Title here

# Abstract

# Introduction

Quantitative genetics aims to quantify genetic diversity; this diversity has broad implications for adaptation, it is well described how diversity enhances efficiency of adaptation; adaptation with more Va = faster, more efficient; particularly seen in the case of quantitative genetics, where stabilizing selection is often assumed; different story with the maintenance of variation around a fitness optimum, i.e. after the adaptive walk what happens?; several models have appeared over the last 50 years to explain the maintenance of variation; continuum of alleles vs diallelic; within continuum of alleles, the approximate distribution of allelic effects depends on the relative mutation rate to selection strength; in other words, the strength of new mutations to standing genetic variation; models over the last 50 years have failed to explain natural diversity observed in populations;

The ubiquity of adaptation in evolutionary studies is telling of the impact of Darwin’s seminal work. Over 4600 papers featuring the keyword ‘adaptation’ were published in Nature research journals in 2019. The allure of adaptation comes from the power of Darwin’s theory to explain natural diversity both within and between populations (Brady *et al.* 2019). Yet explanations of diversity via Darwin’s theory have been misinterpreted before – prior to Williams publishing his thesis (1966), the theory of ‘group adaptation’, whereby adaptation is driven by altruistic mutations that benefit populations at the cost of the individual gene, was commonplace and well-regarded. The focus on adaptive traits, and the ability of populations to adapt to new situations is wonderfully intuitive, however, populations are never perfectly adapted: trait values are rarely optimal, populations decline, and extinctions are commonplace (Brady *et al.* 2019). The extent of maladaptation, where populations are stable at some distance from a fitness optimum, seems wide, however the extent of maladaptation is rarely discussed (Nesse 2005). The keyword ‘maladaptation’ was mentioned in just 45 papers in Nature research journals published in 2019, yet it is expected to .

**Evidence for stab sel: in gene regulation, Hodgins-Davis et al – this paper, also m/s/d balance for quant gen models;**

The extent of maladaptation in natural populations has been debated for some time – are populations likely to be restricted in their fit to a phenotypic optimum? Is such an optimum attainable for populations, or is there more likely to be local, stable optima that are difficult to escape?

Much of the variation between and within populations is the result of continuous variability in traits individuals in those populations possess. Differences in such quantitative traits lead to adaptation and speciation. Underpinning adaptation by quantitative traits is additive genetic variance (VA), the heritable component of variation; the amount of which has mystified quantitative geneticists for close to 100 years. Predicting levels VA is reliant on mutation rate and selection strength: both of which are notoriously difficult to estimate in natural populations. Many quantitative traits are subject to stabilizing selection, where intermediate trait values have maximum fitness. Theory suggests that adaptation via stabilizing selection should be more efficient with higher standing VA, and that the selective fixation of this standing variation should decrease VA as adaptation takes place (Fisher 1930; Lande 1975). However, these expectations have not always coincided with observed data.

Depletion of VA with stabilizing selection has been shown both experimentally , and analytically, however increasing amounts of more modern work show no effect of selection strength on VA. Sztepancz and Blows (2017) showed that there was no relationship between genetic variation and the strength of stabilizing selection in *Drosophila serrata­­­.* More modern analysis of Fisher’s (1930) geometric model (upon which stabilizing selection is built) has shown that when the individual effect of selection in alleles is weak, stabilizing selection has a minimal effect on VA, as drift at any individual locus may compete with selection to adjust allele frequencies (Barton 2017). Discrepancies such as these can perhaps be explained by the relative effect of stabilizing selection, depending on where a population is relative to the optimum. When populations are far from the optimum, mutations act under a directional selection model, where larger mutations that bring an individual closer to the optimum are more beneficial (Zhang 2012). However, as populations approach that optimum, large effect mutations become costly as they are more likely to drag populations further away from the optimum.

Barton 2017: selection negligible when individual allles is weak and comparable to drift

Thornton 2019: when phenotypes approach optimum, strength of selection on indivual muts decreases effect on Va under infinitesimal model, selection gets more info when selection on

Zhang 2012: when phenotypes near optimum, selection is stab, while far away, closer to dir sel

Sztep: dir sel more common in nature? Populations more commonly maladapted?

At the heart of the evolutionary sciences is the need to understand the natural world’s diversity. Darwin’s (1863; SOURCE) introduction of natural selection some 140 years ago led to increasingly accurate glimpses into the units of evolution, genes, and their movement through a population in response to selection (SOURCE). However, these movements, particularly in a multivariate trait space, become a challenging realm to predict (SOURCE; Lande 1979, 80 etc.). To navigate this space, it is necessary to reduce the predictors of trait trajectories to their principles: how they affect additive genetic variance, the heritable component of trait variability.

Additive genetic variability, VA, is regarded as the most important predictor of a population’s adaptability (Lynch and Lande 1998; Aguirre *et al.* 2014; Careau *et al.* 2015), and hence it’s trajectory through time towards a phenotypic optimum. Although a multitude of stochastic and deterministic processes also contribute to the population’s total trait variability, VA is heritable, and therefore predictive of a population’s trajectory over micro-evolutionary time. The VA of a population determines the phenotypic space that population can explore. Hence, it is predicted that populations with large amounts of VA are best suited to adapting to novel environments (Barton and Charlesworth 1998). Such an example is X. However this is not always the case, standing genetic variation is characterized by a variety of architectural and population-level constraints such as rates of pleiotropy, selection strength, additive effect size, linkage, and deleterious mutation/background selection (SOURCE). For example, under infinitesimal models, selection has a trivial impact on standing variation (Barton 2017).

including genetic drift, selective pressures, additive effect sizes, between- and within-gene interactions, and heritability (SOURCES).

* Natural diversity, population movements in trait space
* Heritable variation
* Stabilising selection, effect on variation/need for variation vs drift
  + Expected to remove variation, mutation alone can’t explain why in natural populations we see so much variation: why?
* Additive effect sizes, effects on variation
* Background selection, effect on variation
* Population genetics expectations of variation under bkg sel, additive effects

# Methods

Using the forward-genetics modelling package SLiM 3.4 (Haller and Messer 2019), we constructed two models to explore a portion of the multivariate parameter space that explains genetic variability in natural populations. These parameters included genome wide recombination rate, the additive effect size distribution, the rate of universal pleiotropy, mutational correlation between traits, and the selection strength multiplier, (Table 1). The relative rate of non-QTL, deleterious mutation compared to trait mutations was also varied across models. This parameter led to two alternate outcomes that could influence variation and adaptation: either the reduction in QTL mutation rate due to increasing deleterious mutation rate could cause observed differences, or the effect of the deleterious mutations on fitness could be attributed to the differences. Preliminary analyses indicated that the ratio of QTL mutations to deleterious mutations remained constant across increasing levels of this parameter (Figure S1). This suggests that a similar deleterious load was experienced across populations, and that the effects of increasing this rate were attributable to changes in QTL mutation rate rather than the deleterious effects of non-QTL mutations. The highest QTL mutation rates were experienced by models with low rates of deleterious mutation, and vice versa. Thus, models with high mutation rate and low selection strength (deleterious mutation rate < 0.33; > 660) approximated the Kimura-Fleming-Lande Gaussian approximation of allelic effects (Kimura 1965; Lande 1975; Fleming 1979), while models with low mutation rates and high selection strength approximated Turelli’s (1984) House-of-Cards model. Among all parameter combinations, multiple conditions and assumptions were shared.

## Common model elements

Both of my experimental models consisted of a SLiM 3.4 model simulating a Wright-Fisher population of 8000 diploid individuals evolving over 100,000 generations. Populations were assumed to be completely allopatric. Populations first were subject to 50,000 generations of burn-in to build standing variation to mutation-drift balance (figure S1). Individuals were characterized by 8 traits, controlled by 100 loci each. Each trait had an identical effect on fitness, forming a ‘mega-trait’ with varying variance-covariance structures depending on pleiotropy rates. Each locus was assumed to have identical length, and each base pair within it mutationally independent – hence mutations occurred at an arbitrary position within the locus. This assumption is supported by a study by Thornton (2019), which found that within-locus differences in linkage had no average effect on either genetic variance or the mean trait value, indicating within-locus independence. In addition, the average number of base pairs per locus is highly conserved within eukaryotes (Xu *et al.* 2006), lending credence to the assumption of equal gene length. Mutations were assumed to be completely additive in effect, with no dominance or epistatic interactions, aside from additive epistasis occurring as a result of the fitness function. Mutational effects were in phenotypic units, an arbitrary unit denoting relative differences in phenotype. All loci were assumed to be on the same chromosome, with genetic distance being determined by the recombination rate parameter, r (Table 1). Both models had a genome-wide germline mutation rate of 8.045x10-6 per locus per generation, based on an average of five groups of eukaryotes (Aston *et al.* 2017).

The effective population size, Ne = 8000, was chosen to compromise between computational performance and the effect of genetic drift on populations under stabilizing selection. This value results in weak genetic drift in comparison with the strength of selection (under strong selection pressures), and appropriate standing genetic variation following burn-in to allow for adaptation (Lynch and Lande 1998).

Mutational effects on trait values were sampled from a normal distribution,

where λ is the additive effect size (Table 1). In the case of pleiotropy, a multivariate normal distribution was used, where n = 8, and

where **Σ** is a covariance matrix with diagonal values equal to λ and non-diagonals pulled from a normal distribution:

where is the parameter value of mutation correlation. **Σ** was ensured to be positive definite by multiplication with its transpose,

Non-trait deleterious mutations had fitness effects sampled from a gamma distribution:

Where and (SLiM Manual). This describes a distribution of weak deleterious mutations on average.

All models were subject to 50,000 generations of burn-in, where mutations accumulate until the population reaches mutation-drift equilibrium. This is tracked as heterozygosity through the simulation, where mutation-drift equilibrium occurs when:

where µ represents the per-locus mutation rate per generation (Kimura and Crow 1964). A population at equilibrium was assumed sufficiently burnt in. Trials indicated that 50,000 generations of burn-in was sufficient for our population size (Figure S2). Deleterious mutation/mutation rate lowered the value of away from expectation in initial burn-in tests, however an alternative equilibrium was reached, satisfying the requirements of burn-in regardless of the parameter (Figure S2). During the simulation run, trait variances, covariances, and trait means were collected every 500 generations to track distances from the optimum and trait variability over time. At the end of the simulation, the allelic effects of segregating mutations in all populations were collected.

## Model specific characteristics

After reaching equilibrium, populations evolved for 100,000 generations of neutral drift or stabilizing selection, depending on the treatment. Neutral drift entailed no change from the properties of the burn-in, whereas stabilizing selection imposed a fitness function on phenotypes, invoking a multivariate optimum a fixed distance from the population mean phenotype post-burn-in. The position of the optimum is defined as:

Where is the vector of phenotype means, is the per-locus, per-generation mutation rate; , is the number of mutational steps to reach the optimum, and is the number of generations of burn-in. For our purposes, 8.045x10-6, 100, and . This distance was close to the original phenotypes, meaning most of the simulation (approximately 98000 generations of the simulation) investigated the maintenance of variation at a fitness optimum.

The fitness of an individual in the population was defined as:

Where s is the selection coefficient, represents the gradient of the selection curve, n is the number of traits, and xn is the phenotype for trait n. To ensure a theoretical minimum and maximum fitness, s was fixed at 0.9, ensuring minimum fitness was , and maximum fitness was 1. This results in individuals at the optimum being at most ten times as fit as those infinitely far from the optimum. The model-specific maximum fitness difference depends on, which adjusts the realized fitness gradient via the curvature of the fitness function.

## Model Parameterization

Five parameters were shared between models, with a sixth for testing selection (Table 1). These were sampled using a Latin hypercube sampling design, with 1024 parameter combinations testing the null model, and 256 for the selection model (Figure S3). The hypercube sampling was necessary to explore the entire parameter space, as simple factorial designs would have been impractical to achieve. Each hypercube sample represents a combination of parameters, with the total set of samples designed to maximize the distance between samples (sampling more of the total space), and minimize correlations between them (Helton and Davis 2003). Hypercube samples were generated using the R packages ‘DoE.Wrapper’ and ‘LHS’, using the maximin algorithm (Melo *et al.* 2015; R Developmental Core Team 2019). Each sample/model was repeated 100 times, using 100 seed values fed to SLiM. These seeds were randomly sampled from a uniform distribution of the total range of unsigned 32 bit integers (1 to 232 – 1) (R Developmental Core Team 2019). The array of parameter combinations and replicates was processed across 1152 cores on the University of Queensland’s Tinaroo high performance computing (HPC) system, using embedded Nimrod scripts to feed parameter/seed combinations to individual SLiM processes.

