we created reflect the natural population structure of D. *melanogaster* leading up to a recent E&R experiment as performed by Hoedjes & colleagues (2019). We verified that populations reached mutation-drift equilibrium by calculating mean LD for adjacent loci and determining the shape of the allele frequency distribution. Furthermore, we calculated Nei's genetic distance (Nei, 1972) between the different populations to prove populations are genetically different. We present this information in supplement 2, along with additional information regarding the distribution of genotypic effect sizes and decay of LD over distance. From this information, we conclude that our model captures the biological characteristics regarding LD and genotypic effect sizes and that populations are at equilibrium.

Genetic distance & the genetic architecture

We investigated the consistency of the evolutionary process by comparing populations that are genetically related to varying degrees. We assessed this consistency both within and between populations. The correlation within experiments was very consistent between different populations, ranging from 0.46 to 0.48. Subsequently, we compared the differing populations to each other. We observe that the consistency of the evolutionary response between populations *ρ(ΔAF)* is lost with increasing genetic distance between two populations (Supplemental Fig. 4).

We proceeded by examining the influence of several components of the genetic architecture on the evolutionary process, starting with the number of loci underlying a trait. We found that the correlation within an experiment did not change significantly as the number of loci increased. However, the consistency between populations ρ(ΔAF) decreased approximately linearly with increasing numbers of loci (Fig. 2A).

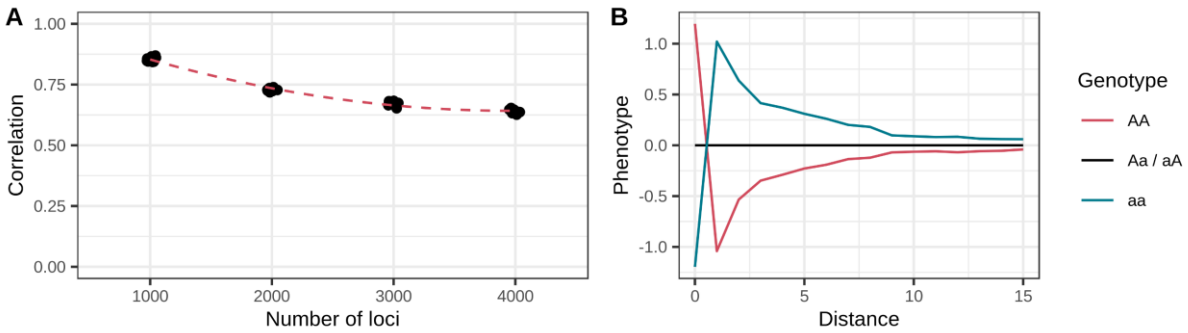


Figure 2 **Chromosomal structure**. **A)** The consistency of the evolutionary response decreases with an increasing number of loci. Simulations comprised 2500 individuals for 100 discrete generations. We simulated 5 experiments and compared between these experiments by calculating ρ(ΔAF) (black dots). The red line indicates a second-degree polynomial fit to the data. **B)** Mean phenotype in the population calculated for a single locus (distance = 0) and expanding with distance in loci in both directions as indicated by the x-axis. The phenotype was normalized with respect to the heterozygous individuals.

Moreover, we compared populations with different allele frequency distributions. We observed that varying the shape parameter ϑ decreased the correlation within experiments (data not shown) and ρ(ΔAF) (Table 1). Overall, increasing ϑ for a more pronounced U-shaped allele frequency distribution decreased consistency within and between experiments. Comparing the evolutionary process between these populations is extremely difficult as the consistency between populations is severely decreased (Table 1). Possibly, more alleles disappear from a population with a more pronounced U-shaped allele frequency distribution because of random drift. Because many alleles have a near-zero allele frequency, random genetic drift may cause alleles to disappear. However, the lost alleles differ per replicate and therefore generate a more heterogeneous response to selective pressure.

Finally, we investigated the importance of the force of selection that the individuals experience. The correlation within an experiment decreased as the selection force decreased. Populations where a higher force of selection was applied thus showed a more consistent response across replicates compared to populations where the force of selection was lower. Pairwise comparisons between populations with different forces of selection show low ρ(ΔAF) values, indicating the difficulty in comparing the evolutionary process between populations that did not experience the same selection forces (Table 3).

Table 2 **Consistency of evolution between populations with different allele frequency distributions**. Values represent ρ(ΔAF) ± standard deviation. Higher values of the shape parameter ϑ generates a more pronounced U-shaped allele frequency distribution.

Shape parameter ϑ	0.02	0.2	0.8
4	0.733 ± 0.009	-	-
6	0.186 ± 0.009	0.645 ± 0.010	-
8	0.200 ± 0.009	0.250 ± 0.010	0.586 ± 0.011

Table 3 **Consistency of evolution between populations that experienced differing forces of selection**. Values represent ρ(ΔAF) ± standard deviation

Width of fitness curve	4	6	8
4	0.731 ± 0.008	-	-
6	0.668 ± 0.011	0.626 ± 0.011	-
8	0.584 ± 0.017	0.546 ± 0.019	0.486 ± 0.024

Recombination rate

When comparing replicates of identical simulations, we noticed some alleles with positive genotypic effects to consistently decrease in allele frequency throughout an E&R experiment. These loci appeared to be under negative selection while their genotypic score was positive (Fig. 1D). We found that this was caused by the formation of haplotype blocks due to non-random segregation of loci. For example, consider a single allele that is surrounded by alleles with near-zero genotypic effects. This allele will hardly be selected for because of genetic drag with neighbouring alleles that probably have near-zero genotypic effect sizes. Instead, alternative haplotypes will be selected that are more advantageous. Inversely, haplotypes may be formed where individuals are missing an allele. Provided that the total genotypic score of the haplotype is higher than the genotypic effect of the single allele, it will have a selective advantage. As a result, the missing allele will decrease in population frequency and eventually disappear. Alleles with positive

genotypic effects may thus consistently decrease in allele frequency because of their neighbouring loci. We show this phenomenon in Fig. 2B, where the individual allele has an above-average positive genotypic effect size. However, separating the population based on their genotype for this allele reveals that individuals with a recessive genotype at this locus have higher phenotypic values when considering neighbouring loci. This suggests that haplotype blocks with sufficiently high genotypic scores have been formed without this allele. Consequently, the allele decreases in allele frequency.

