Significance of linkage disequilibrium and epistasis on the genetic variances and covariance between relatives in non-inbred and inbred populations

José Marcelo Soriano Viana[1*] and Antonio Augusto Franco Garcia[2]

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

[2]Department of Genetics, Luiz de Queiroz College of Agriculture, University of São Paulo, 13418-900, Piracicaba, SP, Brazil.

[*]Corresponding author. E-mail: jmsviana@ufv.br. ORCID: https://orcid.org/0000-0002-5063-4648.

Abstract Because no feasible theoretical model can depict the complexity of phenotype development from a genotype, the joint significance of linkage disequilibrium (LD), epistasis, and inbreeding on the genetic variances remains unclear. The objective of this investigation was to assess the impact of LD and epistasis on the genetic variances and covariances between relatives in non-inbred and inbred populations using simulated data. We provided the theoretical background and simulated grain yield assuming 400 genes in 10 chromosomes of 200 and 50 cM. We generated five populations with low to high LD levels, assuming 10 generations of random cross and selfing. The analysis of the parametric LD in the populations shows that the LD level depends mainly on the gene density. The significance of the LD level is impressive on the magnitude of the genotypic and additive variances, which is the most important component of the genotypic variance, regardless of the LD level and the degree of inbreeding. Regardless of the type of epistasis, the ratio epistatic variance/genotypic variance is proportional to the percentage of the epistatic genes. For the epistatic variances, except for duplicate epistasis and dominant and recessive epistasis, with 100% of epistatic genes, their magnitudes are much lower than the magnitude of the additive variance. The additive x additive variance is the most important epistatic variance. Our results explain why LD for genes and relationship information are key factors affecting the genomic prediction accuracy of complex traits and the efficacy of association studies.

Keywords: linkage disequilibrium, epistasis, inbreeding, genetic variances.

Introduction

Genomic selection or genomic prediction of complex quantitative traits and genome-wide association studies (GWAS) share a common quantitative genetics background: linkage disequilibrium (LD) between genes and single nucleotide polymorphisms (SNPs). Among the thousands of studies that have been published in these areas, only a few have offered a quantitative genetics background, indicating that the accuracy of the genomic prediction and the power to detect a candidate gene depends on the LD between genes and SNPs (Gianola et al. 2009; Goddard 2009; Viana et al. 2016, 2017a). Based on Cockerham (1954), Viana et al. (2016) and Viana et al. (2017a)

provided explicit functions relating SNP additive, dominance, and epistatic effects and variances to the quantitative trait locus (QTL) additive, dominance, and epistatic effects and variances. For a biallelic QTL (alleles A/a) and a SNP (alleles B/b) in LD, the connection between both their effects and variances is the measure of LD in the gametic pool of the population, which is given by the difference between the products of the haplotypes $\Delta = P_{AB}P_{ab} - P_{Ab}P_{aB}$ (Kempthorne, 1973). This LD measure can be expressed as $\Delta = r\sqrt{pqst}$ and $\Delta = D'.\max(e)$, where r is the correlation between the values of alleles in both loci in the same gamete, p and q are the QTL allelic frequencies, s and t are the SNP allelic frequencies, D' is a relative measure of LD (in relation to the maximum value for given allele frequencies), and max(e) is the maximum deviation of the actual gamete frequency from linkage equilibrium (Hill and Robertson 1968; Lewontin 1964).

However, the relationship information is the most important factor affecting the genomic prediction accuracy. For GWAS, the relationship information allows an effective control of the type I error rate and decreases the number of significant associations outside of the QTL intervals (Liu et al. 2015; Pereira et al. 2018). Although the observed relationship information, which is computed from the molecular data, has provided superior prediction accuracies and a more effective control of the false discovery rate (FDR), relative to the expected relationship information (Liu et al. 2016; Munoz et al. 2014), the covariance between relatives is another significant quantitative genetics background in the quantitative genomics era.

Several investigations on genomic selection, genomic prediction of complex traits, and GWAS have included epistasis (Monir and Zhu 2017; Varona et al. 2018). With few exceptions, a common feature of these significant studies is the absence of basic genetics background on the epistasis types (see any standard genetics book). These studies consider the basic quantitative genetics background for modelling epistasis, but there is generally no reference for the key knowledge provided by Kempthorne (1954) and Cockerham (1954) on modelling epistatic effects and defining the epistatic genomic relationship matrices. That is, for Bayesian and random regression (based on SNP effects) and genomic (based on genetic effects) best linear unbiased

prediction (BLUP) approaches, the additive x additive, additive x dominance, dominance x additive, and dominance x dominance SNP/QTL effects are modeled, but in simulation-based studies there is no specification of any type of epistasis as complementary, duplicate, dominant, recessive, dominant and recessive, duplicate genes with cumulative effects, and non-epistatic genic interaction. The significant scientific contributions of these studies rely on the definition of SNP coding for the epistatic effects and specification of the epistatic genomic relationship matrices, which are deduced following the key proposition by VanRaden (2008), aiming maximization of the prediction accuracy and power to detect interacting QTLs (Jiang and Reif 2020; Martini et al. 2017; Vitezica et al. 2017).

Because of negative consequences (nominated inbreeding depression), human, conservative, animal, and cross-pollinated species geneticists agree that inbreeding should be efficiently controlled to maintain adequate genetic diversity in the populations (Hasselgren and Noren 2019; Howard et al. 2017). However, self-pollination has been deliberately used in maize hybrid breeding (currently to a lesser extent due to the doubled-haploid technology). For self-pollinated crops, the development of varieties involves selection over generations of increasing inbreeding. In these populations the inbreeding has an impact on the genetic variances and covariance between relatives (Cockerham 1983).

As previously highlighted, the most important quantitative genetics theory for modelling epistasis was developed by Kempthorne (1954). Cockerham (1954) also provided a significant contribution. If modelling only inbreeding, LD, or epistasis is a difficult task for the quantitative geneticists, jointly modelling the three events is a challenge. An impressive approach for two genes theory in quantitative genetics assuming inbreeding, LD, and epistasis was presented by Weir and Cockerham (1977). 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 these factors on the genetic variability and the degree of relationship in the populations. Because LD, relationship information, epistasis, and inbreeding are

key factors for genomic prediction and association studies, the objective of this investigation was to assess the impact of LD and epistasis on the genetic variances and covariance between relatives in non-inbred and inbred populations, using a simulated data set.

Material and Methods

Additive and dominance genetic values in inbred populations in LD

Assume initially a single biallelic gene (A/a) determining a quantitative trait, where A is the gene that increases the trait expression. The genotype probabilities in a population derived by n generations of selfing from a Hardy-Weinberg equilibrium population (generation 0) are $p^2 + pqF$, $2pq(1-F)$, and $q^2 + pqF$, if AA, Aa, and aa, respectively, where p and q are the allelic frequencies for A and a, respectively, and $F = 1 - (1/2)^n$ is the inbreeding coefficient. The inbred population mean is $M_F = m + (p-q)a + 2pqd - 2Fpqd = M - 2Fpqd$, where m is the mean of the genotypic values of the homozygotes, a is the difference between the genotypic value of the homozygote of greater expression and m, d is the dominance deviation, $-2Fpqd$ is the change in the population mean due to inbreeding, and M is the mean of the non-inbred population (Hallauer and Miranda Filho, 1988). 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 $\alpha = \alpha_A^{(n)} - \alpha_a^{(n)} = [a + (q-p)d]$ is the average effect of an allelic substitution. Note that the average effect of an allelic substitution is the same in the non-inbred and in the inbred populations. 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 non-inbred population. Note that $E\left(A^{(n)}\right) = 4Fpqd$. Expressing the genotypic values in the inbred population as a function of M_F, we have:

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

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

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

Note that in the inbred population, $E\left(A^{(0)}\right) = E\left(D^{(n)}\right) = 0$ but $E\left(D^{(0)}\right) = -2Fpqd$. Note also that the additive value in the non-inbred population is the additive value in the inbred population expressed as deviation from its mean $\left(A^{(0)} = A^{(n)} - 4Fpqd\right)$ and the dominance value in the inbred population is the dominance value in the non-inbred population expressed as deviation from its mean $\left(D^{(n)} = D^{(0)} + 2Fpqd\right)$. This imply 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 non-inbred population in LD (generation 0). Assume dominance but initially no epistasis. The genotype probabilities in generation 0 ($f_{ij}^{(0)}$) are presented by Viana (2004), where i and j (i, j = 0, 1, or 2) are the number of copies of the gene that increase the trait expression (A and B). For example, $f_{22}^{(0)} = p_a^2 p_b^2 + 2p_a p_b \Delta_{ab}^{(-1)} + \left[\Delta_{ab}^{(-1)}\right]^2$, where $\Delta_{ab}^{(-1)} = P_{AB}^{(-1)}.P_{ab}^{(-1)} - P_{Ab}^{(-1)}.P_{aB}^{(-1)}$ is the measure of LD in the gametic pool of generation –1 (Kempthorne, 1973). After n generations of selfing, the genotype probabilities are:

$$f_{22}^{(n)} = f_{22}^{(0)} + (F/2)\left[f_{21}^{(0)} + f_{12}^{(0)}\right] + P_1^{(n)}$$

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

$$f_{20}^{(n)} = f_{20}^{(0)} + (F/2)\left[f_{21}^{(0)} + f_{10}^{(0)}\right] + P_2^{(n)}$$

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

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

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

$$f_{02}^{(n)} = f_{02}^{(0)} + (F/2)\left[f_{01}^{(0)} + f_{12}^{(0)}\right] + P_2^{(n)}$$

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

$$f_{00}^{(n)} = f_{00}^{(0)} + (F/2)\left[f_{01}^{(0)} + f_{10}^{(0)}\right] + P_1^{(n)}$$

where

$$P_1^{(n)} = (1/4)\left\{[F - (1-F)(1-c^n)]f_{11}^{(0)} + c_1(1-2r_{ab})\Delta_{ab}^{(-1)}\right\}$$

$$P_2^{(n)} = (1/4)\left\{[F - (1-F)(1-c^n)]f_{11}^{(0)} - c_1(1-2r_{ab})\Delta_{ab}^{(-1)}\right\}$$

$$c = 1 - 2r_{ab}(1-r_{ab})$$

$$c_1 = 2\{1 - [(1-2r_{ab})/2]^n\}/(1+2r_{ab})$$

and r_{ab} is the recombination frequency.

