

## Supplementary Methods

A novel analytical framework to quantify co-gradient and counter-gradient variation

M.A. Albecker, G.C. Trussell, and K.E. Lotterhos

### CONTENTS

Supplemental Methods.	1. Proof for $CovGE$	2
Supplemental Methods.	2. Population Equations	5
Supplemental Methods.	3. Variance partitioning	6
Supplemental Tables.	Tables	9
Supplemental Tables.	1. Comparison of forms of Covariance	9
Supplemental Tables.	2. Variance Components - Single Common Garden	10
Supplemental Tables.	3. Variance Components - Reciprocal Transplant	11
Supplemental Tables.	4. Imbalanced Designs	12
Supplemental Figures.	Figures	13
Supplemental Figures.	1. Compatible and Incompatible designs	13
Supplemental Figures.	2. Coverage of parameter space	14
Supplemental Figures.	3. Examples of simulated data	15
Supplemental Figures.	4. Population vs. Sample Estimates	16
Supplemental Figures.	5. False negative rates for $\bar{\Delta}_{GxE}$	17
Supplemental Figures.	6. $V_{g*e}$ vs. $V_{cov_{ge}}$	18
Supplemental Figures.	7. Comparison of the percent of variance explained by $CovGE$	19
Supplemental Figures.	8. Sample data for balanced and imbalanced designs	21

### Supplemental Methods. 1. PROOF FOR Cov<sub>GE</sub>

The following equation is the standard approach to measure the covariance (Cov) between a pair of variables (here shown as x and y).

$$(1) \quad \text{Sample Cov}(x, y) = \frac{\sum(x - \bar{x})(y - \bar{y})}{N - 1}$$

$$(2) \quad \text{Population Cov}(x, y) = \frac{\sum(x - \mu_x)(y - \mu_y)}{N}$$

In which N is the total number of samples,  $\bar{x}$  is the sample mean of x,  $\mu_x$  is the population mean of x, while  $\bar{y}$  is the sample mean of y and  $\mu_y$  is the population mean of y.

The standard Cov(X,Y) equation cannot be directly applied to calculating Cov(G,E), because in Cov(G,E) we do not want to measure the covariance among every combination of genotypic and environmental effects, but only among the pairs of genotypic and environmental effects sourced from the same native environment. Therefore, we added the indicator variable ( $I$ ) to ensure that only those data in which genotypes ( $i$ ) are correctly matched to the environments ( $j$ ) to which they were sourced (or are native to) are used to estimate Covariance.  $I_{ij} = 1$  when genotype is correctly matched with the environment to which it is native, and  $I_{ij} = 0$  otherwise (See Fig. 1 in the main text). In keeping, we take the covariance between genotypic and environmental effects on phenotypes, across all genotype-environment pairs ( $n_{gen} = n_{env}$ ). This gives:

$$(3) \quad \text{Population Cov}(G, E) = \frac{1}{\sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} (I_{ij})} \left( \sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} (\mu_i - \mu)(\mu_j - \mu)(I_{ij}) \right)$$

In this equation,  $\mu_i$  is the population mean phenotype of the  $i$ th genotype, and  $\mu_j$  is the population mean phenotype of the  $j$ th environment, and  $\mu$  is the overall population mean phenotype.

To allow for comparisons across taxa, study systems, and environmental gradients, we sought to standardize Cov(G,E). The typical approach to standardize covariance is correlation, which standardizes by the product of the standard deviation of the x and y variable ( $\sigma_x * \sigma_y$ ). Here, we standardize by the variance in group means (instead of variance in all the data) because Cov(G,E) is based on covariance between the group means. We use  $\sigma_{\mu_i}$  to represent the population standard deviation in the group means for genotypic effects , and  $\sigma_{\mu_j}$  to represent the standard deviation in the group means for the environmental effects.

However, the group means for environmental effects and the group means for genotypic effects are not independent of each other, because they are both calculated from the phenotypic means.

For example, the following standardization produces inaccurate results:

$$(4) \quad Cor(G, E) = \frac{1}{\sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} (I_{ij})} \left( \frac{\sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} (\mu_i - \bar{\mu})(\mu_j - \bar{\mu})(I_{ij})}{(\sigma_{\mu_i} * \sigma_{\mu_j})} \right)$$

Equation 4 produces inaccurate results in simulations (e.g., does not produce Cov(G,E) estimates between -1 and 1) because the standardization by  $\sigma_{\mu_i} * \sigma_{\mu_j}$  assumes that G and E are independent, but in this case, they are not. Thus, we multiply this equation by a correction term  $k$ :

$$(5) \quad \left( \frac{1}{\sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} (I_{ij})} \frac{\sum_{j=1}^{n_{env}} (\mu_i - \bar{\mu})(\mu_j - \bar{\mu})(I_{ij})}{(\sigma_{\mu_i} * \sigma_{\mu_j})} \right) * k$$

When the correction factor  $k$  is equal to the following, the result is constrained between -1 and 1:

$$(6) \quad k = \max\left(\frac{\sigma_{\mu_i}}{\sigma_{\mu_j}}, \frac{\sigma_{\mu_j}}{\sigma_{\mu_i}}\right)$$

For example, if  $\sigma_{\mu_i} > \sigma_{\mu_j}$  and we apply  $\frac{\sigma_{\mu_j}}{\sigma_{\mu_i}}$  to the above equation, the  $\sigma_{\mu_j}$  terms cancel each other and the correction becomes:

$$(7) \quad \sigma_{\mu_i} * \sigma_{\mu_i} = \sigma_{\mu_i}^2$$

Which when applied to the correlation equation, can be written as:

$$(8) \quad \frac{1}{\sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} (I_{ij})} \left( \frac{\sum_{j=1}^{n_{env}} (\mu_i - \bar{\mu})(\mu_j - \bar{\mu})(I_{ij})}{\sigma_{\mu_i}^2} \right)$$

Thus, the Cov(G,E) is standardized by the variance in group means for genotypic effects or the variance in group means for environmental effects, whichever is greater. This standardization preserves the relationship between the genotypic and environmental effects in the data, while constraining Cov(G,E) to be between -1 and 1. Thus, the population and estimated measures for standardized covariance (which we call  $Cov_{GE}$ ) are as follows:

$$(9) \quad Population\ Cov_{GE} = \frac{1}{\sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} (I_{ij})} \left( \frac{\sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} (\mu_i - \bar{\mu})(\mu_j - \bar{\mu})(I_{ij})}{\max(\sigma_{\mu_i}^2, \sigma_{\mu_j}^2)} \right)$$

$$(10) \quad Sample\ \hat{Cov}_{GE} = \frac{1}{\sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} (I_{ij}) - 1} \left( \frac{\sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} (\bar{y}_i - \bar{y})(\bar{y}_j - \bar{y}) I_{ij}}{\max(s_{\bar{y}_i}^2, s_{\bar{y}_j}^2)} \right)$$

For the sample estimate,  $\bar{y}_i$  is the estimated marginal mean phenotype for  $i$ th genotype,  $\bar{y}_j$  is the estimated marginal mean phenotype for  $j$ th environment, and  $\bar{y}$  is the estimated overall mean phenotype.  $s^2_{\bar{y}_i}$  is the variance in the group mean phenotype for genotypic effects, while  $s^2_{\bar{y}_j}$  is the variance in the group mean phenotype for environmental effects.

