academicJournals

Vol. 8(11), pp. 257-272, November 2016
DOI: 10.5897/JPBCS2016.0614
Article Number: B93F9CF61326
ISSN 2006-9758

http://www.academicjournals.org/JPBCS

Journal of Plant Breeding and Crop Science

Full Length Research Paper

A method of estimating broad-sense heritability for quantitative traits in the type 2 modified augmented design

Frank M. You[1]*, Gaofeng Jia[1,2], Sylvie Cloutier[3], Helen M. Booker[2], Scott D. Duguid[1] and Khalid Y. Rashid[1]

[1]Morden Research and Development Centre, Agriculture and Agri-Food Canada, Morden, MB, R6M 1Y5, Canada.
[2]Crop Development Centre, Department of Plant Sciences, University of Saskatchewan, 51 Campus Drive, Saskatoon, SK, S7N 5A8, Canada.
[3]Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, K1A 0C6, Canada.

Received 20 September, 2016; Accepted 17 October, 2016

Heritability is a basic genetic parameter for quantitative traits that may determine their selection generation and intensity as well as predict their selection response and efficiency in plant breeding. Estimation of heritability varies based on experimental design. The type 2 modified augmented design (MAD2) as an unbalanced experimental design, has been proposed for evaluating numerous unreplicated test genotypes with several replicated control genotypes to adjust for soil heterogeneity. Here, we define an inter-environment correlation (r_E), that is, the mean Pearson's correlation coefficient of trait performance for test genotypes between all pairs of environments, to approximate broad-sense heritability (H^2). Computer simulation and empirical results demonstrated that r_E was consistent with H^2 estimates on a plot basis by ANOVA for non-missing data sets, and similar to those by the restricted maximum likelihood (REML)-based method for missing data sets. The r_E method was shown to generally outperform the ANOVA- and REML-based methods.

Key words: Broad-sense heritability, analysis of variance, inter-environment correlation, modified augmented design, restricted maximum likelihood, flax.

INTRODUCTION

Heritability is a basic genetic parameter for quantitative traits that may determine their selection generation and intensity as well as predict their selection response and efficiency in plant breeding. Heritability estimation varies depending on the experimental design (Holland et al., 2003). The modified augment design (MAD) has been proposed for square plots (Type 1) (Lin and Poushinsky, 1983) and specifically for rectangular plots (Type 2 - MAD2) (Lin and Poushinsky, 1985) for field evaluation of a large number of breeding lines and used in many crops such as flax (Soto-Cerda et al., 2014a; Soto-Cerda et al., 2014b; Kumar et al., 2015), wheat (Golparvar et al.,

*Corresponding author. E-mail: collinsagho@yahoo.com.

2015), potato (Schaalje et al., 1987), soybean (Lin and Voldeng, 1989), and barley (May et al., 1989; May and Kozub, 1995). This design can accommodate a large number of unreplicated test genotypes with typically only three control genotypes for error control (You et al., 2013).

Our earlier study (You et al., 2013) demonstrates that soil heterogeneity can be sufficiently adjusted for traits in MAD2 trials. The adjusted observations for test and control genotypes are expected to exclude the effect of soil heterogeneity; thus, the variation among replications of each control genotypes should be solely caused by random errors. A method based on analysis of variance (ANOVA) to approximately estimate broad-sense heritability (H^2) for this design has been developed (You et al., 2016b), in which genetic variance is calculated based on the total phenotypic variance estimated from test genotypes minus the error variance estimated from control genotypes. Because H^2 is always overestimated in single trials, joint analysis over multiple environments for heritability estimation was proposed in which the error variance is jointly estimated using the three replicated control genotypes (You et al., 2016b). However, a caveat to this estimation method is the requirement for the same control genotypes to be used in all environments or trials, and this limits its potential use for joint analysis of data from multi-environment trials with different control genotypes or experimental designs.

There are two units for measurement of phenotypic variances in heritability estimation: on a plot basis and on an entry (or genotype)-mean basis. Estimates of phenotypic variance on an entry-mean basis are always larger than those on a plot basis because the error and interaction variance components are divided by the corresponding numbers of observations (You et al., 2016b). From the perspective of breeding, usually our interest is the heritability estimated from a set of trials. Thus, heritability is usually estimated on an entry-mean basis, that is, the genotype × environment variance is divided by the number of environments. However, we found from empirical results that H^2 on an entry-mean basis was overestimated in multi-environment MAD2 trials (You et al., 2016a); thus, H^2 estimates on a plot basis were put forward and used in MAD2 trials of flax (You et al., 2016a) and in this study.

Herein, we define an inter-environment correlation (r_E) to approximately estimate H^2 in multi-environment MAD2 trials and which does not rely on control genotype data. H^2 estimates from two traditional methods, ANOVA and restricted maximum likelihood (REML), are compared with r_E using computer-simulated and empirical data sets with and without missing data.

MATERIALS AND METHODS

Simulation data

For multi-environment MAD2 trials, a linear model for an adjusted observation (y_{ij}) of *g* test genotypes at *e* environments can be written as:

$$y_{ij} = \mu + G_i + E_j + (GE)_{ij} + \varepsilon_{ij}, (i = 1, 2, \ldots, g; j = 1, 2, \ldots, e), \quad (1)$$

Where $y_{ij} \sim N(\mu, \sigma_P^2)$, $G_i \sim N(0, \sigma_G^2)$, $E_j \sim N(0, \sigma_E^2)$, $(GE)_{ij} \sim N(0, \sigma_{GE}^2)$, and $\varepsilon_{ij} \sim N(0, \sigma_e^2)$. σ_P^2, σ_G^2, σ_E^2, σ_{GE}^2, and σ_e^2 are phenotypic, genetic (G), environment (E), genotype-by-environment (G × E) interactions, and error variances, respectively. σ_e^2 is jointly estimated based on *e* environments with *t* replicated control genotypes in each trial (You et al., 2016b). Typically, *t* is three in MAD2 trials.

For a typical multi-environment MAD2 trial, the data set includes observations for *g* unreplicated test genotypes and three control genotypes at each of *e* environments (representing individual trials in multiple years or/and locations). For the three control genotypes, one main plot control has *m* replications and each of two subplot controls are replicated *n* times. The values of *g*, *m*, *n* and *e* depend on the design. Thus, a total of (*g* + *m* + 2**n*)**e* data points will be generated from a typical multi-environment MAD2 trial. For example, an individual MAD2 trial with a grid of 10 rows by 10 columns contained 100 whole plots with each whole plot being split into 5 subplots, resulting in a total of 500 subplots. The main plot control was placed in the center subplot of each whole plot and so *m* = 100. Two additional subplot controls were randomly assigned to two subplots of 5 randomly selected whole plots and thus *n* = 5 or 2*n* = 10. Test genotypes were accommodated to the remaining 390 subplots (*g* = 390), and the ratio of *m* to *g* and that of *m* to all subplots were 25.6 and 20%, respectively (see Figure 1 in You et al., 2013). To simplify the process, all simulations comprised the same percentage of test genotypes for the main plot control replicates (*m* = 0.25 * *g*) and the same number of subplot control replicates (*n* = 5).

To assess the effect of the number of environments (*e*), number of test genotypes (*g*), and theoretical heritability of a trait on H^2 estimation, simulation data sets for a total of 180 parameter combinations of H^2 (0.1 to 0.9 with an interval of 0.1), *e* (2, 4, 6, 8, and 10), and *g* (50, 100, 200, and 300) were generated. For each parameter combination, 500 simulations were replicated, which was sufficient to represent the sampling distribution of a parameter combination based on results of different simulation runs.

Data sets with 5, 10, 15 and 20% missing data for test genotypes were constructed from each complete data set generated for parameter combinations. Missing data were distributed in all environments at random. The R *sample* function was applied to all simulated observations from *e* environments to randomly assign NULL as missing values. Consequently, a total of 450,000 data sets were generated for analysis.

According to empirical results, plot-based H^2 estimates were more accurate than entry-mean-based estimates in MAD2 trials (You et al., 2016a). Accordingly, heritability estimates on a plot basis were used in this study. Given H^2, σ_{GE}^2, and σ_e^2 for a trait, σ_G^2 can be estimated as $\sigma_G^2 = (\sigma_{GE}^2 + \sigma_e^2)H^2/(1 - H^2)$ on a plot basis. Because traits may have different σ_{GE}^2 and σ_e^2, these two variances were randomly and independently generated. In addition, the environmental variance (σ_E^2) was also randomly generated. Thus, the effects of G, E, G × E, and random error can be simulated according to Equation 1 for a multi-environment MAD2 trial. Similar simulation procedures for data generation have been described previously (You et al., 2016b). The R pipeline program for simulation is available at http://probes.pw.usda.gov/bioinformatics_tools/MADPipeline/index.html.

Empirical data

Three flax biparental genetic populations, namely 243 F_6-derived

recombinant inbred lines (RILs) generated from a cross between CDC Bethune and Macbeth (BM), 90 F_6-derived RILs from a cross between E1747 and Viking (EV), and 78 F_1-derived doubled haploid lines from a cross between SP2047 and UGG5-5 (SU), plus a core collection (CC) of 391 accessions, were field evaluated at two locations in Canada (Morden, Manitoba and Kernen Farm near Saskatoon, Saskatchewan) from 2009 to 2012. The same MAD2 design was employed with the same population in all environments but the designs differed across populations. The experimental designs and phenotyping for the 22 traits in CC, 14 in BM, 19 in EV, and 11 in SU over six to eight environments have been previously described (Cloutier et al., 2010; Soto-Cerda et al., 2014b; Kumar et al., 2015; You et al., 2016b). The adjusted observations in each environment were used for H^2 estimation and evaluation.