The genotypic variance for the two genes in the inbred population is $\sigma_G^{2(n)} = \sigma_A^{2(n)} + \sigma_D^{2(n)} + 2\sigma_{A,D}^{(n)}$, where:

$\sigma_A^{2(n)} = (1+F)(2p_aq_a\alpha_a^2 + 2p_bq_b\alpha_b^2) + 2[2 + c_1(1-2r_{ab})]\Delta_{ab}^{(-1)}\alpha_a\alpha_b = (1+F)\sigma_A^{2(0)} + 2[c_1(1-2r_{ab}) - 2F]\Delta_{ab}^{(-1)}\alpha_a\alpha_b$ is the additive variance,

$\sigma_D^{2(n)} = (1-F^2)(4p_a^2q_a^2d_a^2 + 4p_b^2q_b^2d_b^2) + F[4p_aq_a(p_a-q_a)^2d_a^2 + 4p_bq_b(p_b-q_b)^2d_b^2] + 8\{(1-F)(c^n - 1 + F)p_aq_ap_bq_b + (p_a-q_a)(p_b-q_b)[c^n(1-F) - (1-2F)]/2 +$

1−FcnΔab(−1)2dadb=1−F2σD2(0)+FD2+81−Fcn−1+Fpaqapbqb+

pa−qapb−qbcn1−F−1−2F/2+1−Fcn−1−F2Δab(−1)2dadb is the dominance variance, and

$\sigma_{A,D}^{(n)} = 2F[2p_aq_a(p_a-q_a)\alpha_ad_a + 2p_bq_b(p_b-q_b)\alpha_bd_b] + [2F + c_1(1-2r_{ab})]\Delta_{ab}^{(-1)}[(p_b -$

qbαadb+pa−qaabda=2FD1+2F+c11−2rabΔab−1pb− *qbαadb+pa−qaabda* is the covariance between additive and dominance values,

where $\sigma_A^{2(0)}$ and $\sigma_D^{2(0)}$ are the additive and dominance variances in the non-inbred population in LD (Viana 2004), 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 (Cockerham 1983). 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, $\breve{H} = \sigma_D^2$. Thus, $(1 - F^2)\sigma_D^{2(0)} = (1 - F)\sigma_D^{2(0)} + F(1 - F)\breve{H}$. Note that the genotypic variance derived here is a general formulation for the Cockerham's genotypic variance c_{ggg}, 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 non-inbred 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 decreases after n generations of random cross in a non-inbred population. Assuming LD, no inbreeding, and no epistasis, a general formulation for the covariance between relatives is

$$Cov\left(G_X^{(n)}, G_Y^{(n+t)}\right) = 2r_{XY}\sigma_A^{2(n)} + u_{XY}\sigma_D^{2(n)}$$

where r_{XY} is the Malecot's coefficient of relationship between the relatives X in generation n and Y in generation n + t and u_{XY} is the probability that the parents of X (f and m, for the female and male) and Y (f' and m', for the female and male) have genotypes identical by descent ($u_{X,Y} = r_{ff'} \cdot r_{mm'} + r_{fm'} r_{mf'}$). The covariance between parent and offspring and between half- and full-sibs were derived by Viana (2004):

$$Cov\left(P^{(n)}, O^{(n+t)}\right) = (1/2)^t \sigma_A^{2(n)}$$

$$Cov\left(HS^{(n+1)}\right) = (1/4)\sigma_A^{2(n)}$$

$$Cov\left(FS^{(n+1)}\right) = (1/2)\sigma_A^{2(n)} + (1/4)\sigma_D^{2(n)}$$

Assuming LD, inbreeding, and no epistasis, it is not straightforward to develop a general expression for the covariance between relatives from selfing ($c_{tgg'}$) derived by Cockerham (1983), assuming no LD and no epistasis. Note that the covariance between parent (generation 0) and its

selfed progeny (generation 1) derived by Viana (2004) depends also on the covariance between α and d for distinct genes in LD, but the coefficients for the additive and dominance variances and for D_1 are those predicted by Cockerham (1983):

$$c_{001} = \sigma_A^{2(0)} + (1/2)\sigma_D^{2(0)} + (1/2)D_1 + \Delta_{ab}^{(-1)}[(p_b - q_b)\alpha_a d_b + (p_a - q_a)\alpha_b d_a]$$

Epistasis in non-inbred 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 Kempthorne (1954), who assumed a non-inbred population in linkage equilibrium and any number of alleles. We assumed biallelism. It should be highlighted 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 (Kempthorne 1954):

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

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 $\beta = (X'VX)^{-1}X'Vy$, 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). Following Kempthorne (1954), we completed the rank of X with 31 of the following restrictions (not all linearly independent):

i) restrictions for the average effects of genes: 1) $p_a\alpha_A + q_a\alpha_a = 0$ and 2) $p_b\alpha_B + q_b\alpha_b = 0$.

ii) restrictions for the dominance values: 1) $p_a\delta_{AA} + q_a\delta_{Aa} = 0$; 2) $p_a\delta_{Aa} + q_a\delta_{aa} = 0$; 3) $p_b\delta_{BB} + q_b\delta_{Bb} = 0$, and 4) $p_b\delta_{Bb} + q_b\delta_{bb} = 0$.

iii) restrictions for the AA values: 1) $p_b(\alpha_A\alpha_B) + q_b(\alpha_A\alpha_b) = 0$; 2) $p_b(\alpha_a\alpha_B) + q_b(\alpha_a\alpha_b) = 0$; 3) $p_a(\alpha_A\alpha_B) + q_a(\alpha_a\alpha_B) = 0$; and 4) $p_a(\alpha_A\alpha_b) + q_a(\alpha_a\alpha_b) = 0$.

iv) restrictions for the AD values: 1) $p_b(\alpha_A\delta_{BB}) + q_b(\alpha_A\delta_{Bb}) = 0$; 2) $p_b(\alpha_A\delta_{Bb}) + q_b(\alpha_A\delta_{bb}) = 0$; 3) $p_b(\alpha_a\delta_{BB}) + q_b(\alpha_a\delta_{Bb}) = 0$; 4) $p_b(\alpha_a\delta_{Bb}) + q_b(\alpha_a\delta_{bb}) = 0$; 5) $p_a(\alpha_A\delta_{BB}) + q_a(\alpha_a\delta_{BB}) = 0$; 6) $p_a(\alpha_A\delta_{Bb}) + q_a(\alpha_a\delta_{Bb}) = 0$; and 7) $p_a(\alpha_A\delta_{bb}) + q_a(\alpha_a\delta_{bb}) = 0$ (six out of the seven; four out of the seven if there is no dominance for the second gene).

v) restrictions for the DA values: 1) $p_a(\delta_{AA}\alpha_B) + q_a(\delta_{Aa}\alpha_B) = 0$; 2) $p_a(\delta_{Aa}\alpha_B) + q_a(\delta_{aa}\alpha_B) = 0$; 3) $p_a(\delta_{AA}\alpha_b) + q_a(\delta_{Aa}\alpha_b) = 0$; 4) $p_a(\delta_{Aa}\alpha_b) + q_a(\delta_{aa}\alpha_b) = 0$; 5) $p_b(\delta_{AA}\alpha_B) + q_b(\delta_{AA}\alpha_b) = 0$; 6) $p_b(\delta_{Aa}\alpha_B) + q_b(\delta_{Aa}\alpha_b) = 0$; and 7) $p_b(\delta_{aa}\alpha_B) + q_b(\delta_{aa}\alpha_b) = 0$ (six out of the seven; four out of the seven if there is no dominance for the first gene).

vi) restrictions for the DD values: 1) $p_b(\delta_{AA}\delta_{BB}) + q_b(\delta_{AA}\delta_{Bb}) = 0$; 2) $p_b(\delta_{AA}\delta_{Bb}) + q_b(\delta_{AA}\delta_{bb}) = 0$; 3) $p_b(\delta_{Aa}\delta_{BB}) + q_b(\delta_{Aa}\delta_{Bb}) = 0$; 4) $p_b(\delta_{Aa}\delta_{Bb}) + q_b(\delta_{Aa}\delta_{bb}) = 0$; 5) $p_b(\delta_{aa}\delta_{BB}) + q_b(\delta_{aa}\delta_{Bb}) = 0$; 6) $p_b(\delta_{aa}\delta_{Bb}) + q_b(\delta_{aa}\delta_{bb}) = 0$; 7) $p_a(\delta_{AA}\delta_{BB}) + q_a(\delta_{Aa}\delta_{BB}) = 0$; 8) $p_a(\delta_{Aa}\delta_{BB}) + q_a(\delta_{aa}\delta_{BB}) = 0$; 9) $p_a(\delta_{AA}\delta_{Bb}) + q_a(\delta_{Aa}\delta_{Bb}) = 0$; 10) $p_a(\delta_{Aa}\delta_{Bb}) + q_a(\delta_{aa}\delta_{Bb}) = 0$; 11) $p_a(\delta_{AA}\delta_{bb}) + q_a(\delta_{Aa}\delta_{bb}) = 0$; and 12) $p_a(\delta_{Aa}\delta_{bb}) + q_a(\delta_{aa}\delta_{bb}) = 0$ (nine out of the 12; eight out of the 12 if there is no dominance for both genes).

Kempthorne (1954) 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 $E(D)^{(n)} = p_aq_aF(\delta_{AA} - 2\delta_{Aa}+\delta_{aa}) + p_bq_bF(\delta_{BB} - 2\delta_{Bb}+\delta_{bb})$; then, defining the dominance value in an inbred population as the dominance value expressed as deviation from its mean $\left(D^{(n)} = D - E(D)^{(n)}\right)$, $E\left(D^{(n)}\right) = 0$.

3) the expectation of the additive x additive value is

$$E(AA)^{(n)} = \left(4f_{22}^{(n)} + 2f_{21}^{(n)} + 2f_{12}^{(n)} + f_{11}^{(n)}\right)(\alpha_A\alpha_B) + \left(4f_{20}^{(n)} + 2f_{21}^{(n)} + 2f_{10}^{(n)} + f_{11}^{(n)}\right)(\alpha_A\alpha_b) + \left(4f_{02}^{(n)} + 2f_{12}^{(n)} + 2f_{01}^{(n)} + f_{11}^{(n)}\right)(\alpha_a\alpha_B) + \left(4f_{00}^{(n)} + 2f_{10}^{(n)} + 2f_{01}^{(n)} + f_{11}^{(n)}\right)(\alpha_a\alpha_b)$$; this expectation is zero only if there is no LD.

4) the expectation of the additive x dominance value is

$$E(AD)^{(n)} = \left(2f_{22}^{(n)} + f_{12}^{(n)}\right)(\alpha_A\delta_{BB}) + \left(2f_{21}^{(n)} + f_{11}^{(n)}\right)(\alpha_A\delta_{Bb}) + \left(2f_{20}^{(n)} + f_{10}^{(n)}\right)(\alpha_A\delta_{bb}) + \left(2f_{02}^{(n)} + f_{12}^{(n)}\right)(\alpha_a\delta_{BB}) + \left(2f_{01}^{(n)} + f_{11}^{(n)}\right)(\alpha_a\delta_{Bb}) + \left(2f_{00}^{(n)} + f_{10}^{(n)}\right)(\alpha_a\delta_{bb})$$; this expectation is zero only if F = 0 or if there is no LD.

5) the expectation of the dominance x additive value is

$$E(DA)^{(n)} = \left(2f_{22}^{(n)} + f_{21}^{(n)}\right)(\delta_{AA}\alpha_B) + \left(2f_{12}^{(n)} + f_{11}^{(n)}\right)(\delta_{Aa}\alpha_B) + \left(2f_{02}^{(n)} + f_{01}^{(n)}\right)(\delta_{aa}\alpha_B) + \left(2f_{20}^{(n)} + f_{21}^{(n)}\right)(\delta_{AA}\alpha_b) + \left(2f_{10}^{(n)} + f_{11}^{(n)}\right)(\delta_{Aa}\alpha_b) + \left(2f_{00}^{(n)} + f_{01}^{(n)}\right)(\delta_{aa}\alpha_b)$$; this expectation is zero only if F = 0 or if there is no LD.

6) the expectation of the dominance x dominance value is

$E(DD)^{(n)} = f_{22}^{(n)}(\delta_{AA}\delta_{BB}) + \cdots + f_{00}^{(n)}(\delta_{aa}\delta_{bb})$; this expectation is zero only if 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

$$E(AA) = \left(4f_{22}^{(0)} + 2f_{21}^{(0)} + 2f_{12}^{(0)} + f_{11}^{(0)}\right)(\alpha_A\alpha_B) + \cdots + \left(4f_{00}^{(0)} + 2f_{10}^{(0)} + 2f_{01}^{(0)} + f_{11}^{(0)}\right)(\alpha_a\alpha_b)$$

and $E(DD) = f_{22}^{(0)}(\delta_{AA}\delta_{BB}) + \cdots + f_{00}^{(0)}(\delta_{aa}\delta_{bb})$.