## Supplemental Methods. 2. POPULATION EQUATIONS

We used simulated data to measure sampling estimates of  $\hat{Cov}_{GE}$  and  $\bar{\Delta}_{GxE}$ , which are shown in the main document. However, to validate these measures, we also calculated population parameters based on observed  $Cov_{GE}$  when  $\epsilon_{ijk} = 0$  (Eqn. 4 in main text) and used the following equations to measure population parameters of  $Cov_{GE}$  and  $\bar{\Delta}_{GxE}$ . See Supplemental Figure S2 below for simulation results showing agreement between sampling estimates and population parameter.

$$(11) \quad Population Cov_{GE} = \frac{1}{\sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} (I_{ij})} \left( \frac{\sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} (\mu_i - \bar{\mu})(\mu_j - \bar{\mu})(I_{ij})}{\max(\sigma_{\mu_i}^2, \sigma_{\mu_j}^2)} \right)$$

In this equation,  $n_{gen}$  is the number of genotypes and  $n_{env}$  is the number of environments.  $\mu_i$  is the mean phenotype for genotype  $i$ ,  $\mu_j$  is the mean phenotype for environment  $j$ , and  $\bar{\mu}$  is the overall mean phenotype.  $I_{ij}$  is the indicator variable which equals one when genotype  $i$  is correctly matched with the environment  $j$  to which it is native, and zero otherwise. Standard deviation of genotypic group means is indicated by  $\sigma_{\mu_i}$  and standard deviation of environmental group means is  $\sigma_{\mu_j}$ .

$$(12) \quad Population \bar{\Delta}_{GxE} = \frac{1}{(n_{gen})(n_{env})} \sum_{i=1}^{n_{gen}} \sum_{j=1}^{n_{env}} |\mu_{ij} - \mu_i - \mu_j + \bar{\mu}|$$

In this equation,  $n_{gen}$  is the number of genotypes and  $n_{env}$  is the number of environments.  $\mu_{ij}$  is the mean phenotype for genotype  $i$  in environment  $j$ ,  $\mu_i$  is the mean phenotype for each genotype,  $\mu_j$  is the mean phenotype for each environment, and  $\bar{\mu}$  is the overall mean phenotype. There is no indicator variable, as all combinations of genotype and environment are used to calculate  $\bar{\Delta}_{GxE}$ .

### Supplemental Methods. 3. VARIANCE PARTITIONING

In the main manuscript, we focus on estimating the effect size of  $Cov_{GE}$  and  $\bar{\Delta}_{GxE}$  and their significance. The effect size provides information about the strength of the pattern, which is a distinct type of information from the percent of variation in the phenotype explained by the following components: (i) genetic effects on phenotype, (ii) environmental effects on phenotypes, (iii) genetic x environment interactions, (iv) covariance between genetic and environmental effects on phenotypes, and (iv) residual error. Both Falconer (1989) and Conover and Schultz (1995) have previously discussed  $Cov_{GE}$  in this more traditional sense as the percent of variation in the phenotype explained by different variance components:

$$(13) \quad V_P = V_G + V_E + V_{GxE} + xCov_{GE}$$

We follow the definition of environmental variance from Falconer 1989, which states that environmental variance includes all sources of non-genetic variance, including those caused by external causes such as environmental factors, maternal effects, and “untangible” (residual) variation that cannot be eliminated (Falconer 1989). Falconer also defines  $Cov_{GE}$  as a correlation between environmental deviation and genotypic values: “genotypic values and environmental deviations may be correlated, in which case  $V_P$  will be increased by twice the covariance of G with E”. We think this terminology is confusing, and that it is more accurate to say: the correlation between ”phenotypic deviation due to genotype (genotypic deviation/effect)” and ”phenotypic deviation due to environment (environmental deviation/effect)” across spatial locations. What can be inferred about the causes of environmental variance (specifically whether it is due to environment or to “untangible” error) depends on the experimental design and how samples are chosen for the study.

In this equation, increases in  $V_{GxE}$  will reduce the potential magnitude of  $xCov_{GE}$  just as increases in  $V_G$  or  $V_E$  will reduce the potential magnitude of  $V_{GxE}$  or  $xCov_{GE}$ .

In a 2x2 reciprocal transplant or common garden design with 2 genotypes in 2 environments,  $x = 2$ . This factor of  $x = 2$  does not extend to more complex designs, as explained below.

Here, we show how to extend sums of squares ( $SS$ ) calculations from a traditional analysis of variance to incorporate  $SS_{Cov_{GE}}$ , which can in turn be used to understand the percent of variation in phenotypes explained by different components. These calculations assume a fully factorial reciprocal transplant design. We do not advocate that these  $SS$  be used to test the significance of the variance components with a traditional  $F$ -test, because the presence of  $Cov_{GE}$  likely violates the assumption of independence among samples and complicates calculations of degrees of freedom (violations that are addressed in the main manuscript by using bootstrap and permutation for hypothesis testing). It is, however, useful to compare the percent of variation explained by different components to their effect sizes, because it furthers understanding of the relative influence of genetic differentiation and plasticity on the evolved patterns in the population.

In a reciprocal transplant experiment, there are  $g$  genotypes transplanted into  $e$  environmental patches, for a total of  $g * e = n_{ge}$  genotype-environment combinations. In a fully factorial reciprocal transplant experiment,  $g = e$ ,  $g$  is the number of levels of genotypes from  $i = 1, 2...g$ , and  $e$  is the number of levels of environments from  $j = 1, 2...e$ .

Assuming the equal sample sizes  $r$  ( $k = 1, 2, ...r$ ) within each genotype-environment combination, the following sums of squares can be estimated as:

$$(14) \quad V_G = SS_G = re \sum_{i=1}^g (\bar{y}_i - \bar{y})^2$$

$$(15) \quad V_E = SS_E = rg \sum_{j=1}^e (\bar{y}_j - \bar{y})^2$$

$$(16) \quad V_{GxE} = SS_{GE} = r \sum_{i=1}^g \sum_{j=1}^e (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2$$

$$(17) \quad V_{Cov_{GE}} = SS_{Cov_{GE}} = |xr \sum_{i=1}^g \sum_{j=1}^e (\bar{y}_i - \bar{y})(\bar{y}_j - \bar{y}) I_{ij}|$$

$$(18) \quad V_{error} = SS_{Error} = \sum_{i=1}^g \sum_{j=1}^e \sum_{k=1}^r (y_{ijk} - \bar{y}_{ij})^2$$

where  $x = \frac{ge}{(\sum_{i=1}^g \sum_{j=1}^e I_{ij})}$ , and  $I_{ij}$  is an indicator variable that is 1 when the genotype  $i$  originated from environment  $j$  and 0 otherwise. In a 2x2 reciprocal transplant design,  $g = 2$ ,  $e = 2$ , and  $\sum_{i=1}^g \sum_{j=1}^e I_{ij} = 2$ , so  $x = 2$  as is assumed in Eq. S1. However in a 4x4 reciprocal transplant design,  $g = 4$ ,  $e = 4$ , and  $\sum_{i=1}^g \sum_{j=1}^e I_{ij} = 4$ , so  $x = 4$ . The factor  $x$  ensures that the  $SS_{Cov_{GE}}$  scales appropriately with the  $SS$  of the other components with the size of the experiment. Finally, since  $Cov_{GE}$  can be negative under countergradient scenarios, we take the absolute value for partitioning variance.

The percent of variation explained by each component ( $Comp$ ) can then be estimated as

$$(19) \quad \eta_{Comp}^2 = \frac{SS_{Comp}}{SS_G + SS_E + SS_{GxE} + SS_{CovGE} + SS_{Error}}$$

We recognize that this is a crude approach. However, it provides a reasonable way to compare to the percent of variation explained by different components to their effect size estimates in the main text. For example, when

$|Cov_{GE}|$  is maximized, equal amounts of non-residual variance are explained by  $V_G$ ,  $V_E$ , and  $V_{Cov_{GE}}$  for the fully factorial reciprocal transplant experiment with an arbitrary number of populations (Supplementary Fig 6a). For this reason, when residual variation is minimized the maximum percent of phenotypic variance explained by  $V_{Cov_{GE}}$  is rarely greater than 1/3 of the total phenotypic variance (Supplementary Fig. S5, Fig.S6a).

