The impact of linkage disequilibrium and epistasis in the studies of inheritance based on Hayman's diallel and generation mean analysis

José Marcelo Soriano Viana (jmsviana@ufv.br)
Universidade Federal de Viçosa https://orcid.org/0000-0002-5063-4648

Research Article

Keywords: linkage disequilibrium, epistasis, diallel, generation mean analysis

Posted Date: November 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1025104/v1

Version of Record: A version of this preprint was published at Euphytica on April 25th, 2022. See the published version at https://doi.org/10.1007/s10681-022-03010-2.

The impact of linkage disequilibrium and epistasis in the studies of inheritance based on Hayman's diallel and generation mean analysis

José Marcelo Soriano Viana

Federal University of Viçosa, Department of General Biology, 36570-900, Viçosa, MG, Brazil.

E-mail: jmsviana@ufv.br. ORCID: https://orcid.org/0000-0002-5063-4648.

Abstract This simulation-based study assessed the impact of linkage disequilibrium (LD) and epistasis on Hayman's diallel and generation mean analysis, assuming hundreds of genes, variable degree of dominance, and seven types of digenic epistasis. The diallel parents were 15 doubled-haploid lines from a high LD population. The generation mean analysis was based on seven generations, assuming association. Under low LD and no epistasis, the diallel analysis provided confident results about the inheritance of the quantitative trait and high correlation between number of recessive genes and $W_r + V_r$, but biased estimates of the dominance components and genetic parameters. The additional consequences of high LD under no epistasis were rejection of the additive-dominance model assuming high heritability and lower correlation. Assuming 100% of epistatic genes, for four epistasis types there was evidence of inadequacy of the additive-dominance model. Assuming 30% of epistatic genes, there was a tendency for accepting the additive-dominance model for low heritability traits and for rejecting for high heritability traits. Linkage and epistasis affects the estimates of the genetic components of the generation means. Even assuming 100% of interacting genes, for most epistasis types there was no statistical evidence of epistasis. Assuming positive partial dominance, the signs of the epistatic components do not allow discriminate complementary, recessive, dominant and recessive, duplicate genes with cumulative effects, and non-epistatic genic interaction. Negative epistatic components evidence dominant epistasis. When the additive x additive and dominance x dominance components are positive and the additive x dominance component is negative, there is duplicate epistasis.

Keywords linkage disequilibrium, epistasis, diallel, generation mean analysis.

Acknowledgements I thank the National Council for Scientific and Technological Development (CNPq), the Brazilian Federal Agency for Support and Evaluation of Graduate Education (Capes; Finance Code 001), and the Foundation for Research Support of Minas Gerais State (Fapemig) for financial support.

Author Contribution Statement The author planned the study, developed the theory, programmed the software, processed the data, and wrote the manuscript.

Statements and Declarations The author has no relevant financial or non-financial interests to disclose. The author has no competing interests to declare that are relevant to the content of this article. The author certifies that he has no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The author has no financial or proprietary interests in any material discussed in this article.

Introduction

The genetic design diallel is regularly used in plant breeding. It is commonly employed to provide interpopulation crosses or single crosses/F_1's from inbred/pure/doubled haploid (DH) lines, for assessing heterosis, combining ability, or molecular genetic diversity, for predicting non-assessed testcrosses and single-crosses, among other applications (Mowers et al. 2018; Kadam et al. 2016; Yu et al. 2020; Leng et al. 2019). There are several methods of analysis but most investigations are based on the models (fixed or random) and methods proposed by Griffing (1956a). The main reasons that explain the general choice by breeders for the Griffing's combining ability analysis are: the methodology can be used for any crop and trait and the computation and interpretation of the genetic parameters are simple. Concerning the study of inheritance of quantitative traits, both heterosis and combining ability analyses allows testing non-additive effects, if there is genetic variability between the parents. But it is not possible to test for epistasis.

In the context of inheritance, an interesting approach based on diallel cross was proposed by Hayman (1954). However, very few inheritance studies based on Hayman's method were published

in the last 10 years (de Lima et al. 2019; Makumbi et al. 2018; Shahadati-Moghaddam et al. 2017; Kalinina and Lyakh 2011). The Hayman's method has indeed some limitations, as the assumptions of no epistasis (independent action of non-allelic genes) and no linkage disequilibrium (LD) (genes independently distributed in the parents). But it has positive aspects as testing the adequacy of the additive-dominance model. Then, if there is epistasis, the breeder can ignore the trait or take into account the influence of epistasis on the analysis. Why, then, the Hayman's method has been almost ignored in the studies of inheritance of quantitative traits? For sure, this cannot be attributable to a regular evidence of epistasis since most of the empirical data show mainly additive genetic variation (Hill et al. 2008). In my opinion, the main reasons are: the method seems very complex for breeders in regard to computation and interpretation; the breeders guess that the Griffing's method provides the same inferences; and the method is restricted for diploids and homozygous parents. However, Hayman's approach is not very complex for computing and interpreting and it provides some information on inheritance that a combining ability analysis does not offer.

Another interesting approach for the study of inheritance of quantitative traits is the generation mean analysis, proposed by Mather and Jinks (1971). Similar to the heterosis and combining ability analysis, this biometrical genetics methodology is based on the estimation of linear components of means of populations derived from homozygous parents. This approach has been commonly used by breeders of self- and cross-pollinated crops since its proposition (Rai et al. 2020; Addy et al. 2020; Verma and Singh 2018). Its main advantages are: it is easily performed for crops with inbred/pure/DH lines; it is applicable to any trait but few studies involved grain yield (Mohammed et al. 2018); it allows for testing dominance and epistasis separately; and the computation and the interpretation are simple.

Assuming absence of epistasis, the linear components of means do not depend on the LD. However, the genotypic variance and its genetic components are affected by LD (Hill and Maki-Tanila 2015). Because joint modelling epistasis and LD is a challenger for quantitative and biometrical geneticists, in the most important theoretical papers on heterosis and combining ability

analysis there is only superficial information on the influence of non-allelic interaction (Gardner and Eberhart 1966; Kempthorne 1956; Griffing 1956b). Because his method is based on components of the genotypic variance for parents and F_1's, Hayman (1954) makes some statements on how LD and epistasis affects the diallel analysis (see sections 4.4. Correlated gene distributions and 4.5. Non-allelic gene interaction). The knowledge provided by Hayman (1954) was extended in the studies of Hill (1964), Nassar (1965), Mather (1967), and Coughtrey and Mather (1970). In regard to generation mean analysis, Mather and Jinks (1971) included theory and some general conclusions for the analysis assuming LD and epistasis (see section 5.18 Linkage of interacting genes). Hill (1964) and Nassar (1965), based on 3- and 10-gene models, provided contrasting results regarding the consequences of LD on the W_r/V_r graph and the order of dominance. Mather (1967) and Coughtrey and Mather (1970) assumed only complementary and duplicate epistasis in a two-gene model with no LD. Thus, none previous published study provided general information on the joint impact of LD and epistasis on the Hayman's diallel and generation mean analyses. Then, the objective of this study was to provide significant additional knowledge about the influence of LD and epistasis on the Hayman's diallel and generation mean analyses, based on simulated data. I assumed hundreds of genes, variable degree of dominance, variable gene frequencies, LD, and seven types of digenic epistasis. In the first part of this study I present the theoretical background that supports the software used for simulating the data set.

