# Polygenic selection to a changing optimum under self–fertilisation

Matthew Hartfield[1,*] and Sylvain Glémin[2]

[1] Institute of Ecology and Evolution, The University of Edinburgh, Edinburgh EH9 3FL, United Kingdom.

[2] Université de Rennes, Centre National de la Recherche Scientifique (CNRS), ECOBIO (Ecosystèmes, Biodiversité, Evolution) – Unité Mixte de Recherche (UMR) 6553, F–35000 Rennes, France.

∗ *Corresponding author:* m.hartfield@ed.ac.uk

*Running Head:* Polygenic selection under self–fertilisation.

*Key words:* Adaptation, self–fertilisation, polygenic selection, changing environment, mating systems.

*Website for simulation code*:

`https://github.com/MattHartfield/Polygenic-Selfing`

**Abstract**

Many traits are polygenic, affected by multiple genetic variants throughout the genome. Selection acting on these traits involves co–ordinated allele–frequency changes at these underlying variants, and this process has been extensively studied in random–mating populations. Yet many species self–fertilise to some degree, which incurs changes to genetic diversity, recombination and genome segregation. These factors cumulatively influence how polygenic selection is realised in nature. Here, we use simulations to investigate to what extent self–fertilisation affects polygenic adaptation to a new environment. We find that self–fertilisation can incur a fitness advantage through purging maladaptive genotypes. While the degree of self–fertilisation has little influence on how quickly a population adapts to a new optimum, those with the highest long–term fitness tended to have high levels of selfing (at least 90%). Very high levels of self–fertilisation (99% and above) causes fixation of variants with compensatory effects in elevated linkage disequilibrium across the genome. Conversely, if mutations are pleiotropic then only a few major–effect variants fix along with many neutral hitchhikers, with a transient increase in linkage disequilibrium. If deleterious mutations are also present then linkage–block formation is disrupted in highly–selfing species, with fewer mutations contributing to adaptation and less chance of high–linkage genomes persisting over time. We also show how these outcomes differ in outcrossing populations with rescaled mutation and recombination rates, demonstrating that elevated homozygosity also influences the selection response. These results show potential advantages to self–fertilisation when adapting to a new environment, and how the mating system affects the genetic composition of polygenic selection.

# Introduction

Detecting where selection acts in the genome, along with the mechanisms at play, is a major research goal in evolutionary genetics. Many evolutionarily–important traits are complex and polygenic, meaning they are affected by multiple variants present genome–wide (Falconer and Mackay 1996). Selection acting on these traits will incur allele–frequency changes acting on these variants simultaneously. Recent years have seen major interest in polygenic selection research, aided by the appearance of large genome datasets that can identify many variants with small effect sizes, and new methods developed specifically for detecting it (reviewed by Pritchard and Di Rienzo (2010); Pritchard *et al.* (2010); Sella and Barton (2019); Barghi *et al.* (2020); Fagny and Austerlitz (2021)).

In particular, there are an increasing number of studies being performed on self–fertilising organisms. These are species where hermaphrodite individuals produce both male and female sex–cells that can be used to reproduce uniparentally. It is frequent in nature; while commonly associated with plants where up to 50% self–fertilise to some degree (Igic and Kohn 2006), it is also present in animals (Jarne and Auld 2006), fungi (Billiard *et al.* 2011) and algae (Hanschen *et al.* 2018). Polygenic selection has been detected in the highly selfing plant *Arabidopsis thaliana* (Zan and Carlborg 2018; Exposito-Alonso *et al.* 2018; Tsuchimatsu *et al.* 2020; Wieters *et al.* 2021) but also in mixed–mating organisms (e.g., *Mimulus guttatus*; Troth *et al.* (2018)). There have also been examples of multiple beneficial mutations fixing in linkage in highly selfing organisms, in line with the idea that they co-segregate due to reduced recombination. These include the localised presence of genes affecting multiple traits (including climate adaptation,

root growth and flowering time) on Chromosome 3 of *A. thaliana* (Alonso-Blanco *et al.* 2016), and linked toxin-antidote pairs in self–fertilising *Caenorhabditis* nematodes (Ben-David *et al.* 2017, 2021). Given that most existing theory and inference methods are based on the assumption of randomly–mating populations, it may lead to inaccurate interpretations of results and inaccurate predictions.

Empirical findings additionally raise questions regarding the long–term evolutionary fate of self–fertilising species. The transition to self–fertilisation is widely observed in nature but tends to occur at the tips of species' trees, indicating that they are short–lived (reviewed by Wright *et al.* (2013); Barrett *et al.* (2014)). Self–fertilisation incurs several genetic effects (Burgarella and Glémin 2017; Glémin *et al.* 2019); most notably, the resulting inbreeding increases homozygosity and hence also the effects of genetic drift, which manifests itself as a reduction in $N_e$ thus weakening the efficacy of selection (Pollak 1987; Charlesworth 1992; Caballero and Hill 1992; Nordborg and Donnelly 1997). The resulting homozygosity also reduces effective recombination rates and, hence, the creation of new genetic combinations (Nordborg *et al.* 1996; Nordborg 2000). Due to these effects, self–fertilising species are hypothesised to suffer extinction risks due to deleterious mutation accumulating, although there is equivocal empirical evidence for this mechanism (Wright *et al.* 2013; Barrett *et al.* 2014; Hartfield 2016).

These factors can affect how such species adapt to new environments (Hartfield *et al.* 2017). Stebbins (1957) argued that the various genetic changes incurred by self–fertilisation will limit species' adaptive potential. In particular, selection acting on favourable mutations can be hampered through reduced diversity and increased genetic interference caused by decreased recombination (Glémin and Ronfort 2013; Kamran-Disfani and Agrawal 2014; Hartfield and Glémin 2014;

Hartfield and Glémin 2016). Conversely, due to reduced recombination and quicker fixation time of favourable alleles, selection footprints can be more pronounced in selfers as they are present across longer genetic distances (Glémin 2012; Hartfield and Bataillon 2020). Previous theoretical studies have tended to focus on either one– or two–locus models for mathematical tractability (but multi–locus simulation studies exist, e.g., Kamran-Disfani and Agrawal (2014)). However, the effect on positive selection acting over the whole genome is currently unclear. Under extremely high self–fertilisation, recombination can be so rare so as to effectively cause selection to act on the entire genome as a single linked unit. For example, selective sweeps affecting entire chromosomes have been detected in the selfing nematode *Caenorhabditis elegans* (Andersen *et al.* 2012). This process is related to the 'genotype selection' model of Neher and Shraiman (2009); when recombination is sufficiently weak and epistasis exists between mutations, a subset of genotypes have the highest fitness and proliferate in the population. In this scenario, reduced recombination can offer individuals a competitive advantage.

While there is a wide body of theoretical work investigating the mechanism of polygenic selection (de Vladar and Barton 2014; Stephan 2016; Jain and Stephan 2017a,b; Stetter *et al.* 2018; Simons *et al.* 2018; Thornton 2019; Höllinger *et al.* 2019; Hayward and Sella 2022; Stephan and John 2020; John and Stephan 2020; Harpak and Przeworski 2021; Szép *et al.* 2021; Sachdeva 2022), we lack clear theoretical predictions as to how these mating–system properties shape the response to polygenic selection, especially in light of the subsequent reduced heterozygosity and effective recombination. Sachdeva (2019) determined that self–fertilisation can help establish a population in a new location following a bottleneck. Clo *et al.* (2020) found that genetic associations formed due to reduced recombination in

self–fertilisers create cryptic genetic variation that can aid initial adaptation to a new environment, following occasional outcrossing. Hodgins and Yeaman (2019) investigated how self–fertilisation affects the clustering of polygenic variants underpinning local adaptation within a genome. Given these disparate studies, there is a need for a general theoretical investigation into the dynamics of polygenic selection under self–fertilisation, especially those that can be used to interpret the genetic signatures of polygenic selection from next–generation sequence data. In addition, these aforementioned studies only considered selection on a single trait, while adaptation to a new environment likely involve several traits, where mutations affect all of them (i.e., mutations are pleiotropic).

Here, we investigate the general behaviour of polygenic adaptation to a new environment under self–fertilisation. Specifically, we investigate whether increased self–fertilisation does indeed trigger a transition towards the formation of clonal genotypes, and what genetic properties either aid or prevent this transition from occurring.

# Results

We used the SLiM software (Haller and Messer 2019) to simulate a large population ($N = 5,000$) where each individual consisted of a 25Mb chromosome consisting of alternating neutral and gene regions (a similar setup was used by Hodgins and Yeaman (2019)). Mutations in genes can either be neutral, affect the quantitative traits determining fitness, or (in some simulations) always deleterious. We defined mutation and recombination rates that were realistic compared to empirical estimates ($4 \times 10^{-8}$ and $1.98 \times 10^{-7}$, respectively), but also ensured sufficient trait–

variants arose per generation to allow a polygenic selection mechanism (Höllinger *et al.* 2019). Fitness was determined by a Gaussian function, and trait mutations either affected a single trait (no pleiotropy) or five traits (pleiotropy present). There was a burn–in period of $10,000$ generations to an initial optimum value, then the optimum shifted and the population adapted to the new value. More information is available in the Methods section.

Given the large number of parameters in our model, we will introduce results in a piecewise manner. Specifically, we first investigate the general behaviour of polygenic selection under different self–fertilisation levels, then investigate how deviations from this core model affect the main results.

# Basic results – sudden optimum shift, no background selection

## Mean fitness and inbreeding depression

Following the optimum shift, populations quickly adapt to the new optimum irrespective of the frequency of self–fertilisation (Figure 1(a)). Despite this similarity, the mean fitness is highest and inbreeding depression lowest for highly self–fertilising populations (greater than 99%). One reason for this behaviour is that the variance in individual fitness is reduced under high self–fertilisation, limiting the proportion of maladaptive individuals with extreme trait values, and reducing the mutation load and inbreeding depression (Lande and Porcher 2015; Abu Awad and Roze 2018). Results are qualitatively the same when pleiotropy is present (Figure 1(b)), albeit the time to reach the optimum, and subsequent

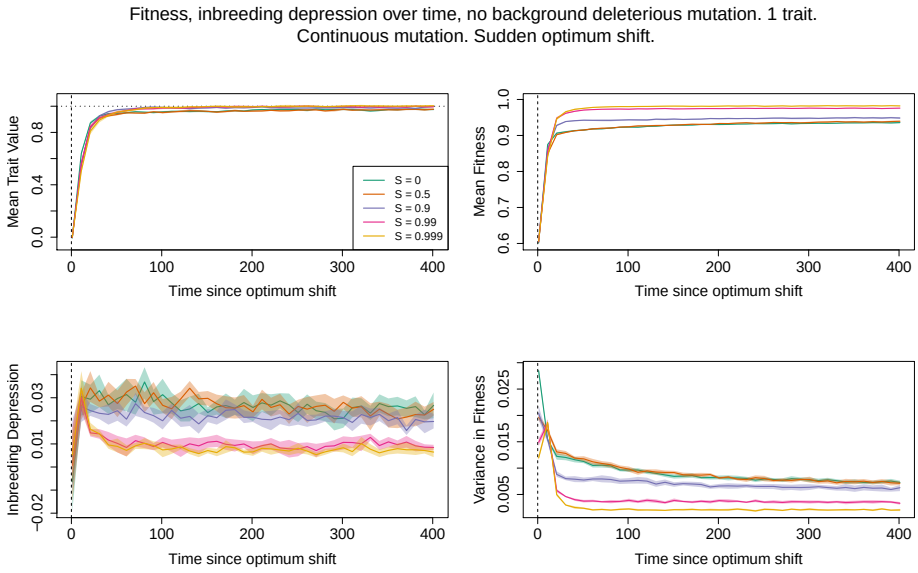
decays in inbreeding depression and variance in fitness, take a longer amount of time. Inbreeding depression also increases with pleiotropy, as previously observed by Ronce *et al.* (2009).