## Analysis

Despite not all data conforming to normality, no data was transformed owing to the large sample sizes. Previous work into the robustness of t-tests, and F-tests have shown that departures from normality can usually still provide reliable estimates, provided the number of observations is large enough that coefficient estimates are approximately normally distributed due to the central limit theorem (Lumley *et al.* 2002). This was verified with diagnostic tools in the R package “jtools” (Long 2020). In terms of regression analysis, heteroscedasticity can still remain a problem, even with large sample sizes. To account for this, we used Eicker-Huber-White (EHW) HC2 or HC3 robust standard errors in our linear regression models via the ‘estimatr’ package in R (Eicker 1967; Huber 1967; White 1980; Hayes and Cai 2007; Blair 2020). Due to the large sample sizes (128000 total models), we were able to find significant differences between groups with extremely small effect sizes. To ensure we focused only on biologically meaningful differences, we calculated the relative contributions of factors to the appropriate regression, using the Lindeman, Merenda and Gold method (Lindeman 1980), explaining only the factors that contributed meaningfully to variation.

For analysis, the interaction between and mutation rate was treated as a ‘model’ parameter, indicating whether the hypercube sample approximated House-of-Cards allelic effects, or Gaussian effects. An additional model type, ‘Null’, summarized the models with no selection and any mutation rate. Remaining models with intermediate selection strengths and deleterious mutation rates were not considered for analysis, although that remains an exciting prospect for the future. Additive effect size, recombination rate, pleiotropy rate, and mutational correlation hypercube values were binned into low, medium, and high categories for simpler analysis.

We compared responses at the final generation of the simulation (100,000) across all analyses. Trait variances and covariances were pooled and averaged to form a ‘mega-trait’ average variance and covariance, since traits were functionally identical. In addition, we computed the population mean Euclidean distance from the optimum for each replicate and model:

Where pi and qi are the population mean and optimum value, respectively, for trait *i*.

To determine the effects of CoA model on adaptation, we explored the distribution of final distances from the optimum, finding a distinct ‘dead zone’ where distances were not represented. We used this dead zone to classify models into two categories: adapted, or maladapted. Adapted models had distances from the optimum less than 16 units, and maladapted with distances greater than 16 units. We used a Chi-square test followed by an odds-ratio post-hoc to determine the differences in representation among CoA models in adapted and maladapted categories. Following this, we discarded maladapted populations, choosing to focus on investigating the genetic architectures underlying adapted populations.

To evaluate the effects of genetic architecture on adaptation under the CoA models, we used EHW-error multiple regression models to determine the effects of CoA model type, additive effect size, recombination rate, pleiotropy rate, and mutational correlations between traits on distance from the optimum, additive variance, and trait covariances. We compared estimated marginal means with Tukey correction to assess differences between Continuum of Alleles models, and parameter levels.

We also collected the mutational effects of segregating alleles at the end of the simulation. With this, we compared mean distributions of allelic effect sizes in adapted populations according to additive effect size with multivariate multiple regression. Responses included mean allelic effect, variance, and kurtosis of the distribution, as well as the mutation counts contributing to VA within each model. We adjusted for heteroskedasticity with EHW robust standard errors. Multiple regressions were calculated across 50 replicates owing to RAM limitations.

# Results

## Tracking population dynamics over time

To determine whether populations were under mutation-selection-drift balance by the end of the simulation, we plotted additive variance and covariance over time across selection strengths. We reasoned that the joint effects of mutation, which creates variance, and drift and selection, which remove variance, would lead to stable levels of genetic variability over long periods of time. We found that after 100,000 generations (2 days of run-time per model), variance increased asymptotically in all models (Figure 3) suggesting that levels of genetic variability were unlikely to change significantly in longer model runs. Mean additive variance was consistently higher under a Gaussian model, whereas it remained low and almost constant in the House-of-Cards models (Figure 3A). Both selection models clearly behaved differently from a null model where genetic drift was expected to dominate. Covariance between traits acted similarly across models (Figure 3B). Knowing that by generation 100,000 we are at mutation-selection-drift equilibrium, we can now investigate whether populations are well-adapted under different selection and genetic models.

## General patterns of adaptation with Continuum of Alleles models

We explored the distribution of Euclidean distances around a phenotypic optimum under House-of-Cards and Gaussian models of allelic effects and compared them to a null model without selection (Fig. 4A). Both Gaussian and House-of-Cards models showed a small proportion of populations that came within 16 phenotypic units from the optimum, with a visible division between adapted and maladapted populations (Fig. 4). The ‘dead space’ that separated these populations did not exist in the null model. To further explore this bimodality, we examined the differences between models in their ability to reach the adapted space. Populations were more likely to be found in the adapted zone if they belonged to either selection model over the null model (χ2 = 9602.1, df = 2, p < 0.0001). 15.23% of Gaussian populations reached the adapted space, while House-of-Cards populations reached this 16.1% of the time. By contrast, 0.53% of null populations reached the adapted space. A post-hoc odds ratio test found significant differences between null and Gaussian (OR = 33.566, 95% C.I. = 29.5, 38.2, p < 0.0001) and null and House-of-Cards (OR = 35.872, 95% C.I. = 31.5, 40.85, p < 0.0001), but not between Gaussian and House-of-Cards (OR = 1.069, 95% C.I. = 0.93, 1.23, p = 0.35). To understand the underlying genetic architectures of populations that were able to come close to the phenotypic optimum, we compared the effects of genetic architecture on distance to the optimum (Fig. 5; Table 2), mean trait variance (Fig. 6), and mean trait covariance (Fig. 7) across the two selection models.

## Genetic architecture effects on adaptation with Continuum of Alleles models

We compared the effects of varying additive effect sizes, recombination rates, pleiotropy rates, and mutational correlations on Euclidean distances of populations close to the phenotypic optimum under Gaussian or House-of-Cards mutational models. Table 2 shows the mean effects of these variables on how close populations get to the phenotypic optimum, as well as the effects on trait variance and covariance. Although all genetic architecture parameters had significant effects on distance, variance and/or covariance, most of these effects were small in magnitude. For brevity, we discuss only the parameters that explain the most variation in distance, variance, and covariance. Variation in distance was explained mostly by pleiotropy (explaining 8.6% of total variation among models), model type (explaining 5.4% of variation, and additive effect size (explaining 2.8% of variation). Mean distance from the optimum was lowest when additive effect sizes were low (0.841 ± 0.181; Fig. 4); this did not change between CoA models (t921 = -0.422, p = 0.998). However, House-of-Cards models were more robust to changes in additive effect size than Gaussian models (t921 = -2.583, p = 0.01). When increasing effect size from low to high under a Gaussian mutation model, adapted populations’ mean distance from the optimum increased by 2.203 ± 0.232 phenotypic units (t921 = -9.504, p < 0.0001). The same change in effect size under a House-of-Cards model resulted in no significant change to mean distance (t921 = -0.587, p = 0.827). Figure 5 shows how patterns of adaptation varied between Continuum of Alleles models when increasing the variance of allelic effect sizes. Pleiotropy rate increased distance, however, there was no interaction between pleiotropy and model type (t921 = 0.843, p = 0.399; Fig. 4C). Increasing pleiotropy rate from low to high led to an average 1.261 ± 0.178 unit decrease in distance from the optimum (t921 = 7.099, p < 0.0001). These effects on distance were not necessarily mirrored with the effects of genetic architecture on trait variance, which was explained by additive effect size (45.1% of variation in trait variance), and its interaction with model type (explaining 14.4% of this variation).

On average, House-of-Cards models near the optimum had considerably more additive variance than Gaussian models (40.4 ± 18.72 units vs 2.6 ± 0.07 units; t921 = -2.019, p = 0.044). Under a Gaussian model, increasing the additive effect size of populations in the adapted zone marginally increased trait variance (t921 = -14.386, p < 0.0001; Fig. 5A), however this was not the case under a House-of-Cards model (t921 = -1.958, p = 0.123). Figure 6 shows how additive effect size interacts with Gaussian and House-of-Cards models to drive differences in variance in adapted populations. Note that several outliers were removed from Figures 6 and 7 owing to their distortion of the figures. These outliers had variance greater than 50 and covariance less than -5 (Fig. S3, S4). Similarly to variance, differences in covariance could be explained mainly be differences in additive effect size (explaining 46.4% of variation), and the interaction between effect size and the Continuum of Alleles model type (explaining 15.6% of variation).

Average trait covariance differed between models (t921 = 2.147, p = 0.032; Fig. 6), with Gaussian models carrying very little genetic covariance amongst traits (0.014 ± 0.005), and House-of-Cards models carrying slightly more (-3.616 ± 1.691). Increasing additive effect size from low to high in Gaussian models led to slight declines in covariance (a decrease of 0.039 ± 0.005; t921 = 7.526, p < 0.0001; Fig. 6A). No significant effect of increasing additive effect size on covariance was seen in House-of-Cards models (t921 = 1.937, p = 0.129). The difference in response to additive effect size between models was marginally insignificant (t921 = -1.929, p = 0.054). Figure 7 shows the effects of increasing additive effect variance and Continuum of Alleles model type on covariance.

These analyses therefore suggest that additive variance and covariance are rather robust under House-of-Cards models, and less so under Gaussian models. Additive effect size in particular seems important to understanding the interplay between adaptation and additive variance. We compared the proportions of CoA models that reached the optimum according to their additive effect size, finding 36.12% of low additive effect size models were adapted, versus 2.29% of medium-effect populations, and 0.19% of high-effect populations (χ2 = 1572.13, df = 2, p < 0.0001). To analyze the underlying cause of these variances, covariances, and by extension, distance to the phenotypic optimum, we need to study the underlying allelic effect size distributions of the models. We compared the means, variances, kurtosis, and count of mutations contributing to these distributions across models to understand the mutational limitations imposed by genetic architectures under the two CoA models.

## Allelic effect size distributions with Continuum of Alleles models

The distributions of allelic effects are dependent on several parameters: the mean of the effects, which may be biased in some direction by genetic architectures, the variance of the distribution, indicating the variability in size-effects that are sampled, and the kurtosis of the distribution, indicating the rarity of large-effect alleles. To assess the mutational bias of models, we first compared the means of distributions across models and genetic architectures. The resulting regression was insignificant (F17, 411 = 1.127, p = 0.325, Adjusted R2 = 0.189), indicating a lack of directional mutational bias. We then turned our attention to the variance of distributions to understand the constraints that genetic architectures may apply to mutational models (F17, 411 = 55.04, p < 0.0001, Adjusted R2 = 0.851). Additive effect size explained 66.2% of total variability between models. Under a Gaussian model, increasing additive effect size from low to medium significantly increased allelic effect variance by 6.02 ± 0.372 phenotypic units (t411 = -16.188, p < 0.0001; Fig. 8), however no significant difference occurred for increasing variance from low to high or from medium to high. No significant changes to variance with increasing effect size were seen under House-of-Cards models. Figure 8 shows the distributions of allelic effects with changing additive effect size variability under Gaussian and House-of-Cards models. Leading on from the variance of allelic effects is the kurtosis, which describes the rarity of large-effect alleles.

Kurtosis differed significantly across models and genetic architectures with additive effect size variance explaining 31.9% of variability between models, and the interaction between additive effect size and model type contributing another 15.5% (F17, 411 = 12.36, p < 0.0001, Adjusted R2 = 0.6). Under the Gaussian model, increasing additive effect size from low to medium increased kurtosis by 0.985 ± 0.159 (t411 = -6.206, p < 0.0001). No analogous effect was seen under the House-of-Cards model (t411 = -0.944, p = 0.6130). As well as the distributions of allelic effects, the absolute counts of mutations contributing to each distribution gives an indicator of the genetic diversity of populations.