To test if the formation of haplotype blocks was indeed caused by the non-random segregation of loci, we performed simulations with increasing recombination rates. Increasing recombination rates breaks down the chromosomal structure and allows loci to segregate more independent from each other. We found that increasing the recombination rate influences both the correlation within experiments (data not shown) and severely decreases low ρ(ΔAF) (Fig. 3A). This resembles to some degree the effect of simulating more genetic loci. To verify that the haplotype blocks are broken up by recombination, we calculated the correlation between a locus' genotypic effect and its mean difference between control and selection lines. This correlation increases with increasing recombination rates, indicating that alleles with larger genotypic effects tend to increase in frequency more so than smaller-effect alleles.

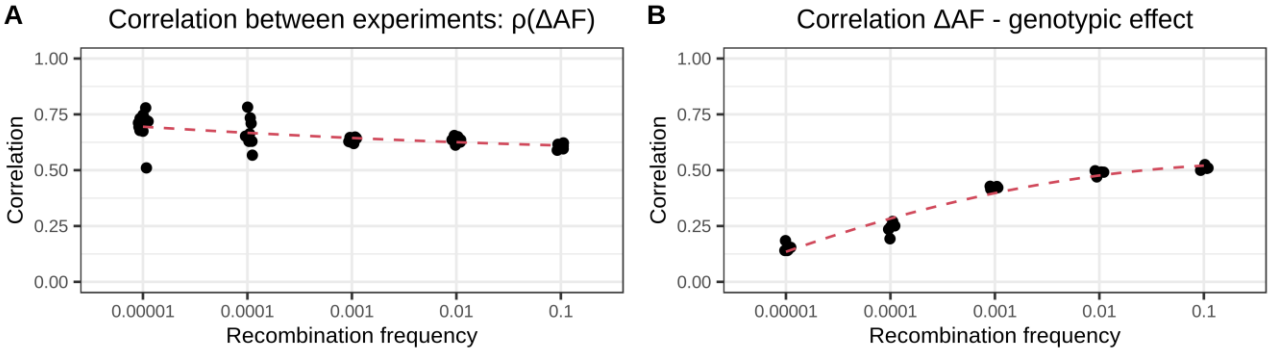


Figure 3 **Recombination breaks up the chromosomal structure**. Simulations comprised 2500 individuals for 100 discrete generations. The red lined represent a second-degree polynomial fit to the data. **A)** Increased recombination rate decreases the correlation between experiments ρ(ΔAF). The black dots represent all pairwise comparisons between 5 replicates. **B)** Recombination breaks up the haplotype blocks. The correlation between a locus' genotypic effect and the difference in allele frequency between control and selection lines increases with increasing recombination rates. Black dots represent this correlation for all pairwise combinations between 5 replicates

Comparison of simulated to experimental results

Subsequently, we investigated the likeliness of results generated by our additive model with recent experimental results. We calculated the correlation between control and selection lines for all E&R studies separately and found correlation values ranging from 0.4 to 0.8 (Supplemental Fig. 5B). In our simulations, correlations within experiments ranged from 0.45 to 0.48. Furthermore, we still observed a moderate degree of consistency between different populations, whereas the E&R studies show a lack of consistency.

Comparison of recent E&R experiments

Given our simulated results, we asked how these components of the genetic architecture differed between recent E&R experiments. All studies have uncovered a differing number of variants, but this number is influenced by the power of the experiment which differed across studies. Our simulations show clear effects of the number of loci on the predictability of evolution, but we can only assume the true number of variants is equal across studies.

However, we can compare the allele frequency distribution between populations. We compared the empirical cumulative distribution function for each experiment, which appear to be substantially different (Supplemental Fig. 5A). Especially the populations from Fabian (2018) show remarkable peaks as a result of variants fixating in a single line out of 4 replicates while disappearing in other lines. Furthermore, the allele frequencies from populations used by Remolina (2012) appear to be normally distributed instead of showing a U-shape.

Finally, we compared the selection regimes across studies. We calculated the correlation between all pairs of control and selection lines for each study separately to get an estimate of drift and selection force in each experiment. We then asked whether the recent E&R studies employed similar selection regimes. Hoedjes (2019) selected for fecundity, while Remolina (2012) favoured late-age reproduction and survival. Fabian (2018) and Carnes (2015) increased lifespan by selecting on late reproduction. Furthermore, the correlation between selection lines also differed per study (Supplemental Fig. 5B). Even though all studies had the common goal to increase lifespan in populations of D. *melanogaster*, it is very likely that the populations experienced differing forces of selection across studies.

Discussion

We aimed to investigate the consistency of polygenic adaptation at the genetic level. To achieve this, we simulated E&R experiments and compared the genetic make-up of individuals after the selection process. First, we used both identical and genetically different populations to prove that consistency of an evolutionary response is lost with genetic distance between populations. In the process, we demonstrated the influence of the chromosomal structure on this evolutionary response. Moreover, we investigated several components of the genetic architecture and showed that these components have the potential to confound results between E&R experiments at the genetic level. Finally, we asked how the genetic architecture differed between recent E&R studies and identified substantial differences.

In population genetics, it is easy to think of genetic loci as units that move independently and without interactions to neighbouring loci. In this study, we showcased an extreme example of the influence loci have on each other. In this study, we showed an extreme example where the influence of neighbouring loci effectively masked the positive genotypic effect of an allele, resulting in a decrease in population frequency for this allele. We anticipate that this is not a problem for alleles with large genotypic sizes, because selection on these positions will be stronger. However, it is important to realize that loci are always influenced by neighbouring loci. Especially loci with small to medium genotypic effect sizes depend heavily on nearby loci and their genotypic effect sizes. When recombination is limited – as in D. *melanogaster* – loci may find themselves 'trapped' in haplotype blocks and become completely dependent on neighbouring loci. At any rate, the allele frequency difference between control and selection lines in E&R experiments does not only depend on the individual allele's genotypic effect size but rather on a combined genotypic effect size. When the allele frequency difference between control and selection lines is not informative of the individual allele's genotypic effect size, interpreting results from E&R results becomes increasingly difficult.