## Genetic variance and its decomposition

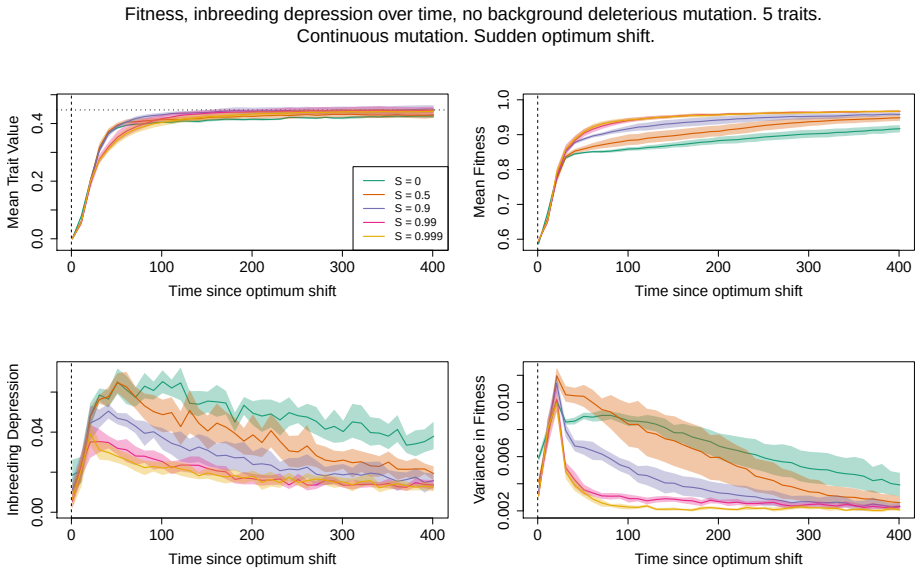
With one trait (no pleiotropy; Figure 2(a)), genetic variance increases after the optimum shift and remains at elevated values for most self–fertilisation values. These results reflect a process where large–effect mutations, both favourable and maladaptive, experience sudden frequency changes before reaching a new steady–state but not necessarily fix. However, with very high self–fertilisation (99% and above), variance instead spikes and subsequently reduces to zero. This likely reflects a transition away from minor allele changes (Höllinger *et al.* 2019) towards a process where all selected mutations are either fixed or are lost (see de Vladar and Barton (2014) for related results on genetic variation following polygenic selection).

Under pleiotropy (Figure 2(b)), genetic variance spikes then decreases after the optimum shift for all selfing values, showing how pleiotropy reduces the probability of polygenic selection occurring, and instead trait mutations are more likely to be fixed or lost rather than maintained at intermediate frequencies (de Vladar and Barton 2014). Note also that some variance components (e.g., inbreeding and LD covariance) take a long time to reach equilibrium values, even after the new optimum has been reached.

Our results for the different variance components generally reflect those obtained by Clo *et al.* (2020), who studied genetic variance in selfing species under adaptation. Both the genic and inbreeding variance increases with the degree

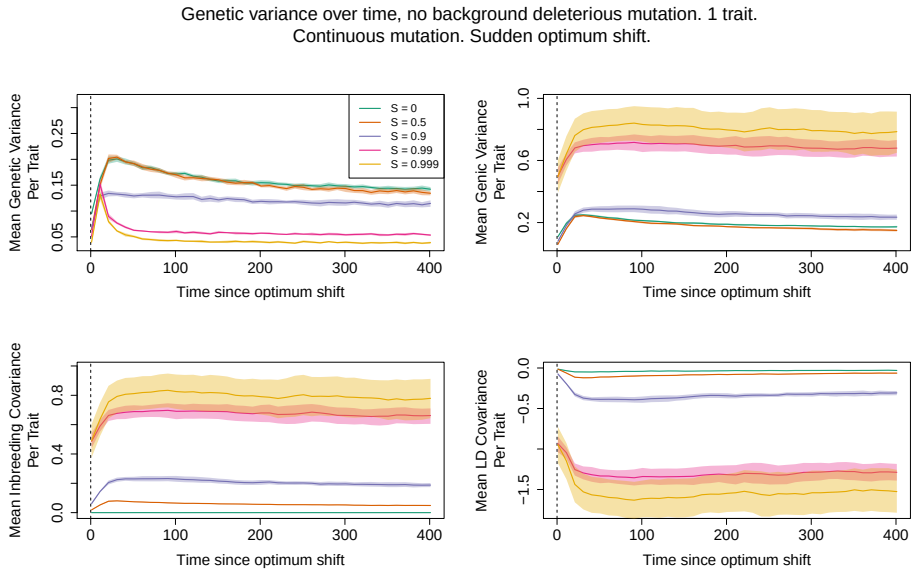



(a)

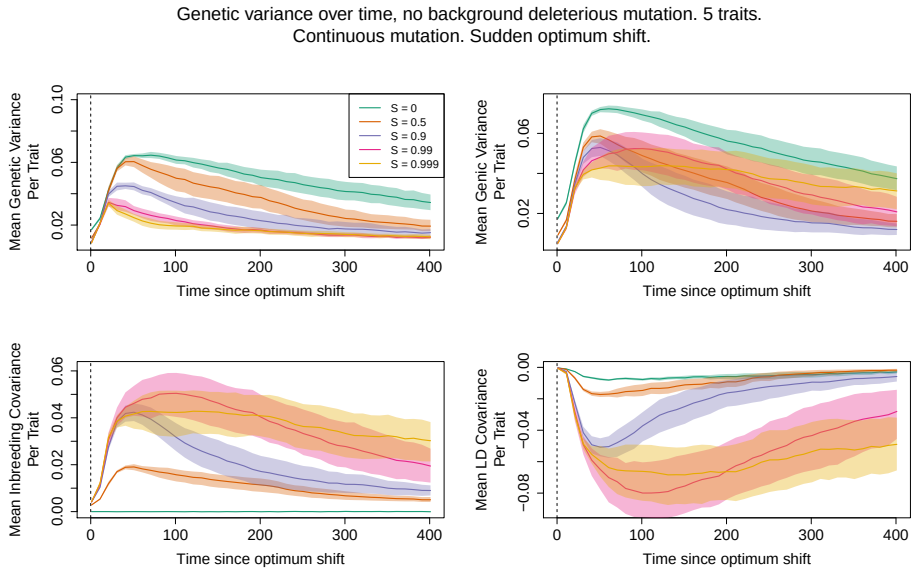



(b)

**Figure 1.** Mean trait value (top–left; horizontal dashed lines denote new trait optima); mean fitness (top–right); inbreeding depression (bottom–left); and variance in fitness (bottom–right) after an optimum shift. Different colours represent different fractions of self–fertilisation, as denoted by the legend. Solid lines are mean values, while bands denote 95% confidence intervals. Each trait variant affected either 1 (a) or 5 (b) traits.




(a)




(b)

**Figure 2.** Genetic variance (top–left); genic variance (top–right); inbreeding covariance (bottom–left); and LD covariance (bottom–right) after an optimum shift. Different colours represent different fractions of self–fertilisation, as denoted by the plot legend. Solid lines are mean values, while bands denote 95% confidence intervals. Each trait mutation affected either 1 (a) or 5 (b) traits.

of self–fertilisation, while the LD covariance is strongly negative. These results show how trait variants are more likely to be present as homozygotes at individual sites under high selfing (as reflected by increased inbreeding covariance), while also being in high linkage disequilibrium. Yet one key difference is that Clo *et al.* (2020) found adaptation proceeds through a reassortment of cryptic variation under high self–fertilisation, caused by residual outcrossing creating new genetic combinations. This process is characterised by a spike in additive variance and a decrease in the magnitude of LD covariance just after an optimum shift. However, we do not observe this outcome in our results. There are several possible reasons for this discrepancy due to differences in modelling assumptions. These include that Clo *et al.* (2020) assumed free recombination between sites, whereas linkage in our model could alter the selection response; stabilising selection is stricter in their model leading to negative LD covariance at equilibrium; cryptic reassortment occurred over very short timescales (20 generations); or that their optimum shift was much stronger ($z_1 = 2.5$, as opposed to no more than 1 here).

## Haplotype structure

### Trait variant frequencies

Haplotype plots demonstrate how trait variants spread following the optimum shift (going from left to right in Figure 3). To determine the effect of the mating system, we compare results under the cases of complete outcrossing against high self–fertilisation (99.9%). With one trait and outcrossing (Figure 3(a)), we see the typical polygenic selection regime at work; there are many variants present in the population, and as adaptation proceeds then while some of them increase

in frequency, we do not observe any particular mutation going to fixation via a selective sweep. Under high self–fertilisation (99.9% Figure 3(b)) we observe a drastic change in genome structure. Before the optimum shift there are instead many selected variants that have already fixed, with compensatory effects due to stabilising selection. As adaptation proceeds, we see other mutations segregating in the background and contributing to adaptation.

The presence of pleiotropy also affects these results. Under outcrossing (Figure 3(c)) then while the polygenic regime is still present, the density of mutations has been reduced. This reflects the known property that mutations are more likely to be deleterious with higher pleiotropy, so fewer are maintained in the population (Tenaillon 2014; Abu Awad and Roze 2018). Another major change occurs when both pleiotropy and self–fertilisation are present (Figure 3(d)); we observe a clear block–like structure within sampled genomes in line with a transition to genome–level selection. As the optimum shifts, we see positive–effect variants go to fixation in line with a traditional selective sweep, but with apparent genome–wide hitchhiking of mostly neutral mutations with the occasional linked trait variant.