To assess the effects of genetic architecture and models on promoting many or few alleles, we compared mutation counts between models, finding significant differences among models (F17, 411 = 580.2, p < 0.0001, Adjusted R2 = 0.94). Model type contributed the most to explaining mutation count variation among models, describing 58.4% of among-model variation. Pleiotropy rate explained 10% of variation, however this is explicated by each pleiotropic mutation contributing multiple effects with a single mutation. The mean number of mutations in Gaussian models was considerably higher than that of House-of-Cards models, but over a large range of values: 1516 ± 6608 mutations for Gaussian models versus 374 ± 114 for House-of-Cards (t411 = 0.173, p = 0.863).

# Discussion

Our findings show that populations under Gaussian or House-of-Cards models can adapt to fitness optima under stabilizing selection, however the chance of doing so is quite low (a maximum of 16.1% of House-of-Cards populations reached distances close to the optimum; Fig. 4A). Hence, maladaptation seems quite common, at least under a population size of 8000 and the levels of drift associated with that. This supports previous predictions of maladaptation prevalence: maladaptation should be common given the capacities of selection, drift and inbreeding depression to remove additive variation, and hence the ability of populations to respond quickly to environmental changes (Crespi 2000; Aguirre *et al.* 2014; Brady *et al.* 2019). Among maladapted populations, House-of-Cards and Gaussian models had high variability in their final distance to the optimum (Fig. 4A), comparable to null models. In these populations, drift is likely to overcome selection strength; a result of a drift-barrier (Lynch *et al.* 2016).

Drift-barriers arise when weakly selected loci are unable to overcome the strength of drift (Lynch *et al.* 2016) (Gardon *et al.* 2020). This problem is especially prevalent in small populations where drift is expected to dominate, however large populations can also experience this if these loci are weak enough (Lynch 2010; Gardon *et al.* 2020). Evidence for these barriers are scarce in natural populations, however Gardon et al. (2020) found evidence for relaxed selection in genes inherited from small ancestral clades in *Prochlorococcus marinus.* In comparison, evidence for strong negative selection was found in more recent genes, arising in the much larger derived population (Gardon *et al.* 2020). The large variability in distances from the optimum in maladapted populations here is analogous to Gardon’s findings, indicating strong drift among both House-of-Cards and Gaussian populations. Since most traits are well adapted (Orr 1998), this suggests that selection must be reasonably strong to drive populations away from mildly maladapted phenotypes, particularly if population sizes are small.

Even among adapted populations, the effect of the drift-barrier might be pronounced in future responses to selection. Houle (1998) pointed out that selection can cause spatial variation in Ne across the genome by removing genetic variation, the key to initial responses to selection (Agashe *et al.* 2011). While the strength of selection seems necessary for driving adaptation past drift-barriers, we found no significant difference in the number of House-of-Cards (strong selection) and Gaussian (weak selection) populations that reached the optimum. Selection alone is not enough: mutational input must provide the variation for selection to act on without swamping the population with strongly deleterious large-effect alleles (Fisher 1930; Franssen *et al.* 2017).

In tandem with selection strength, mutation rate defines the differences between Gaussian and House-of-Cards models (Walsh and Lynch 2018). Gaussian models have higher mutation rates relative to selection strength (Lande 1975). This raises the expectation that Gaussian models should maintain more variability following adaptation and carry more mutations of small effect (Hodgins-Davis *et al.* 2015; Walsh and Lynch 2018). This is contrasted by the House-of-Cards model which is defined by strong selection and low mutation rates, leading to more intermediate sized effects being the most common to increase in frequency in the population (Turelli 1984; Hodgins-Davis *et al.* 2015). The variation in the size of mutational input is therefore extremely important to the expectations of these models: Gaussian models are expected to function with small effect sizes, while House-of-Cards are assumed to function by selecting moderately-sized alleles (Turelli 1984; Walsh and Lynch 2018). Adjusting effect size variation has implications for the efficacy of adaptation under these different models, as we have shown Gaussian and House-of-Cards models are not equally sensitive to changes in effect size variation.

House-of-Cards models were generally robust to changes in additive effect size, with distance from the optimum, variance, and covariance remaining similar across effect size variation treatments (Fig. 5, 6, 7). Gaussian models on the other hand were perturbed by increases to mutational effects, with wider distributions, and more maladaptation occurring under high mutational variance scenarios. This is due to differences in selection strengths between models. While at the optimum, most new mutations are deleterious under House-of-Cards models (Turelli 1984): the strong selective pressure on these populations leads to a constant mutational load that is unchanged by increasing mutational variance – new, large effect mutations are efficiently removed from the population regardless of if they are rare or common (Figure 6). Under Gaussian models, large-effect mutations are less deleterious and more common, and so persist in greater numbers, driving increases in additive variance (Hodgins-Davis *et al.* 2015), as seen in figures 6 and 8.

The phenotypic volatility of populations under high-variance mutation has implications for adaptation to new environments. For example, Gilbert and Whitlock (2017) showed through simulations that adaptation could occur through genetic architectures containing many genes of small-effect or few genes of large-effect, however adaptation in the populations under the few-genes-large-effect architecture took longer to achieve. In addition, they found that adaptation could succeed under two cases: (1) the classical example, where high genetic variation and small-effect alleles drive adaptation, and (2), where genetic variation may be low, but there are sufficient large effect alleles to drive adaptation (Gilbert and Whitlock 2017). If Gaussian populations move towards an optimum with high additive effect sizes, they fall in the middle of this: high expected additive variance from higher mutation rates (Walsh and Lynch 2018), and many large effect alleles that aid in the initial directional push towards an optimum (Zhang 2012). Thus, rapid movement towards the optimum is expected. However, these large effects might become a liability once the population arrives at the optimum.

Large effect alleles are likely to lower population fitness considerably under Gaussian models post-adaptive walk (Walsh and Lynch 2018). With small effect mutations, adaptation is likely to be slower (Gilbert and Whitlock 2017), but maladaptation post-walk will be considerably weaker: it will take many more mutations to move the population away from the optimum the same amount as a single large-effect mutation, and in this time, the weak selection of Gaussian models will be more able to reign in these effects. The balance between selection strength and mutation rate are critical for the Gaussian response to mutations of varying effect, however this balance under House-of-Cards models should lead to robustness against increased mutational variance when populations are at a phenotypic optimum.

Under a House-of-Cards model, populations face stronger selection relative to mutation rates (Turelli 1984), meaning that adaptation is driven by mutational variance rather than standing genetic variation (Walsh and Lynch 2018). While populations are at a phenotypic optimum, mutations are likely to be strongly deleterious, pulling populations towards maladaptation. Under House-of-Cards, mutation rates are low, reducing the chance of this happening. Furthermore, selection is strong: should a large-effect mutation arise, it is likely to be removed from the population quickly (Zhang 2012). This means that regardless of the mutational input, House-of-Cards populations can efficiently remove deleterious alleles, maintaining their position in phenotype space much more effectively than the ‘hotter’ Gaussian models.

To illustrate this theory, figure 9 represents the adherence of populations to an optima given their genetic architecture and Continuum of Alleles assumptions. Gaussian populations are poor at self-regulating their mutational distributions due to the combination of high mutation rates and weak selection. Under small effect sizes, Gaussian populations can reach the optimum, however they are more likely to become maladapted over time, due to the inefficiency of selection in removing weakly-deleterious mutations (Ohta 1973). House-of-Cards models on the other hand can maintain their mutational distributions, withstanding these large effects without being swamped by overwhelming numbers of large-effect mutations (Fig. 8, 9). Hence there is a trade-off: Gaussian models may be able to bring populations to the optimum quickly by using standing variation (Gilbert and Whitlock 2017), however under large additive effects, these populations are more likely to be maladapted. House-of-Cards models may adhere to the optimum more closely, however due to the reliance on new mutations, it will take longer for them to reach the optimum. Evidence for similar speed-accuracy trade-offs exist in gene network studies. Malcom (2011) found that a trade-off between adaptive accuracy and speed occurred in a simulation between two species competing in a variable environment. Smaller gene networks produced a competitive advantage in more temporally variable environments, whereas large gene networks resulted in increased accuracy when environments were more stable over time (Malcom 2011). Similarly, tropical diatom species have shown the ability to quickly adapt to increasing ocean temperatures, with the trade-off of reducing their photosynthetic efficiency and growth rate {Jin, 2018 #254}. But which side of this adaptability versus adaptedness {Leigh, 1970 #374} trade-off is most advantageous? The variability of the environments to which populations adapt will determine which model is most advantageous.

In spatially and/or temporally heterogeneous environments, Gaussian models should fare better than House-of-Cards: the rapid evolution towards the optimum offsets any accuracy costs, as these inaccuracies will be nullified by a new range shift, or drive populations towards a new local optimum {Malcom, 2011 #226}. Indeed, evidence for higher mutation rates in heterogeneous environments has been observed in experimental populations: Sniegowski {, 1997 #275} found in experimental populations of *Escherichia coli* that mutator phenotypes (which promote increased mutation rates through modifier genes) evolved in populations adapting to new environments. Simulations support this finding, with mutation rates controlled by temporal environmental variance, and being driven to low or high mutation rates depending on the degree of environmental variability {Gillespie, 1981 #299}. The greater additive variance introduced by increased mutation rates {Walsh, 2018 #26} could also provide Gaussian populations with a ‘head-start’ to begin adaptation quickly after an environmental event {Malcom, 2011 #226}, or a greater ability to radiate to new niches in the case of spatial environmental variation {Marques, 2019 #302}. In fact, under spatial gradients, large variability in effect sizes could seed populations with variation that allows their members to colonize differential micro-environments {Kagawa, 2018 #324}. In more homogeneous environments, where any movement from the current phenotype tends to be deleterious, House-of-Cards models should be favored.

Populations evolving by strong selection and low mutation should be advantaged in static environments. Without environmental change to perturb the optimum, almost all mutations are deleterious: not only by lethal mutations or non-focal trait mutations, but by almost all focal trait mutations moving populations away from the optimum {Matic, 1997 #262}{Walsh, 2018 #26}. Hence, the genetic load felt by populations under Gaussian models would be higher than that of House-of-Cards models under a stable environment. The finer control that House-of-Cards mutation-selection balances have on allelic frequencies (Fig. 8) allows for a better fit to the optimum, at the cost of slower adaptation (due to relying on new mutational variance to drive adaptation) {Walsh, 2018 #26}{Malcom, 2011 #226}. In addition, lower rates of mutation mean that populations consume less energy maintaining an adapted state relative to Gaussian populations {Lan, 2012 #322}, explaining the ‘colder’, less reactive behavior of adapted House-of-Cards models relative to Gaussian models (Fig. 3, 5, 6, 8, 9).

While additive effect size had strong effects on models, quantitative genetics theory also has predictions for the effects of pleiotropy, recombination, and mutational correlations that were either absent or weak (Table 2). This could be due to differences between expectations while maintaining variation post-adaptation versus approaching the optimum on the adaptive walk itself {Walsh, 2018 #26}. Zhang and Hill {Zhang, 2002 #325} found that the genetic variance maintained in a population depended very little on pleiotropy, and more so on the strength of realized stabilizing selection. While recombination is expected to increase additive variation {Barton, 1998 #24} and reduce covariation among traits {Lande, 1975 #168}, the parameterization may have been too narrow to see this effect over the much larger effect of additive effect size variation.

Among the limitations of this model include the chosen ranges of several parameters. While efforts were made to choose biologically meaningful ranges (Table 1), it was not always possible owing to performance restrictions. Our simulations took around 2 days to complete each, and although we were able to parallelize runs on a multi-node computing cluster, limitations on time and the number of parallel jobs led us to sample a smaller parameter space than originally intended. Recombination rate was sampled from 0 to 9.22x10-8 cM/Mb, which is a relatively high recombination rate in plants {Stapley, 2017 #95}, however small in comparison to some of the mutation rates in other taxa (for example, fungi can reach upwards of 100 cM/Mb of recombination rate {Stapley, 2017 #95}). We were unable to vary population size due to difficulties in effectively sampling a larger-dimensional hyperspace with the time necessary to run simulations, and with increased the computational requirements associated with increasing population sizes in individual-based models {Haller, 2019 #19}. The implementation of deleterious mutation rate also confounds with QTL mutation rate; although we were able to confirm deleterious mutation effects were constant across treatments (Fig. S1). Our implementation of a mega-trait also limited insight into mutational correlations among traits and the effect of pleiotropy. Since fitness effects were identical across traits, the effects of pleiotropy and mutational correlation would be averaged across traits, minimising the signal. These shortcomings however, lead to exciting expansions of our methodology in the future.