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 Kempthorne (1954). This implies that $E(G) = M$.

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 non-inbred populations, biallelic genes, and only digenic epistasis. In respect to the types of digenic epistasis, the following can be defined (Viana 2000, 2005):

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).

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. 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).

Simulated data sets

Because the magnitude of the components of the genotypic variance generally cannot be inferred from the previous functions, all means and genetic variances and covariances were computed from the analyses of simulated data sets provided by the software *REALbreeding* (available upon request). This software uses the quantitative genetics theory that was described in

the previous sections and in Viana (2004). *REALbreeding* has been used to provide simulated data in investigations in the areas of genomic selection (Viana et al. 2019), GWAS (Pereira et al. 2018), QTL mapping (Viana et al. 2017b), linkage disequilibrium (Andrade et al. 2019), population structure (Viana et al. 2013), and heterotic grouping/genetic diversity (Viana et al. 2020).

The software simulates individual genotypes for genes and molecular markers and phenotypes in three steps using user inputs. The first step (genome simulation) 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, 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 (up to 10), including pleiotropy. The population mean (M), additive (A), dominance (D), and epistatic (AA, AD, DA, and DD) genetic values or general and specific combining ability effects (GCA and SCA) 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 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 quantitative genetics theory for epistasis does not solve the challenge of studying genetic variability and covariance between relatives in populations, using simulated data sets, even assuming simplified scenarios such as linkage equilibrium and no inbreeding. This is why there are few significant publications about the influence of epistasis on the genetic variability in populations based on the definitive theory proposed by Kempthorne (1954). 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 interacting genes the software samples a random value for the epistatic value I_{22} (the epistatic value for the genotype AABB), assuming $I_{22} \sim N(0, V(I))$. For complementary epistasis, for example, the other epistatic effects and genotypic values are computed as described below, checking if the genotypic values meet the specifications that were previously defined:

$$G_{22} = (m_a + m_b + a_a + a_b) + I_{22}$$

$$I_{21} = G_{22} - (m_a + m_b + a_a + d_b)$$

$$G_{21} = (m_a + m_b + a_a + d_b) + I_{21}$$

$$I_{12} = G_{22} - (m_a + m_b + d_a + a_b)$$

$$G_{12} = (m_a + m_b + d_a + a_b) + I_{12}$$

$$I_{11} = G_{22} - (m_a + m_b + d_a + d_b)$$

$$G_{11} = (m_a + m_b + d_a + d_b) + I_{11}$$

$$I_{20} = -I_{22} - I_{21}$$

$$G_{20} = (m_a + m_b + a_a - a_b) + I_{20}$$

$$I_{10} = G_{20} - (m_a + m_b + d_a - a_b)$$

$$G_{10} = (m_a + m_b + d_a - a_b) + I_{10}$$

$$I_{02} = G_{20} - (m_a + m_b - a_a + a_b)$$

$$G_{02} = (m_a + m_b - a_a + a_b) + I_{02}$$

$$I_{01} = G_{20} - (m_a + m_b - a_a + d_b)$$

$$G_{01} = (m_a + m_b - a_a + d_b) + I_{01}$$

$$I_{00} = G_{20} - (m_a + m_b - a_a - a_b)$$

$$G_{00} = (m_a + m_b - a_a - a_b) + I_{00}$$

We simulated grain yield assuming 400 genes in 10 chromosomes of 200 and 50 cM (40 genes/chromosome). The average density was approximately one gene each five and one cM, respectively. We generated five populations, two with high LD level and one with low LD level, all three with an average minor allelic frequency (maf) of 0.5, and two populations with intermediate LD level and an average maf of 0.3 but with contrasting average frequency for the favorable genes (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-dominant model and additive-dominant with digenic epistasis model, defining 100% and 30% of interacting genes and a ratio V(I)/(V(A) + V(D)) equal to 1.0. 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 5,000 per generation.

The characterization of the LD in the populations was based on the parametric D, 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.

Results

The analysis of the parametric LD in the populations shows that the LD level depends mainly on the gene density (Figure 1). The higher LD level was observed under high gene density (one gene each cM). Regardless of the gene density, the LD level is generally higher for the closest genes. As expected, 10 generations of random cross significantly decreased the LD level of the populations. The decrease was higher for the density of one gene each five cM, regardless of the

population (on average, approximately −95% for r^2). The average decrease in the r^2 for the density of one gene each cM was −81%. The LD level showed only a slight decrease after 10 generations of selfing (on average, approximately −14% for r^2, regardless of the population).

To characterize the magnitude of the genotypic variance components in non-inbred and inbred populations with contrasting LD levels, assuming no epistasis, we assumed a density of one gene each five cM. Because the populations with high and low LD levels have an average maf of 0.5, the decrease in the population mean due to inbreeding and the genotypic and additive variances are maximized, relative to populations with average maf lower or higher than 0.5. The same is true for the dominance variance in non-inbred populations. After 10 generations of selfing, the decreases in the population means were −15 and −17% for the populations with low and high LD level, respectively (Figure 2). The significance of the LD level is impressive on the magnitude of the genotypic and additive variances. Regardless of the LD level and the degree of inbreeding, the additive variance is the most important component of the genotypic variance,. The additive variance in the population with high LD is 6.8 times greater than the additive variance in the population with low LD in generation 0, 2.9 times greater after 10 generations of random cross, and 5.7 times greater after 10 generations of selfing. In the populations with intermediate LD level, the decreases in the population mean due to inbreeding are similar in magnitude. In both improved and not improved populations, there is also a decrease in the magnitude of the additive variance with random crosses and an increase with selfing. The additive variance is greater in the not improved population, regardless of the generation. In both populations the magnitude of the additive variance is intermediate to the values observed for the populations with high and low LD level.

To characterize the components of the genotypic variance in non-inbred and inbred populations with high LD level, assuming epistasis, we also assumed the density of one gene each five cM. Regardless of the type of epistasis and the percentage of interacting genes, there are non-significant changes in the population mean along 10 generations of random cross (remember that the average decrease in the r^2 values was approximately –95%) (Figure 3). With 10 generations of

selfing, regardless of the percentage of epistatic genes, except for recessive epistasis, the inbreeding decreased the population mean in −19 to −20% with dominant epistasis to −9 to −10% with dominant and recessive epistasis (remember that the decrease assuming no epistasis was −17%).

Regardless of the type of epistasis, the ratio epistatic variance/genotypic variance is proportional to the percentage of the epistatic genes. The epistatic variance in generation 0 corresponded to 2 to 7% (dominant epistasis) of the genotypic variance with 30% of epistatic genes, but it corresponded to 10 to 64% (duplicate epistasis) assuming 100% of epistatic genes (Figures 4 to 11). Irrespective of the type of epistasis and the percentage of epistatic genes, after 10 generations of random cross or selfing the ratio epistatic variance/genotypic variance increased in the range of 11 to 660%. This occurred because the decrease in the genotypic variance was much higher than the decrease in the epistatic variance with random cross and because the increase in the genotypic variance was much lower than the increase in the epistatic variance with selfing. With three exceptions, regardless of the type of epistasis and the percentage of epistatic genes, the most important component of the genotypic variance is also the additive variance. The magnitude of the additive variance decreased with random cross and increased with selfing. With duplicate epistasis and 100% of epistatic genes, the additive x additive variance was higher than the additive variance, throughout 10 generations of random cross or selfing. Assuming dominant epistasis and dominant and recessive epistasis, 100% of epistatic genes, and inbred population, the sum of several epistatic covariances (after four selfings) and the additive x additive variance were greater than the additive variance, respectively. Except for duplicate genes with cumulative effects and non-epistatic genic interaction, the magnitude of the additive variance was two to 16 times greater assuming 30% of epistatic genes, compared with 100% of epistatic genes, for both random cross and selfing. Assuming duplicate genes with cumulative effects and non-epistatic genic interaction, the magnitude of the additive variance was 1.1 and 1.4 times higher with 100% of interacting genes, respectively.

For the epistatic variances, except for duplicate epistasis and dominant and recessive epistasis, 100% of epistatic genes, their magnitudes are much lower than the magnitude of the additive variance. The additive x additive variance is the most important epistatic variance. Generally, a non-significant variation in the magnitudes of the epistatic variances was observed throughout 10 generations of random cross, regardless of the type of epistasis and the percentage of the epistatic genes. A significant increase in the magnitude of the additive x additive, additive x dominant, and dominant x additive variances occurred with selfing, regardless of the percentage of epistatic genes and the type of epistasis. When inbreeding increased, the dominant x dominant variance decreased in the population with high LD but increased in the other populations.

For the populations with intermediate and low LD levels, the previous inferences generally holds but the magnitudes of the genotypic and genetic variances are generally lower than the values for the population of high LD level, regardless of generation, type of epistasis, and percentage of epistatic genes, as exemplified assuming 30% of epistatic genes showing all types of epistasis (Figure 12). With no exception, the additive variance is also the most important component of the genotypic variance, regardless of the generation.

Discussion

Comstock and Robinson (1948) provided the basic knowledge on Quantitative Genetics, partitioning the genotypic variance in the variance due to the average effects of the genes (additive) and the variance due to allelic interaction (dominance). Assuming effects between non-allelic genes, Kempthorne (1954) and Cockerham (1954) demonstrated that the genotypic variance and the covariance between relatives depends on the additive, dominance, and four epistatic variances, assuming digenic epistasis. Cockerham and Weir (1977) derived the very complex functions for the components of the genotypic variance assuming a two-gene model with inbreeding, LD, and epistasis.

Meuwissen et al. (2001) proposed a method for predicting the additive values of non-phenotyped individuals who were genotyped for thousands of SNPs, from the model fitted by

analyzing a limited number of phenotyped and genotyped individuals. Since then, the knowledge on gene effects, LD, and covariance between relatives has been directly or implicitly used for modelling additive, dominance, and epistatic SNP and QTL effects, deriving observed genomic relationship matrices, and assessing prediction accuracy, power of QTL detection, and rate of FDR in thousands of genomic prediction and association studies (Stich and Gebhardt 2011; Varona et al. 2018; Vitezica et al. 2017).

Unfortunately, for over 70 years the joint significance of LD, epistasis, and inbreeding on the genotypic variance and the covariance between relatives remained unclear. This is because the theory available, even assuming only two loci, is of "little use" (Cockerham and Weir, 1977). Additionally, no feasible theoretical model can depict the complexity of the development of a phenotype from a genotype (Robinson et al. 2014). Thus, our simulation-based study provides a better understanding of the influence of LD and epistasis on the genotypic variance and the covariance between relatives in non-inbred and inbred populations.

We assumed low to high LD levels for genes, under a relatively low gene density, and digenic epistasis. In maize, for example, the genome size is approximately 2 Gb, including approximately 42,000 to 56,000 genes. The density is approximately 1-11 genes per 100 kb over a relatively even distribution (Haberer et al. 2005). Because grain yield is affected by most of these genes, the gene density for this very complex quantitative trait should be higher than the gene density for other less complex quantitative traits. Although there is evidence for higher-order epistasis, pairwise epistasis can contribute substantially to phenotypic variation between individuals (Domingo et al. 2019).

