


RESEARCH

Open Access



Significance of linkage disequilibrium and epistasis on genetic variances in noninbred and inbred populations

José Marcelo Soriano Viana^{1*}  and Antonio Augusto Franco Garcia²

Abstract

Background: The influence of linkage disequilibrium (LD), epistasis, and inbreeding on genotypic variance continues to be an important area of investigation in genetics and evolution. Although the current knowledge about biological pathways and gene networks indicates that epistasis is important in determining quantitative traits, the empirical evidence for a range of species and traits is that the genotypic variance is most additive. This has been confirmed by some recent theoretical studies. However, because these investigations assumed linkage equilibrium, considered only additive effects, or used simplified assumptions for two- and higher-order epistatic effects, the objective of this investigation was to provide additional information about the impact of LD and epistasis on genetic variances in noninbred and inbred populations, using a simulated dataset.

Results: In general, the most important component of the genotypic variance was additive variance. Because of positive LD values, after 10 generations of random crosses there was generally a decrease in all genetic variances and covariances, especially the nonepistatic variances. Thus, the epistatic variance/genotypic variance ratio is inversely proportional to the LD level. Increasing inbreeding increased the magnitude of the additive, additive x additive, additive x dominance, and dominance x additive variances, and decreased the dominance and dominance x dominance variances. Except for duplicate epistasis with 100% interacting genes, the epistatic variance/genotypic variance ratio was proportional to the inbreeding level. In general, the additive x additive variance was the most important component of the epistatic variance. Concerning the genetic covariances, in general, they showed lower magnitudes relative to the genetic variances and positive and negative signs. The epistatic variance/genotypic variance ratio was maximized under duplicate and dominant epistasis and minimized assuming recessive and complementary epistasis. Increasing the percentage of epistatic genes from 30 to 100% increased the epistatic variance/genotypic variance ratio by a rate of 1.3 to 12.6, especially in inbred populations. The epistatic variance/genotypic variance ratio was maximized in the noninbred and inbred populations with intermediate LD and an average allelic frequency of the dominant genes of 0.3 and in the noninbred and inbred populations with low LD and an average allelic frequency of 0.5.

Conclusions: Additive variance is in general the most important component of genotypic variance. LD and inbreeding have a significant effect on the magnitude of the genetic variances and covariances. In general, the additive x additive variance is the most important component of epistatic variance. The maximization of the epistatic variance/

*Correspondence: jmsviana@ufv.br

¹ Department of General Biology, Federal University of Viçosa, Viçosa, MG 36570-900, Brazil

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

genotypic variance ratio depends on the LD level, degree of inbreeding, epistasis type, percentage of interacting genes, and average allelic frequency.

Keywords: Linkage disequilibrium, Epistasis, Inbreeding, Genetic variances

Background

Basic knowledge on the genetics of quantitative traits was provided by RA Fisher [1], including the partitioning of the genotypic value in effects due to individual genes, allelic interactions (dominance), and nonallelic interaction (epistasis). Furthermore, he also recognized the significance of the linkage phase between genes on the population variance and on the correlation between relatives. The influence of linkage disequilibrium (LD), epistasis, and inbreeding on genotypic variance continues to be an important area of investigation in genetics and evolution [2–4]. Assuming linkage equilibrium, multilocus model, and three to five loci interactions, A Maki-Tanila and WG Hill [4] concluded that most genotypic variance is additive, regardless of the order of interaction, allelic frequencies, and type and magnitude of interaction effects. Another main finding was that the majority of the epistatic variance is due to digenic interactions. Assuming LD and a two- to three-locus model, WG Hill and A Maki-Tanila [3] showed that variances are generally higher with positive LD and that the epistatic variance/genotypic variance ratio is largest with negative LD. Both studies showed that epistatic variance is increased by increasing heterozygosity. However, this has no impact on the relative magnitude of the epistatic variance because the additive and epistatic variances increase in similar proportions.

Based on the additive model, J Clo, J Ronfort and D Abu Awad [2] showed that assuming stabilizing selection and high mutation rates, self-pollinated populations are able to accumulate genetic variation through negative LD. Using a meta-analysis of quantitative trait heritability, J Clo, L Gay and J Ronfort [5] confirmed previous theoretical and empirical evidence that self-pollinated populations exhibit lower levels of additive variance for quantitative traits. However, the decrease in the additive variance is compensated by the nonadditive components of genotypic variance. Because of negative consequences (inbreeding depression), geneticists agree that inbreeding should be efficiently controlled to maintain adequate genetic diversity in populations [6, 7]. However, self-pollination has been deliberately used in maize hybrid breeding (currently to a lesser extent due to doubled-haploid technology). For self-pollinated crops, the development of varieties involves selection

over generations of increasing inbreeding. In these populations, inbreeding has an impact on the genetic variances and covariance between relatives [8].

Although the current knowledge about biological pathways and gene networks implies that epistasis is important in determining quantitative traits, the empirical evidence for a range of species and traits indicates that genotypic variance is most additive [9, 10]. Based on theoretical models, WG Hill, ME Goddard and PM Visscher [10] concluded that this occurs because of high differences in allelic frequencies. They also concluded that in outbred populations, the detection of epistasis is difficult unless the epistatic effects are large and the allelic frequencies are intermediate. TFC Mackay [9] emphasized that because epistasis regularly determines quantitative traits, it has consequences for plant and animal breeding, evolutionary biology, and human genetics. Recent studies on genomic selection and GWAS, including epistasis, have confirmed that most genetic variance is additive [11–14]. However, incomplete LD at low marker density can indicate epistasis when trait determination is purely additive [15].

The most important quantitative genetics theory for modeling epistasis was developed by O Kempthorne [16]. CC Cockerham [17] also provided a significant contribution. If modeling only inbreeding, LD, or epistasis is a difficult task for quantitative geneticists, jointly modeling the three events is a challenge. An impressive approach for two-genes theory in quantitative genetics assuming inbreeding, LD, and epistasis was presented by BS Weir and CC Cockerham [18]. Because of the complexity of the expressions for the genetic variances and covariance between relatives, they concluded that “the result is of little use”. That is, the functions do not allow assessing the influence of LD, epistasis, and inbreeding on the genetic variability and the degree of relationship in the populations. Furthermore, because recent investigations based on theoretical models assumed linkage equilibrium, considered only additive effects, or used simplified assumptions for two- and higher-order epistatic effects, the objective of this study was to provide additional information about the impact of LD and epistasis on the genetic variances in noninbred and inbred populations, using a simulated dataset.