Material and Methods

Genetic variances and covariances of the Hayman's diallel assuming LD and epistasis

Consider n inbred/pure/DH lines (n > 3). Assume initially LD but no epistasis. In regard to two genes, the genotype probabilities are $P(AABB) = f_{22} = u_a u_b + \Delta_{ab}$, $P(AAbb) = f_{20} = u_a v_b - \Delta_{ab}$, $P(aaBB) = f_{02} = v_a u_b - \Delta_{ab}$, and $P(aabb) = f_{00} = v_a v_b + \Delta_{ab}$, where u and v are the allelic frequencies and Δ_{ab} is the measure of LD in the gametic pool. Note that I did not state that the parents are a sample from a reference population or that the two genes are linked. Using the same definition of Hayman (1954) for the genotypic values of the r-th parent and the F_1 between

parents r and s, the variance of the parents is $V_{0L0} = (1 - w_a^2)d_a^2 + (1 - w_b^2)d_b^2 + 8\Delta_{ab}d_a d_b = D$, where d is the deviation between the genotypic value of the homozygote of greater expression and the mean of the genotypic values of the homozygotes (m) and $w = u - v$. Because the means of the parents and the F_1 are not affected by the LD (in the absence of epistasis), the difference between the mean of the n^2 progeny (m_{L1}) and the mean of their parents (m_{L0}) is equal to the function derived by Hayman (1954): $(1/2)[(1 - w_a^2)h_a + (1 - w_b^2)h_b] = (1/2)h$, where h_i is the dominance deviation. The variance of the r-th array is $V_r = (1/4)D + (1/4)H_1 - (1/4)F_r - (1/4)F_{1r}$, where

$H_1 = (1 - w_a^2)h_a^2 + (1 - w_b^2)h_b^2 + 8\Delta_{ab}h_a h_b$,

$F_r = 2[(1 - w_a^2)d_a\theta_a h_a + (1 - w_b^2)d_b\theta_b h_b + 4\Delta_{ab}d_a\theta_b h_b + 4\Delta_{ab}d_b\theta_a h_a]$, where $\theta = 1$ if AA or BB and $\theta = -1$ if aa or bb, and

$F_{1r} = 16k\Delta_{ab}h_a h_b$, where $k = 0$ if AABB or aabb and $k = 1$ if AAbb or aaBB.

Thus, the average variance is $V_{1L1} = (1/4)D + (1/4)H_1 - (1/4)F - (1/4)F_1$, where $F = 2[w_a(1 - w_a^2)d_a h_a + w_b(1 - w_b^2)d_b h_b + 4\Delta_{ab}w_b d_a h_b + 4\Delta_{ab}w_a d_b h_a]$ and $F_1 = 16\Delta_{ab}(u_a v_b + u_b v_a - 2\Delta_{ab})h_a h_b$.

The covariance between the non-recurrent parents and their offspring in the r-th array is $W_r = (1/2)D - (1/4)F_r$. Thus, the average covariance is $W_{0L01} = (1/2)D - (1/4)F$. Then, if there is dominance and LD, the difference between the covariance and the variance in the arrays is not a constant value. The difference is $(1/4)(D - H_1) + k(1/4)F_{1r}$. This implies that the points (V_r, W_r) does not lie on a straight line of unit slope through their mean point (V_{1L1}, W_{0L01}). However, the function does not allow realizing how much LD affects the deviation from 1. The variance of the array means is $V_{0L1} = (1/4)D + (1/4)H_1 - (1/4)H_2 - (1/4)F$, where $H_2 = (1 - w_a^2)^2 h_a^2 + (1 - w_b^2)^2 h_b^2 + 8\Delta_{ab}(1 - w_a w_b)h_a h_b$. Finally, the total genotypic variance is $V_{0L1} + V_{1L1} = (1/2)D + (1/2)H_1 - (1/4)H_2 - (1/2)F - (1/4)F_1$. Note that, because LD, the four equations independent of r allows the estimation of the parameters D, $H_1 - F_1$, $H_2 - F_1$, and F. This implies that the genetic parameters that are dependent of H_1 and/or H_2 – average degree of dominance,

mean value of uv (symmetry) for dominant genes, proportion between dominant and recessive genes, and number of dominant genes – are biased. Because the functions for these parameters are complex, the magnitude of the bias can only be assed using simulated data. Further, it is not also clear how LD affects the order of dominance of the parents in the graph (W_r, V_r).

Assume now that the two genes are epistatic. Assuming an epistatic effect (I_{ij}; i, j = 2, 1, and 0) for each genotype, the variance of the parents is $V^*_{0L0} = V_{0L0} + V(I)_0 + 2Cov_0$, where $V(I)_0 = f_{22}I^2_{22} + f_{20}I^2_{20} + f_{02}I^2_{02} + f_{00}I^2_{00} - [E(I)_0]^2$ is the epistatic variance of the parents and $Cov_0 = f_{22}(d_a + d_b)I_{22} + f_{20}(d_a - d_b)I_{20} + f_{02}(-d_a + d_b)I_{02} + f_{00}(-d_a - d_b)I_{00} - (m_{L0} - m_a - m_b)E(I)_0$ is the covariance between the sum of additive deviations and the epistatic effect for the parents. The diallel mean is $m^*_{L1} = m_{L1} + E(I)_1$, where $E(I)_1 = f^2_{22}I_{22} + 2f_{22}f_{20}I_{21} + \cdots + f^2_{00}I_{00}$. Because $m^*_{L0} = m_{L0} + E(I)_0$, $m^*_{L1} - m^*_{L0} = (1/2)h + E(I)_1 - E(I)_0$. The variance in the r-th array is $V^*_r = V_r + V(I)_r + 2Cov_r$, where, for example, for the array of the parent AABB, $V(I)_{22} = f_{22}I^2_{22} + f_{20}I^2_{21} + f_{02}I^2_{12} + f_{00}I^2_{11} - [E(I)_{22}]^2$ is the epistatic variance in the array and $Cov_{22} = f_{22}(d_a + d_b)I_{22} + f_{20}(d_a + h_b)I_{21} + f_{02}(h_a + d_b)I_{12} + f_{00}(h_a + h_b)I_{11} - (E(G)_{22} - m_a - m_b)E(I)_{22}$ is the covariance between the non-epistatic deviation and the epistatic effect in the array. Thus, $V^*_{1L1} = V_{1L1} + V(I)_1 - V(\bar{I})_1 + 2Cov^1_1 - 2Cov^2_1$, where $V(I)_1 = f^2_{22}I^2_{22} + 2f_{22}f_{20}I^2_{21} + \cdots + f^2_{00}I^2_{00} - [E(I)_1]^2$ is the F_1 epistatic variance, $V(\bar{I})_1 = f_{22}[E(I)_{22}]^2 + f_{20}[E(I)_{20}]^2 + f_{02}[E(I)_{02}]^2 + f_{00}[E(I)_{00}]^2 - [E(I)_1]^2$ is the variance of the average epistatic values of the array means, $Cov^1_1 = f^2_{22}(d_a + d_b)I_{22} + 2f_{22}f_{20}(d_a + h_b)I_{20} + \cdots + f^2_{00}(-d_a - d_b)I_{00} - (m_{L1} - m_a - m_b)E(I)_1$ is the covariance between the non-epistatic deviation and the epistatic effect in the F_1, and $Cov^2_1 = f_{22}(E(G)_{22} - m_a - m_b)E(I)_{22} + f_{20}(E(G)_{20} - m_a - m_b)E(I)_{20} + f_{02}(E(G)_{02} - m_a - m_b)E(I)_{02} + f_{00}(E(G)_{00} - m_a - m_b)E(I)_{00} - (m_{L1} - m_a - m_b)E(I)_1$ is the covariance between the average non-epistatic and the epistatic values of the array means.

