

RESEARCH ARTICLE

Open Access



Vector space algebra for scaling and centering relationship matrices under non-Hardy–Weinberg equilibrium conditions

Luis Gomez-Raya^{1,2*} , Wendy M. Rauw^{1,2} and Jack C. M. Dekkers¹

Abstract

Background: Scales are linear combinations of variables with coefficients that add up to zero and have a similar meaning to “contrast” in the analysis of variance. Scales are necessary in order to incorporate genomic information into relationship matrices for genomic selection. Statistical and biological parameterizations using scales under different assumptions have been proposed to construct alternative genomic relationship matrices. Except for the natural and orthogonal interactions approach (NOIA) method, current methods to construct relationship matrices assume Hardy–Weinberg equilibrium (HWE). The objective of this paper is to apply vector algebra to center and scale relationship matrices under non-HWE conditions, including orthogonalization by the Gram-Schmidt process.

Theory and methods: Vector space algebra provides an evaluation of current orthogonality between additive and dominance vectors of additive and dominance scales for each marker. Three alternative methods to center and scale additive and dominance relationship matrices based on the Gram-Schmidt process (GSP-A, GSP-D, and GSP-N) are proposed. GSP-A removes additive-dominance co-variation by first fitting the additive and then the dominance scales. GSP-D fits scales in the opposite order. We show that GSP-A is algebraically the same as the NOIA model. GSP-N orthonormalizes the additive and dominance scales that result from GSP-A. An example with genotype information on 32,645 single nucleotide polymorphisms from 903 Large-White × Landrace crossbred pigs is used to construct existing and newly proposed additive and dominance relationship matrices.

Results: An exact test for departures from HWE showed that a majority of loci were not in HWE in crossbred pigs. All methods, except the one that assumes HWE, performed well to attain an average of diagonal elements equal to one and an average of off diagonal elements equal to zero. Variance component estimation for a recorded quantitative phenotype showed that orthogonal methods (NOIA, GSP-A, GSP-N) can adjust for the additive-dominance co-variation when estimating the additive genetic variance, whereas GSP-D does it when estimating dominance components. However, different methods to orthogonalize relationship matrices resulted in different proportions of additive and dominance components of variance.

Conclusions: Vector space methodology can be applied to measure orthogonality between vectors of additive and dominance scales and to construct alternative orthogonal models such as GSP-A, GSP-D and an orthonormal model such as GSP-N. Under non-HWE conditions, GSP-A is algebraically the same as the previously developed NOIA model.

*Correspondence: gomez.luis@inia.es

² Dept Mejora Genética Animal, INIA, Madrid, Spain

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

The first author dedicates this paper to Prof. Larry R. Schaeffer at the University of Guelph, Canada.



Background

Currently, massive single nucleotide polymorphism (SNP) genotyping of animals allows genomic prediction [1] with increased accuracy and response to selection compared to pedigree-based prediction of estimated breeding values. A commonly used technique is genomic best linear unbiased prediction (GBLUP) of breeding values, which incorporates a marker-based additive genomic relationship matrix (**G**-matrix) instead of a relationship matrix based on pedigree [2]. There are two alternative parameterizations when constructing a genomic relationship matrix: statistical and biological. The statistical or classical parameterization describes breeding values in terms of the average substitution effect of a locus at the population level [3]. The classical parameterization is widely used in breeding value estimation for farm animals because it provides information on the impact of a progenitor on the expected performance of its offspring. The alternative is the biological parameterization in which the effects of a locus are given in terms of genotypic values, which is more intuitive and practical when analyzing variability in natural populations. The distinction between these two alternatives has been previously acknowledged [4] and implemented in the construction of the **G** matrix for the statistical [2, 5] and biological [6] parameterizations. Most current applications assume that populations are in Hardy–Weinberg equilibrium (HWE). Conditions for HWE are: a very large breeding population, random mating, no change in allele frequencies due to mutation, and absence of migration and selection. The HWE conditions are usually assumed in the construction of the **G** matrix for simplicity and because departures from those conditions may not be important, in particular if dominance effects are not considered. However, some populations of commercial animals (e.g. pigs and poultry) result from crosses between distant populations, for which HWE does not apply. Recently, Vitezica et al. [7] proposed the construction of genomic relationship matrices with orthogonality between additive and dominance scales based on the NOIA (natural and orthogonal interactions approach) method of Alvarez-Castro and Carlborg [8]. Orthogonality means that additive and dominance effects are uncorrelated. The NOIA method does not require the assumption of HWE. The numerator of **G** matrices in VanRaden and NOIA are equivalent but the denominators (scaling) are different [9]. Varona et al. [10] reviewed methodology for the construction of relationship matrices that incorporate non-additive effects.

Vector spaces are mathematical objects that abstractly capture the geometry and algebra of linear equations. Some techniques of the algebra of vector spaces, such

as the Gram-Schmidt process, have not been applied to the construction of relationship matrices. The objective of this paper is to present and characterize methods to construct orthogonal additive (**G**) and dominance (**D**) genomic relationship matrices using vector space algebra (Gram-Schmidt process) without requiring HWE. A comparison of the newly proposed methods to construct **G** and **D** matrices to existing methods is carried out using a dataset of Large-White × Landrace crossbred pigs.

Theory and methods

Scaling and centering the genomic relationship matrix

In the statistical parameterization, breeding values are modeled using the average of allele substitution effects at genotyped loci, rather than genotypic values. This parameterization requires the assumption of HWE. In 1941, Fisher already pointed out that computation of breeding values must assume random mating [11]. Later and along the same line, Falconer stated: “The concept of breeding value is shown to have no useful meaning when mating is not random” [12]. Therefore, the statistical parameterization will not be considered further and only the biological parameterization will be addressed in the next sections.

The statistical model for genomic prediction with additive and dominance effects in the biological parameterization [6, 13] is:

$$y_i = \mu + \sum_{j=1}^m z_{ij}a_j + \sum_{j=1}^m s_{ij}d_j + e_i, \quad (1)$$

where y_i is the phenotypic record of the i -th individual; μ is the population mean; m is the number of markers; a_j and d_j are the additive and dominance effects of the j -th marker; z_{ij} is 1, 0, and -1 for the i -th individual with genotype at the j -th marker AA , Aa and aa ; s_{ij} is 0, 1, and 0 for the i -th individual with genotype at the j -th marker AA , Aa and aa ; respectively; and e_i is the error. In matrix notation, Model (1) is:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_a\mathbf{a} + \mathbf{S}_d\mathbf{d} + \mathbf{e},$$

where \mathbf{y} is a vector of phenotypes, μ is the mean; \mathbf{Z}_a and \mathbf{S}_d are matrices with n rows (number of individuals) and m columns (number of markers) with values as defined above for z_{ij} and s_{ij} relating information on each individual genotype with additive and dominance effects; \mathbf{a} and \mathbf{d} are vectors of additive and dominance effects, respectively; and \mathbf{e} is a vector of errors. The additive and dominance covariance matrices associated with the random additive and dominance effects in this model are:

$$\text{Cov}(\mathbf{a}) = \frac{\mathbf{H}_a \mathbf{H}_a'}{SC_a} \sigma_a^2 = \mathbf{G} \sigma_a^2,$$

$$\text{Cov}(\mathbf{d}) = \frac{\mathbf{H}_d \mathbf{H}_d'}{SC_d} \sigma_d^2 = \mathbf{D} \sigma_d^2,$$

where \mathbf{G} and \mathbf{D} are additive and dominance genomic relationship matrices; σ_a^2 and σ_d^2 are the additive and dominance variance, respectively; \mathbf{H}_a and \mathbf{H}_d are matrices with n rows and m columns for additive and dominance scales. These scales are used to center all markers so the mean contributed by each marker is zero. SC_a and SC_d are used for scaling additive and dominance relationship matrices. Different assumptions and methods to center and scale relationships lead to alternative \mathbf{G} and \mathbf{D} matrices. First, we present several existing models to construct additive and dominance relationships matrices. Second, we propose methods to test for the lack of orthogonality between the additive and dominance scales. Third, we show a general method from vector space algebra, the Gram-Schmidt process, to generate alternative orthogonal models. Last, we present the results obtained by applying these methods to a dataset consisting of cross-bred animals.

Hardy-Weinberg equilibrium parameterization

Assuming Model (1), the population is segregating for the three genotypes $AA_{(j)}$, $Aa_{(j)}$ and $aa_{(j)}$ at the j -th locus, with alleles $A_{(j)}$ and $a_{(j)}$ and corresponding frequencies $p_{(j)}$ and $q_{(j)}$. The genotypic values under an additive-dominance model of the three genotypes are $a_{(j)}$, $d_{(j)}$ and $-a_{(j)}$, respectively. VanRaden [2] proposed centering of additive marker effects by subtracting the mean additive effect, resulting in the additive genetic values: $(2 - 2p_{(j)})a_{(j)}$, $(1 - 2p_{(j)})a_{(j)}$, $(-2p_{(j)})a_{(j)}$, for individuals with genotypes $AA_{(j)}$, $Aa_{(j)}$ and $aa_{(j)}$, respectively. Therefore, the scales at the column corresponding to the j -th marker in \mathbf{H}_a are:

$$\begin{aligned} u_{AA_{(j)}}^{hw} &= (2 - 2p_{(j)}), u_{Aa_{(j)}}^{hw} = (1 - 2p_{(j)}), \text{ and} \\ u_{aa_{(j)}}^{hw} &= (-2p_{(j)}) \end{aligned} \tag{2}$$

Assuming HWE, Su et al. [6] and Vitezica et al. [5] proposed centering of dominance marker effects by subtracting the mean dominance effect, resulting in $(-2p_{(j)}q_{(j)})d_{(j)}$, $(1 - 2p_{(j)}q_{(j)})d_{(j)}$, and $(-2p_{(j)}q_{(j)})d_{(j)}$ for individuals with genotypes $AA_{(j)}$, $Aa_{(j)}$ and $aa_{(j)}$, respectively. Therefore, the scales at the column corresponding to the j -th marker in \mathbf{H}_d are:

$$\begin{aligned} v_{AA_{(j)}}^{hw} &= (-2p_{(j)}q_{(j)}), v_{Aa_{(j)}}^{hw} = (1 - 2p_{(j)}q_{(j)}), \text{ and} \\ v_{aa_{(j)}}^{hw} &= (-2p_{(j)}q_{(j)}) \end{aligned} \tag{3}$$

Assuming HWE, VanRaden [2] scaled the additive relationship matrix constructed with scales of Eq. (2) by $\sum_{j=1}^m (2p_{(j)}q_{(j)})$, while Su et al. [6], and Vitezica et al. [5] proposed scaling the dominance relationship matrix under the assumption of HWE and constructed with scales in Eq. (3) by $\sum_{j=1}^m (2p_{(j)}q_{(j)})(1 - 2p_{(j)}q_{(j)})$. This approach has two problems. The first one is that it results in functional rather than statistical values associated with locus genotypes, which can be used to derive genotypic values of the individual but not breeding values that predict performance of progeny, as discussed before. Second, centering of the dominance incidence matrix is not necessarily achieved when the population is not in HWE. For example, the mean for the j -th marker for the centered dominance matrix is equal to $p_{AA_{(j)}}(-2p_{(j)}q_{(j)}) + p_{Aa_{(j)}}(1 - 2p_{(j)}q_{(j)}) + p_{aa_{(j)}}(-2p_{(j)}q_{(j)})$, where $p_{AA_{(j)}}$, $p_{Aa_{(j)}}$, and $p_{aa_{(j)}}$ are the frequencies of AA , Aa and aa genotypes at the j -th marker, respectively. It reduces to $p_{Aa_{(j)}} - 2p_{(j)}q_{(j)}$ and is not equal to 0 if the population is not in HWE. For example, an F_1 cross may have an excess of heterozygotes compared to HWE frequencies.