The influence of genetic drag in the context of E&R studies becomes clearer when we look at a recent study that performed identical E&R experiments with two different model organisms (Barghi et al., 2017). Populations of D. *melanogaster* and D. *simulans* were exposed to high temperatures and their genomic response was compared after 60 generations. D. *simulans* is closely related to D. *melanogaster*, but lacks segregating inversions, has higher recombination rates and reduced recombination depression close to centromeres and telomeres (Barghi et al., 2017). They observed a far more distinct response in D. *simulans*, with fewer false positives. Barghi et al argue that the cause for this is the increased recombination rate, which reduces linkage disequilibrium between variants and allows loci to move more independent from each other. As a result of this, fewer loci tend to co-segregate with a position under active selection. Our results agree with this in the sense that higher recombination rates improve the correlation between a locus' genotypic score and the allele frequency difference between control and selection lines, leading to less false positives.

Furthermore, we showed that the genetic architecture determines the direction of evolution in a population. Populations that are genetically less related have different genetic architectures, and therefore their genetic response to selective pressure differs. We were able to generate genetically differing populations by relying on random drift yet achieved only relatively low genetic distances. Probably, this was caused by the large population sizes we used. Nevertheless, we observed a marked decrease in consistency $\rho(\Delta AF)$ between populations. Subsequently, we asked whether populations of D. *melanogaster* used in recent E&R studies also showed genetic differentiation. Without performing exact calculations, it appears that the history of the populations differs substantially. The country of origin, the moment of capture and, by extent, the number of generations under laboratory conditions differ across all studies.

Therefore, it is reasonable to assume that these four populations of D. *melanogaster* show great genetic differentiation as a result of genetic drift and mutations private to each population.

We now turn to the genetic architecture and especially the differences in genetic architecture between populations. We observed a decrease in consistency as the number of loci we simulated increased. This is in line with recent research suggesting that genetic redundancy – two or more genes having identical effects on a quantitative trait – is expected to cause divergent evolution (Barghi et al., 2019). Barghi et al argue that divergent evolution is expected because of the many genetic paths to explore, all leading to the same optimum. Our results are consistent with this theory: as we increased genetic redundancy (e.g. the number of loci), the consistency between populations ρ(ΔAF) decreased. This indicates a more heterogeneous response at the genetic level, while the ultimate results are identical across replicates. However, we still found a consistent response, even when simulating traits with as much as 4000 underlying loci (ρ(ΔAF) ~ 0.640). We observed that the cause for this is the formation of haplotype blocks (Wall & Pritchard, 2003).Haplotype blocks essentially reduce genetic redundancy because the segregation of a locus in a haplotype block heavily depends on surrounding loci. This implies that the total genetic redundancy does not only depend on the number of loci underlying a trait, but also on the recombination rate. Higher recombination rates allow loci to segregate independently, thereby increasing genetic redundancy.

Afterwards, we investigated the shape of the allele frequency distribution. Our simulations show that this distribution greatly determines the direction of selection. The change in allele frequency caused by selection depends on the initial allele frequency (Falconer & Mackay, 1996). Therefore, large allele frequency differences per locus between populations are likely to cause a heterogeneous response to selective pressure. When we compared the allele frequency distributions in recent E&R studies, we found some substantial differences. Especially Fabian et al (2018) reported variants fixating in a single population while disappearing from other lines. This may indicate either large amounts of drift or other effects private to each population. In the recent E&R studies, populations were sequenced at the end of the artificial selection process, during which the allele frequencies change. However, such drastic changes are not expected, and we can therefore assume that the allele frequency distributions also differed at the beginning of the experiment. Our simulations show that it is difficult to compare an evolutionary response between populations when the allele frequency distributions are not identical.

The final component of the genetic architecture we investigated is the force of selection. We have shown that a higher selection force drives a more consistent response and that there is almost no consistent response between populations where different selection regimes are applied. We then asked whether the recent E&R studies employed similar selection regimes, which was not the case. Moreover, we assessed the correlated response for each study, which differed (Supplemental Fig. 5B). This is likely due to multiple factors such as effective population size, length of artificial selection process but also the different selection regimes. Altogether, it is likely that these populations have experienced differing selection forces. In this study, we have shown that differing selection forces can fuel a heterogeneous response, making comparisons between populations extremely difficult.

In summary, we analysed four E&R experiments that sought to investigate the genetic basis of longevity in D. *melanogaster*. A recent meta-analysis reveals that there is no consistency at the genetic level between these experiments (Heuvel et al., personal communication). Based on earlier research (Barghi et al., 2017, 2019) and results presented here, we question whether consistency at the genetic level is expected for such experiments. Looking at the population histories, it is likely that there is a great amount of genetic differentiation between these populations. Consequently, the genetic architecture of a complex trait such as longevity is likely to differ between these populations.

In this study, we identified two components that differ significantly. Simulated results show that the evolutionary response depends heavily on the genetic architecture and that the consistency of the response is lost rather quickly when the genetic architecture differs.

The question then arises: What do we need to do to improve consistency among E&R experiments? In past E&R studies, a lot of attention has always been directed towards analysis at the genetic level: identification of variants that underlie variation in the trait of interest. Instead, it may be more useful to analyse results at a higher biological level. Analyses such as GO-term enrichment in selection lines may provide insight into pathways that are altered by the selection process(Heuvel et al., personal communication). From there, it is possible to compare and find overlap between studies. This approach will yield more meaningful results in comparison to analysis at the genetic level when no consistency is expected there due to genetic differences in populations. However, any overlap in variants between studies might indicate preferred targets of selection or alleles with exceptionally high genotypic effect sizes and make a good starting point for further investigation.

Furthermore, scientists should reconsider their choice of model organism and setup of future E&R experiments. As discussed earlier, D. *simulans* may be a better choice compared to D. *melanogaster*, although the latter is widely used in E&R studies. Naturally, D. *simulans* has its own problem such as reduced availability of genomic resources although some genomic assemblies and annotations are available (Hu et al., 2013; Palmieri et al., 2015). Nonetheless, this species has proven to increase the quality of results generated by E&R experiments (Barghi et al., 2017). Furthermore, the setup of E&R experiments needs to be reconsidered. When analysing results from recent E&R studies, we noticed that not all populations show the expected U-shaped allele frequency distribution. Distorted allele frequency distributions may indicate recent population expansion or bottlenecks (Chen, 2018). Possibly, these populations need to be kept under laboratory conditions for an extended period of time before starting the E&R experiment to allow the population to return to equilibrium.