The covariance in the r-th array is $W^*_r = W_r + Cov^1_{01(r)} + Cov^2_{01(r)} + Cov^3_{01(r)}$, where, for example, for the array of the parent AABB,

$Cov^1_{01(22)} = f_{22}(d_a + d_b)I_{22} + f_{20}(d_a - d_b)I_{21} + f_{02}(-d_a + d_b)I_{12} + f_{00}(-d_a - d_b)I_{11} -$ $(m_{L0} - m_a - m_b)E(I)_{22}$,

$Cov^2_{01(22)} = f_{22}I_{22}(d_a + d_b) + f_{20}I_{20}(d_a + h_b) + f_{02}I_{02}(h_a + d_b) + f_{00}I_{00}(h_a + h_b) -$ $E(I)_0(E(G)_{22} - m_a - m_b)$, and

$Cov^3_{01(22)} = f_{22}I^2_{22} + f_{20}I_{20}I_{21} + f_{02}I_{02}I_{12} + f_{00}I_{00}I_{11} - E(I)_0E(I)_{22}$

are the covariances between the non-epistatic deviation of non-recurrent parent and the epistatic effect of F_1, between the non-epistatic deviation of F_1 and the epistatic effect of non-recurrent parent, and between the epistatic effects of non-recurrent parent and F_1, respectively. The average covariance is $W^*_{0L01} = W_{0L01} + Cov^1_{01} + Cov^2_{01} + Cov^3_{01}$, where

$Cov^1_{01} = f_{22}(d_a + d_b)E(I)_{22} + f_{20}(d_a - d_b)E(I)_{20} + f_{02}(-d_a + d_b)E(I)_{02} + f_{00}(-d_a -$ $d_b)E(I)_{00} - (m_{L0} - m_a - m_b)E(I)_1$,

$Cov^2_{01} = f_{22}I_{22}(E(G)_{22} - m_a - m_b) + f_{20}I_{20}(E(G)_{20} - m_a - m_b) + f_{02}I_{02}(E(G)_{02} - m_a -$ $m_b) + f_{00}I_{00}(E(G)_{00} - m_a - m_b) - E(I)_0(m_{L1} - m_a - m_b)$, and

$Cov^3_{01} = f_{22}I_{22}E(I)_{22} + f_{20}I_{20}E(I)_{20} + f_{02}I_{02}E(I)_{02} + f_{00}I_{00}E(I)_{00} - E(I)_0E(I)_1$.

Note that the difference between the covariance and the variance in the arrays is $W^*_r - V^*_r =$ $(1/4)(D - H_1) + k(1/4)F_{1r} + Cov^1_{01(r)} + Cov^2_{01(r)} + Cov^3_{01(r)} - V(I)_r - 2Cov_r$. Thus, the epistasis is an additional factor that deviate the points (W^*_r, V^*_r) from a straight line of unit slope through the mean point (V_{1L1}, W_{0L01}). However, the function does not also allow realizing how much epistasis affects the deviation from 1. The variance of the array means is $V^*_{0L1} = V_{0L1} +$ $V(\bar{I})_1 + 2Cov^2_1$. Finally, the total genotypic variance is $V^*_{0L1} + V^*_{1L1} = V_{0L1} + V_{1L1} + V(I)_1 +$ $2Cov^1_1$. Thus, epistasis introduces an additional bias in the estimates of the Hayman's genetic parameters, which can only be assessed using simulated data.

Generation mean analysis with LD and epistasis

Assuming two linked epistatic genes, an epistatic effect for each genotype, and parents AABB and aabb (association), the genotypic values of the parents and the F_1 are $G_{22} = m_a + m_b + d_a +$

$d_b + I_{22}$, $G_{00} = m_a + m_b - d_a - d_b + I_{00}$, and $G_{11} = m_a + m_b + h_a + h_b + I_{11}$. Using the notation of Mather and Jinks (1971) for the additive and dominance components, $P_1 = m + [d] + I_{22}$, $P_2 = m - [d] + I_{22}$, and $F_1 = m + [h] + I_{11}$. Because epistasis, m is not the F_∞ mean, given by $F_\infty = m + [1/2(1+2r)](I_{22} + I_{00}) + [1/(1+2r)](I_{20} + I_{02}) = m + E(I)^{(\infty)}$, where r is the recombination frequency. Note that in F_2, linked genes with r lower than 0.5 are in LD. The absolute LD value in the gametic pool of F_1 is $(1-2r)/4$. The value is positive with coupling and negative with repulsion. The mean of the F_2 generation is $F_2 = m + (1/2)[h] + E(I)^{(0)}$, where $E(I)^{(0)} = [(1-r)/2]^2 I_{22} + 2[(1-r)/2](r/2)I_{21} + \cdots + [(1-r)/2]^2 I_{00}$ is the expectation of the epistatic values. The F_{n+2} mean is $F_{n+2} = m + (1/2)^{n+1}[h] + E(I)^{(n)}$, where n is the number of selfing generations. It is interestingly to note that the expectation of the epistatic values in a generation is not directly proportional to the expectation in the F_2 generation, since $E(I)^{(n)} = E(I)^{(0)} + deviation$ (see the deviation for the F_{n+2} generation in the Appendix).

The average genotypic values of the two backcross generations are $BC_1 = m + (1/2)[d] + (1/2)[h] + (1/2)E(I)_1$ and $BC_2 = m - (1/2)[d] + (1/2)[h] + (1/2)E(I)_2$, where $E(I)_1 = (1-r)(I_{22} + I_{11}) + r(I_{21} + I_{12})$ and $E(I)_2 = (1-r)(I_{11} + I_{00}) + r(I_{10} + I_{01})$. Thus, assuming digenic epistasis, the means of the parents, F_1, F_2, F_3, and backcrosses (seven equations) depend on 10 genetic linear components (seven epistatic components). A very known approach for allowing testing epistasis and estimating and testing epistatic components was provided by Mather and Jinks (1971). This simplified approach has been used for modelling epistasis in genomic selection, GWAS (genome-wide association studies), and QTL mapping. Mather and Jinks (1971) assumed $I_{22} = I_{00} = -I_{20} = -I_{02} = [i]$, $I_{21} = -I_{01} = [j]$, $I_{12} = -I_{10} = [j']$, and $I_{11} = [l]$. However, because linkage (LD), $E(I)^{(0)} = [(1-2r)/2][i] + (1/4)[1 + (1-2r)^2][l]$, $E(I)_1 = (1-r)([i] + [l]) + r([j] + [j'])$, and $E(I)_2 = (1-r)([i] + [l]) - r([j] + [j'])$. Note that I wrote simplified approach because the assumptions for the additive x additive, additive x dominance, and dominance x additive components do not necessarily met for any type of digenic epistasis, including complementary, duplicate, dominant, recessive, dominant and recessive, duplicate genes

with cumulative effects, and non-epistatic genic interaction, regardless of the degree of dominance. To characterize complementary epistasis it is necessary to assume $d_a = d_b = h_a = h_b = i_{ab} = j_{ab} = j_{ba} = l_{ab}$; for duplicate epistasis, $d_a = d_b = h_a = h_b = -i_{ab} = -j_{ab} = -j_{ba} = -l_{ab}$; assuming recessive epistasis implies $d_a = h_a = i_{ab} = j_{ba}$ and $d_b = h_b = j_{ab} = l_{ab}$ or $d_a = -h_a = -i_{ab} = j_{ba}$ and $d_b = -h_b = j_{ab} = -l_{ab}$; in case of dominant epistasis, $d_a = h_a = -i_{ab} = -j_{ba}$ and $d_b = h_b = -j_{ab} = -l_{ab}$ or $d_a = -h_a = i_{ab} = -j_{ba}$ and $d_b = -h_b = -j_{ab} = l_{ab}$ (Mather 1967). Then, it is assumed complete dominance (|h/d| = 1).