Non-Hardy-Weinberg equilibrium parameterization

Here, we propose a non-HWE parameterization using functional marker effects based on the mean and the variance of additive and dominance genotypic values. Thus, multiplying genotypic frequencies by their corresponding values and summing, the additive mean for the j -th marker is:

$$\mu_{A_{(j)}} = a_{(j)}p_{AA_{(j)}} - a_{(j)}p_{aa_{(j)}}.$$

Similarly, the variance of the additive genotypic effects is obtained by multiplying genotypic frequencies by the square of their corresponding values, summing, and subtracting the square of the mean:

$$V_{A_{(j)}} = a_j^2 p_{AA_{(j)}} + a_j^2 p_{aa_{(j)}} - \mu_{A_{(j)}}^2.$$

After some algebra, the additive variance becomes $V_{A_{(j)}} = a_j^2 [p_{AA_{(j)}} + p_{aa_{(j)}} - (p_{AA_{(j)}} - p_{aa_{(j)}})^2]$. The additive variance contributed by the m markers is then:

$$V_A = \sum_{j=1}^m \left[p_{AA_{(j)}} + p_{aa_{(j)}} - (p_{AA_{(j)}} - p_{aa_{(j)}})^2 \right] a_j^2 \tag{4}$$

When subtracting the additive mean from the genotypic values, the centered genotypic values for individuals with genotypes AA , Aa and aa at the j -th marker in absence of HWE are $\left(1 - (p_{AA(j)} - p_{aa(j)})\right)a_{(j)}$, $\left(- (p_{AA(j)} - p_{aa(j)})\right)a_{(j)}$, and $\left(-1 - (p_{AA(j)} - p_{aa(j)})\right)a_{(j)}$. Therefore, in the absence of HWE, the scales at the column corresponding to the j -th marker in \mathbf{H}_a are:

$$\begin{aligned} u_{AA(j)}^{NO-HW} &= 1 - (p_{AA(j)} - p_{aa(j)}), \\ u_{Aa(j)}^{NO-HW} &= - (p_{AA(j)} - p_{aa(j)}), \text{ and} \\ u_{aa(j)}^{NO-HW} &= -1 - (p_{AA(j)} - p_{aa(j)}) \end{aligned} \tag{5}$$

Note that these centered scales reduce to those of VanRaden [2] because $p_{(j)} = p_{AA(j)} + \frac{1}{2}p_{Aa(j)}$.

However, the scaling of the additive relationships can be based on Eq. (4), which is different from the classical scaling of matrix \mathbf{G} in, e.g., VanRaden [2], which is $\sum_{j=1}^m (2p_{(j)}q_{(j)})$.

The dominance genotypic effects for the j -th marker without the assumption of HWE are as first given by Alvarez-Castro and Carlborg [8]: $-p_{Aa(j)}d_{(j)}$, $(1 - p_{Aa(j)})d_{(j)}$, and $-p_{Aa(j)}d_{(j)}$ for individuals with genotypes AA , Aa and aa , respectively. In the absence of HWE, the scales at the column corresponding to the j -th marker in \mathbf{H}_d are:

$$\begin{aligned} v_{AA(j)}^{NO-hw} &= -p_{Aa(j)}, v_{Aa(j)}^{NO-hw} = (1 - p_{Aa(j)}), \text{ and} \\ v_{aa(j)}^{NO-hw} &= -p_{Aa(j)} \end{aligned} \tag{6}$$

The variance contributed by the j -th marker is:

$$\begin{aligned} V_{D(j)} &= \left[p_{AA(j)}p_{Aa(j)}^2 + p_{Aa(j)}(1 - p_{Aa(j)})^2 + p_{aa(j)}p_{Aa(j)}^2 \right] d_{(j)}^2 \\ &= \left[p_{Aa(j)} - p_{Aa(j)}^2 \right] d_{(j)}^2. \end{aligned}$$

Summing over loci, $V_D = \sum_{j=1}^m \left[p_{Aa(j)} - p_{Aa(j)}^2 \right] d_{(j)}^2$, which renders the scaling of the dominance relationship matrix with $\sum_{j=1}^m \left[p_{Aa(j)} - p_{Aa(j)}^2 \right]$. Under HWE, the variance of the dominance effects reduces to $V_{D(j)} = \left[2p_{(j)}q_{(j)}(1 - 2p_{(j)}q_{(j)}) \right] d_{(j)}^2$, which is as given by Su et al. [6]. The resulting additive and dominance scales might be correlated (non-orthogonal), which makes estimation and interpretation of the results more difficult.

Orthogonal parameterization based on the NOIA method

Based on Cockerham [14], an orthogonal decomposition of additive and dominance variance components was proposed by Alvarez-Castro and Carlborg [8] and termed the natural and orthogonal interactions approach (NOIA). The requirements for the orthogonal partition of variance are:

$$\sum_{j=1}^m p_{AA(j)}u_{AA(j)}^{NOIA} + p_{Aa(j)}u_{Aa(j)}^{NOIA} + p_{aa(j)}u_{aa(j)}^{NOIA} = 0, \tag{7}$$

$$\sum_{j=1}^m p_{AA(j)}v_{AA(j)}^{NOIA} + p_{Aa(j)}v_{Aa(j)}^{NOIA} + p_{aa(j)}v_{aa(j)}^{NOIA} = 0, \tag{8}$$

$$\begin{aligned} \sum_{j=1}^m p_{AA(j)}u_{AA(j)}^{NOIA}v_{AA(j)}^{NOIA} + p_{Aa(j)}u_{Aa(j)}^{NOIA}v_{Aa(j)}^{NOIA} \\ + p_{aa(j)}u_{aa(j)}^{NOIA}v_{aa(j)}^{NOIA} = 0, \end{aligned} \tag{9}$$

where $u_{g(j)}^{NOIA}$ and $v_{g(j)}^{NOIA}$ are the additive and dominance scales for the g -th genotype ($g = AA, Aa, aa$) at the j -th marker. Requirements (7) and (8) force a comparison of deviations around the mean for additive and dominance scales. Requirement (9) forces the additive and dominance scales to be uncorrelated (orthogonal).

Vitezica et al. [7] implemented the orthogonalization of NOIA to construct additive and dominance relationships. Thus, the proposed orthogonal scales of Alvarez-Castro and Carlborg [8] and Vitezica et al. [7] for individuals with genotypes AA , Aa , and aa are:

$$\begin{aligned} u_{AA(j)}^{NOIA} &= -(-p_{Aa(j)} - 2p_{aa(j)}), \\ u_{Aa(j)}^{NOIA} &= -(1 - p_{Aa(j)} - 2p_{aa(j)}), \\ u_{aa(j)}^{NOIA} &= -(2 - p_{Aa(j)} - 2p_{aa(j)}) \end{aligned} \tag{10}$$

After some algebra, scales for the additive component in the absence of HWE become the same as for the NOIA method:

$$\begin{aligned} u_{AA(j)}^{NO-HW} &= 1 - (p_{AA(j)} - p_{aa(j)}) = -(-p_{Aa(j)} - 2p_{aa(j)}), \\ u_{Aa(j)}^{NO-HW} &= -(p_{AA(j)} - p_{aa(j)}) = -(1 - p_{Aa(j)} - 2p_{aa(j)}), \\ u_{aa(j)}^{NO-HW} &= -1 - (p_{AA(j)} - p_{aa(j)}) \\ &= -(2 - p_{Aa(j)} - 2p_{aa(j)}). \end{aligned}$$

They also reduce to the VanRaden scales, as shown by Joshi et al. [9].

The dominance scales in the NOIA method are:

$$\begin{aligned}
 v_{AA(j)}^{NOIA} &= \frac{-\left(2p_{Aa(j)}p_{aa(j)}\right)}{\left(p_{AA(j)} + p_{aa(j)}\right) - \left(p_{AA(j)} - p_{aa(j)}\right)^2}, \\
 v_{Aa(j)}^{NOIA} &= \frac{\left(4p_{AA(j)}p_{aa(j)}\right)}{\left(p_{AA(j)} + p_{aa(j)}\right) - \left(p_{AA(j)} - p_{aa(j)}\right)^2}, \\
 v_{aa(j)}^{NOIA} &= \frac{-\left(2p_{AA(j)}p_{Aa(j)}\right)}{\left(p_{AA(j)} + p_{aa(j)}\right) - \left(p_{AA(j)} - p_{aa(j)}\right)^2}
 \end{aligned}
 \tag{11}$$

Equations (10) and (11) satisfy conditions of orthogonality of Eqs. (7), (8), and (9). Vitezica et al. [7] implemented orthogonalization of the NOIA approach by scaling the **G** and **D** matrices by $tr(\mathbf{H}_a\mathbf{H}_a')/n$ and $tr(\mathbf{H}_d\mathbf{H}_d')/n$, respectively, where **H_a** and **H_d** include the scales for individuals according to their genotypes in Eqs. (10) and (11), respectively. Therefore, after scaling, genomic and dominance relationship matrices become:

$$\mathbf{G} = \frac{\mathbf{H}_a\mathbf{H}_a'}{tr(\mathbf{H}_a\mathbf{H}_a')/n},
 \tag{12}$$

$$\mathbf{D} = \frac{\mathbf{H}_d\mathbf{H}_d'}{tr(\mathbf{H}_d\mathbf{H}_d')/n}.
 \tag{13}$$

Vector space algebra for orthogonal parameterization using the Gram-Schmidt process

We propose to use algebra of vector spaces to construct genomic and dominance relationship matrices. The Gram-Schmidt process takes several non-orthogonal linearly independent functions to construct an orthogonal basis over an arbitrary weighting function [15]. First, we will use vector space algebra to measure orthogonality between the additive and dominance scales.