Our main results explain the general knowledge on the efficacy of genomic prediction and GWAS provided by numerous published papers. Because the LD has a significant positive effect on the magnitude of the genotypic and genetic variances, especially the additive variance, the efficacy of the additive value genomic prediction and GWAS are proportional to the LD level for genes (Forneris et al. 2017) and to the degree of inbreeding (Martini et al. 2017). The accuracy of future predictions in non-inbred populations, such as human populations and in animal breeding, decreases

mainly due to the decrease in the LD, because the covariance between relatives in distinct generations depends on the components of the genotypic variance in the first generation. Updating training data after some generations implies in defining a new training population with lower LD and, consequently, lower components of the genotypic variance, but it makes the individuals in the training and selection generations more related. However, in animal breeding, probable because changes in the SNP and QTL frequencies due to selection, increasing the LD level for SNPs and QTLs, field and simulation-based studies have showed increase in the prediction accuracy (Neyhart et al. 2017).

Because the epistatic variances have a lower magnitude relative to the additive variance, some field and simulation-based investigations have not shown an increase in the additive value prediction accuracy by fitting SNP epistatic effects (Chen et al. 2019; Forneris et al. 2017; Vitezica et al. 2018). However, including epistasis can improve accuracy and unbiasedness of genomic predictions (Su et al. 2012) as well as QTL detection power and control of the type I error in GWAS (Monir and Zhu 2017; Pecanka et al. 2017; Stich and Gebhardt 2011). However, epistatic variances can be the most important components of the genotypic variance with higher-order epistasis, especially assuming a high number of interacting genes (Viana 2000). Because the additive x additive variance is the most important epistatic variance, predicting the additive value of inbred plants over generations of selfing, including epistasis, can increase the selection efficiency (Jiang and Reif 2015).

In conclusion, our main results from a simulation-based study supported by quantitative genetics theory involving LD, epistasis, and inbreeding were: 1) the LD level for genes, even under a relatively low gene density, has a significant positive effect on the magnitude of the components of the genotypic variance in non-inbred and inbred populations; 2) assuming digenic epistasis, the additive variance is the most important component of the genotypic variance in non-inbred and inbred populations; 3) the magnitude of the epistatic variance is proportional to the percentage of interacting genes, regardless of the degree of inbreeding; 4) except for duplicate and dominant and

recessive epistasis, with 100% of epistatic genes, the additive variance is greater than the epistatic variance; and 5) regardless of the degree of inbreeding, the additive x additive variance is the most important component of the epistatic variance. This explains why LD for genes and relationship information are key factors affecting the genomic prediction accuracy of complex traits and the efficacy of GWAS.

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.

Data Availability The dataset is available at https://doi.org/10.6084/m9.figshare.13607306.v1.

Conflict of Interest The authors declare that they have no conflicts of interest.

Author Contributions Both authors contributed equally.

References

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

Chen ZQ, Baison J, Pan J, Westin J, Gil MRG, Wu HX (2019) Increased Prediction Ability in Norway Spruce Trials Using a Marker X Environment Interaction and Non-Additive Genomic Selection Model. Journal of Heredity 110:830-843

Cockerham CC (1954) An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics:859-882

Cockerham CC (1983) Covariances of relatives from self-fertilization Crop Science 23:1177-1180

Cockerham CC, Weir BS (1977) Two-locus theory in quantitative genetics. In: Pollak E, Kempthorne O, Bailey Jr TB (eds) Proceedings of the International Conference on Quantitative Genetics. Iowa State University Press, Ames, pp 247-269.

Comstock RE, Robinson HF (1948) The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics 4:254-266

Domingo J, Baeza-Centurion P, Lehner B (2019) The Causes and Consequences of Genetic Interactions (Epistasis). Annual Review of Genomics and Human Genetics, Vol 20, 2019 20:433-460

Forneris NS, Vitezica ZG, Legarra A, Perez-Enciso M (2017) Influence of epistasis on response to genomic selection using complete sequence data. Genetics Selection Evolution 49

Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R (2009) Additive genetic variability and the Bayesian alphabet. Genetics 183:347-363

Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245-257

Haberer G, Young S, Bharti AK, Gundlach H, Raymond C, Fuks G, Butler E, Wing RA, Rounsley S, Birren B, Nusbaum C, Mayer KFX, Messing J (2005) Structure and architecture of the maize genome. Plant Physiology 139:1612-1624

Hallauer AR, Miranda Filho JB (1988) Quantitative Genetics in Maize Breeding. 2nd edition. Iowa State University Press, Ames, 468 pp.

Hasselgren M, Noren K (2019) Inbreeding in natural mammal populations: historical perspectives and future challenges. Mammal Review 49:369-383

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theoretical and Applied Genetics 38:226–231

Howard JT, Pryce JE, Baes C, Maltecca C (2017) Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. Journal of dairy science 100:6009-6024

Jiang Y, Reif JC (2015) Modeling Epistasis in Genomic Selection. Genetics 201:759-+

Jiang Y, Reif JC (2020) Efficient Algorithms for Calculating Epistatic Genomic Relationship Matrices. Genetics 216:651-669

Kempthorne O (1973) An Introduction to Genetic Statistics. Iowa State University Press, Ames, 545 pp.

Kempthorne O (1954) The theoretical values of correlations between relatives in random mating populations. Genetics 40:153-167

Lewontin RC (1964) The interaction of selection and linkage. I. general considerations; heterotic models. Genetics 49:49-67

Liu H, Zhou H, Wu Y, Li X, Zhao J, Zuo T, Zhang X, Zhang Y, Liu S, Shen Y, Lin H, Zhang Z, Huang K, Luebberstedt T, Pan G (2015) The Impact of Genetic Relationship and Linkage Disequilibrium on Genomic Selection. PloS one 10

Liu XL, Huang M, Fan B, Buckler ES, Zhang ZW (2016) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. Plos Genetics 12

Martini JW, Gao N, Cardoso DF, Wimmer V, Erbe M, Cantet RJ, Simianer H (2017) Genomic prediction with epistasis models: on the marker-coding-dependent performance of the extended GBLUP and properties of the categorical epistasis model (CE). BMC Bioinformatics 18:3

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157:1819–1829

Monir MM, Zhu J (2017) Comparing GWAS Results of Complex Traits Using Full Genetic Model and Additive Models for Revealing Genetic Architecture. Scientific Reports 7

Munoz PR, Resende MFR, Jr., Gezan SA, Vilela Resende MD, de los Campos G, Kirst M, Huber D, Peter GF (2014) Unraveling Additive from Nonadditive Effects Using Genomic Relationship Matrices. Genetics 198:1759-+

Neyhart JL, Tiede T, Lorenz AJ, Smith KP (2017) Evaluating Methods of Updating Training Data in Long-Term Genomewide Selection. G3-Genes Genomes Genetics 7:1499-1510

Pecanka J, Jonker MA, Bochdanovits Z, Van der Vaart AW, Int Parkinson's Dis G (2017) A powerful and efficient two-stage method for detecting gene-to-gene interactions in GWAS. Biostatistics 18:477-494

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

Robinson MR, Wray NR, Visscher PM (2014) Explaining additional genetic variation in complex traits. Trends in Genetics 30:124-132

Stich B, Gebhardt C (2011) Detection of epistatic interactions in association mapping populations: an example from tetraploid potato. Heredity 107:537-547

Su G, Christensen OF, Ostersen T, Henryon M, Lund MS (2012) Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PloS one 7:e45293

VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. Journal of dairy science 91:4414-4423

Varona L, Legarra A, Toro MA, Vitezica ZG (2018) Non-additive Effects in Genomic Selection. Frontiers in Genetics 9

Viana JMS (2000) Components of variation of polygenic systems with digenic epistasis. Genetics and Molecular Biology 23:883-892

Viana JMS (2004) Quantitative genetics theory for non-inbred populations in linkage disequilibrium. Genetics and Molecular Biology 27:594-601

Viana JMS (2005) Dominance, epistasis, heritabilities and expected genetic gains. Genetics and Molecular Biology 28:67-74

Viana JMS, Pereira HD, Piepho HP, Silva FFE (2019) Efficiency of Genomic Prediction of Nonassessed Testcrosses. Crop Science 59:2020-2027

Viana JMS, Piepho HP, Silva FF (2016) Quantitative genetics theory for genomic selection and efficiency of breeding value prediction in open-pollinated populations. Scientia Agricola 73:243-251

Viana JMS, Piepho HP, Silva FF (2017a) Quantitative genetics theory for genomic selection and efficiency of genotypic value prediction in open-pollinated populations. Scientia Agricola 74:41-50

Viana JMS, Risso LA, Oliveira deLima R, Fonseca e Silva F (2020) Factors affecting heterotic grouping with cross□pollinating crops. Agronomy Journal

Viana JMS, Silva FF, Mundim GB, Azevedo CF, Jan HU (2017b) Efficiency of low heritability QTL mapping under high SNP density. Euphytica 213

Viana JMS, Valente MSF, Silva FF, Mundim GB, Paes GP (2013) Efficacy of population structure analysis with breeding populations and inbred lines. Genetica 141:389-399

Vitezica ZG, Legarra A, Toro MA, Varona L (2017) Orthogonal Estimates of Variances for Additive, Dominance, and Epistatic Effects in Populations. Genetics 206:1297-1307

Vitezica ZG, Reverter A, Herring W, Legarra A (2018) Dominance and epistatic genetic variances for litter size in pigs using genomic models. Genetics Selection Evolution 50