These haplotype plots hint that the nature of polygenic selection differs as populations become highly selfing, however they are only snapshots from individual simulation replicates. To obtain a more systematic overview of how polygenic selection differs between the two scenarios, we calculated polygenic scores (Wray *et al.* 2007) for each population sample over simulation replicates, which is a weighted score of effect–sizes present and can be used to predict the trait response to selection. We also calculated other genetic properties that define these scores (e.g., mean variant frequency; see Methods for details). Figure 4 shows the surprising result that the polygenic score, and hence overall response to selection, is essentially

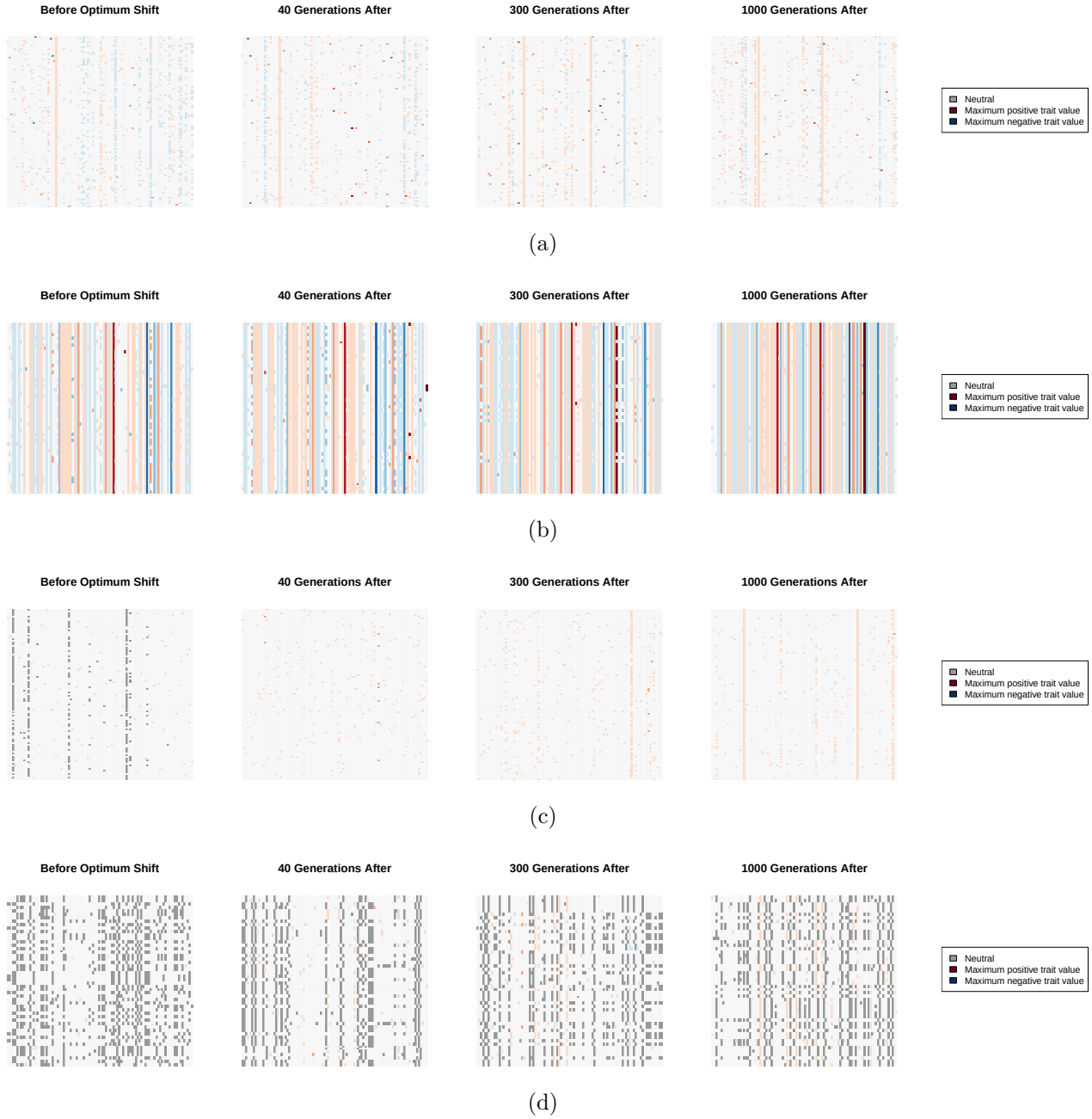
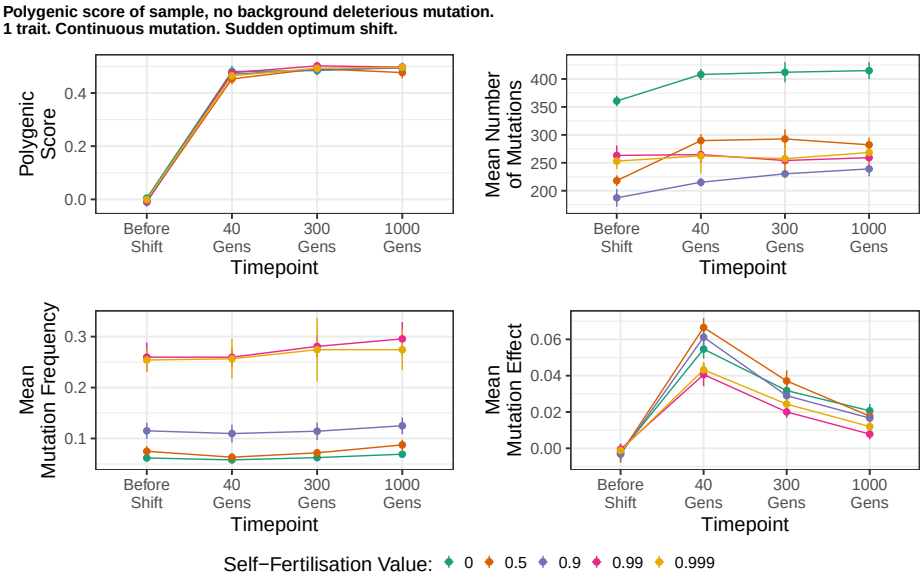



**Figure 3.** Haplotype plots over several time-steps. The X–axis correspond to genomic positions along the chromosome and the Y–axes to sampled individuals, taken at different times at denoted above each plot. Different colours represent mutation types; black points denote neutral variants, red points are trait mutations with a positive effect when averaged over all traits, and blue points are trait mutations with a negative–average effect. Darker colours denote a larger magnitude of average effect. Populations are either outcrossing (a), (c) or self–fertilising with $S = 0.999$ (b), (d). Mutations either affect 1 trait (a), (b) or 5 traits (c), (d).

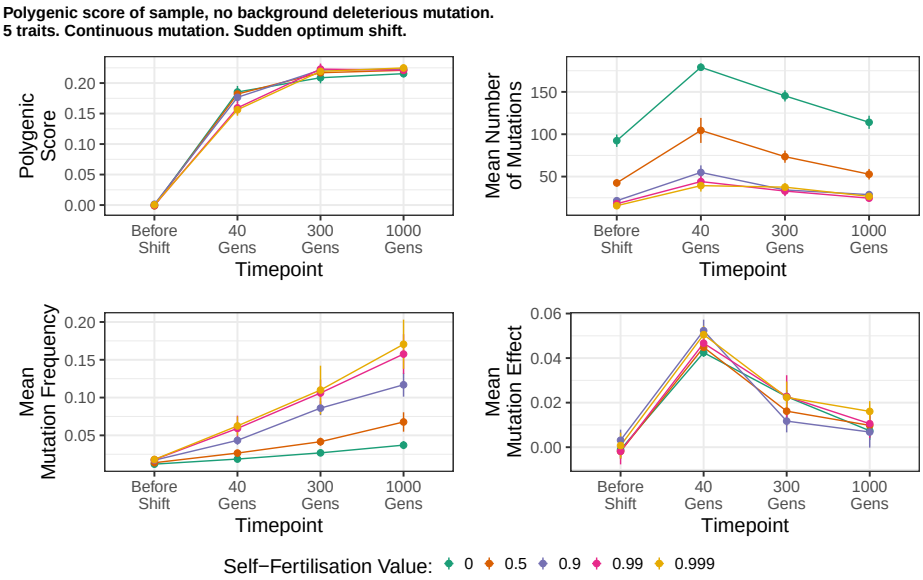
similar irrespective of the self–fertilisation rate. However, when inspecting the underlying properties of these scores, several differences become apparent. The number of variants generally decreases with the selfing rate, albeit it becomes slightly elevated for high selfing rates if pleiotropy is absent (Figure 4(b)). Conversely, the mean variant frequency is elevated under high selfing, and is noticeably higher when pleiotropy is absent. While the mean effect size spikes at 40 generations as adaptation is ongoing, the does not seem to be a noticeable association with the selfing rate but tends to be lower under the highest selfing values. Overall, these plots demonstrate that while the overall polygenic selection response is unaffected by the mating system, highly self–fertilising populations show a transition towards fewer, higher–frequency variants.

## Linkage disequilibrium

We can investigate the transition to a more 'clonal'–like genotype by plotting pairwise LD along the genome. Figure 5 shows examples of genome–wide heatmaps of LD for individual simulation replicates. In the outcrossing cases (Figure 5(a), 5(c)) most LD values are very low, indicating that variants segregate independently over long genetic distances in line with traditional polygenic selection mechanisms. With high selfing and no pleiotropy (Figure 5(b)) we observe regions of high LD being maintained over parts of the genome, with distinct genetic 'blocks' forming as adaptation occurs, reflecting how selection affects larger linked regions of the genome. When pleiotropy is added (Figure 5(d)) these LD blocks can extend across the genome as adaptation is occurring. Also note in this particular example is that LD blocks are formed long after the new optimum has been reached, in line with preceding results demonstrating that LD covariance can take long periods of




(a)




(b)

**Figure 4.** Polygenic scores (top–left), and their properties (number of mutations, top–right; mutation frequency, bottom–left; and mean mutation effect, bottom–right), as a function of time and averaged over all simulations. Variants affect either (a) one trait or (b) five traits.

time to equilibrate (Figure 2).

We can gain additional insight into how linkage blocks are formed by selfing by inspecting how LD changes as a function of distance. With outcrossing, LD is maintained at a near-zero level (Figure 6(a)), indicating little LD between SNPs (for this sampling level of 0.5Mb apart). Conversely, under very high self–fertilisation without pleiotropy, LD is maintained at high levels (between 0.25–0.75) before the optimum shift and continue at these high levels as adaptation is ongoing (Figure 6(b)). When pleiotropy is present, then while LD is also initially at a high value, it noticeably elevates as adaptation proceeds before reducing back towards pre-adaptation levels (Figure 6(c)).

There is also a wide variation in outcomes over individual simulation replicates, potentially indicating that the process of adaptation is greatly different between simulation runs. However, LD measurements are known to be affected by the underlying allele frequencies, even in measurements like $r^2$ that correct for them (Hedrick 1987; Lewontin 1988). We hence also investigated LD decay using the $|D'|$ statistic, which while still being affected by the underlying allele frequencies, normalises LD by the maximum possible value so it can yield insight into to what extent the range of possible LD values are present. In this case, $|D'|$ values are close to one over the observed distances in the 99.9% self–fertilisation case (Figure A in the Supplementary Material). Hence, most variants over a genome are in extremely tight linkage, with variations in $r^2$ reflecting differences in the underlying allele frequencies.