Vectors space algebra to measure orthogonality

Define vector spaces for additive (\vec{u}_j) and dominance (\vec{v}_j) for the *j*-th marker with dimensions equal to the number of individuals. The elements of \vec{u}_j and \vec{v}_j are the scales to

center the additive and dominance relationship matrices, respectively. Then, **H_a** and **H_d** can be constructed as:

$$\mathbf{H}_a = [\vec{u}_1\vec{u}_2 \dots \vec{u}_j\vec{u}_{j+1} \dots \vec{u}_m],$$

$$\mathbf{H}_d = [\vec{v}_1\vec{v}_2 \dots \vec{v}_j\vec{v}_{j+1} \dots \vec{v}_m].$$

The vector spaces for the additive scale, \vec{u}_j , at the *j*-th marker for individuals with genotypes *AA*, *Aa* and *aa* under non-HWE conditions based on Eq. (5) are:

$$u_{AA(j)} = 1 - \left(p_{AA(j)} - p_{aa(j)}\right)$$

$$u_{Aa(j)} = -\left(p_{AA(j)} - p_{aa(j)}\right),$$

$$u_{aa(j)} = -1 - \left(p_{AA(j)} - p_{aa(j)}\right).$$

The elements of \vec{v}_j are dominance scales for individuals with genotypes *AA*, *Aa* and *aa* under non-HWE conditions from Eq. (6):

$$v_{AA(j)} = -p_{Aa(j)},$$

$$v_{Aa(j)} = \left(1 - p_{Aa(j)}\right),$$

$$v_{aa(j)} = -p_{Aa(j)}.$$

For a given marker, the set of vectors, \vec{u}_j and \vec{v}_j , form a basis since both vectors span the vector space and the vectors in the set are linearly independent. However, the set of vectors in this basis is not necessarily orthogonal. In this setting, Alvarez-Castro and Carlborg [8] showed that orthogonality is only achieved when either the two homozygotes are at the same frequency or there is not any heterozygote. The angle, θ_j , between the two vectors, \vec{u}_j and \vec{v}_j , provides a measure of the degree of orthogonality. From the definition of inner product (p 519 in [15]), $\cos \theta_j$ for the *j*-th marker is given by:

$$\cos \theta_j = \frac{\langle \vec{u}_j, \vec{v}_j \rangle}{\|\vec{u}_j\| \|\vec{v}_j\|},$$

where $\langle \vec{u}_j, \vec{v}_j \rangle$ is the inner product of the vectors \vec{u}_j and \vec{v}_j and has an expected value equal to:

$$\begin{aligned} \langle \vec{u}_j, \vec{v}_j \rangle &= n \left[p_{AA(j)} \left[1 - (p_{AA(j)} - p_{aa(j)}) \right] (-p_{Aa(j)}) - p_{Aa(j)} \left[p_{AA(j)} - p_{aa(j)} \right] (1 - p_{Aa(j)}) \right. \\ &\quad \left. + p_{aa(j)} \left[-1 - (p_{AA(j)} - p_{aa(j)}) \right] (-p_{Aa(j)}) \right] \\ &= n (-p_{Aa(j)}) (p_{AA(j)} - p_{aa(j)}). \end{aligned}$$

and

$$\begin{aligned} \|\vec{u}_j\| &= \sqrt{n \left[p_{AA(j)} \left[1 - (p_{AA(j)} - p_{aa(j)}) \right]^2 + \left[p_{Aa(j)} \left[-(p_{AA(j)} - p_{aa(j)}) \right]^2 + p_{aa(j)} \left[-1 - (p_{AA(j)} - p_{aa(j)}) \right]^2 \right]} \\ &= \sqrt{n \left[p_{AA(j)} + p_{aa(j)} - \left[(p_{AA(j)} - p_{aa(j)}) \right]^2 \right]}, \end{aligned}$$

$$\begin{aligned} \|\vec{v}_j\| &= \sqrt{n \left[p_{AA(j)} \left[-p_{Aa(j)} \right]^2 + p_{Aa(j)} \left[1 - p_{Aa(j)} \right]^2 + p_{aa(j)} \left[-p_{Aa(j)} \right]^2 \right]} \\ &= \sqrt{n \left[p_{Aa(j)} \left[1 - p_{Aa(j)} \right] \right]} \end{aligned}$$

Taking arc cos in the above formula renders θ_j in radians. For $\theta_j=90^\circ$, the two vectors are orthogonal. For $\theta_j \neq 90^\circ$, the two vectors are non-orthogonal, with the level of dependency being larger for values of θ_j near 0° or 180° .

Construction of orthogonal G and D matrices using the Gram-Schmidt process

The Gram-Schmidt process can be used to construct an orthogonal basis of the additive and dominance scales for each marker. The basic idea behind orthogonalization by the Gram-Schmidt process is that the first vector is kept unchanged, whereas the common component to both vectors is removed in the second vector. We explored three alternatives for applying the Gram-Schmidt process in the context of genomic relationship matrices. The first, GSP-A, initiates the process with \vec{u} (additive), whereas the second, GSP-D, initiates the process with \vec{v} (dominance). In the third alternative, GSP-N, additive and dominance vector of scales are forced to be of length 1 after using the scales from GSP-A. Conditions of orthogonality from Eqs. (7), (8), and (9) for GSP-A and GSP-D are verified in Appendix 1.

GSP-A Gram-Schmidt process The goal of the process is to obtain orthogonal vectors $\vec{\tau}_A$ and $\vec{\tau}_D$ for the additive and dominance scales, respectively. The Gram-Schmidt process for GSP-A is:

$$\vec{\tau}_{A(j)} = \vec{u}_{(j)}.$$

Therefore, the elements of $\vec{\tau}_{A(j)}$ are the additive scales from Eq. (5) for the individuals depending on their genotype, AA, Aa, or aa:

$$\begin{aligned} \tau_{A(j)}(AA) &= 1 - (p_{AA(j)} - p_{aa(j)}), \\ \tau_{A(j)}(Aa) &= -(p_{AA(j)} - p_{aa(j)}), \\ \tau_{A(j)}(aa) &= -1 - (p_{AA(j)} - p_{aa(j)}) \end{aligned} \tag{14}$$

For the dominance scales,

$$\vec{\tau}_{D(j)} = \vec{v}_{(j)} - \text{proj}_{\vec{\tau}_{A(j)}}(\vec{v}_{(j)}),$$

where $\text{proj}_{\vec{\tau}_{A(j)}}(\vec{v}_{(j)}) = \frac{\langle \vec{\tau}_{A(j)}, \vec{v}_{(j)} \rangle}{\langle \vec{\tau}_{A(j)}, \vec{\tau}_{A(j)} \rangle} \vec{\tau}_{A(j)}$

with $\langle \vec{\tau}_{A(j)}, \vec{v}_{(j)} \rangle = n (-p_{Aa(j)}) (p_{AA(j)} - p_{aa(j)})$
 and, $\langle \vec{\tau}_{A(j)}, \vec{\tau}_{A(j)} \rangle = n \left[p_{AA(j)} + p_{aa(j)} - \left[(p_{AA(j)} - p_{aa(j)}) \right]^2 \right]$.

Substituting $\langle \vec{\tau}_{A(j)}, \vec{v}_{(j)} \rangle$ and $\langle \vec{\tau}_{A(j)}, \vec{\tau}_{A(j)} \rangle$ into the above equation yields:

$$\vec{\tau}_{D(j)} = \vec{v}_{(j)} - \frac{(-p_{Aa(j)}) (p_{AA(j)} - p_{aa(j)})}{p_{AA(j)} + p_{aa(j)} - \left[(p_{AA(j)} - p_{aa(j)}) \right]^2} \vec{\tau}_{A(j)}.$$

Therefore, the elements of the orthogonal basis for the dominance vector, $\vec{\tau}_{D(j)}$, for the three genotypes are:

$$\begin{aligned}
 \tau_{D(j)}(AA) &= -p_{Aa(j)} - \frac{-p_{Aa(j)}(p_{AA(j)} - p_{aa(j)})}{p_{AA(j)} + p_{aa(j)} - [(p_{AA(j)} - p_{aa(j)})]^2} [1 - (p_{AA(j)} - p_{aa(j)})], \\
 \tau_{D(j)}(Aa) &= (1 - p_{Aa(j)}) - \frac{-p_{Aa(j)}(p_{AA(j)} - p_{aa(j)})}{p_{AA(j)} + p_{aa(j)} - [(p_{AA(j)} - p_{aa(j)})]^2} [-(p_{AA(j)} - p_{aa(j)})], \\
 \tau_{D(j)}(aa) &= -p_{Aa(j)} - \frac{-p_{Aa(j)}(p_{AA(j)} - p_{aa(j)})}{p_{AA(j)} + p_{aa(j)} - [(p_{AA(j)} - p_{aa(j)})]^2} [-1 - (p_{AA(j)} - p_{aa(j)})]
 \end{aligned}
 \tag{15}$$

As shown in [Appendix 2](#), the elements for centering the three genotypes from Eq. (15) are algebraically identical to those derived under the NOIA method:

$$\begin{aligned}
 \tau_{D(j)}(AA) &= \frac{-(2p_{Aa(j)}p_{aa(j)})}{(p_{AA(j)} + p_{aa(j)}) - (p_{AA(j)} - p_{aa(j)})^2}, \\
 \tau_{D(j)}(Aa) &= \frac{(4p_{AA(j)}p_{aa(j)})}{(p_{AA(j)} + p_{aa(j)}) - (p_{AA(j)} - p_{aa(j)})^2}, \\
 \tau_{D(j)}(aa) &= \frac{-(2p_{AA(j)}p_{Aa(j)})}{(p_{AA(j)} + p_{aa(j)}) - (p_{AA(j)} - p_{aa(j)})^2}.
 \end{aligned}$$