Traditional estimation of H^2

The simulated and empirical adjusted observations were used to estimate H^2 on a plot basis (You et al., 2016a; You et al., 2016b). H^2 was estimated as $\hat{H}^2 = \hat{\sigma}_G^2 /(\hat{\sigma}_G^2 + \hat{\sigma}_{GE}^2 + \hat{\sigma}_e^2)$, where $\hat{\sigma}_G^2$, $\hat{\sigma}_{GE}^2$, and $\hat{\sigma}_e^2$ are the genetic, G × E, and error variance, respectively. These variance components were estimated using the method of moments based on both ANOVA (You et al., 2016b) and REML (Holland, 2006; Piepho and Möhring, 2011). The R (https://www.r-project.org/) package *lme4* (Bates et al., 2015) was used to calculate the REML-based variance components. A test for homogeneity of error variance across environments was performed before parameter estimations to satisfy the assumption of the model (Equation 1).

Inter-environment correlation as a H^2 statistic

Based on Equation 1, the inter-environment correlation of adjusted observations ($\boldsymbol{y}$) for two environments (E_1 and E_2) was defined as:

$$r_E(\boldsymbol{y}_1, \boldsymbol{y}_2) = \frac{COV(\boldsymbol{y}_1,\boldsymbol{y}_2)}{\sigma_{y_1}\sigma_{y_2}} = \frac{COV(\boldsymbol{\mu}+\boldsymbol{G}+\boldsymbol{E}_1+(\boldsymbol{GE})_1+\boldsymbol{\varepsilon}_1,\boldsymbol{\mu}+\boldsymbol{G}+\boldsymbol{E}_2+(\boldsymbol{GE})_2+\boldsymbol{\varepsilon}_2)}{\sigma_{y_1}\sigma_{y_2}} = \frac{\sigma_G^2}{(\sigma_G^2+\sigma_{GE}^2+\sigma_e^2)^{\frac{1}{2}}(\sigma_G^2+\sigma_{GE}^2+\sigma_e^2)^{\frac{1}{2}}} = \frac{\sigma_G^2}{\sigma_P^2} = H^2. \quad (2)$$

Here, $COV(\boldsymbol{y}_1,\boldsymbol{y}_2)$ is the covariance of adjusted observations between two environments. It is equal to σ_G^2 because G, E, and random error are hypothesized to be independent of each other and thus the covariance between them is zero, whereas $\sigma_{y_1}\sigma_{y_2} = \sigma_P^2$ because, for a single environment, the environmental variance $\sigma_E^2 = 0$. Therefore, $r_E(\boldsymbol{y}_1, \boldsymbol{y}_2)$ corresponds to an H^2estimate. To validate this definition, independent observations were simulated in two-environment MAD2 trials at given H^2 values (0.1 to 0.9 with 0.1 intervals) and for a number of test genotypes (50, 100, 200 and 300).

For a multi-environment MAD2 trial, r_E was defined as the mean inter-environment correlation coefficient of all possible pairs of environments to estimate the H^2 of a trait:

$$r_E = \sum_{i<j}^{e} r_E(\boldsymbol{y}_i, \boldsymbol{y}_j)/(\frac{e(e-1)}{2}) \ (i, j = 1, 2, \ldots, e), \quad (3)$$

Where e is the number of environments and $r_E = 0$ if $r_E < 0$.

Standard error of H^2 estimates

The delta method was implemented for MAD2 (You et al., 2016b) and used to estimate the standard error (SE) of $\hat{H}^2$ by ANOVA and REML for the empirical data sets. For simulation data, the standard deviation of 500 replicates for each parameter combination was calculated to represent the SE of three H^2 estimates ($s(\hat{H}^2)$ or $s(r_E)$).

Owing to the properties of r_E as a mean Pearson's correlation coefficient, the SE of r_E can also be approximated based on the SE formula of Pearson's correlation (Bowley, 1928; Kendall and Stuart, 1977) using:

$$SE(r_E) = \frac{1-r_E^2}{g^{1/2}}/(\frac{e(e-1)}{2})^{1/2}, \quad (4)$$

Where e is the number of environments and g is the number of test genotypes.

The bootstrap method (Efron, 1979; Xie and Mosjidis, 1997) was also used to estimate the standard error of three heritability estimates for either simulated or empirical data. Bootstrapping was performed by resampling test genotypes with replacement from the original population, and each bootstrap sample had the same size as the original population. The R *sample* function was used to generate the bootstrap samples. The standard deviation of r_E in 500 bootstrap samples was calculated to represent the standard error of r_E, denoted as $s_b(r_E)$ to distinguish from and compare with $SE(r_E)$ in Equation 4, $s(\hat{H}^2)$, or $s(r_E)$ estimated by simulation.

An R pipeline program was developed to automate the simulations and, consequently, the H^2 estimation for three heritability estimation methods. A separate R program for calculating r_E and its SE (bootstrap method) for empirical test genotype data was also created and is available at http://probes.pw.usda.gov/bioinformatics_tools/MADPipeline/index.html.

Statistical power

Statistical power (P) of three heritability estimation methods (r_E, ANOVA, and REML) was calculated for all parameter combinations to determine the minimum number of environments and test genotypes required for heritability estimation and for method comparison. The power of a trait heritability estimate is the probability of correctly rejecting the null hypothesis that heritability is zero when the true heritability of the trait is greater than zero. First, the Z score for the power of a heritability estimate was calculated using (Klein, 1974):

$$Z = Z_\alpha - E(\hat{H}^2)/\sigma_{\hat{H}^2}, \quad (5)$$

Where Z_α is the Z score of the significance level α used for a one-tailed test ($Z_{0.05} = 1.645$); $E(\hat{H}^2)$ is the expected value of the heritability, here the given or parametric heritability value; and $\sigma_{\hat{H}^2}$ is the SE of the expected $\hat{H}^2$ value, estimated by the standard deviation of 500 simulated samples. The one-tailed test was adopted to test whether a heritability estimate is greater than 0 because heritability should be always equal to or greater than zero. The statistical power was defined as the area under the standard normal curve from Z to plus infinity, and can be calculated using an R function:

$$P = 1 - pnorm(-Z), \quad (6)$$

Where *pnorm* is the R function for calculating the cumulative density of the normal distribution.

RESULTS

Genetic and phenotypic variance estimation in the definition of r_E

Simulation results showed a perfectly consistent

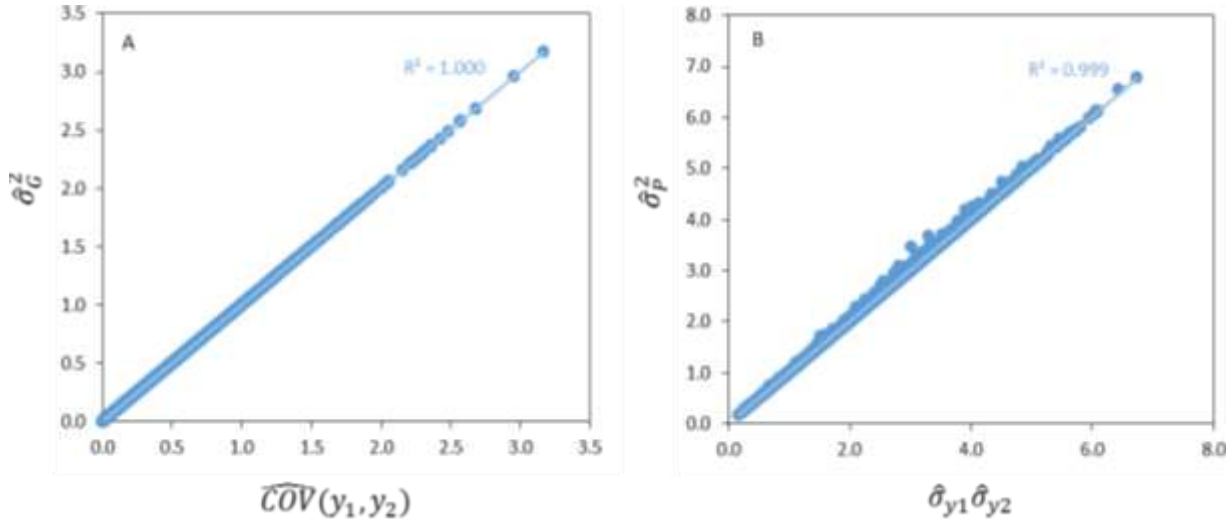

Figure 1. Covariances ($\widehat{COV}(y_1, y_2)$) of trait performance between two environments in relation to genetic variances (σ_G^2) estimated by ANOVA (A), and the product of standard deviations of trait performance from two environments ($\hat{\sigma}_{y1}\hat{\sigma}_{y2}$) in relation to phenotypic variances (σ_P^2) on a plot basis estimated by ANOVA (B).

relationship between $\widehat{COV}(\boldsymbol{y}_1, \boldsymbol{y}_2)$ and $\hat{\sigma}_G^2$ (r = 1.000) (Figure 1A) and between $\hat{\sigma}_{\boldsymbol{y1}}\hat{\sigma}_{\boldsymbol{y2}}$ and $\hat{\sigma}_P^2$ (r = 0.999) (Figure 1B) for the ANOVA-based method, confirming the definition in Equation 2.