Results

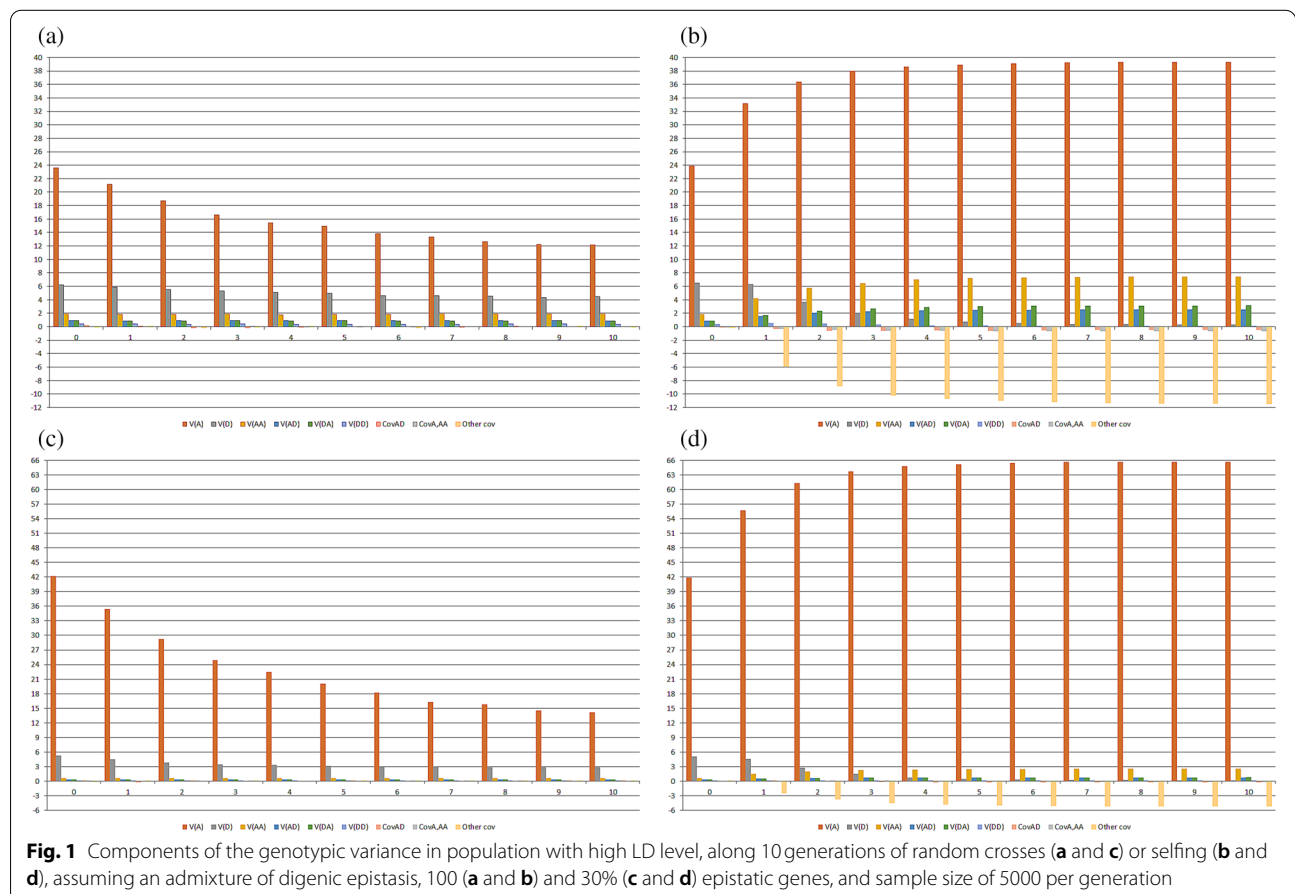
The analysis of the parametric LD in the populations shows that the LD level depends mainly on the gene density (Additional Fig. 1). A higher LD level was observed under high gene density (one gene/cM). Regardless of the gene density, the LD level was generally higher for the closest genes. Because the LD was positive, 10 generations of random crosses significantly decreased the LD level of the populations. The decrease was higher for the density of one gene/5 cM, regardless of the population (approximately -95% for r^2 , on average). The average decrease in r^2 for the density of one gene/cM was -81%. The LD level showed only a slight decrease after 10 generations of selfing regardless of the population (approximately -14% for r^2 , on average).

To assess the significance of LD and epistasis on the magnitude of the genotypic variance components, we assumed the density of one gene/5 cM. In general, regardless of the type of epistasis, percentage of interacting genes, LD level, and degree of inbreeding, the most important component of the genotypic variance was the additive variance (Additional Figs. 2 to 8 and Fig. 1). Only under duplicate epistasis, 100% epistatic genes, and

$F \geq 7/8$ was the additive x additive variance higher than the additive variance (Additional Fig. 3). The impact of LD on the genetic variances and covariances is shown by the changes in their magnitudes over the random cross generations (Additional Figs. 2 to 8 and Fig. 1). Because of positive LD values, after 10 generations of random crosses there was generally a decrease in all genetic variances and covariances, especially the nonepistatic variances. The decreases in the additive and dominance variances ranged between -28 and -70% and -12 to -62%, respectively, depending on the type of epistasis and the percentage of interacting genes. The changes in the epistatic variances were much lower, ranging from 0.5 to -13%. Thus, the epistatic variance/genotypic variance ratio is inversely proportional to the LD level.

Because there was only a slight decrease in the LD level with inbreeding, the changes in the magnitudes of the genetic variances and covariances over generations of selfing are mainly attributable to inbreeding. Increasing inbreeding increased the magnitude of the additive,

additive x additive, additive x dominance, and dominance x additive variances, and decreased the dominance and dominance x dominance variances (Additional



Figs. 2 to 8 and Fig. 1). The additive variance increased from 50 to 76% and the epistatic variances increased in the range 114 to 863%, depending on the type of epistasis and the percentage of epistatic genes. The decreases in the dominance and dominance \times dominance variances were similar, in the range -76 to -98%. Except for duplicate epistasis with 100% interacting genes, the epistatic variance/genotypic variance ratio was proportional to the level of inbreeding.