The first part of the Gram-Schmidt process leads to an orthogonal basis of additive and dominance scales for each marker. The scaling of additive relationship matrix is as shown in Eq 4. For the dominance component, the variance contributed by the *j*-th marker is:

$$\begin{aligned}
 V_{D(j)} &= p_{AA(j)} \tau_{D(j)}^2(AA) d_{(j)}^2 + p_{Aa(j)} \tau_{D(j)}^2(Aa) d_{(j)}^2 + p_{aa(j)} \tau_{D(j)}^2(aa) d_{(j)}^2 \\
 &= \frac{4p_{AA(j)}p_{Aa(j)}p_{aa(j)}(p_{Aa(j)}p_{aa(j)} + 4p_{AA(j)}p_{aa(j)} + p_{AA(j)}p_{Aa(j)})}{[p_{AA(j)} + p_{aa(j)} - [(p_{AA(j)} - p_{aa(j)})]^2]^2} d_{(j)}^2 \\
 &= \frac{4p_{AA(j)}p_{Aa(j)}p_{aa(j)} [4p_{AA(j)}p_{aa(j)} + p_{AA(j)} + p_{aa(j)} - (p_{AA(j)} + p_{aa(j)})^2]}{[p_{AA(j)} + p_{aa(j)} - [(p_{AA(j)} - p_{aa(j)})]^2]^2} d_{(j)}^2.
 \end{aligned}$$

After substituting $[4p_{AA(j)}p_{aa(j)} + p_{AA(j)} + p_{aa(j)} - (p_{AA(j)} + p_{aa(j)})^2]$ by its value: $p_{AA(j)} + p_{aa(j)} - [(p_{AA(j)} - p_{aa(j)})]^2$, the above equation reduces to:

$$V_{D(j)} = \frac{4p_{AA(j)}p_{Aa(j)}p_{aa(j)}}{p_{AA(j)} + p_{aa(j)} - [(p_{AA(j)} - p_{aa(j)})]^2} d_{(j)}^2.$$

The only difference between GSP-A and the NOIA orthogonalization is in the scaling of the **G** and **D** matrices. The scaling in Eqs. (12) and (13) of Vitezica et al. [7] was based on the realized trace of $\mathbf{H}_a\mathbf{H}'_a$ or $\mathbf{H}_d\mathbf{H}'_d$. In GSP-A, the expected values of those expressions are used for the scaling:

$$E \left[\frac{\text{tr}(\mathbf{H}_a\mathbf{H}'_a)}{n} \right] = \sum_{j=1}^m [p_{AA(j)} + p_{aa(j)} - (p_{AA(j)} - p_{aa(j)})^2],$$

$$\begin{aligned}
 E \left[\frac{\text{tr}(\mathbf{H}_d\mathbf{H}'_d)}{n} \right] &= \\
 &= \sum_{j=1}^m \left[\frac{4p_{AA(j)}p_{Aa(j)}p_{aa(j)}}{p_{AA(j)} + p_{aa(j)} - [(p_{AA(j)} - p_{aa(j)})]^2} \right].
 \end{aligned}$$

For very large *n*, the left and the right-hand sides of the above equations will be the same and also, the relationship matrices constructed using either method.

GSP-D Gram-Schmidt process The Gram-Schmidt process for GSP-D is initiated with the vector for the dominance effects and then the common part of the additive and dominance effects is removed when constructing an orthogonal basis:

$$\vec{\tau}_{D(j)} = \vec{v}_{(j)}.$$

Therefore, dominance scales for the three genotypes in $\vec{\tau}_{D(j)}$ using Eq. 6 are:

$$\text{with } \langle \tau_{D(j)}, \vec{u}_{(j)} \rangle = n(-p_{Aa(j)}) (p_{AA(j)} - p_{aa(j)}),$$

$$\text{and } \langle \vec{\tau}_{D(j)}, \vec{\tau}_{D(j)} \rangle = np_{Aa(j)} [1 - p_{Aa(j)}].$$

Substituting the above equations into Eq. (17) yields:

$$\vec{\tau}_{A(j)} = \vec{u}_{(j)} - \frac{(-p_{Aa(j)}) (p_{AA(j)} - p_{aa(j)})}{p_{Aa(j)} [1 - p_{Aa(j)}]} \vec{\tau}_{D(j)}.$$

Therefore, the elements of $\vec{\tau}_{A(j)}$ for the three genotypes are:

$$\begin{aligned} \tau_{A(j)}(AA) &= \left[1 - (p_{AA(j)} - p_{aa(j)}) \right] - \frac{-p_{Aa(j)} (p_{AA(j)} - p_{aa(j)})}{p_{Aa(j)} [1 - p_{Aa(j)}]} (-p_{Aa(j)}) \\ &= \left[1 - (p_{AA(j)} - p_{aa(j)}) \right] + \frac{-p_{Aa(j)} (p_{AA(j)} - p_{aa(j)})}{[1 - p_{Aa(j)}]} = \frac{2p_{aa(j)}}{[1 - p_{Aa(j)}]} \end{aligned} \tag{18}$$

$$\begin{aligned} \tau_{A(j)}(Aa) &= \left[-(p_{AA(j)} - p_{aa(j)}) \right] - \frac{-p_{Aa(j)} (p_{AA(j)} - p_{aa(j)})}{p_{Aa(j)} [1 - p_{Aa(j)}]} (1 - p_{Aa(j)}) \\ &= -(p_{AA(j)} - p_{aa(j)}) + (p_{AA(j)} - p_{aa(j)}) = 0 \end{aligned} \tag{19}$$

$$\begin{aligned} \tau_{A(j)}(aa) &= \left[-1 - (p_{AA(j)} - p_{aa(j)}) \right] - \frac{-p_{Aa(j)} (p_{AA(j)} - p_{aa(j)})}{p_{Aa(j)} [1 - p_{Aa(j)}]} (-p_{Aa(j)}) \\ &= \left[-1 - (p_{AA(j)} - p_{aa(j)}) \right] + \frac{-p_{Aa(j)} (p_{AA(j)} - p_{aa(j)})}{[1 - p_{Aa(j)}]} = \frac{-2p_{AA(j)}}{[1 - p_{Aa(j)}]} \end{aligned} \tag{20}$$

$$\begin{aligned} \tau_{D(j)}(AA) &= -p_{Aa(j)}, \\ \tau_{D(j)}(Aa) &= (1 - p_{Aa(j)}), \\ \tau_{D(j)}(aa) &= -p_{Aa(j)}. \end{aligned} \tag{16}$$

The scaling for **D** is the same as for the scaling in the non-Hardy-Weinberg parameterization, $\sum_{j=1}^m [p_{Aa(j)} - p_{Aa(j)}^2]$.

The additive scales are obtained as

$$\vec{\tau}_{A(j)} = \vec{u}_{(j)} - \text{proj}_{\vec{\tau}_{D(j)}}(\vec{u}_{(j)}), \tag{17}$$

where $\text{proj}_{\vec{\tau}_{D(j)}}(\vec{u}_{(j)}) = \frac{\langle \vec{\tau}_{D(j)}, \vec{u}_{(j)} \rangle}{\langle \vec{\tau}_{D(j)}, \vec{\tau}_{D(j)} \rangle} \vec{\tau}_{D(j)}$

The additive variance contributed for the m markers is:

$$\begin{aligned} V_{A(j)} &= \sum_{j=1}^m \left[\frac{4p_{AA(j)}p_{aa(j)}^2 + 4p_{aa(j)}p_{AA(j)}^2}{[1 - p_{Aa(j)}]^2} \right] a_{(j)}^2 \\ &= \sum_{j=1}^m \left[\frac{4p_{AA(j)}p_{aa(j)}}{[p_{AA(j)} + p_{aa(j)}]} \right] a_{(j)}^2. \end{aligned}$$

GSP-N Gram-Schmidt process The first step of the GSP-N process leads to an orthogonal basis of additive and dominance scales based on GSP-A. The next and final step of the GSP-N process is to obtain an orthonormal basis for additive and dominance scales by dividing $\vec{\tau}_{A(j)}$

and $\vec{\tau}_{D(j)}$ by the norm of each vector. The length of resulting vectors in GSP-N is unity, which implies that all markers contribute equally when constructing the genomic relationship matrices. Using the scales for $\vec{\tau}_{A(j)}$ and $\vec{\tau}_{D(j)}$ as described in Eqs. (14) and (15), the norm of the additive and dominance vectors are obtained as follows:

$$\begin{aligned} \|\vec{\tau}_{A(j)}\| &= \sqrt{n \left[p_{AA(j)} + p_{aa(j)} - \left[(p_{AA(j)} - p_{aa(j)}) \right]^2 \right]} \\ \|\vec{\tau}_{D(j)}\| &= \sqrt{\frac{4p_{AA(j)}p_{Aa(j)}^2p_{aa(j)}^2 + 16p_{AA(j)}^2p_{Aa(j)}p_{aa(j)}^2 + 4p_{AA(j)}^2p_{Aa(j)}^2p_{aa(j)}}{n \left[p_{AA(j)} + p_{aa(j)} - \left[p_{AA(j)} - p_{aa(j)} \right]^2 \right]}} \\ &= \sqrt{\frac{4p_{AA(j)}p_{Aa(j)}p_{aa(j)}}{n \left[p_{AA(j)} + p_{aa(j)} - \left[p_{AA(j)} - p_{aa(j)} \right]^2 \right]}} \end{aligned}$$

Then, matrices \mathbf{H}_a and \mathbf{H}_d become:

$$\begin{aligned} \mathbf{H}_a &= \begin{bmatrix} \frac{\vec{\tau}_{A_1}}{\|\vec{\tau}_{A_1}\|} & \frac{\vec{\tau}_{A_2}}{\|\vec{\tau}_{A_2}\|} & \dots & \frac{\vec{\tau}_{A_j}}{\|\vec{\tau}_{A_j}\|} & \frac{\vec{\tau}_{A_{j+1}}}{\|\vec{\tau}_{A_{j+1}}\|} & \dots & \frac{\vec{\tau}_{A_m}}{\|\vec{\tau}_{A_m}\|} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}, \\ \mathbf{H}_d &= \begin{bmatrix} \frac{\vec{\tau}_{D_1}}{\|\vec{\tau}_{D_1}\|} & \frac{\vec{\tau}_{D_2}}{\|\vec{\tau}_{D_2}\|} & \dots & \frac{\vec{\tau}_{D_j}}{\|\vec{\tau}_{D_j}\|} & \frac{\vec{\tau}_{D_{j+1}}}{\|\vec{\tau}_{D_{j+1}}\|} & \dots & \frac{\vec{\tau}_{D_m}}{\|\vec{\tau}_{D_m}\|} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}. \end{aligned}$$

The resulting relationship matrices still need to be scaled, which can be easily done by dividing \mathbf{G} and \mathbf{D} by $\left[\frac{\text{tr}(\mathbf{H}_a\mathbf{H}_a')}{n} \right]$ and $\left[\frac{\text{tr}(\mathbf{H}_d\mathbf{H}_d')}{n} \right]$, respectively.