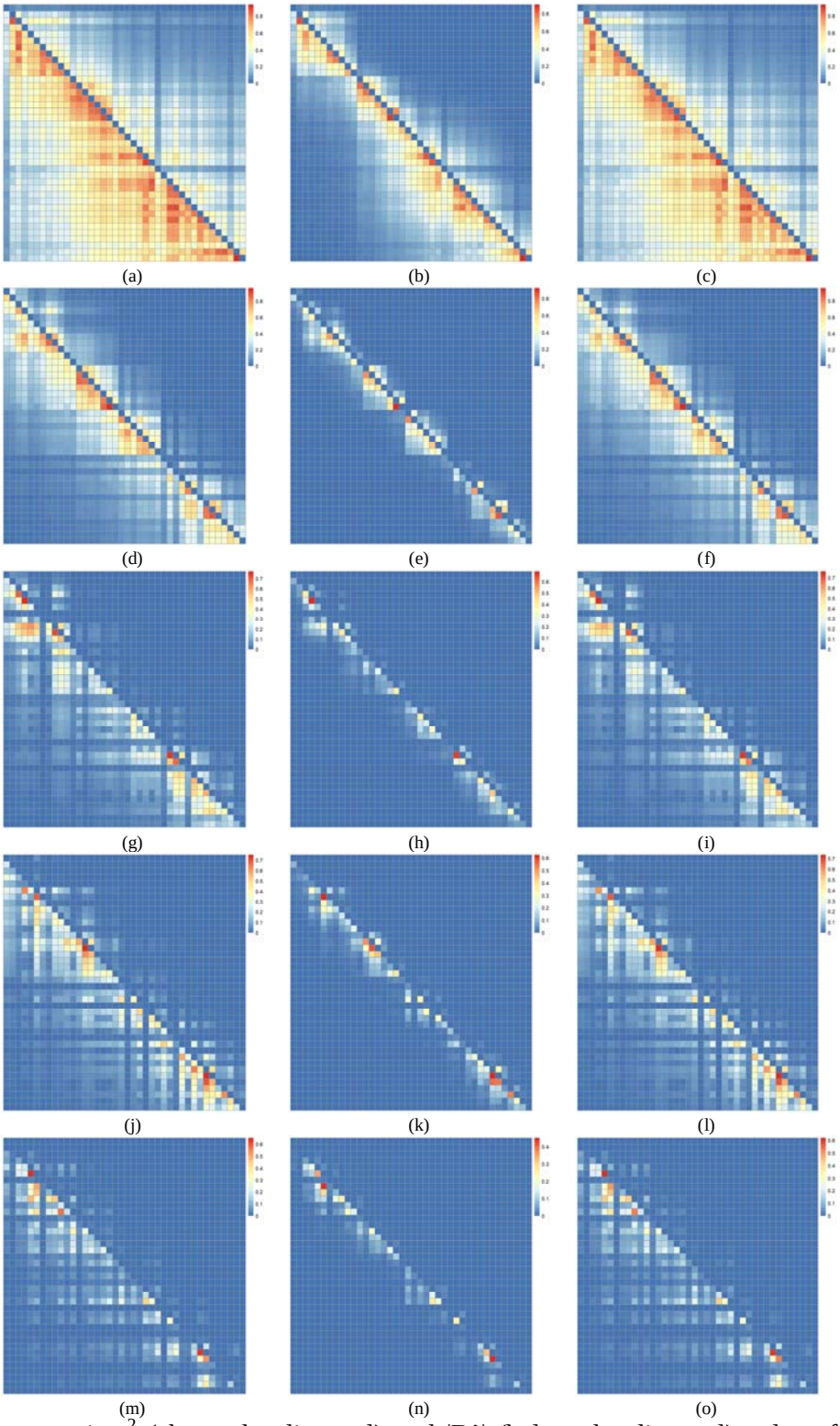


Figure 1. Parametric r^2 (above the diagonal) and |D'| (below the diagonal) values for 40 genes along chromosome 1 in the populations with higher gene density and high LD (a, b, c), and lower gene density and high (d, e, and f), intermediate (g, h, i, j, k, l), and low (m, n, and o) LD levels, in the generations 0 (a, d, g, j, and m) and 10, assuming random cross (b, e, h, k, and n) or selfing (c, f, i, l, and o).

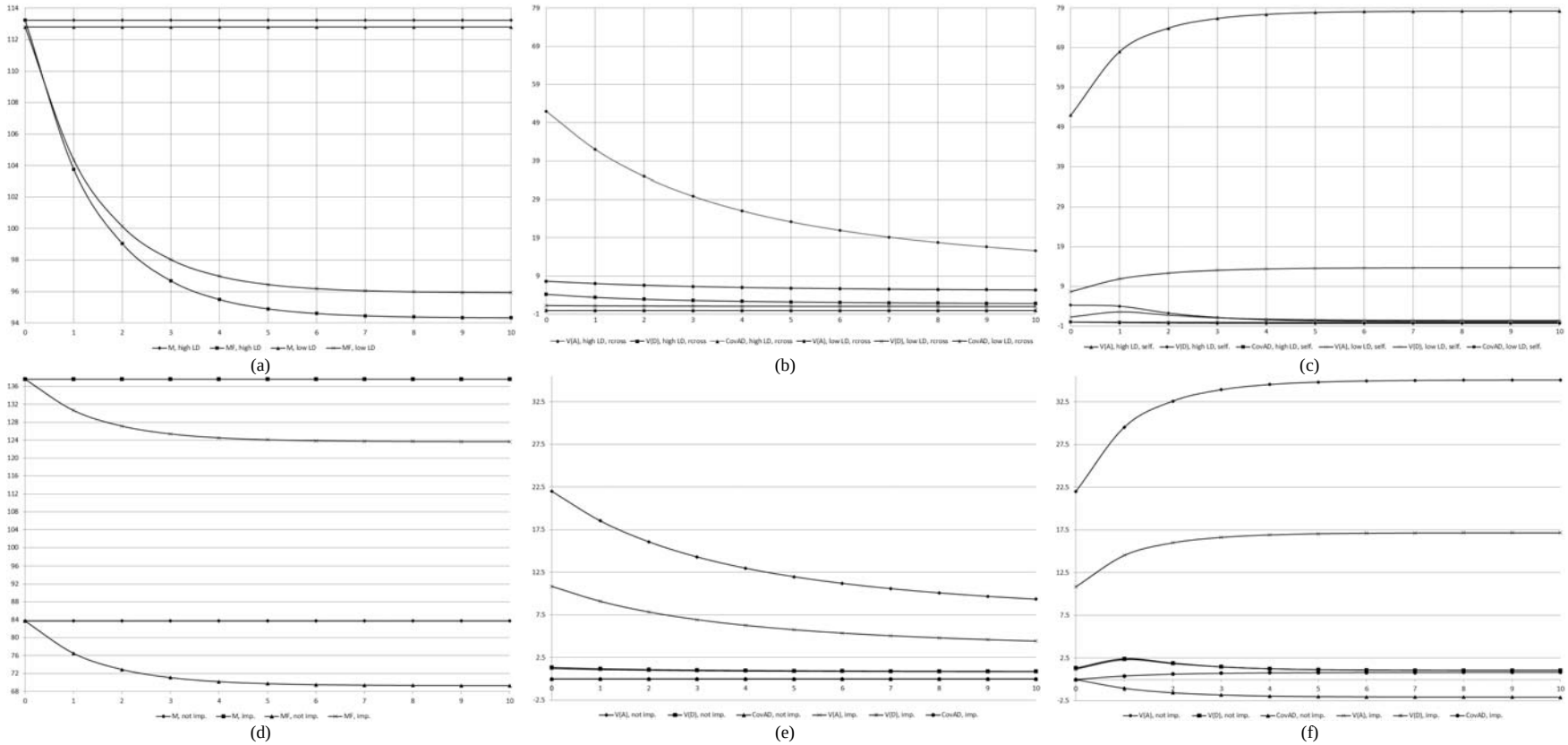
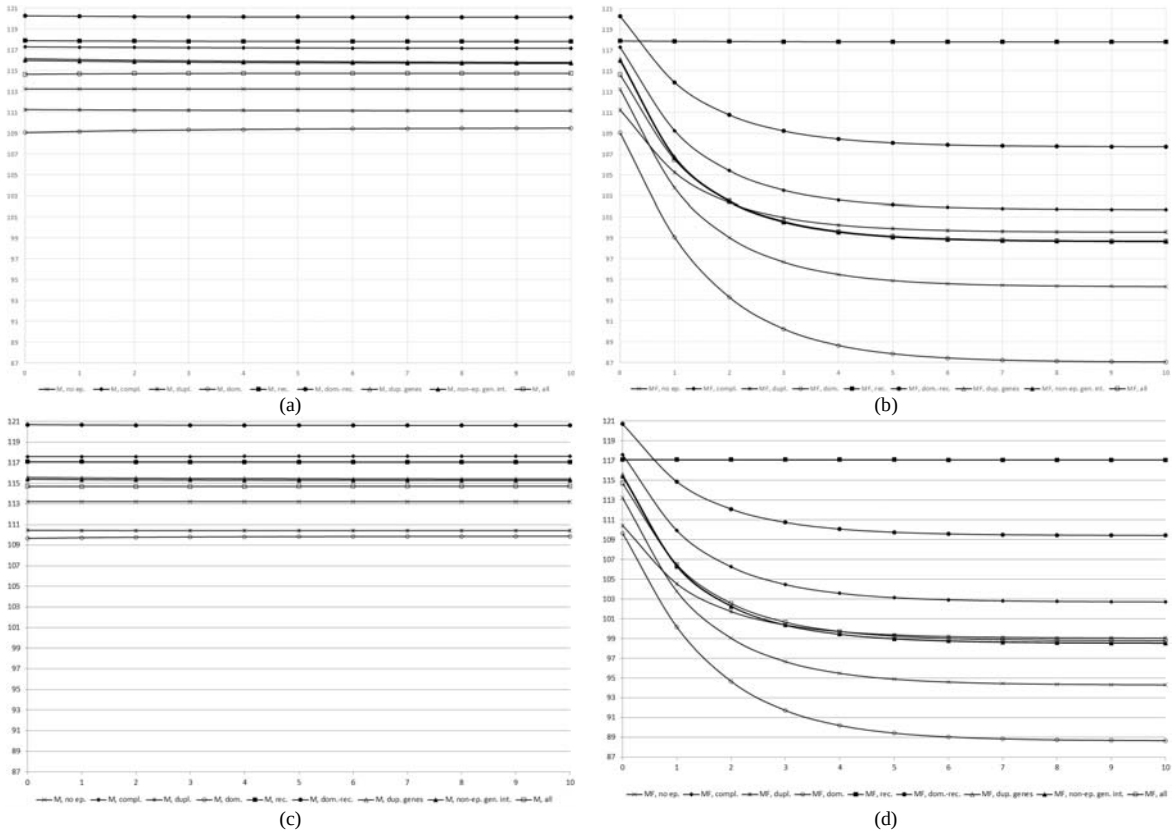


Figure 2. Parametric population means (M and MF; a and d), additive (V(A)) and dominance (V(D)) variances, and covariance between additive and dominance values (CovAD), in populations with high and low LD levels (b and c) and in populations with intermediate LD level (not improved and improved) (e and f), along 10 generations of random cross (b and e) and 10 generations of selfing (c and f).



(a) (b) (c) (d)

Figure 3. Parametric population means (M and MF) in population with high LD level, along 10 generations of random cross (a and c) and 10 generations of selfing (b and d), assuming no epistasis, a single type of digenic epistasis (complementary, duplicate, dominant, recessive, dominant and recessive, duplicate genes with cumulative effects, and non-epistatic genic interaction) or an admixture of the seven digenic types (all), and 100 (a and b) and 30% (c and d) of epistatic genes.