Data simulation

The simulated data set was generated using the software *REALbreeding* (available by request). *REALbreeding* has been used in studies related to genomic selection (Viana et al. 2018), GWAS (Pereira et al. 2018), QTL mapping (Viana et al. 2017), LD (Andrade et al. 2019), population structure (Viana et al. 2013b), heterotic grouping/genetic diversity (Viana et al. 2020), and plant breeding (Viana et al. 2013a). In summary, the software simulates individual genotypes for genes and molecular markers and phenotypes in three phases using inputs from the user. The first phase (genome simulation) is the specification of the number of chromosomes, molecular markers, and genes as well as marker type (dominant and/or codominant; bi- or multi-allelic) and density. The second phase (population simulation) is the specification of the population(s) and sample size or progeny number and size. A population is characterized by the average frequency for the genes (biallelic) and markers (first allele). The last phase (trait simulation) is the specification of the minimum and maximum genotypic values for homozygotes, the minimum and maximum phenotypic values (to avoid outliers), the direction and degree of dominance, and the broad sense heritability. The current version allows the inclusion of digenic epistasis, gene x environment interaction, and multiple traits, including pleiotropy. The population mean (M), additive (A), dominance (D), and epistatic (additive x additive (AA), additive x dominance (AD), dominance x additive (DA), and dominance x dominance (DD)) genetic values or GCA and SCA effects or genotypic values (G) and epistatic values (I), depending on the population, are calculated from the

parametric gene effects and frequencies and the parametric LD values. The population in LD is generated by crossing two populations in linkage equilibrium followed by a generation of random cross. The parametric LD is $\Delta_{ab}^{(-1)} = [(1 - 2r_{ab})/4](p_{a1} - p_{a2})(p_{b1} - p_{b2})$, where the indexes 1 and 2 stand for the gene frequencies in the parental populations. The phenotypic values (P) are computed assuming error effects (E) sampled from a normal distribution ($P = M + A + D + AA + AD + DA + DD + E = G + E$ or $P = M + GCA1 + GCA2 + SCA + I + E = G + E$).

The types of digenic epistasis are: complementary ($G_{22} = G_{21} = G_{12} = G_{11}$ and $G_{20} = G_{10} = G_{02} = G_{01} = G_{00}$; proportion of 9:7 in a F_2), duplicate ($G_{22} = G_{21} = G_{20} = G_{12} = G_{11} = G_{10} = G_{02} = G_{01}$; proportion of 15:1 in a F_2), dominant ($G_{22} = G_{21} = G_{20} = G_{12} = G_{11} = G_{10}$ and $G_{02} = G_{01}$; proportion of 12:3:1 in a F_2), recessive ($G_{22} = G_{21} = G_{12} = G_{11}$, $G_{02} = G_{01}$, and $G_{20} = G_{10} = G_{00}$; proportion of 9:3:4 in a F_2), dominant and recessive ($G_{22} = G_{21} = G_{12} = G_{11} = G_{20} = G_{10} = G_{00}$ and $G_{02} = G_{01}$; proportion of 13:3 in a F_2), duplicate genes with cumulative effects ($G_{22} = G_{21} = G_{12} = G_{11}$, and $G_{20} = G_{10} = G_{02} = G_{01}$; proportion of 9:6:1 in a F_2), and non-epistatic genic interaction ($G_{22} = G_{21} = G_{12} = G_{11}$, $G_{20} = G_{10}$, and $G_{02} = G_{01}$; proportion of 9:3:3:1 in a F_2). Because the genotypic values for any two interacting genes are not known, there are infinite genotypic values that satisfy the specifications of each type of digenic epistasis. For example, fixing the gene frequencies (the population) and the parameters m, d, h, and h/d (degree of dominance) for each gene (the trait), the solutions $G_{22} = G_{21} = G_{12} = G_{11} = 5.25$ and $G_{20} = G_{10} = G_{02} = G_{01} = G_{00} = 5.71$ or $G_{22} = G_{21} = G_{12} = G_{11} = 6.75$ and $G_{20} = G_{10} = G_{02} = G_{01} = G_{00} = 2.71$ define complementary epistasis but the genotypic values are not the same. The solution implemented in the software allows the user to control the magnitude of the epistatic variance (V(I)), relative to the magnitudes of the additive and dominance variances (V(A) and V(D)). As an input for the user, the software requires the ratio V(I)/(V(A) + V(D)) for each pair of interacting genes (a single value; for example, 1.0). Then, for each pair of interacting genes the software samples a random value for the epistatic value I_{22} (the epistatic value for the genotype AABB), assuming $I_{22} \sim N(0, V(I))$. Then, the other epistatic effects and genotypic values are computed. In this study, I assumed ratio 1.

Increasing the ratio increases the magnitude of the additive, dominance, and epistatic genetic values.

For the diallel cross I generated 1,000 DH lines from a population with high LD and selected 15 DHs. The criterion for selecting the parents was two to three DHs at random from six classes for the number of dominant genes: 0 to 140 up to 261 to 290. The number of dominant genes in the selected DHs ranged from 161 to 246. Then, the selected DHs were crossed in a complete diallel without reciprocals. For the generation mean analysis I generated the contrasting parental inbred lines (P_1 and P_2), assuming all genes in association, and the generations F_1, F_2, F_3, BC_1, and BC_2. The numbers of plants per generation were 50, 50, 50, 400, 400, 400, and 400, respectively. The number of genes was 400, distributed in 10 chromosomes of 200 cM. I simulated grain yield (g/plant), assuming positive dominance (average degree of dominance of 0.6), and expansion volume (a measure of popcorn quality; ml/g), assuming bidirectional dominance (average degree of dominance of 0.0). For grain yield, the minimum and maximum genotypic values for homozygotes were 30 and 160. The minimum and maximum phenotypic values for homozygotes were 10 and 180. For expansion volume I assumed 5 and 55 as the minimum and maximum genotypic values for homozygotes. The minimum and maximum phenotypic values were 0 and 60. I assumed no epistasis (but LD), seven types of digenic epistasis and an admixture of these types, defining 30 and 100% of epistatic genes. I also generated a scenario of low LD and no epistasis, assuming 10 independent genes and 10 DHs sampled from 1,000 DHs. The criterion for selecting the 10 DHs was minimization of the LD. The LD values ($\Delta_{ab} = p_{AB} - p_A p_B$) ranged from –0.11 to 0.09 (absolute average value 0.04). For comparison purpose, I computed the LD values for the 80 genes in the chromosomes 1 and 2. The ranges and the means of the absolute values for the genes in chromosomes 1, 2 and for the independent genes were, respectively, –0.15 to 0.25 and 0.07, –0.16 to 0.21 and 0.05, and –0.20 to 0.20 and 0.05. The broad sense heritability at the plant level was 20%. For the progeny level were 40, 60, and 80%. To avoid the influence of the experimental error,

all analyses were based on the parametric genotypic values and variances and covariances, provided by *REALbreeding*.

Results

The impact of LD on the Hayman's diallel analysis is evident even in the scenario of independent genes. The analysis of variance of $W_r - V_r$ indicated adequacy of the additive-dominance model (P value of 1.00) even assuming a heritability of 80%. The regression of W_r on V_r indicated partial dominance ($\beta_0 = 0.02$; P value of 0.00) and β_1 equal to 1 (P value of 0.38). The coefficient of determination was 0.99. The conclusion of partial dominance is correct since the average degree of dominance is 0.52 (in the range 0.01 to 1.16). However, the LD between the independent genes led to a bias of −22.5 and −27.3% in the estimates of the dominance components and bias in the range −13.8 to −47.7% for three of the genetic parameters, ignoring the estimate of the number of dominant genes that is always sub estimated (4 versus 9, because the DHs have one fixed gene). The correlation between the number of recessive genes and $W_r + V_r$ was 0.92 and the number of dominant genes for the parents' order of dominance was 6, 6, 6, 6, 5, 5, 4, 5, 4, and 2. These results indicates that when there is low LD and no epistasis, the Hayman's diallel analysis provides confident results about the inheritance of the quantitative trait but biased estimates of the dominance components and the genetic parameters. The negative consequences of high LD under no epistasis are also biased estimates of the dominance components and genetic parameters plus a biased estimate of the covariance F, a significant test for the homogeneity of $W_r - V_r$ assuming a heritability of 80%, and an intermediate value for the correlation between the number of recessive genes and $W_r + V_r$ (Table 1). Thus, high LD can lead to inadequacy of the additive-dominance model for a high heritability trait. The number of dominant genes for the parents' order of dominance was 245, 176, 246, 195, 230, 229, 203, 189, 191, 217, 166, 176, 214, 203, and 161. Regardless of the heritability trait, there was evidence of partial dominance and a slight deviation from 1.0 for the regression coefficient.

