Pleiotropy or linkage? Their relative contributions to the genetic correlation of quantitative traits and detection by multi-trait GWA studies.

Jobran Chebib and Frédéric Guillaume

Department of Evolutionary Biology and Environmental Studies, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. email: frederic.guillaume@ieu.uzh.ch

Abstract

Genetic correlations between traits may cause correlated responses to selection depending on the source of those genetic dependencies. Previous models described the conditions under which genetic correlations were expected to be maintained. Selection, mutation and migration are all proposed to have similar effects on genetic correlation, regardless of whether the underlying genetic architecture consists of pleiotropic or tightly-linked loci affecting the traits. Here, we investigate the conditions under which pleiotropy and linkage have differential effects on the genetic correlations between traits by explicitly modeling multiple genetic architectures to look at the effects of selection strength, degree of correlational selection, mutation rate, mutational variance, recombination rate, and migration rate. We show that at mutation-selection(-migration) balance, mutation rates differentially affect the equilibrium levels of genetic correlation when architectures are composed of pairs of physically linked loci compared to architectures of pleiotropic loci. Even when there is perfect linkage (no recombination within pairs of linked loci), a lower genetic correlation is maintained than with pleiotropy, with a lower mutation rate leading to a larger decrease. These results imply that the detection of causal loci in multi-trait association studies will be affected by the type of underlying architectures, whereby pleiotropic variants are more

likely to be underlying multiple detected associations. We also confirm that tighter linkage between non-pleiotropic causal loci maintains higher genetic correlations at the traits and leads to a greater proportion of false positives in association analyses.

Keywords: Pleiotropy, Linkage, Genetic Architecture, Correlational Selection, Migration, Mutation

---

Introduction

Both pleiotropy and linkage disequilibrium create genetic correlations between traits so that traits do not act independently of one another (Wright, 1977; Arnold, 1992; Walsh and Blows, 2009). Under natural selection, this process can prevent a combination of traits from reaching their respective optimal potentials (Falconer and Mackay, 1996). Likewise, under artificial selection it can constrain breeders from improving desired change in one trait due to undesired changes in another, and in medical gene targeted therapy treatments it can cause adverse side-effects (Wright, 1977; Parkes et al., 2013; Visscher et al., 2017). Pleiotropy may cause genetic correlation because one gene's product (e.g. an enzyme or a transcription factor) has more than one target and therefore affects more than one trait or because one gene's product belongs to a metabolic pathway that has more than one downstream effect (Hodgkin, 1998; Stearns, 2010; Wagner and Zhang, 2011). Linkage disequilibrium (LD) may be the result of a set of loci in close physical proximity on a chromosome that makes a set of alleles at those loci less likely to be split up by recombination and therefore more likely to get passed on together from one generation to the next. But other mechanisms leading to the transmission of one combination of alleles at separate loci over another combination, can also generate LD and create genetic correlations between traits that those loci affect (e.g., assortative mating, environmental correlations) (Falconer and Mackay, 1996).

One of the main objectives of a genome-wide association study (GWAS)

is to identify causal genetic variants underlying one or more traits. GWASes leverage the explosion in genomic sequencing to find correlations between traits and genotypes, and their success is dependent on the size of the effect and the distinction between phenotypes. GWASes have had success in associating genetic variants with traits of interest, which have allowed researchers to find the molecular underpinnings of trait change (Visscher et al., 2017). Moving from one trait to two or more trait associations can lead to discovering pleiotropic loci (Saltz et al., 2017). One GWAS using 1094 traits and 14,459 genes, found that 44% of genes were "pleiotropic", but this was determined by assigning genetic variants to the closest gene and even to both flanking genes when the genetic variant was intergenic (Chesmore et al., 2018). This conflates linkage and pleiotropy, and the chain of causality (Platt et al., 2010). Another study, found 90% of genes and 32.4% of SNPs were associated with more than one trait domain, but they could not rule out SNPs associated with traits due to linkage disequilibrium (Watanabe et al., 2018). Unfortunately, determining whether genetic variant associations and trait correlations are actually the result of pleiotropy or linkage is difficult since they often map to large regions of genomes, or are in intergenic regions and don't associate with the closest genes (Flint and Mackay, 2009; Zhu et al., 2016; Peichel and Marques, 2017; Visscher et al., 2017). Distinguishing between the two types of genetic architectures is important for understanding the underlying molecular functions of the traits, and determining how the traits may be deferentially affected by selection (Lynch et al., 1998; Barrett and Hoekstra, 2011; Saltz et al., 2017). This is salient at a time when an increasing number of traits of interest (e.g. human diseases) appear to be affected by loci that affect other traits, and especially when targeted gene therapy clinical trials are more widespread than ever (Edelstein et al., 2007; Cai et al., 2016; Pickrell et al., 2016; Visscher and Yang, 2016; Chesmore et al., 2018; Ginn et al., 2018). There are potentially negative implications for gene therapy because fixing a gene underlying one disease might increase

risk for another disease. For example, some genetic variants that are associated with greater risk of Ankylosing spondylitis are also associated with less risk of Rheumatoid arthritis, and so "fixing" one gene would have undesired side-effects in this case (Parkes et al., 2013; Gratten and Visscher, 2016).

But the evolutionary dynamics of pleiotropic versus linked loci in creating genetic correlations are expected to be different, since pleiotropy requires only one mutation to affect multiple traits and build-up genetic correlations, and linked pairs require two. Also, recombination can break up associations between alleles at linked loci, reducing genetic correlations between the traits they affect, but the same cannot occur with a pleiotropic locus. Polygenic analytical models attempting to approximate the level of genetic variance and covariance at mutation-selection balance in a population suggest that tight linkage between pairs of loci affecting separate traits "is nearly equivalent to" pleiotropic loci affecting both traits (Lande, 1984). Therefore, genetic correlations between traits can be approximated using previously elucidated pleiotropic models under certain conditions (Lande, 1980, 1984; Turelli, 1985). On the other hand, more recent extensions of Fisher's Geometric Model (Fisher, 1930) predict that pleiotropic mutations, compared to mutations that affect on only one trait, are less likely to be beneficial overall since a beneficial effect on one trait may be detrimental to others (Orr, 1998; Otto, 2004). The detriment of pleiotropic effects is exacerbated when increasing the strength of selection or the correlational selection between traits, since both reduce the amount of phenotypic space where mutations are beneficial. This detriment is not present for linked loci affecting separate traits since their beneficial mutations will not have the collateral effects of pleiotropy. These, therefore suggest that linkage and pleiotropy may have differential effects on genetic variance and covariances depending on selection regimes but this comparison was not fully explored in any previous model.

Lande (1984) predicted that when loci affecting *different* traits are tightly linked, and there is strong correlational selection between traits, recombina-

tion rates between loci affecting different traits can strongly affect genetic correlations between traits, when selection is weak. In an extreme case where there is complete linkage between pairs of loci affecting different traits (the recombination rate is 0), and no linkage between sets of these pairs of linked loci (the recombination rate is 0.5), then he determined that the maximum genetic correlation due to linkage may be almost as large as the extent of correlational selection, which can be calculated from the (per linkage group) genetic covariance between traits and the genetic variances, respectively, as:

$$genetic\,covariance\ (b) = \frac{\rho\omega^2\mu\alpha^2}{2c}, \tag{1}$$

$$genetic\,variance\ (c) = \sqrt{(1+\sqrt{1-\rho^2})\omega^2\frac{\mu\alpha^2}{2}}, \tag{2}$$

where ρ is the extent of correlational selection acting between the traits, ω^2 is the strength of selection (with lower values representing stronger selection), μ is the per-locus mutation rate, and α^2 is the per-locus mutation variance. If there is equal variances among traits then the genetic correlation is calculated as:

$$genetic\,correlation = \frac{b}{c} = \frac{\rho}{1+\sqrt{1-\rho^2}}. \tag{3}$$

From these equations we see that, even in the absence of pleiotropy, genetic correlation may arise from linkage disequilibrium, and depends on both the strength of correlational selection between traits and selection on each trait, as well as on the mutational inputs (mutation rates and mutational variances) of the genes affecting those traits. Lande goes on further to state that the case of complete linkage between pairs of loci affecting different traits is "equivalent to a lesser number of loci with pleiotropic effects", but this is not quantified nor is the scaling of the two examined. We seek to quantify the equivalence of pleiotropy and linkage in their ability to maintain equilibrium levels of genetic (co)variation under the same conditions. We also wish to extend this to look at a range of linkage distances, selection variances and

correlations, and mutation rates and variances, to look at the relative effects of each.