Furthermore, an important factor in any E&R experiment is the length of the artificial selection process. In our simulations, there appeared to be a trade-off between the number of generations needed to increase a beneficial allele's frequency and the amount of genetic drift that the population experiences. As a result of genetic drift, the allele frequency differences by the selection process start to fade. Consequently, consistency between populations starts to fade. Similar mechanisms appear to have played a role in the recent E&R experiments. Populations from Fabian et al (2018) and Carnes et al (2015) have experienced selection for prolonged periods of time – 144 and 170 generations, respectively. Accumulation of genetic drift caused lower and more irregular correlation values between selection lines. Remolina et al (2012) did not select for such long periods and achieved higher correlations. However, the populations were smaller and therefore the effects of genetic drift are more pronounced. This may explain the range in correlation values. Hoedjes et al (2019) selected for < 20 generations and had the largest populations. Consequently, selection lines are more correlated and more consistent. The optimal number of generations for any specific E&R experiment likely depends on the effective population size, number of replicate lines and the selection regime that is applied. Scientists should use models such as the one presented here to test different scenarios and get an estimate of the ideal length of the selection process.

In this study, we stressed the importance of biologically realistic models, dominance genetic variance and interaction genetic variance. It is important to realize that the model used in this study does not cover all these aspects. Here, we outline future additions to the model that we believe will make this model more realistic. A first factor is the lack of dominance and epistatic interactions. Supplemental information 3 describes the framework we have set up to

simulate these interactions thus far. Furthermore, the model does currently not allow sex chromosomes and therefore no genders can be simulated. Subsequently, it should be possible to simulate chromosome-specific recombination rates, especially on the sex chromosomes (Comeron et al., 2012). Additionally, this model may be improved by modelling biologically more realistic recombination process. Locations of crossovers events are currently randomly placed along the chromosome – implying the absence of a centromere. However, studies have shown that crossovers tend to happen more distally, away from the centromere (Comeron et al., 2012). Furthermore, the location of a crossover event can be influenced by other crossover events happening simultaneously – crossover interference (Hillers, 2004; Stevison & Noor, 2010). Finally, simulations should include the usage of recombination maps to enable simulations of recombination hotspots and coldspots. Fine-scale recombination maps to achieve this for D. *melanogaster* are readily available (Fiston-Lavier et al., 2010; MacKay et al., 2012). Simulations in earlier studies have shown some interesting dynamics regarding haplotype block formation, LD between loci and recombination hotspots (Wall & Pritchard, 2003). Aside from additions to the model, implementations of this model in a high-level language like Java or C++ should significantly decrease computational time. This enables users to simulate more elaborate scenarios (e.g. more loci, more replicates, larger populations), making the simulations more realistic.

In this study, we used a comprehensive measure of the consistency of evolutionary processes within and between studies. However, this measure depends quite heavily on the number of replicates used in the experiment. Consequently, it is difficult to faithfully compare experiments with differing numbers of replicates. Furthermore, comparing correlations within and between studies is not possible because of the differing numbers of replicates (2x2 comparison within studies, but 4x4 comparisons between studies). While this measure is easy to comprehend and gives a good estimation of the consistency of evolution across replicates, there are limitations to be aware of. Further research may be able to identify a more sophisticates measure of consistency that allows comparison across uneven replicates.

In conclusion, this study has proven that a consistent evolutionary response between two populations depends on the degree of genetic differentiation between these populations. Afterwards, we focused on the genetic architecture and show their influence on the evolutionary response. We then analysed four recent E&R studies and show that the genetic architecture differs in two out of three components. Finally, we discussed how to improve the quality of results generated by E&R experiments. This study will impact the setup, analysis and comparison of future E&R experiments. Furthermore, it will hopefully inspire scientists to use in silico models to test the setup of their next E&R experiment to maximize the quality of the results.