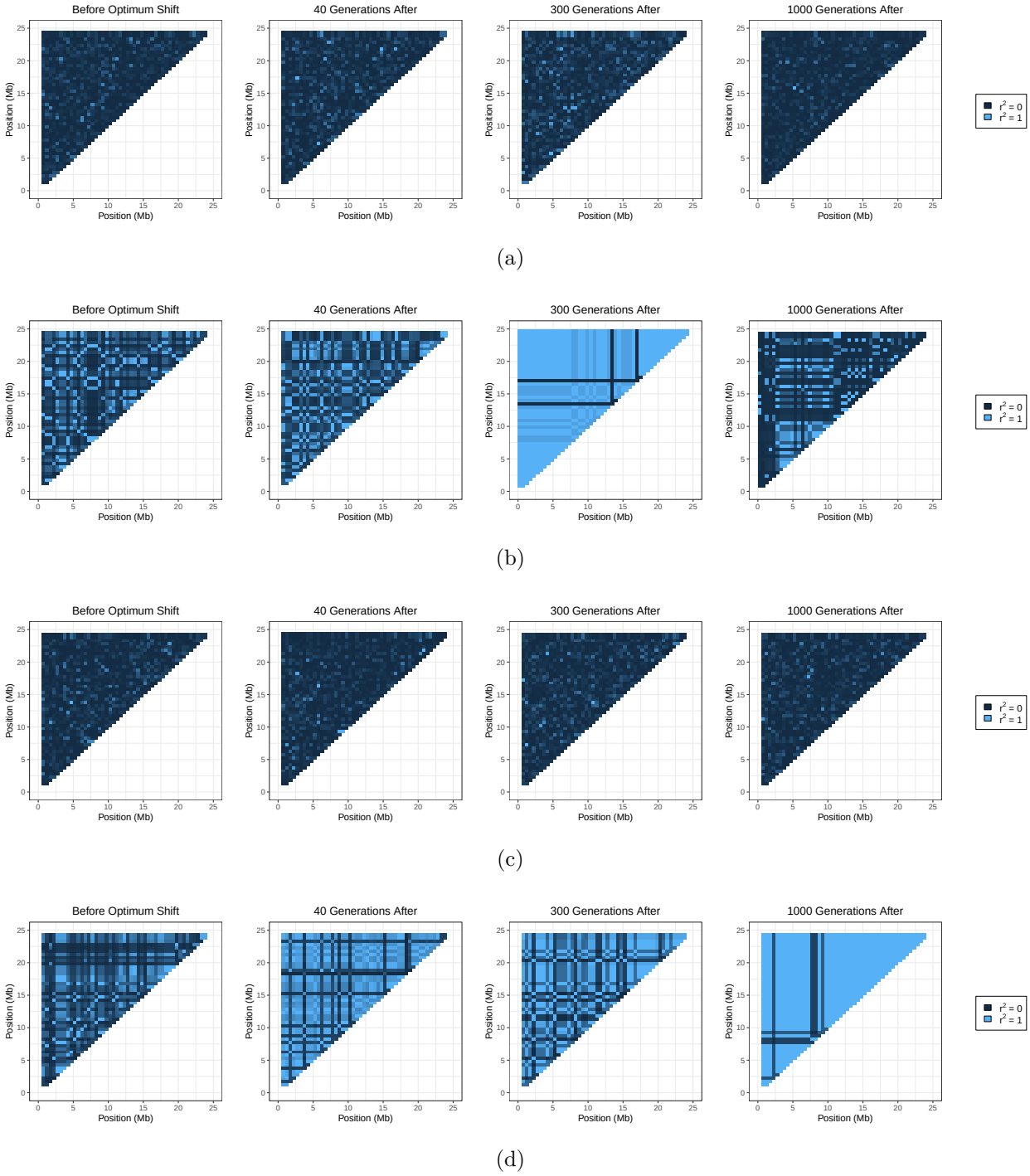
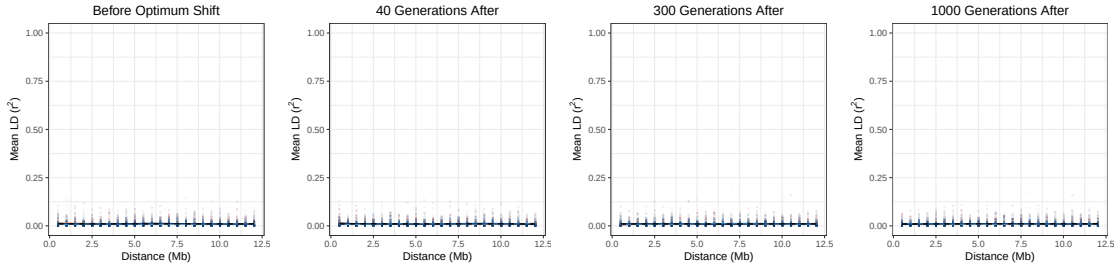
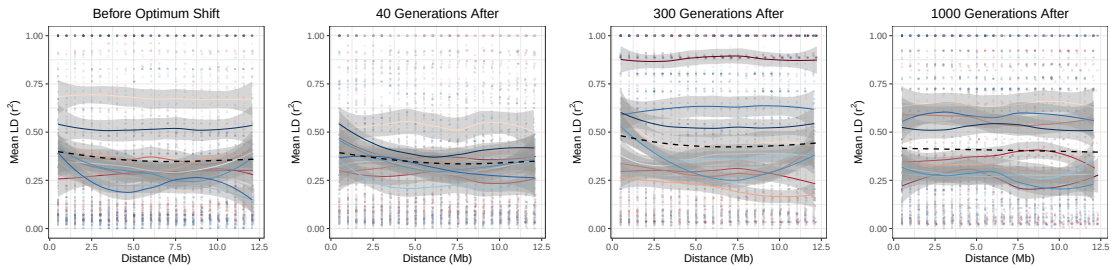


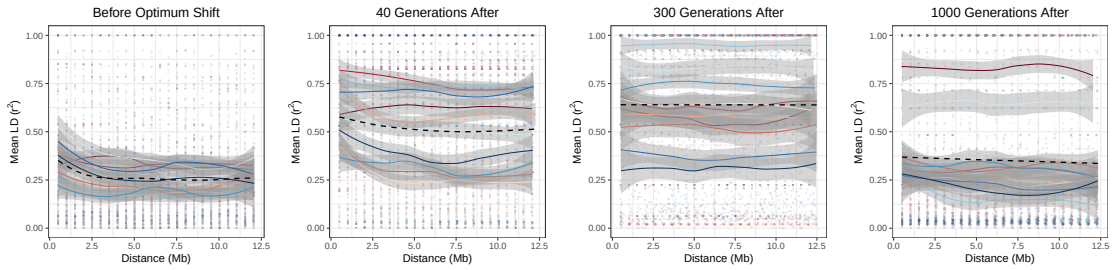

**Figure 5.** Heatmaps showing linkage disequilibrium, as measured by $r^2$, across the sampled genomes and time points shown in Figure 3. Colours range from dark blue representing $r^2 = 0$, to light blue representing $r^2 = 1$.




(a)




(b)




(c)

**Figure 6.** LD, as measured by $r^2$, as a function of distance (note that only distances up to 12.5Mb, covering half the simulated genome, are considered). Different colours represent different simulation replicates, with points being individual samples, and lines representing the mean value over distance (bands are 95% confidence intervals). The black dashed line shows the mean value over all simulations. The following cases are plotted: (a) outcrossing, no pleiotropy; (b) 99.9% selfing, no pleiotropy; (c) 99.9% selfing, pleiotropy present.

### Gradual optimum shift

In the Supplementary Material, we show results to cases where there is a gradual change in the optimum, as opposed to a sudden change. Results are outlined in Figures B–D, and are generally the same as for the instant–shift case with the main exception being that the formation of elevated polygenic scores occurs later after the optimum starts changing.

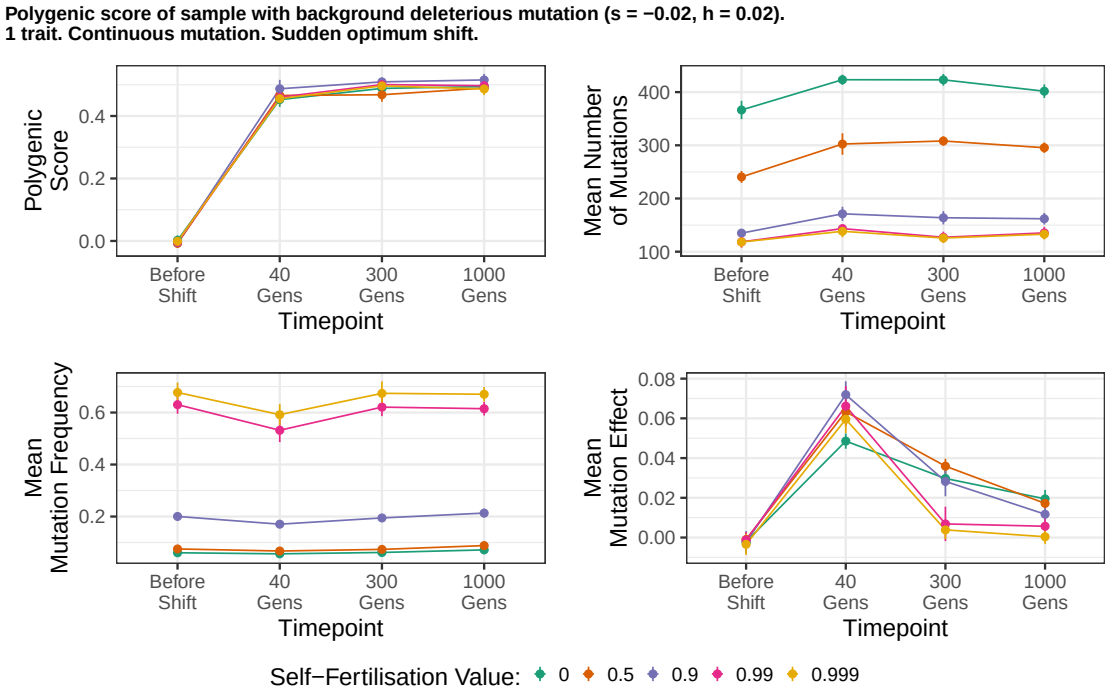
## Including background selection

Deleterious mutations are prevalent in nature (Keightley 2012; Charlesworth 2015), with empirical and theoretical work estimating that they have recessive effects on average (Manna *et al.* 2011). Recessive deleterious mutations are hypothesised to be a major contributor to inbreeding depression (Charlesworth and Willis 2009; Winn *et al.* 2011), and could potentially affect the role of polygenic adaptation through influencing trait variants via selective interference. We hence added unconditionally–recessive deleterious mutations to the simulations, either with dominance coefficient $h = 0.02$ or $h = 0.2$, to determine how they affect polygenic adaptation and the formation of linkage blocks.

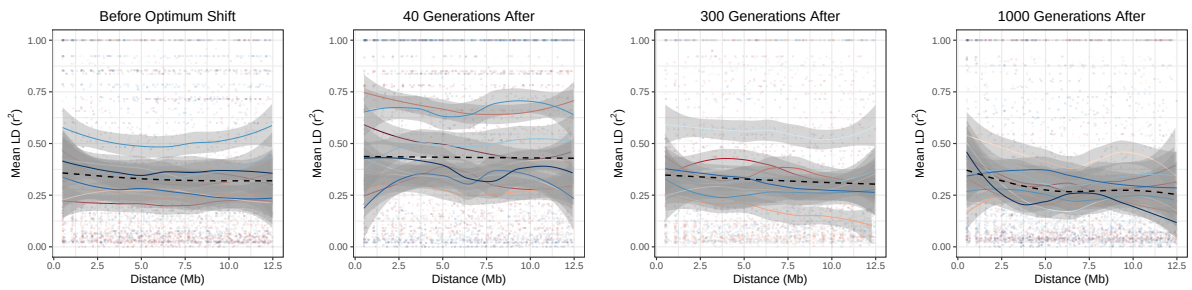
When recessive deleterious mutations are added ($h = 0.02$; Figure E in the Supplementary Material) then highly–selfing populations no longer have the largest mean fitness, despite exhibiting the lowest variance in fitness. This demonstrates how the presence of recessive deleterious mutations that are exposed to selection under selfing reduces the benefits of the mating system in strengthening stabilising selection. Inbreeding depression reduces with increased selfing, demonstrating how these deleterious mutations are purged from the population. The quantitative

genetic variance measurements do not greatly differ compared to the case without deleterious mutations (compare Figures 2(a) with E(b) in the Supplementary Material). One key difference is that the variance components for the highest selfing case revert back to zero over time, possibly indicating that the presence of deleterious mutations breaks down the formation of linkage blocks, and instead leads to adaptation via individual sweeps. Similar behaviour is observed when $h = 0.2$ (Figure F in the Supplementary Material).

