Dissecting genetic correlation through recombinant perturbations: the role of developmental bias

Haoran Cai[1], Kerry Geiler-Samerotte[2,3], David L. Des Marais[1]

[1]Department of Civil and Environmental Engineering, MIT.
15 Vassar St., Cambridge, MA, 02139 USA

[2]Center for Mechanisms of Evolution, Biodesign Institute, Arizona State University, USA

[3]School of Life Sciences, Arizona State University, USA

Abstract

Despite the tremendous diversity and complexity of life forms, there are certain forms of life that are never observed. Organisms like angels might not emerge because of developmental constraint, or because they have low fitness in any environment yet available on Earth. Given that both developmental constraint and selection may create similar phenotypes, it is difficult to distinguish between the two causes of evolutionary stasis among related taxa. For example, remarkably invariant traits are observed spanning million years, such as wing shape in *Drosophila* wherein qualitative differences are rare within genera. We thus ask whether the absence of combinations of traits, indicated by genetic correlation, reflects developmental bias limiting the possibility of change. However, much confusion and controversy remain over definitions of developmental bias and quantifying it is challenging. We present an approach aiming to estimate developmental bias by leveraging recombination in genetic mapping populations. We reason that information rendered by such mild perturbations captures inherent interdependencies between traits – developmental bias. Through empirical analyses, we find that our developmental bias metric is a strong indicator of genetic correlation stability across conditions. Our framework presents a feasible way to quantify developmental bias between traits and opens up the possibility to dissect patterns of genetic correlation.

Keywords: Pleiotropy; Developmental constraint; Genetic correlation; Developmental bias

Introduction

> *When the morphology of a species remains virtually unchanged for millions of years, we would like to know whether this reflects developmental constraints limiting the possibility of change or, conversely, the maintenance of uniformity by stabilizing selection.* — *Maynard Smith* et al. *1985*

Genetic correlation represents an important class of evolutionary constraints (Maynard Smith *et al.*, 1985; Clark, 1987), affecting future evolutionary trajectories. Yet, genetic correlations are themselves evolvable (Doroszuk *et al.*, 2008; Dugand *et al.*, 2021; Delph *et al.*, 2011; Conner, 2002; Uller *et al.*, 2018) and reflect both the past selection of trait combinations and, in some cases, developmental constraint (Dugand *et al.*, 2021; Arnold, 1992). Natural selection may favor certain combinations of traits and thereby lead to genetic correlation via pleiotropy and linkage equilibruim. Pleiotropy and LD may then further inhibit traits from evolving independently towards a theoretical phenotypic optimum (Schluter, 1996). On the other hand, genetic correlation can be shaped by bias due to intrinsic attributes of the organism, energy, or the laws of physics, relative to the assumption of isotropic variation. This latter concept has been decribed as developmental bias (Maynard Smith *et al.*, 1985; Arnold, 1992; Cheverud, 1984), which may account for the observation that perturbations, such as mutation or environmental variation, to biological systems will tend to produce some phenotypic outcomes more readily (Uller *et al.*, 2018; Waddington, 1957). In spite of the numerous studies that address genetic correlation as an evolutionary constraint, much confusion and controversy remains over definitions of different types of constraint, the mechanism(s) causing constraint, and the relative importance of different mechanisms in shaping evolutionary trajectories (Muir *et al.*,

2022; Conner *et al.*, 2011).

The theoretical underpinnings for genetic covariance as an evolutionary constraint are well-developed (Lande, 1979; Lande & Arnold, 1983). Genetic covariance specifically describes trait covariance due to pleiotropic alleles, where a single locus has effects on two traits, or due to linkage disequilibrium of two loci, each of which affects a single trait but are physically so close that these two traits are strongly associated in populations (Lande, 1980; Lynch *et al.*, 1998; Falconer *et al.*, 1996; Conner *et al.*, 2004). The genetic information summarized by genetic covariance is connected to evolutionary processes in complex ways. For example, evolution toward a phenotypic optimum for two traits may be restricted if selection favors two traits antagonistically but the traits are positively correlated. That is, adaptive evolution can be limited if the joint vector of selection is antagonistic to the trait correlations. In some cases, such evolutionary constraint may persist over long time scales (McGlothlin *et al.*, 2018; Opedal *et al.*, 2023).

Straightforward applications of evolutionary quantitative genetic theory dealing with the joint evolution of a pair of traits generally assumes an invariant genetic covariance structure (**G** matrix) over the time frame of interest. However, the stability of genetic covariance structure and how it evolves remains unclear and contentious (Turelli, 1988; Bürger & Lande, 1994; Arnold *et al.*, 2008; Steppan *et al.*, 2002; Milocco & Salazar-Ciudad, 2022; Loeschcke, 1987; Barton & Turelli, 1989). Empirical studies of the evolution of genetic covariance structure have found mixed results on whether genetic covariance changes rapidly or slowly. Some comparisons of **G** matrices between natural populations found no evidence of change in **G** (Delahaie *et al.*, 2017; Arnold *et al.*, 2008; Hangartner *et al.*, 2020; Henry & Stinchcombe,

2023), while others have found changes in genetic covariance in only a few generations, across populations, in response to selection, or across environmental conditions (Chakrabarty & Schielzeth, 2020; Milocco & Salazar-Ciudad, 2022; Eroukhmanoff & Svensson, 2011; Walter *et al.*, 2018; Wood & Brodie III, 2015; Henry & Stinchcombe, 2022; Hudson *et al.*, 2022; Scoville *et al.*, 2009). We might also predict that the genetic covariance among some suites of traits is stable, while it is unstable for others (Jones *et al.*, 2003). Generally, it is largely unknown what determines the stability of genetic covariance (Wood & Brodie III, 2015) and this uncertainty challenges our ability to predict the outcome of selection.

The persistence of correlational constraint and whether genetic correlation is a good predictor of long-term evolutionary divergence ultimately hinge on our understandings of the underlying mechanisms that cause genetic correlation (Loeschcke, 1987; Conner *et al.*, 2011, 2004). For example, genetic correlation due to pleiotropy or tight linkage are much more likely to cause constraint than those caused by linkage disequilibrium between loosely linked loci (Conner *et al.*, 2011, 2004; Conner, 2002). Correlations due to pleiotropy or tight linkage may persist in the absence of selection, while correlations caused by linkage disequilibrium can be changed quickly by recombination and selection (Conner, 2002; Conner *et al.*, 2004, 2011). We here reason that genetic correlation due to developmental bias is more likely to impose constraint on evolutionary change and may be more stable than other factors. Insight into the role of developmental bias may reveal why genetic correlations between some traits are more constant over a long period as compared to other pairs of traits and why, in some cases, genetic constraints can be readily degraded by natural or artificial selection. However, formally discriminating between 'developmental' bias and other mechanisms of genetic correlation is

notoriously difficult (Maynard Smith *et al.*, 1985).

Here, we provide an approximate measure of developmental bias by exploiting recombinant genetic perturbations. We distinguish between two kinds of pleiotropy: horizontal pleiotropy, which describes a locus that has an effect on two traits, where such pleiotropic effect deviates from the rest of the loci across a genome (Fig. 1 right). Conversely, developmental bias between traits describes the observation of consistent pleiotropic effect of loci throughout the genome on a given trait pair (Fig. 1 left). We use this consistency of pleiotropic effect throughout the genome to indicate developmental bias r_D. We reason that if two traits are correlated because of developmental or physiological constraint, these two traits should be correlated regardless of which specific variant causes the effect. Intuitively, genetic correlation shaped by developmental bias may last longer or be more stable than those trait pairs shaped by several individual horizontal pleiotropic loci. Besides, evidence showed that large, pleiotropic effect mutations may undergo strong selection (Denver *et al.*, 2005; McGuigan *et al.*, 2014), which suggest that horizontal pleiotropic loci, instead of the developmental bias, is more likely to be shaped by selection

Our primary goal in the present work is to dissect genetic correlations to account for the component of developmental bias and horizontal pleiotropy. Specifically, our method demonstrates the possibility of identifying horizontal pleiotropy loci without knowing within-line correlation across clonal cells, which requires considerable experimental work and is infeasible for many organisms (Geiler-Samerotte *et al.*, 2020). The second goal of our study is to ask whether such decomposition can extend our understanding of pleiotropy, genetic correlation, and evolutionary constraint. We ask whether a genetic constraint that arises principally from a

developmental bias is more persistent across conditions. We may then use this principle to predict the divergence of genetic correlation over a long-term period. Both numerical simulations and empirical datasets were analyzed to demonstrate how identifying the role of developmental bias in trait correlation could distinguish among trait pairs, despite their similar genetic correlation. We thus argue that readily accessible QTL mapping data using recombinant population are under-utilized and can help to deepen our understandings of mechanisms causing genetic correlation.

Results

We here define developmental bias as phenotypic bias that triggered by non-directed perturbations which have not been encountered in the recent evolutionary past. We quantify developmental bias by examining the additive effect of loci for trait pairs throughout the genome. During a genetic association study, each genetic marker is assigned an odds likelihood ratio along with a effect size for the trait of interest. Rather than identifying statistically significant loci in such conventional genetic association studies, the essential idea, here, is to examine the consistency of pleiotropic effects across genetic backgrounds. We define a locus with effects that deviate from the overall bivariate trend throughout the genome as a horizontal pleiotropic (HP) locus (Fig. 1b). We diagnose developmental bias (r_D) as the consistency of pleiotropy across genetic backgrounds excluding HP loci. In a mapping population, allele substitutions at each locus represent non-directed (i.e., random) perturbations of varying directions and magnitudes. We reason that if the effect size of these perturbations on two traits are highly correlated (i.e., the effect size correlation across all background variants except HP

loci), the developmental bias between the two traits is likely to be strong. To better illustrate the framework we propose, the bivariate effect size distributions under two scenarios are shown (Fig. 1a, b), where both pairs of traits are associated with a major pleiotropic locus. We define a locus with a major phenotypic effect that deviates from the overall trend of other loci throughout the genome as a horizontal pleiotropic locus, and the consistency of pleiotropy (effect size correlation) with the rest of the loci, excluding those HP loci, as r_D.

Simulation demonstrating relationships between r_D and r_G

To examine how the estimated developmental bias r_D – characterized by the effect size correlation among loci – relates to genetic correlation (r_G), we first simulated two thousand trait pairs for a given simulated population with 500 individual genotypes. For each trait pair, genetic architecture with 226 loci was generated, with additive effect sizes sampled from a multivariate exponential distribution. The genetic values are obtained by multiplying the genotypes with allelic effect sizes, assuming no epistasis and no linkage disequilibruim. r_G is calculated by correlating the genetic values between two traits following standard protocols (Falconer *et al.*, 1996). We calculated r_D and corresponding r_G for each pair of traits. Notably, genetic correlation r_G is a correlation across a population of individuals while r_D is a correlation across a population of loci in a genome.

We performed these simulations under six different scenarios to examine the effects of HP, LD, polygenicity, and effect size on the correlation structure (Fig. 2). Assuming no horizontal pleiotropy (HP) and linkage disequilibrium (LD), we expect r_D and r_G to be equal. As expected, without accounting for HP and LD, r_D strongly correlates with r_G regardless of

the genetic architecture regimes (high versus low kurtosis)(Fig. 2b,c). Notice that the points representing the effect sizes of an allele on two traits are not exactly on the unity line because allele frequency influences r_G but not r_D. Next, we repeated our simulations under conditions with HP or LD to understand how these forces would affect the correlations (Fig. 2d,e). Under HP scenario, randomly selected n ($0 < n < 10$) SNPs are forced to have an HP effect (either concordant to or antagonistic with the rest of loci). The genetic correlation r_G is not perfectly correlated with r_D under scenarios with HP, especially when the kurtosis of the effect size distribution is high. In an extreme case, a single large-effect locus can dominate the trait correlation despite the low r_D (Fig. S1). Collectively, these observations suggest that r_D and HP loci are two components of r_G, and that a single large-effect HP locus can drive r_G without overall consistency of pleiotropy throughout the genome. Note that empirical examples of large effect variants with putative causal roles in driving trait covariation are common (Hall *et al.*, 2006; Wessinger *et al.*, 2014; Bradshaw Jr *et al.*, 1998; Saltz *et al.*, 2017; Hine *et al.*, 2022; Lee *et al.*, 2014; Stinchcombe *et al.*, 2009).