H^2 estimates for simulated data sets without missing data

H^2 estimates by r_E, ANOVA ($\widehat{H}^2$(A)), and REML ($\widehat{H}^2$(R)) were calculated for all complete simulated data sets. r_E was perfectly correlated with the given H^2 values (r = 1.000) independent of the number of environments (Figure 2A) or test genotypes (Figure 2C). However, standard errors of r_E ($s(r_E)$) declined with increasing number of environments or test genotypes (Figure 2B and D). For a trait with H^2 equal to or greater than 0.5, the heritability can be correctly estimated at as few as two environments with 50 test genotypes at a statistical power of > 0.999 (Table 1). For traits with low heritability, a larger number of environments and test genotypes contribute positively to more reliable H^2 estimates through increasing statistical power. For an H^2 estimate of 0.2, a statistical power in excess of 0.95 can be achieved from data sets with 300 test genotypes at two environments, 100 test genotypes at four environments, or 50 test genotypes at ≥ six environments when there was no missing data (Table 1).

r_E was consistent with $\widehat{H}^2$(A) for different numbers of environments and test genotypes (Table 2) (r = 1.000). r_E also highly correlated with $\widehat{H}^2$(R) (r > 0.993) (Table 2). High similarity between $\widehat{H}^2$(A) and $\widehat{H}^2$(R) (r > 0.993) was observed (Table 2).

H^2 estimates for simulated data sets with missing data

H^2 estimates by r_E, ANOVA, and REML were calculated for all simulated data sets with missing data of 5 to 25%. The SEs of H^2 estimates by all three methods increased with the rate of missing data. Among the three estimation methods, REML yielded smaller SEs than the r_E- and ANOVA-based methods for both non-missing and missing data sets (Table 3). $\widehat{H}^2$(A) had SEs consistent with r_E for non-missing data sets (R^2 = 1.000) but larger SEs for missing data sets (Figure 3A); r_E had more consistent REs with $\widehat{H}^2$(R) (R^2 = 0.974) for missing data sets (Figure 3B).

The statistical power of H^2 estimates was markedly affected by missing data rates (Tables 3 and 4). Owing to the small SEs, H^2 estimates by REML had higher power than those by r_E, which was also higher than those by ANOVA. The power of r_E and $\widehat{H}^2$(R) was relatively less affected by missing data. For an H^2 estimate of 0.2 in data sets with a missing data rate of ≤5%, a statistical power over 0.95 can be achieved from data sets with 300 test genotypes at two environments, 100 test genotypes at ≤ four environments, or 50 test genotypes at ≤ eight environments for all three estimation methods (Table 4).

Estimates of r_E and $\widehat{H}^2$(R) were less affected by missing data, remaining largely constant with increasing missing data rate at different given H^2 values (Table 3). r_E was consistent with $\widehat{H}^2$ (R) for different numbers of environments and test genotypes (Table 5) (r = 0.995-0.999) but less correlated with $\widehat{H}^2$(A) (r = 0.955-0.996) (Table 5). Bias of $\widehat{H}^2$ (A) and $\widehat{H}^2$ (R) from their true heritability values was observed. At a given H^2 ≤ 0.3,

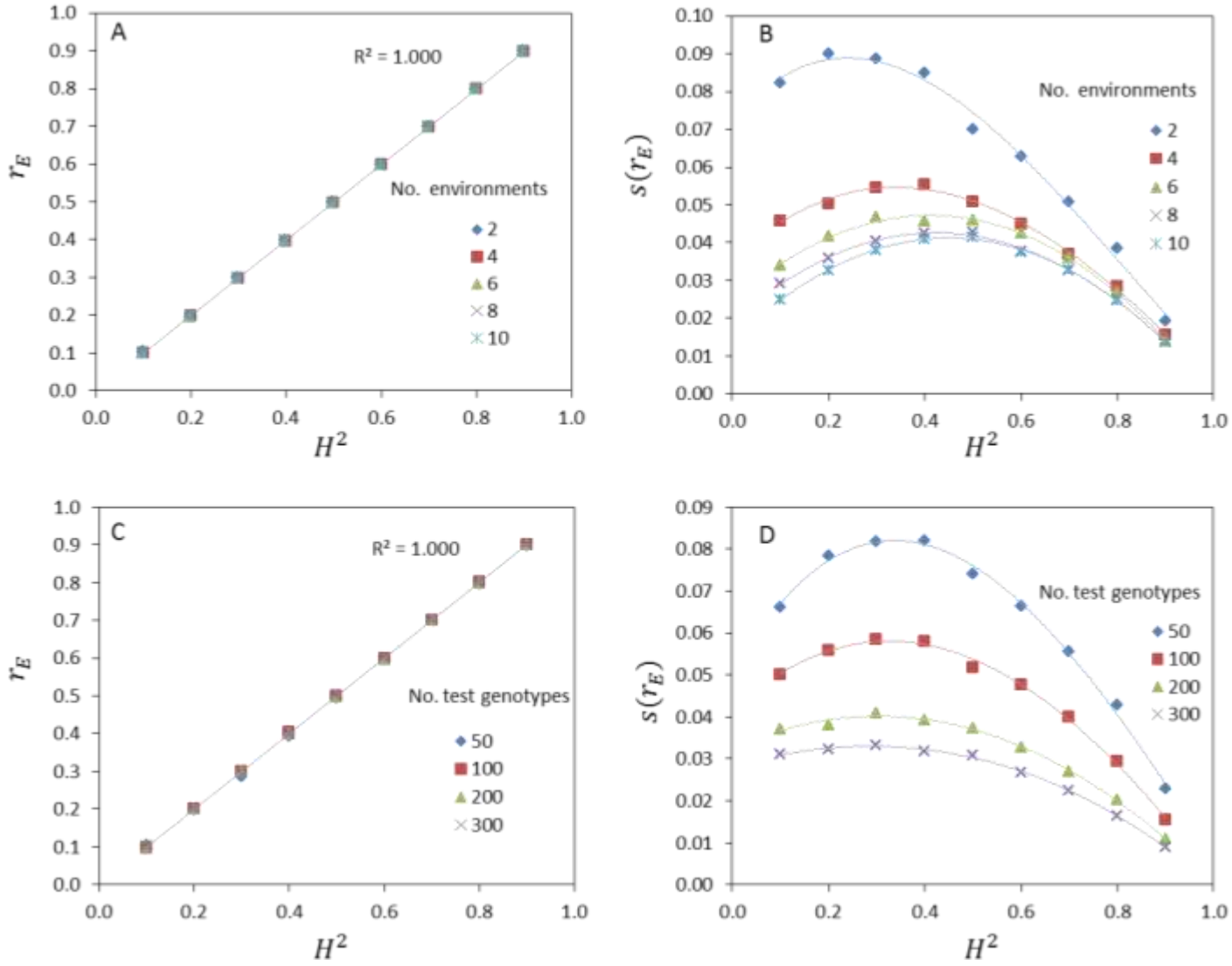

Figure 2. Inter-environment correlations (r_E) and their standard deviations ($s(r_E)$) in terms of heritability (H^2) for different numbers of environments (A and B) and test genotypes (C and D), respectively.

$\widehat{H}^2$(A) was overestimated (Table 3). However, at a given H^2 ≥ 0.4 for $\widehat{H}^2$(A) and ≥ 0.2 for $\widehat{H}^2$(R), $\widehat{H}^2$ values were slightly underestimated (Table 3). A negligible deviation of r_E estimates from their true values was also observed when given H^2 ≥ 0.3 (Table 3).

H^2 estimates for empirical data

As a case study, r_E, $\widehat{H}^2$(A), and $\widehat{H}^2$(R) and their SEs were estimated for 22 traits in the CC, 14 in the BM population, 19 in the EV population, and 11 in the SU population (Table S1). These traits varied greatly in estimated heritability (0.00-0.94), were phenotyped in a different number of environments ranging from two to eight, and evaluated in different populations, that is, one natural and three biparental populations of varying size (Table 6). Here, heritability of some traits was estimated to be 0 due to large experimental errors in data and the ANOVA based model used which resulted in negative genetic variances. The negative genetic variance was treated as 0 in calculation although the real genetic variance should be greater than 0. r_E was highly correlated to both $\widehat{H}^2$(A) (r = 0.948-0.998) and $\widehat{H}^2$ (RP) (r = 0.974-0.998) in individual populations (Table 6). Similar relationships among r_E, $\widehat{H}^2$(A), and $\widehat{H}^2$(R) were observed in different populations despite different numbers of test genotypes (Table 6). Figure 4 depicts the similar overall relationship of r_E with $\widehat{H}^2$ (A) and $\widehat{H}^2$ (R) for all 66 data points. In addition, a strong correlation between $\widehat{H}^2$(A) and $\widehat{H}^2$(R) (r = 0.995-1.000), similar to the results in the simulation data sets, was also observed in the different populations

Table 1. Statistical power (*P*) of broad-sense heritability estimates (H^2) by inter-environment correlation (r_E), ANOVA and REML for α = 0.05 (one-tailed test) as a function of number of environments and test genotypes.