The idea that deleterious mutations prevents the formation of high–fitness linkage blocks under high selfing is reinforced by looking at polygenic scores and LD decay (Figure 7). When looking at the components of polygenic scores (Figure 7(a)), we see that in very high selfing cases ($\geq 99\%$) both the mean number of mutations and the mean mutation effect after fixation are lower compared to cases with less than 90% selfing. However, the mean mutation frequency is very high (around 0.6). Under 99.9% selfing there are a lower variance in values over time; in particular, there are fewer high–LD cases at later timepoints (300 generations and afterwards), unlike results without deleterious mutations. Similar results arise when $h = 0.2$ (Figure G in the Supplementary Material). Overall, these results seem consistent with background selection disrupting adaptation in highly–selfing species, removing high–effect trait mutations so fewer fix together in linkage. Figures H, I in the Supplementary Material provide examples of haplotype plots and LD heatmaps for individual simulation runs.




(a)




(b)

**Figure 7.** (a) Polygenic scores and (b) LD decay of population samples under 99.9% selfing, when background deleterious mutations are also included, with dominance coefficient $h = 0.02$. Trait mutations are non–pleiotropic and there is a sudden optimum shift.

## Comparing the effects of self–fertilisation and low recombination in outcrossing species

The formation of linkage blocks under self–fertilisation is likely a consequence of reduced recombination, due to the long runs of homozygosity formed under inbreeding. Similar linkage effects could also arise in outcrossing genomes with a low overall recombination rate. To determine to what extent observed results reflect altered effective recombination and mutation rates under self–fertilisation, we compared the highly selfing results to an outcrossing population with rescaled recombination and mutation rates to reflect their effective values under high selfing (Nordborg and Donnelly (1997); Roze (2009); see Methods for further details).

Results are outlined in the Supplementary Material. We demonstrate that the genetic variance components are different between the two cases; in particular, there is an elevated inbreeding covariance in the highly–selfing case, reflecting how the clustering effects reflect increased homozygosity and purging of trait variance (Abu Awad and Roze 2018). When recessive deleterious mutations are aded, then low–recombination outcrossing populations exhibit stronger negative LD covariance, reflecting how selective interference is more prevalent in this scenario (Hill and Robertson 1966; Otto 2009). Finally, LD generally decays over shorter distances than in the highly–selfing cases, although some long–range LD can be formed if deleterious mutations are present. Hence long–range LD in selfing species is also a consequence of increased homozygosity.

# Discussion

## Summary of results

Self–fertilisation, like any form of inbreeding, changes how genetic variants are inherited between parents and offspring, which also affects how diversity is structured. These changes, including increased homozygosity and reduced effective recombination, also affects how polygenic selection is realised. Here, we ran a comprehensive set of simulations to investigate how polygenic adaptation is affected by self–fertilisation. In particular, we demonstrate that high self–fertilisation changes how polygenic selection is realised. A large number of compensatory mutations fix when pleiotropy is absent, but if mutations are pleiotropic then genetic associations are created that stretch across large parts of the genome.

Overall, we observe that self–fertilisation has a greatest influence on the long–term effects of genetic adaptation once an optimum has been reached. The mating–system has little influence on the time needed for the population to reach its new optimum, both in the simplest case where only trait variants are present (Figure 1) and also when deleterious alleles are present (Figures E, F in the Supplementary Material). In contrast, we observe that once the optimum has been reached, highly–selfing populations exhibit the highest fitness and lowest inbreeding depression (Figure 1). This advantage likely arises as these populations also exhibit the lowest variance in fitness through purging maladaptive genotypes (Abu Awad and Roze 2018). Similar results were obtained by Sachdeva (2019) in the context of tracking a population invading a new region; highly–selfing populations were more able to establish if the source population is also highly selfing and was able to purge its inbreeding depression.

By inspecting genetic variance and haplotype structure (Figures 2–6), we observe that self–fertilisation creates genome–wide linked blocks of trait variants consisting of compensatory mutations. Polygenic scores are generally unaffected by the mating system suggesting that the overall selection response is unaffected by it. However, we observe that the number of trait mutations are reduced under high selfing but are present at higher frequencies (Figure 4). If mutations are pleiotropic then we observe a transition from a polygenic–selection regime to a monogenic selection regime, where only a few large–effect variants go to fixation. LD is extensive in this case, but it seems that many hitchhiking mutations are neutral.

The presence of recessive deleterious mutations removes the long–term fitness advantage that highly self–fertilising organisms otherwise obtain under stabilising selection, likely because many deleterious mutation are exposed to selection (Figures E, F in the Supplementary Material). Their presence also weakens the LD structure that is otherwise formed during polygenic adaptation, with fewer high–LD populations present after the optimum has been reached (Figure 7). These results are similar to those observed by Hodgins and Yeaman (2019) in a related model of local adaptation under self–fertilisation, where the addition of deleterious mutations lead to clustering of variants that underlie local adaptation over shorter physical distances than in simulations when they were absent.

We hypothesised that the formation of these long–distance linkage blocks under self–fertilisation were principally caused by the weakened effect of recombination. We hence compared the highly–selfing results to an outcrossing model with rescaled recombination and mutation rates. However, the two cases do not yield equivalent results (details in the Supplementary Material). There appears to be a greater

purging of trait variation under the selfing case, but also a prevalence of selection interference in the outcrossing case when deleterious mutations are added. Long–range LD is still present under the rescaled outcrossing case but is not as strong as under the selfing case, but can be formed when deleterious mutations are present. Overall, these results suggest that it is not just the reduced effects of recombination causing a transition to genome–level selection, but increased homozygosity also plays an important role.

## Implications for genetic evolution and genome inference

One of the earliest predictions regarding adaptation in self–fertilising species is that selfers should be less able to adapt to their environment, compared to outcrossing species (Stebbins 1957). Our findings run contrary to these early, verbal predictions. The mating system has little effect on how long it takes species to adapt to new environments; once the new optimum is reached, then the highest fitness is usually observed in species with a high degree of self–fertilisation. These advantages seem to arise due to the purging effect of selfing, which removes both large–effect variants that lead to maladaptation, along with recessive deleterious mutations. Other theoretical studies have also shown advantages to self–fertilisation when adapting to new environments (Sachdeva 2019; Clo *et al.* 2020), highlighting how it does not always impedes adaptive potential. However, negative effects of selfing can adversely affect long–term population fitness (see 'Open Questions' below).

Traditional models of polygenic adaptation, and methods for detecting it, generally assume that individuals are idealised outcrossers and that polygenic vari-

ants segregate independently from one another. The results presented in this manuscript suggest that these assumptions are violated under high levels of self–fertilisation (typically above 90%). Clusters of selected mutations in high linkage disequilibrium are likely to form; in extreme cases large regions of the genome forming a block consisting of many trait variants in linkage with one another. Furthermore, the presence of pleiotropy causes many selected variants to be purged, meaning that adaptation is more likely to proceed via the fixation of fewer major–effect mutations (Figures 2, 3).

Given that blocks are more likely to form during polygenic selection under self–fertilisation, then their presence needs to be accounted for when testing for it in genome studies. Empirical polygenic selection studies, especially in humans, usually consider a single variant per linked region to correct for linkage effects (Field *et al.* 2016; Racimo *et al.* 2018; Berg *et al.* 2019; Sohail *et al.* 2019). Hence, it will be even more important to account for these regions before performing polygenic selection tests in self–fertilising organisms. LD, as measured by $r^2$, is a standard population–genetic statistic for identifying linked regions and so can be readily used in the first instance to identify co–linked regions. In addition, these results show that these outcomes are not just a consequence of reduced recombination in outcrossers. Self–fertilisation also leads to purging of deleterious mutations, as exemplified by reduced inbreeding depression (Figure E in the Supplementary Material). Future research can aim to disentangle the roles of reduced recombination and increased homozygosity in forming linked clusters.

Another challenge with analysing data from self–fertilising species, as highlighted by this study, is subsequently identifying what selective mutations are present in each block; whether there are many selected variants or just one vari-

ant with many hitchhikers. A first test could be to measure the length of swept haplotype blocks; longer regions are more likely to contain functional variants that contribute to polygenic selection, but such a test does not fully resolve the nature of selection within these haplotypes (e.g., whether mutations are major- or minor effects, or with compensatory effects). Given that pleiotropy is associated with a lower density of adaptive trait variants and more neutral hitchhikers, then measuring the pleiotropy of each gene in a block (through, for example, measuring how connected genes are in a regulatory network; Josephs *et al.* (2017); Mähler *et al.* (2017); Hämälä *et al.* (2020)) could provide information on whether adaptation is driven by polygenic or monogenic adaptation. Further information (e.g., looking at prevalence of non–synonymous variants or genes with known functions) can then shed light on possible selection mechanisms.

## Open Questions

A major question arising from this work is to what extent neutral diversity actually reflects neutral processes, given that linked selection can be extensive during polygenic adaptation. For example, linkage disequilibrium covariance takes a long time to equilibrate after the optimum is reached (Figure 2), which appears to result in long-distance LD formed many generations after the population reaches the new optimum (Figure 5). In evolutionary genetic studies, it is often important to separate out neutral genome regions from selected sites, but if LD is extensive then analyses of these putative neutral sites can be affected by selection. One example is estimating a population's demographic history from genome data, which can be misinferred if neutral regions have been affected by linked selection (Schrider

*et al.* 2016). Highly–selfing species can hence be affected by these linkage effects [e.g., *Caenorhabditis elegans*, which exhibits a self–fertilisation rate of at least 99% (Cutter *et al.* 2019) and for which chromosome–scale selective sweeps have been detected (Andersen *et al.* 2012)]. Similarly, LD–based approaches to infer population–scale recombination rates should be interpreted with caution in high selfers, as inferred recombination landscapes could also reflect selection landscapes. In these cases it will be important to filter genome data over longer regions to ensure putatively neutral SNPs are not in linkage disequilibrium, otherwise it may be necessary to jointly account for selection and demography (Li *et al.* 2012). If time–series data were available, one could measure synonymous diversity before and after adaptation to directly measure to what extent neutral diversity is affected by linked selection.

These findings also shed new light into how the mating–system, mutation rates and pleiotropy affect genetic variance under inbreeding. The earliest models concluded that inbreeding would increase additive variance in the absence of selection (Fisher 1918; Wright 1921). Lande (1977) subsequently showed that while stabilising selection would reduce genetic variance, the mating–system has little effect on the variance maintained in a population. Subsequent work has clarified how self–fertilisation does decrease genetic variance, but there is debate over whether this reduction is gradual Charlesworth and Charlesworth (1995); Abu Awad and Roze (2018) or drastically reduced ('purged') above a threshold selfing rate Lande and Porcher (2015). Our results show a mixture of both outcomes, depending on the pleiotropy present. In the simplest model (no background deleterious mutation or pleiotropy; Figure 2(a)) genetic variance is greatly reduced when selfing becomes really high, in line with the purging result of Lande and Porcher (2015).

However, once pleiotropy is present (Figure 2(b)) then genetic variance reduces more gradually with the selfing rate, in line with the models of Charlesworth and Charlesworth (1995) and Abu Awad and Roze (2018). This transition was likely caused by pleiotropy reducing the trait mutations present, so there was less trait variation to purge. However, additive variance was also elevated with high selfing when pleiotropy was absent, with strongly negative LD covariance reducing the overall genetic variance (see also Clo *et al.* (2020)). These results hint that the genetic variance present in selfers is affected by many factors, and elucidating how these all affect the long–term response to selection will be a rich area of future research.