Animals and data

A dataset comprising five trials (PHGC17, PHGC21, PHGC23, PHGC24, and PHGC25) aimed at investigating the genetic basis of resistance to porcine reproductive and respiratory syndrome virus (PRRSV) after natural infection was used for variance component estimation.

The LOESS function sets a low-degree polynomial at each point using weighted least squares and gives more weight to observations that are near the point for which response is being estimated and less weight to observations further away. This fitting was necessary to account for the natural variation in the concentration of viremia due to sampling and the methodology used to measure viremia in serum. The area under the curve for viremia for each individual, which will be referred to as viral load (VL), was the phenotype used for estimating

variance components.

A tissue sample obtained from each pig was used for SNP genotyping. DNA extraction and genotyping was carried out using the Infinium HD Assay Ultra protocol (Illumina Inc.) and the Illumina Porcine SNP60 BeadChip [16]. In total, 32,645 single nucleotide polymorphisms were used to construct relationship matrices. Details are in Gomez-Raya et al. [17].

Construction of G and D matrices and variance component estimation

An exact test for HWE conditions was carried out for all SNPs in the dataset. This test was performed using the Hardy–Weinberg package in the R language (<https://cran.r-project.org/web/packages/HardyWeinberg/HardyWeinberg.pdf>).

Construction of the \mathbf{G} and \mathbf{D} matrices was performed using the six methods described in the previous sections. The Kullback–Leibler divergence [18] was used to quantify the divergence between \mathbf{G} and \mathbf{D} . The Kullback–Leibler divergence [18] from \mathbf{Q} to \mathbf{P} was computed as:

$$D_{KL}(\mathbf{P} \parallel \mathbf{Q}) = 0.5 \left[\text{trace}(\mathbf{Q}^{-1}\mathbf{P}) + (\boldsymbol{\mu}_Q - \boldsymbol{\mu}_P)' \mathbf{Q}^{-1} (\boldsymbol{\mu}_Q - \boldsymbol{\mu}_P) - n + \ln \frac{|\mathbf{Q}|}{|\mathbf{P}|} \right],$$

Right after weaning, 903 Landrace \times Large White barrows were moved to farms with a history of PRRSV infections. Blood was drawn weekly and curves of viremia over time were constructed using a LOESS (LOcal regrESSion)

where $\boldsymbol{\mu}_P$ and $\boldsymbol{\mu}_Q$ are vectors of the means of the n individuals in the \mathbf{P} and \mathbf{Q} matrices, respectively. \mathbf{P} represents a normal multivariate distribution. Multivariate normal distribution \mathbf{Q} represents an approximation to \mathbf{P} . The Kullback–Leibler divergence is the average difference of the number of bits required for encoding samples of \mathbf{P}

using a code optimized for **Q** rather than one optimized for **P**. The unit of Kullback–Leibler divergence is the natural unit of information (nats) [19]. Values of $D_{KL}(\mathbf{P}||\mathbf{Q})$ equal to zero means that **P** and **Q** are the same distributions. The Kullback–Leibler divergence is not a true metric since it is not symmetrical and it does not obey the triangle inequality. Thus, Kullback–Leibler's divergence from **P** to **Q** is different to divergence from **Q** to **P**. Asymmetrical $D_{KL}(\mathbf{P}||\mathbf{Q}) > D_{KL}(\mathbf{Q}||\mathbf{P})$ implies that more information is needed to approximate **P** with **Q** than the other way around. In our analysis, **P** and **Q** were either two additive or dominance relationship matrices as derived for the six methods investigated in this paper.

The statistical model to analyze *VL* included the fixed effects of the mean and trial. Random variables were the additive and the dominance scales of the biological parameterization, as described in the Theory and methods section. Heritability (h^2) was estimated as the ratio of the estimate of additive genomic variance over the sum of estimates of the variance of each random component. The proportion of dominance variance (d^2) was estimated by dividing the estimate of the dominance variance component by the sum of the estimates of all random components in the model. The mixed linear models were fitted using ASReml [20], with **G** and **D** matrices as described in the previous sections.

Results

A genome-wide Fisher's exact test for Hardy–Weinberg departures was performed using all SNPs jointly for all five trials. A Manhattan plot showed that disequilibrium is common in the Landrace \times Large-White crosses, although the location of SNPs that were in disequilibrium did not appear to be random (Fig. 1). There was an average excess of 4.7% of heterozygotes across the genome. Thus, these data are appropriate for investigating the properties and comparison of alternative genomic and dominance relationships matrices with departures from HWE.

A first look at the performance of relationship matrices of the six methods revealed that all except the HW approach performed well to attain an average of diagonal elements equal to one and an average off diagonal elements equal to zero (Table 1). The average of the diagonal elements of the **G** matrix of the HW approach was 0.94 (Table 1). All eigenvalues were positive for all six methods used to construct genomic relationship matrices.

Kullback–Leibler's divergence for pairs of combinations of either **G** or **D** matrices is in Table 2. Divergence from **G** (GSP-D) to **G** created by other methods was larger than the divergence between other pairs of **G** matrices. It was also slightly asymmetrical, meaning that more information is needed to approximate **G** from

GSP-D using **G** from the other methods than the other way around. For the dominance relationship matrices, the Kullback–Leibler divergence using HW relationship distribution is strongly asymmetrical since relationship matrices NOIA, GSP-A, GSP-D, or GSP-N require a much larger number of bits when using a code optimized for HW than the other way around. This could be attributed again to the increase in heterozygosity in the cross-breeds, resulting in the scales in **D** from the HW method not being actually centered to zero. Divergence from **G** (or **D**) matrices in NOIA and GSP-A methods to **G** (or **D**) created by other methods were very similar because they are equivalent. The only difference between these two methods is that NOIA uses $tr(\mathbf{H}_a\mathbf{H}_a')/n$ and $tr(\mathbf{H}_d\mathbf{H}_d')/n$ as the denominators of the **G** and **D**, whereas GSP-A uses just the expected values of these expressions.

The methodology of the algebra of vector spaces allows to investigate orthogonality between additive and dominance vectors. An angle of 90° between the two vectors implies that the elements of the additive and dominance scales are orthogonal. Figure 2 shows a density plot of the estimates of the angle, θ , between additive and dominance vectors for all markers. Results show that a majority of markers had an angle between the additive and dominance vectors that was close to 90°. In Fig. 3, the $-\log_{10}$ (p-value) of Fisher's exact test for departures from HWE is plotted against the allele frequency and the angle between additive and dominance vectors for each marker when using NO-HW method to construct **G** and **D** matrices. Markers with intermediate allele frequencies tended to show significant departures from HWE and tended to have vectors of additive and dominance scales that were orthogonal.

Table 3 shows variance component estimates when using the six alternative methods to construct **G** and **D** matrices. The statistical analysis yielded a highly significant trial effect, which is attributable to the uncontrolled conditions in each trial. The NO-HW method tended to have lower estimates of the additive and dominance variance components than HW method, which is consistent with the HW method being upwards biased. Comparing estimates from the orthogonal methods, it can be summarized that (a) NOIA and GSP-A resulted in nearly identical estimates for both additive and dominance variance components, as expected; (b) GSP-D resulted in a larger estimate of the dominance variance component than any of the other methods, which was contrary to estimates from the NOIA and GSP-A methods; and (c) GSP-N resulted in an estimate of zero for dominance variance.

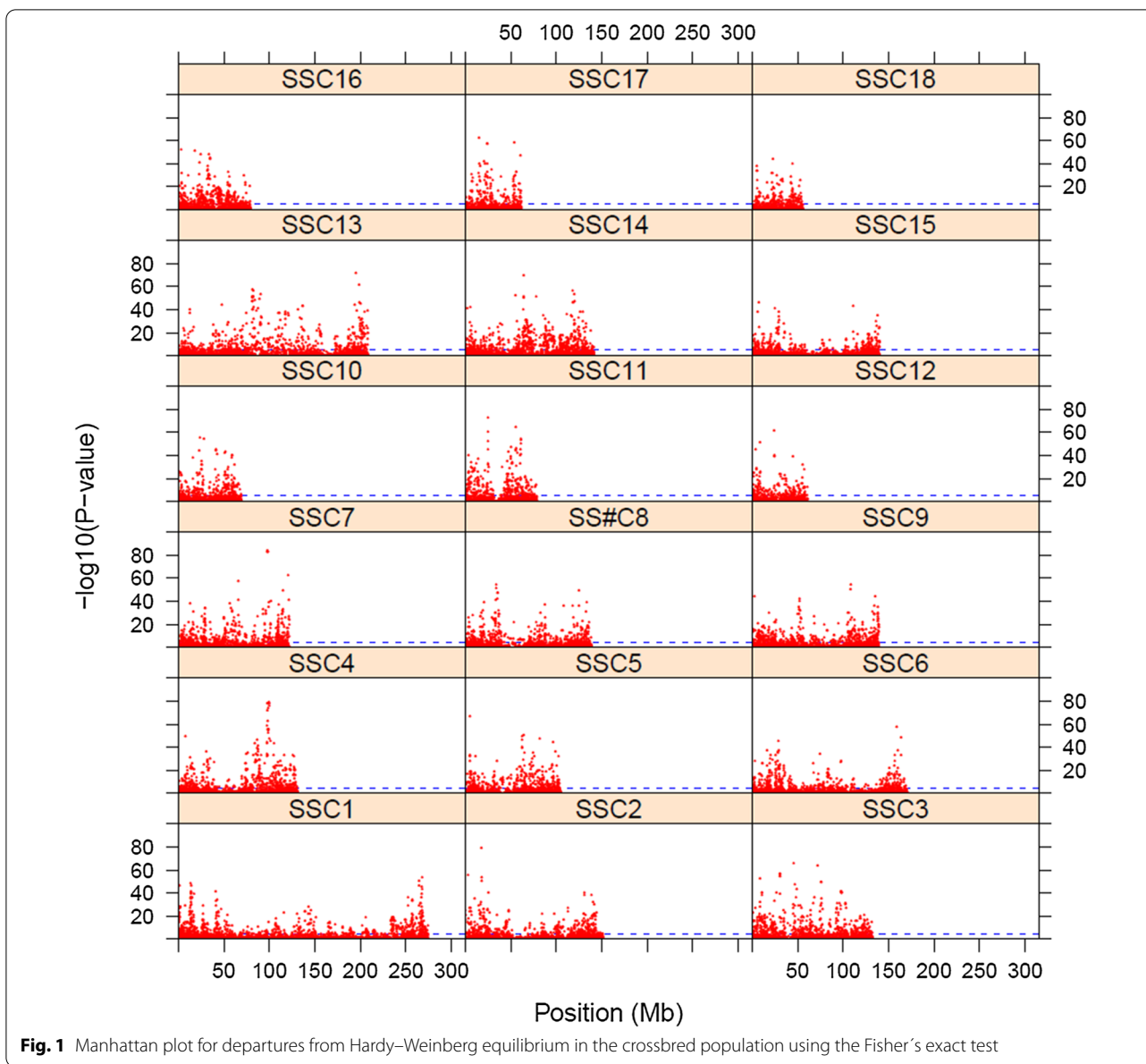


Table 1 Average elements of the genomic (**G**) and dominance (**D**) relationship matrix using biological parameterization when assuming Hardy-Weinberg equilibrium (**HW**), non-Hardy-Weinberg equilibrium (**NO-HW**), orthogonal NOIA (**NOIA**), or the Gram-Schmidt process for additive (**GSP-A**), dominance (**GSP-D**), and orthonormal additive (**GSP-N**) scales