No. environments	No. test genotypes	Method	*P* for H^2				
			0.1	0.2	0.3	0.4	≥ 0.5
2	50	r_E	0.765	0.533	0.769	0.940	≃1
		ANOVA	0.715	0.533	0.786	0.935	≃1
		REML	0.699	0.544	0.795	0.951	≃1
	100	r_E	0.692	0.686	0.940	0.998	≃1
		ANOVA	0.670	0.685	0.941	0.998	≃1
		REML	0.652	0.711	0.947	0.999	≃1
	200	r_E	0.555	0.890	0.999	≃1	≃1
		ANOVA	0.549	0.889	0.999	≃1	≃1
		REML	0.540	0.899	0.999	≃1	≃1
	300	r_E	0.551	0.977	1.000	≃1	≃1
		ANOVA	0.558	0.977	1.000	≃1	≃1
		REML	0.543	0.974	1.000	≃1	≃1
4	50	r_E	0.549	0.872	0.986	≃1	≃1
		ANOVA	0.567	0.872	0.983	0.999	≃1
		REML	0.535	0.875	0.989	≃1	≃1
	100	r_E	0.685	0.986	≃1	≃1	≃1
		ANOVA	0.698	0.986	≃1	≃1	≃1
		REML	0.692	0.990	≃1	≃1	≃1
	200	r_E	0.912	≃1	≃1	≃1	≃1
		ANOVA	0.892	≃1	≃1	≃1	≃1
		REML	0.896	≃1	≃1	≃1	≃1
	300	r_E	0.979	≃1	≃1	≃1	≃1
		ANOVA	0.981	≃1	≃1	≃1	≃1
		REML	0.981	≃1	≃1	≃1	≃1
6	50	r_E	0.649	0.962	0.996	≃1	≃1
		ANOVA	0.673	0.964	0.995	≃1	≃1
		REML	0.673	0.969	0.997	≃1	≃1
	100	r_E	0.908	0.999	≃1	≃1	≃1
		ANOVA	0.905	0.999	≃1	≃1	≃1
		REML	0.909	0.999	≃1	≃1	≃1
	200	r_E	0.991	≃1	≃1	≃1	≃1
		ANOVA	0.990	≃1	≃1	≃1	≃1
		REML	0.992	≃1	≃1	≃1	≃1
	300	r_E	0.999	≃1	≃1	≃1	≃1
		ANOVA	0.999	≃1	≃1	≃1	≃1
		REML	0.999	≃1	≃1	≃1	≃1
8	50	r_E	0.757	0.990	≃1	≃1	≃1
		ANOVA	0.779	0.990	≃1	≃1	≃1
		REML	0.805	0.994	≃1	≃1	≃1
	100	r_E	0.958	≃1	≃1	≃1	≃1
		ANOVA	0.953	≃1	≃1	≃1	≃1
		REML	0.959	≃1	≃1	≃1	≃1
	200	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	0.999	≃1	≃1	≃1	≃1
		REML	0.999	≃1	≃1	≃1	≃1
	300	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	≃1	≃1	≃1	≃1	≃1

Table 1. Contd.

		REML	≃1	≃1	≃1	≃1	≃1
10	50	r_E	0.876	0.996	≃1	≃1	≃1
		ANOVA	0.869	0.997	≃1	≃1	≃1
		REML	0.867	0.996	≃1	≃1	≃1
	100	r_E	0.992	≃1	≃1	≃1	≃1
		ANOVA	0.991	≃1	≃1	≃1	≃1
		REML	0.990	≃1	≃1	≃1	≃1
	200	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	≃1	≃1	≃1	≃1	≃1
		REML	≃1	≃1	≃1	≃1	≃1
	300	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	≃1	≃1	≃1	≃1	≃1
		REML	≃1	≃1	≃1	≃1	≃1

Data sets had no missing data.

Table 2. Correlation of broad-sense heritability ($\widehat{H}^2$) estimated by three methods inter-environment correlation (r_E), ANOVA (A), and REML (R) for simulated data sets without missing data.

No. of environments	No. of test genotypes	$r(r_E, \widehat{H}^2$(A))	$r(r_E, \widehat{H}^2$(R))	$r(\widehat{H}^2$(A), $\widehat{H}^2$(R))
2	50 - 300	1.000	0.992	0.992
4	50 - 300	1.000	0.997	0.997
6	50 - 300	1.000	0.998	0.998
8	50 - 300	1.000	0.998	0.998
10	50 - 300	1.000	0.998	0.998
2 - 10	50	1.000	0.991	0.991
2 - 10	100	1.000	0.997	0.997
2 - 10	200	1.000	0.999	0.999
2 - 10	300	1.000	0.999	0.999
2 - 10	50- 300	1.000	0.997	0.997

$\widehat{H}^2$(A): $\widehat{H}^2$ on a plot basis estimated by ANOVA; $\widehat{H}^2$(R): $\widehat{H}^2$ on a plot basis estimated by REML. Simulated data includes data points generated from combinations of environments (2, 4, 6, and 8), test genotypes (50, 100, 200 and 300), and heritability (0.1-0.9 in 0.1 increments) with 500 replicates for each combination.

Table 3. Effects of missing data on estimation of broad-sense heritability (H^2) and statistical power (*P*).

H^2	MDR (%)	$\widehat{H}^2 \pm s$			*P*		
		r_E	ANOVA	REML	r_E	ANOVA	REML
0.1	0	0.101 ± 0.047	0.101 ± 0.047	0.100 ± 0.045	0.835	0.836	0.836
	5	0.101 ± 0.049	0.121 ± 0.063	0.100 ± 0.047	0.821	0.792	0.821
	10	0.101 ± 0.051	0.136 ± 0.081	0.100 ± 0.048	0.830	0.823	0.831
	15	0.102 ± 0.053	0.150 ± 0.095	0.101 ± 0.050	0.812	0.749	0.815
	20	0.102 ± 0.057	0.163 ± 0.108	0.101 ± 0.053	0.806	0.710	0.811
0.2	0	0.200 ± 0.054	0.200 ± 0.053	0.199 ± 0.052	0.944	0.945	0.948
	5	0.200 ± 0.056	0.212 ± 0.064	0.198 ± 0.054	0.931	0.913	0.935
	10	0.200 ± 0.058	0.221 ± 0.077	0.198 ± 0.055	0.936	0.935	0.940
	15	0.201 ± 0.061	0.229 ± 0.088	0.199 ± 0.057	0.926	0.893	0.931
	20	0.200 ± 0.064	0.235 ± 0.099	0.198 ± 0.060	0.911	0.866	0.918
0.3	0	0.298 ± 0.057	0.298 ± 0.057	0.296 ± 0.056	0.984	0.984	0.985
	5	0.298 ± 0.059	0.303 ± 0.063	0.296 ± 0.057	0.977	0.972	0.980

Table 3. Contd.

	10	0.298 ± 0.062	0.306 ± 0.072	0.296 ± 0.059	0.982	0.981	0.984
	15	0.298 ± 0.064	0.308 ± 0.081	0.296 ± 0.061	0.973	0.963	0.978
	20	0.298 ± 0.067	0.310 ± 0.089	0.296 ± 0.064	0.968	0.951	0.973
0.4	0	0.398 ± 0.055	0.397 ± 0.054	0.395 ± 0.053	0.998	0.997	0.998
	5	0.398 ± 0.057	0.397 ± 0.058	0.395 ± 0.055	0.995	0.996	0.996
	10	0.398 ± 0.060	0.395 ± 0.064	0.394 ± 0.057	0.996	0.997	0.998
	15	0.398 ± 0.062	0.393 ± 0.071	0.395 ± 0.058	0.993	0.993	0.996
	20	0.397 ± 0.066	0.389 ± 0.078	0.394 ± 0.061	0.989	0.989	0.994
0.5	0	0.499 ± 0.052	0.498 ± 0.051	0.494 ± 0.051	≃1	≃1	≃1
	5	0.499 ± 0.053	0.493 ± 0.054	0.494 ± 0.052	≃1	≃1	≃1
	10	0.499 ± 0.056	0.488 ± 0.058	0.494 ± 0.053	≃1	≃1	≃1
	15	0.499 ± 0.059	0.482 ± 0.063	0.494 ± 0.055	0.999	≃1	≃1
	20	0.498 ± 0.061	0.475 ± 0.069	0.493 ± 0.057	0.999	0.999	≃1
0.6	0	0.598 ± 0.048	0.597 ± 0.048	0.591 ± 0.047	≃1	≃1	≃1
	5	0.598 ± 0.050	0.590 ± 0.050	0.591 ± 0.048	≃1	≃1	≃1
	10	0.597 ± 0.051	0.582 ± 0.053	0.591 ± 0.050	≃1	≃1	≃1
	15	0.597 ± 0.054	0.574 ± 0.056	0.591 ± 0.051	≃1	≃1	≃1
	20	0.597 ± 0.056	0.566 ± 0.060	0.591 ± 0.052	≃1	≃1	≃1
0.7	0	0.698 ± 0.039	0.697 ± 0.039	0.688 ± 0.040	≃1	≃1	≃1
	5	0.698 ± 0.040	0.689 ± 0.040	0.688 ± 0.041	≃1	≃1	≃1
	10	0.698 ± 0.041	0.681 ± 0.043	0.688 ± 0.042	≃1	≃1	≃1
	15	0.698 ± 0.043	0.672 ± 0.045	0.688 ± 0.042	≃1	≃1	≃1
	20	0.698 ± 0.045	0.663 ± 0.048	0.689 ± 0.043	≃1	≃1	≃1
0.8	0	0.799 ± 0.028	0.797 ± 0.028	0.784 ± 0.034	≃1	≃1	≃1
	5	0.799 ± 0.029	0.791 ± 0.030	0.784 ± 0.034	≃1	≃1	≃1
	10	0.798 ± 0.030	0.783 ± 0.031	0.785 ± 0.034	≃1	≃1	≃1
	15	0.798 ± 0.032	0.775 ± 0.033	0.785 ± 0.034	≃1	≃1	≃1
	20	0.798 ± 0.033	0.767 ± 0.036	0.785 ± 0.035	≃1	≃1	≃1
0.9	0	0.899 ± 0.016	0.898 ± 0.016	0.877 ± 0.030	≃1	≃1	≃1
	5	0.899 ± 0.016	0.894 ± 0.017	0.878 ± 0.030	≃1	≃1	≃1
	10	0.899 ± 0.017	0.889 ± 0.018	0.878 ± 0.030	≃1	≃1	≃1
	15	0.899 ± 0.018	0.884 ± 0.019	0.878 ± 0.030	≃1	≃1	≃1
	20	0.899 ± 0.018	0.879 ± 0.020	0.879 ± 0.030	≃1	≃1	≃1

MDR, Missing data rate on an entire multi-environment trial basis; H^2 was calculated based on 10,000 data points generated from combinations of various numbers of environments (2, 4, 6, 8 and 10) and test genotypes (50, 100, 200 and 300) with 500 replicates for each combination; Statistical power was the average of power estimates from 20 parametric sets of different numbers of environments and test genotypes.