The comparison of effect sizes and variance components is also useful for understanding the relative influences of  $V_G$  and  $V_E$  on a particular value of  $|Cov_{GE}|$ ; for instance, similarly intermediate values of  $|Cov_{GE}|$  can be driven by higher  $V_E$  and lower  $V_G$  (Supplementary Fig. 6B), or lower  $V_E$  and higher  $V_G$  (Supplementary Fig. S6c).

## Supplemental Tables. TABLES

### Supplemental Tables. 1. COMPARISON OF FORMS OF COVARIANCE

Phrase	Verbal	Intrinsic vs. Emergent?	References	Experimental design and estimation from data	GxE interpretation from data
Spatial CovGE	CovGE is the covariance between phenotypic deviations due to genotype (genotypic effects) and phenotypic deviations due to environment (environmental effects) across spatial locations. Stated differently, CovGE occurs when the environmental and genotypic effects on the phenotype across spatial locations covary positively or negatively.	Emergent metapopulation-scale phenomenon.	This paper, Levins 1968, Falconer 1989, Conover and Schultz 1995	This paper	Evolution of different mean reaction norms among individuals sampled from different spatial locations
Covariance between phenotype and environment	The association between phenotypic expression of a single genotype and the environment of growth (i.e., the reaction norm)	evolved state of environmentally induced production of alternate phenotypes from a given genotype	DeWitt and Scheiner 2004	To estimate reaction norms, several progenies from each family, clone, or genotype are grown in a discrete number of environmental treatments	Evolution of different mean reaction norms among progeny from different families, clones, or genotypes
"Genetic covariance in character states"	"In the character state approach, the reaction norm for a particular character is modelled as the set of phenotypic values that would be expressed in each environment by a given genotype and evolutionary models are based on population means and genetic (co)variances of these character states." Via et al. 1995.  The means of this bivariate normal distribution represents the mean (i) phenotype at one value of the environment and (ii) mean phenotype at another value of the same environment, and the diagonals in the covariance matrix represent the additive genetic variance for (i) and (ii), respectively. The off diagonal in the covariance matrix is an additive genetic covariance that represents the association between the additive effect of a mutation on the phenotype at one value of the environment and the same phenotype at the other value of the environment.	Intrinsic property of the genotype-phenotype map and the external environment	Reviewed in Via et al. 1995	"Because measurements of different character states must be made on separate individuals, the usual statistical methods for calculating the genetic correlation are not applicable. In this case alternatives may be employed, such as the correlation of family means or the re-expression of genotype-environment interaction as a genetic covariance between character states expressed in different environments." (Via and Lande 1985)	Unclear (de Jong 1995)
"Genetic covariance in polynomial coefficients"	"In the polynomial approach, the reaction norm is described by a polynomial function of the phenotypic values expressed by a genotype across a range of environments, and evolutionary models are based on the population means and genetic (co)variances of coefficients of the polynomial." (Via et al. 1995)  The means of this bivariate normal distribution represents the mean parameter values in the polynomial, and the diagonals in the covariance matrix represent the additive genetic variance for each parameter. The off diagonal in the covariance matrix is an additive genetic covariance that represents the association between the additive effect of a mutation on one parameter and another parameter in the polynomial function.	Intrinsic property of the genotype-phenotype map and the external environment	Reviewed in Via et al. 1995	Genetic and phenotypic variances and covariances can be estimated using the animal model (Lynch and Walsh 1998)	Genetic variance for reaction norm slope > 0
"bivariate normal distribution of slope (plasticity) and elevation (breeding value) in the reference environment"	The means of this bivariate normal distribution represents the mean (i) reaction norm intercept and (ii) slope in the reference environment. The diagonals on this covariance matrix in this bivariate normal distribution represent the additive genetic variance for the (i) reaction norm intercept and (ii) slope. The off diagonal in this bivariate normal distribution is an additive genetic covariance that represents the association between the additive effect of a mutation on the reaction norm intercept and the reaction norm slope.  Linear norms of reaction for each genotype are sampled at random from a population with this bivariate normal distribution.	Intrinsic property of the genotype-phenotype map and the external environment	Lande 2009	Genetic and phenotypic variances and covariances can be estimated using the animal model (Lynch and Walsh 1998)	Additive genetic variance for reaction norm slope > 0
"bivariate distribution of a phenotype split into its genetic and environmental components"	In this IBM-based approach, the phenotype is modeled as a combination of additive genetic and environmental components with a bivariate distribution.  The means of this bivariate normal distribution represents the mean (i) reaction norm intercept and (ii) slope in the reference environment. The diagonals on this covariance matrix in this bivariate normal distribution represent the additive genetic variance for the (i) reaction norm intercept and (ii) slope. The off diagonal in this bivariate normal distribution is an additive genetic covariance that represents the association between the additive effect of a mutation on the reaction norm intercept and the reaction norm slope.	Intrinsic property of the genotype-phenotype map and the external environment	Coulson et al. 2017 Coulson et al. 2021	If relatedness information is available and the infinitesimal model is assumed, genetic and phenotypic variances and covariances can be estimated using the animal model (Lynch and Walsh 1998).	Additive genetic variance for reaction norm slope > 0

TABLE 1. Many different types of variance and covariance can be found in the evolutionary literature, which can easily lead to confusion about *CovGE*. This table clarifies the different types of covariance in the literature.

## Supplemental Tables. 2. VARIANCE COMPONENTS - SINGLE COMMON GARDEN

Mathematical notation	Variance component name	Biological interpretation if multiple inbred lines are used in common garden (genetic variation removed within lines)	Biological interpretation if multiple wild-collected ecotypes are used in common garden (genetic variation not removed within ecotypes)
$V_G = SS_G = r \sum_{i=1}^g (\bar{y}_i - \bar{y})^2$	Variation among genotypes	Reflects phenotypic variation due to genetic differences among inbred lines. Often interpreted as $V_G$	Reflects phenotypic variation due to genetic differences among ecotypes. Not directly interpretable as $V_G$
$V_{error} = SS_{Error} = \sum_{i=1}^g \sum_{k=1}^r (y_{ik} - \bar{y}_i)^2$	Residual variation (Variation within genotypes)	Reflects phenotypic variation within inbred lines due to environmental noise. Often interpreted as $V_E$ .	Reflects phenotypic variation within ecotypes due to environmental noise and genetic variation within ecotypes. Not interpreted as $V_E$ because it includes environmental noise and genetic variation (confounds them).

TABLE 2. Biological interpretations are informed from Falconer (1989). Falconer also defines  $Cov_{GE}$  as a correlation between environmental deviation and genotypic values: “genotypic values and environmental deviations may be correlated, in which case  $V_P$  will be increased by twice the covariance of  $G$  with  $E$ ”. We think this terminology is confusing, and that it is more accurate to say: the correlation between “phenotypic deviation due to genotype (genotypic deviation/effect)” and “phenotypic deviation due to environment (environmental deviation/effect)” across spatial locations. What can be inferred about the causes of environmental variance (specifically whether it is due to environment or to “untangible” error) depends on the experimental design and how samples are chosen for the study.

In the equations,  $r$  ( $k = 1, 2, \dots, r$ ) refers to the number of samples collected within each genotype-environment combination (assuming equal sample sizes).  $SS$  refers to Sums of Squares,  $g$  indicates the genotypes ( $i = 1, 2, \dots, i$ ),  $\bar{y}_i$  is the average mean phenotype for a particular genotype,  $y_{ik}$  is the phenotype for each sample collected,  $\bar{y}$  is the average mean phenotype across all genotypes. See Supplemental Methods 3 above for more details.