In general, additive \times additive was the most important component of epistatic variance, regardless of the type of epistasis, percentage of epistatic genes, LD level, and degree of inbreeding (Additional Figs. 2 to 8 and Fig. 1). This variance corresponded to 41 to 48% and 25 to 64% of the epistatic variance in the noninbred and inbred populations, respectively. Concerning the genetic covariances, in general, they showed lower magnitudes relative to the genetic variances and positive and negative signs, regardless of the type of epistasis, percentage of interacting genes, LD level, and degree of inbreeding (Additional Figs. 2 to 8 and Fig. 1). The sum of the covariances achieved significant values with inbreeding and 100% epistatic genes. The total value was positive under dominant epistasis and negative for the other epistasis types.

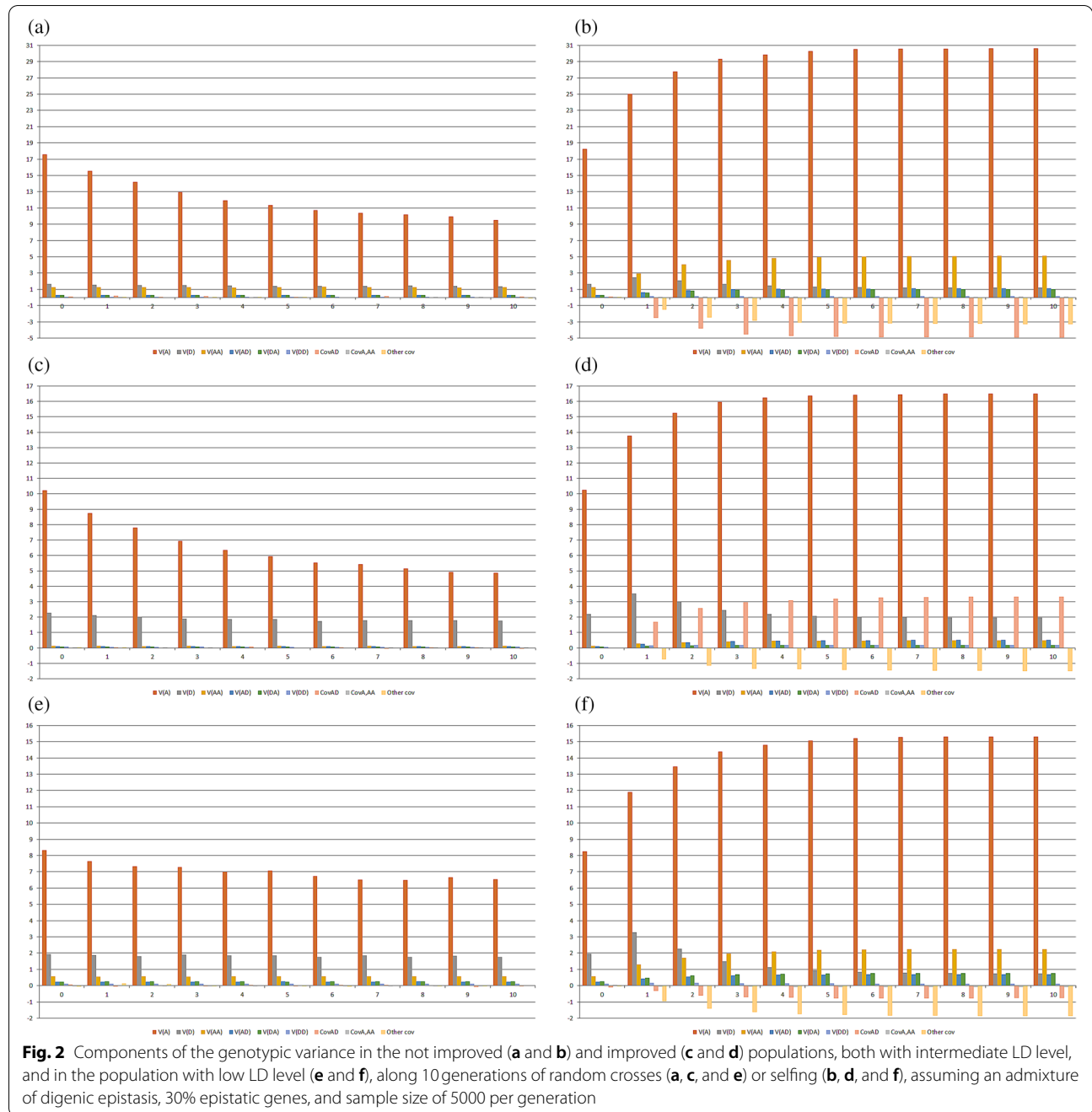
In addition to the LD level and degree of inbreeding, the type of epistasis, percentage of interacting genes, and average allelic frequencies affect the magnitude of the epistatic variances. The epistatic variance/genotypic variance ratio was maximized under duplicate and dominant epistasis and minimized assuming recessive and complementary epistasis irrespective of the percentage of interacting genes (Additional Figs. 2 to 8 and Fig. 1). Increasing the percentage of epistatic genes from 30 to 100% increased the epistatic variance/genotypic variance ratio by a rate of 1.3 to 12.6, especially in inbred populations. Fixing the LD level at an intermediate level and assuming an admixture of epistasis types and 30% interacting genes, the epistatic variance/genotypic variance ratio was maximized in the population with an average allelic frequency for the dominant genes of 0.3, relative to the population with an average allelic frequency of 0.7, especially in the noninbred populations (Fig. 2). Thus, increasing the average allelic frequency from 0.3 to 0.7 decreased in approximately -70% the epistatic variance/genotypic variance ratio in the noninbred populations but lead to a slight increase of the ratio in the inbred populations. The epistatic variance/genotypic variance ratios in the noninbred and inbred populations (ranges of 9 to 15% and 9 to 10%, respectively) were greater than the ratios in the population with high LD and an average allelic frequency of 0.5 (ranges of 3 to 8% for both random cross and selfing

generations). But they are comparable to the ratios in the noninbred population with low LD and an average allelic frequency of 0.5 along the generations of random crosses (9 to 12%). In the inbred generations from the low LD population with an average allelic frequency of 0.5, the epistatic variance/genotypic variance ratio ranged from 10 to 22%. Decreasing the LD level from high to low, under an average allelic frequency of 0.5, maximized the epistatic variance/genotypic variance ratio in both noninbred and inbred populations. The increases ranged from 66 to 238%, especially in inbred populations. Thus, under positive dominance, the epistatic variance/genotypic variance ratio was maximized in the noninbred and inbred populations with intermediate LD and an average allelic frequency for the dominant genes of 0.3 and in the noninbred and inbred populations with low LD and an average allelic frequency of 0.5.