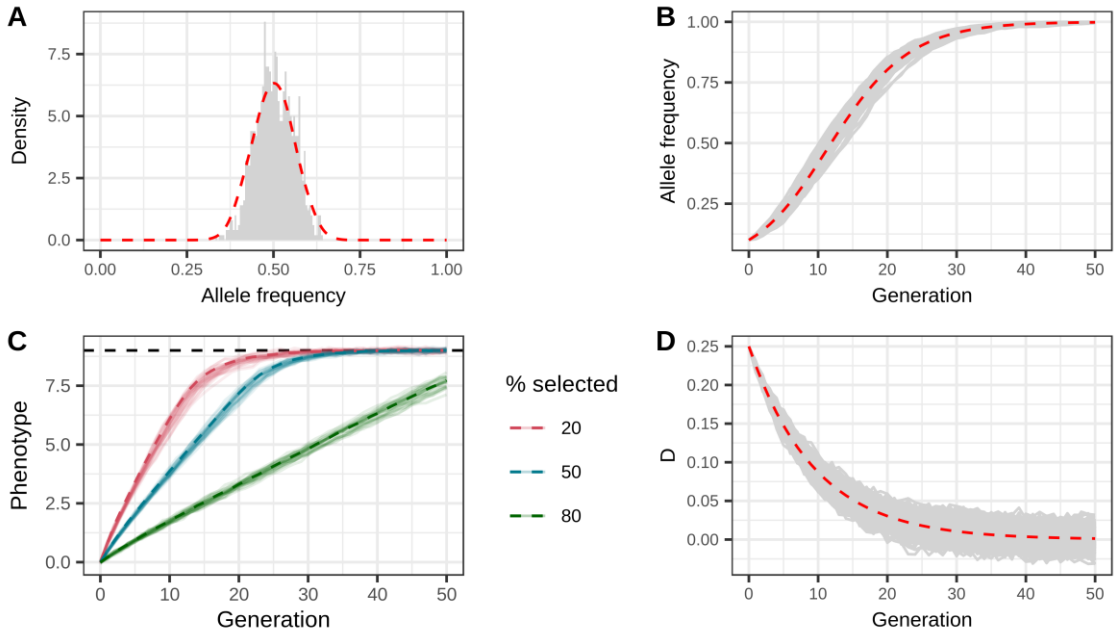
Supplemental information

Supplement 1: Model validation

We recognize genetic drift, selection and recombination as the major forces that shape the outcome of an evolutionary process. Here, we validate the model we use to simulate E&R experiments by simulating smaller test scenarios and comparing results to theoretical expectations (Supplemental Figure 1). To simulate genetic drift, individuals are given 10.000 codominant unlinked loci with starting allele frequencies of 0.5 and identical α. The theoretical distribution after 50 generations of drift was generated using a Wright-Fisher model, where alleles have a random chance to be passed on to the next generation. The two-sided Kolmogorov-Smirnov test was used to check for equality of distributions. Following, selection was simulated by giving individuals a single pair of codominant alleles with a starting population allele frequency of 0.1. Theoretical expectations were derived using the *formula* $Q_t = (1 - ½\ sq - ½\ sq^2) / 1\text{-}sq$ (Falconer & Mackay, 1996), and statistical significance was calculated every 10th generation by using a one-sample t-test.

Subsequently, the population response to varying intensities of selection was evaluated by simulating truncation selection. Alleles were distributed randomly and given random genotypic effect sizes from a negative exponential distribution to create standing genetic variation for selection to act upon. Truncation selection then allowed the fittest individuals to reproduce. The theoretical response was calculated using the breeder's equation $R = H_2 * S$, which permits to calculate the response to selection as a function of heritability and selection differential (Falconer & Mackay, 1996). Ultimately, we test whether recombination is correctly implemented by assessing the decay of LD over time. Initially, all loci were in complete disequilibrium ($D_0 = 0.25$). Theoretical decay was calculated iteratively using the formula $D_t = D_0 * (1\text{-}c)^t$ (Falconer & Mackay, 1996).

We find no significant differences between theoretical and simulated results for genetic drift (p-value: 0.314), selection on a quantitative trait (5 p-values > 0.407) and LD decay (5 p-values > 0.409). However, the population response to selection showed some deviations from theory. The cause for this is the implicit assumption that heritability is constant when calculating the population response at each generation using the formula $R = H_2 * S$. However, it has been shown that additive genetic variance declines over time when a population experiences selection. As a result of this, heritability is also reduced (Falconer & Mackay, 1996). To assert correct functionality, we calculate actual heritability values by rewriting the breeder's equation to $H_2 = R / S$ and fit a linear regression model to these values. Predicted values indeed showed a decrease over time and repeating the simulations with the predicted heritability values showed no significant deviations from theoretical expectations (3 p-values > 0.784). We conclude that genetic drift, selection and recombination are correctly modelled.

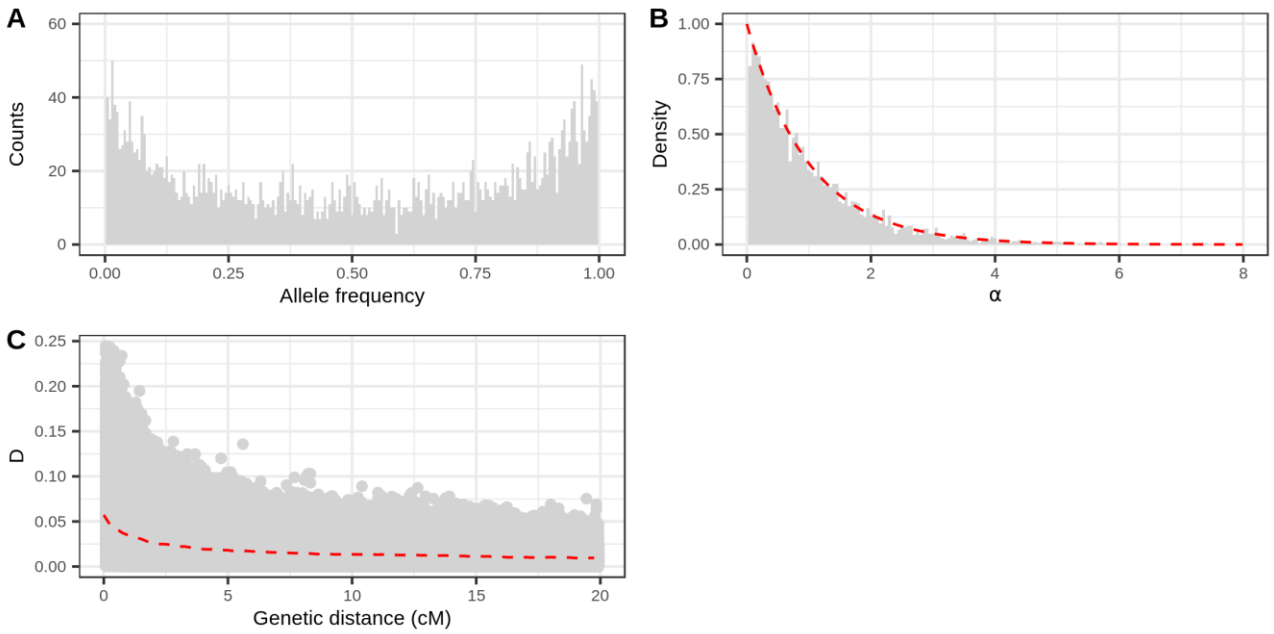


Supplemental Figure 1 **Model validation**.Unless specified otherwise, N=2000 individuals were simulated for a total of 50 discrete generations. Each individual has 1000 codominant (h=0.5) loci, with a starting allele frequency of 0.5 and identical locus-specific additive genetic variance. Mutations were not considered. A) Allele frequency distribution at generation 50, compared to theoretical expectations as represented by the red dashed line. The distributions are not significantly different (Kolmogorov-Smirnov test; p-value = 0.3136). B) A single codominant locus under selective advantage (s = 0.3) is considered for 25 separate simulations. Values were compared to theoretical values (red dashed line) at t = 10, 20, 30, 40 and 50 generations. Comparisons yielded no significant differences (5 t-test; all p-values > 0.407). C) The response to differing selection intensities. All loci were given random genotypic effects taken from a negative exponential distribution and random starting allele frequencies to create a range of phenotypes in the starting population. Truncation selection was applied in 25 separate simulations to allow differing proportions of the population to reproduce in order to create subsequent generations. Comparison to theoretical values (red dashed lines – mean of all 25 predictions) yielded significant differences. See supplement 1 for an explanation. D) Linkage disequilibrium decay for adjacent loci over time, compared to theoretical expectations (red dashed line). Initially, all loci were in complete linkage disequilibrium. Comparison at t = 10, 20, 30, 40 and 50 generations yield no significant differences (5 t-test; all p-values > 0.409).