## Supplemental Tables. 3. VARIANCE COMPONENTS - RECIPROCAL TRANSPLANT

Mathematical notation	Variance component name	Biological interpretation if multiple inbred lines are used in experiment (genetic variation removed within lines)	Biological interpretation if multiple wild-collected ecotypes are used in experiment (genetic variation not removed within ecotypes)
$V_G = SS_G = re \sum_{i=1}^g (\bar{y}_i - \bar{y})^2$	Variation among phenotypic deviations due to genotypes	Reflects phenotypic variation due to genetic differences among inbred lines, averaged across environments. Often interpreted as $V_G$	Reflects phenotypic variation due to genetic differences among ecotypes, averaged across environments. Often interpreted as $V_G$
$V_E = SS_E = rg \sum_{j=1}^e (\bar{y}_j - \bar{y})^2$	Variation among phenotypic deviations due to environments	Reflects phenotypic variation due to environmental differences among sites, averaged across genotypes. Often interpreted as $V_E$	Reflects phenotypic variation due to environmental differences among sites, averaged across genotypes. Often interpreted as $V_E$
$V_{GxE} = SS_{GE} = r \sum_{i=1}^g \sum_{j=1}^e (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2$	Interaction between genotypic and environmental deviations on phenotypes	Reflects phenotypic variation due to the environmental phenotypic deviation not being equal across genotypes. Often interpreted as $V_{GxE}$	Reflects phenotypic variation due to the environmental phenotypic deviation not being equal across genotypes. Often interpreted as $V_{GxE}$
$V_{CovGE} = SS_{CovGE} =  xr \sum_{i=1}^g \sum_{j=1}^e (\bar{y}_i - \bar{y})(\bar{y}_j - \bar{y})I_{ij} $	Covariance between genotypic and environmental influences on phenotypes	Not often included in model calculation.	Not often included in model calculation.
$V_{error} = SS_{Error} = \sum_{i=1}^g \sum_{j=1}^e \sum_{k=1}^r (y_{ijk} - \bar{y}_{ij})^2$	Residual variation (Variation within groups)	Reflects phenotypic variation within inbred lines due to environmental noise.	Reflects phenotypic variation within ecotypes due to environmental noise and genetic variation within ecotypes.

TABLE 3. Biological interpretations are informed from Falconer (1989) or KEL, personal observation.

In the equations,  $r$  ( $k = 1, 2, \dots, r$ ) refers to the number of samples collected within each genotype-environment combination (assuming equal sample sizes).  $SS$  refers to Sums of Squares,  $e$  indicates the environments ( $j = 1, 2, \dots, j$ ),  $g$  indicates the genotypes ( $i = 1, 2, \dots, i$ ),  $\bar{y}_i$  is the average mean phenotype for a particular genotype,  $\bar{y}_j$  is the average mean phenotype for a particular environment,  $y_{ijk}$  is the phenotype for each sample in each genotype and environment.  $\bar{y}$  is the average mean phenotype across all genotypes. In a 2x2 reciprocal transplant or common garden design with 2 genotypes in 2 environments,  $x = 2$ . This factor of  $x = 2$  does not extend to more complex designs, as explained in Supplemental Methods 3.

#### Supplemental Tables. 4. IMBALANCED DESIGNS

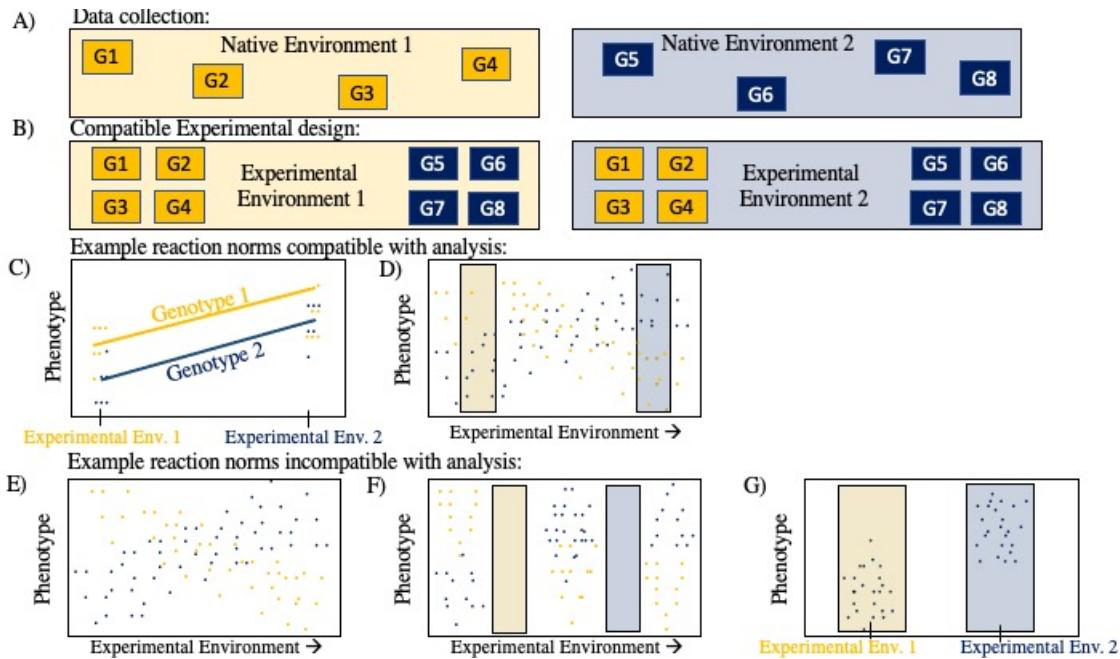
For common garden experimental designs, there may be imbalance where there are unequal sample sizes or unequal numbers of genotypes for each native environment. Estimated marginal means reduce bias that may arise from unequal sample sizes within groups, thus do not affect  $\hat{Cov}_{GE}$ . However, because environment and genotype are treated as factors in the ANOVA to extract estimated marginal means, imbalanced numbers of genotypes (example shown in Supplemental Figure S7B below) introduce bias in the environmental means ( $\bar{y}_j$  or  $\mu_j$ ) and overall mean phenotype ( $\bar{y}$  or  $\mu$ ) (See B in the table below) In fact, both studies analyzed in the main text have imbalanced numbers of genotypes. To correct for the bias, a second ANOVA model with native environment substituted for genotype can then be used to extract estimated marginal means to calculate unbiased ( $\bar{y}_j$  or  $\mu_j$ ) and ( $\bar{y}$  or  $\mu$ ). The original model should be used to generate estimated marginal means for ( $\bar{y}_i$  or  $\mu_i$ ). As shown in column C in the table below, this approach returns  $\mu_j$  and  $\mu$  to the values in balanced designs (A).