To understand how LD affects the relationship between r_G and r_D, we also performed simulations using actual recombinant genotypes from a yeast mapping population (Geiler-Samerotte *et al.*, 2020) (Fig. 2f,g). Similarly to our simulations above, we sampled the effect size for each SNP on each trait from bivariate Laplace distribution with the same γ and then propagated the effect size for a given SNP by "contaminating" its effect size according to the effect sizes of the SNPs in LD with it. (This procedure only accounts for weak linkage, See Supplementary Note 1.) In these simulated outcome, LD appears to affect r_G with a given effect size correlation even for cases in which the genetic architecture is highly polygenic (Fig. 2f). These results

imply that LD does not always strengthen r_G; it could also weaken r_G when, for example the effect of two loci in LD are antagonistic with the overall trend of pleiotropy across the genome. To summarize the numerical simulations, LD, HP, and r_D together shape r_G. The effect of LD and HP can become stronger under a more 'Mendelian' genetic architecture with lower polygenicity.

Identifying horizontal pleiotropic loci and delineating developmental bias for yeast morphological traits

We next applied our approach a yeast morphology dataset, where 374 recombinant strains of yeast cells were imaged for, on average, 800 fixed, stained cells per strain, using high-throughput microscopy (Geiler-Samerotte *et al.*, 2020). In total, measurements of 167 morpholgical traits were acquired. We expect that the patterns in this large dataset offer a general picture of how r_D, HP, and LD affect effect size correlations and how r_D contributes to our understanding of trait relationship in complement to r_G.

As described above, we define developmental bias r_D as the effect size correlation for a subset of variants where outliers are removed. When calculating effect size correlation for outlier-corrected subset of variants (r_D), we first calculate the correlation by individual level product (Lea *et al.*, 2019) for each trait pair across each locus. Outliers are then identified as the product falling outside 1.5 times the interquartile range above the upper quartile and below the lower quartile of the distribution. Since LD can also potentially affect the correlation of effect size, we conducted LD pruning to subset the variants to $r < 0.5$ under LD corrected setting (See Materials and Methods). Therefore, in total, we present the effect size correlation

against r_G in three settings: Default (using all genotyped variants), LD corrected, and outlier corrected.

Fig. 3a presents the distribution of effect size correlations using all genotyped variants, only LD pruned variants, or outlier-corrected variants (r_D). LD does not exert effect on the patterns of effect size correlation in this dataset (3a, two-sample Kolmogorov-Smirnov test, D = 0.017006, p-value = 0.3879), but horizontal pleiotropic loci, which we identified as outliers, appear to strengthen the effect size correlation (Fig. 3a,b). In principle, an outlier can either weaken or strengthen the effect size correlation. However, our results suggest a bias towards concordant effects between outliers and other loci, given that the distribution under the outlier-corrected setting (r_D) has smaller variance (Fig. 3a) and effect size correlation is generally weaker under the outlier corrected setting (Fig. 3b). Notably, points that deviate more from the unity line ($y = x$) may represent trait pairs which are strongly affected by horizontal pleiotropy.

To further investigate horizontal pleiotropy (HP), we identified those trait pairs significantly affected by the outliers (i.e., yellow dots in Fig. 3b). Outlier loci for these trait pairs are likely indicative of horizontal pleiotropy. Indeed, we confirmed that our method identifies two loci (L15.9 and L13.7, See Table 1) that presented the strongest evidence of horizontal pleiotropy in an earlier study using stronger genetic correlation than within-line environmental correltion as an indicator (Geiler-Samerotte *et al.*, 2020). Conversely, we found trait pairs with extremely high effect size correlations (Fig. 3a, b); an example effect size distribution of a trait pair with exceptionally high effect correlation – possibly reflecting a strong developmental bias between two traits – is shown in Fig. S2. In summary, we show that horizontally pleiotropic loci may indeed affect r_G (unless a exceptionally strong developmental bias is present) and that our

method can be used to identify horizontal pleiotropic loci and further delineate r_G.

To assess how the effect size correlation including estimated developmental bias r_D (effect size correlation in an outlier-corrected setting) relates to the genetic correlation r_G, we calculated the effect size correlations between pairwise traits and plotted them against r_G for each trait pair (Fig. 3c,d,e). Under all settings, we find no cases where trait pairs with no r_G exhibit a strong effect size correlation. Qualitatively similar results are observed for an additional *Brassica* dataset with 11 floral, vegetative, and phenology traits (Supplementary Note 2, Fig. S5, Fig. S6).

Notably, without LD correction and removing horizontal pleiotropic loci as outliers, the results here using recombinant genetic perturbation (Fig. 3c) are similar to a study using mutational accumulation line to examine the contribution of mutation to genetic correlation, where mutational correlations between traits were found to be stronger than genetic correlations (Dugand *et al.*, 2021). Only after removing horizontal pleiotropic loci as outliers, the scatter plot pattern (Fig. 3e) appears aligned with simulation outcomes without HP (Fig. 2, scenarios without HP), where dots spread evenly around the unity line. Prominently, to show why it is important to calculate r_D, examples of the additive effect for a pair of traits are shown in Fig. 4, where these two pairs of traits have a similar and moderately high genetic correlation r_G but contrasting levels of developmental bias r_D. These two examples, empirically, echoes hypothetical scenarios in Fig. 1a and b, implying that estimated developmental bias derived from our method can unveil extra information between traits despite similar level of r_G .

r_D predicts the stability of r_G following environmental perturbation

Genetic correlations between traits may alter the evolutionary trajectory of either trait (Schluter, 1996). Predicting the trajectory of trait evolution therefore can depend upon the stability of genetic correlations (Jones *et al.*, 2003). To test whether estimated r_D under unperturbed condition provides information of r_G stability upon perturbations, we used a second yeast dataset, which comprises changes of r_G across three "environments." Here, the environmental conditions are three concentrations of geldanamycin (GdA), a small-molecule inhibitor that binds the ATP-binding site of the chaperonin Hsp90, rendering it unable to perform its cellular function. We plotted absolute r_D with changes of genetic correlation (r_G) for each pair of traits at the three drug concentrations (Fig. 5). The results show that, as r_D becomes greater, the changes of r_G become smaller.

Since r_G is highly correlated with r_D, to formally test whether under a given r_G, r_D is informative in determining changes of r_G upon environmental perturbations, we conducted multivariable linear regression ($\Delta r_G \sim r_G + r_D$, all variables are transformed to absolute value). The regression results (Fig. S3a) demonstrate that conditioning on r_G of a trait pair, r_D significantly negatively correlates with the changes of r_G across drug concentrations. In other words, given a set of yeast morphology trait pairs with the same level of r_G, the changes of magnitude of r_G would be smaller for trait pairs with larger r_D, on average. Furthermore, we found that this effect of r_D is strongest under mild perturbation but becomes diluted as the drug concentration becomes higher and, presumably, more stressful for the cells (Geiler-Samerotte *et al.*, 2016). To account for the effect of collinearity between r_D and r_G (PCC = 0.939) on regression outcomes, we also report, here, null model simulated results(Fig. S3b

and Fig. S4). The null model expectation Overall, these results together indicate that the estimated r_D under the normal condition may indeed predict the stability of r_G following environmental perturbtions. This is expected because when r_G are driven by developmental biases, we expect them to be more stable, as opposed to when they are driven by HP.

Discussions

It has long been recognized that developmental integration is one cause of multivariate genetic constraint (Klingenberg, 2005; Pigliucci & Preston, 2004). On the other hand, genetic constraint can also reflect correlational selection. However, dissecting the underlying mechanism(s) causing genetic correlation is challenging. Here, we exploited a hidden source of data that has been overlooked to quantify the contribution of developmental component in creating genetic correlation. Assessing consistency of pleiotropy across many genomic loci provides a possible framework to explore the mechanisms of genetic correlation. The central messages from our analyses are three-fold.

First, developmental bias estimated from recombinant genetic perturbations provides an indicator of genetic correlation stability following environmental perturbations. Our stability analyses in empirical datasets (Fig. 5, Fig. S5c, and Fig. S3) suggest that the higher a developmental bias, the more likely a genetic correlation between two traits remains stable across environmental conditions. In other words, higher developmental bias leads to smaller response of genetic correlation to environmental changes.

Second, Mendelian genetic architecture for a given trait pair will increase the chance of devi-

ation of genetic correlation from developmental bias (Fig. 2). Under such scenario, genetic correlation as a summary statistic can not fully reflect the complex genetic architecture underlying a genetic correlation. In fact, evidence of discrepancies of effect between genetic background and major loci abound (Albert *et al.*, 2008; Hall *et al.*, 2006; Scoville *et al.*, 2009; Stinchcombe *et al.*, 2009). For example, in *Mimulus*, a major QTL contributes a negative covariance between stigma–anther separation and pollen viability, which is antagonistic to the overall positive genetic covariance between these two traits (Scoville *et al.*, 2009). Furthermore, previous work suggest that we might expect to see more changes of **G**-matrix during evolution if traits have an oligogenic genetic basis rather than aligning the infinitesimal model (Barton & Turelli, 1987, 1989; Lande, 1979). For example, Lande (Lande, 1979) emphasized that trait means typically change much more rapidly than trait (co)variances. Yet, changes of (co)variances can be quite rapid if there are underlying loci with large contributions to (co)variation. Similarly, in our present work, we show that under a infinitesimal model, genetic correlation mainly arises from developmental bias (Fig. 2), which as our stability tests suggest might also be more stable across conditions (Fig. 5).

Third, our method allows us to identify horizontal pleiotropic loci without measuring phenotype across clonal cells. There is a long standing interest in identifying horizontal pleiotropy in nature (Verbanck *et al.*, 2018; Jordan *et al.*, 2019; Bowden *et al.*, 2018). One motivation for doing so is that evolutionary theory predicts that natural selection should limit horizontal pleiotropy because as the number of traits that a mutation influences increases, the probability of that mutation having a positive fitness effect decreases (Zhang & Wagner, 2013; Orr, 2000; Pavlicev & Wagner, 2012; McGuigan *et al.*, 2014). However, identifying cases of

horizontal pleiotropy is difficult because genetic correlations do not always indicate horizontal pleiotropy. By discovering a way to disentangle the portion of genetic correlation caused by developmental bias, we have also discovered a novel way to identify candidate loci that act via horizontal pleiotropy. Our method of identifying horizontal pleiotropy can be broadly useful because, unlike previous methods (Geiler-Samerotte *et al.*, 2020), it does not require measuring the trait correlations that are present across clonal cells. Thus, while previous methods are mainly useful for organisms that propagated clonally, e.g., microbes, our method can be applied more broadly.

Can recombinant mapping population characterize M matrix?

Numerious previous studies have used mutational accumulation lines to derive mutational matrix (**M** matrix) as a means to understand the influence of mutations on shaping genetic correlation (Dugand *et al.*, 2021; Houle *et al.*, 2017). Dugand *et al.* (2021) discovered a significant similarity between **G** and **M** matrix, suggesting that mutation directly shape **G**. On the other hand, mutational correlations using mutational accumulation lines consistently surpass genetic correlations (Dugand *et al.*, 2021), which is aligned with our findings (Fig. 3c), where the recombinant mutational correlations, indicated by the effect size correlation across all loci, are stronger than the genetic correlation r_G. This naturally raises several questions related to mutational accumulation line, recombinant mutation, and developmental bias: Firstly, to what extent do a recombinant genetic perturbation reflect the effect of mutation in a mutational accumulation experiment? There are nowadays increasingly accessible resources available for recombinant mapping populations, such as the recently developed multiparent panels and advanced intercross lines (Kover *et al.*, 2009; Gage *et al.*, 2020). These resources provide a

promising avenue to investigate the respective roles of mutation and selection. Secondly, to what extent do mutational perturbations reflect the strength of developmental bias? Are mild perturbations more representative of developmental bias that shapes the divergence of traits over millions of years?