Discussion

WG Hill, ME Goddard and PM Visscher [10] emphasize that knowledge about the relative magnitudes of the additive, dominance, and epistatic variances is important in evolutionary biology, medicine, and agriculture. However, the theoretical investigation about the joint significance of LD, epistasis, and inbreeding on the genetic variances for a quantitative trait is a challenge, even when fixing the number of genes, the allelic frequencies, and the degrees of dominance. One main reason is that the theory available is too complex to allow the assessment of the relative magnitudes of the genetic variances [3, 4, 10, 19]. The other main reason is the large number of combinations between levels of LD (low to high) and inbreeding (not inbred to completely inbred) with distinct percentages of epistatic genes (for example, 30 to 100%), degree of epistasis (digenic to a high order), and type of epistasis (up to seven types of digenic epistasis, complementary or duplicate trigenic or high-order epistasis, or an admixture of types).

BS Weir and CC Cockerham [18] derived very complex functions for the components of the genotypic variance assuming a two-gene model with inbreeding, LD, and epistasis and concluded that the functions are of "little use". T Wang and ZB Zeng [19] highlight that their theoretical results serve only as a framework to understand and properly interpret estimates of the genetic effects and variance components in a QTL mapping experiment. The theoretical models investigated by WG Hill, ME Goddard and PM Visscher [10], assuming linkage equilibrium, predict high proportions of additive variance even in the presence of non-additive gene action. Assuming linkage equilibrium, the theoretical results from A Maki-Tanila and WG Hill [4] showed that the epistatic variance is of



low magnitude compared with the additive variance, even when assuming high heterozygosity. They also emphasize that the majority of the epistatic variance is due to two-locus interactions. Based on theoretical models including LD, WG Hill and A Maki-Tanila [3] confirmed that most of the genotypic variance in a segregating population is additive.

Because the main conclusion from the previously described studies is that most of the genotypic variance is additive, we believe that our simulation-based

study provides significant additional knowledge about the influence of LD, inbreeding and epistasis on genetic variances. Our study has a strong theoretical background on quantitative genetics for modelling LD and epistasis. We assumed low to high LD levels for genes, not inbred to completely inbred populations, 30 and 100% epistatic genes, and the seven types of digenic epistasis. Although there is evidence for higher-order epistasis, pairwise interactions contribute substantially to phenotypic variation between individuals [4, 20].

Our results agree with the main conclusions from WG Hill and A Maki-Tanila [3], A Maki-Tanila and WG Hill [4], and WG Hill, ME Goddard and PM Visscher [10], showing that LD significantly affects genetic variances and that most of the genotypic variance is additive. However, from the analyses assuming an admixture of the types of epistasis and 30% of interacting genes, the epistatic variance/genotypic variance ratio was maximized in the populations with intermediate LD and an average allelic frequency of 0.3 for the dominant genes and with low LD and an average frequency of 0.5 regardless of the generation and degree of inbreeding. The ratio was minimized in the populations with intermediate LD and an average allelic frequency for the dominant genes of 0.7 and with high LD and an average frequency of 0.5. Our results also give support to the main conclusions of J Clo, J Ronfort and D Abu Awad [2], who assumed an additive model under LD and distinct selfing rates. The differences observed for outcrossing species rely on their assumption of negative LD.

This study and the investigations of WG Hill and A Maki-Tanila [3], A Maki-Tanila and WG Hill [4], and WG Hill, ME Goddard and PM Visscher [10] show that the impact of LD, epistasis, and inbreeding on genetic variances depends on the LD level, predominant type and order of epistasis, percentage of epistatic genes, magnitude of the epistatic effects, average allelic frequency, and degree of inbreeding. The results based on simulated datasets agree with the results from QTL mapping based on field data and with the current knowledge about biological pathways and gene networks. They indicate that epistasis is important in determining quantitative traits for a range of species and traits, and that the genotypic variance is most additive [9]. Because only in specific situations can the epistatic variance constitute a high proportion of the genotypic variance, the simulated results explain why is difficult to detect epistasis, especially under small sample size [9].

In random cross populations, as human subpopulations, there are variable degree of LD [21] but inbreeding coefficients close to zero even when the subpopulation has a limited effective size or a rate of consanguineous matings higher than that expected under random crosses [22]. This imply that in random cross populations, LD but not inbreeding can significantly affects the genotypic variance and the covariance between relatives. In noninbred populations under recurrent selection, this implies that the emphasis should continue to be selecting based on estimated/predicted breeding values even when fitting the additive-model. In self-pollinated crops, the simulated

results show that the epistatic variance can constitute a significant fraction of the genotypic variance, especially under a high percentage of interacting genes. But, the main component of the epistatic variance is due to additive x additive effects. Because the covariance between parent and offspring depends mainly on additive and additive x additive effects, selecting based on estimated/predicted breeding values from fitting the additive-dominance with additive x additive epistasis model should be effective. Recognizing that epistasis can be an important effect determining quantitative traits, several recently studies on genomic selection and GWAS included epistatic effects, aiming to increase the prediction accuracy and the power of QTL detection [11–14]. In these studies, the epistatic variance ranged from 0 to 9.5% of the phenotypic variance.

Conclusions

Additive variance is in general the most important component of genotypic variance. LD and inbreeding have a significant effect on the magnitude of the genetic variances and covariances. In general, the additive x additive variance is the most important component of the epistatic variance. The maximization of the epistatic variance/genotypic variance ratio depends on the LD level, degree of inbreeding, epistasis type, percentage of interacting genes, and average allelic frequency.