(Table 6). Because the empirical data had missing values in some environments (Table S1), a slightly weaker relationship among r_E, $\hat{H}^2$(A) and $\hat{H}^2$(R), and a stronger relation between r_E and $\hat{H}^2$(R) than that between r_E and $\hat{H}^2$(A) (Table 6) were observed, which resembled the results obtained in missing data sets.

Standard error of r_E

To perform a significance test for r_E estimates, the SEs of r_E determined by Eq. 4 ($SE(r_E)$) and by the bootstrap method ($s_b(r_E)$) were compared to the corresponding simulated SEs ($s(r_E)$). Although a strong correlation of $s(r_E)$ to $SE(r_E)$ was observed (R^2 = 0.815) (Figure 5A), a high correlation (R^2 = 0.995) was obtained for $s(r_E)$ with $s_b(r_E)$ (Figure 5B) indicating the bootstrap method yielded SE estimates for r_E that were highly consistent with those obtained by simulation. The SE estimates obtained by the bootstrap method were systematically smaller (by 0.0015 on average) than those obtained by simulation. A linear regression equation, $s(r_E) = 1.0802^* s_b(r_E) - 0.0017$ (Figure 5B), may be used to adjust the bootstrap estimates.

DISCUSSION

H^2 is the ratio of genetic variance to total phenotypic variance, representing the extent with which genotypes

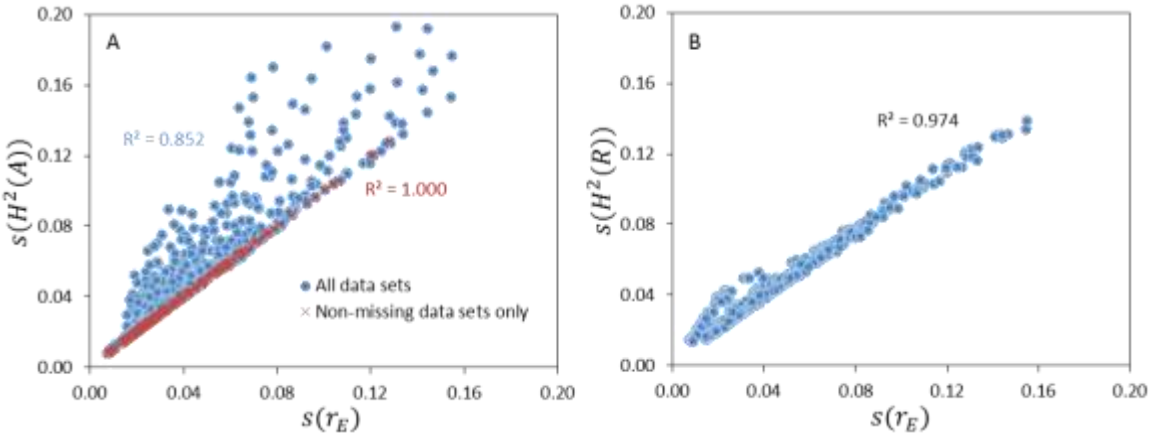

Figure 3. Relationship of standard errors (SEs) of inter-environment correlation (r_E) with SEs of H^2estimates by ANOVA (A) and REML (R). (A) SEs of r_E estimated from simulations ($s(r_E)$) vs. SEs of H^2estimates by ANOVA ($s(\hat{H}^2(A)$) using all data sets (missing and non-missing data) (blue) and non-missing data only (red). (B) $s(r_E)$) vs. $s(\hat{H}^2(R))$ using all data sets.

Table 4. Statistical power (*P*) of broad-sense heritability estimates ($\hat{H}^2$) at a given heritability of 0.2 by inter-environment correlation (r_E), ANOVA, and REML for α = 0.05 (one-tailed test) as a function of the number of environments, and test genotypes, and missing data rates.

No. of environments	No. of test genotypes	Method	*P* for missing data rate (%)				
			0	5	10	15	20
2	50	r_E	0.533	0.503	0.518	0.543	0.571
		ANOVA	0.533	0.527	0.507	0.500	0.528
		REML	0.544	0.516	0.502	0.523	0.543
	100	r_E	0.686	0.640	0.606	0.611	0.534
		ANOVA	0.685	0.673	0.653	0.619	0.557
		REML	0.711	0.666	0.643	0.623	0.563
	200	r_E	0.890	0.859	0.837	0.818	0.733
		ANOVA	0.889	0.812	0.705	0.639	0.596
		REML	0.899	0.875	0.858	0.835	0.755
	300	r_E	0.977	0.959	0.953	0.929	0.908
		ANOVA	0.977	0.889	0.775	0.703	0.642
		REML	0.974	0.956	0.947	0.929	0.903
4	50	r_E	0.872	0.849	0.821	0.793	0.733
		ANOVA	0.872	0.886	0.835	0.807	0.741
		REML	0.875	0.859	0.832	0.832	0.763
	100	r_E	0.986	0.976	0.976	0.955	0.913
		ANOVA	0.986	0.982	0.960	0.918	0.858
		REML	0.990	0.984	0.982	0.969	0.947
	200	r_E	≃1	≃1	≃1	0.999	0.998
		ANOVA	≃1	0.999	0.979	0.942	0.888
		REML	≃1	≃1	≃1	≃1	0.999
	300	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	≃1	≃1	0.993	0.966	0.917
		REML	≃1	≃1	≃1	≃1	≃1
6	50	r_E	0.962	0.946	0.934	0.916	0.894
		ANOVA	0.964	0.957	0.929	0.897	0.866
		REML	0.969	0.959	0.949	0.939	0.925
	100	r_E	0.999	0.998	0.997	0.994	0.992

Table 4. Contd.

		ANOVA	0.999	0.996	0.976	0.938	0.903
		REML	0.999	0.999	0.998	0.997	0.996
	200	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	≃1	≃1	≃1	0.999	0.994
		REML	≃1	≃1	≃1	≃1	≃1
	300	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	≃1	≃1	0.999	0.988	0.960
		REML	≃1	≃1	≃1	≃1	≃1
8	50	r_E	0.990	0.988	0.979	0.971	0.958
		ANOVA	0.990	0.985	0.970	0.960	0.929
		REML	0.994	0.992	0.990	0.987	0.977
	100	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	≃1	≃1	0.999	0.995	0.986
		REML	≃1	≃1	≃1	≃1	≃1
	200	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	≃1	≃1	≃1	0.999	0.995
		REML	≃1	≃1	≃1	≃1	≃1
	300	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	≃1	≃1	≃1	≃1	0.999
		REML	≃1	≃1	≃1	≃1	≃1
10	50	r_E	0.996	0.995	0.990	0.988	0.988
		ANOVA	0.997	0.994	0.988	0.983	0.971
		REML	0.996	0.996	0.994	0.992	0.994
	100	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	≃1	≃1	≃1	0.999	0.995
		REML	≃1	≃1	≃1	≃1	≃1
	200	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	≃1	≃1	≃1	≃1	≃1
		REML	≃1	≃1	≃1	≃1	≃1
	300	r_E	≃1	≃1	≃1	≃1	≃1
		ANOVA	≃1	≃1	≃1	≃1	≃1
		REML	≃1	≃1	≃1	≃1	≃1

are affected by environments and random error. H^2 also represents the repeatability of trait performance in different environments. For highly heritable traits, the performance of genotypes in one environment has a high repeatability in other environments; in other words, the trait performance of the genotypes is strongly correlated between any pair of environments. In contrast, low heritability traits display low correlations of trait performance between any two environments. As such, the inter-environment correlation (r_E) of trait performance should be an indicator of trait heritability. The theoretical derivation confirms that the r_E between two environments corresponds to H^2 (Equation 2). Both simulation and empirical results demonstrate that r_E is an accurate and stable estimate of H^2 on a plot basis.

The r_E method has two significant advantages. First, because r_E is a mean correlation coefficient between pairs of environments, its calculation is simple. Second, only the test genotype data is required for calculation of r_E, eliminating the prerequisite for the use of the same control genotypes across all trials and hence permitting flexible field designs. Thus, the r_E method allows joint H^2 estimation over multiple environments for genetic populations that may differ in their MAD2 designs and which may not necessarily include the same control genotypes. The r_E method is equally applicable to trials with the same test genotypes but where different control genotypes are used in the different environments. A practical example was the joint H^2estimation for three flax biparental populations: BM, EV, and SU. These three populations were evaluated in the same six to eight environments using MAD2 designs but with different

Table 5. Correlation of broad-sense heritability ($\hat{H}^2$) estimated by three methods inter-environment correlation (r_E), ANOVA (A), and REML (R) for simulated data sets with missing data rates of 5 to 20%.