Additionally, levels of trait genetic covariation can be influenced by other evolutionary processes that affect allele frequencies, and the covariation of allelic values in a population (e.g., migration (Guillaume and Whitlock, 2007), drift (Griswold et al., 2007), inbreeding (Lande, 1984), and phenotypic plasticity (Draghi and Whitlock, 2012)). Migration affects genetic covariation because when it is sufficiently high (relative to selection in the focal population), then combinations of alleles coming from a source population will also be maintained in the focal population. This can lead to higher genetic covariation between traits in the focal populations, whether the combinations of alleles immigrating are (more likely to be) correlated in their effects on those traits or not (Guillaume and Whitlock, 2007). Migration may also have different effects depending on whether the genetic architecture is pleiotropic or made up of linked loci, but this has not been explored.

Here, we are interested in the conditions in which pleiotropic architectures behave similarly or differently to architectures with tight physical linkage between loci affecting different traits, with respect to their effects on genetic correlations between the traits. We use computer simulations to investigate whether the evolutionary forces known to affect genetic correlations between traits are dependent on the type of genetic architecture, and how. We then investigate the relative contributions of selection, mutation and migration on the genetic correlation between traits with different genetic architectures. We show that unless mutation rates are high, genetic architectures with tight linkage between loci maintain much lower equilibrium genetic correlations than pleiotropic architectures. Even when mutation rates are high, other evolutionary forces affecting equilibrium levels of genetic correlation still show a difference between architectures but to a much lesser extent. Additionally, we simulate genomic single-nucleotide polymorphism (SNP) data sets using the different architectures, and show that linkage distances affect

false positive proportions in GWA analyses.

Methods

We modeled four different genetic architectures in a modified version of the individual-based, forward-in-time, population genetics simulation software NEMO (Guillaume and Rougemont, 2006; Chebib and Guillaume, 2017). NEMO was modified to allow single non-pleiotropic loci to affect different quantitative traits. To compare how pleiotropy and linkage differentially affect the genetic correlation between traits, we modeled three different sets of genetic architecture by varying the distance between 120 pairs of additive loci affecting two quantitative traits (Figure 1). Each pair of loci was located on its own chromosome (i.e., unlinked to other pairs) and the recombination distance between each pair on a chromosome was either 0cM, 0.1cM, or 1cM apart for a particular genetic architecture representing no recombination between linked loci, as well as an average and an extreme value of recombination at "hotspots" in the human genome, respectively (Myers et al., 2006). A fourth set of genetic architecture consisted of 120 unlinked, additive, pleiotropic loci that affected both quantitative traits.

Unless otherwise specified, each simulation was run with 5,000 initially monomorphic (variation is gradually introduced through mutations), diploid individuals for 10,000 generations achieving mutation-selection(-migration) balance in order to observe general patterns of genetic correlation in the near-absence of drift. Individuals were hermaphrodites mating at random within a population, with non-overlapping generations. Phenotypes were calculated for each of the two traits modeled by summing the allelic values of all loci affecting one trait. Gaussian stabilizing selection was applied and determined the survival probability of juveniles, whose fitness was calculated as $w = \exp\left[-\frac{1}{2}\left((\mathbf{z}-\theta)^{\mathrm{T}} \cdot \mathbf{\Omega}^{-1} \cdot (\mathbf{z}-\theta)\right)\right]$, where $\mathbf{z}$ is the individual phenotype vector (initialized to the optimum values), θ is the vector of local optimal trait values (set to 10 for both traits in the focal population), and $\mathbf{\Omega}$ is the

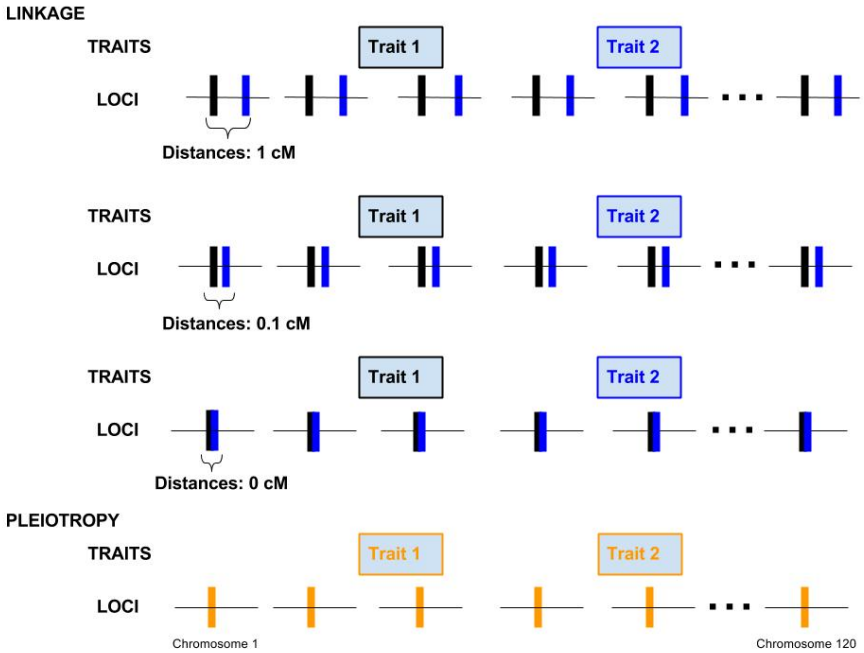

Figure 1: Four genetic architectures showing the distribution of loci on 120 chromosomes. In the case of linkage architectures, pairs of loci affecting the two different traits on each chromosome are either 1, 0.1 or 0 cM apart. In the case of the pleiotropic architecture, each locus on each chromosome affects both traits.

selection variance-covariance matrix ($n \times n$, for n traits) describing the multivariate Gaussian selection surface. To examine the effects of the strength of stabilizing selection on each trait and strength of correlational selection between traits, different sets of simulations were run with the diagonal elements of the $\mathbf{\Omega}$ matrix set as $\omega^2 = 50$, or 100 (selection strength), and off-diagonal set to $\omega^2 \times 0.5$ or $\omega^2 \times 0.9$ (selection covariance). The strength of selection scales inversely with ω^2 where a value of 100 corresponds to weak selection as opposed to correlational selection, ρ_ω, where a value of 0.9 corresponds to strong correlational selection between traits (Lande, 1984).

To examine the effects of mutational input on genetic correlation between traits, different sets of simulations were run with mutation rates (μ) of 0.001, 0.0001, or 0.00001, and mutational effect sizes (α^2) of 0.1, 0.01, or 0.001.

Mutational effects at each locus were drawn from a univariate normal distribution (with a mean of zero) in the three sets of genetic architecture with linked loci or a bivariate normal distribution (with means of zero and a covariance of 0) in the fourth set with pleiotropic architecture. Mutational effects were then added to the existing allelic values (continuum-of-alleles model; Crow and Kimura, 1964). All loci were assumed to have equal mutational variance and no environmental effects on the traits was included.

To examine the effects of migration from a source population on genetic correlation between traits, different sets of simulations were run with unidirectional migration from a second population with backward migration rates (m) of 0.1, 0.01, and 0.001. The backward migration rate represents the average proportion of new individuals in the focal population whose parent is from the source population. The local optimum values for the two traits in the source population were set at $\theta = \left[\sqrt{50}, \sqrt{50}\right]$ (10 units distance from the focal population's local optimum). Both focal and source populations had weak stabilizing selection with a strength of $\omega^2 = 100$, the focal population had no correlational selection between the two traits and the source population had a correlational selection of $\rho_\omega = 0$ or 0.9. Fifty replicate simulations were run for each set of parameter values and statistics were averaged over replicates. Averages were also compared against analytical expectations laid out by Lande (1984) and reproduced here in Equations 1–3.