	Relationship matrix					
	HW	NO-HW	NOIA	GSP-A	GSP-D	GSP-N
Average diagonal of G	0.942	0.994	1.000	0.999	0.994	1.000
Average off-diagonal of G	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
Range of eigenvalues of G	0.000-74.34	0.000-78.45	0.000-78.92	-0.000 to 78.45	0.001-69.97	0.000-72.17
Average diagonal of D	1.003	0.994	1.000	0.994	1.000	1.000
Average off-diagonal of D	0.012	-0.001	-0.001	-0.001	-0.001	-0.001

Table 2 Kullback–Leibler’s divergence for the genomic ($D_{KL}(G_1||G_2)$) and dominance ($D_{KL}(D_1||D_2)$) relationship matrices based on different parameterizations

G_2	G_1					
	HW	NO-HW	NOIA	GSP-A	GSP-D	GSP-N
HW	0.00	0.90	0.82	0.69	23.58	3.69
NO-HW	0.81	0.00	0.17	0.18	21.71	2.36
NOIA	0.86	0.32	0.00	0.03	21.51	2.50
GSP-A	0.72	0.32	0.03	0.00	21.66	2.54
GSP-D	31.22	26.67	26.11	26.56	0.00	27.88
GSP-N	4.17	2.58	2.61	2.68	24.69	0.00

D_2	D_1					
	HW	NO-HW	NOIA	GSP-A	GSP-D	GSP-N
HW	0.00	1938720.00	928502.90	930729.70	1938720.00	324210.90
NO-HW	7.19	0.00	20.30	20.24	0.00	27.38
NOIA	26.55	19.69	0.00	0.05	19.69	6.17
GSP-A	26.23	19.39	0.05	0.00	19.39	6.13
GSP-D	7.19	0.00	20.30	20.24	0.00	27.38
GSP-N	35.63	30.74	7.03	7.06	30.74	0.00

Values in the upper diagonal are $D_{KL}(G_1||G_2)$ (or $D_{KL}(D_1||D_2)$), while values in the lower diagonal are $D_{KL}(G_2||G_1)$ (or $D_{KL}(D_2||D_1)$). Parametrizations of relationship matrices are: Hardy–Weinberg equilibrium (HW), non-Hardy–Weinberg equilibrium (NO-HW), natural and orthogonal interactions approach (NOIA), and Gram-Schmidt process for additive (GSP-A), dominance (GSP-D) and orthonormal additive (GSP-N)

Discussion

Most studies in genetics applied to animal breeding or human genetics assume that the populations under study are in HWE. Advances in molecular technologies in recent years have allowed for a renewed interest in the impact of HWE assumptions in genetic analyses [21]. Modern animal breeding now relies on genomic prediction [1]. One of the most widely used methods of genomic prediction is GBLUP, which consists in replacing the traditional pedigree-based relationship matrix by a genomic relationship matrix that incorporates genotype information on SNPs. This requires genotype contributions for each SNP to be centered and scaled for both **G** and **D** matrices. The first attempt to construct **G** matrices from marker genotypes was by VanRaden [2], whose method did not consider dominance and assumed HWE for the scaling of the relationship matrix. Later on, a distinction between genomic relationships constructed using biological and statistical parameterization methods was proposed [5, 6], which also assumed HWE. A new method, NOIA, developed by Alvarez-Castro and Carlborg [8] and applied to the construction of genomic relationship by Vitezica et al. [7] uses an orthogonal partition of additive and dominance effects, and does not require the assumption of HWE. In this paper, we show that vector space algebra can be helpful in the construction of relationships matrices and in the evaluation of the level of departures from orthogonality between additive and

dominance vectors of scales. We showed that in our data, vectors of additive and dominance scales constructed using HW-NO are often orthogonal ($\theta=90^\circ$). We also show that markers at intermediate frequencies tend to have significant departures from HWE and their vectors of additive and dominant scales are orthogonal. This is because the numerator of $\cos \theta_j$ is $n(-p_{Aa(j)})(p_{AA(j)} - p_{aa(j)})$, which becomes zero (orthogonality) at intermediate frequencies ($p_{AA(j)} = p_{aa(j)}$).

We show in this paper that centering and scaling the **G** matrix using the NOIA method coincides with the GSP-A method, which is based on orthogonalization by the Gram-Schmidt process. The Gram-Schmidt process

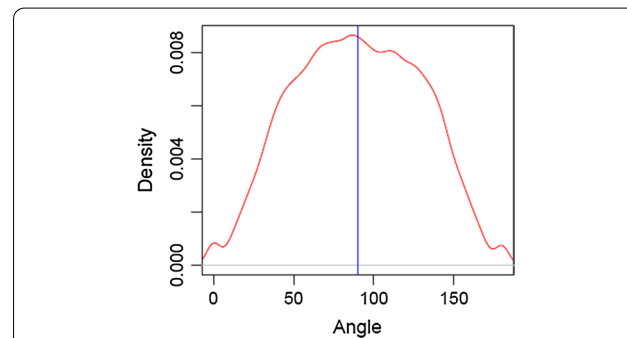
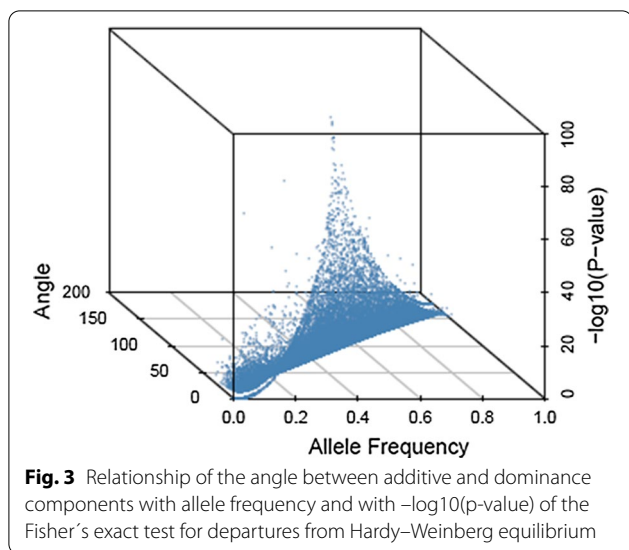


Fig. 2 Density of the angle between additive and dominance components across the genome of the crossbred population. The blue vertical bar shows the angle of orthogonality



converts vectors into an orthogonal system. This is done by taking one of the vectors and finding the projection of the next vector that is orthogonal to the former vector. We also showed that algebraically, GSP-A, and consequently, NOIA only centers and scales by accounting for the lack of HWE. However, orthogonality is achieved in the construction of the **D** matrix after removing the variation that is common to the **G** and **D** matrices. Thus, the proposed applications of the Gram-Schmidt process are equivalent to removing the additive-dominance co-variation from the other relationship matrix by linear regression.

Another alternative to deal with additive-dominance co-variation is inclusion of a covariance term between additive and dominance effects. This was explored by Xiang et al. [22] based on an equivalent statistical model, as proposed by Fernandez et al. [23]. More work is needed to compare this model with models that use NOIA or GSP-A to construct **G** and **D** matrices.

One of the most common situations where HWE does not hold is in crossbred populations. Lo et al. [24] first described how to use data on crossbreds and their corresponding purebreds to estimate breeding values and variance components. In their model, each individual has two breeding values; one for purebred performance and one for crossbred performance. Ibañez-Escriche et al. [25] first implemented a crossbred model that incorporates genomic information. More recently, Vitezica et al. [26] showed how additive and dominance components can be implemented in genomic prediction using purebred and crossbred performance to estimate breeding values for purebred animals and their crosses. Our analyses differ from those of Vitezica et al. [26] in that we do not use SNP genotype information on purebreds and just incorporate SNP genotype information from crossbreds into genomic relationship matrices, as an example with extreme departures from HWE (as expected and observed in our analyses). We did not differentiate between allele substitution effects according to the breed origin of the alleles either. In the analysis of crossbred data, the method of Vitezica et al. [26] is more appropriate if the goal is to estimate breeding values of purebreds, and SNP genotype information is available on the purebred parents. Nevertheless, their method also assumes HWE within each of the purebred parental populations, which may affect estimates of variance components.

We observed that roughly between 15 and 25% of all the variation for viral load following PRRSV infection is of genetic origin. Different methods to center and scale relationship matrices provided a different answer to the relative proportion of additive and dominance variation. HW, NO-HW, and GSP-D obtained a much higher estimate of dominance variance than of additive variance, whereas NOIA, GSP-A, and GSP-N resulted in the opposite. This is expected because of the way GSP-A, NOIA, and GSP-N are constructed, i.e. by removing common

Table 3 Estimates of variance components for viral load for the crossbred dataset, using genomic relationship matrices constructed with alternative biological parameterizations: Hardy-Weinberg equilibrium (HW), non-Hardy-Weinberg equilibrium (NO-HW), natural and orthogonal interactions approach (NOIA) and Gram-Schmidt process for additive (GSP-A), dominance (GSP-D) and orthonormal additive (GSP-N)

Parameters	HW	NO-HW	NOIA	GSP-A	GSP-D	GSP-N
h^2	4.30	4.01	16.08	16.16	0.00	15.86
d^2	20.35	20.44	1.79	1.80	25.11	0.00
V_A	58.68	54.63	220.47	221.82	0.0032	217.10
V_D	277.90	278.31	24.57	24.71	340.86	0.000
$V_{RESIDUAL}$	1029.28	1028.70	1125.84	1125.84	1016.80	1152.07
$V_{PHENOTYPIC}$	1365.86	1361.64	1370.88	1372.37	1357.66	1369.17

h^2 : heritability $\times 100$; d^2 : dominance $\times 100$; heritability and dominance are the additive or dominance variance divided by the sum of all variance components; V_A : Additive variance; V_D : dominance variance; $V_{RESIDUAL}$ residual variance; $V_{PHENOTYPIC} = V_A + V_D + V_{RESIDUAL}$

additive-dominance covariance in the centering of the dominance relationship matrix.