No. of environments	No. of test genotypes	$r(r_E, \hat{H}^2$(A))	$r(r_E, \hat{H}^2$(R))	$r(\hat{H}^2$(A), $\hat{H}^2$(R))
2	50 - 300	0.955	0.988	0.945
4	50 - 300	0.984	0.995	0.980
6	50 - 300	0.991	0.997	0.989
8	50 - 300	0.995	0.998	0.993
10	50 - 300	0.996	0.998	0.994
2 - 10	50	0.980	0.998	0.970
2 - 10	100	0.982	0.986	0.979
2 - 10	200	0.984	0.998	0.983
2 - 10	300	0.985	0.999	0.984
2 - 10	50 - 300	0.983	0.995	0.979

$\hat{H}^2$(A): $\hat{H}^2$ on a plot basis estimated by ANOVA; $\hat{H}^2$(R): $\hat{H}^2$ on a plot basis estimated by REML. Simulated data includes data points generated from combinations of environments (2, 4, 6, and 8), test genotypes (50, 100, 200 and 300), and heritability (0.1-0.9 in 0.1 increments) with 500 replicates for each combination.

Table 6. Correlation of broad-sense heritability ($\hat{H}^2$) estimated by three methods inter-environment correlation (r_E), ANOVA (A), and REML (R) for empirical data.

Population	No. of test genotypes	No. of traits	$r(r_E, \hat{H}^2$(A))	$r(r_E, \hat{H}^2$(R))	$r(\hat{H}^2$(A), $\hat{H}^2$(R))
CC	391	22	0.986	0.992	0.995
BM	243	14	0.982	0.982	1.000
EV	90	19	0.948	0.974	0.980
SU	78	11	0.998	0.998	1.000
Total	-	66	0.975	0.985	0.989

CC, Core collection; BM, CDC Bethune/Macbeth; EV, E1747/Viking; SU, SP2047/UGG5-5; $\hat{H}^2$(A): $\hat{H}^2$ on a plot basis estimated by ANOVA; $\hat{H}^2$(R): $\hat{H}^2$ on a plot basis estimated by REML. The estimate for each trait was used as a single data point to calculate the correlation among r_E, $\hat{H}^2$(A), and $\hat{H}^2$(R). No. of traits in each population represent the number of data points used for calculation of correlation between any two methods. Total represents all 66 data points from four individual populations for correlation calculation.

control genotypes. In each MAD2 design, CDC Bethune was used as the main plot control but the subplot controls were the two parents of each of the three biparental populations to improve error control because the two parents shared the genetic background of their offspring. As a consequence of the use of different control genotypes in the three population trials, neither ANOVA- nor REML-based methods could estimate joint H^2 values, but this was achievable with the r_E method using pooled data. The results will be reported separately.

In the ANOVA-based H^2 estimation, the error variance of unreplicated test genotypes is estimated by duplicated control genotypes. This is based on the assumption that control genotypes share the same random error variance with all test genotypes. Theoretically, the total mean square (MS) of the test genotypes will be greater than the error MS in a single trial or the G × E interaction variance in a multi-environment trial. As such, the genetic variance of the test genotypes can be estimated as the total MS minus the error MS of a single trial or by the total variance minus the G × E interaction MS in a multi-environment trial. However, the sampling bias caused by a limited number of control genotypes (typically three in MAD2) may occasionally result in negative genetic variance estimates and failure to correctly estimate genetic parameters, especially when the heritability of a trait is very low. In this case, the r_E method can avoid this potential drawback because the genetic variance in r_E corresponds to the covariance of trait performance between two environments ($COV(y_1, y_2)$) (Equation 2 and Figure 1A). If the $COV(y_1, y_2)$ is less than zero then the genetic variance can be considered zero, whereas negative or null genetic variance obtained by ANOVA or REML might be an incorrect estimate in some cases. In the core collection, for example, both oil and linolenic acid contents have H^2 estimates of zero because their genetic variances were estimated to be zero by both ANOVA- and REML-based methods. This result is

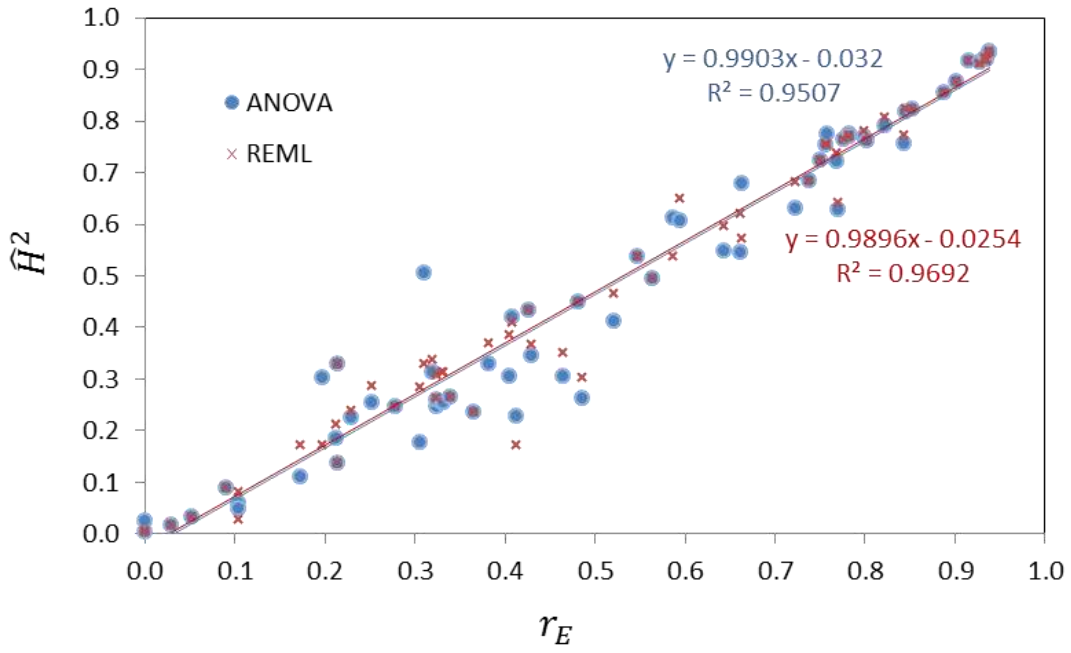

Figure 4. Relationship of inter-environment correlations (r_E) with broad-sense heritability ($\hat{H}^2$) estimated from empirical data by ANOVA and REML. Data points include estimates for r_E and $\hat{H}^2$ from 22 traits in the core collection, 14 in the CDC Bethune/Macbeth, 19 in the E1747/Viking, and 11 in the SP20147/UGG5-5 populations.

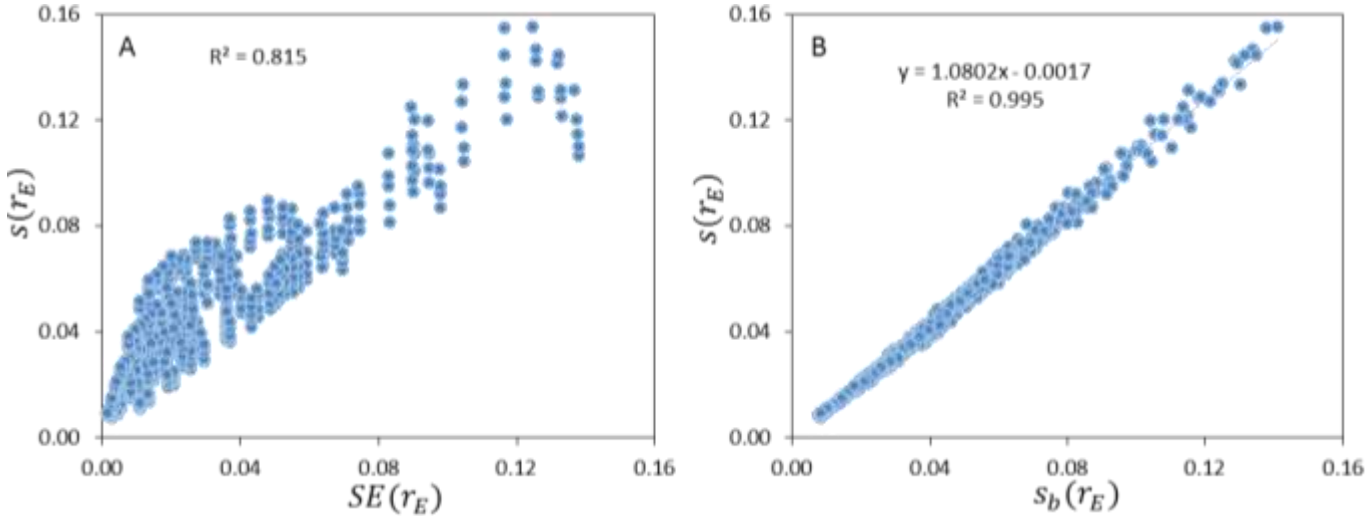

Figure 5. Relationship of standard error (*SE*) of inter-environment correlation (r_E) with *SE* estimates by other methods. (A) $s(r_E)$ vs. *SE* of r_E calculated by Equation 4 ($SE(r_E)$). (B)$s(r_E)$ vs. *SE* estimated by the bootstrap method ($s_b(r_E)$). All data sets including missing and non-missing data) were used.

obviously incorrect because oil and linolenic acid content are traits of moderate to high heritability (You et al., 2016a). However, r_E-based H^2 estimates for the two traits were 0.387 and 0.661, respectively, although these estimates are still smaller than expected.