A) Original Design				B) Imbalanced Design				C) Imbalanced Design using correction			
Model for $\mu_j$ and $\mu_i$ and $\mu$ : {Phenotype ~ Gen. * Experimental Env.}				Model for $\mu_j$ and $\mu_i$ and $\mu$ : {Phenotype ~ Gen. * Experimental Env.}				Model for $\mu_j$ and $\mu$ : {Phenotype ~ Native Env. * Experimental Env.} Model for $\mu_i$ : {Phenotype ~ Gen. * Experimental Env.}			
$\mu_i$	Gen.	Native Env.	$\mu_j$	$\mu_i$	Gen.	Native Env.	$\mu_j$	$\mu_i$	Gen.	Native Env.	$\mu_j$
0.68	G_1	E_1	-0.68	0.68	G_1	E_1	-0.98	0.68	G_1	E_1	-0.68
0.68	G_2	E_1	-0.68								
0.68	G_3	E_1	-0.68								
0.68	G_4	E_1	-0.68								
-0.68	G_5	E_2	0.68	-0.68	G_5	E_2	0.27	-0.68	G_5	E_2	0.68
-0.68	G_6	E_2	0.68	-0.68	G_6	E_2	0.27	-0.68	G_6	E_2	0.68
-0.68	G_7	E_2	0.68	-0.68	G_7	E_2	0.27	-0.68	G_7	E_2	0.68
-0.68	G_8	E_2	0.68	-0.68	G_8	E_2	0.27	-0.68	G_8	E_2	0.68
$\mu = 0.00$				$\mu = -0.21$				$\mu = 0.00$			

TABLE 4. (A) shows example data for a strong negative  $Cov_{GE}$  pattern for a Common-garden design. When the design is balanced, four genotypes native to two different environments were measured across two experimental environments (Supplemental Figure S7 below for depiction of corresponding sample data, Fig. S7A shows data for panels A in the above table, Fig. S7B shows data for panels B and C in above table). When the design is balanced, the mean phenotypes for genotype ( $\mu_i$ ), environment ( $\mu_j$ ), and overall ( $\mu$ ) are calculated by extracting estimated marginal means from a single model. When the design is unbalanced (as shown in B), continuing to use the same model to generate estimated marginal means produces bias in environmental means ( $\mu_j$ ), and overall means ( $\mu$ ). As shown in (C), the bias is corrected by conducting a secondary ANOVA with native environment replacing genotype to calculate environmental means ( $\mu_j$ ), and overall means ( $\mu$ ). Genotypic means ( $\mu_i$ ) are calculated using the original ANOVA model.