As demonstrated from the theory, epistasis is a significant additional factor negatively affecting the Hayman's diallel analysis, especially assuming a high proportion of interacting genes. Assuming 100% of epistatic genes and dominant, recessive, duplicate genes with cumulative effects, and non-epistatic genic interaction, there was evidence of inadequacy of the additive-dominance model, regardless of the trait heritability (Table 1). Irrespective of the type of epistasis and significance of the test of adequacy of the additive-dominance model, there was a tendency for concluding in favor of overdominance, which is a wrong inference. Further, the correlation between the number of recessive genes and $W_r + V_r$ tends to be lower (in the range −0.01 to 0.67; 0.39 on average) and all genetic components and parameters will be very biased. Assuming 30% of epistatic genes, there was a tendency for accepting the additive-dominance model for low heritability traits and for rejecting the model for high heritability traits (Table 1 and Online Resource Table 1). Regardless of the significance level of the adherence test, there was evidence of partial dominance and the correlation between the number of recessive genes and $W_r + V_r$ had intermediate magnitude (in the range 0.31 to 0.62; 0.50 on average). However, as demonstrated from the theory, the genetic components and parameters will be very biased. Note that, in general, the coefficient of regression has a magnitude close to 1.0, but only assuming an admixture of epistasis it is consistently not statistically different from 1.0. In regard to the heterosis, that is unaffected by LD, note that the epistasis tends to decrease it, except assuming duplicate genes with cumulative effects and non-epistatic genic interaction (Table 1 and Online Resource Table 1).

In the absence of epistasis, as theoretically demonstrated, linkage (LD) does not affect the generation mean analysis, respecting that this method cannot distinguish absence of dominance from symmetrical bidirectional dominance. Under the sample sizes and heritability (a low value, 20%) assumed, the experimental evidence is that there is additive variability, dominance, and no epistasis, for both traits (Table 2). Additionally, there was evidence of (predominantly) positive dominance for grain yield and (predominantly) negative dominance for expansion volume. Note also that, regardless of the trait, heritability, and genetic control, if the sample sizes for parents and

derived generations are representative, the analysis will provide confident estimates of the parametric means (correlations close to 1.0 between the estimated and true means). However, linkage and epistasis significantly affects the estimates of the genetic components of the generation means, regardless of the type of epistasis, proportional the percentage of epistatic genes (Table 2). One impressive result by fitting the epistatic components ([i], [j], and [l]) is that, even assuming 100% of interacting genes, for most epistasis types there were evidence of no epistasis (P values in the range 0.052 to 0.15). Another remarkable result is a null or negative correlation between the epistatic components of my model and of the Mather and Jinks' model, for most of the epistasis types. Note that the two higher correlations (among four) are associated with the less biased estimates of the non-epistatic components, under duplicate genes with cumulative effects and non-epistatic genic interaction. Fitting the Mather and Jinks' complete model when there is statistical evidence of epistasis or fitting the additive-dominance model when there is epistasis but no statistical evidence of epistasis provides biased estimates of the genetic components, proportional to the percentage of epistatic genes (Table 2 and Online Resource Table 2). Under duplicate and dominant and recessive epistasis, a wrong inference about dominance occurred (negative dominance and absence of dominance, respectively, instead of positive dominance).

Discussion

Probably because Hill (1964), Nassar (1965), Mather (1967), and Coughtrey and Mather (1970) used simplified genetic models (2 to 10 genes, complete dominance, only LD or epistasis) for investigating the effects of LD or epistasis on Hayman's diallel analysis, their main findings are very limited and only partially confirmed in this study. Hill (1964) concluded that LD led to a significant upward curvature on the W_r/V_r graph and alters the level of dominance. My results do not support his conclusions. I observed a downward curvature and a change in the level of dominance only under a high proportion of epistatic genes. Nassar (1965) observed that LD determines serious deviations in the slope of the W_r/V_r regression line (consistently lower than one) and an intercept predominantly below the origin (indicating overdominance). I generally observed a

slope with magnitude close to 1 but statistically lower than one in 50% of the scenarios, but there was evidence of overdominance only under a high proportion of epistatic genes. I also did not observed a W_r/V_r graph concave upwards with complementary epistasis, regardless of the percentage of interacting genes, as emphasized by Mather (1967) and Coughtrey and Mather (1970).

Concerning the impact of LD and epistasis on the Hayman's diallel analysis, the results describe an apparently very negative effect: biased estimates of the non-additive components and, consequently, biased estimates of the genetic parameters. But in my opinion, average degree of dominance, symmetry, proportion of dominant to recessive genes and, especially, number of dominant genes has no current significance for breeders. Assuming no epistasis or 30% of interacting genes the results show a correct inference on the inheritance of a low heritability quantitative trait: partial positive dominance. The decrease in the correlation between the number of dominant genes and $W_r + V_r$ is not so serious to make the Hayman's diallel analysis useless. However, breeders should be conscious to process and interpret only the traits that show adequacy of the additive-dominance model (homogeneity of the difference $W_r - V_r$), functional relationship between W_r and V_r (slope of the regression statistically different from zero), and especially a reasonable coefficient of determination of the regression analysis (say, greater than 75%). Unfortunately, this was not observed in some recently published studies (Pessoa et al. 2019; de Lima et al. 2019; Shahadati-Moghaddam et al. 2017). Thus, breeders must realize that LD and epistasis are not their major problem when processing a Hayman's diallel analysis, but the experimental error.

As theoretically demonstrated, only in a very particular case the Mather and Jinks' model will provide unbiased estimates of the genetic components. This will occur only if $I_{22} = I_{00}$, $I_{21} + I_{12} = -I_{01} - I_{10}$, and $E(I)^{(n)} = (1/4)^{n+1} I_{11}$. But, because epistasis implies in more genetic components than generations, the only option is to fit the Mather and Jinks' model. In my opinion, biased estimates of the genetic components are not a serious problem if the inferences on the inheritance of

the quantitative trait are correct. In general, even assuming 100% of epistatic genes, the analyses correctly evidenced positive dominance for grain yield but, unfortunately, in 5 out of 10 scenarios the epistasis was not detected. Thus, a more serious problem of the generation mean analysis is not allowing the detection of epistasis even when the percentage of interacting genes is high. When there is evidence of epistasis, another serious problem is the correct attribution of the predominant type. Regardless of the percentage of interacting genes, the signs of the seven epistatic components do not separate complementary, recessive, and dominant and recessive types, as well as duplicate genes with cumulative effects and non-epistatic genic interaction. Assuming positive partial dominance and 100% of interacting genes, the signs of the epistatic components [i], [j], and [l] do not allow discriminate complementary, recessive, dominant and recessive, duplicate genes with cumulative effects, and non-epistatic genic interaction (all positive). When the three components are negative, there is predominantly dominant epistasis. If the additive x additive and dominance x dominance components are positive and the additive x dominance component is negative, there is duplicate epistasis, as emphasized by Mather and Jinks (1971).

Recent published generation mean analyses based on field data, involving diverse crops and traits, six to seven generations, one to six crosses, one or more seasons or environments, evidenced epistasis in many cases but not all epistatic components were statistically different from zero (Shirinpour et al. 2020; Rai et al. 2020; Pal et al. 2020; Verma and Singh 2019, 2018; Mohammed et al. 2018). This is not a problem since the non-additive components are a sum of effects that take sign. In few studies, the authors concluded in favor of complementary or duplicate epistasis but assuming diverse combinations of signs for the epistatic components. This is not in agreement with Mather and Jinks (1971). In two investigations also involving a qualitative analysis, the two analyses provided inconsistent results. For example, the qualitative analysis showed dominant and recessive epistasis and the generation mean analysis evidenced complementary epistasis.