We have produced a framework to understand polygenic adaptation in the context of both quantitative and population genetics, as well as population genomics. Expanding this model to explain differences in the effects of drift in mutation-selection balance models will lend insight into the effects of heightened drift-barriers on restricting adaptation under the two models. In addition, varying the number of loci contributing to traits will give an indication of the robustness of variation under changing polygenicity. While here we have explored the maintenance of variation, a natural progression is to quantify how these models differ in their adaptive walks, giving evidence for a speed versus accuracy trade-off in polygenic adaptation. Similarly, integrating moving optima and heterogeneous environments into the model will test the predictions of where Gaussian and House-of-Cards mutation-selection balances are expected to be advantageous.

Overall, this study has shown that in an evolutionary context, Latin hypercube sampling is a robust tool for exploring complex parameter spaces, such as those underpinning polygenic adaptation. In addition, the ability to not only track the mutational effects underlying quantitative characters, provides great insight into the mechanics controlling population-level dynamics. The dynamics of House-of-Cards and Gaussian mutation-selection-drift balance models are clearly affected by mutational effect sizes differently, suggesting trade-offs between adaptability and adaptedness are common, and may answer why maladaptation appears so prevalent in natural populations.

# Snippets

Underpinning this model is the continuum of alleles model of allelic effects, suggesting large numbers of alleles at many loci forming a continuous distribution of effect sizes, usually Normal in shape (Lande).

Pleiotropy fundamentally alters the signatures of HCA vs Gaussian approximation in COA so they approach each other - other parameters may as well?

Most effort in understanding stabilizing selection has focused on assuming either a Gaussian (as in this paper) or quadratic fitness function

Pleiotropy also had strong effects, due to contributing more than one trait value per mutation. Increasing pleiotropy rate by 10% increased RAF by 59.366 ± 2.531 alleles under no selection (t63937 = 23.458, p < 0.0001). Increasing deleterious mutation with pleiotropy rate significantly reduced this effect, with a 10% increase in pleiotropy rate and deleterious mutation rate simultaneously leading to a total loss of 22.555 ± 5.994 alleles (t63937 = -17.795, p < 0.0001). Under stabilising selection, a simultaneous 10% increase in pleiotropy rate and deleterious mutation rate led to an increase of 23.553 ± 8.567 alleles (t63937 = 19.193, p < 0.0001).

Among the distinctions between Gaussian and House-of-Cards models are their assumptions regarding the relative importance of standing genetic variation and mutational variance. Under Gaussian models, standing genetic variation provides most of the genetic variance, owing to higher mutation rates relative to selection, which in turn leads to higher numbers of segregating alleles (Walsh and Lynch 2018). Indeed, our results support this, with House-of-Cards models maintaining lower variation than Gaussian models (Figure 5). It should be noted that it is difficult to ascertain if this trend remains at high additive effect sizes as both model types have considerably fewer adapted populations than under lower additive effect size variance, which fuels large standard errors of means (Table 2).

Both models impose restrictions on which mutations are viable, and the strength of those restrictions defines the distribution of allelic effects. We found that the distribution of allelic effects became significantly wider with increasing additive effect size variance, but only under a Gaussian model, indicating a sensitivity to additive effect variation that is not present under a House-of-Cards model (Figure 8). We believe the mechanism underpinning this sensitivity to additive effect size change lies in the underlying assumptions of the models.

It is well understood in population genetics that background selection reduces effective population size, reducing the effectiveness of selection and increasing the strength of genetic drift (Charlesworth *et al.* 1997; Houle 1998). As deleterious mutations are removed from the population, close-by linked QTLs are also removed (Charlesworth and Charlesworth 2010). The effect of this is decreased genetic diversity. In population genetics studies this is usually expressed in terms of FST or , whereas in quantitative genetics the analog is additive genetic variance (Falconer 1996; Charlesworth *et al.* 1997). Reductions in VA with increasing background selection were observed in this study, supporting this expectation (Figure 3A, 4). The expected effect of this on adaptation is quite clear when considering the initial approach towards the optimum: in quantitative genetics models, genetic variability is expected to increase the trait space that populations are able to explore, improving their ability to travel towards an optimum (Fisher 1930; Charlesworth and Charlesworth 2010; Aguirre *et al.* 2014). Indeed, these theoretical expectations have been found in natural populations: for example, Pujol and Pannell (2008) showed that populations of annual mercury, *Mercualis annua,* were able to respond to selection for pollen production when standing genetic variation was higher. Similarly, studies into the adaptation of red flour beetle (*Tribolium castaneum*) populations to new niches found high standing variation decreased the likelihood of extinction, and increased rates of niche expansion (Agashe and Bolnick 2010; Agashe *et al.* 2011). However, these expectations do not describe what we found in the current study: the most well-adapted populations consistently have higher rates of deleterious mutation, and hence lower standing genetic variance. The key to this lies in the expectations of the *maintenance* of variation and fitness around an optimum rather than the *approach* towards said optimum. The expectations surrounding this temporal space is considerably less extensive than that of the adaptive walk.

While reduced standing variation is expected to increase the time a population takes to reach an optimum (or perhaps prevent populations from reaching it at all), once a population has reached its optimum or stabilizes around its ‘local optimum’, the closest position it can maintain given the selected traits’ genetic architectures, mutation rates, and the population size – where does the population go?

However, even among the adapted subset of populations, very few were directly at the optimum: there was always some level of maladaptation,

Genetic architecture controls ability to adapt and stay there; Gaussian vs HoC are similar in ability to get there, very different from null (16 vs 0.5% chance); among models that are maladapted, there is plenty of variance, to the point where much of the null model distribution overlaps with that of the CoA models; Among the few null models that did get into that adapted range, they never got to the perfect adaptation level, where distance = 0; suggests there is a drift barrier to clear to be able to overcome mutation/drift and reach the optimum; example of barrier (Gardon et al. 2020).

The position of this barrier differs depending on the expectations of where the m/s/d equilibrium is; this is likely different between models, as shown in fig 4A where the distribution of distances is much smaller in HoC vs Gaussian, due to fewer mutations that may cause populations to drift further from their optimum; Among those adapted populations, what was the predictor underpinning that ability? How did these populations adapt? What were then signatures associated with that? HoC and Gaussian were pretty similar in how close they got – but their responses to changes in additive effect size were different; HoC remained unaffected, Gaussian moved further away; insensitivity caused by the relative strengths of selection and mutation swamping of high mutation rates; with high effect sizes and weak selection, Gaussian models fluctuate more around optimum, as selection cannot effectively reduce this standing variation; HoC on the other hand is efficient in removing new mutations that are mostly deleterious, results in lower standing variance and increased reliance on big new mutations to drive further adaptation; trade-off in the case of a changing optimum – Gaussian may be able to get to the general range of an optimum faster, but will not be able to truly get there; HoC needs to wait for the appropriate mutations to arise to drive adaptation, will take longer but the stronger selection means they are more likely to pinpoint their location very close to the optimum; different strategies that are likely to be beneficial in different environments; Gaussian in heterogeneous environments; HoC in homogeneous, stable environments; the models act like ‘hot’ or ‘cold’, with Gaussian being a hot excited molecule dashing around the optimum imprecisely, while HoC are more cold, and move less far from the optimum over the same time; over time, modifiers of mutation rate may be beneficial if you need to react to a new event, then those will slowly go away; example;

Demonstrated that genetic architecture contributes greatly to the ability of populations to adapt, and that the effect of genetic architecture on adaptability depends on the relative strength of mutation to selection, via the HoC vs Gaussian model. Particularly the variation in additive effect size, the ‘precision’ of mutation to drive populations to the optimum, had large influences (Fig. 5, 6, 7, 8). However, HoC models appeared more robust to changes in effect size in general, with variance, distance, and covariance not being perturbed as much as under Gaussian models. Why? Comes down to assumptions of these models

Gaussian assumes mutation > selection, most additive variation comes from standing variation. HC assumes selection > mutation, so mutational variance is greater than standing variance. Because selection strong, the allele with the highest frequency is expected to have a value close to the optimum. So new mutations are deleterious and tend to disappear quickly, resulting in most of the gen var being due to rare alleles with large effects

standing genetic variation (in the Gaussian model, where selection is much weaker than mutation; Figure 2), or new mutational variance (in the House-of-Cards model, where selection is much stronger than mutation).

Very little variation is expected to be maintained by adapted populations, may change with

Why is effect size so important? Introduces mutational variance, increases additive variance which selection can act on to go more efficiently, or in this case, can reduce ability of populations to maintain positions around optima because effects are all over the place – swamping effect. This is why Gaussian vs HoC so important, the relative rates of mutation to selection define the positions of models.

Driver of effect size, importance of mutational variance vs standing variation

Weird outliers in variance and covariance: could be that these populations did have high variance and covariance due to wildly different individual phenotypes, but when you took the mean distance of the population, that mean was somewhere in the middle of all that variation, which happened to be close to the optimum.

Robustness of HoC vs Gaussian to changes in Effect size

Truly adapted ones – at distance = 0

Really rare to be close under null

Estes and Arnold 2007 – 64% populations at least 1 sd from the optimum

Accuracy vs speed of adaptation – we are looking at accuracy

Fig 4A: when you’re far from the optimum there is a lot of stochasticity – Gaussian and HoC overlap with Null distributions indicating they aren’t able to escape the drift barrier (Nes < 1)

Fig 4B: HoC/Gaus pops are very different from null

At the optimum (Po = 1), the Gaus and HoC are not like Null

Relative importance of mutational variance vs standing variation drives differences in responses to selection

This is the power of pop gen models – get to see these allele frequencies, the whole spectrum

In turn, these changes in variance-covariance structure could affect the ability of populations to adapt to future environmental changes (Arnold *et al.* 2008; Otto 2009).

Loss of fitness due to variation around optimum: expected to be 1/4Ne without any background selection (will vary with Ne due to effect on local Ne ) – Lande 1976

# Tables

**Table 1:** Model parameters for both null and stabilizing selection models. The range of values is based on literature, but values are adjusted to be practical for the time of the experiment.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameter** | **Symbol** | **Range** | **Description** | **Source(s)** |
| Genome wide recombination rate | r | 0 to 1.241x10-4 per locus | The singular recombination rate used across the entire simulated genome. | Stapley et al. 2017 |
| Background selection rate | δ | 0 to 1 | The number of non-trait deleterious mutations that occur relative to trait mutations. |  |
| Rate of universal pleiotropy | ϖ | 0 to 0.5 | The proportion of trait mutations that affect all traits rather than a single trait. While 100 loci control a trait independently by default, this may be changed by this parameter. However, ratios of loci affecting each trait will remain constant, especially across multiple replicates. | Chesmore et al. 2017; |
| Mutational pleiotropic correlation | m | 0 to 0.5 | The mutational correlation between additive effects of pleiotropic mutations determines the similarity of trait effects between traits for the same pleiotropic mutation. |  |
| Additive effect size | λ | 0.1 to 10 | Additive effect size controls the variance of trait effect size around mean 0, so that N(0, λ). | Albert et al. 2008; |
| Selection strength (selection model only) |  | 10 to 10000 | The parameter that controls the curve of the fitness function (eq. 3), with higher values resulting in a smaller difference in fitness between trait-differing individuals. |  |