The debate regarding the extent of pleiotropy

Pleiotropy describes the phenomenon in which a gene or a mutation affects more than one phenotypic trait. The concept and nuance of pleiotropy has had a prominent role and broad implications on genetics, evolution, and medicine (Klingenberg, 2008; Stearns, 2010; Promislow, 2004; Williams, 2001; Barton, 1990; He & Zhang, 2006; Otto, 2004; Wagner & Zhang, 2011; Des Marais & Juenger, 2010). Conceptually, many possible scenarios can result in a pleiotropic effect, including mediated pleiotropy (i.e., vertical pleiotropy), horizontal pleiotropy, and other spurious pleiotropy such as linkage, (Wagner & Zhang, 2011; Solovieff *et al.*, 2013). One major debate on pleiotropy is what are the extent of pleiotropy. That is, we lack consensus about how pleiotropic the natural system is (Paaby & Rockman, 2013; Zhang & Wagner, 2013). A key challenge is whether the effect of a single locus on correlated traits can be counted as pleiotropic effect, for instance, as pointed out by Wagner & Zhang 2011; e.g., are the depth and the width of a bird beak two characters? Thus, ignoring trait correlations may bias the estimation of pleiotropy. One possible solution is to consider the effective number of traits by looking at the eigenvalue variance of the phenotypic correlation matrix (Wagner & Zhang, 2011; Pavlicev *et al.*, 2009; Wagner *et al.*, 2008): The more dispersed the eigenvalues, the more interdependency of the traits. However, this approach likely biases the interdependency of the traits, especially in the presence of major pleiotropic effect loci. For example, as an extreme

case, even a single pleiotropic locus alone can drive the trait correlation in spite of the low consistency of pleiotropy (Fig. S1 and also see (Agrawal *et al.*, 2001)). Hence, in this case, the effective number of traits calculated via the phenotypic correlation matrix will be overestimated simply because there is a major effective pleiotropic locus – this does not necessarily mean two traits are inherently interrelated. Similarly, the bias is present if there is a major antagonistic loci against overall correlation of effect size. Instead, our analyses demonstrated that the consistancy of effect sizes may provide a more appropriate way to measure inherent trait correlation and hence effective trait dimensions.

Materials and Methods

Sophisticated tools in the field of quantitative genetics have been developed to identify genetic loci which statistically explain phenotypic variance in quantitative traits to regions of chromosomes, so-called quantitative trait loci (QTLs). One of the fundamental metrics of quantitative genetics is the additive effect of a QTL, which represents the change in the average phenotype produced by substituting one allele for another (Lynch *et al.*, 1998; Falconer *et al.*, 1996). To better illustrate what we could exploit through the additive effect distribution, bivariate effect size distribution under two scenarios are shown (Fig. 1a,b), where both two pairs of traits have a major pleiotropic locus. In contrast, the consistency of pleiotropic effect throughout genome is different. This illustrative example may be extreme, but it implies that only analyzing the summary statistics such as genetic correlation or statistically significant loci in a genetic association study may lose information behind the genetic architecture. Such hidden information could be valuable when assessing the strength of developmental bias r_D: If two traits

are correlated because of developmental or physiological constraint, these two traits should be correlated regardless of which specific variant is causing the effect. i.e., there is consistency of the pleiotropic effect across genetic background (Fig. 1a). On the other hand, if two traits are genetically correlated simply because of several major pleiotropic loci for a given population, those small loci can have inconsistent effect between traits (Fig. 1b). We term such consistency of pleiotropy as developmental bias r_D and those loci with effect deviated from overall trend throughout the genome as horizontal pleiotropic (HP) loci. Note that the way we quantify developmental bias is similar to the definition of vertical pleiotropy or mediated pleiotropy Geiler-Samerotte *et al.* (2020). Indeed, the high consistency of pleiotropic effect implies vertical or mediated pleiotropic nature of loci. Yet, we here define the developmental bias as a trait-level metric, whereas the vertical or mediated pleiotropy is most often used to the effects of variants. In vertical pleiotropy, the traits themselves are biologically related, such that a variant's effect on trait A inevitably causes the effect on trait B. Likewise, horizontal pleiotropy is defined as a variant/mutation causing the effect for two traits that are otherwise independent. Another distinction between developmental bias and vertical pleiotropy is that vertical pleiotropy frequently refers to a part of causal cascade, as exemplified by low-density lipoprotein (LDL) cholesterol levels causing the risk of heart disease (Geiler-Samerotte *et al.*, 2020). Developmental bias, on the other hand, depicts the correlational structure among traits since many traits (e.g., morphological traits) are not necessarily exhibited direct causal relationships. We thus apply the term vertical pleiotropy to variants that share the effect for inherently related traits without considering the causal direction.

Numerical simulations

To simulate multiple pairs of traits within a population, first genotypes were simulated through the function *simulateGenotypes* in PhenotypeSimulator (Meyer & Birney, 2018) with 226 SNPs (mimic actual number of loci in an empirical dataset; Geiler-Samerotte *et al.* (2020)) and 500 individuals, where the allele frequencies are either sampled from 0.05, 0.1, 0.2, and 0.5 or constant value 0.3 for a mapping population. For each pair of traits, the additive effect for each SNP is sampled from a bivariate exponential distribution (bivariate Laplace distribution) with $\mu = (0, 0)$ and $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. ρ is drawn from the uniform distribution $(-1, 1)$. A shape parameter γ determines the distribution, where a smaller γ represents genetic architecture approximating one or a small number of large-effect loci (Mendelian genetic architecture, high kurtosis for effect size distributions) while a larger γ trends towards a polygenic infinitesimal model (low kurtosis for effect size distributions). We used γ of 1.0 and 0.5 in Fig. 2 left and right, respectively. Under horizontal pleiotropy scenario, randomly selected n $(0 < n < 10)$ SNPs are forced to have horizontal pleiotropic effect (either concordant to or antagonistic with the rest of loci). The genetic correlation, r_G between traits M and N, was calculated as the Pearson correlation $\rho_{(X\beta_{[M]}^{\intercal}, X\beta_{[N]}^{\intercal})}$, where $\beta_{[M]}$ and $\beta_{[N]}$ represents the effect size for trait M and N across genome, and $X\beta_{[M]}^{\intercal}$ and $X\beta_{[N]}^{\intercal}$ represent the genotypic values of trait M and trait N, respectively. The developmental bias, r_D, is calculated as the Pearson correlation coeffecient of effect size for trait M and N, where summations are taken over all loci except those assigned as horizontal pleiotropic SNPs (n loci):

$$D_{\beta[M,N]} = \frac{cov(\beta_{[M]}, \beta_{[N]})}{\sigma_{\beta_{[M]}}\sigma_{\beta_{[N]}}} = \frac{n\sum\beta_{i[M]}\beta_{i[N]} - \sum\beta_{i[M]}\sum\beta_{i[N]}}{\sqrt{n\sum\beta_{i[M]}^2 - (\sum\beta_{i[M]})^2}\sqrt{n\sum\beta_{i[N]}^2 - (\sum\beta_{i[N]})^2}} \quad (1)$$

To account for linkage disequilibrium (LD), simulating trait pairs with actual genotype information, the additive effect size of 226 SNPs are simulate similarly as above except that there are no horizontal pleiotropic loci. The effect size is then propagated through the LD block defined by $r > 0.5$.

Dataset retrievals and genetic correlations r_G

Two empirical datasets were used. The first dataset comprises single cell morphology data for budding yeast *Saccharomyces cerevisiae*, where for each of 374 recombinant strains of yeast cells about 800 fixed, stained cells were imaged using high-throughput microscopy (Geiler-Samerotte *et al.*, 2020). 167 morphological features were estimated, including these representative examples: cell size, bud size, bud angle. The original dataset contains genetic (between-strain) and environment (within-strain) correlation, using a multilevel correlation partitioning method (Bliese, 2013). The authors found that using this approach to estimate correlations has similar results as compared to a linear mixed model and variance component analysis. The second dataset contains phenological, floral and vegetative traits for a recombinant inbred population of *B. rapa L.* created from a cross between yellow sarson, R500, and the rapid cycling IMB211 inbred lines (Brock *et al.*, 2010). The QTL mapping was conducted with 223 markers in 131 individuals (field condition) and 132 individuals (greenhouse condition). Eleven phenotypes were included, here, as we excluded branch length in the field and leaf width in the greenhouse, which were not measured in both conditions. When calculating

r_D, we use either all genome-wide variants or an LD-pruned subset of variants. During LD pruning, we removed within-choromosome QTLs to $r < 0.5$ for both datasets.

Horizontal pleiotropic loci identification and empirical calculation of r_D

If we ignore dominance, epistasis, and linkage disequilibrium, and assume two alleles per QTL, the covariance components of a **G** matrix can be written as (Kelly, 2009):

$$C_{\alpha[M,N]} = \sum_i 2q_i(1-q_i)\alpha_{i[M]}\alpha_{i[N]} \tag{2}$$

where $C_{\alpha[M,N]}$ is the additive genetic covariance between trait M and trait N. q_i is the frequency of first allele at loci i within a given population; $\alpha_{i[M]}$ and $\alpha_{i[N]}$ are the additive effects of that allele on trait M and N, respectively; summations are taken over all loci. Accordingly, a large effect QTL for trait M (high $\alpha_{i[M]}$) can make a minor contribution to the genetic covariance structure if allele frequency q_i is small.

We developed an approach to evaluate the horizontal pleiotropy and calculate r_D empirically. In brief, our method has three components: (a) detection of horizontal pleiotropic loci; (b) calculating r_D, via outlier removal correcting for horitonal pleiotropy; (c) testing pairs with significant difference after outlier removal, identified as horizontal trait pairs.

First, we use the following procedures to exclude horizontal pleiotropic loci: first calculate the normalized, demeaned, and element-wise product of outcome for each locus (Steiger, 1980):

$$\frac{(\alpha_{i[M]} - \overline{\alpha_{[M]}})(\alpha_{i[N]} - \overline{\alpha_{[N]}})}{\sigma_{\alpha_{i[M]}}\sigma_{\alpha_{i[N]}}} \tag{3}$$

where $\overline{\alpha_{[M]}}$ and $\overline{\alpha_{[N]}}$ are the mean genome-wide additive effect size for trait M and N, respectively, and $\sigma_{\alpha_{i[M]}}$ and $\sigma_{\alpha_{i[N]}}$ are standard deviations of additive effect size for trait M and N, respectively. The horizontal pleiotropic loci are defined as loci with the product falling outside 1.5 times the interquartile range above the upper quartile and below the lower quartile. The Pearson correlation coefficient is equal to the above average element-wise product of two measured traits. In other words, the outliers of these element-wise product represent the outliers when calculating the correlation, substantially deviating from the overall trend of bivariate effect size distribution.

Second, developmental bias (r_D) is calculated as the Pearson correlation coefficient among the rest of loci written by:

$$r_{D_{\alpha[M,N]}} = \frac{cov(\alpha_{[M]}, \alpha_{[N]})}{\sigma_{\alpha_{[M]}}\sigma_{\alpha_{[N]}}} = \frac{n\sum_i \alpha_{i[M]}\alpha_{i[N]} - \sum \alpha_{i[M]} \sum \alpha_{i[N]}}{\sqrt{n\sum \alpha_{i[M]}^2 - (\sum \alpha_{i[M]})^2}\sqrt{n\sum \alpha_{i[N]}^2 - (\sum \alpha_{i[N]})^2}} \tag{4}$$

The additive effect size α estimated empirically from inbred line crosses and experimental mapping populations is calculated as: for the trait M, $\alpha = \frac{\overline{M_{AA}} - \overline{M_{BB}}}{2}$ (Falconer *et al.*, 1996), where A and B are two alleles of a locus. The summations of equation (4) are taken over across all loci (with equal probabilities) excluding major horizontal pleiotropic loci.