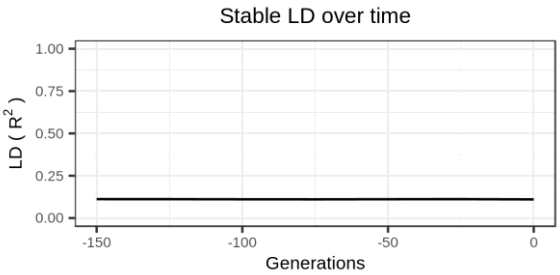
Supplement 2: Population characteristics

Here, we present information to ensure that populations have reached mutation-drift equilibrium. We show the shape of the allele frequency distribution and calculate mean LD between loci. The allele frequency distribution shows a U-shape, as expected from a population under equilibrium (Davoudi et al., 2018). Furthermore, mean LD between loci is stable over time, indicating that the population has reached an equilibrium between forces that create LD (genetic drift, mutations) and forces that break down LD (recombination).

Additionally, we show the distribution of genotypic effects (Supplemental Fig 2B). Values are taken from a negative exponential distribution with scaling parameter $\beta = 1 / \lambda = 1$. This distribution approximates genotypic effects found in real data, so that few loci have large genotypic effects, and most loci have small or near-zero genotypic effects. Finally, we verify that populations are genetically different using Nei's genetic distance (Nei, 1972). We find that the genetic distance between two populations increases approximately linearly with evolutionary time (data not shown).

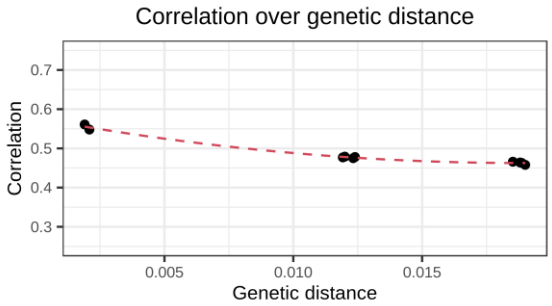


Supplemental Figure 2 **Simulated chromosomes capture biological reality. A)** Allele frequencies at the start of the E&R experiment. A U-shaped allele frequency distribution is expected from a population at equilibrium. **B)** Genotypic effect sizes per locus are taken from a negative exponential distribution with scaling parameter λ= 1. The grey histogram shows actual values used in simulations, whereas the red dashed line shows the underlying distribution. **C)** Linkage disequilibrium decay over distance on a single chromosome. The red dashed line indicates the mean linkage disequilibrium.

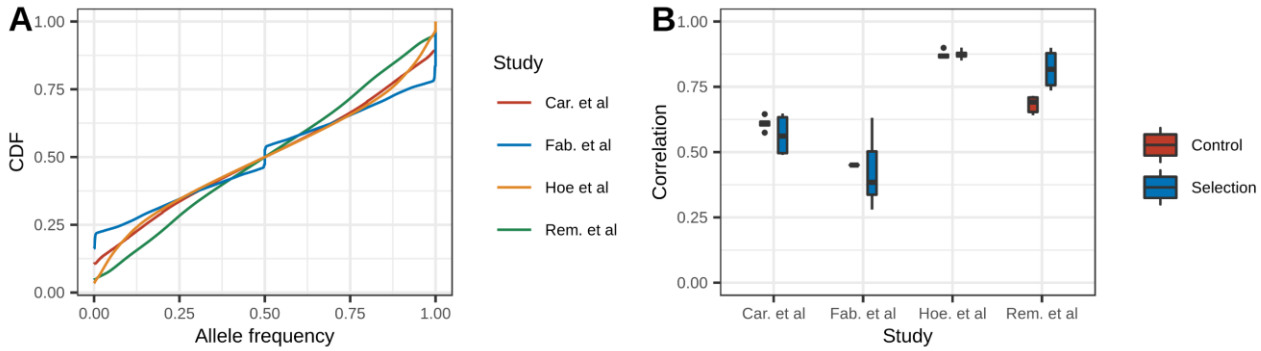


Supplemental Figure 3 **Linkage disequilibrium constant over time.** Mean linkage disequilibrium was calculated every 25th generation for 150 generations leading up to the start of the E&R experiment.

Supplemental figures



Supplemental Figure 4 **Correlation between population decreases with genetic distance**. We find that the correlation ρ(ΔAF) decreases as the genetic distance between populations increases. Black dots represent all pairwise comparisons between 5 replicates. The red line is a second-degree polynomial fitted to the data



Supplemental Figure 5 **Comparison between recent E&R studies. A)** The empirical cumulative distribution function of the allele frequencies per study. **B)** Correlations for all control and selection lines per study.