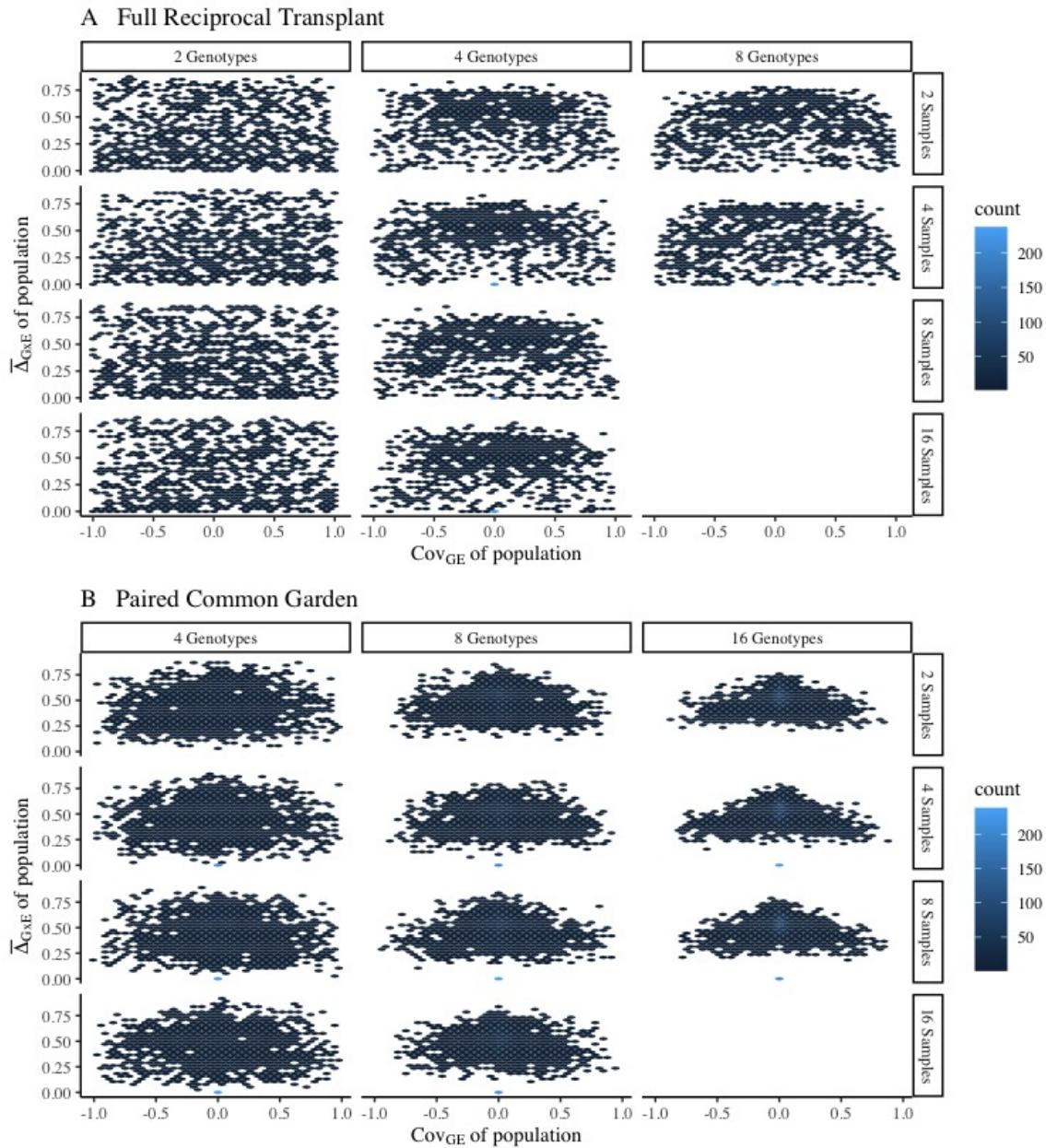
## Supplemental Figures. FIGURES

## Supplemental Figures. 1. COMPATIBLE AND INCOMPATIBLE DESIGNS



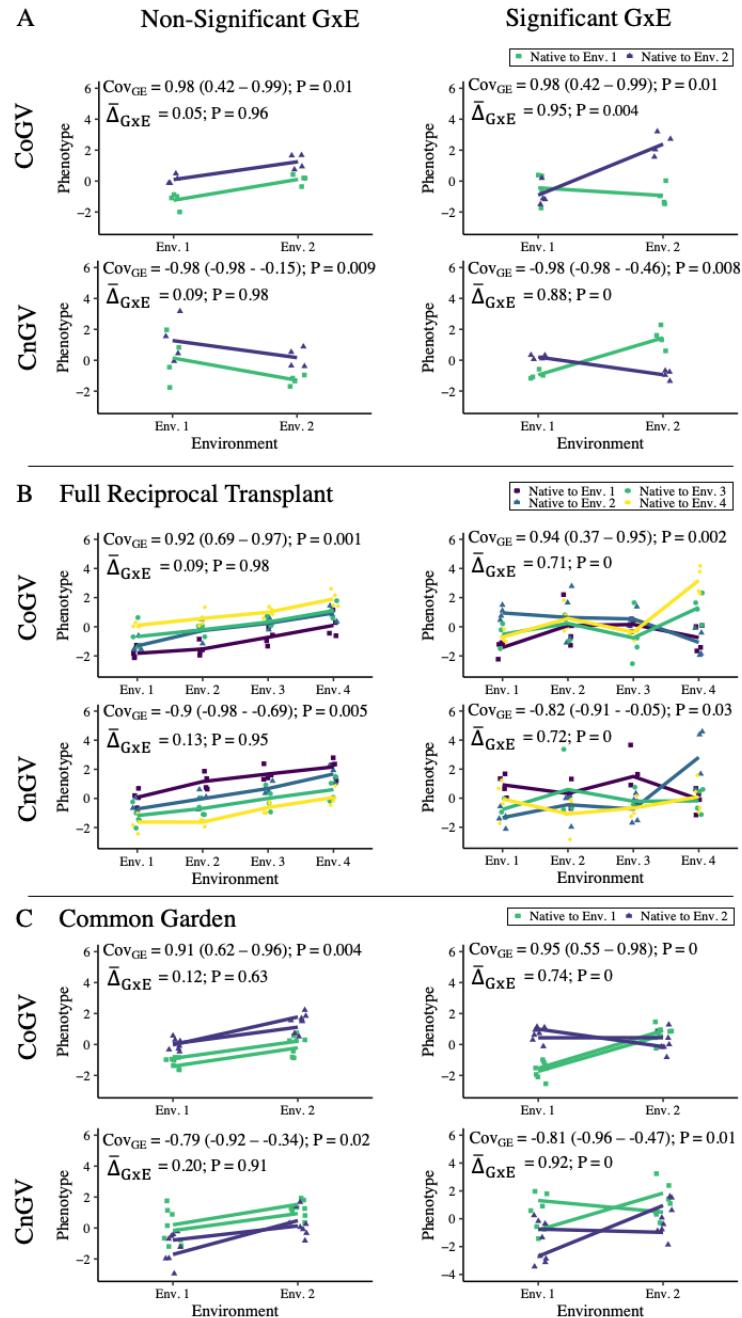
SUPPLEMENTARY FIGURE S1. Examples of experimental designs that are compatible or incompatible with  $CovGE$ . Plot A shows how phenotypic data are collected. Each box represents a different environment (orange, blue). Four genotypes are shown within each environment, with boxes representing the genotypes from which individuals (samples) are collected. Plot B shows an example of a fully factorial experimental design in which genotypes native to each environment are exposed to both environmental conditions. Plots C and D indicate phenotypic data that may result from the experimental design in panel B that are compatible with the analysis presented herein. Plot C shows categorical phenotypic data collected in either Environment 1 (orange), or Environment 2 (blue). Plot D shows phenotypic collected from an experimental design that used a continuous environmental gradient. However, to be analyzed, continuous phenotypic data would need to be binned. Note that in panel D, there are uneven numbers of samples in each bin. This is okay, because estimated marginal means reduce bias from unequal sample sizes. Plots E, F, and G show designs that are not incompatible with the  $CovGE$  equation. Panel E shows phenotypic data along a continuous axis but there is no information on which environment each genotype is native to. Because our analysis requires that information on the native environment of each genotype is known, this design would remain incompatible even if environment was categorical (like in plot C). Plot F demonstrates phenotypic data in which the native environment is known, but was not included as a treatment in the experiment. Indeed, not only is information about the native environment required, but it also must be included as a treatment. Plot G shows an instance where a single genotype is tested across multiple environments. This design is incompatible because the  $CovGE$  calculation requires at least 2 genotypes from at least 2 environments (but note that 2x2 designs are often underpowered - see Fig. 4 in the main text).

## Supplemental Figures. 2. COVERAGE OF PARAMETER SPACE



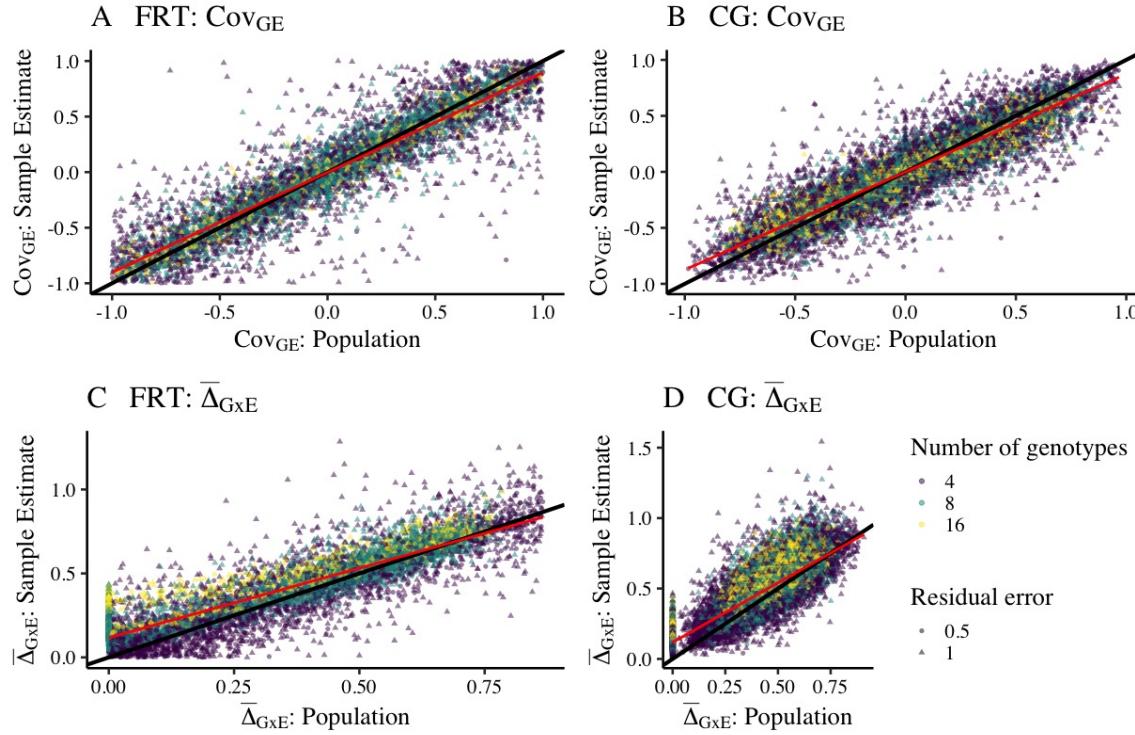
SUPPLEMENTARY FIGURE S2. Coverage of parameter space of  $Cov_{GE}$  and  $\bar{\Delta}_{GxE}$  for full reciprocal transplant (A) and paired common garden designs (B). Hexagons are colored according to the density of observations in each bin.