Effects of genetic architecture on false positive/negative proportions in association studies

In order to elucidate the differential effects of pleiotropy and linkage on the detection of true causal genetic variants in association studies, a genome-wide association (GWA) analysis was performed on data simulated as described above, except that diallelic loci were used instead of a continuum-of-alleles model to better represent SNPs. Correlational selection values were chosen that provided equal on-average genetic correlations between traits for

all genetic architectures of 0.3 and 0.4, values frequently observed in both morphological and life-history traits (Roff, 1996). In the association study, a per-locus regression of trait values was performed over genotypes, and the (negative log 10) p-values of regression slopes were plotted with a Benjamini-Hochberg False Discovery Rate (FDR) cutoff to adjust significance levels for multiple tests (Benjamini and Hochberg, 1995). From this, we observed the number (and proportion) of false positives (linked loci that had no effect on a trait but whose regression slope p-values were above the FDR cutoff for that same trait) and false negatives (pleiotropic loci that had an effect on both traits but whose regression slope p-values were below the FDR cutoff for either trait). Linkage disequilibrium values of $D^{'}$ and R^2 between pairs of linked traits were also calculated using the **R** package genetics (v1.3.8.1) (Warnes et al., 2013). Statistics for number and proportion of false positives and negatives were obtained from the average over 20 replicate simulations of each genetic architecture.

Results

Effects of genetic architecture on genetic correlation

By generation 10,000, when mutation-selection balance is reached, simulations with the pleiotropic architecture maintain a higher average genetic correlation at equilibrium than those with linkage architectures, even when recombination is absent (linkage distance of 0cM between pairs of loci) (Figure 2). A higher genetic correlation with pleiotropy compared to 0cM linkage is the result of lower total genetic variation, since they maintain the same level of genetic covariance between the two traits. As the distance between linked loci increases, the average genetic correlation it maintains decreases.

Effects of selection strength and correlational selection on genetic correlation

In all simulations, an increase in selection strength decreases total variance and genetic covariance, but slightly increases genetic correlation since

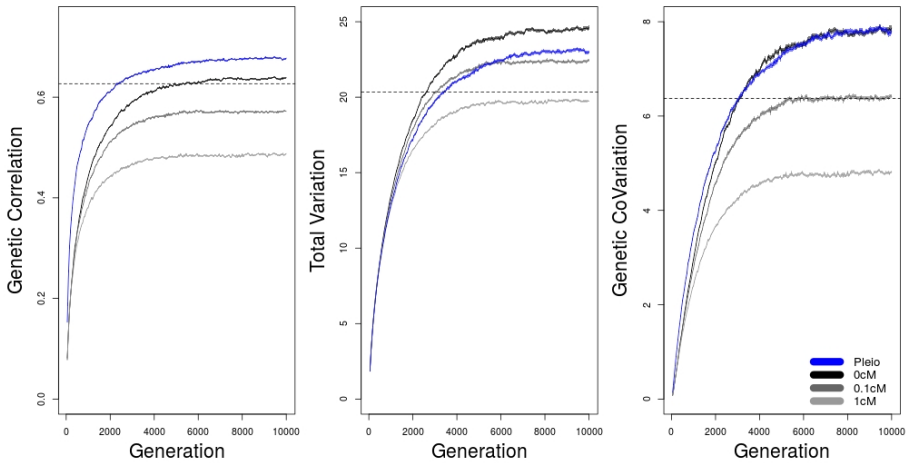

Figure 2: Average genetic correlation, total genetic variation and genetic covariation (and their standard deviations) over 10,000 generations reaching mutation-selection equilibrium for four different genetic architectures: pairs of linked loci affecting two different traits with 0, 0.1 or 1cM between loci, or pleiotropic loci affecting both traits. $N = 5000$, $\omega^2 = 100$, $\rho_\omega = 0.9$, $\alpha^2 = 0.1$, and $\mu = 0.001$. Dashed line represents Lande's 1984 expectations for completely linked loci (0 cm).

total variance decreases to a larger extent than genetic covariance (Figure 3). A decrease in correlational selection increases total variation and decreases genetic covariance and genetic correlation, regardless of genetic architecture (Figure 4). This increase in total variation is more pronounced in simulations with genetic architecture having looser linkage. A decrease in correlational selection from $\rho_\omega = 0.9$ to 0.5 has a larger effect on maintaining strong genetic correlations than a decrease in selection strength from $\omega^2 = 50$ to 100. In the absence of correlational selection little to no genetic correlation persists, regardless of genetic architecture (not shown).

Effects of mutation rate and mutational variance on genetic correlation

A smaller mutation rate leads to a lower equilibrium average genetic correlation in all populations, regardless of genetic architecture (Figure 5). This decrease in mutation rate has a larger effect on genetic correlation when genetic architecture is made up of linked pairs of loci compared with the

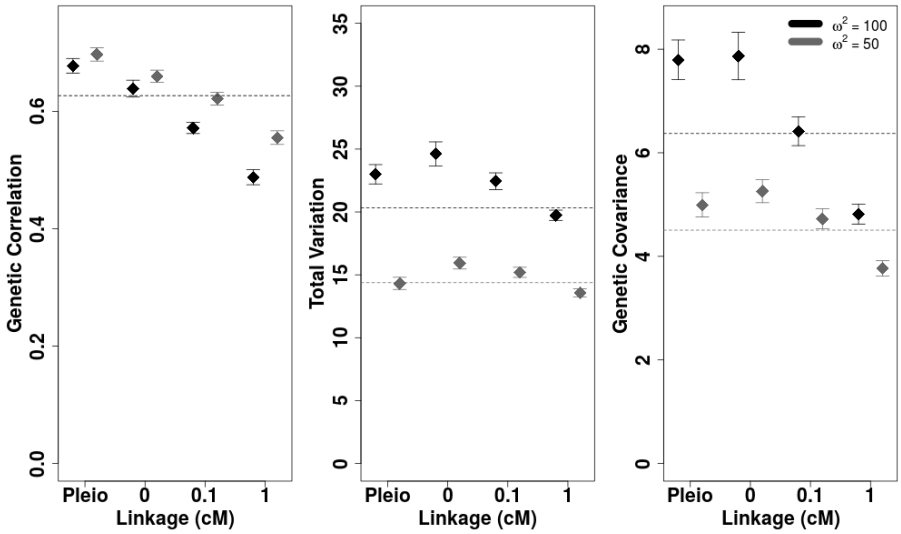

Figure 3: Effect of selection variance (ω^2) on average genetic correlation, total variance and genetic covariance (and their standard deviations) after 10,000 generations of correlated, stabilizing selection for four different genetic architectures. Dashed lines represents Lande's 1984 expectations for completely linked loci (0 cm).

pleiotropic genetic architecture. With lower mutation rate there is also a lower total genetic variance and lower genetic covariance. The higher genetic correlation in the pleiotropic genetic architecture compared to architectures with linkage is due to a lower total genetic variance when the mutation rate is high ($\mu = 0.001$), but due to higher genetic covariance when mutation rate is lower ($\mu = 0.0001$ or 0.00001).

A smaller mutational variance leads to a lower equilibrium average genetic correlation in all populations regardless of genetic architecture, but the reduction is less drastic than that of mutation rate decrease (Figure 6). A decrease in mutational variance leads to a decrease in genetic correlation by a similar amount regardless of genetic architecture (though loose linkage is affected the most). Populations with linkage architectures need both high mutation rates and high mutational variance to maintain strong genetic correlation, whereas the pleiotropic architecture just needs high mutational variance.