**Table 2:** Means of distance from the optimum (, variance (VA), and covariance among traits for levels of additive effect size, recombination rate, pleiotropy rate, and mutational correlations for Gaussian and House-of-Cards models. Values in bold are mentioned in the main text and featured in Figures 5, 6, or 7. Values in italics indicate means that include outliers that were excluded from figures 6 and 7 for better readability. \* denotes values of interest.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | δ | | | VA | | | Covariance | | |
| Null | House-of-Cards | Gaussian | Null | House-of-Cards | Gaussian | Null | House-of-Cards | Gaussian |
| Additive effect size (α) | |  |  |  |  |  |  |  |  |  |  |
|  | Low |  | **11.151** | **0.715** | **1.384** | **1.035** | **0.241** | **1.747** | **0.069** | **0.002** | **0.015** |
| S.E. | 0.133 | 0.033 | 0.065 | 0.067 | 0.013 | 0.025 | 0.008 | 0.0002 | 0.002 |
| n | 545 | 458 | 445 | 545 | 458 | 445 | 545 | 458 | 445 |
| Medium |  | -\* | **3.109** | **2.085** | - | **9.733** | **3.120** | - | **-0.699** | **0.034** |
| S.E. | - | 1.110 | 0.213 | - | 8.590 | 0.163 | - | 0.709 | 0.010 |
| n | 0 | 6 | 26 | 0 | 6 | 26 | 0 | 6 | 26 |
| High |  | -\* | **1.202** | **2.699** | - | ***110.240*** | **2.943** | - | ***-9.932*** | **-0.022** |
| S.E. | - | 0.258 | - | - | 57.536 | - | - | 5.077 | - |
| n | 0 | 3 | 1 | 0 | 3 | 1 | 0 | 3 | 1 |
| Recombination rate | |  |  |  |  |  |  |  |  |  |  |
|  | Low |  | 10.349 | 0.664 | 1.543 | 0.822 | 4.151 | 1.595 | 0.066 | -0.452 | 0.006 |
| S.E. | 0.197 | 0.066 | 0.101 | 0.063 | 3.080 | 0.026 | 0.013 | 0.371 | 0.003 |
| n | 267 | 46 | 234 | 267 | 46 | 234 | 267 | 46 | 234 |
| Medium |  | 11.452 | 1.372 | 1.467 | 0.898 | 5.941 | 3.412 | 0.068 | -0.355 | 0.043 |
| S.E. | 0.191 | 0.289 | 0.219 | 0.079 | 5.297 | 0.408 | 0.009 | 0.357 | 0.011 |
| n | 219 | 37 | 2 | 219 | 37 | 2 | 219 | 37 | 2 |
| High |  | 13.658 | 0.699 | 1.310 | 2.507\* | 0.232 | 2.048 | 0.089 | 0.002 | 0.024 |
| S.E. | 0.241 | 0.033 | 0.076 | 0.418 | 0.012 | 0.048 | 0.022 | 0.0002 | 0.002 |
| n | 59 | 37 | 236 | 59 | 384 | 236 | 59 | 384 | 236 |
|  | | | δ | | | VA | | | Covariance | | |
| Null | House-of-Cards | Gaussian | Null | House-of-Cards | Gaussian | Null | House-of-Cards | Gaussian |
| Pleiotropy rate | |  |  |  |  |  |  |  |  |  |  |
|  | Low |  | 10.965 | 1.073 | 1.736 | 0.846 | 2.011 | 1.912 | 0.056 | -0.194 | 0.003 |
| S.E. | 0.137 | 0.127 | 0.100 | 0.037 | 1.559 | 0.046 | 0.007 | 0.196 | 0.002 |
| n | 512 | 85 | 270 | 512 | 85 | 270 | 512 | 85 | 270 |
| Medium |  | 14.139 | 1.024 | 0.966 | 4.472 | 3.458 | 1.786 | 0.465 | -0.207 | 0.042 |
| S.E. | 0.288 | 0.114 | 0.077 | 0.623 | 2.438 | 0.040 | 0.098 | 0.167 | 0.003 |
| n | 15 | 83 | 109 | 15 | 83 | 109 | 15 | 83 | 109 |
| High |  | 13.952 | 0.581 | 1.063 | 3.539 | 0.140 | 1.639 | 0.113 | 0.002 | 0.019 |
| S.E. | 0.308 | 0.028 | 0.067 | 1.362 | 0.006 | 0.047 | 0.060 | 0.0002 | 0.003 |
| n | 18 | 299 | 93 | 18 | 299 | 93 | 18 | 299 | 93 |
| Mutational correlation | |  |  |  |  |  |  |  |  |  |  |
|  | Low |  | 13.722 | 0.629 | 1.499 | 1.559 | 0.414 | 1.981 | 0.005 | -0.016 | 0.0003 |
| S.E. | 0.161 | 0.039 | 0.076 | 0.098 | 0.208 | 0.051 | 0.009 | 0.017 | 0.002 |
| n | 117 | 253 | 238 | 117 | 253 | 238 | 117 | 253 | 238 |
| Medium |  | 13.885 | 2.521 | 1.118 | 3.378 | 22.616 | 1.681 | 0.196 | -1.461 | 0.016 |
| S.E. | 0.240 | 0.778 | 0.100 | 0.509 | 21.752 | 0.054 | 0.036 | 1.468 | 0.002 |
| n | 51 | 9 | 83 | 51 | 9 | 83 | 51 | 9 | 83 |
| High |  | 9.983\* | 0.820 | 1.479 | 0.555 | 0.934 | 1.672 | 0.072 | -0.077 | 0.039 |
| S.E. | 0.148 | 0.054 | 0.145 | 0.031 | 0.647 | 0.030 | 0.009 | 0.081 | 0.004 |
| n | 377 | 205 | 151 | 377 | 205 | 151 | 377 | 205 | 151 |

# Figure legends

**Figure 3:** Mean additive variance (VA; panel A) and mean between-trait covariance (B) over 100,000 generations of stabilizing selection of different strengths (). 256 total models were sampled across the spectrum of selection strengths () with an additional 1024 models sampling the null space of parameters ().

**Figure 4:** Euclidean distances from the optimum over models. (A): total distributions of all models. (B): distributions of adapted models with small distance to the optimum.

**Figure 5:** Euclidean distances from the optimum () among adapted populations with increasing additive effect size. Note that there was only one adapted Gaussian population with high additive effect size, and three House-of-Cards with high effect size. Bars indicate S.E.M.

**Figure 6:** Mean additive variance (VA) among adapted populations with increasing additive effect size. Note that there was only one adapted Gaussian population with high additive effect size, and three House-of-Cards with high effect size. Several outliers were removed for improved readability. Bars indicate S.E.M.

**Figure 7:** Mean trait covariance among adapted populations with increasing additive effect size. Note that there was only one adapted Gaussian population with high additive effect size, and three House-of-Cards with high effect size. Several outliers were removed for improved readability. Bars indicate S.E.M.

**Figure 8:** Density estimates of mutational effect sizes for adapted populations at generation 100,000 under House-of-Cards and Gaussian models, with differing additive effect size distributions.

**Figure 9**: Population adherence to a two-trait phenotypic optimum over time. Xs indicate population positions in phenotype space, with the size of the X corresponding to the magnitude of mutational variance in the population. Blue Xs represent populations under House-of-Cards models of allelic effects, where mutation rates are low relative to selection strength. Red Xs represent populations under Gaussian models, where mutation rates are high relative to selection.

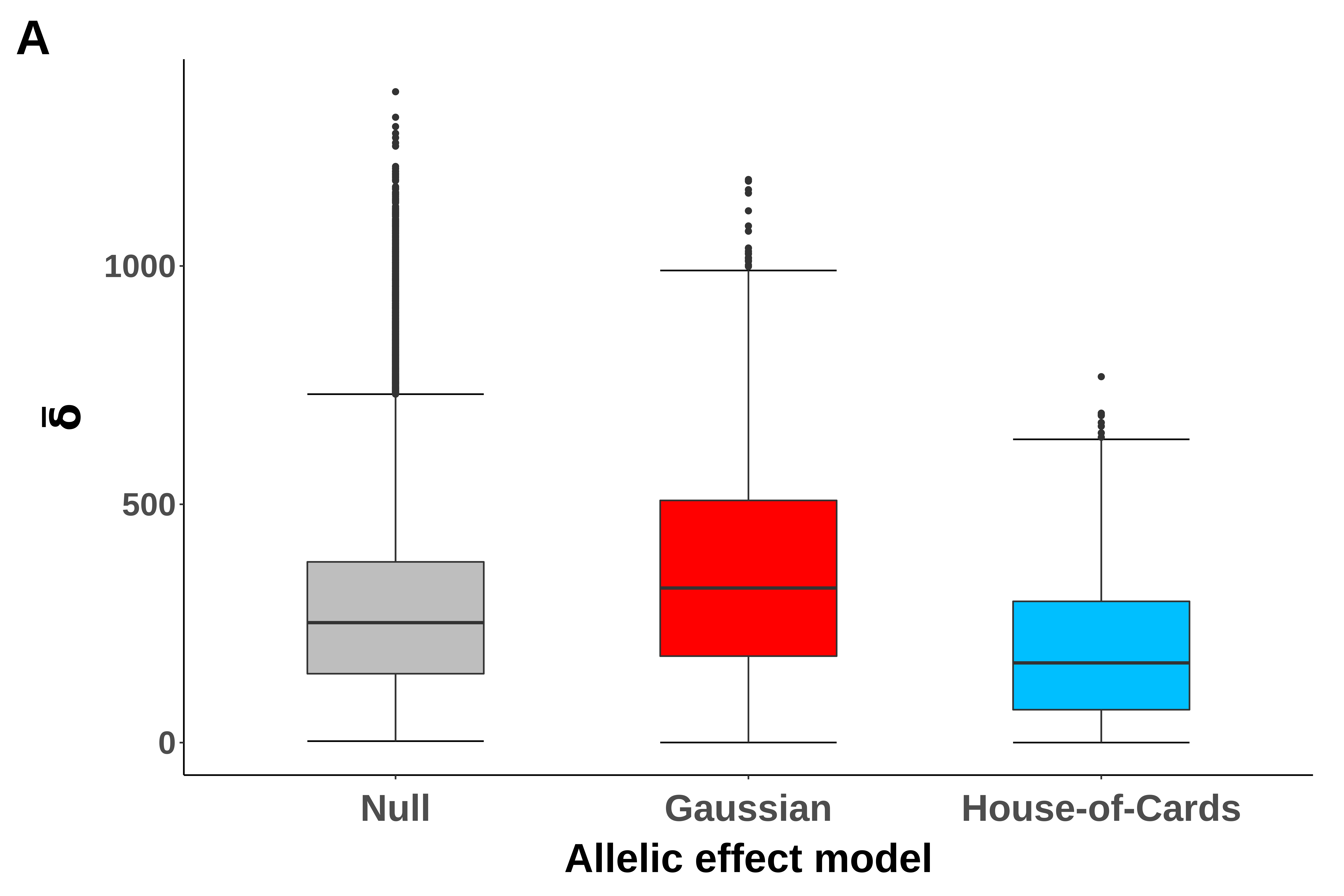
**Figure S1:** Ratio of deleterious mutations to QTL mutations with increasing deleterious mutation rate. Note that odds of deleterious mutation to QTL go from 100% QTL at x = 0 to 50% QTL 50% deleterious at x = 1.

**Figure S2**: Preliminary analysis of mean heterozygosity over time with changing population size. Solid lines represent mean trajectories of 20 replicates, with ribbons representing standard errors. Dotted lines represent expected heterozygosities ± 5%, given by .