Finally, to test the significant difference of effect size correlation before and after outlier removal, we use a cutoff of 1% FDR for all pairs of traits included in the yeast dataset. Those

trait pairs significantly deviating from the mean are identified as the horizontal trait pair, indicating that horizontal pleiotropic loci may contribute to the genetic correlation r_G.

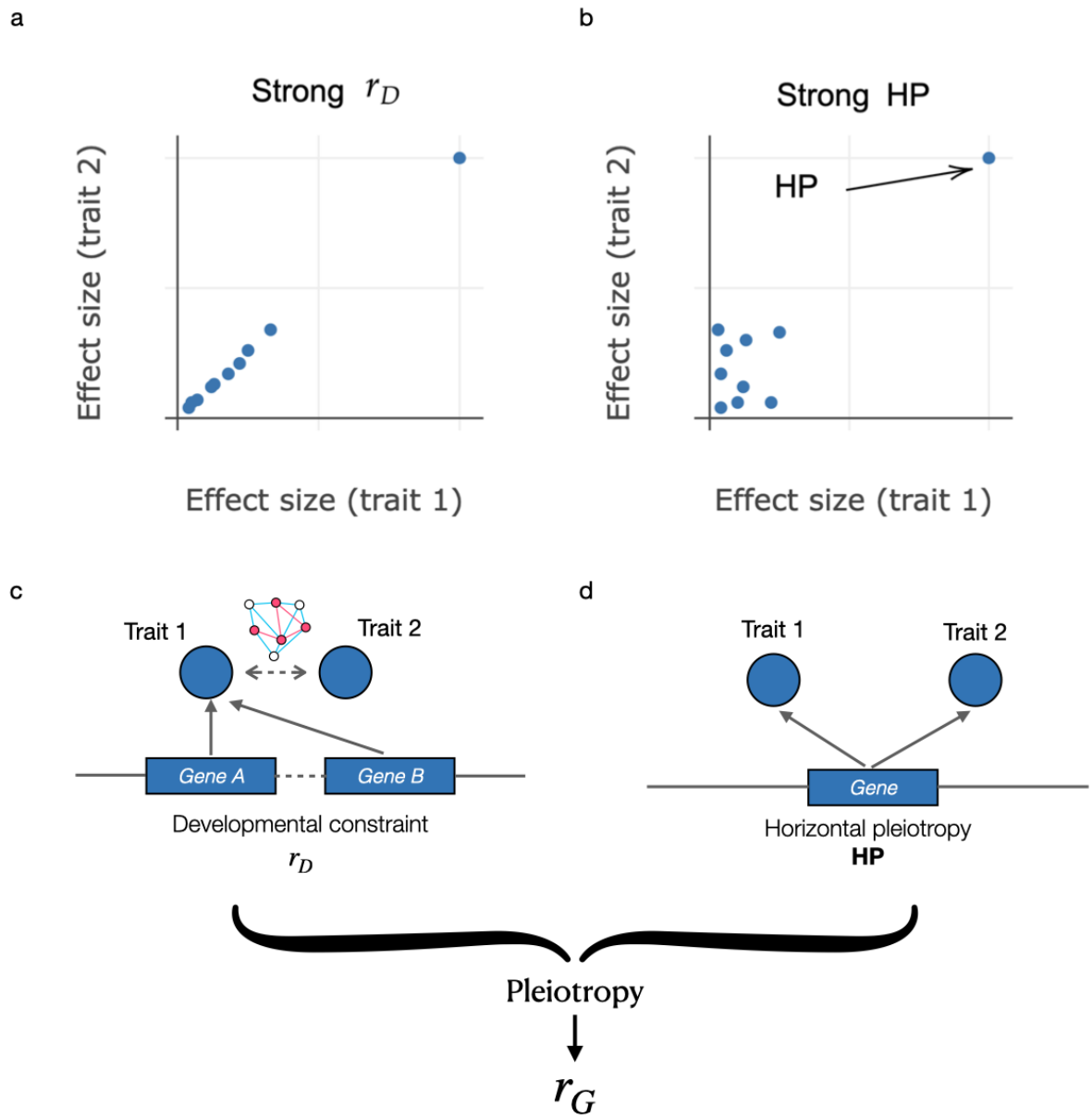

Figure 1: Conceptual framework for distinguishing between developmental bias and horizontal pleiotropy as drivers of the genetic correlation between two traits. **a.** and **b.** Bi-plots showing the correlation of effect sizes of ten genetic loci on hypothetical traits 1 and 2. **a.** Strong developmental bias and low horizontal pleiotropy, as seen by coherent and consistent pleiotropic effect on the two traits across the sampled loci. **b.** One large-effect pleiotropic locus appears to drive the genetic correlation between traits 1 and 2, showing a strong horizontal pleiotropic (HP) effect. **c.** and **d.** Suggest genetic mechanisms for the observed effect correlations in **a.** and **b.**.**c.** A developmental bias r_D, where each locus that affects one trait will inevitably affect the other trait, meaning that the traits are inherent correlated regardless of the type and directions of perturbations. **d.** Horizontal pleiotropy (HP), where a locus can have a direct effect to two traits. A third cause of genetic correlation is linkage disequilibrium (LD), where two loci, each determining a trait, are physically so closed that they are associated within a population more often. LD is not considered in detail in the current study and therefore not presented.

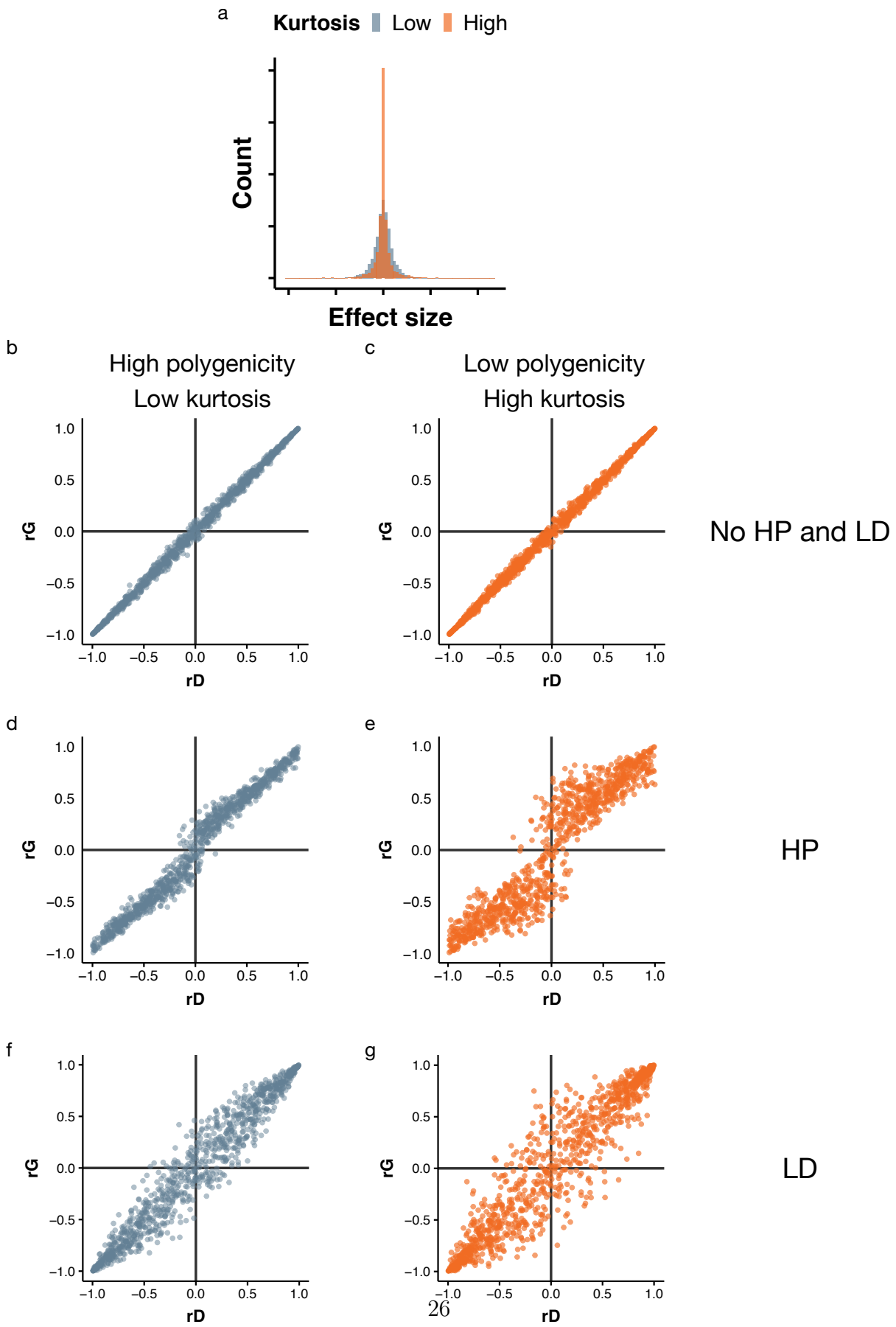
a
Kurtosis
Low
High
Count
Effect size
b
High polygenicity
Low kurtosis
c
Low polygenicity
High kurtosis
No HP and LD
d
e
HP
f
g
LD
rG
rD
1.0
0.5
0.0
−0.5
−1.0

Figure 2: Simulations showing the relationship between developmental bias,r_D, and genetic correlation, r_G. We simulated 2,000 pairs of traits using an exponential model, with varying σ among the plots shown. σ describes the correlation of effect size when generating effect sizes for each trait pair. A second parameter of the exponential model, γ, indicates regimes of varying kurtosis of the effect size distribution: large γ represents low kurtosis while smaller γ represents a regime with high kurtosis, as shown in **a.**. Two regimes of genetic architectures are considered, with γ equals 1.0 (plots **b.**, **d.**, and **f.**) and 0.5 (plots **c.**, **e.**, and **g.**) Three scenarios are simulated. **b.** and **c.** Under the first scenario, no horizontal pleiotropy (HP) and linkage disequilibrium (LD), the pattern does not change with the kurtosis and r_G nearly perfectly represents r_D. **d.** and **e.** Under the second scenario, to introduce HP, 0 to 10 randomly chosen SNP are introduced for each trait pair, with a shared pleiotropic effect for two traits, regardless of the pleiotropic effect of remaining loci. **f.** and **g.** For the third scenario, we used actual genotypes from a yeast mapping population to account for LD, without introducing any horizontal pleiotropic loci. To simulate trait pairs with actual genotype information, effect sizes are first generated as above, but are propagated through the LD block defined by $r > 0.5$. Note that the correlation of effect size is then calculated by excluding those 'repeated' loci within LD blocks.