## Supplemental Figures. 3. EXAMPLES OF SIMULATED DATA



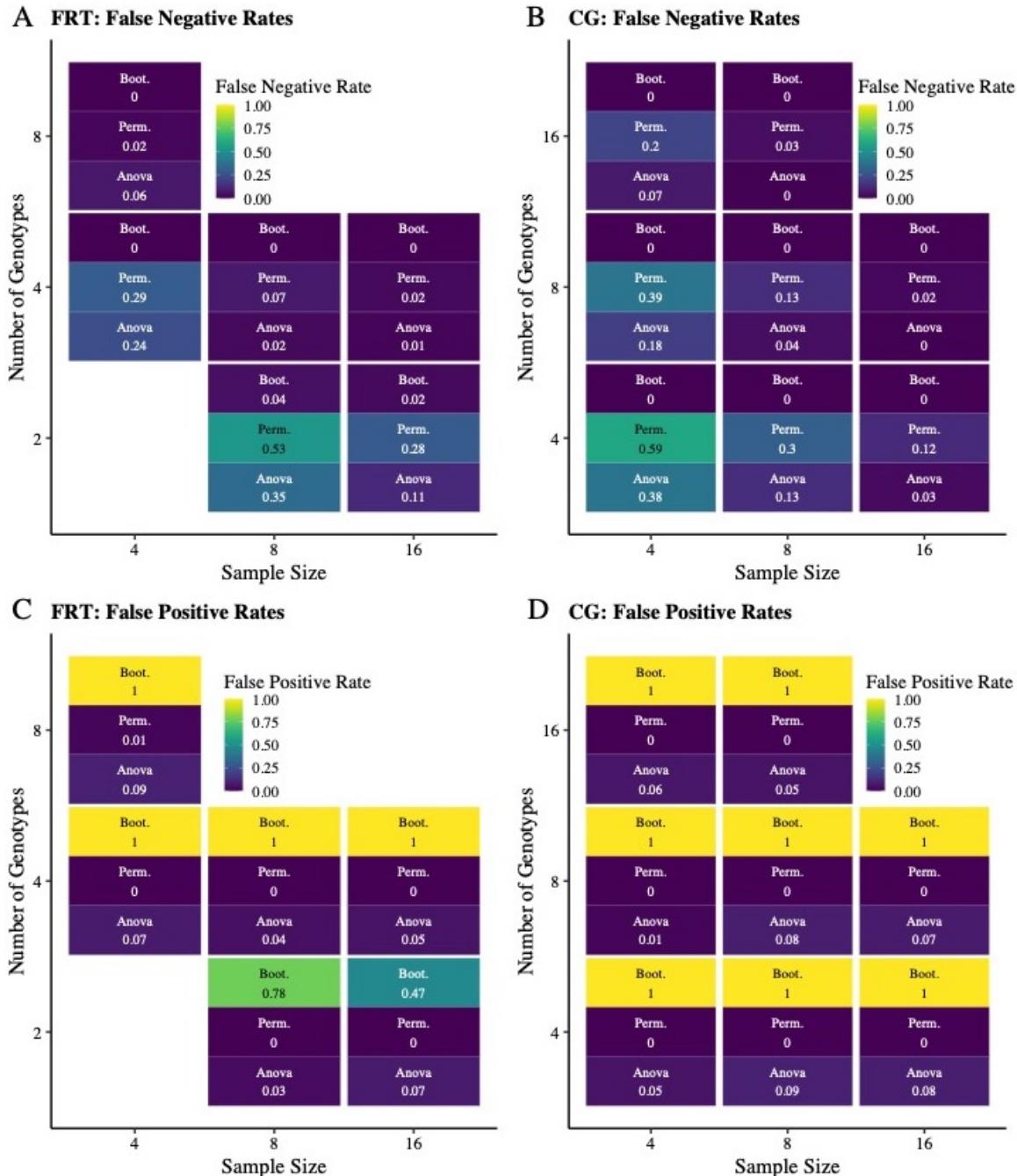
SUPPLEMENTARY FIGURE S3. A) Possible outcomes from a two-genotype by two-environment design. B) Possible outcomes from a full reciprocal transplant design in which four distinct genotypes are transplanted into four distinct environments. C) Possible outcomes from a common garden design in which two pairs of genotypes share the same native environment. In plots A and B, reaction norms signify different genotypes and are colored accordingly. In panel C, separate genotypes that share the same native environment share the same color.

## Supplemental Figures. 4. POPULATION VS. SAMPLE ESTIMATES

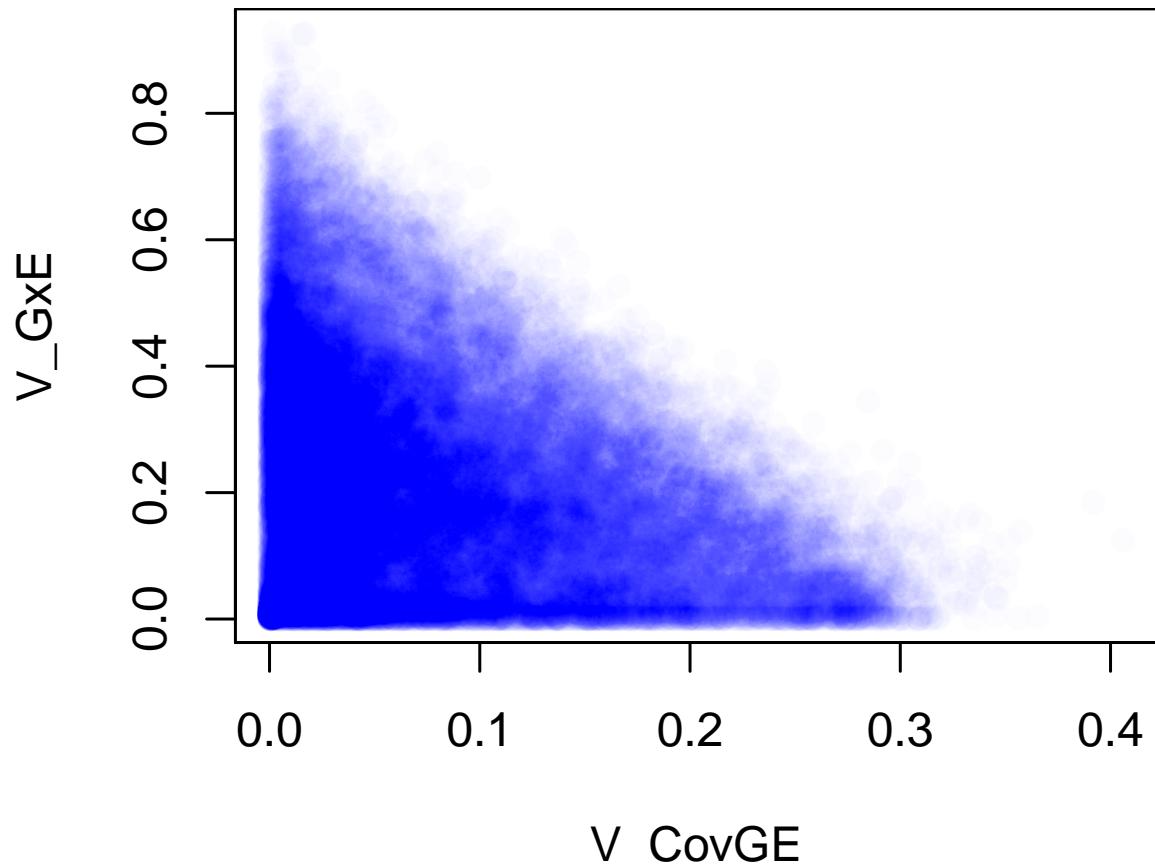


SUPPLEMENTARY FIGURE S4. Agreement between population measures and sample estimates of  $\hat{Cov}_{GE}$  and  $\bar{\Delta}_{GxE}$  for Full Reciprocal Transplant (A, C) and paired Common Garden designs (B, D). The black line falls along a 1:1 line while the red line reflects the pattern of the data. Point colors indicate the number of genotypes, while point shapes indicate the level of residual variation. As expected, sample estimates deviate more from the population measure in situations with low sample sizes and higher residual variation.