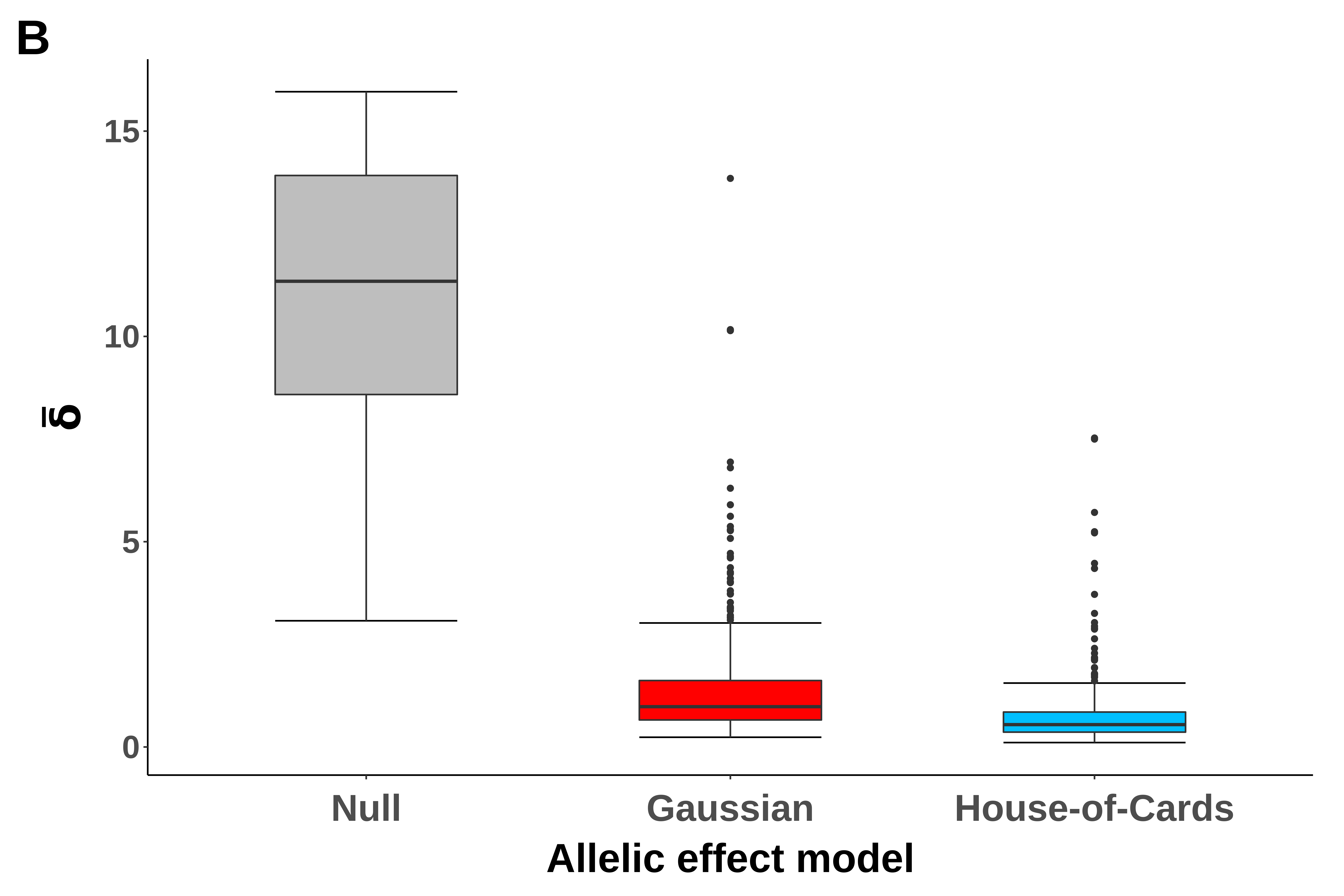
**Figure S3:** Latin hypercube sampling of null models (A) and selection models (B). Diagonals represent distributions of samples, which are uniform across the parameter range. Points in bottom off-diagonal indicate a single sample in the parameter space. Each sample was replicated 100 times with unique seed values. Correlations in upper off-diagonal indicate maximum correlations between samples.