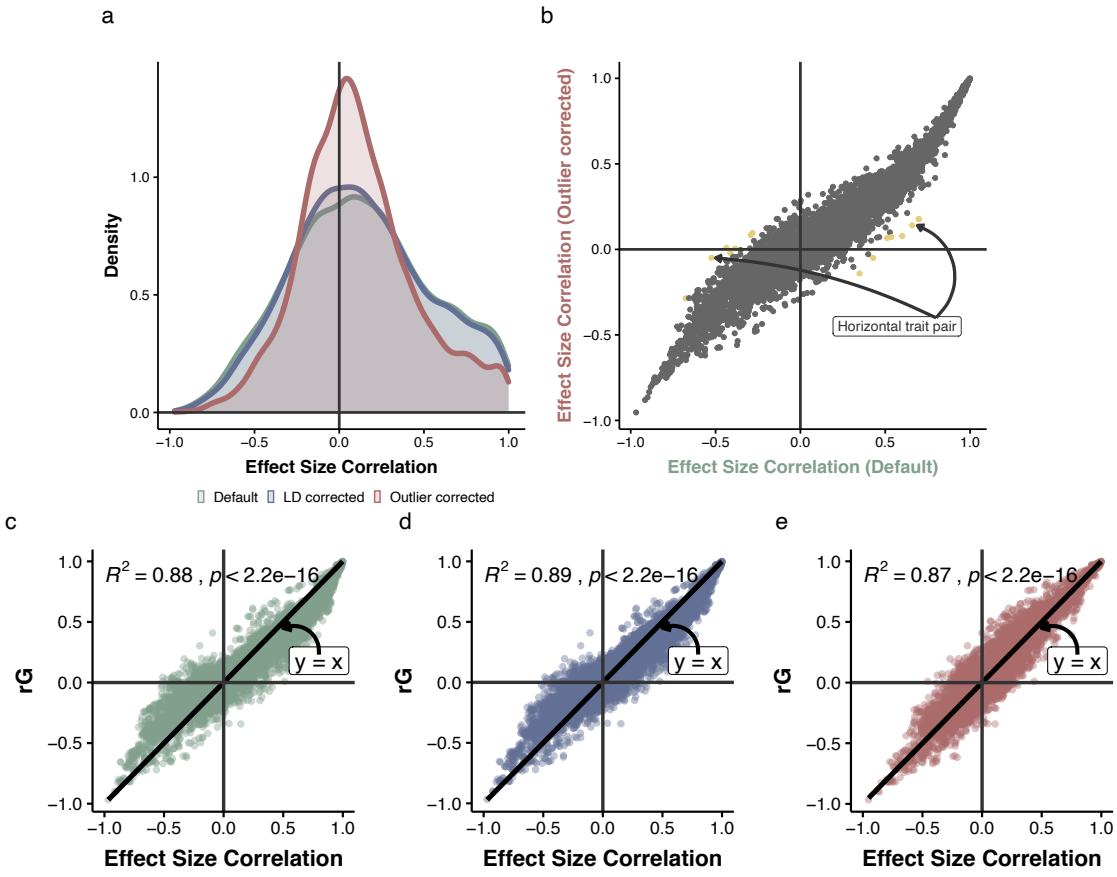

Figure 3: The relationship between rD and rG from a re-analysis of phenotypes measured in 374 recombinant strains of yeast cells (Geiler-Samerotte *et al.*, 2020). Each point in **b.-d.** represents a genetic locus from this empirical dataset. We consider three settings, summarized in **a.** and **b.**. **a.** Distribution of effect size correlation under three settings. Under the default setting, we included genome-wide markers to calculate correlation of effect size. LD pruned results only included the loci to the $r < 0.5$ within a chromosome. For outlier corrected effect size correlation, we excluded those outlier horizontal loci when calculating the correlation coefficient (See Materials and Methods). **b.** The effect size correlation under default versus outlier corrected settings. Yellow dots denote trait pairs that significantly affected by outliers correction (p-value <0.025). **c.-e.** Each dot in the scatter plot representing, for a given pair of traits in the yeast dataset which consists of 167 traits, the correlation of additive effect (on the x axis) and the genetic correlation (r_G, on the y axis). **c.** All loci across the genome. **d.** LD pruned variants. **e.** Outlier corrected variants.

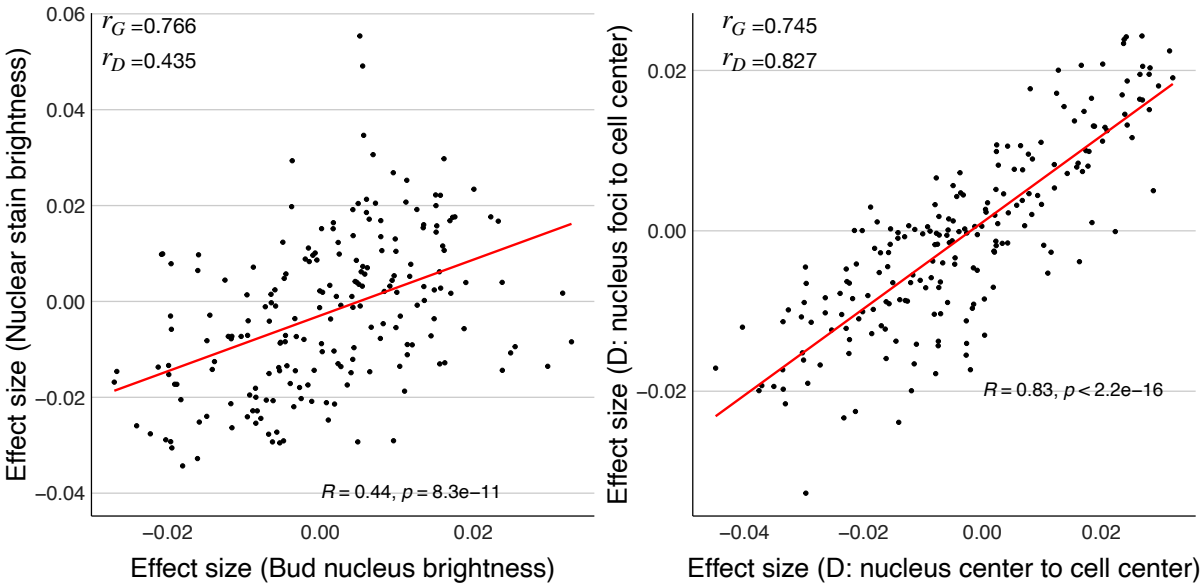

Figure 4: Two trait pairs with similar strength of r_G (0.766 and 0.745) exhibit difference in r_D (0.435 and 0.827). **a.** The r_G between nuclear stain brightness and bud nucleus brightness is marginally higher than the **b.** r_G between nucleus foci-to-cell center and nucleus center-to-cell center. However, r_D between nuclear stain brightness and bud nucleus brightness is lower than r_D between nucleus foci to cell center and nucleus center to cell center (0.435 vs 0.827), suggesting a higher developmental bias and inherent correlation between nucleus center-to-cell center and nucleus foci-to-cell center. Each point represents the additive effects for a single locus on each of two traits shown.

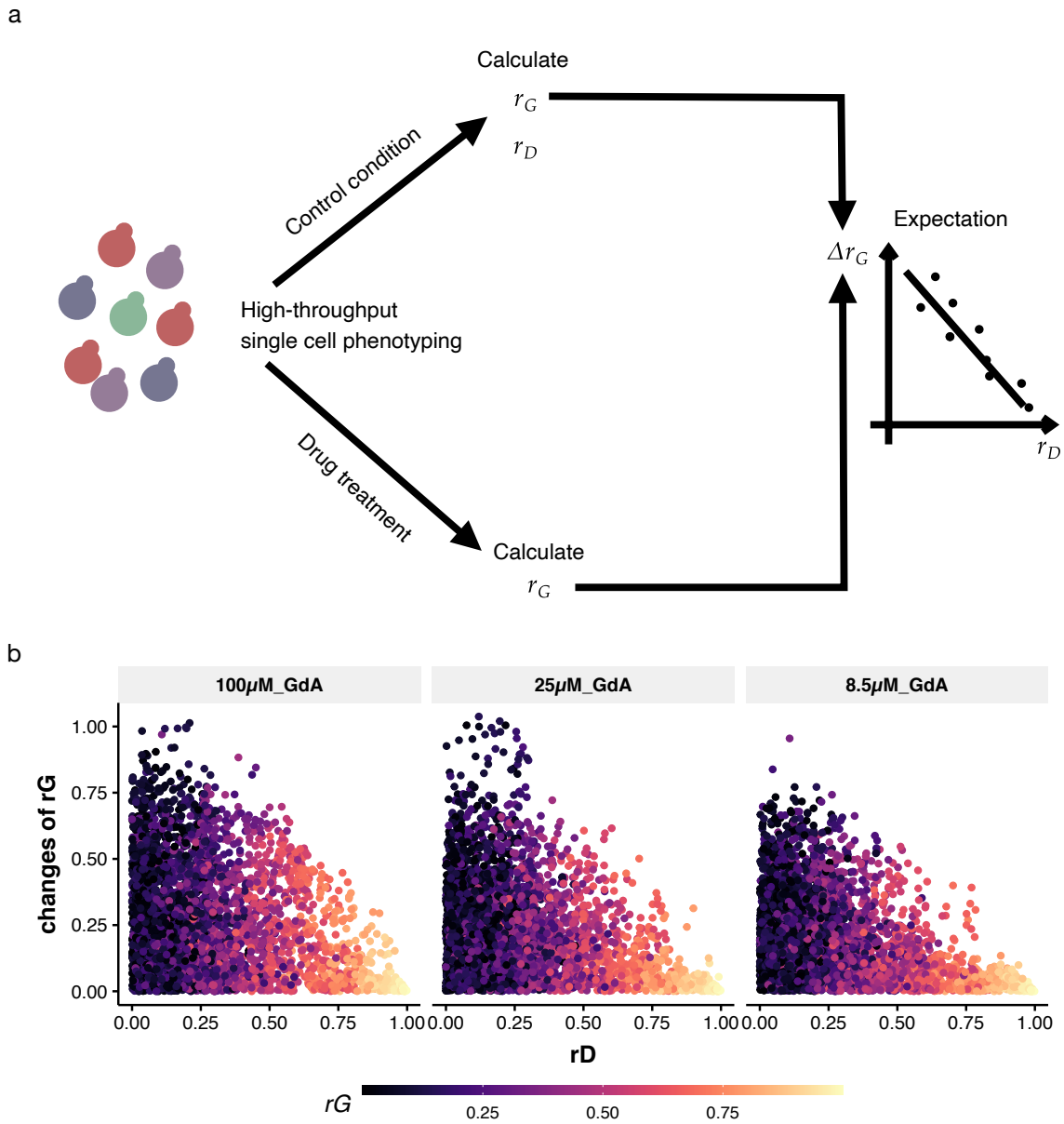

Figure 5: **a** Conceptual representation showing how high r_D could lead to stability of r_G. Many morphological traits are measured in a yeast biparental population using single cell phenotyping (Geiler-Samerotte *et al.*, 2020). These data are used to estimate r_G and r_D between traits. A subset of the yeast population is subjected to three levels of drug concentrations, representing three environmental conditions, and r_G is calculated for traits expressed in each of these treatments. We ask whether r_D is an indicator of Δr_G by using a multivariable linear model: $\Delta r_G \sim r_D + r_G$. We include r_G as a predictor because r_G itself can reflect its own plasticity. **b.** Scatter plots showing how r_G plasiticity in response to different drug concentration relates to r_D. The color bar indicates r_G in control condition.

Table 1: Details of horizontal trait pairs and nearest markers of driver loci

Phenotype1	Phenotype2	rG	rD_outlier_corrected	rD_default	Nearest marker for the biggest outlier
D14.3_A1B	D15.3_A1B	0.458	0.127	0.663	L15.9
D15.3_A1B	D175_A1B	0.447	0.193	0.701	L15.9
D15.3_A1B	D178_A1B	0.448	0.166	0.702	L15.9
D15.3_A1B	D181_A1B	0.372	0.087	0.607	L15.9
C13_C	C103_C	-0.332	0.017	-0.388	L13.7
C13_C	C118_C	-0.087	0.092	-0.29	L13.7
C13_C	D109_C	-0.215	0.017	-0.44	L13.7
C13_C	D131_C	-0.23	-0.055	-0.442	L13.7
D15.1_C	D17.1_C	0.188	-0.052	0.431	L15.9
D15.1_C	D17.2_C	-0.038	-0.023	0.407	L15.9
D15.1_C	D182_C	0.218	0.097	0.529	L15.9
D15.3_C	D17.1_C	0.106	-0.15	0.352	L15.9
D15.3_C	D188_C	0.111	0.096	0.542	L15.9
D17.1_C	D117_C	-0.379	-0.049	-0.523	L15.9
D17.1_C	D169_C	-0.307	0.035	-0.408	L15.9

References

Agrawal, A. F., Brodie, E. D. & Rieseberg, L. H. (2001). Possible consequences of genes of major effect: transient changes in the g-matrix. In: *Microevolution rate, pattern, process*. Springer, pp. 33–43.

Albert, A. Y., Sawaya, S., Vines, T. H., Knecht, A. K., Miller, C. T., Summers, B. R., Balabhadra, S., Kingsley, D. M. & Schluter, D. (2008). The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution: International Journal of Organic Evolution*, 62, 76–85.