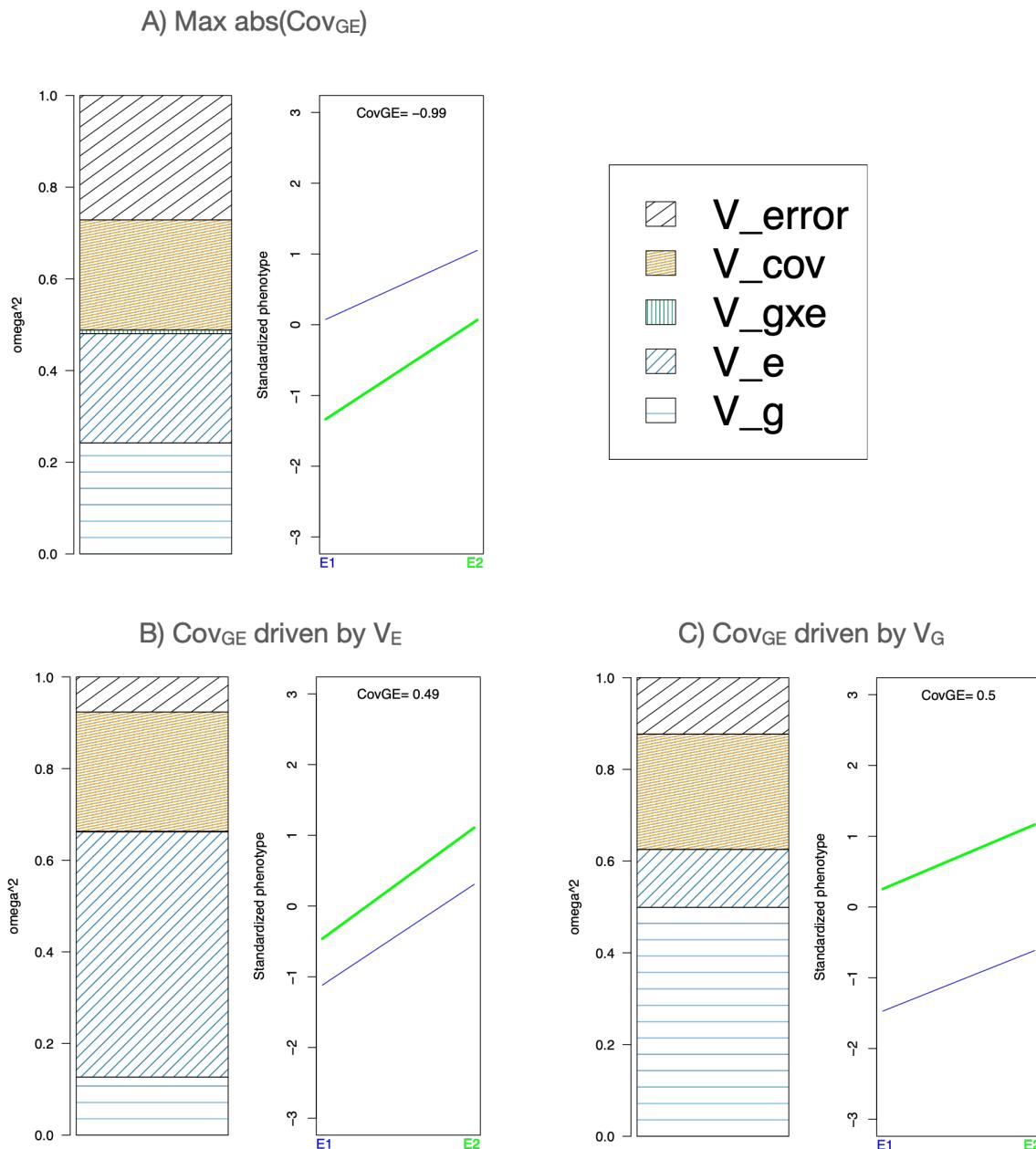
Supplemental Figures. 5. FALSE NEGATIVE RATES FOR  $\bar{\Delta}_{GxE}$



SUPPLEMENTARY FIGURE S5. Heat map showing false negative rates for  $\bar{\Delta}_{GxE}$  for Full Reciprocal Transplant (A) and paired Common Garden designs (B) and false positive rates for full reciprocal transplant (C) and paired Common Garden designs (D). Tiles are split to show rates for bootstrapping (top), permutation (middle) and ANOVA (bottom) approaches. This figure complements Fig. 4 in the main text and is meant to demonstrate that bootstrapping produces unreliable confidence intervals for  $\bar{\Delta}_{GxE}$ .

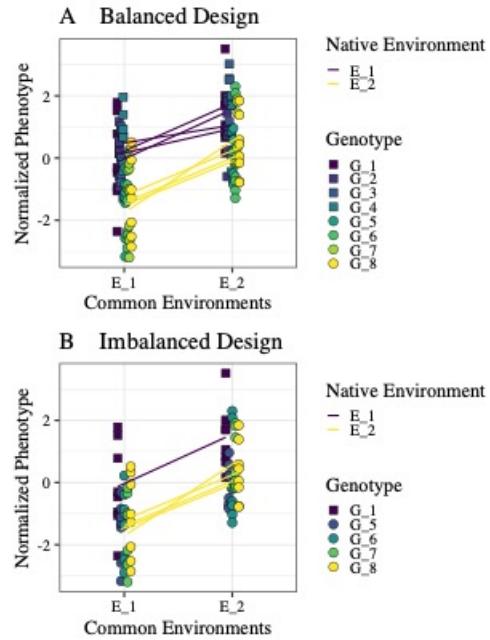
Supplemental Figures. 6.  $V_{g*e}$  vs.  $V_{cov_{ge}}$ SUPPLEMENTARY FIGURE S6. Comparison of the percent of variance explained by  $V_{GxE}$  and  $V_{Cov_{GE}}$  across all simulations.

Supplemental Figures. 7. COMPARISON OF THE PERCENT OF VARIANCE EXPLAINED BY  $Cov_{GE}$



SUPPLEMENTARY FIGURE S7. Comparison of the percent of variance explained by  $V_{Cov_{GE}}$  with the effect size of  $Cov_{GE}$  for three example simulations from a 2 genotype x 2 environment reciprocal transplant design. In each subpanel, the stacked barplot on the left of each panel shows the proportion of variance explained by each component corresponding to the legend (See Supplemental Materials 4), while the plot on the right shows the mean standardized reaction norms for each genotype (the effect size of  $Cov_{GE}$  is shown at the top). A) When  $|Cov_{GE}| = 1$  (i.e., maximal  $Cov_{GE}$ ), equal proportions of phenotypic variance is explained by  $V_g$ ,  $V_e$ , and  $V_{Cov_{GE}}$ . B) An intermediate effect size of  $Cov_{GE} = 0.5$  arising from larger  $V_e$  compared to  $V_g$ . C) An intermediate effect size of  $Cov_{GE} = 0.5$  arising from larger  $V_g$  compared to  $V_e$ .

Supplemental Figures. 8. SAMPLE DATA FOR BALANCED AND IMBALANCED DESIGNS



SUPPLEMENTARY FIGURE S8. (A) An example of normalized phenotypic data for a balanced common garden design, meaning that both native environments have equal numbers of genotypes (in this case, 4 genotypes each). Points show phenotypic data colored according to genotype, while reaction norms are colored according to the environment to which the different genotypes are native. Data point shapes correspond to the native environment as well. (B) Normalized phenotypic for an imbalanced design in which there are 4 genotypes native to Environment 2, but only 1 genotype native to Environment 1. Imbalanced designs can generate bias in  $CovGE$  unless corrected. See Supplemental Methods 3 for details on the correction.

SUPPLEMENTARY FIGURE : LITERATURE CITED.

Conover, D.O. and Schultz, E.T., 1995. Phenotypic similarity and the evolutionary significance of countergradient variation. *Trends in Ecology and Evolution*, 10(6), pp.248-252.

Coulson, T., 2021. We live in a changing world, but that shouldn't mean we abandon the concept of equilibrium. *Ecology Letters*, 24(1), pp.3-5.

Coulson, T., Kendall, B.E., Barthold, J., Plard, F., Schindler, S., Ozgul, A. and Gaillard, J.M., 2017. Modeling adaptive and nonadaptive responses of populations to environmental change. *The American Naturalist*, 190(3), pp.313-336.

De Jong, G., 1995. Phenotypic plasticity as a product of selection in a variable environment. *The American Naturalist*, 145(4), pp.493-512.

DeWitt, T.J. and Scheiner, S.M. eds., 2004. Phenotypic plasticity: functional and conceptual approaches. *Oxford University Press*.

Falconer, D.S., 1996. Introduction to quantitative genetics. *Pearson Education India*.

Lande, R., 2009. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of evolutionary biology*, 22(7), pp.1435-1446.

Levins, R., 1968. Evolution in changing environments. *Princeton University Press*.

Lynch, M. and Walsh, B., 1998. Genetics and analysis of quantitative traits. *Sinauer Associates, Inc.*. Sunderland, Massachusetts.

Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S.M., Schlichting, C.D. and Van Tienderen, P.H., 1995. Adaptive phenotypic plasticity: consensus and controversy. *Trends in ecology and evolution*, 10(5), pp.212-217.

Via, S. and Lande, R., 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, 39(3), pp.505-522.