Throughout, we have focussed on the creation of linked genotypes as measured by linkage disequilibrium (principally $r^2$ but also using $|D'|$). In mixed-mating populations, different levels of self–fertilisation in a population lead to 'identity disequilibrium', characterised by long runs of homozygosity (Haldane 1949; Weir and Cockerham 1973). Building on models by Kelly (1999a,b), Lande and Porcher (2015) used selfing age–classes in quantitative genetics models to account for identity–disequilibrium, so it is a likely cause of the variance purging they observe. We hence foresee that investigating the joint role of linkage and identity disequilibria will prove useful when developing novel methods for detecting polygenic selection under self–fertilisation. For example, Buffalo and Coop (2019, 2020) recently developed methods for detecting polygenic selection from temporal data, based on the covariance in allele frequency changes over time as formed via linkage disequilibrium. We hypothesise that it will be important to include identity disequilibrium effects if applying these methods to non–outcrossing organisms.

The results could also change with the modelling assumptions. We assumed 'universal pleiotropy' where each variant affected all traits simultaneously (Paaby and Rockman 2013). Partial pleiotropy, where only a few traits are affected by each mutation, could change the outcome of these simulations. However, since some pleiotropy will be present then we expect similar results to arise (i.e., monogenic selection with long–range linkage effects under high self–fertilisation). Fitness effects of mutations can also change if we alter the fitness landscape that determines the optimum (Tenaillon 2014). We also assumed a genome structure consisting of alternating genic and non–genic regions; the size and distribution of these could also affect results through changing the prevalence of trait variants and deleterious mutations.

A major open question is with regards to the long–term fate of outcrossing and self–fertilising species. In these simulations, populations with highest fitness after reaching the new optimum tended to be those with very high levels of self–fertilisation, and these advantages are not completely removed in the presence of deleterious mutations. Similar results were observed by Sachdeva (2019) who showed that the purging of inbreeding depression in selfers may aid colonization and adaptation to a new habitat. However, once the new optimum is reached then there is a risk that highly selfing populations could go extinct due to weakened selection against deleterious mutations (Wright *et al.* 2013; Barrett 2013; Barrett *et al.* 2014). Furthermore, selfing not only alters the genetic functioning of a species but is also expected to strongly affect its ecology and population dynamics, through creating more frequent bottlenecks along with smaller and more ephemeral populations. It is feasible that these demographic effects interact with linked selection, lowering the long–term adaptive potential of selfing species. It

remains to be tested whether these theoretical short–term advantages are present in nature and whether they are sufficient to maintain self–fertilising populations, and whether they interact with demography to create a long–term disadvantage to selfing lineages.

# Methods

## Simulation Procedure

Simulations were ran using SLiM version 3.3.2 (Haller and Messer 2019). A diploid population of $5,000$ individuals was simulated. We used a per–nucleotide mutation rate of $\mu = 4 \times 10^{-8}$; although this is approximately six–fold higher than that measured in *Arabidopsis thaliana* ($7 \times 10^{-9}$; Ossowski *et al.* 2010), the estimated $N_e$ for *A. thaliana* populations from southern Sweden is $198,400$, which is nearly 40-fold higher than that used here, so the net input of new mutations (as measured by $N_e\mu$) will actually be lower than expected in wild populations. The recombination rate was set to $\sim 1.98 \times 10^{-7}$; this value was obtained by converting the estimated genomic recombination rate of 3.6cM/Mb (Salome *et al.* 2012) to a per-nucleotide rate using Haldane's mapping formula (Haldane 1919), then rescaling so that the mutation–to–recombination rate ratio was the same as empirical estimates for *A. thaliana*.

Each individual had a genetic architecture similar to that used by Hodgins and Yeaman (2019). There was a single chromosome of length 25Mb long, consisting of 4000bp–long neutral buffer space and 1000bp–long gene regions. All mutations in buffer regions were neutral. In genes, 25% of mutations were neutral; 65% of mutations are deleterious, and 10% affect the quantitative trait. The neutral fraction reflects the number of changes affecting presumed–neutral synonymous sites (Hodgins and Yeaman 2019). With this setup, an average of 100 new trait mutations will be introduced into the population every generation. In a theoretical analysis of polygenic adaptation, Höllinger *et al.* (2019) showed that this level of mutation input will lead to polygenic adaptation via subtle frequency shifts. Note

that the Gaussian fitness function used here (see below) is different from that in Höllinger *et al.* (2019), so we employ their model as a general guide to understand simulation results, rather than trying to exactly replicate their findings. The frequency of self–fertilisation was set to one of 0, 0.5, 0.9, 0.99, or 0.999. The three first values allow to compare outcrossing, mixed mating and high selfing, whereas the two last ones focus on the presumed critical zone between high and very high selfing rates where clonal genotypes are likely to form.

We investigated cases where the only selected mutations were those that affected the trait, or where a fraction of mutations were uniformly deleterious. If present, these deleterious mutations had a selection coefficient $s = -0.02$, and a dominance coefficient of either $h = 0.02$ or 0.2. Otherwise, we set $s = 0$ to disable their presence. Each trait mutation affected $n$ sub–traits, which we either set to 1 (no pleiotropy) or 5 (pleiotropy present). For $n = 1$, mutation effect sizes were drawn from a normal distribution with mean 0 and standard deviation $\sigma = 0.25$. If $n > 1$ then effect sizes were drawn from a multinomial distribution, where each trait had mean 0 and variance $(0.25)^2/\sqrt{n}$. When pleiotropy is present, the mean selection coefficient of mutations is proportional to $n\sigma^2$ (Tenaillon 2014); rescaling the trait variance by $\sqrt{n}$ will therefore enable the mean selective effects to be approximately the same as in the univariate case, enabling us to test for the effect of pleiotropy alone. Trait effects were additive within– and between loci; hence, homozygote mutations had twice the effect size as heterozygotes, and the value of trait $i$ within an individual, $z_i$, is summed over all mutations. Note that trait values were not standardised so that homozygote genotypes had the same magnitude of effect but in opposite directions, as is often used in quantitative genetics models (Falconer and Mackay 1996).

Individuals' fitness were calculated from the product of those contributed by both background deleterious mutations and trait mutations. The fitness contributed by deleterious mutations is $w_d = (1 - hs)^{nhet}(1 - s)^{nhom}$ for $nhet$ the number of heterozygote sites and $nhom$ the number of homozygote sites. The fitness contributed by trait mutations, $w_q$, is calculated over all $n$ traits using a multivariate Gaussian fitness function:

$$w_q = \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(z_i - z_0)^2}{V_s}\right) \tag{1}$$

where $z_0$ is the optimal trait value, and $V_s$ is the variance in the Gaussian fitness curve, which determines how strict fitness is. $V_s$ is set to 1 throughout. Note that although mutation effects are additive within and between loci, stabilising selection naturally induces dominance and epistasis on fitness, although mean epistasis at the optimum is zero under this fitness function (Tenaillon 2014).

Simulations were ran for a a burn–in period of $2N = 10{,}000$ generations with $z_0 = 0$ for all traits. In the main simulation set, the optimum was subsequently and instantaneously changed to $z_0 = 1/\sqrt{n}$; this new value ensures that the initial drop in fitness due to maladaptation is the same irrespective of $n$ (Tenaillon 2014). We also ran simulations where the optimum gradually changed from 0 to $1/\sqrt{n}$ over the first 100 generations after the burn–in, with the optimum increasing by $1/(100\sqrt{n})$ each time. All simulations were ran for 1001 generations following the burn–in to enable the population to adapt to the new optimum.

We also simulated an outcrossing population but with rescaled parameters to match those expected under self–fertilisation, to test whether the observed results are caused principally by the reduced recombination incurred. In these simulations,

the mutation rate was divided by $(1 + F)$ for $F$ the inbreeding coefficient, which equals $\sigma/(2-\sigma) \approx 0.998$ for a selfing fraction $\sigma = 0.999$. This rescaling reflects the reduction in effective population size $N_e$ caused by selfing, and hence the effect on the net introduction of mutations $N_e\mu$ (Nordborg and Donnelly 1997). Rescaling the recombination rate is tricker; for a genome tract with total recombination rate $r$ acting over it, the effective recombination rate under selfing equals $r(1 - 2F + \Phi)$, where $\Phi$ is the probability of joint identity–by–descent at two loci and is a function of both the selfing and recombination rates; $\Phi \approx F$ unless self–fertilisation and recombination rates are both large (Roze 2009). By default, SLiM defines recombination rates on a per–basepair basis, but defining $\Phi$ using this value is ineffective if we want to rescale recombination across the whole genome. To this end, we rescaled the recombination rate using $\Phi$ defined using $r = 1/2$; the justification here is that the mean number of breakpoints per individual, $r(L - 1)$ for $L$ the number of basepairs (25Mb), will subsequently be rescaled by $(1-2F+\Phi)$. In this case, $\Phi = [\sigma(2+\sigma)]/[(4-\sigma)(2-\sigma)] \approx 0.997$ for $\sigma = 0.999$, and the rescaled recombination rate approximately equals $2.65 \times 10^{-10}$.

## Calculating statistics

All statistics, except haplotype diversity, were outputted every 10 generations after the optimum shift. We were interested in capturing three broad patterns of adaptation: (i) the change in fitness over time; (ii) the genetic variance over time; (iii) haplotype structure, and how many trait variants contribute to adapted genotypes.

To measure fitness effects, we outputted the population mean trait value; the

mean and variance in fitness; and inbreeding depression. The latter was calculated by creating two new subpopulations, each of size 500, where one was obligately outcrossing and one was obligately self–fertilising. Inbreeding depression was subsequently calculated as $1 - (w_S/w_O)$, for $w_S$ the mean fitness of the self–fertilising subpopulation and $w_O$ the mean fitness of the outcrossing populations. Mutation rates were also set to zero in each subpopulation, so fitness calculations were based on variation that was only present in the main population. Both subpopulations were subsequently removed until the next check.

We also calculated the genetic variance, and decomposed it to determine which factors most strongly contribute to adaptation (a similar approach was used by Clo *et al.* (2020)). Let $p_j$ denote the frequency of trait mutation $j$; $f_{0,j}$, $f_{1,j}$ and $f_{2,j}$ be the frequency of wild-type homozygotes, heterozygotes, and derived homozygotes; and $a_{i,j}$ the effect size of mutation $j$ affecting trait $i$. Then we calculate the genetic variance of trait $i$, $V_{G,i}$, as:

$$V_{G,i} = \sum_j \left( f_{1,j} a_{i,j}^2 + f_{2,j} (2a_{i,j})^2 \right) - \mu_i \tag{2}$$

$$\text{for mean } \mu_i = \sum_j \left( f_{1,j} a_{i,j} + f_{2,j} (2a_{i,j}) \right) \tag{3}$$

We further decompose the genetic variance into the genic variance $V_g$, which is the variance assuming Hardy–Weinberg expectations and no linkage; the inbreeding covariance $C_I$ reflecting deviations from Hardy–Weinberg equilibria; and the linkage covariance $C_{LD}$. These terms were calculated using equations provided in Bulmer (1976), modified to take different effect sizes into account:

$$V_{g,i} = \sum_j 2p_j(1-p_j)a_{i,j}^2 \tag{4}$$

$$C_{I,i} = \sum_j (2f_{2,j}f_{0,j} - (1/2)f_{1,j})a_{i,j}^2 \tag{5}$$

$$C_{LD,i} = V_{G,i} - V_{g,i} - C_{I,i} \tag{6}$$

We plot the mean of the variance values over all traits. To check the simulation code, we tested whether the pre–shift genic variance $V_g$ matched the house–of–cards expectation, $4NU_t$ for $U_t$ the total mutation–rate of trait variants (Turelli 1984), in the simplest case of outcrossing, no pleiotropy or background deleterious mutations. Given the genome setup and mutation rates, $V_g = 0.08$ for default parameters. We observed that the actual $V_g$ exceeded this expectation (Supplementary Figure R), which is a known mis–match if mutation rates are high (Abu Awad and Roze 2018). However, if we reduced the mutation and recombination–rates 10–fold so expected $V_g = 0.008$, then the simulations did match this value. We hence conclude that the simulation seems to work as expected. 95% confidence intervals for fitness and variance measurements were calculated using 1,000 bootstrap replicates.