Methods

Additive and dominance genetic values in inbred populations

Assume initially a single biallelic gene (A/a) determining a quantitative trait, where A is the gene that increases the trait expression, and a population derived by n generations of selfing from a Hardy-Weinberg equilibrium population (generation 0). Defining M_F^1 and M_F^2 as the means of the inbred population after an allelic substitution for the genes A and a, respectively, the average effect of the allelic genes in the inbred population are $\alpha_A^{(n)} = M_F^1 - M_F = q\alpha + 2Fpqd$ and $\alpha_a^{(n)} = M_F^2 - M_F = -p\alpha + 2Fpqd$, where $M_F = m + (p - q)a + 2pqd - 2Fpqd = M - 2Fpqd$ is the inbred population mean, p and q are the allelic frequencies, α is the average effect of an allelic substitution, F is the inbreeding coefficient, and M is the noninbred population mean. Thus, the additive values in the inbred population are $A_{AA}^{(n)} = 2q\alpha + 4Fpqd = A_{AA}^{(0)} + 4Fpqd$, $A_{Aa}^{(n)} = (q - p)\alpha + 4Fpqd = A_{Aa}^{(0)} + 4Fpqd$, and $A_{aa}^{(n)} = -2p\alpha + 4Fpqd = A_{aa}^{(0)} + 4Fpqd$, where $A^{(0)}$ is the additive value in the noninbred population. Note that $E(A^{(n)}) = 4Fpqd$. Expressing the genotypic values in the inbred population as a function of M_F , we have:

$$G_{AA} = M_F + A_{AA}^{(0)} + (-2q^2d + 2Fpqd) = M_F + A_{AA}^{(0)} + (D_{AA}^{(0)} + 2Fpqd) = M_F + A_{AA}^{(0)} + D_{AA}^{(n)}$$

$$G_{Aa} = M_F + A_{Aa}^{(0)} + (2pqd + 2Fpqd) = M_F + A_{Aa}^{(0)} + (D_{Aa}^{(0)} + 2Fpqd) = M_F + A_{Aa}^{(0)} + D_{Aa}^{(n)}$$

$$G_{aa} = M_F + A_{aa}^{(0)} + (-2p^2d + 2Fpqd) = M_F + A_{aa}^{(0)} + (D_{aa}^{(0)} + 2Fpqd) = M_F + A_{aa}^{(0)} + D_{aa}^{(n)}$$

Note that in the inbred population, $E(A^{(0)}) = E(D^{(n)}) = 0$ but $E(D^{(0)}) = -2Fpqd$. Note also that the additive value in the noninbred population is the additive value in the inbred population expressed as deviation from its mean ($A^{(0)} = A^{(n)} - 4Fpqd$) and the dominance value in the inbred population is the dominance value in the noninbred population expressed as deviation from its mean ($D^{(n)} = D^{(0)} + 2Fpqd$). This implies that, in the inbred population, $E(G) = M_F$.

Genetic variances in inbred populations in LD

Assume now two linked biallelic genes (A/a and B/b) determining a quantitative trait and a noninbred population in LD (generation 0). Assume dominance but initially no epistasis. After n generations of selfing, the genotypic variance for the two genes in the inbred population is (see the genotype probabilities in the [Additional File Appendix](#)) $\sigma_G^{2(n)} = \sigma_A^{2(n)} + \sigma_D^{2(n)} + 2\sigma_{A,D}^{(n)}$, where:

$$\begin{aligned} \sigma_A^{2(n)} &= (1 + F)(2p_a q_a \alpha_a^2 + 2p_b q_b \alpha_b^2) \\ &\quad + 2[2 + c_1(1 - 2r_{ab})] \Delta_{ab}^{(-1)} \alpha_a \alpha_b \\ &= (1 + F)\sigma_A^{2(0)} + 2[c_1(1 - 2r_{ab}) - 2F] \Delta_{ab}^{(-1)} \alpha_a \alpha_b \end{aligned}$$

is the additive variance,

$$\begin{aligned} \sigma_D^{2(n)} &= (1 - F^2)(4p_a^2 q_a^2 d_a^2 + 4p_b^2 q_b^2 d_b^2) + F[4p_a q_a (p_a - q_a)^2 d_a^2 + 4p_b q_b (p_b - q_b)^2 d_b^2] \\ &\quad + 8\left\{ (1 - F)(c^n - 1 + F)p_a q_a p_b q_b + (p_a - q_a)(p_b - q_b)[(1 - F)c^n - (1 - 2F) + c_1(1 - 2r_{ab})/2] \Delta_{ab}^{(-1)}/2 + (1 - F)c^n \Delta_{ab}^{(-1)2} \right\} d_a \\ d_b &= (1 - F^2)\sigma_D^{2(0)} + FD_2 + 8\left\{ (1 - F)(c^n - 1 + F)p_a q_a p_b q_b + (p_a - q_a)(p_b - q_b)[(1 - F)c^n - (1 - 2F) + c_1(1 - 2r_{ab})/2] \Delta_{ab}^{(-1)}/2 + [(1 - F)c^n - (1 - F^2)] \Delta_{ab}^{(-1)2} \right\} d_a d_b \end{aligned}$$

is the dominance variance, and

$$\begin{aligned} \sigma_{A,D}^{(n)} &= 2F[2p_a q_a (p_a - q_a) \alpha_a d_a + 2p_b q_b (p_b - q_b) \alpha_b d_b] \\ &\quad + [2F + c_1(1 - 2r_{ab})] \Delta_{ab}^{(-1)} [(p_b - q_b) \alpha_a d_b + (p_a - q_a) \alpha_b d_a] \\ &= 2FD_1 + [2F + c_1(1 - 2r_{ab})] \Delta_{ab}^{(-1)} [(p_b - q_b) \alpha_a d_b + (p_a - q_a) \alpha_b d_a] \end{aligned}$$

is the covariance between additive and dominance values,

where $\Delta_{ab}^{(-1)} = P_{AB}^{(-1)} \cdot P_{ab}^{(-1)} - P_{Ab}^{(-1)} \cdot P_{aB}^{(-1)}$ is the measure of LD in the gametic pool of generation -1 [23], where $P^{(-1)}$ is a haplotype probability, r_{ab} is the recombination frequency, $c_1 = 2\{1 - [(1 - 2r_{ab})/2]^n\}/(1 + 2r_{ab})$, $c = 1 - 2r_{ab}(1 - r_{ab})$, $\sigma_A^{2(0)} = 2p_a q_a \alpha_a^2 + 2p_b q_b \alpha_b^2 + 4\Delta_{ab}^{(-1)} \alpha_a \alpha_b$ and $\sigma_D^{2(0)} = 4p_a^2 q_a^2 d_a^2 + 4p_b^2 q_b^2 d_b^2 + 8\Delta_{ab}^{(-1)} d_a d_b$ are the additive and dominance variances in the noninbred population in LD [24], and D_1 (covariance of a and d) and D_2 (variance of d) are the components of the covariance of relatives from self-fertilization, assuming linkage equilibrium [8]. The other terms are the covariances between the average effects of an allelic substitution, between dominance deviations, and between the average effect of an allelic substitution and dominance deviation, for genes in LD. Because we assumed biallelic genes, $H = \sigma_D^2$. Thus, $(1 - F^2)\sigma_D^{2(0)} = (1 - F)\sigma_D^{2(0)} + F(1 - F)H$. Note that the genotypic variance derived here is a general formulation for the Cockerham's genotypic variance c_{ggg} [8], assuming LD. If $p = q$, $\sigma_{A,D}^{(n)} = 0$.