Concluding, taking into account the theory presented and the results from the simulation study, LD and epistasis can have negative effects on the Hayman's diallel analysis and on the

Mather and Jinks' generation mean analysis, proportional to the LD level and the percentage of epistatic genes, and depending on the predominant type of epistasis. However, biased estimates of quadratic or linear genetic parameters are not so serious if the inheritance of the quantitative trait is correctly inferred, at least partially. Note that, excepting for a high proportion of epistatic genes under high LD, the general correct conclusion was partial positive dominance for grain yield from both analysis. Further, the order of dominance provides a good discrimination between the parents, regardless of the type of epistasis, percentage of interacting genes, and level of LD. Unfortunately, the detection of epistasis from both analyses is highly affected by the trait heritability, predominant type of epistasis, and percentage of interacting genes.

Data Availability Statement: The dataset is available at https://doi.org/10.6084/m9.figshare.16732888.v1.

Supplementary Material Two additional tables.

Appendix

The expectation of the epistatic effects in the F_{n+2} generation is $E(I)^{(n)} = E(I)^{(0)} + p_{22}^{(n)}I_{22} + p_{21}^{(n)}I_{21} + p_{20}^{(n)}I_{20} + p_{12}^{(n)}I_{12} + p_{11}^{(n)}I_{11} + p_{10}^{(n)}I_{10} + p_{02}^{(n)}I_{02} + p_{01}^{(n)}I_{01} + p_{00}^{(n)}I_{00}$, where n is the number of selfing generations and

$$p_{22}^{(n)} = (F/2)(f_{21}^{(0)} + f_{12}^{(0)}) + P_1^{(n)}$$

$$p_{21}^{(n)} = (1-F)\left[f_{21}^{(0)} + (1-c^n)f_{11}^{(0)}/2\right] - f_{21}^{(0)}$$

$$p_{20}^{(n)} = (F/2)(f_{21}^{(0)} + f_{10}^{(0)}) + P_2^{(n)}$$

$$p_{12}^{(n)} = (1-F)\left[f_{12}^{(0)} + (1-c^n)f_{11}^{(0)}/2\right] - f_{12}^{(0)}$$

$$p_{11}^{(n)} = [(1-F)c^n - 1]f_{11}^{(0)}$$

$$p_{10}^{(n)} = (1-F)\left[f_{10}^{(0)} + (1-c^n)f_{11}^{(0)}/2\right] - f_{10}^{(0)}$$

$$p_{02}^{(n)} = (F/2)(f_{01}^{(0)} + f_{12}^{(0)}) + P_2^{(n)}$$

$$p_{01}^{(n)} = (1-F)\left[f_{01}^{(0)} + (1-c^n)f_{11}^{(0)}/2\right] - f_{01}^{(0)}$$

$p_{00}^{(n)} = (F/2)(f_{01}^{(0)} + f_{10}^{(0)}) + P_1^{(n)}$

where $F = 1 - (1/2)^n$ is the inbreeding coefficient and $f_{ij}^{(0)}$ is a genotype probability in F_2 (i and j = 2, 1, or 0). Defining $P_{11}^{(0)}$, $P_{10}^{(0)}$, $P_{01}^{(0)}$, and $P_{00}^{(0)}$ as the gamete probabilities of the F_1, where, for example, $P_{11}^{(0)} = P_{00}^{(0)} = (1-r)/2$ and $P_{10}^{(0)} = P_{01}^{(0)} = r/2$ for coupling, the probability of the genotype AABb in F_2 is $f_{21}^{(0)} = 2P_{11}^{(0)}P_{10}^{(0)}$. Further, $c = 1 - 2r(1-r)$, $P_1^{(n)} = (1/4)\{[F - (1 - F)(1 - c^n)]f_{11}^{(0)} + c_1^{(n)}(1-2r)\Delta\}$, $P_2^{(n)} = (1/4)\{[F - (1-F)(1-c^n)]f_{11}^{(0)} - c_1^{(n)}(1-2r)\Delta\}$, $c_1^{(n)} = \{2[1 - ((1-2r)/2)^n]/(1+2r)\}$, and $|\Delta| = (1-2r)/4$ (positive for coupling and negative for repulsion).

References

Addy S, Cichy KA, Adu-Dapaah H, Asante IK, Emmanuel A, Offei SK (2020) Genetic Studies on the Inheritance of Storage-Induced Cooking Time in Cowpeas Vigna unguiculata (L.) Walp. Frontiers in Plant Science 11. doi:10.3389/fpls.2020.00444

Andrade ACB, Viana JMS, Pereira HD, Pinto VB, Fonseca ESF (2019) Linkage disequilibrium and haplotype block patterns in popcorn populations. PloS one 14 (9):e0219417. doi:10.1371/journal.pone.0219417

Coughtrey A, Mather K (1970) Interaction and gene association and dispersion in diallel crosses where gene frequencies are unequal. Heredity 25:79-+. doi:10.1038/hdy.1970.7

de Lima VJ, Amaral ATD, Kamphorst SH, Bispo RB, Leite JT, Santos TD, Schmitt KFM, Chaves MM, de Oliveira UA, Santos P, Goncalves GMB, Khan S, Guimaraes LJM (2019) Combined Dominance and Additive Gene Effects in Trait Inheritance of Drought-Stressed and Full Irrigated Popcorn. Agronomy-Basel 9 (12). doi:10.3390/agronomy9120782

Gardner CO, Eberhart SA (1966) Analysis and Interpretation of the Variety Cross Diallel and Related Populations. Biometrics 22 (3):439-452

Griffing B (1956a) Concept of general and specific combining ability in relation to diallel crossing systems. Australian Journal of Biological Sciences 9:463-493

Griffing B (1956b) A generalised treatment of the use of diallel crosses in quantitative inheritance. Heredity 10:31-50

Hayman BI (1954) The theory and analysis of diallel crosses. Genetics 39:789-809

Hill J (1964) Effects of correlated gene distributions in analysis of diallel crosses. Heredity 19 (1):27-&. doi:10.1038/hdy.1964.3

Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. Plos Genetics 4 (2). doi:10.1371/journal.pgen.1000008

Hill WG, Maki-Tanila A (2015) Expected influence of linkage disequilibrium on genetic variance caused by dominance and epistasis on quantitative traits. Journal of Animal Breeding and Genetics 132 (2):176-186. doi:10.1111/jbg.12140

Kadam DC, Potts SM, Bohn MO, Lipka AE, Lorenz AJ (2016) Genomic Prediction of Single Crosses in the Early Stages of a Maize Hybrid Breeding Pipeline. G3-Genes Genomes Genetics 6 (11):3443-3453. doi:10.1534/g3.116.031286

Kalinina EY, Lyakh VA (2011) Genetic control systems of stem branching and plant height in linseed. Cytology and Genetics 45 (2):101-105. doi:10.3103/s0095452711020058

Kempthorne O (1956) The theory of the diallel cross. Genetics 41 (4):451–459

Leng YF, Lv CX, Li LJ, Xiang Y, Xia C, Wei RJ, Rong TZ, Lan H (2019) Heterotic grouping based on genetic variation and population structure of maize inbred lines from current breeding program in Sichuan province, Southwest China using genotyping by sequencing (GBS). Molecular Breeding 39 (3). doi:10.1007/s11032-019-0946-y

Makumbi D, Alvarado G, Crossa J, Burgueno J (2018) SASHAYDIALL: A SAS Program for Hayman's Diallel Analysis. Crop Science 58 (4):1605-1615. doi:10.2135/cropsci2018.01.0047

Mather K (1967) Complementary and duplicate gene interactions in biometrical genetics. Heredity 22:97-&. doi:10.1038/hdy.1967.8

Mather K, Jinks JJ (1971) Biometrical Genetics. Cornell University Press, New York