Arnold, S. J. (1992). Constraints on phenotypic evolution. *The American Naturalist*, 140, S85–S107.

Arnold, S. J., Bürger, R., Hohenlohe, P. A., Ajie, B. C. & Jones, A. G. (2008). Understanding the evolution and stability of the g-matrix. *Evolution: International Journal of Organic Evolution*, 62, 2451–2461.

Barton, N. (1990). Pleiotropic models of quantitative variation. *Genetics*, 124, 773–782.

Barton, N. H. & Turelli, M. (1987). Adaptive landscapes, genetic distance and the evolution of quantitative characters. *Genetics Research*, 49, 157–173.

Barton, N. H. & Turelli, M. (1989). Evolutionary quantitative genetics: how little do we know. *Annual review of genetics*, 23, 337–370.

Bliese, P. (2013). Multilevel modeling in r (2.6). *Retrieved September*, 3, 2013.

Bowden, J., Hemani, G. & Davey Smith, G. (2018). Invited commentary: detecting individual and global horizontal pleiotropy in mendelian randomization—a job for the humble heterogeneity statistic? *American journal of epidemiology*, 187, 2681–2685.

Bradshaw Jr, H., Otto, K. G., Frewen, B. E., McKay, J. K. & Schemske, D. W. (1998). Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (mimulus). *Genetics*, 149, 367–382.

Brock, M. T., Dechaine, J. M., Iniguez-Luy, F. L., Maloof, J. N., Stinchcombe, J. R. & Weinig, C. (2010). Floral genetic architecture: an examination of qtl architecture underlying floral (co) variation across environments. *Genetics*, 186, 1451–1465.

Bürger, R. & Lande, R. (1994). On the distribution of the mean and variance of a quantitative trait under mutation-selection-drift balance. *Genetics*, 138, 901–912.

Chakrabarty, A. & Schielzeth, H. (2020). Comparative analysis of the multivariate genetic architecture of morphological traits in three species of gomphocerine grasshoppers. *Heredity*, 124, 367–382.

Chebib, J. & Guillaume, F. (2021). Pleiotropy or linkage? their relative contributions to the genetic correlation of quantitative traits and detection by multitrait gwa studies. *Genetics*, 219, iyab159.

Cheverud, J. M. (1984). Quantitative genetics and developmental constraints on evolution by selection. *Journal of theoretical biology*, 110, 155–171.

Clark, A. (1987). Genetic correlations: the quantitative genetics of evolutionary constraints. In: *Genetic constraints on adaptive evolution*. Springer.

Conner, J. K. (2002). Genetic mechanisms of floral trait correlations in a natural population. *Nature*, 420, 407–410.

Conner, J. K., Hartl, D. L. *et al.* (2004). *A primer of ecological genetics*, vol. 425. Sinauer Associates Sunderland, MA.

Conner, J. K., Karoly, K., Stewart, C., Koelling, V. A., Sahli, H. F. & Shaw, F. H. (2011). Rapid independent trait evolution despite a strong pleiotropic genetic correlation. *The American Naturalist*, 178, 429–441.

Delahaie, B., Charmantier, A., Chantepie, S., Garant, D., Porlier, M. & Teplitsky, C. (2017). Conserved g-matrices of morphological and life-history traits among continental and island blue tit populations. *Heredity*, 119, 76–87.

Delph, L. F., Steven, J. C., Anderson, I. A., Herlihy, C. R. & Brodie III, E. D. (2011). Elimination of a genetic correlation between the sexes via artificial correlational selection. *Evolution*, 65, 2872–2880.

Denver, D. R., Morris, K., Streelman, J. T., Kim, S. K., Lynch, M. & Thomas, W. K. (2005). The transcriptional consequences of mutation and natural selection in caenorhabditis elegans. *Nature genetics*, 37, 544–548.

Des Marais, D. L. & Juenger, T. E. (2010). Pleiotropy, plasticity, and the evolution of plant abiotic stress tolerance. *Annals of the New York Academy of Sciences*, 1206, 56–79.

Doroszuk, A., Wojewodzic, M. W., Gort, G. & Kammenga, J. E. (2008). Rapid divergence of genetic variance-covariance matrix within a natural population. *The American Naturalist*, 171, 291–304.

Dugand, R. J., Aguirre, J. D., Hine, E., Blows, M. W. & McGuigan, K. (2021). The contribution of mutation and selection to multivariate quantitative genetic variance in an outbred population of drosophila serrata. *Proceedings of the National Academy of Sciences*, 118, e2026217118.

Eroukhmanoff, F. & Svensson, E. (2011). Evolution and stability of the g-matrix during the colonization of a novel environment. *Journal of evolutionary biology*, 24, 1363–1373.

Falconer, D. S., Mackay, T. F. & Frankham, R. (1996). Introduction to quantitative genetics (4th edn). *Trends in Genetics*, 12, 280.

Gage, J. L., Monier, B., Giri, A. & Buckler, E. S. (2020). Ten years of the maize nested association mapping population: impact, limitations, and future directions. *The Plant Cell*, 32, 2083–2093.

Geiler-Samerotte, K. A., Li, S., Lazaris, C., Taylor, A., Ziv, N., Ramjeawan, C., Paaby,

A. B. & Siegal, M. L. (2020). Extent and context dependence of pleiotropy revealed by high-throughput single-cell phenotyping. *PLoS biology*, 18, e3000836.

Geiler-Samerotte, K. A., Zhu, Y. O., Goulet, B. E., Hall, D. W. & Siegal, M. L. (2016). Selection transforms the landscape of genetic variation interacting with hsp90. *PLoS biology*, 14, e2000465.

Hall, M. C., Basten, C. J. & Willis, J. H. (2006). Pleiotropic quantitative trait loci contribute to population divergence in traits associated with life-history variation in mimulus guttatus. *Genetics*, 172, 1829–1844.

Hangartner, S., Lasne, C., Sgrò, C. M., Connallon, T. & Monro, K. (2020). Genetic covariances promote climatic adaptation in australian drosophila. *Evolution*, 74, 326–337.

He, X. & Zhang, J. (2006). Toward a molecular understanding of pleiotropy. *Genetics*, 173, 1885–1891.

Henry, G. A. & Stinchcombe, J. R. (2022). G matrix stability of clinally diverging populations of an annual weed. *bioRxiv*.

Henry, G. A. & Stinchcombe, J. R. (2023). G-matrix stability in clinally diverging populations of an annual weed. *Evolution*, 77, 49–62.

Hine, E., Runcie, D. E., Allen, S. L., Wang, Y., Chenoweth, S. F., Blows, M. W. & McGuigan, K. (2022). Maintenance of quantitative genetic variance in complex, multi-trait phenotypes: The contribution of rare, large effect variants in two drosophila species. *bioRxiv*.

Houle, D., Bolstad, G. H., van der Linde, K. & Hansen, T. F. (2017). Mutation predicts 40 million years of fly wing evolution. *Nature*, 548, 447–450.

Hudson, A. I., Odell, S. G., Dubreuil, P., Tixier, M.-H., Praud, S., Runcie, D. E. & Ross-Ibarra, J. (2022). Analysis of genotype-by-environment interactions in a maize mapping population. *G3*, 12, jkac013.

Jones, A. G., Arnold, S. J. & Bürger, R. (2003). Stability of the g-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution*, 57, 1747–1760.

Jordan, D. M., Verbanck, M. & Do, R. (2019). Hops: a quantitative score reveals pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. *Genome biology*, 20, 1–18.

Juenger, T., Pérez-Pérez, J. M., Bernal, S. & Micol, J. L. (2005). Quantitative trait loci mapping of floral and leaf morphology traits in arabidopsis thaliana: evidence for modular genetic architecture. *Evolution & development*, 7, 259–271.

Kelly, J. K. (2009). Connecting qtls to the g-matrix of evolutionary quantitative genetics. *Evolution: International Journal of Organic Evolution*, 63, 813–825.

Klingenberg, C. P. (2005). Developmental constraints, modules, and evolvability. In: *Variation*. Elsevier, pp. 219–247.

Klingenberg, C. P. (2008). Morphological integration and developmental modularity. *Annual review of ecology, evolution, and systematics*, 39, 115–132.

Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., Purugganan, M. D., Durrant, C. & Mott, R. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in arabidopsis thaliana. *PLoS genetics*, 5, e1000551.

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution*, 402–416.

Lande, R. (1980). The genetic covariance between characters maintained by pleiotropic mutations. *Genetics*, 94, 203–215.

Lande, R. & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, 1210–1226.

Lea, A., Subramaniam, M., Ko, A., Lehtimäki, T., Raitoharju, E., Kähönen, M., Seppälä, I., Mononen, N., Raitakari, O. T., Ala-Korpela, M. *et al.* (2019). Genetic and environmental perturbations lead to regulatory decoherence. *eLife*, 8, e40538.

Lee, C.-R., Anderson, J. T. & Mitchell-Olds, T. (2014). Unifying genetic canalization, genetic constraint, and genotype-by-environment interaction: Qtl by genomic background by environment interaction of flowering time in boechera stricta. *PLoS genetics*, 10, e1004727.

Loeschcke, V. (1987). *Genetic constraints on adaptive evolution*. Springer.

Lynch, M., Walsh, B. *et al.* (1998). Genetics and analysis of quantitative traits.

Maynard Smith, J., Burian, R., Kauffman, S., Alberch, P., Campbell, J., Goodwin, B., Lande, R., Raup, D. & Wolpert, L. (1985). Developmental constraints and evolution: a perspective from the mountain lake conference on development and evolution. *The Quarterly Review of Biology*, 60, 265–287.

McGlothlin, J. W., Kobiela, M. E., Wright, H. V., Mahler, D. L., Kolbe, J. J., Losos, J. B. & Brodie III, E. D. (2018). Adaptive radiation along a deeply conserved genetic line of least resistance in anolis lizards. *Evolution Letters*, 2, 310–322.

McGuigan, K., Collet, J. M., Allen, S. L., Chenoweth, S. F. & Blows, M. W. (2014). Pleiotropic mutations are subject to strong stabilizing selection. *Genetics*, 197, 1051–1062.

Meyer, H. V. & Birney, E. (2018). Phenotypesimulator: a comprehensive framework for simulating multi-trait, multi-locus genotype to phenotype relationships. *Bioinformatics*, 34, 2951–2956.

Milocco, L. & Salazar-Ciudad, I. (2022). Evolution of the g matrix under nonlinear genotype-phenotype maps. *The American Naturalist*, 199, 000–000.

Muir, C. D., Conesa, M. À., Galmés, J., Pathare, V., Rivera, P., Rodríguez, R. L., Terrazas, T. & Xiong, D. (2022). How important are functional and developmental constraints on phenotypic evolution? an empirical test with the stomatal anatomy of flowering plants.

Opedal, Ø. H., Armbruster, W. S., Hansen, T. F., Holstad, A., Pélabon, C., Andersson, S., Campbell, D. R., Caruso, C. M., Delph, L. F., Eckert, C. G. *et al.* (2023). Evolvability

and trait function predict phenotypic divergence of plant populations. *Proceedings of the National Academy of Sciences*, 120, e2203228120.

Orr, H. A. (2000). Adaptation and the cost of complexity. *Evolution*, 54, 13–20.

Otto, S. P. (2004). Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271, 705–714.

Paaby, A. B. & Rockman, M. V. (2013). Pleiotropy: what do you mean? reply to zhang and wagner. *Trends in Genetics*, 29, 384.

Pavlicev, M., Cheverud, J. M. & Wagner, G. P. (2009). Measuring morphological integration using eigenvalue variance. *Evolutionary Biology*, 36, 157–170.

Pavlicev, M. & Wagner, G. P. (2012). Coming to grips with evolvability. *Evolution: Education and Outreach*, 5, 231–244.

Pigliucci, M. & Preston, K. (2004). *Phenotypic integration: studying the ecology and evolution of complex phenotypes*. Oxford University Press.