Assuming LD but no inbreeding, the genotypic variance after n generations of random cross in the noninbred population in LD is $\sigma_G^{2(n)} = \sigma_A^{2(n)} + \sigma_D^{2(n)}$, because $\sigma_{A,D}^{(n)} = 0$, where:

$$\sigma_A^{2(n)} = 2p_a q_a \alpha_a^2 + 2p_b q_b \alpha_b^2 + 4(1 - r_{ab})^n \Delta_{ab}^{(-1)} \alpha_a \alpha_b$$

$$\sigma_D^{2(n)} = 4p_a^2 q_a^2 d_a^2 + 4p_b^2 q_b^2 d_b^2 + 8\left[(1 - r_{ab})^n \Delta_{ab}^{(-1)} \right]^2 d_a d_b$$

Thus, the genotypic variance can increase or decrease after n generations of random cross in a noninbred population, depending on the sign of the LD measure. The LD value is positive for genes in coupling phase and negative for genes in repulsion phase.

Epistasis in noninbred and inbred populations in LD

The quantitative genetics theory for modelling epistasis in a population in LD is a generalization of the theory proposed by O Kempthorne [16], who assumed a noninbred population in linkage equilibrium and any number of alleles. We assumed biallelism. It should be emphasized that the Kempthorne’s theory allows a generalization from two to three or more interacting genes. But fitting three or more interacting genes in a population in LD is a challenge because the genotype probabilities for three or more genes in LD are too complex to derive. Furthermore, only complementary and duplicate epistasis can be easily defined for three or more epistatic genes.

Assume now that the two previous defined genes are epistatic. The genotypic value is [16]:

$$G_{ijkl} = M + \alpha_i^1 + \alpha_j^1 + \alpha_k^2 + \alpha_l^2 + \delta_{ij}^1 + \delta_{kl}^2 + (\alpha^1 \alpha^2)_{ik} + (\alpha^1 \alpha^2)_{jk} + (\alpha^1 \alpha^2)_{il} + (\alpha^1 \alpha^2)_{jl} + (\alpha^1 \delta^2)_{ikl} + (\alpha^1 \delta^2)_{jkl} + (\delta^1 \alpha^2)_{ijk} + (\delta^1 \alpha^2)_{ijl} + (\delta^1 \delta^2)_{ijkl} = M + A + D + AA + AD + DA + DD$$

where AA, AD, DA, and DD are the additive x additive, additive x dominance, dominance x additive, and dominance x dominance epistatic genetic values.

The parametric values of the 36 parameters for the nine genotypic values are obtained by solving the equations $\beta = (X' VX)^{-1} X' Vy$, under the restrictions defined by O Kempthorne [16], where X is the incidence matrix, $V = diagonal \left\{ f_{ij}^{(n)} \right\}$ is the diagonal matrix of the genotype probabilities, and y is the vector of the genotypic values (G_{ij}) (i, j=0, 1, and 2).

O Kempthorne [16] provided explicit functions for all effects because he assumed linkage equilibrium. Assuming LD makes very difficult to derive such functions but the following results hold:

- 1) the expectation of the breeding value is zero regardless of the degree of inbreeding in the population.
- 2) the expectation of the dominance value is $E(D)^{(n)} = p_a q_a F(\delta_{AA} - 2\delta_{Aa} + \delta_{aa}) + p_b q_b F(\delta_{BB} - 2\delta_{Bb} + \delta_{bb})$; then, defining the dominance value in an inbred population as the dominance value expressed as deviation from its mean ($D^{(n)} = D - E(D)^{(n)}$), $E(D^{(n)}) = 0$.
- 3) the expectation of the additive x additive value is zero only if there is no LD.
- 4) the expectation of the additive x dominance value is zero only if $F = 0$ or $p = q$ for all genes.

- 5) the expectation of the dominance x additive value is zero only if $F = 0$ or $p = q$ for all genes.
- 6) the expectation of the dominance x dominance value is zero only if $F = 0$ and there is no LD.

Thus, defining the additive x additive, additive x dominance, dominance x additive, and dominance x dominance epistatic values as the values expressed as deviation from its mean, $AA^{(n)} = AA - E(AA)^{(n)}$, $AD^{(n)} = AD - E(AD)^{(n)}$, $DA^{(n)} = DA - E(DA)^{(n)}$, and $DD^{(n)} = DD - E(DD)^{(n)}$, the genotypic value in an inbred population can be expressed as

$$G = M + E(D)^{(n)} + E(AA)^{(n)} + E(AD)^{(n)} + E(DA)^{(n)} + E(DD)^{(n)} + A + D^{(n)} + AA^{(n)} + AD^{(n)} + DA^{(n)} + DD^{(n)} = M_F + A + D^{(n)} + AA^{(n)} + AD^{(n)} + DA^{(n)} + DD^{(n)}$$

This implies that $E(G) = M_F$. If $F = 0$ then

$$G = M + E(AA) + E(DD) + A + D + [AA - E(AA)] + AD + DA + [DD - E(DD)] = M^* + A + D + AA^* + AD + DA + DD^*$$

where,

$$E(AA) = 2\Delta_{ab}^{(-1)}(\alpha_A \alpha_B - \alpha_A \alpha_b - \alpha_a \alpha_B + \alpha_a \alpha_b)$$

and

$$E(DD) = \left[\Delta_{ab}^{(-1)} \right]^2 (\delta_{AA} \delta_{BB} - 2\delta_{AA} \delta_{Bb} + \delta_{AA} \delta_{bb} - 2\delta_{Aa} \delta_{BB} + 4\delta_{Aa} \delta_{Bb} - 2\delta_{Aa} \delta_{bb} + \delta_{aa} \delta_{BB} - 2\delta_{aa} \delta_{Bb} + \delta_{aa} \delta_{bb})$$

This implies that $E(G) = M^*$. If $F = 0$ and there is no LD,

$$G = M + A + D + AA + AD + DA + DD$$

where the linear components are those defined by O Kempthorne [16]. This implies that $E(G) = M$.