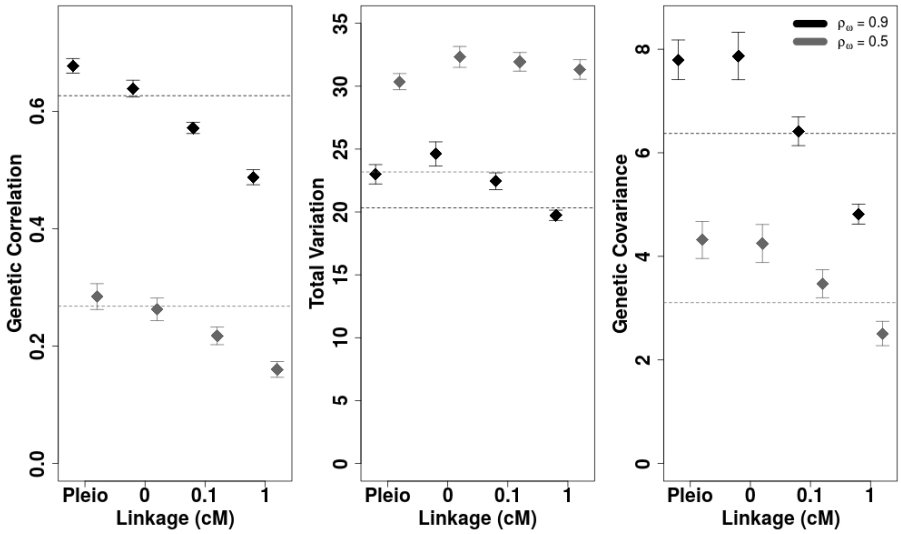

Figure 4: Effect of correlational selection (ρ_ω) on average genetic correlation, total variance and genetic covariance (and their standard deviations) after 10,000 generations of correlated, stabilizing selection for four different genetic architectures. Dashed lines represents Lande's 1984 expectations for completely linked loci (0 cm).

Effects of migration on genetic correlation

A higher migration rate from a source population, whose traits are under correlational selection, leads to higher genetic correlations in the focal population than with no migration regardless of the genetic architecture (Figure 7A). The effect of migration increases with tighter linkage and is highest with pleiotropic architecture. This effect on genetic correlation is still observed when there is no correlational selection on the traits in the source population, but to a largely reduced degree (Figure 7B).

Effects of linkage and pleiotropy on proportion of false positives/negatives and linkage disequilibrium in multi-trait GWASes

In simulations where there is linkage between SNPs and equivalent levels of genetic correlation between traits, the number and proportion of loci that are false positives (above FDR cutoff but no effect on trait) increase as linkage distance decreases between SNPs affecting different traits (shown in

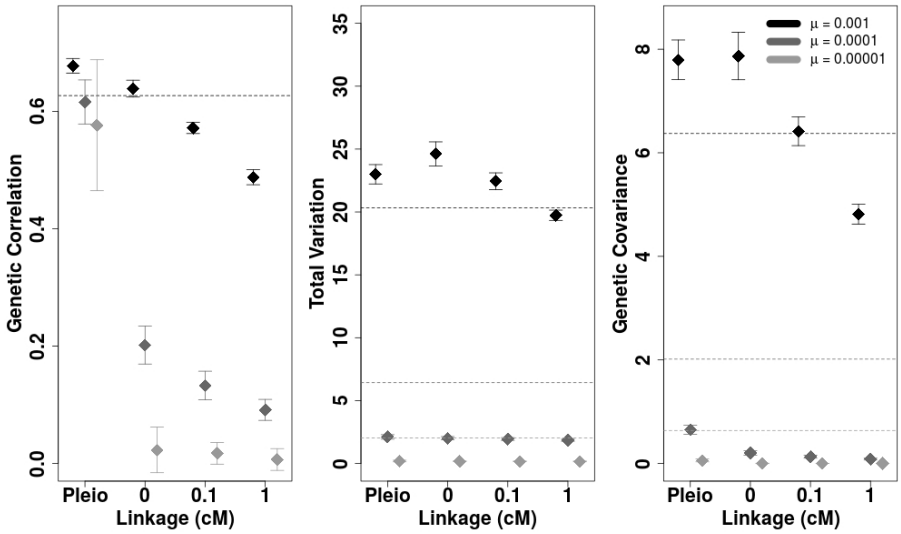

Figure 5: Effect of mutation rate (μ) on average genetic correlation, total variance and genetic covariance (and their standard deviations) after 10,000 generations of correlated, stabilizing selection for four different genetic architectures. Dashed lines represents Lande's 1984 expectations for completely linked loci (0 cm).

Figure 8 and summarized in Table 1). When genetic correlation is higher (due to stronger correlational selection), linkage distance has a greater impact on the proportion of false positives. In simulations where SNPs are pleiotropic, genetic correlation due to correlational selection has little impact on the number and proportion of false negatives (below FDR cutoff but does affect the traits). Linkage disequilibrium between pairs of linked SNPs decreases as distance between SNPs increases regardless of genetic correlation (Table 1 and Supplemental Figure S1). Long-distance linkage disequilibrium between unlinked SNPs increases when the distance between pairs of linked SNPs increases (when measured with $D^{'}$), and is higher with higher genetic correlation when comparing the same genetic architectures (Supplemental Figure S2). In simulations where SNPs are pleiotropic, long-distance linkage disequilibrium does not seem to be affected by change in genetic correlation.

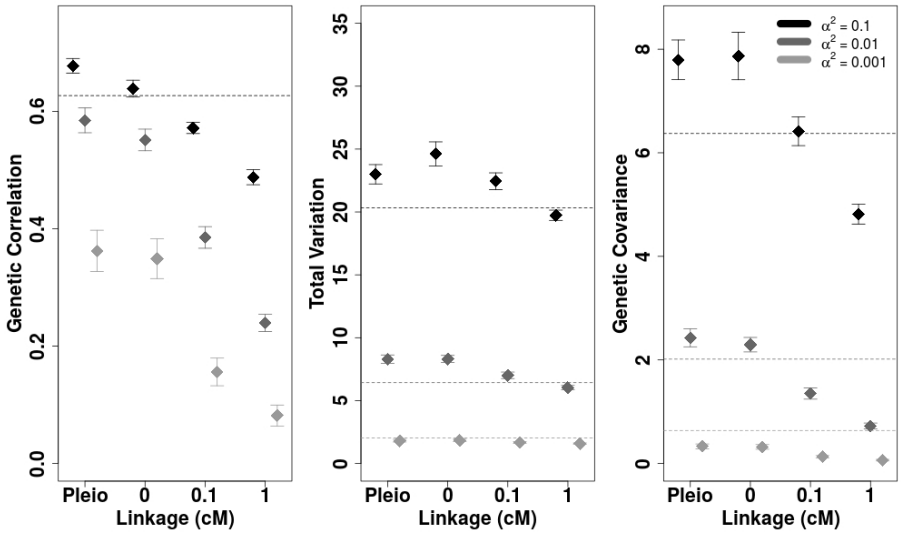

Figure 6: Effect of mutation variance (α^2) on average genetic correlation, total variance and genetic covariance (and their standard deviations) after 10,000 generations of correlated, stabilizing selection for four different genetic architectures. Dashed lines represents Lande's 1984 expectations for completely linked loci (0 cm).