Mohammed R, Are AK, Munghate RS, Gaddameedi A, Bilhan KKP, Sharma HC (2018) Pattern of genetic inheritance of morphological and agronomic traits of sorghum associated with resistance to sorghum shoot fly, Atherigona soccata. Euphytica 214 (2). doi:10.1007/s10681-018-2111-9

Mowers RP, Nyhus KA, Tragesser SL, Foster DJ (2018) Doubled Haploid Testcross Design for Predicting Corn Hybrid Performance Using Molecular Markers. Crop Science 58 (2):521-528. doi:10.2135/cropsci2017.02.0086

Nassar RF (1965) Effect of correlated gene distribution due to sampling on diallel analysis. Genetics 52 (1):9-&

Pal S, Rao ES, Sriram S (2020) Genetic analysis of resistance to Fusarium oxysporum f. sp. niveum race 2 in cultivated watermelon Citrullus lanatus (Thunb.) Matsum & Nakai. Australasian Plant Pathology. doi:10.1007/s13313-020-00708-6

Pereira HD, Viana JMS, Andrade ACB, Silva FFE, Paes GP (2018) Relevance of genetic relationship in GWAS and genomic prediction. Journal of Applied Genetics 59 (1):1-8. doi:10.1007/s13353-017-0417-2

Pessoa AMS, Rego ER, dos Santos CAP, de Carvalho MG, do Rego MM (2019) Inheritance of seedling and plant traits in ornamental pepper (Capsicum annuum). Genetics and Molecular Research 18 (1). doi:10.4238/gmr18120

Rai AK, Sadashiva AT, Basavaraj YB, Venugopalan R, Rao ES, Nandeesha P (2020) Genetic analysis of bud necrosis disease caused by groundnut bud necrosis virus (GBNV) in tomato (Solanum lycopersicumL.). Euphytica 216 (8). doi:10.1007/s10681-020-02657-z

Shahadati-Moghaddam Z, Bagheri N, Jelodar NB, Kiani G, Hosseininejad SA (2017) Inheritance of southern root-knot nematode resistance in air-cured tobacco. Tropical Plant Pathology 42 (1):32-38. doi:10.1007/s40858-016-0122-4

Shirinpour M, Asghari A, Aharizad S, Rasoulzadeh A, Khorasani SK (2020) Genetic interaction and inheritance of biochemical traits can predict tolerance of hybrid maize cv. SC704 to drought. Acta Physiologiae Plantarum 42 (7). doi:10.1007/s11738-020-03110-3

Verma A, Singh Y (2018) Generation mean analysis of horticultural traits in mid-late cauliflower (Brassica oleracea L. var. botrytis) under sub temperate conditions of Western Himalayas. Plant Breeding 137 (1):97-108. doi:10.1111/pbr.12556

Verma A, Singh Y (2019) Estimation of genetic architecture of biochemical traits in mid-late cauliflower (<bold>Brassica oleracea</bold> L. var. <bold>botrytis</bold>) under sub-temperate conditions of north western Himalayas. Journal of Genetics 98 (1). doi:10.1007/s12041-019-1073-6

Viana JMS, DeLima RO, Mundim GB, Teixeira Conde AB, Vilarinho AA (2013a) Relative efficiency of the genotypic value and combining ability effects on reciprocal recurrent selection. Theoretical and Applied Genetics 126 (4):889-899. doi:10.1007/s00122-012-2023-3

Viana JMS, Pereira HD, Mundim GB, Piepho HP, Silva FFE (2018) Efficiency of genomic prediction of non-assessed single crosses. Heredity 120 (4):283-295. doi:10.1038/s41437-017-0027-0

Viana JMS, Risso LA, Oliveira deLima R, Fonseca e Silva F (2020) Factors affecting heterotic grouping with cross-pollinating crops. Agronomy Journal. doi:10.1002/agj2.20485

Viana JMS, Silva FF, Mundim GB, Azevedo CF, Jan HU (2017) Efficiency of low heritability QTL mapping under high SNP density. Euphytica 213 (1). doi:10.1007/s10681-016-1800-5

Viana JMS, Valente MSF, Silva FF, Mundim GB, Paes GP (2013b) Efficacy of population structure analysis with breeding populations and inbred lines. Genetica 141 (7-9):389-399. doi:10.1007/s10709-013-9738-1

Yu KC, Wang H, Liu XG, Xu C, Li ZW, Xu XJ, Liu JC, Wang ZH, Xu YB (2020) Large-Scale Analysis of Combining Ability and Heterosis for Development of Hybrid Maize Breeding Strategies Using Diverse Germplasm Resources. Frontiers in Plant Science 11. doi:10.3389/fpls.2020.00660

Table 1 Probability values for the hypothesis of homogeneity of $W_r - V_r$ (P1), assuming heritabilities of 40, 60, and 80, respectively, $\beta_0 = 0$ (P2), $\beta_1 = 0$ (P3), and $\beta_1 = 1$ (P4), parametric values of the coefficients of the regression of W_r on V_r (β_0 and β_1) and the coefficient of determination (R^2), parametric average heterosis (h) and correlation between the number of recessive genes and $W_r + V_r$ (r1), parametric[a] and estimated components of the genotypic variance for parents and F_1s, and total genotypic variance (V), assuming no epistasis (No), seven types of digenic epistasis[b] and an admixture of types, and 100 and 30% of epistatic genes (% eg)