The REML-based method, as an alternative to the more traditional ANOVA techniques, has been widely used for estimation of heritability and genetic correlations (Holland, 2006; Piepho and Möhring, 2011). The advantages compared to ANOVA methods are that REML estimates of variance and covariance components have known asymptotic distributional properties and can

efficiently handle both unbalanced data and complex experimental designs (Meyer, 1985; Holland, 2006). REML's main drawback is that it is much more computationally intensive than ANOVA, a disadvantage that is exacerbated with large data sets. In this study, these two H^2 estimation methods were compared to the r_E method using both simulation and empirical data of multi-environment MAD2 trials. All results corroborate the agreements between the ANOVA- and REML-based methods with $r > 0.99$ in the simulated data sets without missing data, especially in the empirical data of BM and SU populations where their estimates are nearly identical (Table 2). The ANOVA-based H^2estimates are consistent with the r_E estimates when no missing data exist in the data sets. Nevertheless, when missing values occur, the r_E- and REML-based methods show higher power than the ANOVA-based method, confirming that the REML-based method is efficient for tackling unbalanced data. Overall, r_E-, ANOVA-, and REML-based methods can be used for non-missing data sets; the r_E- and REML-based methods are suitable for missing data sets; and the r_E method is versatile for all cases of practical data sets in multi-environment MAD2 trials.

A significance test for r_E estimates requires SE values. Generally, SEs ($s(r_E)$) estimated by simulation with a large sample size (500 in this study) provide a good estimate of the sampling error. Thus, they were used as a relative standard for comparisons. Two potential methods were assessed: the approximate standard error ($SE(r_E)$) of mean simple correlation coefficients (Equation 4) and bootstrap. The non-parametric bootstrap is an effective alternative for determining distribution of an estimator with an unknown probability density, and has been used to estimate standard errors for heritability (Xie and Mosjidis, 1997). Results demonstrate that the bootstrap method outperforms the SE formula of simple correlation coefficient (Equation 4), and provides perfectly consistent SE estimates with $s(r_E)$. However, a systematic difference between $s(r_E)$ and $s_b(r_E)$ was observed; adjustments may be done using the regression equation in Figure 5B, which was constructed using 900 data points generated from different numbers of environments (2, 4, 6, 8, and 10), test genotypes (50, 100, 200 and 300), and missing data rates (0, 5, 10, 15 and 20%) at given heritability values ranging from 0.1 to 0.9. Therefore, the bootstrap method is recommended to estimate the standard error of r_E in significance tests. The bootstrap estimation of *SE* has been implemented in the R pipeline program (http://probes.pw.usda.gov/bioinformatics_tools/MADPipeline/index.html).

To find the effective sample size for estimating H^2 in multi-environment MAD2 trials, the statistical power of three H^2 estimation methods was calculated. As expected, the power is affected by the number of environments and test genotypes, and missing data rates. For traits with H^2 equal to or greater than 0.5, 50 test genotypes at two environments were sufficient to achieve a statistical power over 0.95 (Table 1); for traits with lower heritability (e.g., 0.2), 300 test genotypes at two environments, 100 test genotypes at four environments, or 50 test genotypes at greater than or equal to six environments are required to obtain the same statistical power (Table 1). Increasing the missing data rate decreased the statistical power (Table 4), but an increase of the number of environments and/or test genotypes markedly improved the statistical power of H^2 estimates.

Notably, heritability estimates in the simulation data sets were slightly biased from their true values when true H^2 values were greater than 0.2 (Table 3). This bias was observed in all three H^2 estimation methods, especially when the ANOVA-based method was used for missing data sets. However, the r_E method has less deviation than other two for both missing and non-missing data sets. This bias may be inherently owing to the MAD2's unbalanced feature. Piepho and Möhring (2007) discussed how estimation of broad-sense heritability in unbalanced trials differs from that in the case of a balanced design. The r_E proposed here, as well as the ANOVA- (You et al., 2016b) and REML- (You et al., 2016a) based methods, provide an approximate H^2 estimate for MAD2 trials. This approximation is due to not only unbalanced data but also approximate assumption of independence for adjusted observations from an MAD2 trial.

For adjustment of observations in an MAD2 trial, there are four different cases for quantitative traits: (1) significant additive soil variation due to row or column effects (M1); (2) significant non-additive soil variation due to row × column interaction effects (M3); (3) M1+M3, and (4) no additive or non-additive soil variation (You et al., 2013). For case 4, as no data adjustment is required, their estimates of heritability will be unbiased, while for the first three cases, data adjustment are needed and the adjusted data may be correlated to some extent, resulting in biased estimates. In this study, the simulation data was completely independent of each other, but the empirical data of some traits were unnecessarily completely independent due to data adjustment. However, we found that highly similar or consistent results were obtained in both data sets, indicating small effect of defective independence assumption to heritability estimation. In our actual trials, especially those with good quality, most of the traits would not need data adjustment or have minor row or/and column or their interaction effects, the bias due to defective independence would be small and may be disregarded in breeding applications.

Conclusion

A H^2 statistic, r_E, representing an inter-environment correlation of a quantitative trait, was presented for multi-environment MAD2 trials. The r_E method provides a simple approach to approximate H^2 in any multi-

environment MAD2 trial without the limitations of ANOVA- or REML-based methods that require the use of the same control genotypes across trials and/or environments.

Conflicts of Interests

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENTS

This work was conducted as part of the A-base project (No. 1142) funded by Agriculture and Agri-Food Canada, the flax breeding database project funded by the Western Grain Research Foundation (WGRF), and the Total Utilization Flax GENomics (TUFGEN) project funded by Genome Canada and other stakeholders.

REFERENCES

Bates D, Mäechler M, Bolker BM, Walker SC (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67(1):1-48.

Bowley AL (1928). The standard deviation of the correlation coefficient. J. Am. Stat. Assoc. 23(161):31-34.

Cloutier S, Ragupathy R, Niu Z, Duguid S (2010). SSR-based linkage map of flax (*Linum usitatissimum* L.) and mapping of QTLs underlying fatty acid composition traits. Mol. Breed. 28(4):437-451.

Efron B (1979). Bootstrap methods: another look at the jacknife. Ann. Stat. 7:1-26.

Golparvar AR, Gheisari MM, Naderi D, Mehrabi AM, Hadipanah A, Salehi S (2015). Determination of the best indirect selection criteria in Iranian durum wheat (*Triticum aestivum* L.) genotypes under irrigated and drought stress conditions. Genetika 47:549-558.

Holland JB (2006). Estimating genotypic correlations and their standard errors using multivariate restricted maximum likelihood estimation with SAS Proc MIXED. Crop Sci. 46(2):642-654.

Holland JB, Nyquist WE, Cervantes-Martinez GT (2003). Estimating and interpreting heritability for plant breeding: an update. Plant Breed. Rev. 22:9-112.

Kendall MG, Stuart A (1977). The advanced theory of statistics, Distribution theory. (London: Griffin).

Klein TW (1974). Heritability and genetic correlation: statistical power, population comparisons, and sample size. Behav. Genet. 4(2):171-189.

Kumar S, You FM, Duguid S, Booker H, Rowland G, Cloutier S (2015). QTL for fatty acid composition and yield in linseed (*Linum usitatissimum* L.). Theor. Appl. Genet. 128(5):965-984.

Lin C, Poushinsky G (1983). A modified augmented design for an early stage of plant selection involving a large number of test lines without replication. Biometrics 39(3):553-561.

Lin CS, Poushinsky G (1985). A modified augmented design (type 2) for rectangular plots. Can. J. Plant Sci. 65(3):743-749.

Lin CS, Voldeng HD (1989). Efficiency of type 2 modified augmented designs in soybean variety trials. Agron. J. 81:512-517.

May KW, Kozub GC (1995). Success of a selection program for increasing grain yield of two-row barley lines and evaluation of the modified augmented design (type 2). Can. J. Plant Sci. 75(4):795-799.

May KW, Kozub GC, Schaalje GB (1989). Field evaluation of a modified augmented design (type 2) for screening barley lines. Can. J. Plant Sci. 69(1):9-15.

Meyer K (1985). Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. Biometrics 41(1):153-165.

Piepho HP, Möhring J (2007). Computing heritability and selection response from unbalanced plant breeding trials. Genetics 177:1881-1888.

Piepho HP, Möhring J (2011). On estimation of genotypic correlations and their standard errors by multivariate REML using the MIXED procedure of the SAS system. Crop Sci. 51(6):2449-2454.

Schaalje GB, Lynch DR, Kozub GC (1987). Field evaluation of a modified augmented design for early stage selection involving a large number of test lines without replication. Potato Res. 30:35-45.

Soto-Cerda BJ, Duguid S, Booker H, Rowland G, Diederichsen A, Cloutier S (2014a). Genomic regions underlying agronomic traits in linseed (*Linum usitatissimum* L.) as revealed by association mapping. J. Integr. Plant Biol. 56(1):75-87.

Soto-Cerda BJ, Duguid S, Booker H, Rowland G, Diederichsen A, Cloutier S (2014b). Association mapping of seed quality traits using the Canadian flax (*Linum usitatissimum* L.) core collection. Theor. Appl. Genet. 127(4):881-896.

Xie C, Mosjidis JA (1997). Influence of sample size on precision of heritability and expected selection response in red clover. Plant Breed. 116:83-88.

You FM, Duguid SD, Thambugala D, Cloutier S (2013). Statistical analysis and field evaluation of the type 2 modified augmented design (MAD) in phenotyping of flax (*Linum usitatissimum*) germplasms in multiple environments. Aust. J. Crop Sci. 7(11):1789-1800.