References

Barghi, N., Tobler, R., Nolte, V., Jakšić, A. M., Mallard, F., Otte, K. A., Dolezal, M., Taus, T., Kofler, R., & Schlötterer, C. (2019). Genetic redundancy fuels polygenic adaptation in Drosophila. *PLOS Biology*, *17*(2), e3000128. https://doi.org/10.1371/journal.pbio.3000128

Barghi, N., Tobler, R., Nolte, V., & Schlötterer, C. (2017). Drosophila simulans: A species with improved resolution in evolve and resequence studies. *G3: Genes, Genomes, Genetics*, *7*(7), 2337–2343. https://doi.org/10.1534/g3.117.043349

Barton, N. H. (2000). Genetic hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *355*(1403), 1553–1562. https://doi.org/10.1098/rstb.2000.0716

Barton, N. H. (2017). How does epistasis influence the response to selection? *Heredity*, *118*(1), 96–109. https://doi.org/10.1038/hdy.2016.109

Barton, N. H., & Turelli, M. (2004). Effects of genetic drift on variance components under a general model of epistasis. *Evolution*, *58*(10), 2111–2132. https://doi.org/10.1111/j.0014-3820.2004.tb01591.x

Berg, J. J., & Coop, G. (2014). A Population Genetic Signature of Polygenic Local Adaptation. *BioRxiv*. https://doi.org/10.1101/000026

Carlborg, Ö., & Haley, C. S. (2004). Epistasis: Too often neglected in complex trait studies? *Nature Reviews Genetics*, *5*(8), 618–625. https://doi.org/10.1038/nrg1407

Carnes, M. U., Campbell, T., Huang, W., Butler, D. G., Carbone, M. A., Duncan, L. H., Harbajan, S. V., King, E. M., Peterson, K. R., Weitzel, A., Zhou, S., & Mackay, T. F. C. (2015). The Genomic Basis of Postponed Senescence in Drosophila melanogaster. *PLOS ONE*, *10*(9), e0138569. https://doi.org/10.1371/journal.pone.0138569

Chen, H. (2018). An Efficient Computational Approach for Constructing the Allele Frequency Spectrum of Populations with Arbitrary Complex History. *BioRxiv*. https://doi.org/10.1101/456335

Comeron, J. M., Ratnappan, R., & Bailin, S. (2012). The Many Landscapes of Recombination in Drosophila melanogaster. *PLoS Genetics*, *8*(10). https://doi.org/10.1371/journal.pgen.1002905

Daub, J. T., Hofer, T., Cutivet, E., Dupanloup, I., Quintana-Murci, L., Robinson-Rechavi, M., & Excoffier, L. (2013). Evidence for Polygenic Adaptation to Pathogens in the Human Genome. *Molecular Biology and Evolution*, *30*(7), 1544–1558. https://doi.org/10.1093/molbev/mst080

Davoudi, P., Abdollahi-Arpanahi, R., & Nejati-Javaremi, A. (2018). The impact of QTL allele frequency distribution on the accuracy of genomic prediction. *Archives Animal Breeding*, *61*(2), 207–213. https://doi.org/10.5194/aab-61-207-2018

Fabian, D. K., Garschall, K., Klepsatel, P., Santos-Matos, G., Sucena, É., Kapun, M., Lemaitre, B., Schlötterer, C., Arking, R., & Flatt, T. (2018). Evolution of longevity improves immunity in Drosophila. *Evolution Letters*, *2*(6), 567–579. https://doi.org/10.1002/evl3.89

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (Fourth). Pearson Prentice Hall.

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

Gazal, S., Finucane, H. K., Furlotte, N. A., Loh, P. R., Palamara, P. F., Liu, X., Schoech, A., Bulik-Sullivan, B., Neale, B. M., Gusev, A., & Price, A. L. (2017). Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nature Genetics*, *49*(10), 1421–1427. https://doi.org/10.1038/ng.3954

Haller, B. C., & Messer, P. W. (2017). SLiM 2: Flexible, interactive forward genetic simulations. *Molecular Biology and Evolution*, *34*(1), 230–240.

Hansen, T. F. (2006). The Evolution of Genetic Architecture. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 123–157. https://doi.org/10.1146/annurev.ecolsys.37.091305.110224

Hansen, T. F. (2013). Why epistasis is important for selection and adaptation. *Evolution*, *67*(12), 3501–3511. https://doi.org/10.1111/evo.12214

Hartl, D. L., & Clark, A. G. (1997). *Principles of population genetics* (Vol. 116). Sinauer associates Sunderland, MA.

Heuvel, J. van den, Pannebakker, B., Kammenga, J., & Zwaan, B. (personal communication). A meta-analysis reveals network gene ontology enrichment as the conversed biological level for lifespan variation upon experimental evolution. *[Personal Communication]*.

Hill, W. G. (2017). "Conversion" of epistatic into additive genetic variance in finite populations and possible impact on long-term selection response. *Journal of Animal Breeding and Genetics*, *134*(3), 196–201. https://doi.org/10.1111/jbg.12270

Hill, W. G., & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, *38*(6), 226–231. https://doi.org/10.1007/BF01245622

Hillers, K. J. (2004). Crossover interference. *Current Biology*, *14*(24), R1036–R1037.

Hoedjes, K. M., Heuvel, J. van den, Kapun, M., Keller, L., Flatt, T., & Zwaan, B. J. (2019). Distinct genomic signals of lifespan and life history evolution in response to postponed reproduction and larval diet in Drosophila. *Evolution Letters*, *3*(6), 598–609. https://doi.org/10.1002/evl3.143

Hu, T. T., Eisen, M. B., Thornton, K. R., & Andolfatto, P. (2013). A second-generation assembly of the Drosophila simulans genome provides new insights into patterns of lineage-specific divergence. *Genome Research*, *23*(1), 89–98.

Jha, A. R., Miles, C. M., Lippert, N. R., Brown, C. D., White, K. P., & Kreitman, M. (2015). Whole-Genome Resequencing of Experimental Populations Reveals Polygenic Basis of Egg-Size Variation in Drosophila melanogaster. *Molecular Biology and Evolution*, *32*(10), 2616–2632. https://doi.org/10.1093/molbev/msv136

Kessner, D., & Novembre, J. (2014). forqs: Forward-in-time simulation of recombination, quantitative traits and selection. *Bioinformatics*, *30*(4), 576–577.