% eg	Model	P1	β_0	P2	β_1	P3	P4	R^2	h	r1	D	H_1	H_2	F	F1	V
0	No	0.40/0.20/0.00	30.47	0.00	0.85	0.00	0.02	0.99	36.9	0.51	208.9[a]	85.8[a]	122.3[a]	–2.9[a]	–47.4[a]	130.0
											208.9	139.8	169.7	3.8	-	130.0
100	Co	0.46/0.05/0.00	–1.88	0.00	0.83	0.00	0.03	0.98	21.2	0.49	38.7	67.6	67.0	–22.8	-	47.8
	Du	0.94/0.53/0.04	–0.12	0.00	1.09	0.00	0.24	0.96	–11.3	0.20	34.3	32.1	22.5	36.6	-	9.3
	Do	0.00/0.00/0.00	2.01	0.00	0.32	0.10	0.00	0.77	–1.6	0.67	70.4	170.9	89.2	82.2	-	57.2
	Re	0.00/0.00/0.00	–1.88	0.00	0.72	0.00	0.00	0.99	33.5	0.55	56.4	125.8	138.2	–42.7	-	77.9
	DR	0.86/0.32/0.00	0.71	0.00	0.51	0.00	0.00	0.91	4.3	–0.1	10.0	17.2	12.1	6.6	-	7.2
	Dg	0.00/0.00/0.00	–67.49	0.00	0.97	0.00	0.36	0.99	62.9	0.43	153.8	444.9	447.3	–159.1	-	267.1
	Ne	0.00/0.00/0.00	–45.01	0.00	0.85	0.00	0.04	0.98	63.9	0.48	157.1	449.8	454.6	–157.8	-	268.7
	All	0.61/0.11/0.00	–5.63	0.00	1.05	0.00	0.37	0.96	21.8	0.43	55.4	71.7	77.0	–2.4	-	45.5
30	Co	0.77/0.22/0.00	14.4	0.00	0.86	0.00	0.02	0.99	32.2	0.46	129.0	105.2	124.7	–9.4	-	90.6
		0.44/0.05/0.00	19.1	0.00	0.83	0.00	0.01	0.99	32.1	0.52	141.9	108.2	124.5	–0.27	-	94.4
		0.36/0.03/0.00	16.6	0.00	0.92	0.00	0.18	0.99	32.9	0.54	155.8	111.4	131.0	–9.9	-	105.8
		0.28/0.02/0.00	21.7	0.00	0.83	0.00	0.01	0.99	31.7	0.52	147.2	104.2	130.7	–10.2	-	98.1
		0.50/0.07/0.00	15.8	0.00	0.90	0.00	0.12	0.99	33.4	0.48	152.6	115.8	139.2	–7.8	-	103.2
	Du	1.00/0.90/0.29	7.16	0.00	0.84	0.00	0.02	0.99	24.2	0.49	67.2	60.1	72.2	–9.9	-	50.4
	Do	0.16/0.00/0.00	14.87	0.00	0.85	0.00	0.01	0.99	25.3	0.52	126.0	87.8	84.0	68.5	-	51.7
	Re	0.59/0.10/0.00	19.75	0.00	0.76	0.00	0.00	0.99	36.4	0.48	141.2	136.0	166.4	–25.5	-	109.8
	DR	0.89/0.45/0.02	9.59	0.00	0.89	0.00	0.10	0.99	27.2	0.31	97.6	79.1	91.8	–3.74	-	62.3
	Dg	0.97/0.67/0.08	5.82	0.00	0.94	0.00	0.20	0.99	43.9	0.49	191.3	195.9	230.2	–57.4	-	164.8
	Ne	0.71/0.17/0.00	15.63	0.00	0.84	0.00	0.01	0.99	45.2	0.51	198.2	211.4	246.4	–61.1	-	173.8
	All	0.37/0.04/0.00	8.96	0.00	1.01	0.00	0.38	0.99	33.1	0.43	152.6	113.3	132.0	13.6	-	93.2
		0.81/0.27/0.01	3.22	0.00	0.97	0.00	0.37	0.99	41.5	0.56	165.8	160.5	178.9	20.8	-	108.0
		0.94/0.06/0.00	12.64	0.00	0.94	0.00	0.28	0.99	31.2	0.39	137.6	101.2	120.5	1.9	-	88.3
		0.79/0.25/0.01	9.17	0.00	0.95	0.00	0.31	0.99	32.0	0.53	131.1	106.6	126.7	–6.3	-	90.3
		0.39/0.04/0.00	14.24	0.00	0.94	0.00	0.22	0.99	30.6	0.44	142.8	100.2	123.7	–4.4	-	92.8
		0.59/0.10/0.00	13.65	0.00	0.93	0.00	0.17	0.99	33.7	0.55	151.1	114.6	136.0	–1.6	-	99.6
		0.32/0.02/0.00	15.92	0.00	0.93	0.00	0.19	0.99	31.9	0.49	151.9	106.2	127.0	–0.6	-	97.6
		0.20/0.01/0.00	12.86	0.00	1.00	0.00	0.39	0.99	31.0	0.53	151.3	100.9	121.7	0.0	-	95.7
		0.79/0.24/0.01	10.2	0.00	0.91	0.00	0.14	0.99	32.3	0.50	127.7	107.8	128.7	–3.2	-	87.2
		0.31/0.02/0.00	16.54	0.00	0.89	0.00	0.16	0.99	33.5	0.49	158.5	120.3	136.2	11.7	-	99.5

[b]Co = complementary, Du = duplicate, Do = dominant, Re = recessive, DR = dominant and recessive, Dg = duplicate genes with cumulative effects, and Ne = non-epistatic genic interaction.

Table 2 Parametric linear components of means, probability values for the tests of additive effects (P1), dominance (P2), and epistasis (P3), correlation between the parametric and estimated means (r_1), and correlation between epistatic components (my model and Mather and Jinks' model; r_2), assuming no epistasis (No), seven types of digenic epistasis[a] and an admixture of types, 100 and 30% of epistatic genes, and grain yield (GY; g/plant) and expansion volume (EV; ml/g)

Trait	Model	m	[d]	[h]	I_{22}	I_{00}	I_{11}	$E(I)^{(0)}$	$E(I)^{(1)}$	E(I)1	E(I)2	[i]	[j]	[l]	P1	P2	P3	r_1	r_2
GY	No	95.0	65.0	40.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.003	0.007	0.320	1.00	-
	Co	95.0	65.0	40.6	–8.1	70.0	16.2	14.2	15.5	8.1	61.1	-	-	-	-	-	-	-	-
		113.3	25.9	26.1	-	-	-	-	-	-	-	12.6	25.1	12.5	0.005	0.009	0.067	1.00	–0.22
	Du	95.0	65.0	40.6	–10.3	73.4	–32.2	–8.8	5.0	–88.8	65.1	-	-	-	-	-	-	-	-
		115.2	23.1	–22.8	-	-	-	-	-	-	-	11.3	–70.1	11.0	0.015	0.025	0.091	1.00	0.78
	Du[2]	95.0	65.0	40.6	–16.2	27.5	–8.2	–2.1	1.1	–24.4	17.0	-	-	-	-	-	-	-	-
		99.5	43.1	26.8	-	-	-	-	-	-	-	1.2	2.4	1.2	0.001	0.001	0.031	1.00	0.00
	Do	95.0	65.0	40.6	–87.7	–1.2	–63.4	–11.8	3.6	–151.1	47.4	-	-	-	-	-	-	-	-
		107.0	21.7	20.9	-	-	-	-	-	-	-	–56.5	–112.1	–55.6	0.052	0.033	0.078	0.97	0.73
	Re	95.0	65.0	40.6	–8.6	71.0	15.7	13.9	15.3	7.1	61.4	-	-	-	-	-	-	-	-
		113.3	25.2	25.4	-	-	-	-	-	-	-	12.9	25.4	12.6	0.004	0.006	0.041	1.00	–0.21
	DR	95.0	65.0	40.6	–9.7	120.3	14.7	25.9	33.7	5.0	110.8	-	-	-	-	-	-	-	-
		138.1	0.0	0.2	-	-	-	-	-	-	-	12.2	24.2	12.0	0.260	0.212	0.043	1.00	–0.30
	DR[2]	95.0	65.0	40.6	–2.6	36.4	5.4	8.2	10.4	2.8	34.1	-	-	-	-	-	-	-	-
		108.0	45.5	29.2	-	-	-	-	-	-	-	3.9	7.7	3.8	0.001	0.003	0.054	1.00	–0.30
	Dg	95.0	65.0	40.6	18.2	44.4	42.6	7.7	–2.5	60.8	9.2	-	-	-	-	-	-	-	-
		87.2	51.9	52.3	-	-	-	-	-	-	-	39.2	77.8	38.7	0.018	0.035	0.122	0.98	0.83
	Ne	95.0	65.0	40.6	17.2	44.5	41.6	7.8	–2.1	58.8	10.5	-	-	-	-	-	-	-	-
		87.7	51.3	51.9	-	-	-	-	-	-	-	38.1	75.7	37.5	0.017	0.028	0.116	0.97	0.82
	All	95.0	65.0	40.6	–20.8	78.5	3.6	6.5	10.4	–17.1	56.4	-	-	-	-	-	-	-	-
		110.9	15.4	15.6	-	-	-	-	-	-	-	13.0	25.7	12.8	0.008	0.016	0.052	1.00	–0.24
	All[2]	95.0	65.0	40.6	–8.7	17.6	–0.6	0.8	1.8	–9.3	14.2	-	-	-	-	-	-	-	-
		98.0	51.9	35.6	-	-	-	-	-	-	-	1.4	2.8	1.4	0.001	0.002	0.115	1.00	–0.34
EV	No	30.0	25.0	–0.81	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.001	0.076	0.201	1.00	-
	All	30.0	25.0	–0.81	–11.6	24.5	14.2	7.6	4.9	2.6	31.3	-	-	-	-	-	-	-	-
		32.7	7.0	7.0	-	-	-	-	-	-	-	3.7	7.4	3.7	0.012	0.025	0.150	1.00	–0.32

[a]Co = complementary, Du = duplicate, Do = dominant, Re = recessive, DR = dominant and recessive, Dg = duplicate genes with cumulative effects, and Ne = non-epistatic genic interaction; [2]30% of epistatic genes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- OnlineResource1.pdf