Promislow, D. E. (2004). Protein networks, pleiotropy and the evolution of senescence. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271, 1225–1234.

Saltz, J. B., Hessel, F. C. & Kelly, M. W. (2017). Trait correlations in the genomics era. *Trends in ecology & evolution*, 32, 279–290.

Schluter, D. (1996). Adaptive radiation along genetic lines of least resistance. *Evolution*, 50, 1766–1774.

Scoville, A., Lee, Y. W., Willis, J. H. & Kelly, J. K. (2009). Contribution of chromosomal polymorphisms to the g-matrix of mimulus guttatus. *New Phytologist*, 183, 803–815.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics*, 14, 483–495.

Stearns, F. W. (2010). One hundred years of pleiotropy: a retrospective. *Genetics*, 186, 767–773.

Steiger, J. H. (1980). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research*, 15, 335–352.

Steppan, S. J., Phillips, P. C. & Houle, D. (2002). Comparative quantitative genetics: evolution of the g matrix. *Trends in Ecology & Evolution*, 17, 320–327.

Stinchcombe, J. R., Weinig, C., Heath, K. D., Brock, M. T. & Schmitt, J. (2009). Polymorphic genes of major effect: consequences for variation, selection and evolution in arabidopsis thaliana. *Genetics*, 182, 911–922.

Turelli, M. (1988). Phenotypic evolution, constant covariances, and the maintenance of additive variance. *Evolution*, 42, 1342–1347.

Uller, T., Moczek, A. P., Watson, R. A., Brakefield, P. M. & Laland, K. N. (2018). Developmental bias and evolution: A regulatory network perspective. *Genetics*, 209, 949–966.

Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nature genetics*, 50, 693–698.

Waddington, C. H. (1957). *The strategy of the genes*. Routledge.

Wagner, G. P., Kenney-Hunt, J. P., Pavlicev, M., Peck, J. R., Waxman, D. & Cheverud, J. M. (2008). Pleiotropic scaling of gene effects and the 'cost of complexity'. *Nature*, 452, 470–472.

Wagner, G. P. & Zhang, J. (2011). The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics*, 12, 204–213.

Walter, G. M., Aguirre, J. D., Blows, M. W. & Ortiz-Barrientos, D. (2018). Evolution of genetic variance during adaptive radiation. *The American Naturalist*, 191, E108–E128.

Wessinger, C. A., Hileman, L. C. & Rausher, M. D. (2014). Identification of major quantitative trait loci underlying floral pollination syndrome divergence in penstemon. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130349.

Williams, G. C. (2001). Pleiotropy, natural selection, and the evolution of senescence: Evolution 11, 398-411 (1957). *Science of Aging Knowledge Environment*, 2001, cp13–cp13.

Wood, C. W. & Brodie III, E. D. (2015). Environmental effects on the structure of the g-matrix. *Evolution*, 69, 2927–2940.

Zhang, J. & Wagner, G. P. (2013). On the definition and measurement of pleiotropy. *Trends in Genetics*, 29, 383–384.

Appendices

Supplementary Note 1: Distinctions between linkage disequilibrium and tight linkage

Our simulation and analyses explictly account for LD that specifically defined as linkage disequilibrium between markers in a genetic mapping study. Another form of linkage, which is the tight linkage in a genomic context (i.e., perhect linkage, no recombination within pairs of linked loci), is implicitly accounted for when removing major horizontal pleiotropic loci as outliers. We are not able to distinguish between tight linkage and pleiotropy (though these two architectures may differ in maintaining genetic correlation, see (Chebib & Guillaume, 2021)). These major horizontal pleiotropic loci are either caused by LD within the marker or pleiotropy. How about LD and horizontal pleiotropy with small effect size loci? In fact, polygenic trait correlations have been suggested to be unlikely to arise from chance LD events, as well as several individual horizontal pleiotropy (Saltz *et al.*, 2017). Therefore, in essence, our metric dissects the genetic correlation by accouting for allele frequency, linkage equilibrium, as well as horizontal pleiotropy, providing an approach to explore the role of developmental constraint (trait-trait relationship) and vertical pleiotropy (developmental constraint in a variant level).

Supplementary Note 2: Additional dataset characterizes r_D under different conditions

We analyzed an additional data set consisting of vegetative and floral traits measured in a population of *B. rapa L.* in the field and in greenhouse (Brock *et al.*, 2010). We present the result based on outlier corrected subset of variants (LD pruned). Fig. S5a shows r_D against r_G (also see Supplementary Table 1). We found, that most trait pairs with high r_G exhibit high r_D. This is expected since those trait pairs with high r_G are mostly floral morphological traits, which have long been assumed under developmental integration. However, we still found a few exceptions: trait pairs that exhibit low r_D but high r_G. Specifically, for instance, petal length and the petal width have a r_G comparable to midpoint length and filament length, petal length and ovary length. However, petal length and petal width has relatively low r_D (especially when we use 95% lower bound of effect size correlation to indicate r_D). This is in contrast to two other pairs of traits, who exhibit higher r_D. Notably, very few length traits appear to have significant r_G with petal width in this dataset (Brock *et al.*, 2010). Juenger *et al.* (2005) also found that, in Arabidopsis, the sets of floral organ lengths (petal length, sepal length, long stamen length, pistil length) or organ widths (petal width, sepal width) were highly correlated, while r_G between length and width measures were generally not significant except petal length - petal width. Therefore, we speculate that instead of r_D as a general cause of r_G, LD or horizontal pleiotropic loci play a critical role in shaping r_G between floral organ lengths and organ widths.

Since the sample size does not allow us to assess the genetic correlation (r_G) change for a given

trait pair, we thus calculated the distribution difference in two environments. As shown, Δr_D appears to have different patterns compared to r_G in response to the environmental change (Fig. S5b, Kolmogorov-Smirnov test p-value = 8.923e-07). Surprisingly, the **G** matrix element has a significant shift (Wilcoxon signed-rank test, p-value = 2.184e-5) while distribution of **D** matrix elements does not (Wilcoxon signed-rank test, p-value = 0.8373). Furthermore, in line with the analyses of yeast morphological data, traits with higher r_D tends to be more conserved in their r_G between conditions: large changes in r_G are more likely to occur for those pairs with lower r_D(Fig. S5c, Fig. S6).

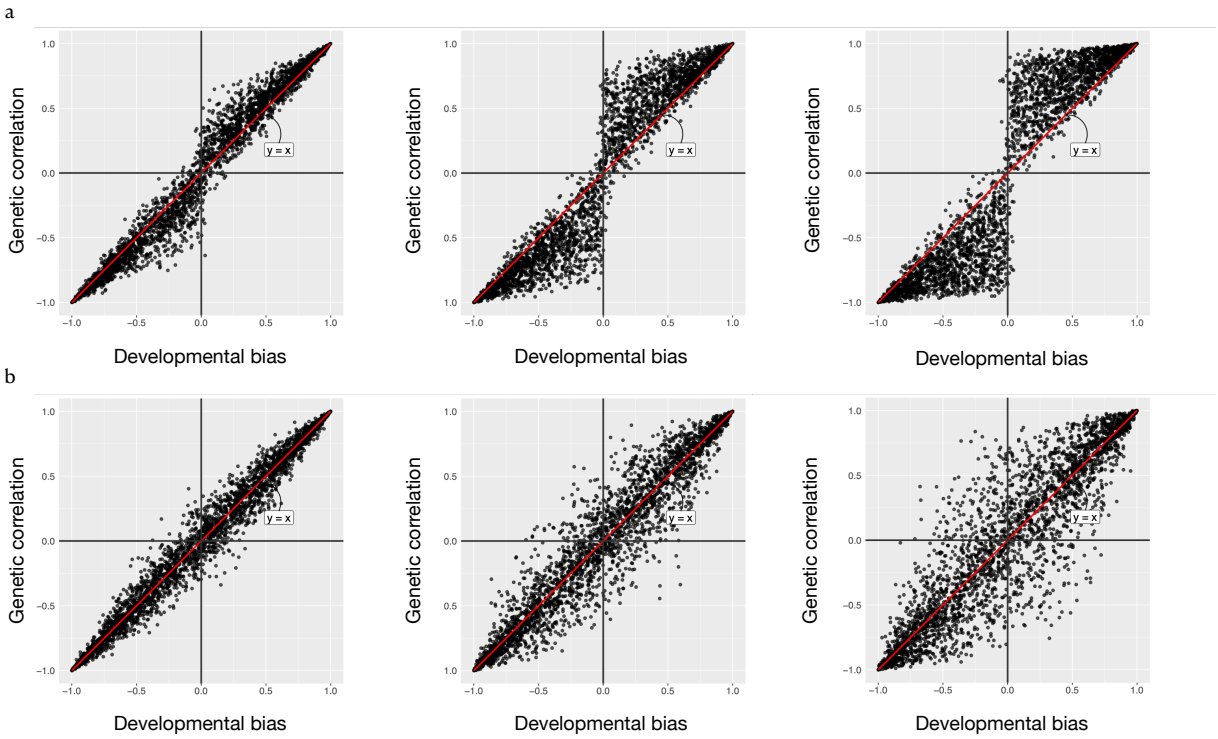

Figure S1: Reproduce the Fig. 2 but using bivariate normal distribution to generate effect sizes for each loci. Each point represents a pair of traits. Given a population, we simulated 3,000 pairs of traits using a model under which each trait pair is consisted of one randomly chosen SNP with a 'large' pleiotropic (additive) effect, and other 999 SNPs with 'small' (additive) effects which are sample from a bivariate normal distribution with $\mu = (0, 0)$ and $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. ρ is sample from uniform distribution $(-1, 1)$. **a.** simulations under concordant model, where large pleiotropic loci has concordant effect with the genetic background. **b.** simulations without concordant assumption, where large pleiotropic loci has either concordant or antagonistic effect with the genetic background. From left to right, effect size of the large pleiotropic loci increases.

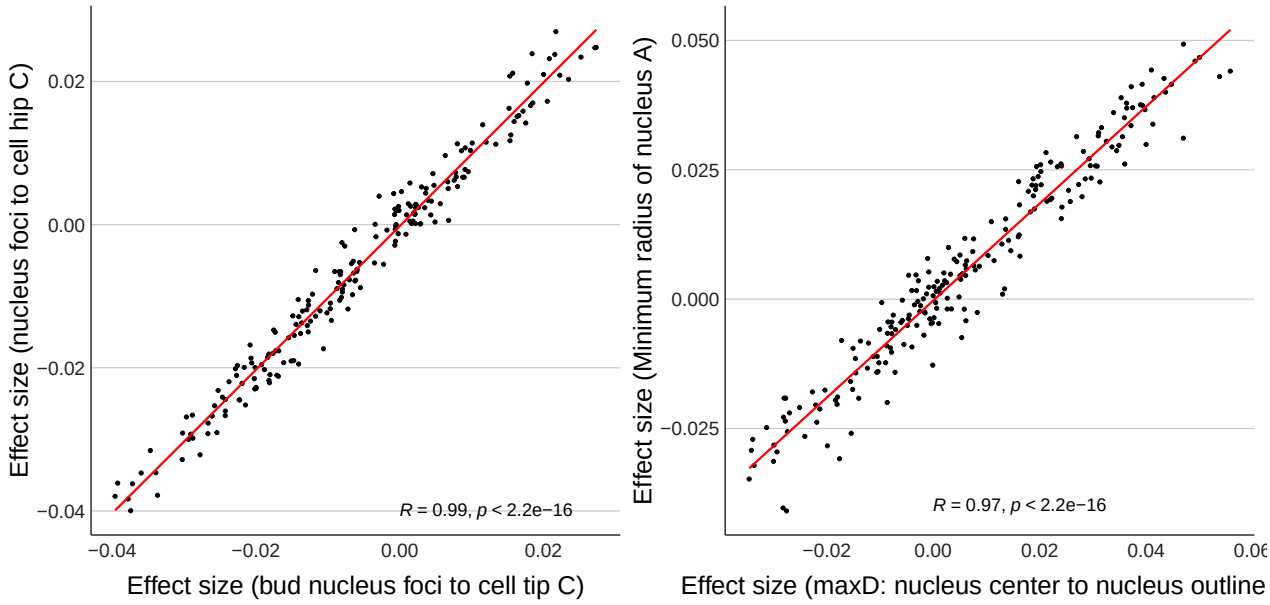

Figure S2: Two example trait pairs showing extremely strong r_D between traits. Each point represents additive effect for a single locus. These trait pairs demonstrate strong inherent redundancy and genetic correlation is likely driven by developmental bias alone for these trait pairs.