Fisher was the first to separate genetic variance into additive, dominance, and epistatic components using the least squares principle [27]. Later on, Cockerham partitioned the epistatic variance into additive \times additive, additive \times dominance, dominance \times additive, and dominance \times dominance interaction components [14]. Cockerham also showed how to scale additive and dominance components using a regression model under the assumption of HWE [14]. He stated “This particular set of scales (among the many others mathematically possible) was chosen for its utility. The scales pertaining to the marginal comparisons of each locus were chosen to separate the marginal variance into the additive (linear) and dominance (quadratic) portions that were long ago shown to be useful for expressing simply the correlation between parent and offspring and between other relatives”. Concerning hybrids, Stuber and Cockerham showed that a proper partitioning can be done in the parental populations, with each being assumed to be in HWE [28]. The NOIA model extends these scales to the situation in which the population is not in HWE [8]. Certainly, NOIA (or GSP-A) reflects better the linear regression nature, for example, of parent–offspring than GSP-D does. Also, it is more appropriate to predict response to selection. However, different scales in GSP-A (or NOIA or GSP-N) versus GSP-D yield different partitions of additive and dominance variance components, which deserves further investigation to address which of the partitions is of interest and for which purpose. In addition, method GSP-N standardized the length of the additive and dominance vectors to one, which implies that all markers weigh equally when constructing genomic relationship matrices, regardless of their frequencies (and/or marker locations). More work is needed to understand the implications and properties of alternative **G** and **D** matrices constructed using vector space algebra.

Conclusions

Vector space theory provides techniques that can be useful for the construction of relationship matrices in populations that are not in HWE. It can provide a measure of the degree of departures from orthogonality between additive and dominance components. It can also be applied to construct orthogonal or orthonormal relationship matrices, such as based on GSP-A, GSP-D, or GSP-N. The GSP-A method coincides with the NOIA method. With the GSP-N method, all markers contribute equally when constructing relationship matrices. Alternative orthogonal models to construct relationship matrices result in different estimates of additive and dominance variances, which requires further research.

Acknowledgements

The first author dedicates this paper to Prof. Larry R. Schaeffer at the University of Guelph, Canada. The authors would like to thank the PRRS Host Genetic Consortium for access and use of the dataset. The authors are grateful to the anonymous reviewers for their thorough review and inputs.

Authors' contributions

LGR conceived the idea, developed the theory and wrote the first version of the manuscript; WMR contributed in the analyses of data and the writing of the manuscript; and JCMD contributed to the writing of the manuscript and was involved in the generation of the data used in this paper. All authors read and approved the final manuscript.

Funding

Generation of these data was supported by the USDA ARS and NIFA award 2013-68004-20362. WMR and LGR acknowledge financial support from (1) the Salvador de Madariaga program of the Ministry of Education of Spain, (2) SusAn ERA-Net project “Sustainability of pig production through improved feed efficiency”, an ERA-Net co-funded under the European Union’s Horizon 2020 research and innovation programme, under Grant Agreement no 696231, and (3) project AGL2016-75942-R, “Caracterización molecular de la eficiencia alimentaria y de los caracteres reproductivos en cerdo Ibérico.”

Availability of data and materials

The data that support the findings of this study are accessible from the PRRS Host Genetics Consortium, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data may be available from the authors upon reasonable request.

Ethics approval and consent to participate

All experimental protocols for these trials were approved by the Kansas State University Institutional Animal Care and Use Committee.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Animal Science, Iowa State University, Ames, IA, USA. ² Dept Mejora Genética Animal, INIA, Madrid, Spain.

Appendices

Appendix 1

Verification of orthogonalization using the Gram-Schmidt process

In this Appendix, the conditions (7), (8), and (9) for orthogonalization of the Gram-Schmidt process are verified for both GSP-A and GSP-D.

Orthogonalization GSP-A

The three requirements for the orthogonal partition of variance components are:

$$\sum_{j=1}^m p_{AA(j)} \tau_{D(j)}(AA) + p_{Aa(j)} \tau_{D(j)}(Aa) + p_{aa(j)} \tau_{D(j)}(aa) = 0, \quad (\text{A1})$$

$$\sum_{j=1}^m p_{AA(j)} \tau_{A(j)}(AA) + p_{Aa(j)} \tau_{A(j)}(Aa) + p_{aa(j)} \tau_{A(j)}(aa) = 0, \quad (\text{A2})$$

$$\sum_{j=1}^m p_{AA(j)} \tau_{A(j)}(AA) \tau_{D(j)}(AA) + p_{Aa(j)} \tau_{A(j)}(Aa) \tau_{D(j)}(Aa) + p_{aa(j)} \tau_{A(j)}(aa) \tau_{D(j)}(aa) = 0 \tag{A3}$$

where m is the number of markers; $\tau_{A(j)}(k)$ and $\tau_{D(j)}(k)$ are the additive and dominance scales for the k -th genotype (AA, Aa, aa) of the j -th marker, respectively. Let $\beta_j = p_{AA(j)} + p_{aa(j)} - [(p_{AA(j)} - p_{aa(j)})]^2$. Substituting $\tau_{D(j)}(k)$ for their values in Eq. (15) into Eq. (A1) yields:

$$\begin{aligned} &= \sum_{j=1}^m p_{AA(j)} \left[-p_{Aa(j)} - \frac{-p_{Aa(j)}(p_{AA(j)} - p_{aa(j)})}{\beta_j} [1 - (p_{AA(j)} - p_{aa(j)})] \right] \\ &+ p_{Aa(j)} \left[1 - p_{Aa(j)} - \frac{-p_{Aa(j)}(p_{AA(j)} - p_{aa(j)})}{\beta_j} [-(p_{AA(j)} - p_{aa(j)})] \right] \\ &+ p_{aa(j)} \left[-p_{Aa(j)} - \frac{-p_{Aa(j)}(p_{AA(j)} - p_{aa(j)})}{\beta_j} [-1 - (p_{AA(j)} - p_{aa(j)})] \right] \\ &= \sum_{j=1}^m -p_{Aa(j)} + p_{Aa(j)} - \frac{-p_{Aa(j)}(p_{AA(j)} - p_{aa(j)})^2}{\beta_j} + \frac{-p_{Aa(j)}(p_{AA(j)} - p_{aa(j)})^2}{\beta_j} = 0. \end{aligned}$$

Substituting $\tau_{A(j)}(k)$ for their values in Eq. (14) into Eq. (A2) yields:

$$\begin{aligned} &\sum_{j=1}^m p_{AA(j)} [1 - (p_{AA(j)} - p_{aa(j)})] + p_{Aa(j)} [-(p_{AA(j)} - p_{aa(j)})] \\ &+ p_{aa(j)} [-1 - (p_{AA(j)} - p_{aa(j)})] \\ &= \sum_{j=1}^m (p_{AA(j)} - p_{aa(j)}) + (p_{AA(j)} + p_{Aa(j)} + p_{aa(j)}) [-(p_{AA(j)} - p_{aa(j)})] = 0. \end{aligned}$$

Substituting $\tau_{A(j)}(k)$ and $\tau_{D(j)}(k)$ for their values in Eqs. (14) and (15) into Eq. (A3) yields:

$$\begin{aligned} &\sum_{j=1}^m p_{AA(j)} [1 - (p_{AA(j)} - p_{aa(j)})] \left[-p_{Aa(j)} - \frac{-p_{Aa(j)}(p_{AA(j)} - p_{aa(j)})}{\beta_j} [1 - (p_{AA(j)} - p_{aa(j)})] \right] \\ &+ p_{Aa(j)} [-(p_{AA(j)} - p_{aa(j)})] \left[1 - p_{Aa(j)} - \frac{-p_{Aa(j)}(p_{AA(j)} - p_{aa(j)})}{\beta_j} [-(p_{AA(j)} - p_{aa(j)})] \right] \\ &+ p_{aa(j)} [-1 - (p_{AA(j)} - p_{aa(j)})] \left[-p_{Aa(j)} - \frac{-p_{Aa(j)}(p_{AA(j)} - p_{aa(j)})}{\beta_j} [-1 - (p_{AA(j)} - p_{aa(j)})] \right]. \end{aligned}$$

Let

$$\lambda_j = -\frac{-p_{Aa(j)}(p_{AA(j)} - p_{aa(j)})}{\beta_j}.$$

Orthogonalization GSP-D

Substituting $\tau_{D(j)}(k)$ for their values in Eq. (16) into Eq. (A1) yields:

Then, the above equation becomes:

$$\begin{aligned} & \sum_{j=1}^m p_{AA(j)} \left[1 - (p_{AA(j)} - p_{aa(j)}) \right] \left[-p_{Aa(j)} + \lambda_j \left[1 - (p_{AA(j)} - p_{aa(j)}) \right] \right] \\ & + p_{Aa(j)} \left[- (p_{AA(j)} - p_{aa(j)}) \right] \left[1 - p_{Aa(j)} + \lambda_j \left[- (p_{AA(j)} - p_{aa(j)}) \right] \right] \\ & + p_{aa(j)} \left[-1 - (p_{AA(j)} - p_{aa(j)}) \right] \left[-p_{Aa(j)} + \lambda_j \left[-1 - (p_{AA(j)} - p_{aa(j)}) \right] \right] \\ & = \sum_{j=1}^m p_{AA(j)} \left[-p_{Aa(j)} + \lambda_j \left[1 - (p_{AA(j)} - p_{aa(j)}) \right] \right] \\ & \quad - p_{aa(j)} \left[-p_{Aa(j)} + \lambda_j \left[-1 - (p_{AA(j)} - p_{aa(j)}) \right] \right] \\ & \quad + p_{AA(j)} \left[- (p_{AA(j)} - p_{aa(j)}) \right] \left[-p_{Aa(j)} + \lambda_j \left[1 - (p_{AA(j)} - p_{aa(j)}) \right] \right] \\ & \quad + p_{Aa(j)} \left[- (p_{AA(j)} - p_{aa(j)}) \right] \left[1 - p_{Aa(j)} + \lambda_j \left[- (p_{AA(j)} - p_{aa(j)}) \right] \right] \\ & \quad + p_{aa(j)} \left[- (p_{AA(j)} - p_{aa(j)}) \right] \left[-p_{Aa(j)} + \lambda_j \left[-1 - (p_{AA(j)} - p_{aa(j)}) \right] \right] \\ & = \sum_{j=1}^m -p_{Aa(j)} (p_{AA(j)} - p_{aa(j)}) + p_{AA(j)} \lambda_j \left[1 - (p_{AA(j)} - p_{aa(j)}) \right] \\ & \quad - p_{aa(j)} \lambda_j \left[-1 - (p_{AA(j)} - p_{aa(j)}) \right] \\ & \quad + p_{Aa(j)} \left(- (p_{AA(j)} - p_{aa(j)}) \right) + (p_{AA(j)} + p_{Aa(j)} + p_{aa(j)}) (-p_{Aa(j)}) \left(- (p_{AA(j)} - p_{aa(j)}) \right) \\ & \quad + (p_{AA(j)} + p_{Aa(j)} + p_{aa(j)}) (p_{AA(j)} - p_{aa(j)})^2 \lambda_j - (p_{AA(j)} + p_{Aa(j)} + p_{aa(j)}) (p_{AA(j)} - p_{aa(j)})^2 \lambda_j \\ & = \sum_{j=1}^m -p_{Aa(j)} (p_{AA(j)} - p_{aa(j)}) + \lambda_j (p_{AA(j)} + p_{aa(j)}) - \lambda_j \left[(p_{AA(j)} - p_{aa(j)})^2 \right] \\ & = \sum_{j=1}^m -p_{Aa(j)} (p_{AA(j)} - p_{aa(j)}) - \frac{-p_{Aa(j)} (p_{AA(j)} - p_{aa(j)})}{\beta_j} \left[p_{AA(j)} + p_{aa(j)} - (p_{AA(j)} - p_{aa(j)})^2 \right] \\ & = \sum_{j=1}^m -p_{Aa(j)} (p_{AA(j)} - p_{aa(j)}) + p_{Aa(j)} (p_{AA(j)} - p_{aa(j)}) = 0. \end{aligned}$$