Haplotype plots were produced for the first simulation replicate of each parameter set. 50 individuals were sampled either the generation before the optimum shift if there was a burn–in; 40 generations after the optimum shift; 300 generations afterwards; and 1,000 generations afterwards. Haplotype information was printed out as a VCF file, along with information on the effect size of each trait variant. The number of mutations in haplotype plots were thinned to 100; if there

were less than 100 trait mutations, then haplotype plots consist of all fixed trait variants; segregating trait variants; then a sample of background mutations, both deleterious and neutral, if space allowed.

Polygenic scores were calculated by taking all the trait–variants in the VCF file, obtaining the mean frequency $p_j$ for mutation $j$ and the mean effect size $\overline{a}_j$ over all traits; the polygenic score is subsequently calculated as $\sum_j p_j \overline{a}_j$ (following Wray *et al.* (2007)). We also measured the mean number of mutations; the mean mutation frequency; and the mean mutation effect (over all mutations) to further understand the components of the polygenic score. 95% confidence intervals were calculated using a Student's t–distribution.

Linkage disequilibrium (hereafter LD), as measured using both $r^2$ (Hill and Robertson 1968) and $|D'|$ (Lewontin 1964), was calculated using *VCFtools* (Danecek *et al.* 2011) after thinning SNPs so that variants are at least 0.5Mb apart and only including those with a minor allele frequency of 0.1. Results were plotted in two ways: first as a heatmap to demonstrate genome-wide clustering of mutants for a single simulation replicate, and as a function of distance between two alleles to show the decay of LD over the genome for all simulation replicates. In the latter case, we follow a similar procedure as in Hartfield *et al.* (2018); we first grouped measurements in bins of 0.5Mb long, based on the distance between the two SNPs. We then only consider distances less than 12.5Mb as measurements become too noisy otherwise due to fewer comparisons. We then thinned the number of entries to equal the smallest number over these remaining bins, to further reduce the effect of within-bin stochasticity. If the smallest number of entries in these bins is less than 10, we omitted that simulation replicate from the plot.

## Other software used

Data was processed and plotted in R 3.6.1 (R Core Team 2021). We also used the following R packages: *cowplot* (Wilke 2020); *gplots* (Warnes *et al.* 2020), *plyr* (Wickham 2011), *RColorBrewer* (Neuwirth 2014), *ggplot* (Wickham 2016) and *tidyverse* (Wickham *et al.* 2019). Output PDFs were combined into single files using *pdfjam* (`https://github.com/DavidFirth/pdfjam`).

**Acknowledgments.** We would like to thank Josselin Clo for providing help with interpreting results. MH is supported by a NERC Independent Research Fellowship (NE/R015686/1). SG is supported by the CNRS.

**Data availability.** Simulation code and data processing scripts are available from `https://github.com/MattHartfield/Polygenic-Selfing`.

# References


Abu Awad, D. and Roze, D. 2018. Effects of partial selfing on the equilibrium genetic variance, mutation load, and inbreeding depression under stabilizing selection. *Evolution*, 72(4): 751–769.

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M., Cao, J., Chae, E., Dezwaan, T. M., Ding, W., Ecker, J. R., Exposito-Alonso, M., Farlow, A., Fitz, J., Gan, X., Grimm, D. G., Hancock, A. M., Henz, S. R., Holm, S., Horton, M., Jarsulic, M., Kerstetter, R. A., Korte, A., Korte, P., Lanz, C., Lee, C.-R., Meng, D., Michael, T. P., Mott, R., Muliyati, N. W., Nägele, T., Nagler, M., Nizhynska, V., Nordborg, M., Novikova, P. Y., Picó, F. X., Platzer, A., Rabanal, F. A., Rodriguez, A., Rowan, B. A., Salomé, P. A., Schmid, K. J., Schmitz, R. J., Seren, Ü., Sperone, F. G., Sudkamp, M., Svardal, H., Tanzer, M. M., Todd, D., Volchenboum, S. L., Wang, C., Wang, G., Wang, X., Weckwerth, W., Weigel, D., and Zhou, X. 2016. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell*, 166(2): 481–491.

Andersen, E. C., Gerke, J. P., Shapiro, J. A., Crissman, J. R., Ghosh, R., Bloom, J. S., Felix, M.-A., and Kruglyak, L. 2012. Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat. Genet.*, 44(3): 285–290.

Barghi, N., Hermisson, J., and Schlötterer, C. 2020. Polygenic adaptation: a unifying framework to understand positive selection. *Nat. Rev. Genet.*

Barrett, S. C. H. 2013. The evolution of plant reproductive systems: how often are transitions irreversible? *Proc. R. Soc. B*, 280(1765).

Barrett, S. C. H., Arunkumar, R., and Wright, S. I. 2014. The demography and population genomics of evolutionary transitions to self-fertilization in plants. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 369(1648): 20130344.

Ben-David, E., Burga, A., and Kruglyak, L. 2017. A maternal-effect selfish genetic element in *Caenorhabditis elegans*. *Science*, 356(6342): 1051–1055.

Ben-David, E., Pliota, P., Widen, S. A., Koreshova, A., Lemus-Vergara, T., Verpukhovskiy, P., Mandali, S., Braendle, C., Burga, A., and Kruglyak, L. 2021. Ubiquitous selfish toxin-antidote elements in *Caenorhabditis* species. *Curr. Biol.*, 31(5): 990–1001.e5.

Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., Boyle, E. A., Zhang, X., Racimo, F., Pritchard, J. K., and Coop, G. 2019. Reduced signal for polygenic adaptation of height in UK Biobank. *eLife*, 8: e39725.

Billiard, S., López-Villavicencio, M., Devier, B., Hood, M. E., Fairhead, C., and Giraud, T. 2011. Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. *Biol. Rev. Camb. Philos. Soc.*, 86(2): 421–442.

Buffalo, V. and Coop, G. 2019. The linked selection signature of rapid adaptation in temporal genomic data. *Genetics*, 213(3): 1007–1045.

Buffalo, V. and Coop, G. 2020. Estimating the genome-wide contribution of selection to temporal allele frequency change. *Proc. Natl. Acad. Sci. USA*, 117(34): 20672–20680.

Bulmer, M. G. 1976. The effect of selection on genetic variability: a simulation study. *Genet. Res.*, 28: 101–117.

Burgarella, C. and Glémin, S. 2017. *Population Genetics and Genome Evolution of Selfing Species*, pages 1–8. John Wiley & Sons, Ltd. doi: 10.1002/9780470015902.a0026804.

Caballero, A. and Hill, W. G. 1992. Effects of partial inbreeding on fixation rates and variation of mutant genes. *Genetics*, 131(2): 493–507.

Charlesworth, B. 1992. Evolutionary rates in partially self-fertilizing species. *Am. Nat.*, 140(1): 126–148.

Charlesworth, B. 2015. Causes of natural variation in fitness: Evidence from studies of *Drosophila* populations. *Proc. Natl. Acad. Sci. USA*, 112(6): 1662–1669.

Charlesworth, D. and Charlesworth, B. 1995. Quantitative genetics in plants: The effect of the breeding system on genetic variability. *Evolution*, 49(5): 911–920.

Charlesworth, D. and Willis, J. H. 2009. The genetics of inbreeding depression. *Nat. Rev. Genet.*, 10(11): 783–796.

Clo, J., Ronfort, J., and Abu Awad, D. 2020. Hidden genetic variance contributes to increase the short-term adaptive potential of selfing populations. *J. Evol. Biol.*, 33(9): 1203–1215.

Cutter, A. D., Morran, L. T., and Phillips, P. C. 2019. Males, outcrossing, and sexual selection in *Caenorhabditis* nematodes. *Genetics*, 213(1): 27–57.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., and Durbin, R. 2011. The variant call format and VCFtools. *Bioinformatics*, 27(15): 2156–2158.

de Vladar, H. P. and Barton, N. 2014. Stability and response of polygenic traits to stabilizing selection and mutation. *Genetics*, 197(2): 749–767.

Exposito-Alonso, M., Vasseur, F., Ding, W., Wang, G., Burbano, H. A., and Weigel, D. 2018. Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*. *Nat. Ecol. Evol.*, 2(2): 352–358.

Fagny, M. and Austerlitz, F. 2021. Polygenic adaptation: Integrating population genetics and gene regulatory networks. *Trends Genet.*, 37(7): 631–638.

Falconer, D. S. and Mackay, T. F. C. 1996. *Introduction to quantitative genetics*. Longman, Harlow, 4th edition.

Field, Y., Boyle, E. A., Telis, N., Gao, Z., Gaulton, K. J., Golan, D., Yengo, L., Rocheleau, G., Froguel, P., McCarthy, M. I., and Pritchard, J. K. 2016. Detection of human adaptation during the past 2000 years. *Science*, 354(6313): 760–764.

Fisher, R. A. 1918. The correlation between relatives on the supposition of mendelian inheritance. *Trans. R. Soc. Edinb.*, 52(2): 399–433.

Glémin, S. 2012. Extinction and fixation times with dominance and inbreeding. *Theor. Popul. Biol.*, 81(4): 310–316.

Glémin, S. and Ronfort, J. 2013. Adaptation and maladaptation in selfing and outcrossing species: New mutations versus standing variation. *Evolution*, 67(1): 225–240.

Glémin, S., François, C. M., and Galtier, N. 2019. Genome evolution in outcrossing vs. selfing vs. asexual species. In M. Anisimova, editor, *Evolutionary Genomics: Statistical and Computational Methods*, pages 331–369. Springer New York, New York, NY.

Haldane, J. B. S. 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics*, 8: 299–309.

Haldane, J. B. S. 1949. The association of characters as a result of inbreeding and linkage. *Ann. Eugen.*, 15(1): 15–23.

Haller, B. C. and Messer, P. W. 2019. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Mol. Biol. Evol.*, 36(3): 632–637.

Hämälä, T., Gorton, A. J., Moeller, D. A., and Tiffin, P. 2020. Pleiotropy facilitates local adaptation to distant optima in common ragweed (*Ambrosia artemisiifolia*). *PLoS Genet.*, 16(3): e1008707.

Hanschen, E. R., Herron, M. D., Wiens, J. J., Nozaki, H., and Michod, R. E. 2018. Repeated evolution and reversibility of self-fertilization in the volvocine green algae. *Evolution*, 72(2): 386–398.