Discussion

Pleiotropy and linkage are not the same

The main expectation under an assumption of weak selection and strong correlational selection is that populations with a genetic architecture consisting of unlinked pairs of two completely linked loci (0cM distance) should maintain similar equilibrium levels of genetic correlation as with a genetic architecture consisting of a lesser number of unlinked pleiotropic loci (Lande, 1984). Our results show that this is the case when there are half as many pleiotropic loci and mutation rates are relatively high. A high rate of mutation (10^{-3}) allows for multiple mutations in both loci in a tightly linked pair to accumulate and maintain levels of genetic covariance near to that of mutations in a single pleiotropic locus, but empirical estimations of mutation rates from varied species like bacteria and humans suggests that *per-nucleotide* mutation rates are in the order of 10^{-8} to 10^{-9} (Nachman and

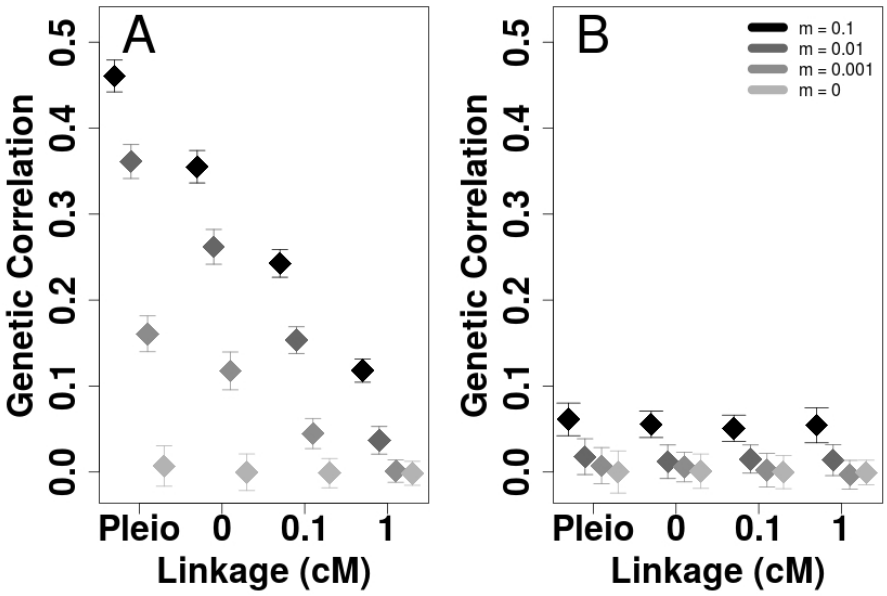

Figure 7: Average genetic correlations in the focal populations (and their standard deviations) after 10,000 generations of migration from a source population with different migration rates (m) for four different genetic architectures. A– Migration from a source population with correlational selection between traits ($\rho_\omega = 0.9$). B– Migration from a source population without correlational selection between traits ($\rho_\omega = 0$).

Crowell, 2000; Lynch, 2010; Ford et al., 2011). If a polygenic locus consists of hundreds or thousands of nucleotides, as in the case of many quantitative trait loci (QTLs), then per-locus mutation rates may be as high as 10^{-5}, but the larger the locus the higher the chance of recombination between within-locus variants that are contributing to genetic correlation. This leads us to believe that with empirically estimated levels of mutation and recombination, strong genetic correlation between traits are more likely to be maintained if there is an underlying pleiotropic architecture affecting them than will be maintained due to tight linkage. Consequently, GWASes that detect associations between multiple traits and single genetic variants are more likely to be detecting pleiotropic loci than linked loci. Also, previous theoretical models suggest that Lande's (1984) equilibrium levels of genetic variation are not well approximated at low per-locus mutation rates (compared to the

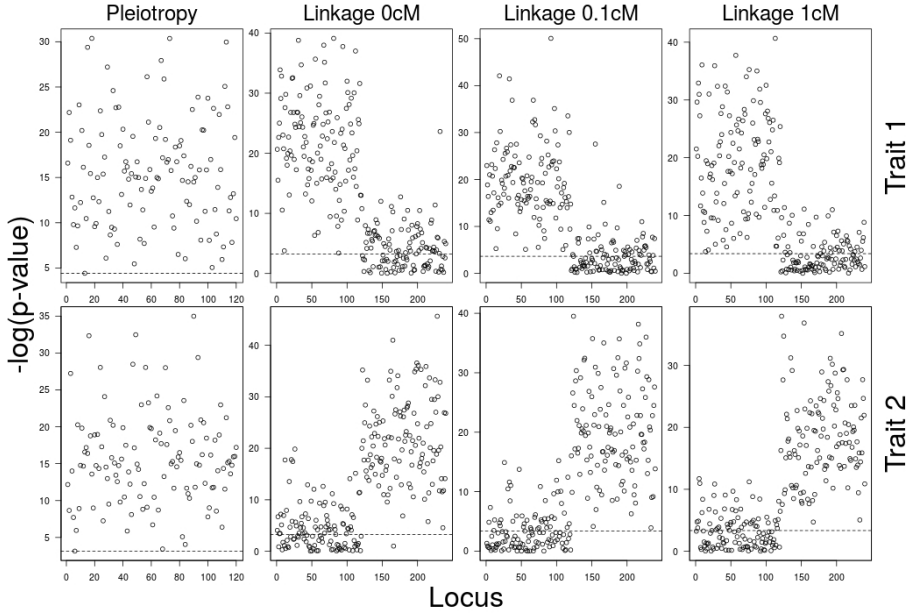

Figure 8: GWA analysis: -log(p-values of slope of regression of trait values on genotypes) from one set of example simulations. In the case of linkage architectures, the first 120 loci only affected trait 1 and the next 120 loci only affected trait 2. In the case of the pleiotropic architecture, all 120 loci affected both traits. The average genetic correlation of ≈ 0.3 was observed by adjusting the correlational selection levels to 0.88, 0.89, 0.93, and 0.965 for pleiotropy, linkage 0cM, linkage 0.1cM, and linkage 1cM, respectively. Dashed lines represent the Benjamini-Hochberg FDR cutoffs for a significance level of 0.05.

strength of selection), which was also true in our simulations (Supplemental Figure S3) (Turelli, 1985; Bürger, 2000).

We find that even under scenarios where pleiotropy and tight linkage maintain similar levels of genetic covariance, pleiotropic architectures have higher genetic correlations because they have lower total genetic variance. This can be explained by understanding the differential fitness effects of loci. Mutations that affect more than one trait are less likely to be beneficial (Orr, 1998; Otto, 2004). The distribution of fitness effects of pleiotropic mutations is shifted towards more negative average values as the number of traits affected increases (Martin and Lenormand, 2006; Chevin et al., 2010). Hence, pleiotropic architectures that affect more traits have less positive mutational

Table 1: Results of GWA analyses for different architectures with average false negatives (Type II errors) for pleiotropic architectures and false positives (Type I errors) for linkage architectures, as well as linkage disequilibrium (LD) measurement averages for short-distance (physically linked loci) and long-distance (unlinked loci) comparisons.

Genetic Architecture	Genetic Cor (SE)	Type I/II Error %	D' short	D' long	R^2 short	R^2 long
Pleiotropy	0.308 (0.0046)	0.35%	NA	0.018	NA	0.00027
Linkage (0cM)	0.300 (0.0055)	22.06%	0.37	0.023	0.089	0.00026
Linkage (0.1cM)	0.300 (0.0045)	20.17%	0.26	0.025	0.047	0.00027
Linkage (1cM)	0.308 (0.0035)	18.28%	0.13	0.030	0.007	0.00027
Pleiotropy	0.407 (0.0048)	0.32%	NA	0.018	NA	0.00027
Linkage (0cM)	0.398 (0.0074)	28.76%	0.43	0.025	0.107	0.00027
Linkage (0.1cM)	0.408 (0.0035)	28.46%	0.30	0.027	0.050	0.00027
Linkage (1cM)	0.404 (0.0029)	25.34%	0.19	0.048	0.006	0.00027

effects on fitness and maintain a lower equilibrium genetic variation when compared to linked architectures (Turelli, 1985). It has been suggested that this might be overcome in more complex organisms with a greater number of traits by modularization of the effects of different pleiotropic genes to separate sets of traits and decrease the pleiotropic degree of the mutations but theoretical models have shown mixed results (Baatz and Wagner, 1997; Hansen, 2003; Welch et al., 2003; Martin and Lenormand, 2006; Chevin et al., 2010; Wagner and Zhang, 2011).

When correlational selection between traits is strong in the simulations with linked architectures, the genetic correlation is dependent on the recombination rates between loci affecting different traits where tightly linked loci can maintain higher levels of genetic correlation from a build-up of positive linkage disequilibrium than loosely linked loci. This matches the analytical predictions put forth in Lande (1984) under the assumption of weak stabilizing selection, strong correlational selection, and loose linkage between loci affecting the same trait. Strong stabilizing selection combined with strong correlational selection, on the other hand, leads to an increased build-up of negative LD as extreme combinations of values between pairs of loci are ex-

cluded and small but opposite-sign values at paired loci remain. This has a disproportionately larger reduction of genetic covariances in more tightly linked loci, a scenario that was not explored in Lande (1984).