$$\sum_{j=1}^m p_{AA(j)} (-p_{Aa(j)}) + p_{Aa(j)} (1 - p_{Aa(j)}) + p_{aa(j)} (-p_{Aa(j)})$$

$$= \sum_{j=1}^m -p_{Aa(j)} (p_{AA(j)} + p_{Aa(j)} + p_{aa(j)}) + p_{Aa(j)} = 0.$$

Substituting $\tau_{A(j)}(k)$ for their values in Eqs. (18), (19) and (20) into Eq. A2 yields:

$$\sum_{j=1}^m p_{AA(j)} \frac{2p_{aa(j)}}{p_{AA(j)} + p_{aa(j)}} + p_{aa(j)} \frac{-2p_{AA(j)}}{p_{AA(j)} + p_{aa(j)}} = 0.$$

Substituting $\tau_{A(j)}(k)$ and $\tau_{D(j)}(k)$ for their values in Eqs. (6), (18), (19) and (20) into Eq. (A3) yields:

$$\sum_{j=1}^m p_{AA(j)} \frac{2p_{aa(j)}}{p_{AA(j)} + p_{aa(j)}} (-p_{Aa(j)})$$

$$+ p_{aa(j)} \frac{-2p_{AA(j)}}{p_{AA(j)} + p_{aa(j)}} (-p_{Aa(j)}) = 0.$$

Appendix 2

Equivalence of NOIA and GSP-A

In this appendix, the equivalence for centering between the NOIA and GSP-A is shown. This will be accomplished by reducing the formulae obtained by the Gram-Schmidt process to the formulae of NOIA method for each of the three genotypes:

Genotypes AA

Let $\beta = \tau_A, \tau_A = p_{AA} + p_{aa} - [(p_{AA} - p_{aa})]^2$,

Equation (15) for individuals with genotypes AA becomes:

$$\tau_D(AA) = -p_{Aa} - \frac{-p_{Aa}(p_{AA} - p_{aa})}{\beta} [1 - (p_{AA} - p_{aa})]$$

$$= -p_{Aa} + \frac{p_{Aa} [(p_{AA} - p_{aa}) - (p_{AA} - p_{aa})^2]}{\beta}$$

$$= -p_{Aa} + \frac{p_{Aa} [(p_{AA} + p_{aa}) - (p_{AA} - p_{aa})^2 - 2p_{aa}]}{\beta}$$

$$= -p_{Aa} + \frac{p_{Aa} [\beta - 2p_{aa}]}{\beta}$$

$$= \frac{-p_{Aa}\beta + p_{Aa} [\beta - 2p_{aa}]}{\beta}$$

$$= \frac{-2p_{Aa}p_{aa}}{p_{AA} + p_{aa} - [(p_{AA} - p_{aa})]^2},$$

which is the same Eq. (11) for genotypes AA as in the NOIA method.

Genotypes Aa

Equation (15) for individuals with genotypes Aa becomes:

$$\tau_D(Aa) = (1 - p_{Aa}) - \frac{-p_{Aa}(p_{AA} - p_{aa})}{\beta} [-(p_{AA} - p_{aa})]$$

$$= \frac{\beta - \beta p_{Aa} - p_{Aa} [(p_{AA} - p_{aa})^2]}{\beta}$$

$$= \frac{\beta - p_{Aa} [\beta + (p_{AA} - p_{aa})^2]}{\beta}$$

$$= \frac{\beta - p_{Aa} [(p_{AA} + p_{aa})]}{\beta}$$

$$= \frac{p_{AA} + p_{aa} - [(p_{AA} - p_{aa})]^2 - p_{Aa} [(p_{AA} + p_{aa})]}{\beta}$$

$$= \frac{(1 - p_{Aa})(p_{AA} + p_{aa}) - [(p_{AA} - p_{aa})]^2}{\beta}$$

$$= \frac{[(p_{AA} + p_{aa})]^2 - [(p_{AA} - p_{aa})]^2}{\beta}$$

$$= \frac{4p_{AA}p_{aa}}{p_{AA} + p_{aa} - [(p_{AA} - p_{aa})]^2},$$

which is the same Eq. (11) for genotypes Aa as in the NOIA method.

Genotypes aa

Equation (15) for individuals with genotypes aa becomes:

$$\tau_D(aa) = -p_{Aa} - \frac{-p_{Aa}(p_{AA} - p_{aa})}{\beta} [-1 - (p_{AA} - p_{aa})]$$

$$= -p_{Aa} + \frac{p_{Aa} [(-p_{AA} + p_{aa}) - (p_{AA} - p_{aa})^2]}{\beta}$$

$$= -p_{Aa} + \frac{p_{Aa} [(p_{AA} + p_{aa}) - (p_{AA} - p_{aa})^2 - 2p_{AA}]}{\beta}$$

$$= -p_{Aa} + \frac{p_{Aa} [\beta - 2p_{AA}]}{\beta}$$

$$= \frac{-p_{Aa}\beta + p_{Aa} [\beta - 2p_{AA}]}{\beta}$$

$$= \frac{-2p_{AA}p_{Aa}}{p_{AA} + p_{aa} - [(p_{AA} - p_{aa})]^2},$$

which is the same Eq. (11) for genotypes aa as in the NOIA method.

Received: 31 January 2019 Accepted: 3 November 2020

Published online: 18 January 2021

References

1. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157:1819–29.
2. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91:4414–23.
3. Falconer DS, Mackay TFC. Introduction to quantitative genetics. Harlow: Pearson Prentice Hall; 2009.

4. Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet*. 2008;4:e1000008.
5. Vitezica ZG, Varona L, Legarra A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics*. 2013;195:1223–30.
6. 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:e45293.
7. Vitezica ZG, Legarra A, Toro MA, Varona L. Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. *Genetics*. 2017;206:1297–307.
8. Alvarez-Castro JM, Carlborg O. A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. *Genetics*. 2007;176:1151–67.
9. Joshi R, Meuwissen THE, Woolliams JA, Gjoen HM. Genomic dissection of maternal, additive and non-additive genetic effects for growth and carcass traits in Nile tilapia. *Genet Sel Evol*. 2020;52:1.
10. Varona L, Legarra A, Toro MA, Vitezica ZG. Non-additive effects in genomic selection. *Front Genet*. 2018;9:78.
11. Fisher RA. Average excess and average effect of a gene substitution. *Ann Eugenics*. 1941;11:53–63.
12. Falconer DS. A note on Fisher's 'average effect' and 'average excess'. *Genet Res*. 1985;46:337–47.
13. Toro MA, Varona L. A note on mate allocation for dominance handling in genomic selection. *Genet Sel Evol*. 2010;42:33.
14. 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.
15. Ward C, Kincaid DR. *Linear algebra: Theory and applications*. Burlington: Jones and Bartlett Publishers; 2009.
16. Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, Beever JE, et al. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One*. 2009;4:e6524.
17. Gomez-Raya L, Rauw WM, Dunkelberger JR, Dekkers JCM. Autozygosity and genetic differentiation of Landrace and Large White pigs as revealed by the genetic analyses of crossbreds. *Front Genet*. 2019;10:739.
18. Duchi J. Derivations for linear algebra and optimization. 2007. http://web.stanford.edu/~jduchi/projects/general_notes.pdf. Accessed 28 October 2020.
19. International Electrotechnical Commission. International standard, IEC 80000-13: Quantities and units - Part 13: Information science and technology; 2008.
20. Gilmour A, Cullis B, Welham S, Thompson R. ASREML discovery reference manual. Hemel Hempstead: VSN International Ltd; 2009.
21. Chen JJ. The Hardy-Weinberg principle and its applications in modern population genetics. *Front Biol*. 2010;5:348–53.
22. Xiang T, Christensen OF, Vitezica ZG, Legarra A. Genomic model with correlation between additive and dominance effects. *Genetics*. 2018;209:711–23.
23. Fernandez EN, Legarra A, Martinez R, Sanchez JP, Baselga M. Pedigree-based estimation of covariance between dominance deviations and additive genetic effects in closed rabbit lines considering inbreeding and using a computationally simpler equivalent model. *J Anim Breed Genet*. 2017;134:184–95.
24. Lo LL, Fernando RL, Cantet RJ, Grossman M. Theory for modelling means and covariances in a two-breed population with dominance inheritance. *Theor Appl Genet*. 1995;90:49–62.
25. Ibanez-Escriche N, Fernando RL, Toosi A, Dekkers JC. Genomic selection of purebreds for crossbred performance. *Genet Sel Evol*. 2009;41:12.
26. Vitezica ZG, Varona L, Elsen JM, Misztal I, Herring W, Legarra A. Genomic BLUP including additive and dominant variation in purebreds and F1 crossbreds, with an application in pigs. *Genet Sel Evol*. 2016;48:6.
27. Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. *Earth Environ Sci Trans Royal Soc Edin*. 1918;52:35.
28. Stuber CW, Cockerham CC. Gene effects and variances in hybrid populations. *Genetics*. 1966;54:1279–86.

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