Harpak, A. and Przeworski, M. 2021. The evolution of group differences in changing environments. *PLoS Biol.*, 19(1): e3001072.

Hartfield, M. 2016. Evolutionary genetic consequences of facultative sex and outcrossing. *J. Evol. Biol.*, 29(1): 5–22.

Hartfield, M. and Bataillon, T. 2020. Selective sweeps under dominance and inbreeding. *G3*, 10(3): 1063–1075.

Hartfield, M. and Glémin, S. 2014. Hitchhiking of deleterious alleles and the cost of adaptation in partially selfing species. *Genetics*, 196(1): 281–293.

Hartfield, M. and Glémin, S. 2016. Limits to adaptation in partially selfing species. *Genetics*, 203(2): 959–974.

Hartfield, M., Bataillon, T., and Glémin, S. 2017. The evolutionary interplay between adaptation and self-fertilization. *Trends Genet.*, 33(6): 420–431.

Hartfield, M., Wright, S. I., and Agrawal, A. F. 2018. Coalescence and linkage disequilibrium in facultatively sexual diploids. *Genetics*, 210(2): 683–701.

Hayward, L. K. and Sella, G. 2022. Polygenic adaptation after a sudden change in environment. *eLife*, 11: e66697.

Hedrick, P. W. 1987. Gametic disequilibrium measures: Proceed with caution. *Genetics*, 117(2): 331–341.

Hill, W. G. and Robertson, A. 1966. The effect of linkage on limits to artificial selection. *Genet. Res.*, 8: 269–294.

Hill, W. G. and Robertson, A. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.*, 38(6): 226–231.

Hodgins, K. A. and Yeaman, S. 2019. Mating system impacts the genetic architecture of adaptation to heterogeneous environments. *New Phytol.*, 224(3): 1201–1214.

Höllinger, I., Pennings, P. S., and Hermisson, J. 2019. Polygenic adaptation: From sweeps to subtle frequency shifts. *PLoS Genet.*, 15(3): e1008035.

Igic, B. and Kohn, J. R. 2006. The distribution of plant mating systems: study bias against obligately outcrossing species. *Evolution*, 60(5): 1098–1103.

Jain, K. and Stephan, W. 2017a. Modes of rapid polygenic adaptation. *Mol. Biol. Evol.*, 34(12): 3169–3175.

Jain, K. and Stephan, W. 2017b. Rapid adaptation of a polygenic trait after a sudden environmental shift. *Genetics*, 206(1): 389–406.

Jarne, P. and Auld, J. R. 2006. Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. *Evolution*, 60(9): 1816–1824.

John, S. and Stephan, W. 2020. Important role of genetic drift in rapid polygenic adaptation. *Ecol. Evol.*, 10(3): 1278–1287.

Josephs, E. B., Wright, S. I., Stinchcombe, J. R., and Schoen, D. J. 2017. The relationship between selection, network connectivity, and regulatory variation within a population of capsella grandiflora. *Genome Biol. Evol.*, 9(4): 1099–1109.

Kamran-Disfani, A. and Agrawal, A. F. 2014. Selfing, adaptation and background selection in finite populations. *J. Evol. Biol.*, 27(7): 1360–1371.

Keightley, P. D. 2012. Rates and fitness consequences of new mutations in humans. *Genetics*, 190(2): 295–304.

Kelly, J. K. 1999a. Response to selection in partially self-fertilizing populations. i. selection on a single trait. *Evolution*, 53(2): 336–349.

Kelly, J. K. 1999b. Response to selection in partially self-fertilizing populations. ii. selection on multiple traits. *Evolution*, 53(2): 350–357.

Lande, R. 1977. The influence of the mating system on the maintenance of genetic variability in polygenic characters. *Genetics*, 86(2): 485–498.

Lande, R. and Porcher, E. 2015. Maintenance of quantitative genetic variance under partial self-fertilization, with implications for evolution of selfing. *Genetics*, 200(3): 891–906.

Lewontin, R. C. 1964. The interaction of selection and linkage. i. general considerations; heterotic models. *Genetics*, 49(1): 49–67.

Lewontin, R. C. 1988. On measures of gametic disequilibrium. *Genetics*, 120(3): 849–852.

Li, J., Li, H., Jakobsson, M., Li, S., Sjödin, P., and Lascoux, M. 2012. Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Mol. Ecol.*, 21(1): 28–44.

Mähler, N., Wang, J., Terebieniec, B. K., Ingvarsson, P. K., Street, N. R., and Hvidsten, T. R. 2017. Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS Genet.*, 13(4): e1006402.

Manna, F., Martin, G., and Lenormand, T. 2011. Fitness landscapes: An alternative theory for the dominance of mutation. *Genetics*, 189(3): 923–937.

Neher, R. A. and Shraiman, B. I. 2009. Competition between recombination and epistasis can cause a transition from allele to genotype selection. *Proc. Natl. Acad. Sci. USA*, 106(16): 6866–6871.

Neuwirth, E. 2014. *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-2.

Nordborg, M. 2000. Linkage disequilibrium, gene trees and selfing: An ancestral recombination graph with partial self-fertilization. *Genetics*, 154(2): 923–929.

Nordborg, M. and Donnelly, P. 1997. The coalescent process with selfing. *Genetics*, 146(3): 1185–1195.

Nordborg, M., Charlesworth, B., and Charlesworth, D. 1996. Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. *Proc. R. Soc. B*, 263(1373): 1033–1039.

Ossowski, S., Schneeberger, K., Lucas-Lledo, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., Weigel, D., and Lynch, M. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, 327(5961): 92–94.

Otto, S. P. 2009. The evolutionary enigma of sex. *Am. Nat.*, 174(s1): S1–S14.

Paaby, A. B. and Rockman, M. V. 2013. The many faces of pleiotropy. *Trends Genet.*, 29(2): 66–73.

Pollak, E. 1987. On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics*, 117(2): 353–360.

Pritchard, J. K. and Di Rienzo, A. 2010. Adaptation - not by sweeps alone. *Nat. Rev. Genet.*, 11(10): 665–667.

Pritchard, J. K., Pickrell, J. K., and Coop, G. 2010. The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.*, 20(4): R208–R215.

R Core Team 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Racimo, F., Berg, J. J., and Pickrell, J. K. 2018. Detecting polygenic adaptation in admixture graphs. *Genetics*, 208(4): 1565–1584.

Ronce, O., Shaw, F. H., Rousset, F., and Shaw, R. G. 2009. Is Inbreeding Depression Lower in Maladapted Populations? A Quantitative Genetics Model. *Evolution*, 63(7): 1807–1819.

Roze, D. 2009. Diploidy, population structure, and the evolution of recombination. *Am. Nat.*, 174(S1): S79–S94.

Sachdeva, H. 2019. Effect of partial selfing and polygenic selection on establishment in a new habitat. *Evolution*, 73(9): 1729–1745.

Sachdeva, H. 2022. Reproductive isolation via polygenic local adaptation in subdivided populations: Effect of linkage disequilibria and drift. *PLoS Genet.*, 18(9): e1010297.

Salome, P. A., Bomblies, K., Fitz, J., Laitinen, R. A. E., Warthmann, N., Yant, L., and Weigel, D. 2012. The recombination landscape in *Arabidopsis thaliana* F2 populations. *Heredity*, 108(4): 447–455.

Schrider, D. R., Shanku, A. G., and Kern, A. D. 2016. Effects of linked selective sweeps on demographic inference and model selection. *Genetics*, 204(3): 1207–1223.

Sella, G. and Barton, N. H. 2019. Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annu. Rev. Genom. Hum. Genet.*

Simons, Y. B., Bullaughey, K., Hudson, R. R., and Sella, G. 2018. A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS Biol.*, 16(3): e2002985.

Sohail, M., Maier, R. M., Ganna, A., Bloemendal, A., Martin, A. R., Turchin, M. C., Chiang, C. W., Hirschhorn, J., Daly, M. J., Patterson, N., Neale, B., Mathieson, I., Reich, D., and Sunyaev, S. R. 2019. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife*, 8: e39702.

Stebbins, G. L. 1957. Self fertilization and population variability in the higher plants. *Am. Nat.*, 91(861): 337–354.

Stephan, W. 2016. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Mol. Ecol.*, 25(1): 79–88.

Stephan, W. and John, S. 2020. Polygenic adaptation in a population of finite size. *Entropy*, 22: 907.

Stetter, M. G., Thornton, K., and Ross-Ibarra, J. 2018. Genetic architecture and

selective sweeps after polygenic adaptation to distant trait optima. *PLoS Genet.*, 14(11): e1007794.

Szép, E., Sachdeva, H., and Barton, N. H. 2021. Polygenic local adaptation in metapopulations: A stochastic eco-evolutionary model. *Evolution*, 75(5): 1030–1045.

Tenaillon, O. 2014. The Utility of Fisher's Geometric Model in Evolutionary Genetics. *Annu. Rev. Ecol. Evol. Syst.*, 45(1): 179–201.

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(4): 1513–1530.

Troth, A., Puzey, J. R., Kim, R. S., Willis, J. H., and Kelly, J. K. 2018. Selective trade-offs maintain alleles underpinning complex trait variation in plants. *Science*, 361(6401): 475–478.

Tsuchimatsu, T., Kakui, H., Yamazaki, M., Marona, C., Tsutsui, H., Hedhly, A., Meng, D., Sato, Y., Städler, T., Grossniklaus, U., Kanaoka, M. M., Lenhard, M., Nordborg, M., and Shimizu, K. K. 2020. Adaptive reduction of male gamete number in the selfing plant *Arabidopsis thaliana*. *Nat. Commun.*, 11(1): 2885.

Turelli, M. 1984. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theor. Popul. Biol.*, 25(2): 138–193.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., and

Venables, B. 2020. *gplots: Various R Programming Tools for Plotting Data*. R package version 3.1.0.

Weir, B. S. and Cockerham, C. C. 1973. Mixed self and random mating at two loci. *Genet. Res.*, 21(3): 247–262.

Wickham, H. 2011. The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1): 1–29.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. 2019. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43): 1686.

Wieters, B., Steige, K. A., He, F., Koch, E. M., Ramos-Onsins, S. E., Gu, H., Guo, Y.-L., Sunyaev, S., and de Meaux, J. 2021. Polygenic adaptation of rosette growth in *Arabidopsis thaliana*. *PLoS Genet.*, 17(1): e1008748.

Wilke, C. O. 2020. *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package version 1.1.1.

Winn, A. A., Elle, E., Kalisz, S., Cheptou, P.-O., Eckert, C. G., Goodwillie, C., Johnston, M. O., Moeller, D. A., Ree, R. H., Sargent, R. D., and Vallejo-Marín, M. 2011. Analysis of inbreeding depression in mixed-mating plants provides

evidence for selective interference and stable mixed mating. *Evolution*, 65(12): 3339–3359.

Wray, N. R., Goddard, M. E., and Visscher, P. M. 2007. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.*, 17(10): 1520–1528.

Wright, S. 1921. Systems of mating. v. general considerations. *Genetics*, 6(2): 167–178.

Wright, S. I., Kalisz, S., and Slotte, T. 2013. Evolutionary consequences of self-fertilization in plants. *Proc. R. Soc. B*, 280(1760): 20130133.

Zan, Y. and Carlborg, Ö. 2018. A polygenic genetic architecture of flowering time in the worldwide *Arabidopsis thaliana* population. *Mol. Biol. Evol.*, 36(1): 141–154.