Kofler, R., & Schlötterer, C. (2014). A Guide for the Design of Evolve and Resequencing Studies. *Molecular Biology and Evolution*, *31*(2), 474–483. https://doi.org/10.1093/molbev/mst221

Mackay, T. F. C. (2001). The Genetic Architecture of Quantitative Traits. *Annual Review of Genetics*, *35*(1), 303–339. https://doi.org/10.1146/annurev.genet.35.102401.090633

MacKay, T. F. C., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., Casillas, S., Han, Y., Magwire, M. M., Cridland, J. M., Richardson, M. F., Anholt, R. R. H., Barrón, M., Bess, C., Blankenburg, K. P., Carbone, M. A., Castellano, D., Chaboub, L., Duncan, L., … Gibbs, R. A. (2012). The Drosophila melanogaster Genetic Reference Panel. *Nature*, *482*(7384), 173–178. https://doi.org/10.1038/nature10811

Nei, M. (1972). Genetic Distance between Populations. *The American Naturalist*, *106*(949), 283–292. https://doi.org/10.1086/282771

Neuenschwander, S., Hospital, F., Guillaume, F., & Goudet, J. (2008). quantiNemo: An individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics*, *24*(13), 1552–1553. https://doi.org/10.1093/bioinformatics/btn219

Nuzhdin, S. V., & Turner, T. L. (2013). Promises and limitations of hitchhiking mapping. *Current Opinion in Genetics & Development*, *23*(6), 694–699. https://doi.org/10.1016/j.gde.2013.10.002

Palmieri, N., Nolte, V., Chen, J., & Schlötterer, C. (2015). Genome assembly and annotation of a Drosophila simulans strain from Madagascar. *Molecular Ecology Resources*, *15*(2), 372–381.

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current Biology*, *20*(4), 208–215. https://doi.org/10.1016/j.cub.2009.11.055

Pritchard, J. K., & Przeworski, M. (2001). Linkage Disequilibrium in Humans: Models and Data. *The American Journal of Human Genetics*, *69*(1), 1–14. https://doi.org/10.1086/321275

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R., & Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature*, *411*(6834), 199–204. https://doi.org/10.1038/35075590

Remolina, S. C., Chang, P. L., Leips, J., Nuzhdin, S. V., & Hughes, K. A. (2012). Genomic basis of aging and life-history evolution in Drosophila Melanogaster. *Evolution*, *66*(11), 3390–3403. https://doi.org/10.1111/j.1558-5646.2012.01710.x

Sargolzaei, M., & Schenkel, F. S. (2009). QMSim: A large-scale genome simulator for livestock. *Bioinformatics*, *25*(5), 680–681. https://doi.org/10.1093/bioinformatics/btp045

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals—Mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, *15*(11), 749–763. https://doi.org/10.1038/nrg3803

Stevison, L. S., & Noor, M. A. F. (2010). Genetic and Evolutionary Correlates of Fine-Scale Recombination Rate Variation in Drosophila persimilis. *Journal of Molecular Evolution*, *71*(5), 332–345. https://doi.org/10.1007/s00239-010-9388-1

Stranger, B. E., Stahl, E. A., & Raj, T. (2011). Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics. *Genetics*, *187*(2), 367–383. https://doi.org/10.1534/genetics.110.120907

Szendro, I. G., Schenk, M. F., Franke, J., Krug, J., & De Visser, J. A. G. M. (2013). Quantitative analyses of empirical fitness landscapes. *Journal of Statistical Mechanics: Theory and Experiment*, *2013*(1). https://doi.org/10.1088/1742-5468/2013/01/P01005

Templeton, A. R. (2000). Epistasis and complex traits. *Epistasis and the Evolutionary Process*, *4157*.

Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J., & Richards, J. B. (2018). Genetic architecture: The shape of the genetic contribution to human traits and disease. *Nature Reviews Genetics*, *19*(2), 110–124. https://doi.org/10.1038/nrg.2017.101

Tobler, R., Franssen, S. U., Kofler, R., Orozco-terWengel, P., Nolte, V., Hermisson, J., & Schlötterer, C. (2014). Massive Habitat-Specific Genomic Response in D. melanogaster Populations during Experimental Evolution in Hot and Cold Environments. *Molecular Biology and Evolution*, *31*(2), 364–375. https://doi.org/10.1093/molbev/mst205

Turner, T. L., & Miller, P. M. (2012). Investigating natural variation in drosophila courtship song by the evolve and resequence approach. *Genetics*, *191*(2), 633–642. https://doi.org/10.1534/genetics.112.139337

Turner, T. L., Stewart, A. D., Fields, A. T., Rice, W. R., & Tarone, A. M. (2011). Population-Based Resequencing of Experimentally Evolved Populations Reveals the Genetic Basis of Body Size Variation in Drosophila melanogaster. *PLoS Genetics*, *7*(3), e1001336. https://doi.org/10.1371/journal.pgen.1001336

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S. J. van der, Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Mulbregt, P. van. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 1–12. https://doi.org/10.1038/s41592-019-0686-2

Vlachos, C., & Kofler, R. (2018). MimicrEE2: Genome-wide forward simulations of Evolve and Resequencing studies. *PLOS Computational Biology*, *14*(8), e1006413. https://doi.org/10.1371/journal.pcbi.1006413

Wall, J. D., & Pritchard, J. K. (2003). Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews Genetics*, *4*(8), 587–597. https://doi.org/10.1038/nrg1123

Weinreich, D. M., Watson, R. A., & Chao, L. (2005). Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*, *59*(6), 1165–1174. https://doi.org/10.1111/j.0014-3820.2005.tb01768.x

Wellenreuther, M., & Hansson, B. (2016). Detecting Polygenic Evolution: Problems, Pitfalls, and Promises. *Trends in Genetics*, *32*(3), 155–164. https://doi.org/10.1016/j.tig.2015.12.004

Whitlock, M. C., Philips, P. C., Moore, F. B., & Tonsor, S. J. (1996). Multiple fitness peaks and epistasis. *Annual Review of Ecological Systems*, *26*, 601–29.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wolf, L. F. of L. S. J. B., Brodie, A. P. in the D. of B. E. D., & Wade, P. in the D. of B. M. J. (2000). *Epistasis and the Evolutionary Process*. Oxford University Press.

Zanini, F., & Neher, R. A. (2012). FFPopSim: An efficient forward simulation package for the evolution of large populations. *Bioinformatics*, *28*(24), 3332–3333. https://doi.org/10.1093/bioinformatics/bts633

Zhou, D., Udpa, N., Gersten, M., Visk, D. A. W., Bashirb, A., Xue, J., Frazer, K. A., Posakony, J. W., Subramaniam, S., Bafna, V., & Haddad, G. G. (2011). Experimental selection of hypoxia-tolerant Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(6), 2349–2354. https://doi.org/10.1073/pnas.1010643108