In noninbred populations in LD, only the additive and dominance values are not correlated. The genotypic variance in these populations is, in simplified form,

$$\sigma_G^{2(0)} = \sigma_A^{2(0)} + \sigma_D^{2(0)} + \sigma_{AA}^{2(0)} + 2\sigma_{A,AA}^{(0)} + 2\sigma_{D,AA}^{(0)} + \dots$$

where

$$\sigma_{AA}^{(0)} = f_{22}^{(0)} [(4\alpha_A\alpha_B)]^2 + \dots + f_{00}^{(0)} [(4\alpha_a\alpha_b)]^2 - [E(AA)^{(0)}]^2$$

$$\sigma_{A,AA}^{(0)} = 2\Delta_{ab}^{(-1)} \left[\alpha^A (\alpha_A\alpha_B - \alpha_A\alpha_b + \alpha_a\alpha_B - \alpha_a\alpha_b) + \alpha^B (\alpha_A\alpha_B - \alpha_a\alpha_B + \alpha_A\alpha_b - \alpha_a\alpha_b) \right]$$

$$\sigma_{D,AA}^{(0)} = -4\Delta_{ab}^{(-1)} [p_a q_a d_a (\alpha_A\alpha_B - \alpha_A\alpha_b - \alpha_a\alpha_B + \alpha_a\alpha_b) + p_b q_b d_b (\alpha_A\alpha_B - \alpha_a\alpha_B - \alpha_A\alpha_b + \alpha_a\alpha_b)]$$

where, to avoid confusion, α^A and α^B are the average effects of an allelic substitution.

The assumption of LD makes very difficult to derive the components of the genotypic variance (additive, dominance, and epistatic variances and the covariances between these effects), even assuming noninbred populations, biallelic genes, and only digenic epistasis. In respect to the types of digenic epistasis, the following can be defined [25, 26]:

1. 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 , assuming independent assortment).
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).
3. 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).
4. 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).
5. 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).
6. 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).
7. Nonepistatic 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).

Simulated datasets

Because the magnitude of the components of genotypic variance generally cannot be inferred from previous functions, all means and genetic variances and covariances were computed from simulated datasets provided by *REALbreeding* software (available upon request). This program uses the quantitative genetics theory that was described in the previous sections and in JMS Viana [24]. *REALbreeding* has been used to provide simulated data in investigations in the areas of genomic selection [27], GWAS [28], QTL mapping [29], linkage disequilibrium [30], population structure [31], and heterotic grouping/genetic diversity [32].

The program simulates individual genotypes for genes and molecular markers and phenotypes in three steps using user inputs. The first step (genome simu-

lation) is the specification of the number of chromosomes, molecular markers, and genes as well as marker type and density. The second step (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 final step (trait simulation) is the specification of the individual phenotypes. In this stage, the user informs the minimum and maximum genotypic values for homozygotes (to compute the deviations), the minimum and maximum phenotypic values (to avoid outliers), the direction and degree of dominance (to compute the dominance deviations/d), and the broad sense heritability. The current version allows the inclusion of digenic epistasis, gene x environment interaction, and multiple traits (up to 10), including pleiotropy. The population mean (M) and additive (A), dominance (D), and epistatic (AA, AD, DA, and DD) genetic values or general and specific combining ability effects (GCA and SCA) and epistatic values (I), or genotypic values (G), depending on the population, are calculated from the parametric gene effects and frequencies and the parametric LD values. 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 population in LD is generated by crossing two populations in linkage equilibrium followed by a generation of random crosses. This generation of random crosses aims to generate a population in Hardy-Weinberg equilibrium. Thus, generation 0 (the founder population) is a population in Hardy-Weinberg equilibrium in LD for linked genes and molecular markers, and the individuals are not related. The parametric LD in this population is $\Delta_{ab}^{(-1)} = [(1 - 2r_{ab})/4](p_{a1} - p_{a2})(p_{b1} - p_{b2})$, where the indices 1 and 2 indicate the parental populations.

The quantitative genetics theory for epistasis does not solve the challenge of studying genetic variability and covariance between relatives in populations, using simulated datasets, even assuming simplified scenarios such as

linkage equilibrium and no inbreeding. 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 , a , d , and d/a (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 epistatic 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.

We simulated grain yield (g/plant) assuming 400 genes in 10 chromosomes of 200 and 50 cM (40 genes/chromosome). The average density was approximately one gene/5 cM and one gene/cM, respectively. Assuming a density of one gene/cM, we simulated a population with an average frequency of 0.5 and high LD. Under the density of one gene/5 cM, we generated four populations, one with a high LD level and one with a low LD level, both with an average frequency of 0.5, and two populations with an intermediate LD level and an average frequency for the favorable genes of 0.3 (not improved) and 0.7 (improved). We defined positive dominance (average degree of dominance of 0.6), maximum and minimum genotypic values for homozygotes of 160 and 30 g.plt⁻¹, and maximum and minimum phenotypic values of 180 and 10 g.plt⁻¹. The broad sense heritability was 20%. For each population, we assumed additive-dominance with digenic epistasis model defining 100 and 30% of interacting genes. Concerning the ratio $V(I)/(V(A) + V(D))$, the analyses assuming ratios 1, 10, and 100 evidenced that increasing the ratio from 1 to 10 and 100 increased the epistatic variances but also increased the additive and dominance variances. Then, because the main conclusions for the greater ratios were essentially the same provided by ratio 1, we will present only the results for ratio 1. With epistasis, we assumed a single type or an admixture of the seven types. We ranged the degree of inbreeding from 0.0 to 1.0, assuming 10 generations of selfing. We also assumed 10 generations of random crosses. The population size was 5000 per generation.