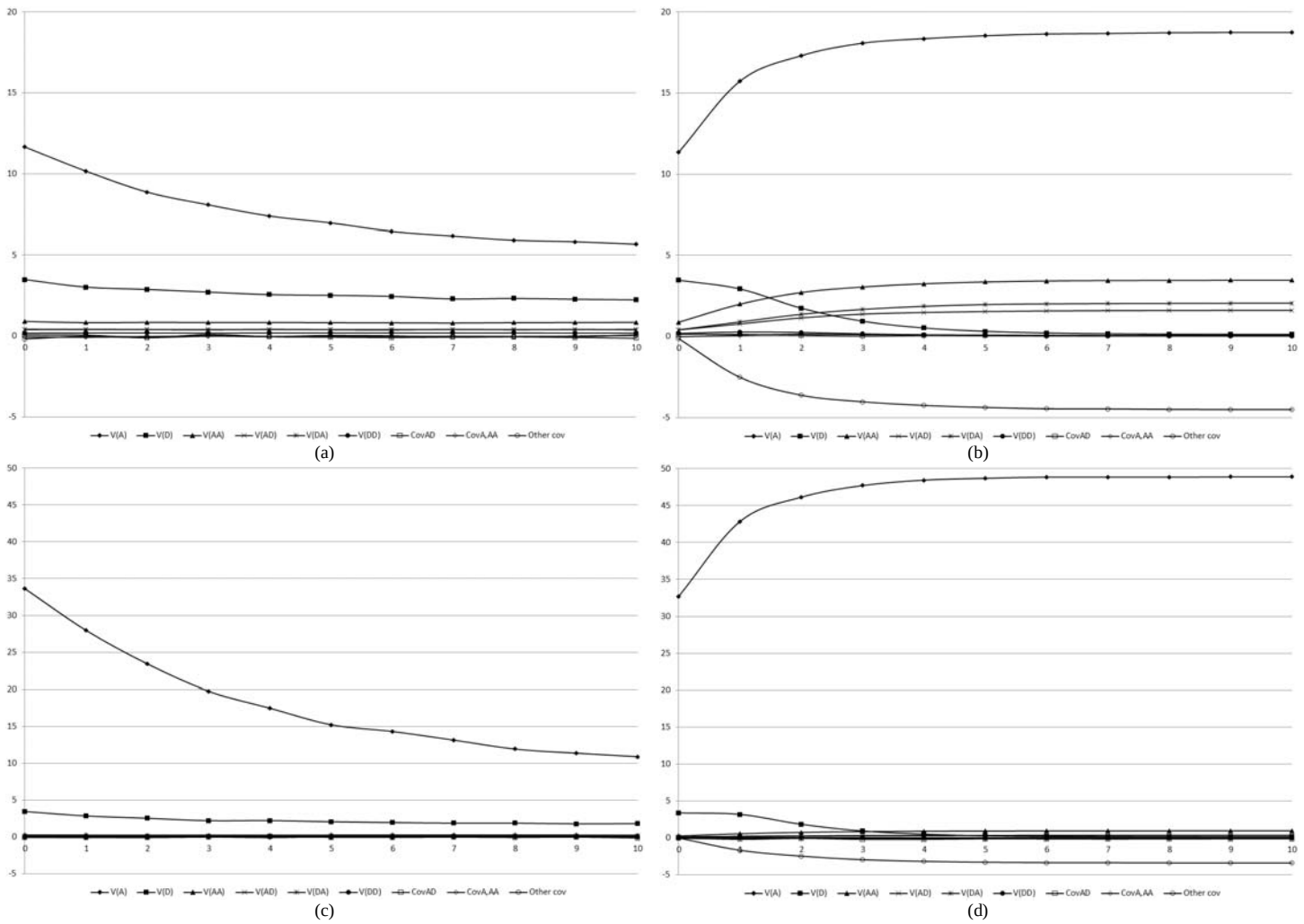


Figure 4. Components of the genotypic variance in population with high LD level, along 10 generations of random cross (a and c) or selfing (b and d), assuming complementary epistasis, 100 (a and b) and 30% (c and d) of epistatic genes, and sample size of 5,000 per generation.

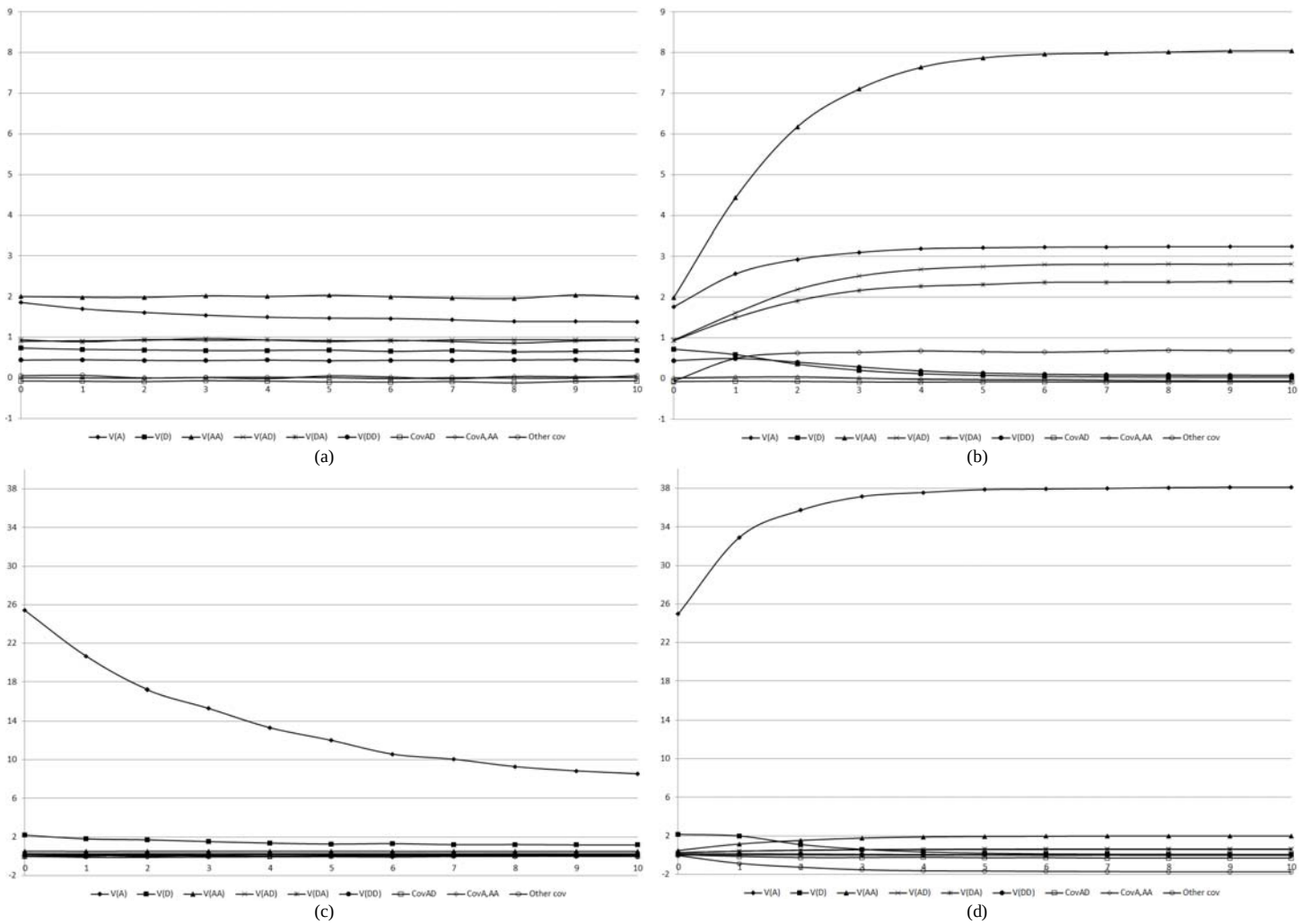


Figure 5. Components of the genotypic variance in population with high LD level, along 10 generations of random cross (a and c) or selfing (b and d), assuming duplicate epistasis, 100 (a and b) and 30% (c and d) of epistatic genes, and sample size of 5,000 per generation.

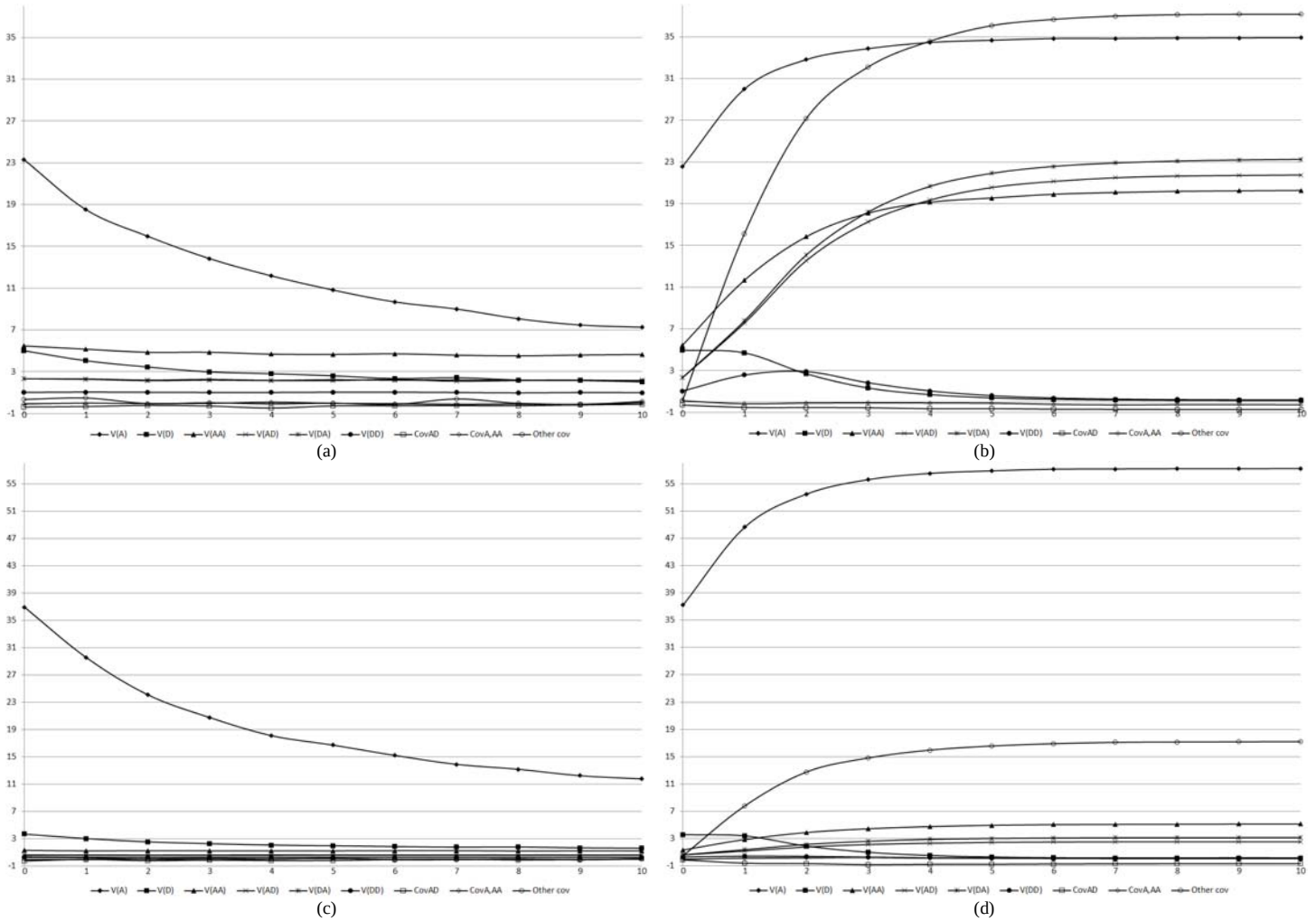


Figure 6. Components of the genotypic variance in population with high LD level, along 10 generations of random cross (a and c) or selfing (b and d), assuming dominant epistasis, 100 (a and b) and 30% (c and d) of epistatic genes, and sample size of 5,000 per generation.