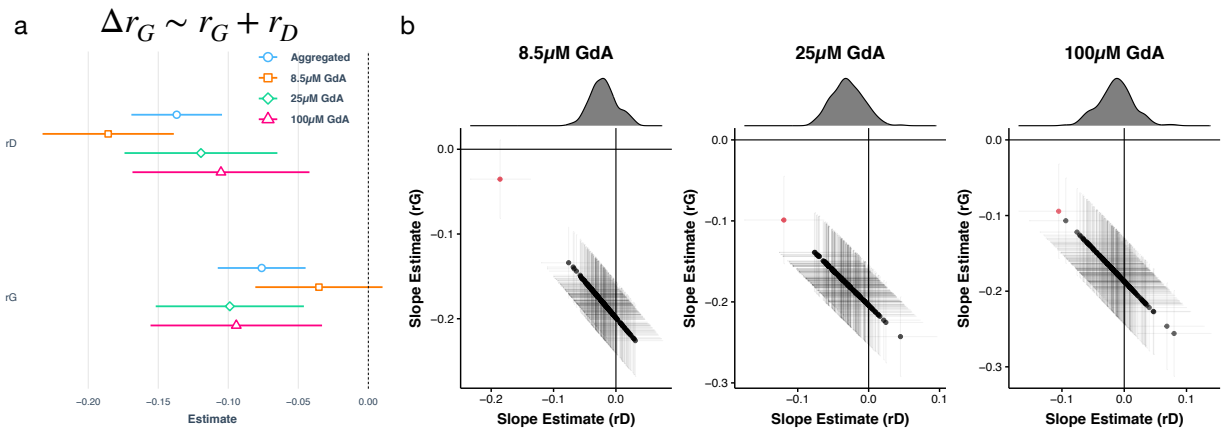

Figure S3: **a** Linear regression results for $\Delta r_G \sim r_D + r_G$. Estimates of coefficient and their 95% confidence intervals under four conditions (use all pooled data at aggregated condition) for r_D and r_G are shown. **b** Since rG and rD are highly correlated, which might cause multicolinearity problems during regressions. Therefore, we did additional analyses by simulating a null model (black dots): $\Delta r_G \sim r_n + r_G$, where r_n is sampled from the bivariate normal distribution with covariance 0.939, conditioning on r_G. The coefficient estimates of r_D under all conditions are significantly different from noise, suggesting rD may provide additional information besides r_G.

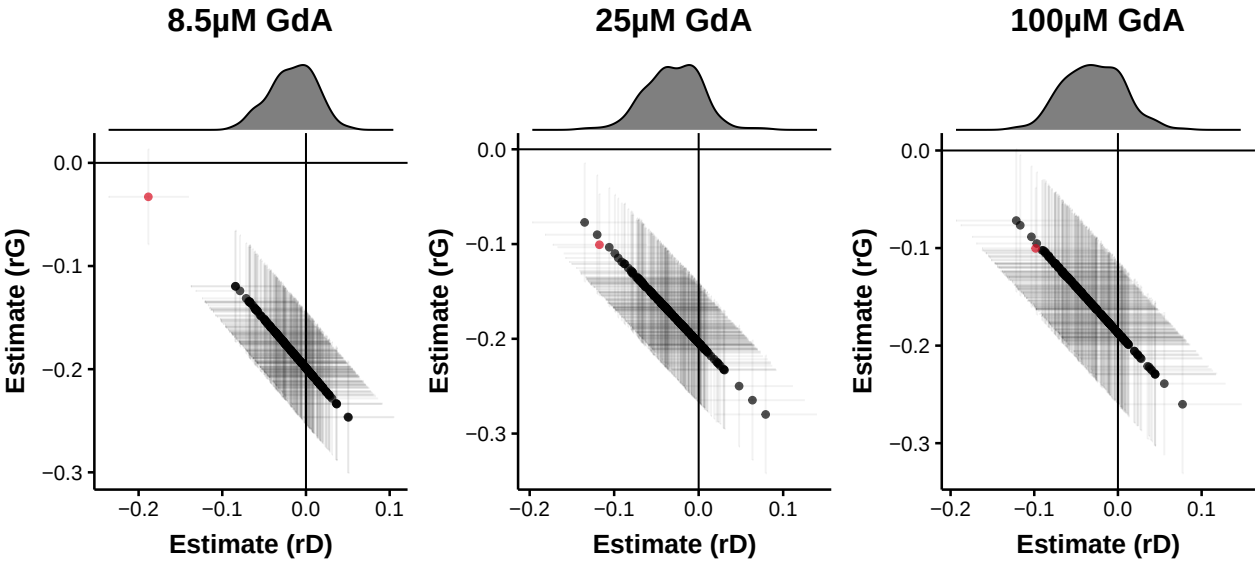

Figure S4: Similarly with Fig. S3b, A null model, where rn is generated by r_G with Gaussian noise, shows qualitatively similar results with Fig. S3.

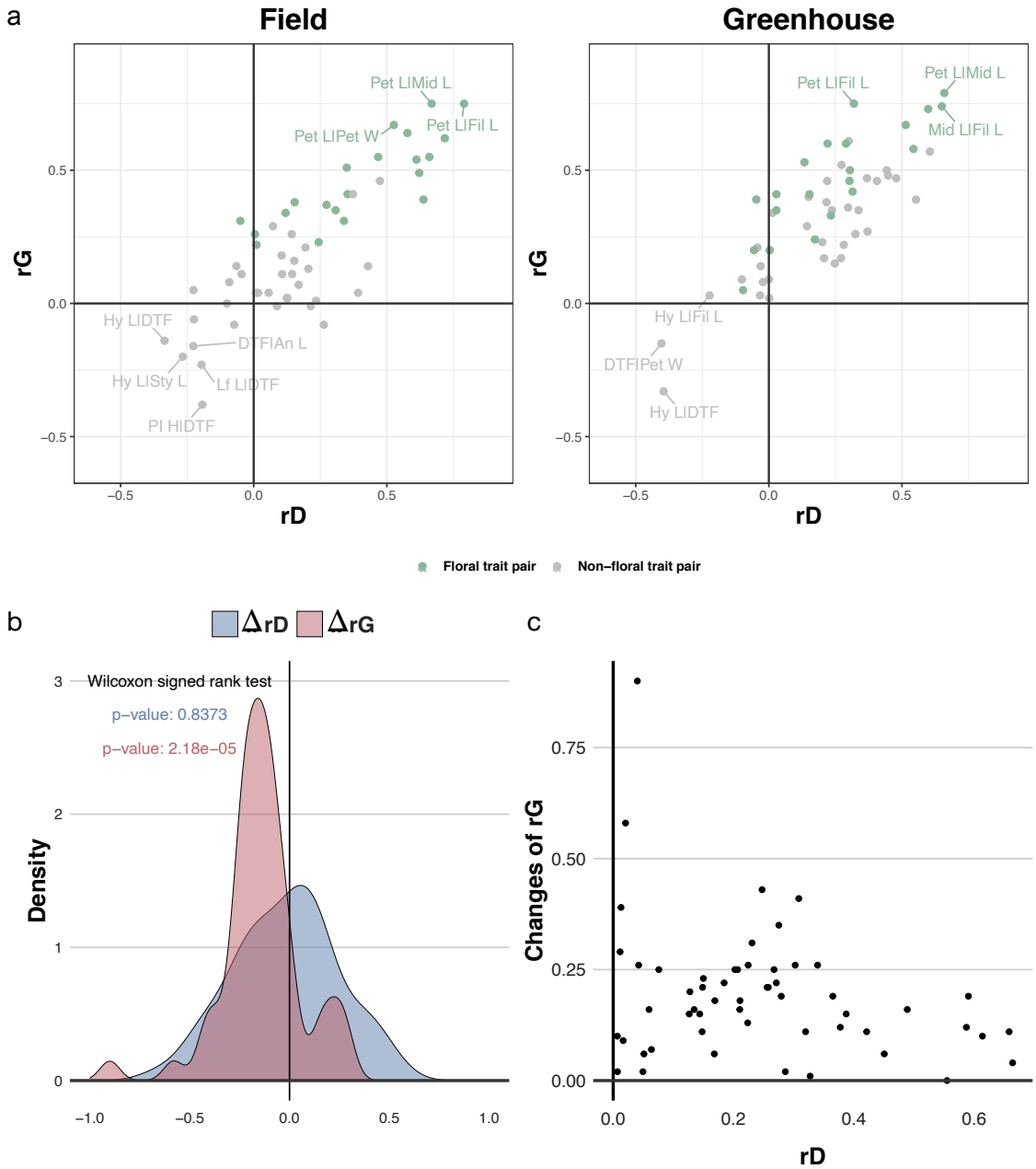

Figure S5: **a** A *Brassica* dataset with 11 floral, vegetative, and phenology traits (Brock *et al.*, 2010) was analysed similarly to Fig. 3, under two environmental conditions(field and greenhouse). The loci used to calculate the correlation of effect size are LD pruned and outlier corrected. Green dots denote pairs of traits that are both floral. **b** The distribution of difference of developmental bias r_D and genetic correlation r_G. The estimated developmental bias has a different pattern responding to environmental perturbation, comparing with genetic correlation (p-value = 8.923e-07). Wilcoxon signed-rank tests show that G matrix elements are significantly changed across conditions (p-value = 2.184e-5) while D matrix elements are not (p-value = 0.8373). **c** Trait pairs with large r_D have more stable genetic correlation The x axis is the averaged developmental bias across two environments (Field and Greenhouse). y axis shows the difference of genetic correlation for a given pair. The plot suggests that the upper bound of differential genetic correlation is decreased as the developmental bias becomes stronger.

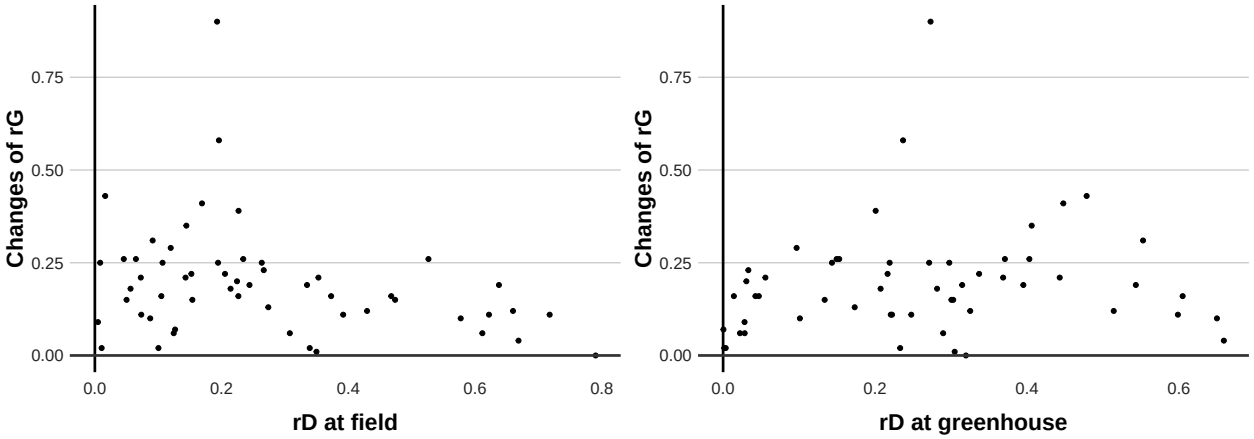

Figure S6: Reproduce of Fig. S5c by using the rD in each environmental condition.