You FM, Booker MH, Duguid DS, Jia G, Cloutier S (2016a). Accuracy of genomic selection in biparental populations of flax (*Linum usitatissimum* L.). Crop J. 4(4):290-303.

You FM, Song Q, Jia G, Cheng Y, Duguid S, Booker H, Cloutier S (2016b). Estimation of genetic parameters and their sampling variances for quantitative traits in the type 2 modified augmented design. Crop J. 4(2):107-118.

Supplementary data

Table S1. Broad-sense heritability estimates ($\widehat{H}^2$) determined by inter-environment correlation (r_E), ANOVA (A), and REML (R), and their standard error (SE) for traits for three flax biparental populations (BM, EV, and SU) and the flax core collection (CC).

Population	Trait	MDR (%)	r_E ± SE	$\widehat{H}^2(A)$ ± SE	$\widehat{H}^2(R)$ ± SE
CC	Bolls (m^{-2})	6.07	0.319 ± 0.051	0.313 ± 0.022	0.337 ± 0.023
CC	Cellulose content (%)	2.23	0.306 ± 0.046	0.178 ± 0.014	0.284 ± 0.020
CC	Cell wall content (%)	2.23	0.173 ± 0.050	0.111 ± 0.012	0.170 ± 0.017
CC	Fiber content (%)	2.23	0.331 ± 0.052	0.253 ± 0.018	0.312 ± 0.021
CC	Days to 5% flowering	2.05	0.643 ± 0.043	0.548 ± 0.020	0.595 ± 0.020
CC	Days to 95% flowering	2.40	0.662 ± 0.038	0.546 ± 0.020	0.619 ± 0.019
CC	Plant height (cm)	2.05	0.594 ± 0.033	0.608 ± 0.019	0.650 ± 0.019
CC	Iodine value	3.11	0.800 ± 0.026	0.769 ± 0.014	0.780 ± 0.014
CC	Lignin content (%)	2.23	0.324 ± 0.051	0.247 ± 0.018	0.307 ± 0.021
CC	Linoleic content (%)	3.11	0.934 ± 0.050	0.919 ± 0.006	0.921 ± 0.006
CC	Lodging	1.98	0.104 ± 0.048	0.060 ± 0.010	0.081 ± 0.014
CC	Days to maturity	2.23	0.212 ± 0.055	0.185 ± 0.017	0.212 ± 0.019
CC	Mildew score	5.53	0.522 ± 0.038	0.411 ± 0.028	0.466 ± 0.031
CC	Oleic content (%)	3.11	0.768 ± 0.028	0.722 ± 0.016	0.736 ± 0.016
CC	Palmitic content (%)	3.11	0.822 ± 0.027	0.791 ± 0.013	0.806 ± 0.012
CC	Pasmo score	3.02	0.253 ± 0.042	0.256 ± 0.022	0.288 ± 0.025
CC	Plant branching score	9.85	0.000 ± 0.053	0.004 ± 0.013	0.006 ± 0.018
CC	Shive content (%)	2.23	0.332 ± 0.052	0.254 ± 0.018	0.312 ± 0.021
CC	Protein content (%)	3.46	0.723 ± 0.061	0.631 ± 0.021	0.680 ± 0.020
CC	Stearic content (%)	3.11	0.845 ± 0.023	0.817 ± 0.012	0.822 ± 0.011
CC	Thousand seed weight (g)	0.38	0.770 ± 0.028	0.629 ± 0.030	0.641 ± 0.030
CC	Seed yield ($T\ H^{-1}$)	2.84	0.405 ± 0.044	0.306 ± 0.020	0.386 ± 0.022
BM	Cell wall content (%)	0.29	0.091 ± 0.064	0.089 ± 0.020	0.089 ± 0.020
BM	Iodine value	0.46	0.783 ± 0.032	0.769 ± 0.018	0.769 ± 0.018
BM	Linoleic content (%)	0.46	0.756 ± 0.030	0.755 ± 0.019	0.755 ± 0.019
BM	Linolenic content (%)	0.46	0.783 ± 0.032	0.774 ± 0.018	0.773 ± 0.018
BM	Days to maturity	0.00	0.427 ± 0.050	0.432 ± 0.031	0.432 ± 0.031
BM	Oil content (%)	0.36	0.564 ± 0.047	0.494 ± 0.027	0.494 ± 0.027
BM	Oleic content (%)	0.46	0.777 ± 0.029	0.764 ± 0.018	0.764 ± 0.018
BM	Palmitic content (%)	0.46	0.803 ± 0.029	0.763 ± 0.018	0.762 ± 0.018
BM	Seeds per boll	0.55	0.366 ± 0.056	0.234 ± 0.028	0.235 ± 0.028
BM	Protein content (%)	0.00	0.215 ± 0.061	0.328 ± 0.029	0.328 ± 0.029
BM	Stearic content (%)	0.46	0.852 ± 0.028	0.822 ± 0.015	0.823 ± 0.014
BM	Straw weight (g)	0.51	0.340 ± 0.055	0.266 ± 0.035	0.266 ± 0.035
BM	Thousand seed weight (g)	0.55	0.324 ± 0.071	0.261 ± 0.028	0.261 ± 0.028
BM	Seed yield ($T\ H^{-1}$)	0.26	0.029 ± 0.061	0.016 ± 0.013	0.016 ± 0.013
EV	Cellulose content (%)	0.00	0.383 ± 0.087	0.329 ± 0.045	0.368 ± 0.043
EV	Cell wall content (%)	0.00	0.052 ± 0.099	0.034 ± 0.027	0.033 ± 0.028
EV	Fiber content (%)	0.00	0.429 ± 0.084	0.344 ± 0.045	0.368 ± 0.048
EV	Days to flowering	0.00	0.000 ± 0.083	0.024 ± 0.089	0.004 ± 0.115
EV	Plant height (cm)	0.00	0.310 ± 0.090	0.506 ± 0.047	0.330 ± 0.078
EV	Iodine value	0.00	0.916 ± 0.031	0.916 ± 0.012	0.917 ± 0.013
EV	Lignin content (%)	0.00	0.465 ± 0.086	0.304 ± 0.048	0.352 ± 0.050
EV	Linoleic content (%)	0.00	0.928 ± 0.029	0.913 ± 0.013	0.912 ± 0.013
EV	Linolenic content (%)	0.00	0.936 ± 0.026	0.918 ± 0.012	0.918 ± 0.012
EV	Days to maturity	0.00	0.230 ± 0.105	0.226 ± 0.045	0.237 ± 0.048
EV	Oil content (%)	0.00	0.587 ± 0.072	0.612 ± 0.041	0.536 ± 0.055
EV	Oleic content (%)	0.00	0.408 ± 0.092	0.419 ± 0.046	0.409 ± 0.050

Table S1. Contd.

EV	Palmitic content (%)	0.00	0.758 ± 0.048	0.775 ± 0.029	0.754 ± 0.036
EV	Protein content (%)	0.00	0.664 ± 0.056	0.680 ± 0.040	0.573 ± 0.059
EV	Shive content (%)	0.00	0.486 ± 0.086	0.262 ± 0.043	0.301 ± 0.046
EV	Stearic content (%)	0.00	0.843 ± 0.042	0.756 ± 0.030	0.773 ± 0.026
EV	Straw weight (g)	0.00	0.198 ± 0.085	0.302 ± 0.086	0.171 ± 0.086
EV	Seed yield (T H^{-1})	0.00	0.414 ± 0.098	0.227 ± 0.041	0.172 ± 0.053
SU	Plant height (cm)	0.00	0.278 ± 0.101	0.245 ± 0.055	0.245 ± 0.055
SU	Iodine value	0.00	0.934 ± 0.033	0.921 ± 0.013	0.921 ± 0.013
SU	Linoleic content (%)	0.00	0.939 ± 0.031	0.935 ± 0.011	0.935 ± 0.011
SU	Linolenic content (%)	0.00	0.939 ± 0.033	0.932 ± 0.012	0.932 ± 0.012
SU	Days to maturity	0.00	0.547 ± 0.078	0.537 ± 0.054	0.537 ± 0.054
SU	Oil content (%)	0.00	0.738 ± 0.049	0.685 ± 0.042	0.685 ± 0.042
SU	Oleic content (%)	0.00	0.752 ± 0.052	0.724 ± 0.039	0.724 ± 0.039
SU	Palmitic content (%)	0.00	0.902 ± 0.020	0.878 ± 0.020	0.878 ± 0.020
SU	Protein content (%)	0.00	0.481 ± 0.089	0.450 ± 0.054	0.450 ± 0.054
SU	Stearic content (%)	0.00	0.888 ± 0.031	0.854 ± 0.023	0.854 ± 0.023
SU	Seed yield (T H^{-1})	0.00	0.214 ± 0.105	0.136 ± 0.044	0.136 ± 0.044

MDR, Missing data rate; BM, CDC Bethune/Macbeth; EV, E1747/Viking; SU: SP2047/UGG5-5; CC, core collection. The population sizes of BM, EV, SU, and CC are 243, 86, 70, and 391, respectively. $\widehat{H}^2$(A): $\widehat{H}^2$ on a plot basis estimated by ANOVA; $\widehat{H}^2$(R): $\widehat{H}^2$on a plot basis estimated by REML. The standard error (SE) for $r_E$was estimated using the bootstrap method and the SE for $\widehat{H}^2$(A) and $\widehat{H}^2$(R) was calculated based on the Delta method implemented in You et al. (2016b).