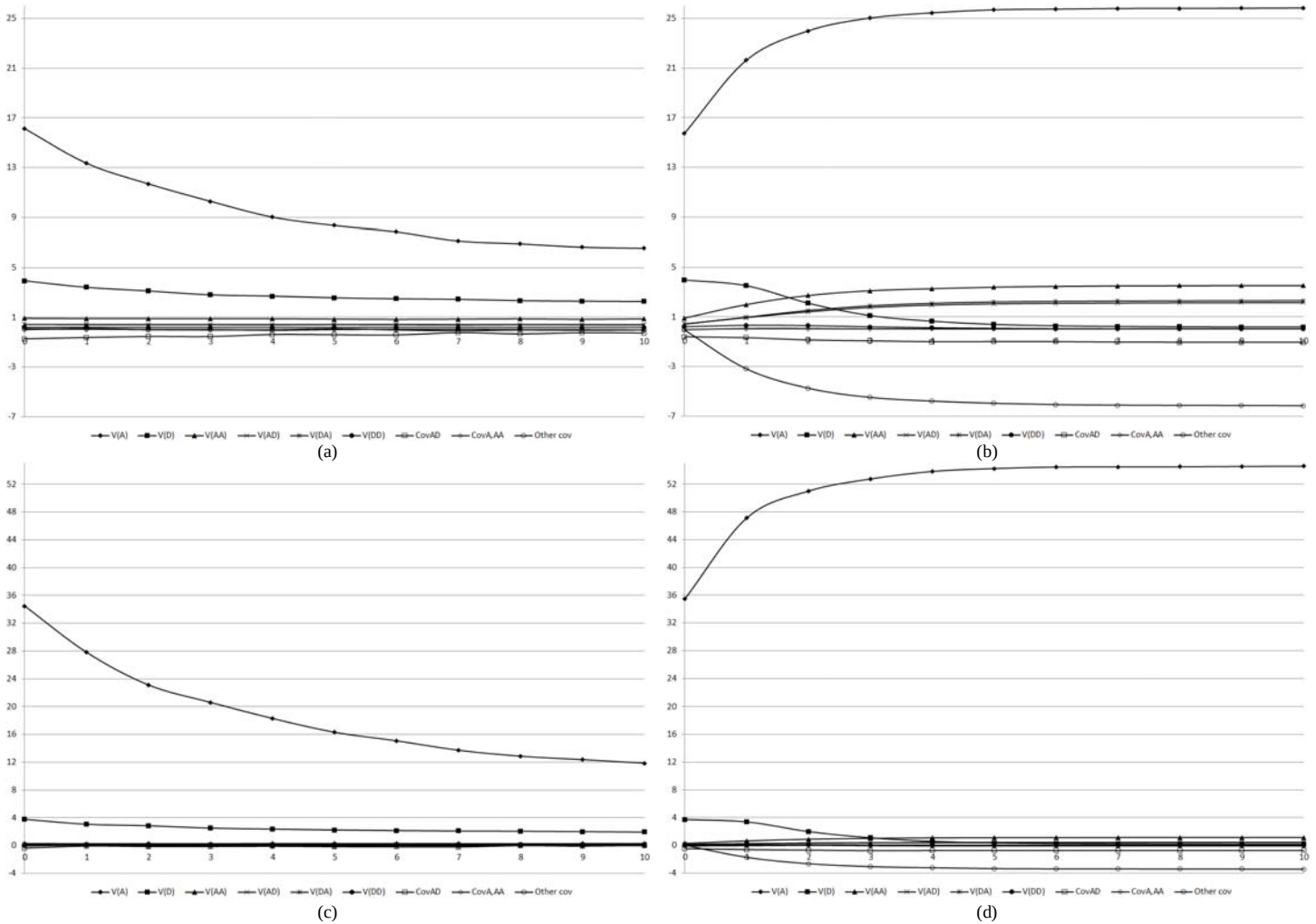


Figure 7. Components of the genotypic variance in population with high LD level, along 10 generations of random cross (a and c) or selfing (b and d), assuming recessive epistasis, 100 (a and b) and 30% (c and d) of epistatic genes, and sample size of 5,000 per generation.

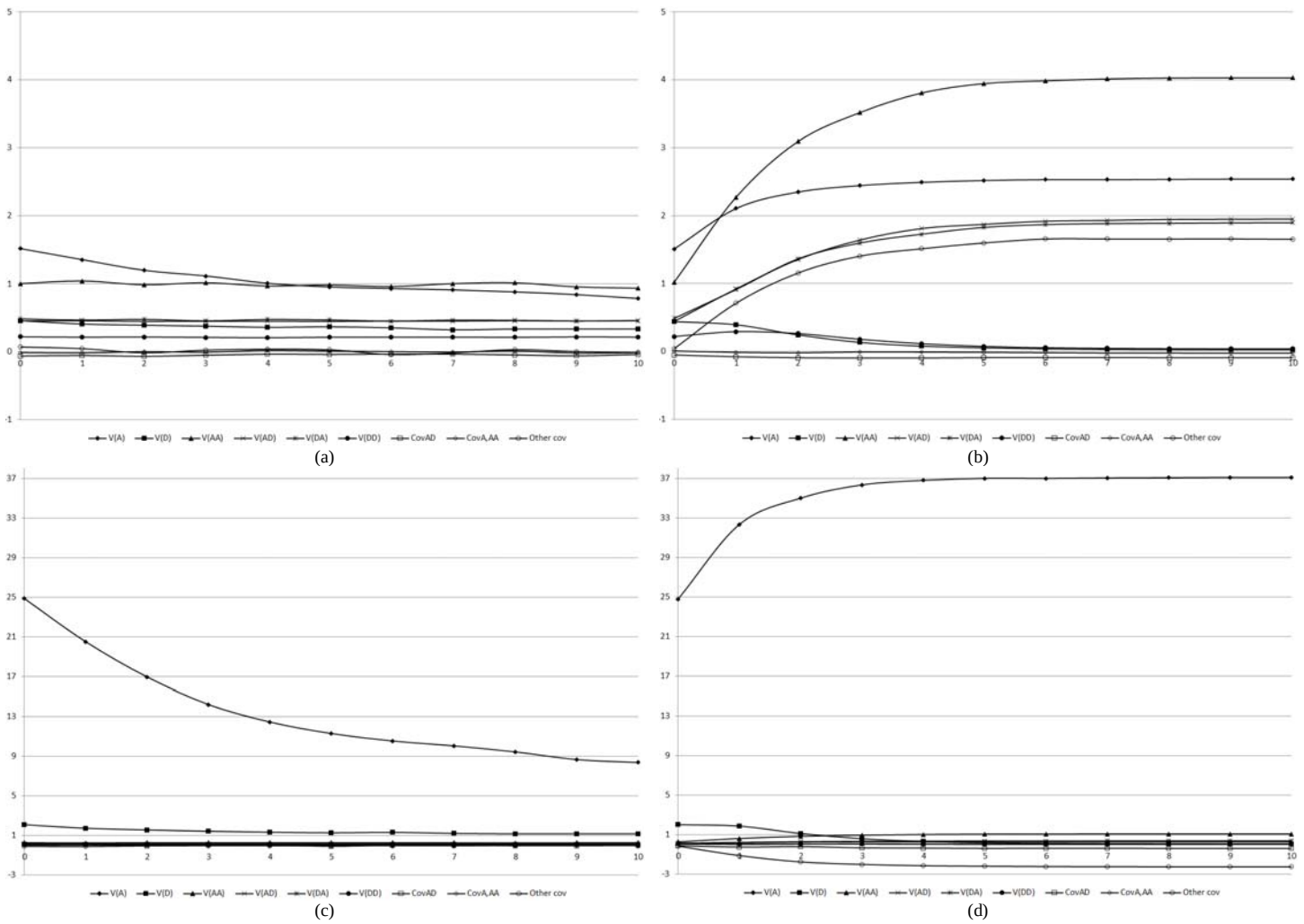


Figure 8. Components of the genotypic variance in population with high LD level, along 10 generations of random cross (a and c) or selfing (b and d), assuming dominant and recessive epistasis, 100 (a and b) and 30% (c and d) of epistatic genes, and sample size of 5,000 per generation.

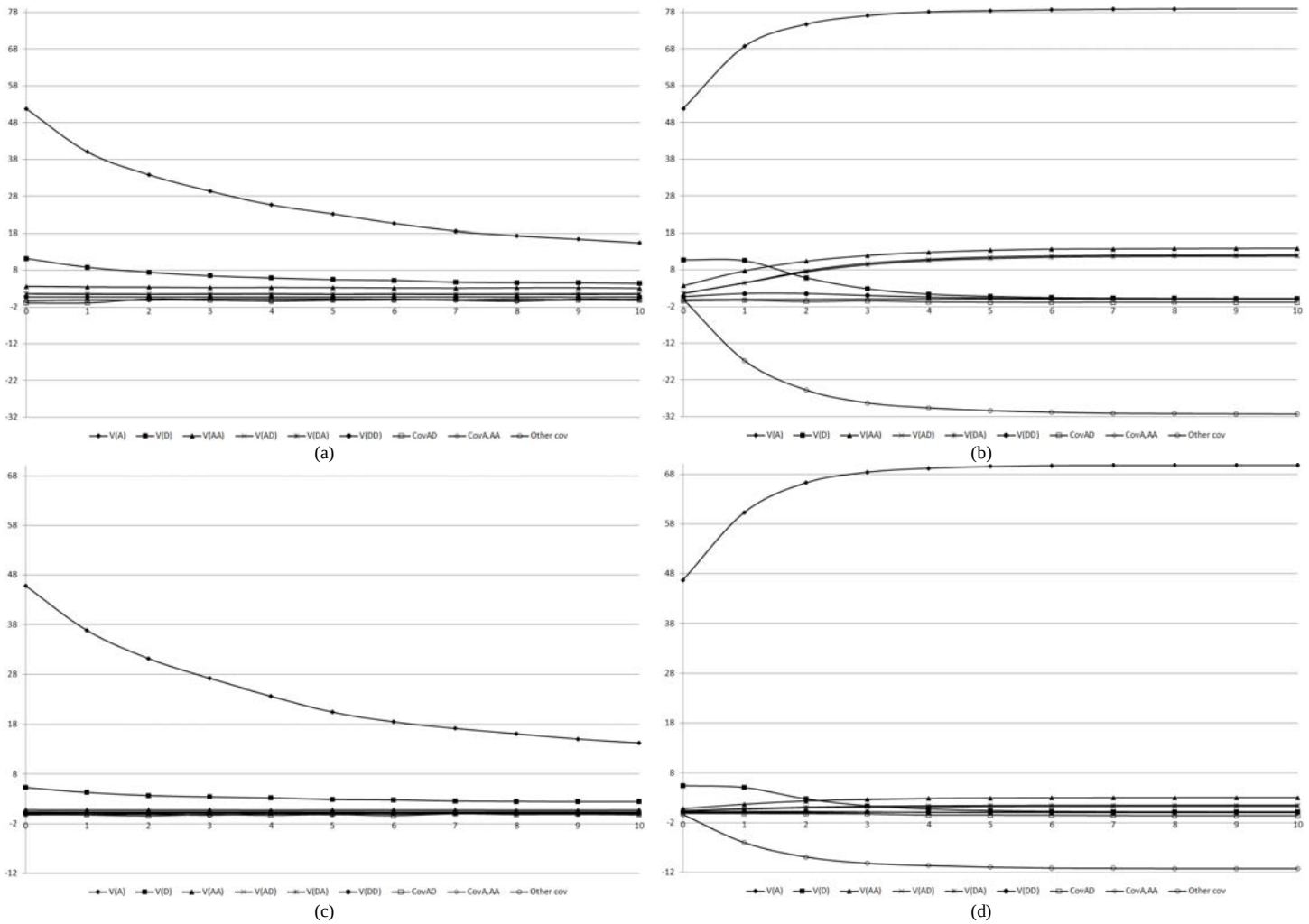


Figure 9. Components of the genotypic variance in population with high LD level, along 10 generations of random cross (a and c) or selfing (b and d), assuming duplicate genes with cumulative effects, 100 (a and b) and 30% (c and d) of epistatic genes, and sample size of 5,000 per generation.