# Figure 3

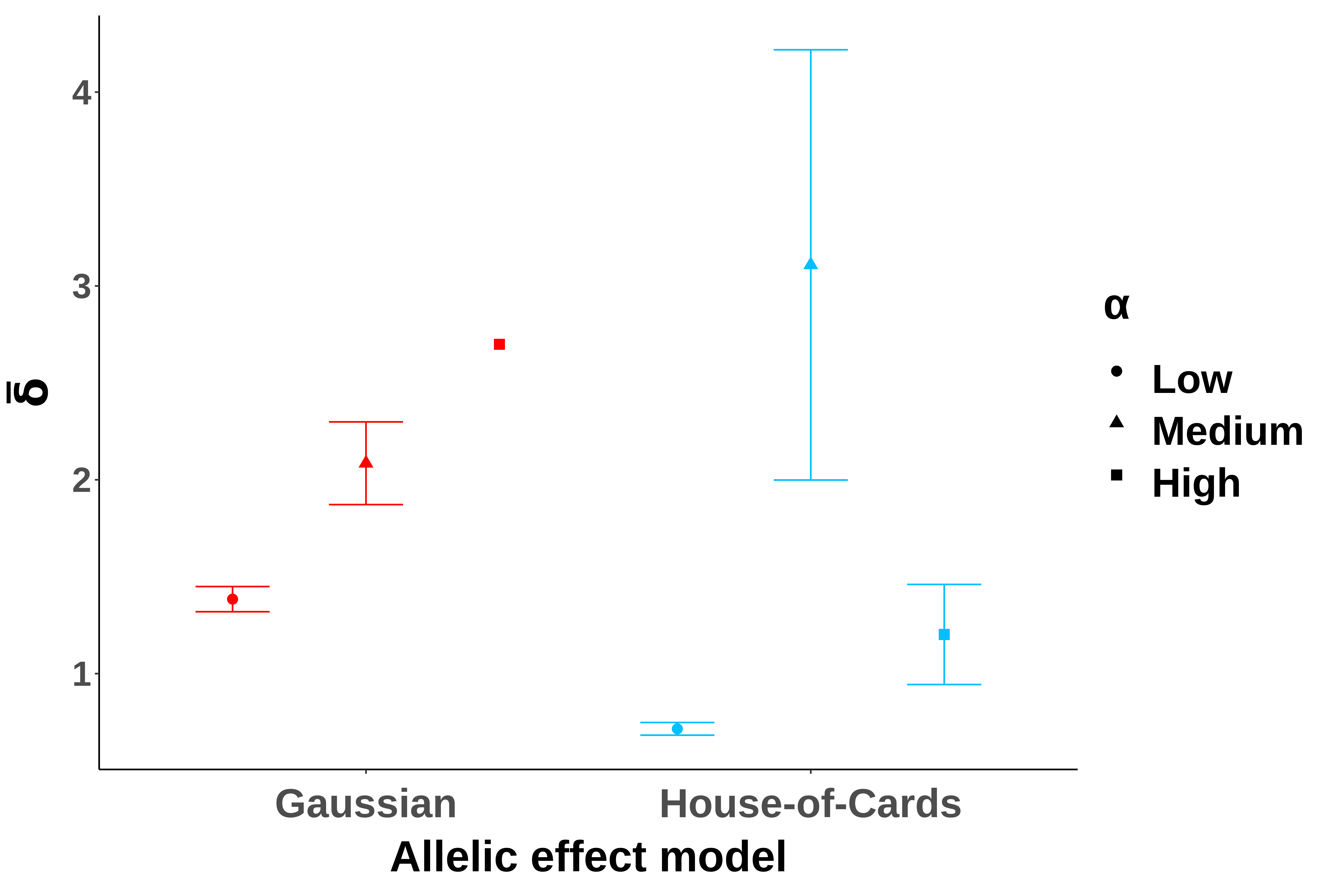
# Figure 4A



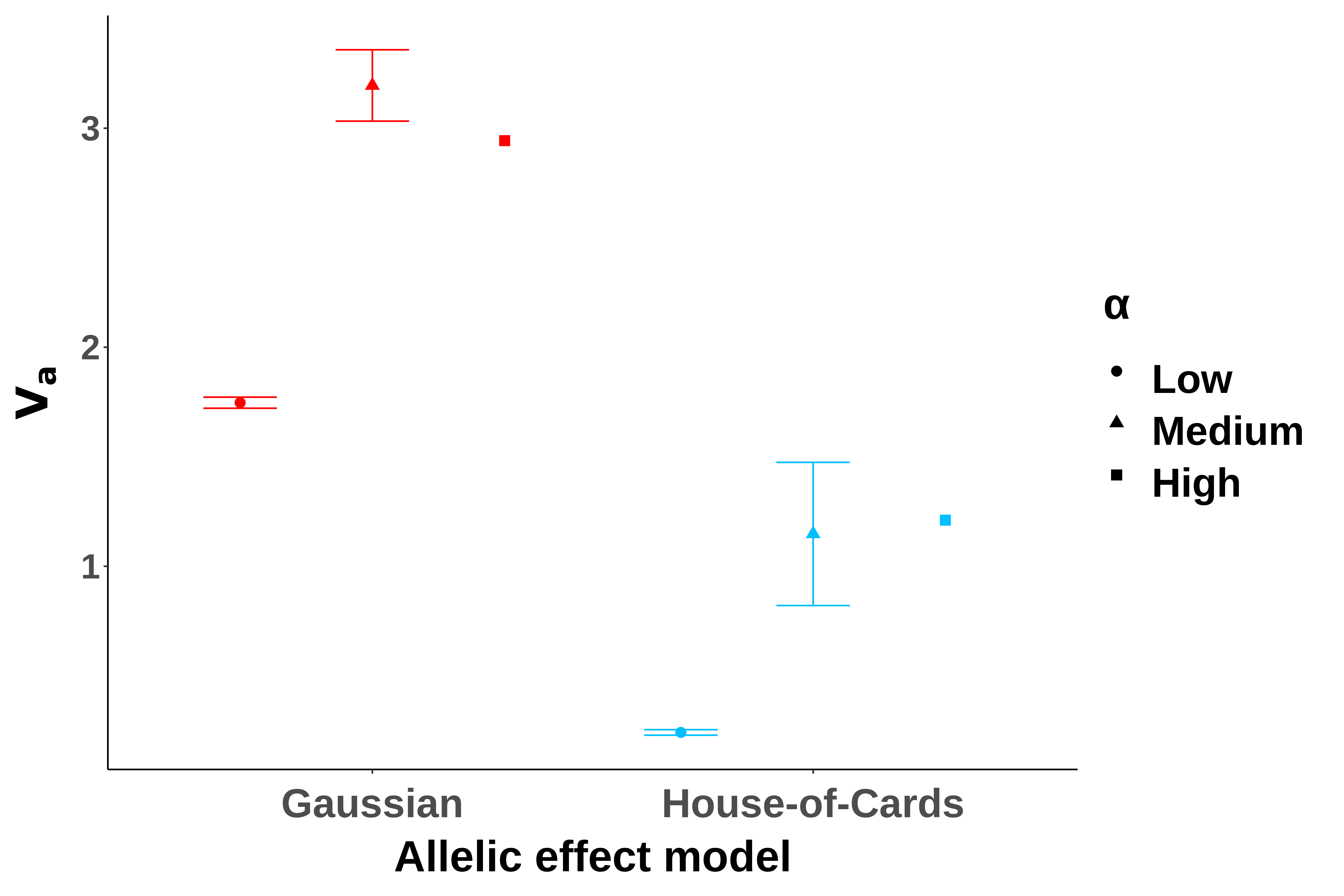
# Figure 4B



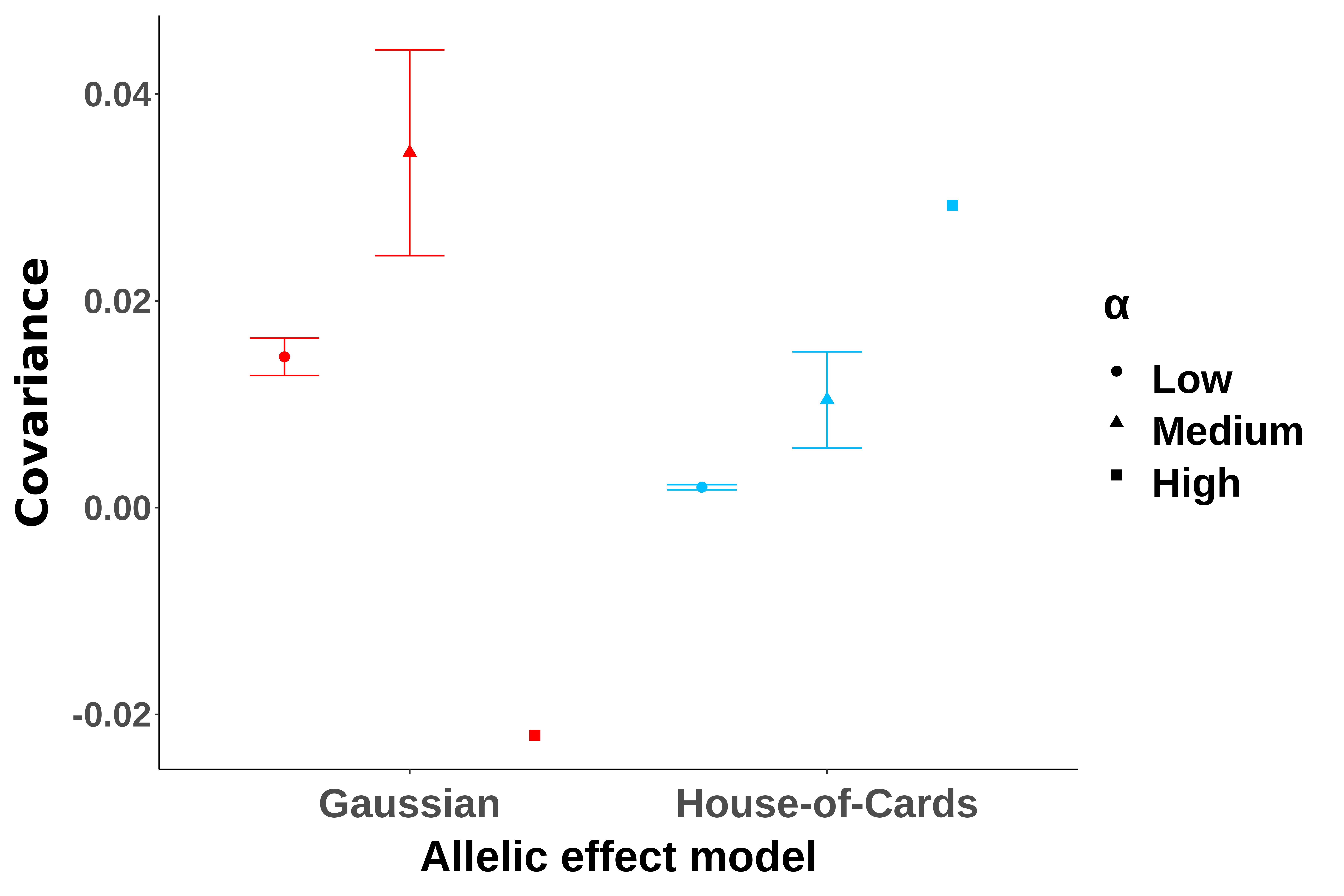
# Figure 5



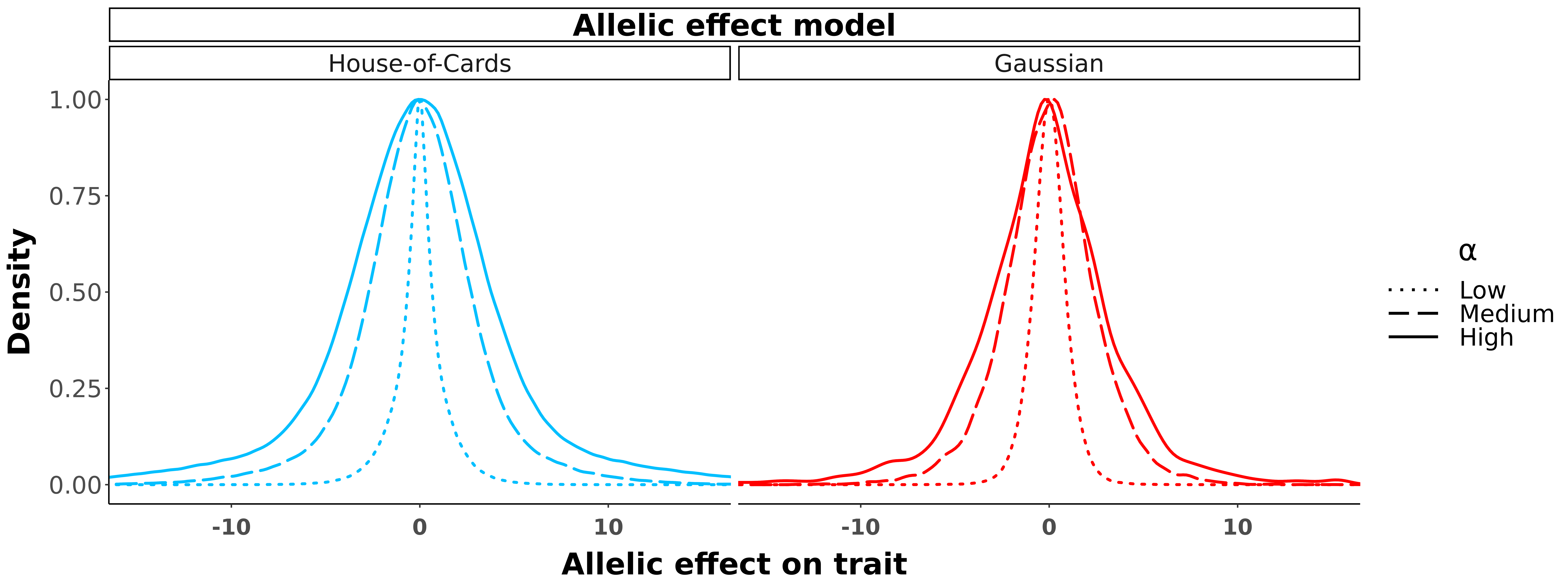
# Figure 6



# Figure 7



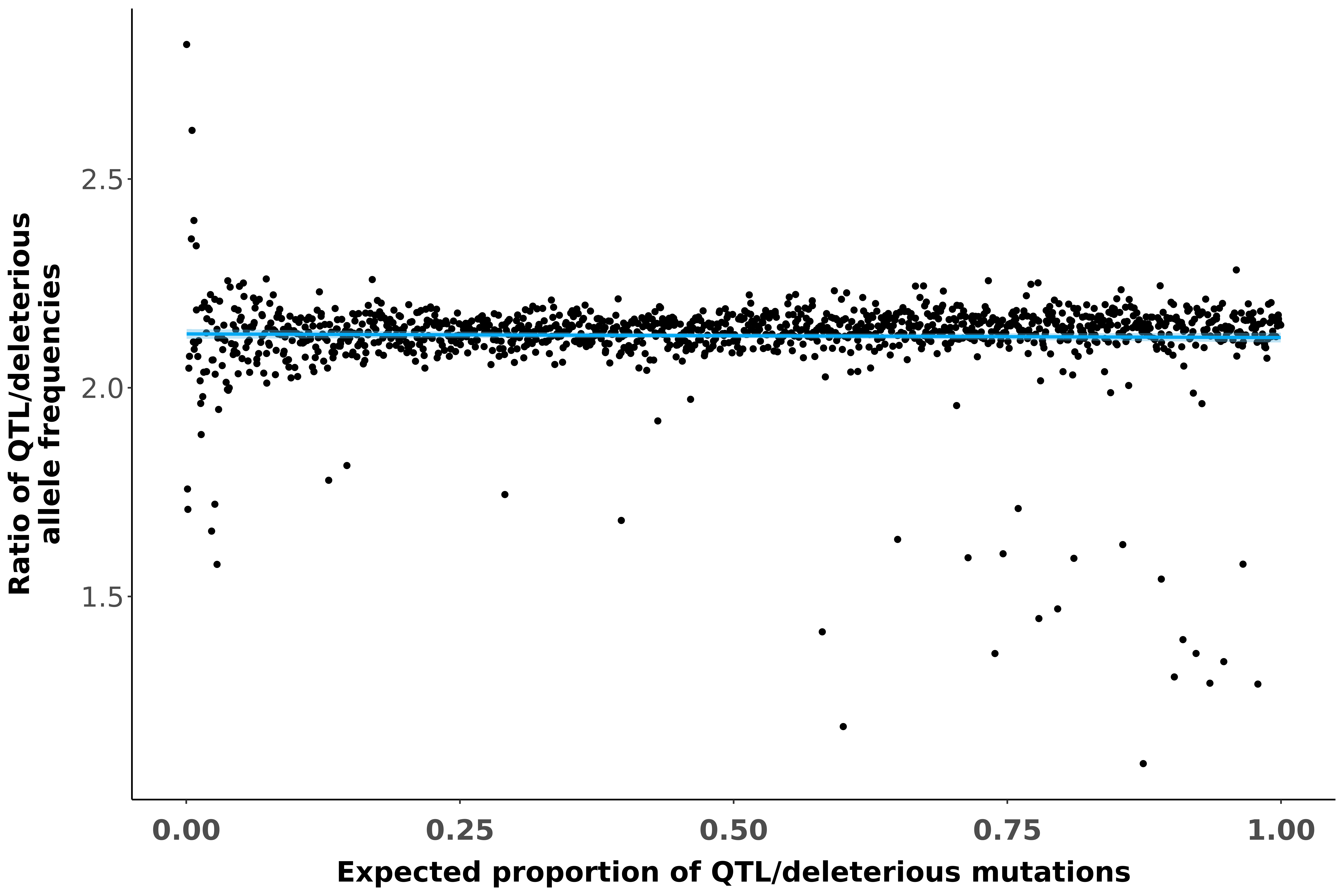
# Figure 8



# Figure 9

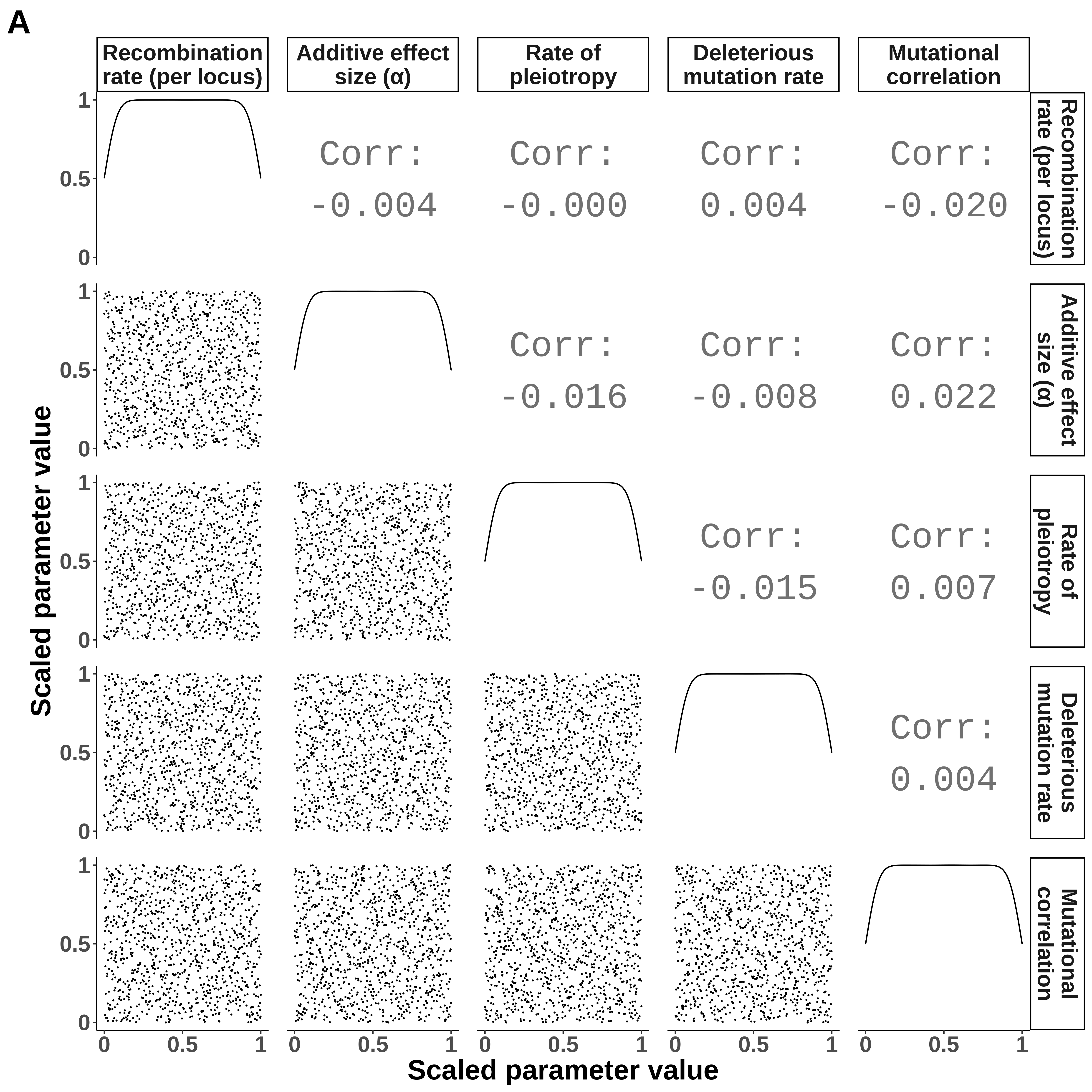
# Supplementary material

# Figure S1

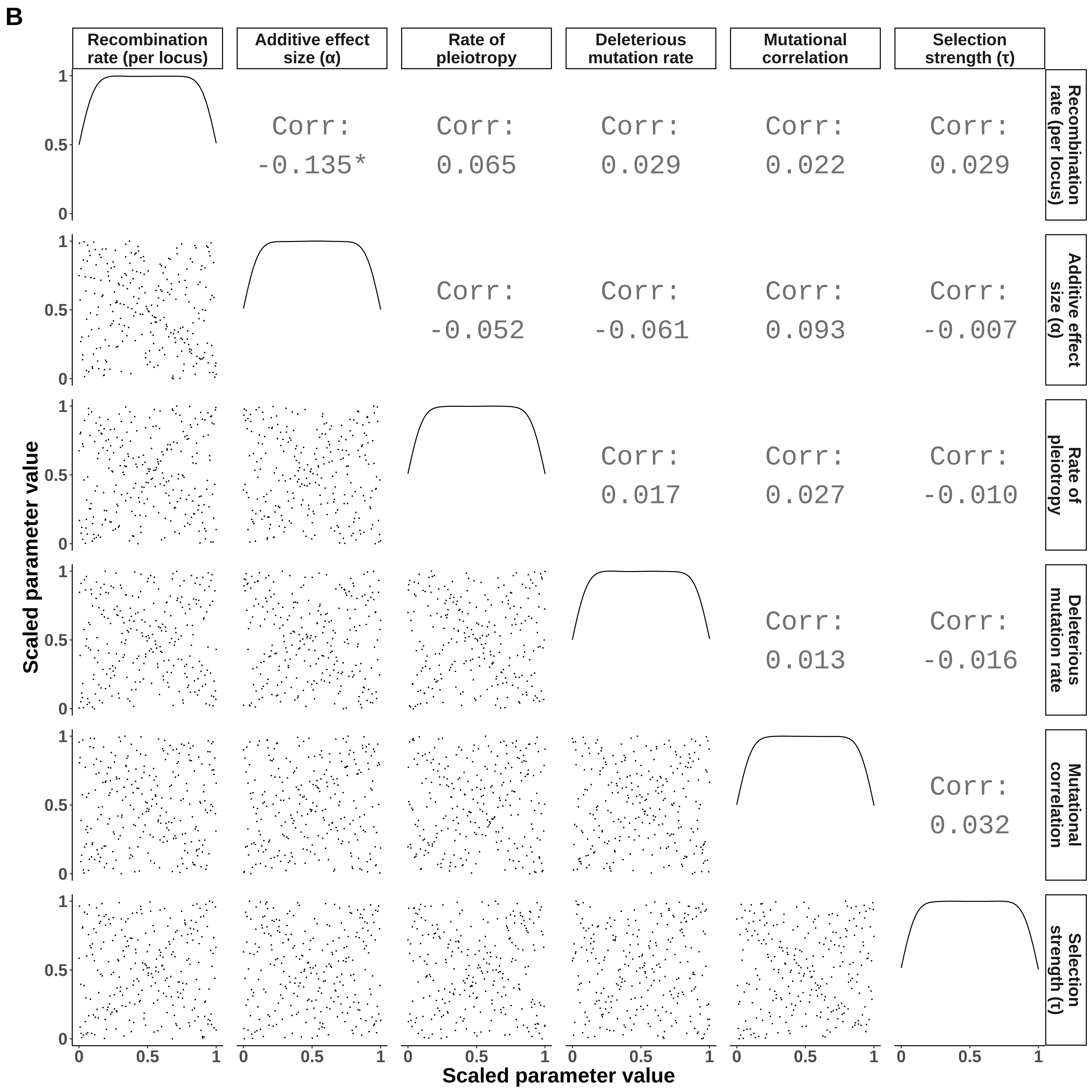


# Figure S2

# Figure S3A



# Figure S3B



# References

Agashe, D., and D. I. Bolnick, 2010 Intraspecific genetic variation and competition interact to influence niche expansion. Proc Biol Sci 277**:** 2915-2924.

Agashe, D., J. J. Falk and D. I. Bolnick, 2011 Effects of founding genetic variation on adaptation to a novel resource. Evolution 65**:** 2481-2491.

Aguirre, J. D., E. Hine, K. McGuigan and M. W. Blows, 2014 Comparing G: multivariate analysis of genetic variation in multiple populations. Heredity 112**:** 21-29.

Arnold, S. J., R. Burger, P. A. Hohenlohe, B. C. Ajie and A. G. Jones, 2008 Understanding the Evolution and Stability of the G-Matrix. Evolution 62**:** 2451-2461.

Aston, E., A. Channon, R. V. Belavkin, D. R. Gifford, R. Krasovec *et al.*, 2017 Critical Mutation Rate has an Exponential Dependence on Population Size for Eukaryotic-length Genomes with Crossover. Sci Rep 7**:** 15519.

Barton, N. H., 2017 How does epistasis influence the response to selection? Heredity (Edinb) 118**:** 96-109.

Barton, N. H., and B. Charlesworth, 1998 Why sex and recombination? Science 281**:** 1986-1990.

Blair, G. C., J. Coppock, A. Humphreys, M. Sonnet, L., 2020 estimatr: Fast Estimators for Design-Based Inference, pp.

Brady, S. P., D. I. Bolnick, A. L. Angert, A. Gonzalez, R. D. H. Barrett *et al.*, 2019 Causes of maladaptation. Evol Appl 12**:** 1229-1242.

Careau, V., M. E. Wolak, P. A. Carter and T. Garland, Jr., 2015 Evolution of the additive genetic variance-covariance matrix under continuous directional selection on a complex behavioural phenotype. Proc Biol Sci 282.

Charlesworth, B., and D. Charlesworth, 2010 *Elements of Evolutionary Genetics*. Roberts and Company, Greenwoord Village, Colorado, USA.

Charlesworth, B., M. Nordborg and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet Res 70**:** 155-174.

Crespi, B. J., 2000 The evolution of maladaptation. Heredity (Edinb) 84 ( Pt 6)**:** 623-629.

Eicker, F., 1967 Limit theorems for regressions with unequal and dependent errors, pp. 59-82 in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. University of California Press, Berkeley, Calif.

Falconer, D. S. M., T. F. C., 1996 *Introduction to Quantitative Genetics*. Pearson Education Limited, Longmans Green, Harlow, Essex, UK.

Fisher, R. A., 1930 *The genetical theory of natural selection*. The Clarendon press, Oxford, UK.

Fleming, W. H., 1979 Equilibrium Distributions of Continuous Polygenic Traits. Siam Journal on Applied Mathematics 36**:** 148-168.

Franssen, S. U., R. Kofler and C. Schlotterer, 2017 Uncovering the genetic signature of quantitative trait evolution with replicated time series data. Heredity (Edinb) 118**:** 42-51.

Gardon, H., C. Biderre-Petit, I. Jouan-Dufournel and G. Bronner, 2020 A drift-barrier model drives the genomic landscape of a structured bacterial population. Mol Ecol.

Gilbert, K. J., and M. C. Whitlock, 2017 The genetics of adaptation to discrete heterogeneous environments: frequent mutation or large-effect alleles can allow range expansion. J Evol Biol 30**:** 591-602.

Haller, B. C., and P. W. Messer, 2019 SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. Molecular Biology and Evolution 36**:** 632-637.

Hayes, A. F., and L. Cai, 2007 Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. Behavior Research Methods 39**:** 709-722.

Helton, J. C., and F. J. Davis, 2003 Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering & System Safety 81**:** 23-69.