The impact of pleiotropy and linkage maintaining different genetic correlations in association studies

When methods like GWA analyses are employed to detect shared genetic influences (pleiotropy or linkage) on multiple traits of interest, they are dependent upon detecting combinations of effect sizes of genetic variants associated with those traits (Hill and Zhang, 2012b,a; Chung et al., 2014; Visscher and Yang, 2016). The success or failure of this endeavor is directly connected to the ability to detect loci with associations to each trait and the strength of genetic correlation between traits (Wei et al., 2014; Pickrell et al., 2016; Chesmore et al., 2018; Verbanck et al., 2018). Our results show that (tight) linkage between loci affecting different traits will lead to "many" false positives. Therefore, GWASes will not be able to distinguish between pleiotropy and linkage, empirically. The proportion of genes associated with two or more phenotypes in the GWAS catalog has increased to around 40% in the last decade (Welter et al., 2013; Pickrell et al., 2016). But it is difficult to determine if this is truly representative of the prevalence of pleiotropy because QTLs are often mapped to loci that can encompass thousands of nucleotides (and more than one gene) and informative SNPs with significant effect sizes are assigned to the closest genes with annotated phenotypes (Chesmore et al., 2018; Liu et al., 2019). Conflating inter-genic pleiotropic SNPs with nearby pleiotropic genes (or loci) can distort the prevalence of pleiotropy and reduce the ability to distinguish pleiotropy and physical linkage, which in turn can reduce the efficacy of treatments dependent on those results (like gene targeted therapies) (Dudley et al., 2005; Gianola et al., 2015). Finding out the true false positive rate in GWA studies due to linkage is difficult because it is almost never known whether the source of genetic correlations between traits is linked loci or not, even when fine-scale sequences are available (for the

reasons mentioned above and because of the way pleiotropy is erroneously defined in GWA studies) (Platt et al., 2010). Watanabe et al. (2018) attempted to break down this issue in a meta-analysis of 558 GWASes by looking at the proportion of genomic loci, genes, and SNPs associated with multiple traits, which may provide a clearer picture of the prevalence of pleiotropic causal variants. They found that 93.3% of loci, 81.0% of genes, and 60.2% of SNPs, were associated with more than one trait. This may seem to provide a better estimate of pleiotropic levels, except that in this study SNPs that were associated with more than one trait could still have been the result of linkage disequilibrium. A point that was brought up by the authors.

When there is tighter linkage between loci affecting separate traits, there is also a higher proportion of loci that will be statistically associated with traits that they do not affect (false positives). This is expected (for the same level of genetic correlation between traits), since the power to detect a locus linked to a causal variant is proportional to the correlation coefficient between loci (given by their linkage disequilibrium) (Siegmund and Yakir, 2007). On the other hand, very few false negatives in pleiotropic loci were observed (regardless of genetic correlation) because we "sampled" the entire population and therefore had the power to find significant associations with (almost) all causal loci. Had we taken smaller samples of our population to perform the GWA analysis, we would have found a greater number of false negatives. The salient consequence is that study design, threshold levels, and genetic correlations between traits will all affect detection of genetic variants, whether the variants are causal themselves or linked to causal variants (Wagner and Zhang, 2011; Hill and Zhang, 2012a). Also, the number of pleiotropic effects a locus has may be under-represented by significance levels in association studies (Hill and Zhang, 2012b). Wagner and Zhang 2011 go a step further to suggest that number or proportions of traits affected may not be as meaningful as describing the distributions of pleiotropic effect sizes on traits.

There is a difference between pleiotropy and linkage at the nucleotide level

Transgenic experiments may differentiate pleiotropy from linkage at the gene level (Mills et al., 2014), but at the nucleotide level does the distinction between two linked loci and one pleiotropic locus go away? There is evidence that even in the same gene, adjacent polymorphisms affecting different traits in *Drosophila* can be in linkage equilibrium due to fine-scale recombination (Carbone et al., 2006; Flint and Mackay, 2009). But imagine a case where a mutation in a single base-pair has an effect on one trait and a mutation in the base-pair right next to the first base-pair has an effect on a second trait. Now imagine a second case where a mutation in a single base-pair has an effect on two traits. There still seems to be a distinction between these two cases because the probability of a change in both traits in the first case is the mutation rate squared compared to the second case where the probability of a change in both traits is just the mutation rate. Depending on the per-locus mutation rate this difference can be quite large (e.g. 10^{-4} versus 10^{-16}). Even in this extreme case, there may indeed still be a gray area in the distinction between pleiotropy and linkage at a mutational level. Mutations may affect the pleiotropic degree (e.g. like enzyme specificity) of a protein-coding gene and the degree to which the gene maintains multi-functionality may itself evolve (Guillaume and Otto, 2012). If there is correlational selection between the catalytic functions of an enzyme, then some pleiotropic mutations that affect more than one catalytic ability will be favoured, and genetic correlations will increase. With this in mind, it makes more sense from a theoretical and functional standpoint to refer to pleiotropy at the nucleotide level (or at the unit of a mutation), than at the gene or larger locus level (but this may depend on the questions of interest (Rockman, 2012; Rausher and Delph, 2015)).

Other factors

Even in the absence of correlational selection it is possible to maintain genetic correlation through continued migration from a source population.

High migration brings individuals whose combination of alleles will expand focal population variation in the direction of the source population. This corroborates previous results that showed that slow introgression of allelic combinations into a population can affect the genetic variance-covariance structure of that population (Guillaume and Whitlock, 2007). Whether genetic covariance will be maintained in real populations depends on the nature of correlational selection on traits in the population of interest, since migration can reduce local fitness (i.e. migration load) if allele combinations are not favoured by selection or increase it if they are (Nosil et al., 2006; Bolnick and Otto, 2013). Migration into a population will also affect false positive rates since immigrating allele combinations will be in LD from the source population and will therefore increase the proportion of certain genotypes, even if there is no strong trait correlation in the source population. Although not investigated in this study, a structured population and/or a continual system of inbreeding in a population where there is correlational selection between polygenic traits can result in increased genetic covariation caused by larger LD Lande (1984), which can in turn increase false positive proportions.

Conclusion

Pleiotropic loci maintain stronger genetic correlations between traits than linked loci affecting different traits even when no recombination occurs between the loci, and especially in the magnitude of empirically estimated mutation rates. Previous models of the maintenance of genetic covariation at mutation-selection equilibrium describe genetic covariation as a function of the product of mutation rate and variance. These models provide similar expectations for pleiotropic and tight linkage architectures. The discrepancy occurs because of the contingency of mutational covariance input on the occurrence of mutations (and hence mutation rate). Without high mutation rates, the ability to create genetic covariance between linked loci is highly diminished because the combined likelihood of mutations in each linked loci

with both mutational effects in the same direction is low. This result will have implications in the type of underlying architecture we expect to find in multi-trait association studies. On the one hand, tighter linkage between causal loci and detected loci maintains higher genetic correlations, leading to a greater proportion of causal variant false positives. More importantly, variants are more likely to have pleiotropic effects on traits than linked effects, when they are found to be associated with multiple, strongly correlated traits.