The characterization of the LD in the populations was based on the parametric Δ , r^2 , and D' values for the 40 genes in chromosome 1, which were provided by *REALbreeding* (it should be similar for the other chromosomes). The heatmaps were processed using the R package *pheatmap*. Assuming no epistasis, the software provides parametric additive and dominance genetic values and parametric genetic variances and covariances. Assuming epistasis, the software provides parametric additive, dominance, and epistatic genetic values. Thus, under epistasis, the genetic variances and covariances were computed from the parametric genetic values, using a sample size of 5000 individuals per generation.

Abbreviations

LD: Linkage disequilibrium; A: Additive value; D: Dominance value; AA: Additive x additive value; AD: Additive x dominance value; AD: Dominance x additive value; DD: Dominance x dominance value; G: Genotypic value; I: Epistatic value.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08335-9>.

Additional file 1. Figures and Appendix.

Acknowledgements

We 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.

Authors' contributions

JMSV designed the study, developed the software, processed the data, and wrote the manuscript. AAFG designed the study, processed the data, and revised the manuscript. The authors read and approved the final manuscript.

Funding

None.

Availability of data and materials

The dataset is available at <https://doi.org/10.6084/m9.figshare.13607306.v2>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of General Biology, Federal University of Viçosa, Viçosa, MG 36570-900, Brazil. ²Department of Genetics, Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, SP 13418-900, Brazil.

Received: 28 June 2021 Accepted: 22 January 2022
Published online: 09 April 2022

References

- Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edinburgh*. 1918;52(2):399–433.
- Clo J, Ronfort J, Abu Awad D. Hidden genetic variance contributes to increase the short-term adaptive potential of selfing populations. *J Evol Biol*. 2020;33(9):1203–15.
- Hill WG, Maki-Tanila A. Expected influence of linkage disequilibrium on genetic variance caused by dominance and epistasis on quantitative traits. *J Anim Breed Genet*. 2015;132(2):176–86.
- Maki-Tanila A, Hill WG. Influence of gene interaction on complex trait variation with multilocus models. *Genetics*. 2014;198(1):355–67.
- Clo J, Gay L, Ronfort J. How does selfing affect the genetic variance of quantitative traits? An updated meta-analysis on empirical results in angiosperm species. *Evolution*. 2019;73(8):1578–90.
- Hasselgren M, Noren K. Inbreeding in natural mammal populations: historical perspectives and future challenges. *Mammal Rev*. 2019;49(4):369–83.
- Howard JT, Pryce JE, Baes C, Maltecca C. Invited review: inbreeding in the genomics era: inbreeding, inbreeding depression, and management of genomic variability. *J Dairy Sci*. 2017;100(8):6009–24.
- Cockerham CC. Covariances of relatives from self-fertilization. *Crop Sci*. 1983;23:1177–80.
- Mackay TFC. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet*. 2014;15(1):22–33.
- Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet*. 2008;4(2):e1000008.
- Vitezica ZG, Legarra A, Toro MA, Varona L. Orthogonal estimates of variances for additive, dominance, and Epistatic effects in populations. *Genetics*. 2017;206(3):1297–307.
- Fornieris NS, Vitezica ZG, Legarra A, Perez-Enciso M. Influence of epistasis on response to genomic selection using complete sequence data. *Genet Sel Evol*. 2017;49:66.
- Su G, Christensen OF, Ostersen T, Henryon M, Lund MS. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. *PLoS One*. 2012;7(9):e45293.
- Monir MM, Zhu J. Comparing GWAS results of complex traits using full genetic model and additive models for revealing genetic architecture. *Sci Rep*. 2017;7:38600.
- Misztal I, Aguilar I, Lourenco D, Ma L, Steibel J, Toro M. Emerging issues in genomic selection. *J Anim Sci*. 2021;99(6):1–14.
- Kempthorne O. The theoretical values of correlations between relatives in random mating populations. *Genetics*. 1954;40:153–67.
- Cockerham CC. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics*. 1954;39:859–82.
- Weir BS, Cockerham CC. Two-locus theory in quantitative genetics. In: *International conference on quantitative genetics: 1976*. Ames: The Iowa State University Press; 1976. p. 247–69.
- Wang T, Zeng ZB. Models and partition of variance for quantitative trait loci with epistasis and linkage disequilibrium. *BMC Genet*. 2006;7.
- Domingo J, Baeza-Centurion P, Lehner B. The Causes and Consequences of Genetic Interactions (Epistasis). *Annu Rev Genomics Hum Genet*. 2019;20:433–60.
- Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet*. 2003;4(8):587–97.
- Santos S, Kok F, Weller M, Paiva FRL, Otto PA. Inbreeding levels in Northeast Brazil: strategies for the prospecting of new genetic disorders. *Genet Mol Biol*. 2010;33:220–3.
- Kempthorne O. *An introduction to genetic statistics*. Ames: The Iowa State University Press; 1973.
- Viana JMS. Quantitative genetics theory for non-inbred populations in linkage disequilibrium. *Genet Mol Biol*. 2004;27(4):594–601.
- Viana JMS. Dominance, epistasis, heritabilities and expected genetic gains. *Genet Mol Biol*. 2005;28(1):67–74.
- Viana JMS. Components of variation of polygenic systems with digenic epistasis. *Genet Mol Biol*. 2000;23(4):883–92.
- Viana JMS, Pereira HD, Piepho HP, Silva FFE. Efficiency of genomic prediction of nonassessed testcrosses. *Crop Sci*. 2019;59(5):2020–7.
- Pereira HD, Viana JMS, Andrade ACB, Silva FFE, Paes GP. Relevance of genetic relationship in GWAS and genomic prediction. *J Appl Genet*. 2018;59(1):1–8.
- Viana JMS, Silva FF, Mundim GB, Azevedo CF, Jan HU. Efficiency of low heritability QTL mapping under high SNP density. *Euphytica*. 2017;213(1):13.
- Andrade ACB, Viana JMS, Pereira HD, Pinto VB, Fonseca ESF. Linkage disequilibrium and haplotype block patterns in popcorn populations. *PLoS One*. 2019;14(9):e0219417.
- Viana JMS, Valente MSF, Silva FF, Mundim GB, Paes GP. Efficacy of population structure analysis with breeding populations and inbred lines. *Genetica*. 2013;141(7–9):389–99.
- Viana JMS, Riso LA, Oliveira deLima R, Fonseca e Silva F. Factors affecting heterotic grouping with cross-pollinating crops. *Agronomy J*. 2021;113:210–23.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