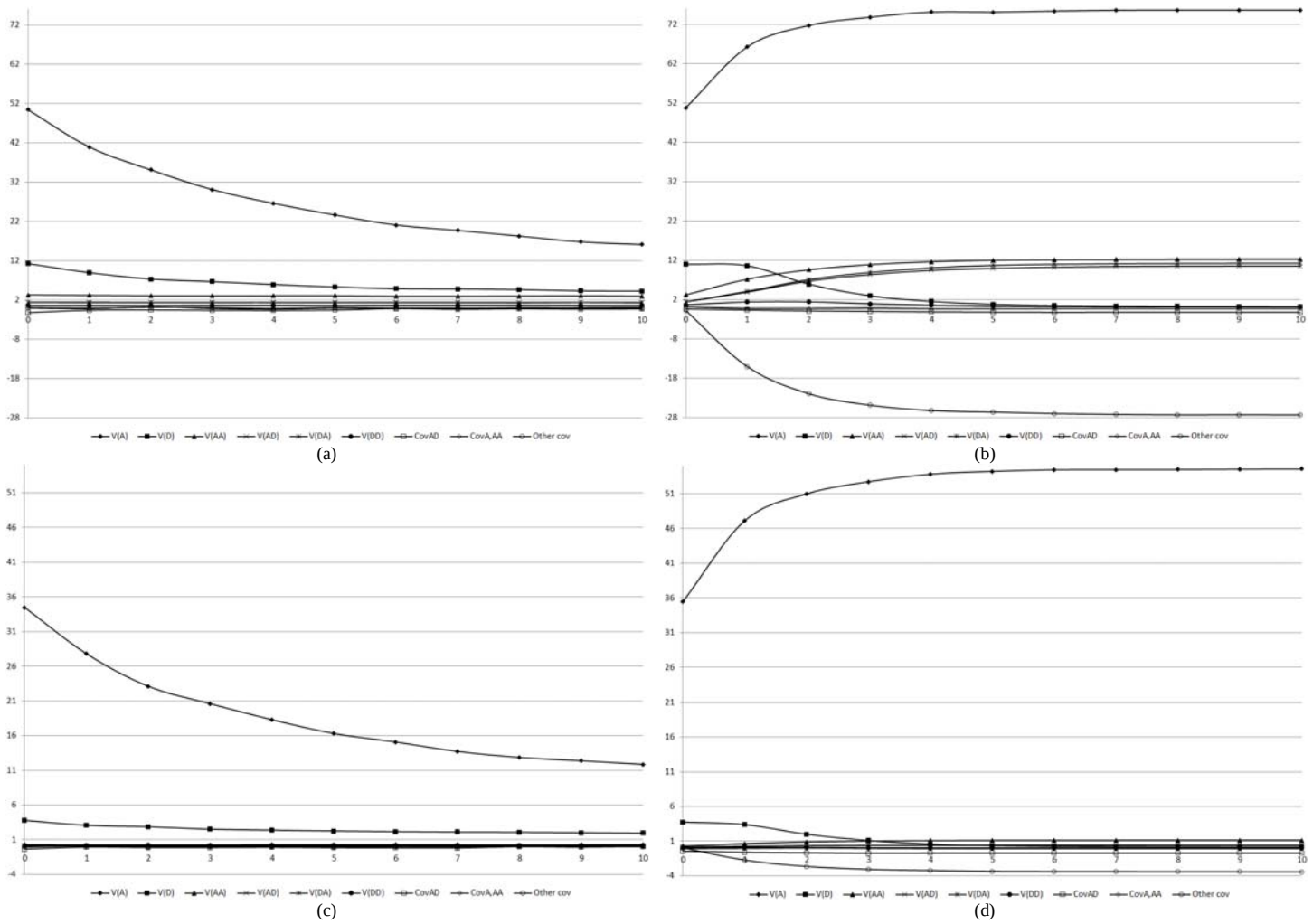


Figure 10. Components of the genotypic variance in population with high LD level, along 10 generations of random cross (a and c) or selfing (b and d), assuming non-epistatic genic interaction, 100 (a and b) and 30% (c and d) of epistatic genes, and sample size of 5,000 per generation.

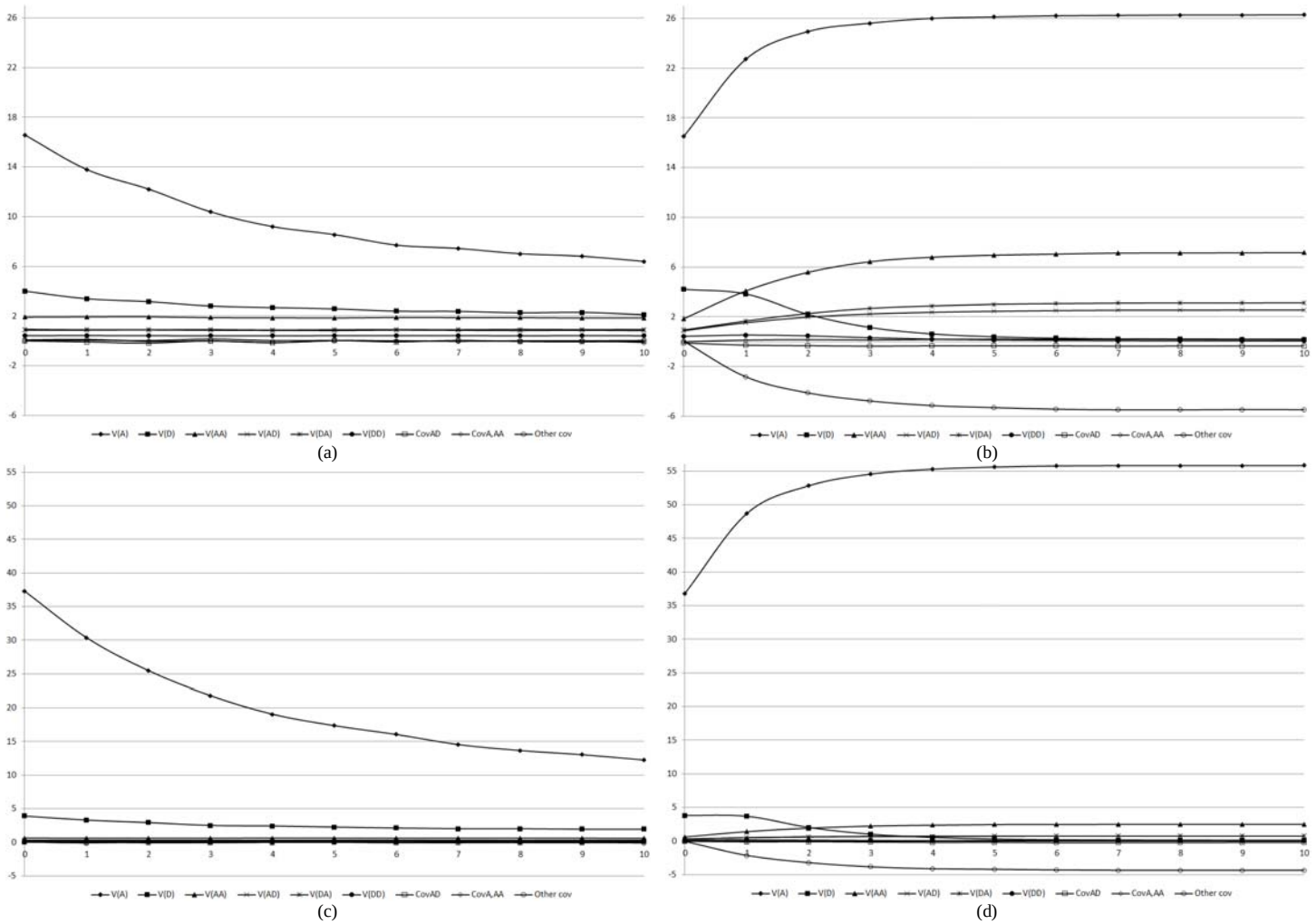


Figure 11. Components of the genotypic variance in population with high LD level, along 10 generations of random cross (a and c) or selfing (b and d), assuming an admixture of digenic epistasis, 100 (a and b) and 30% (c and d) of epistatic genes, and sample size of 5,000 per generation.

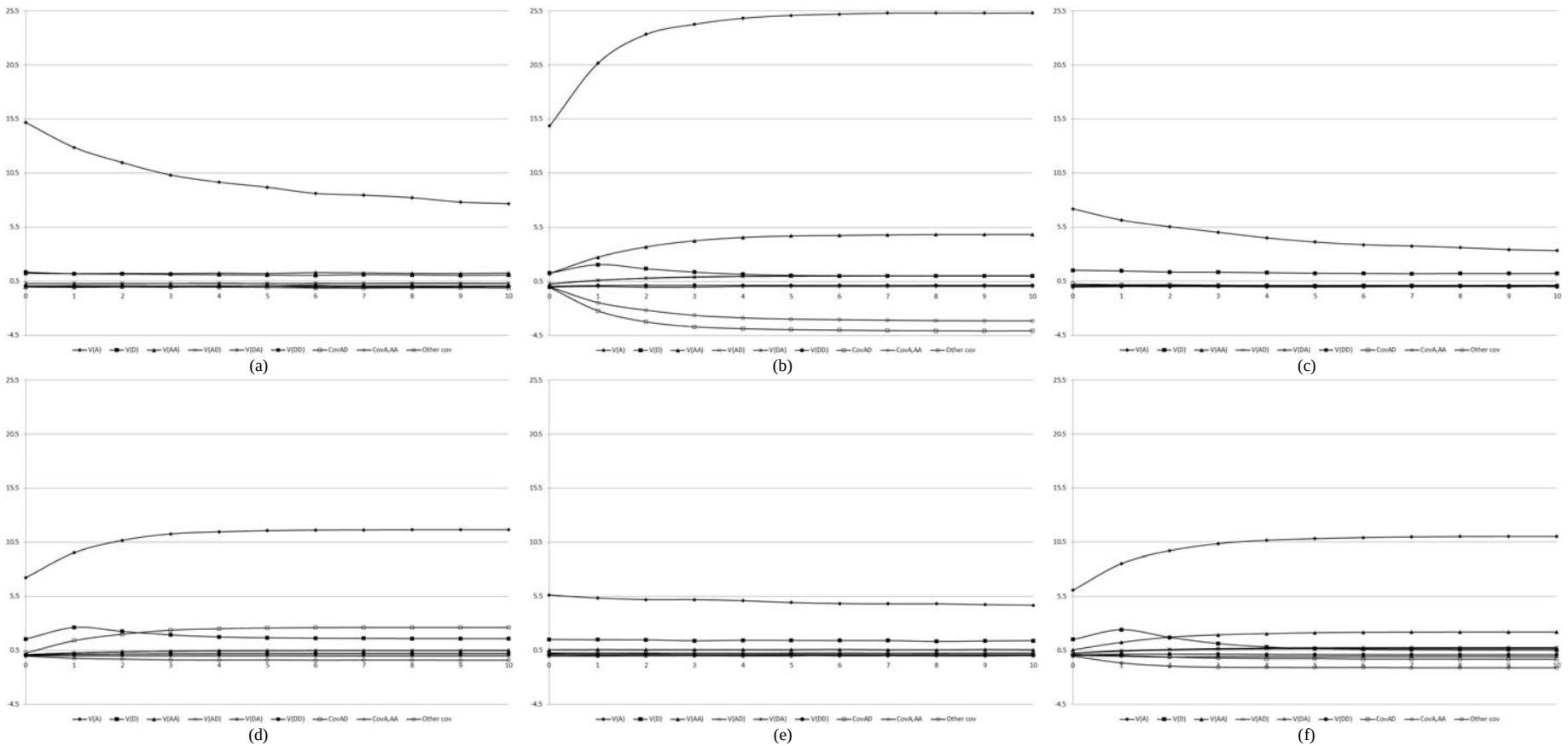


Figure 12. Components of the genotypic variance in the populations not improved (a and b) and improved (c and d), with intermediate LD level, and in the population with low LD level (e and f), along 10 generations of random cross (a, c, and e) or selfing (b, d, and f), assuming an admixture of digenic epistasis, 30% of epistatic genes, and sample size of 5,000 per generation.