Supplemental

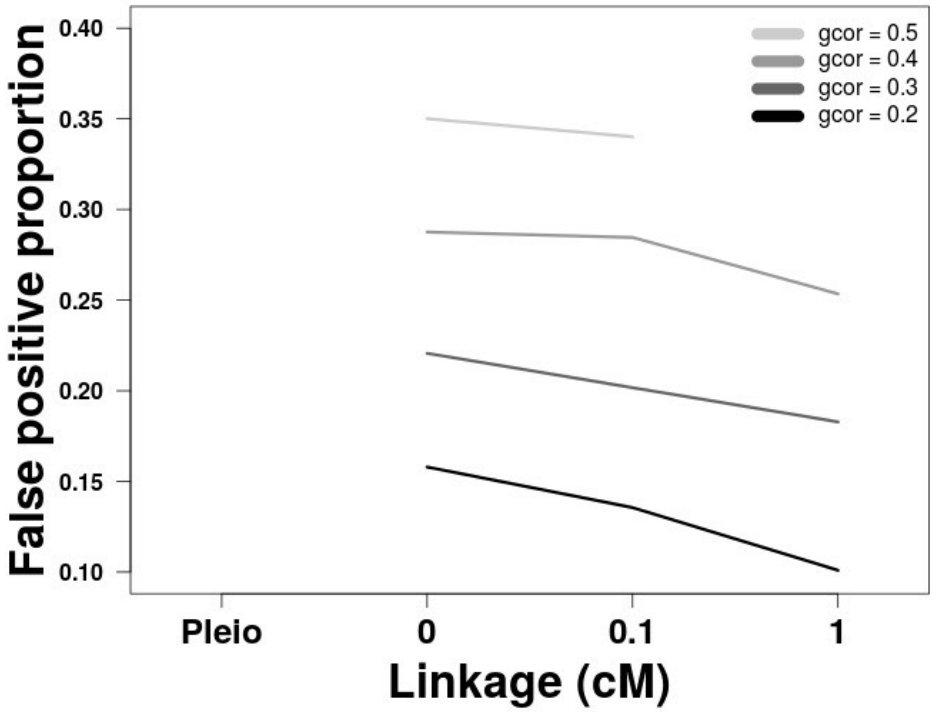

Figure S1: Average proportion of false positives among replicates for different genetic correlations and genetic architectures. N.B. No false positives were possible for pleiotropic architectures.

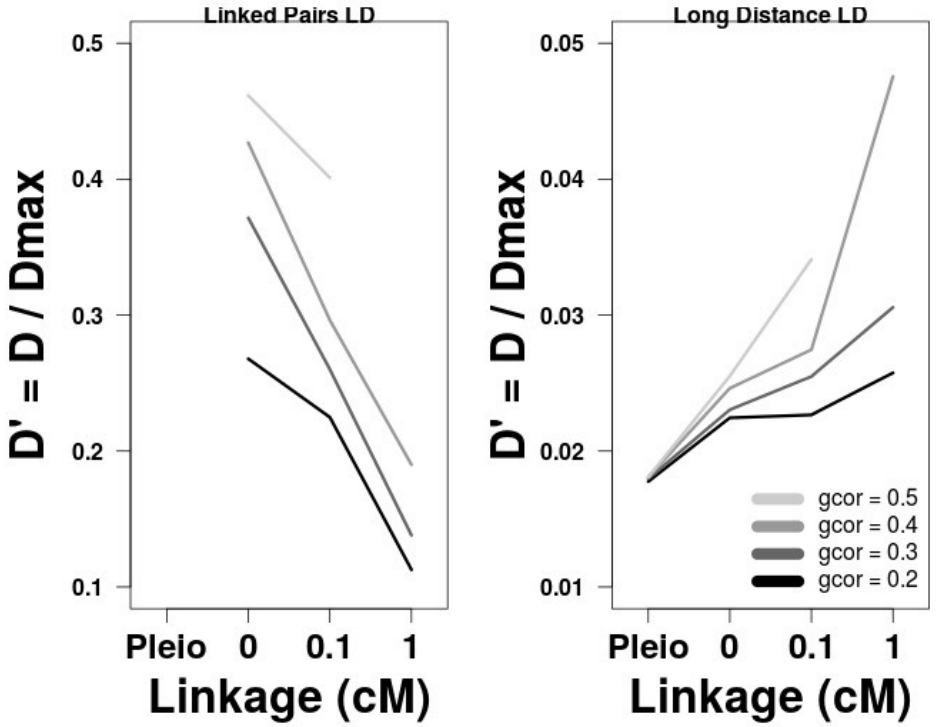

Figure S2: Average linkage disequilibrium (measured by $D^{'}$) between linked pairs (left panel) and between unlinked pairs (right panels) for different genetic correlations and genetic architectures. N.B. No linked pairs existed between pleiotropic loci.

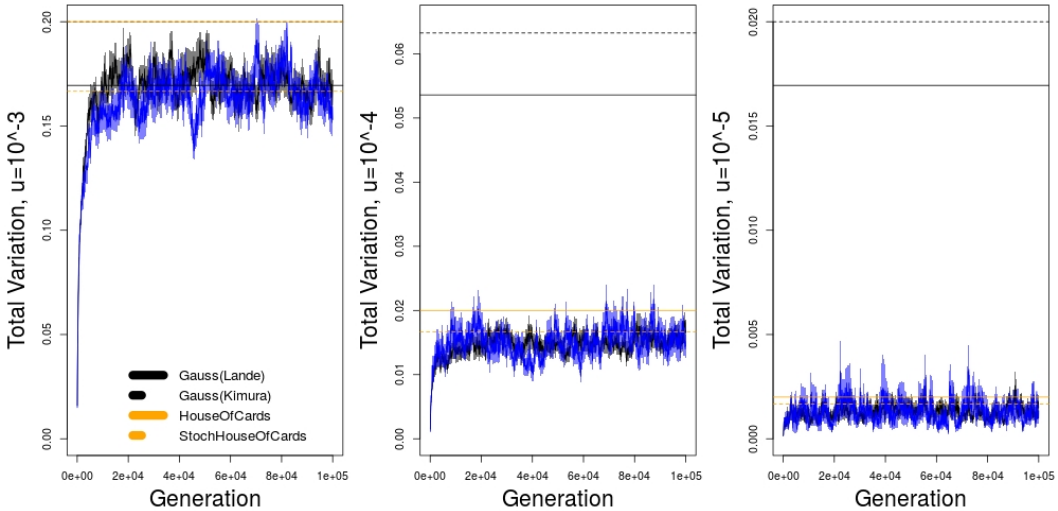

Figure S3: Average genetic variances for different mutation rates and genetic architectures, with either one pleiotropic locus or two completely linked loci, compared against theoretical expectations from several models (Bürger, 2000).

References

Arnold, S. J. (1992). Constraints on phenotypic evolution. *Am. Nat.*, pages S85–S107.

Baatz, M. and Wagner, G. P. (1997). Adaptive inertia caused by hidden pleiotropic effects. *Theor. Popul. Biol.*, 51(1):49–66.

Barrett, R. D. and Hoekstra, H. E. (2011). Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, 12(11):767.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300.

Bolnick, D. I. and Otto, S. P. (2013). The magnitude of local adaptation under genotype-dependent dispersal. *Ecology and evolution*, 3(14):4722–4735.

Bürger, R. (2000). *The mathematical theory of selection, recombination, and mutation*, volume 228. Wiley Chichester.

Cai, L., Fisher, A. L., Huang, H., and Xie, Z. (2016). Crispr-mediated genome editing and human diseases. *Genes & Diseases*, 3(4):244–251.

Carbone, M. A., Jordan, K. W., Lyman, R. F., Harbison, S. T., Leips, J., Morgan, T. J., DeLuca, M., Awadalla, P., and Mackay, T. F. (2006). Phenotypic variation and natural selection at catsup, a pleiotropic quantitative trait gene in drosophila. *Current Biology*, 16(9):912–919.

Chebib, J. and Guillaume, F. (2017). What affects the predictability of evolutionary constraints using a G-matrix? The relative effects of modular pleiotropy and mutational correlation. *Evolution*, 71(10).

Chesmore, K., Bartlett, J., and Williams, S. M. (2018). The ubiquity of pleiotropy in human disease. *Human genetics*, 137(1):39–44.

Chevin, L.-M., Martin, G., and Lenormand, T. (2010). Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution*, 64(11):3213–3231.

Chung, D., Yang, C., Li, C., Gelernter, J., and Zhao, H. (2014). Gpa: a statistical approach to prioritizing gwas results by integrating pleiotropy and annotation. *PLoS genetics*, 10(11):e1004787.

Crow, J. F. and Kimura, M. (1964). The theory of genetic loads. *Proc. XI Int. Congr. Genetics*, 2:495–505.

Draghi, J. A. and Whitlock, M. C. (2012). Phenotypic plasticity facilitates mutational variance, genetic variance, and evolvability along the major axis of environmental variation. *Evolution*, 66(9):2891–2902.