Hodgins-Davis, A., D. P. Rice and J. P. Townsend, 2015 Gene Expression Evolves under a House-of-Cards Model of Stabilizing Selection. Mol Biol Evol 32**:** 2130-2140.

Houle, D., 1998 How should we explain variation in the genetic variance of traits? Genetica 102-103**:** 241-253.

Huber, P. J., 1967 The behavior of maximum likelihood estimates under nonstandard conditions, pp. 221-233 in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. University of California Press, Berkeley, Calif.

Kimura, M., 1965 A stochastic model concerning the maintenance of genetic variability in quantitative characters. Proc Natl Acad Sci U S A 54**:** 731-736.

Kimura, M., and J. F. Crow, 1964 The Number of Alleles That Can Be Maintained in a Finite Population. Genetics 49**:** 725-738.

Lande, R., 1975 The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genet Res 26**:** 221-235.

Lindeman, R. H. M., P.F. Gold, R.Z., 1980 *Introduction to Bivariate and Multivariate Analysis*. Scott, Foresman, Glenview, IL.

Long, J. A., 2020 jtools: Analysis and Presentation of Social Scientific Data

Lumley, T., P. Diehr, S. Emerson and L. Chen, 2002 The importance of the normality assumption in large public health data sets. Annu Rev Public Health 23**:** 151-169.

Lynch, M., 2010 Evolution of the mutation rate. Trends Genet 26**:** 345-352.

Lynch, M., M. S. Ackerman, J. F. Gout, H. Long, W. Sung *et al.*, 2016 Genetic drift, selection and the evolution of the mutation rate. Nat Rev Genet 17**:** 704-714.

Lynch, M., and R. Lande, 1998 The critical effective size for a genetically secure population. Animal Conservation 1**:** 70-72.

Malcom, J. W., 2011 Evolution of Competitive Ability: An Adaptation Speed vs. Accuracy Tradeoff Rooted in Gene Network Size. Plos One 6.

Melo, D., G. Garcia, A. Hubbe, A. P. Assis and G. Marroig, 2015 EvolQG - An R package for evolutionary quantitative genetics. F1000Research 4**:** 925.

Nesse, R. M., 2005 Maladaptation and natural selection. Q Rev Biol 80**:** 62-70.

Ohta, T., 1973 Slightly Deleterious Mutant Substitutions in Evolution. Nature 246**:** 96-98.

Orr, H. A., 1998 The Population Genetics of Adaptation: The Distribution of Factors Fixed during Adaptive Evolution. Evolution 52**:** 935-949.

Otto, S. P., 2009 The Evolutionary Enigma of Sex. American Naturalist 174**:** S1-S14.

Pujol, B., and J. R. Pannell, 2008 Reduced responses to selection after species range expansion. Science 321**:** 96.

R Developmental Core Team, 2019 R: A language and environment for statistical computing, pp. R Foundation for Statistical Computing, Vienna, Austria.

Sztepanacz, J. L., and M. W. Blows, 2017 Artificial Selection to Increase the Phenotypic Variance in gmax Fails. Am Nat 190**:** 707-723.

Thornton, K. R., 2019 Polygenic Adaptation to an Environmental Shift: Temporal Dynamics of Variation Under Gaussian Stabilizing Selection and Additive Effects on a Single Trait. Genetics 213**:** 1513-1530.

Turelli, M., 1984 Heritable Genetic-Variation Via Mutation Selection Balance - Lerch Zeta Meets the Abdominal Bristle. Theoretical Population Biology 25**:** 138-193.

Walsh, B., and M. Lynch, 2018 *Evolution and selection of quantitative traits*. Oxford University Press, New York, NY.

White, H., 1980 A Heteroskedasticity-Consistent Covariance-Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica 48**:** 817-838.

Williams, G. C., 1966 *Adaptation and natural selection; a critique of some current evolutionary thought*. Princeton University Press, Princeton, N.J.,.

Xu, L., H. Chen, X. Hu, R. Zhang, Z. Zhang *et al.*, 2006 Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. Mol Biol Evol 23**:** 1107-1108.

Zhang, X. S., 2012 Fisher's geometrical model of fitness landscape and variance in fitness within a changing environment. Evolution 66**:** 2350-2368.