Dudley, A. M., Janse, D. M., Tanay, A., Shamir, R., and Church, G. M. (2005). A global view of pleiotropy and phenotypically derived gene function in yeast. *Molecular systems biology*, 1(1).

Edelstein, M. L., Abedi, M. R., and Wixon, J. (2007). Gene therapy clinical trials worldwide to 2007an update. *The Journal of Gene Medicine: A cross-disciplinary journal for research on the science of gene transfer and its clinical applications*, 9(10):833–842.

Falconer, D. S. and Mackay, T. F. (1996). *Introduction to quantitative genetics*. Longman, London, fourth edition.

Fisher, R. A. (1930). *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press.

Flint, J. and Mackay, T. F. (2009). Genetic architecture of quantitative traits in mice, flies, and humans. *Genome research*, 19(5):723–733.

Ford, C. B., Lin, P. L., Chase, M. R., Shah, R. R., Iartchouk, O., Galagan, J., Mohaideen, N., Ioerger, T. R., Sacchettini, J. C., Lipsitch, M., et al.

(2011). Use of whole genome sequencing to estimate the mutation rate of mycobacterium tuberculosis during latent infection. *Nature genetics*, 43(5):482–486.

Gianola, D., de los Campos, G., Toro, M. A., Naya, H., Schön, C.-C., and Sorensen, D. (2015). Do molecular markers inform about pleiotropy? *Genetics*, pages genetics–115.

Ginn, S. L., Amaya, A. K., Alexander, I. E., Edelstein, M., and Abedi, M. R. (2018). Gene therapy clinical trials worldwide to 2017: An update. *The journal of gene medicine*, 20(5):e3015.

Gratten, J. and Visscher, P. M. (2016). Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome medicine*, 8(1):78.

Griswold, C. K., Logsdon, B., and Gomulkiewicz, R. (2007). Neutral evolution of multiple quantitative characters: a genealogical approach. *Genetics*, 176(1):455–466.

Guillaume, F. and Otto, S. P. (2012). Gene functional trade-offs and the evolution of pleiotropy. *Genetics*, 192:1389–1409.

Guillaume, F. and Rougemont, J. (2006). Nemo:an evolutionary and population genetics programming framework. *Bioinformatics*, 22:2556–2557.

Guillaume, F. and Whitlock, M. C. (2007). Effects of migration on the genetic covariance matrix. *Evolution*, 61(10):2398–2409.

Hansen, T. F. (2003). Is modularity necessary for evolvability?: Remarks on the relationship between pleiotropy and evolvability. *Biosystems*, 69(2):83–94.

Hill, W. G. and Zhang, X.-S. (2012a). Assessing pleiotropy and its evolutionary consequences: pleiotropy is not necessarily limited, nor need it hinder the evolution of complexity. *Nature Reviews Genetics*, 13(4):296.

Hill, W. G. and Zhang, X.-S. (2012b). On the pleiotropic structure of the genotype-phenotype map and the evolvability of complex organisms. *Genetics*, 190(3):1131–1137.

Hodgkin, J. (1998). Seven types of pleiotropy. *Int. J. Dev. Biol.*, 42(3):501–505.

Lande, R. (1980). The genetic covariance between characters maintained by pleiotropic mutations. *Genetics*, 94(1):203–215.

Lande, R. (1984). The genetic correlation between characters maintained by selection, linkage and inbreeding. *Genetics Research*, 44(3):309–320.

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D., and others, . (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*.

Lynch, M. (2010). Rate, molecular spectrum, and consequences of human mutation. *Proceedings of the National Academy of Sciences*, 107(3):961–968.

Lynch, M., Walsh, B., et al. (1998). *Genetics and analysis of quantitative traits*, volume 1. Sinauer Sunderland, MA.

Martin, G. and Lenormand, T. (2006). A general multivariate extension of fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution*, 60(5):893–907.

Mills, M. G., Greenwood, A. K., and Peichel, C. L. (2014). Pleiotropic effects of a single gene on skeletal development and sensory system patterning in sticklebacks. *EvoDevo*, 5(1):5.

Myers, S., Spencer, C., Auton, A., Bottolo, L., Freeman, C., Donnelly, P., and McVean, G. (2006). The distribution and causes of meiotic recombination in the human genome. *Biochem. Soc. Trans.*, 34(4):536–530.

Nachman, M. W. and Crowell, S. L. (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics*, 156(1):297–304.

Nosil, P., Crespi, B., Sandoval, C., and Kirkpatrick, M. (2006). Migration and the genetic covariance between habitat preference and performance. *The American Naturalist*, 167(3):E66–E78.

Orr, H. A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, 52(4):935–949.

Otto, S. P. (2004). Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proc. R. Soc. Lond (Biol)*, 271(1540):705–714.

Parkes, M., Cortes, A., Van Heel, D. A., and Brown, M. A. (2013). Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nature Reviews Genetics*, 14(9):661.

Peichel, C. L. and Marques, D. A. (2017). The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Phil. Trans. R. Soc. B*, 372(1713):20150486.

Pickrell, J. K., Berisa, T., Liu, J. Z., Ségurel, L., Tung, J. Y., and Hinds, D. A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nature genetics*, 48(7):709.

Platt, A., Vilhjálmsson, B. J., and Nordborg, M. (2010). Conditions under which genome-wide association studies will be positively misleading. *Genetics*.

Rausher, M. D. and Delph, L. F. (2015). Commentary: When does understanding phenotypic evolution require identification of the underlying genes? *Evolution*, 69(7):1655–1664.

Rockman, M. V. (2012). The qtn program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution: International Journal of Organic Evolution*, 66(1):1–17.

Roff, D. A. (1996). The evolution of genetic correlations: an analysis of patterns. *Evolution*, pages 1392–1403.

Saltz, J. B., Hessel, F. C., and Kelly, M. W. (2017). Trait correlations in the genomics era. *Trends in ecology & evolution*, 32(4):279–290.

Siegmund, D. and Yakir, B. (2007). *The statistics of gene mapping*. Springer Science & Business Media.

Stearns, F. W. (2010). One hundred years of pleiotropy: a retrospective. *Genetics*, 186(3):767–773.

Turelli, M. (1985). Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. *Genetics*, 111(1):165–195.

Verbanck, M., Chen, C.-Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nature genetics*, 50(5):693.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., and Yang, J. (2017). 10 years of gwas discovery: biology, function, and translation. *The American Journal of Human Genetics*, 101(1):5–22.

Visscher, P. M. and Yang, J. (2016). A plethora of pleiotropy across complex traits. *Nature genetics*, 48(7):707.

Wagner, G. P. and Zhang, J. (2011). The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.*, 12(3):204–213.

Walsh, B. and Blows, M. W. (2009). Abundant genetic variation + strong selection = multivariate genetic constraints: A geometric view of adaptation. *Annu. Rev. Ecol. Evol. Syst.*, 40(1):41–59.

Warnes, G., with contributions from Gregor Gorjanc, Leisch, F., and Man., M. (2013). *genetics: Population Genetics*. R package version 1.3.8.1.

Watanabe, K., Stringer, S., Frei, O., Mirkov, M. U., Polderman, T. J., van der Sluis, S., Andreassen, O. A., Neale, B. M., and Posthuma, D. (2018). A global view of pleiotropy and genetic architecture in complex traits. *bioRxiv*, page 500090.

Wei, W.-H., Hemani, G., and Haley, C. S. (2014). Detecting epistasis in human complex traits. *Nature Reviews Genetics*, 15(11):722.

Welch, J. J., Waxman, D., and Houle, D. (2003). Modularity and the cost of complexity. *Evolution*, 57(8):1723–1734.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2013). The nhgri gwas catalog, a curated resource of snp-trait associations. *Nucleic acids research*, 42(D1):D1001–D1006.

Wright, S. (1977). Evolution and the genetics of populations. volume 3. experimental results and evolutionary deductions. *Evolution and the genetics of populations. Volume 3. Experimental results and evolutionary deductions.*

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., Montgomery, G. W., Goddard, M. E., Wray, N. R., Visscher, P. M., et al. (2016). Integration of summary data from gwas and eqtl studies predicts complex trait gene targets. *Nature genetics*, 48(5):481.