Genetics Selection Evolution

RESEARCH ARTICLE

Open Access

Segregation between breeds and local breed proportions in genetic and genomic models for crossbreds

Jón H. Eiríksson[1,2*], Guosheng Su[1], Ismo Strandén[3] and Ole F. Christensen[1]

Abstract

Background The breeding value of a crossbred individual can be expressed as the sum of the contributions from each of the contributing pure breeds. In theory, the breeding value should account for segregation between breeds, which results from the difference in the mean contribution of loci between breeds, which in turn is caused by differences in allele frequencies between breeds. However, with multiple generations of crossbreeding, how to account for breed segregation in genomic models that split the breeding value of crossbreds based on breed origin of alleles (BOA) is not known. Furthermore, local breed proportions (LBP) have been modelled based on BOA and is a concept related to breed segregation. The objectives of this study were to explore the theoretical background of the effect of LBP and how it relates to breed segregation and to investigate how to incorporate breed segregation (co)variance in genomic BOA models.

Results We showed that LBP effects result from the difference in the mean contribution of loci between breeds in an additive genetic model, i.e. breed segregation effects. We found that the (co)variance structure for BS effects in genomic BOA models does not lead to relationship matrices that are positive semi-definite in all cases. However, by setting one breed as a reference breed, a valid (co)variance structure can be constructed by including LBP effects for all other breeds and assuming them to be correlated. We successfully estimated variance components for a genomic BOA model with LBP effects in a simulated example.

Conclusions Breed segregation effects and LBP effects are two alternative ways to account for the contribution of differences in the mean effects of loci between breeds. When the covariance between LBP effects across breeds is included in the model, a valid (co)variance structure for LBP effects can be constructed by setting one breed as reference breed and fitting an LBP effect for each of the other breeds.

Background

Estimation of breeding values relies on additive genetic (co)variances between individuals. These (co)variances are usually derived from two types of information, i.e. pedigree or genotypes, where the latter are typically based on genotypes for genome-wide single nucleotide polymorphisms (SNPs) [1]. These two sources of information can also be combined in single-step genomic models [2, 3]. The general methods for estimating these (co)variances have been developed for purebred populations and

*Correspondence:
Jón H. Eiríksson
jonhjalti@lbhi.is
[1] Center for Quantitative Genetics and Genomics, Aarhus University, 8000 Aarhus C, Denmark
[2] Present Address: Faculty of Agricultural Sciences, Agricultural University of Iceland, 311 Borgarnes, Iceland
[3] Natural Resources Institute Finland (Luke), 31600 Jokioinen, Finland

thus modifications are needed for genetic evaluation of crossbred populations [4, 5].

For crossbred populations, other than the first generation of crossbreeding (F1), the additive genetic (co) variances include both contributions from each of the contributing pure breeds and contributions from segregation between breeds (breed segregation, BS) [4]. García-Cortés and Toro [6] presented a pedigree-based model for the genetic evaluation of crossbred populations in which the additive genetic (co)variances for crossbreds of any breed combination were partitioned into breed-specific terms and BS terms. In their model [6], each breed-specific and BS variance–covariance term is the product of a partial genetic variance component and a partial genetic relationship matrix. However, the inclusion of genomic information in that model is not straightforward because of the separate BS terms for each pair of breeds.

For genomic relationships, Strandén and Mäntysaari [7] presented a random regression approximation of the García-Cortés and Toro [6] model by partitioning the breeding value by breed proportions. Other studies suggested that the breed origin of alleles (BOA) should be accounted for in genomic evaluations for crossbreds [5, 8–10]. For genomic BOA models, the breed origins of the marker alleles are traced and their effects are allowed to depend on breed origin to account for breed differences in marker allele effects, e.g. due to differences in the linkage disequilibrium between markers and quantitative trait loci (QTL) and differences in the genetic background between breeds [8, 9]. Thus, the partitioning of breeding values into breed-specific effects is more accurate in BOA models than when the partitioning is based on the breed proportions alone. Christensen et al. [5] considered only the first generation of crossbreeding, for which BS is not present, while Ibánẽz-Escriche et al. [8] ignored BS for three-way and four-way crossbreeding, and Karaman et al. [9] and Eiríksson et al. [10] ignored BS in genomic BOA models for rotational crossbreeding systems. For a three-way terminal crossbreeding system, Christensen et al. [11] presented a single-step genomic model by partitioning the breeding value into breed-specific terms according to BOA and a BS term for segregation between the two maternal breeds. In that paper [11], a BS partial relationship for the segregation between the two maternal breeds was constructed based on information on BOA. For a population of inbred maize lines with lines from two genetic groups and admixed lines, Rio et al. [12] presented a genomic model with group-specific effects and a BS term. Recently, Aase et al. [13] presented an extension of the model of Rio et al. [12], adapted for an admixture of multiple groups in wild animal populations but ignored the BS term, although formulas for BS covariances were given in their Supplementary material [13], but without fully considering the properties of the proposed (co)variance structure. We are not aware of any published studies that explain how to construct BS partial relationships in genomic BOA models for crossbred populations with any breed composition.

A concept that accounts for genetic effects that are not accounted for by the breed-specific genetic effects, similar to BS, is local breed proportion (LBP) effects [14]. For crossbred dairy cattle, Eiríksson et al. [14] fitted LBP effects with random regressions on the proportion of alleles assigned to each breed origin within chromosome segments or for individual SNPs. They reported a low but statistically significant estimate of variance related to the LBP effects for milk production traits. Similarly, Bolormaa et al. [15] investigated the effect of local zebu or taurine ancestry of chromosome segments on phenotypes in Australian composite beef cattle and found that ancestry at some positions affected the studied traits. The effects of LBP, or of local ancestry, have similarities with BS but are defined for individual breeds rather than between pairs of breeds. However, the theoretical background of LBP and the exact relationship between LBP and BS is unclear.

Therefore, the objectives of this study were: (1) to derive and present the theoretical background of the effects of LBP in genetic models and how they are related to BS effects, (2) to investigate how BS effects can be included in genomic BOA models, and (3) to present variance structure that can be included in genetic and genomic models for crossbred populations with any breed composition to account for BS variance.

Methods

We start this section by reviewing the theory behind partitioning the breeding value into breed-specific terms and BS or LBP terms in genetic models. Second, we review the construction of breed-specific partial relationship matrices for these models. Third, we show that the (co) variance structure for BS effects based on BOA does not lead to relationship matrices that have a quadratic form. Finally, we present an alternative method for the inclusion of LBP effects in both pedigree-based and genomic BOA models, which accounts for BS and results in valid relationship matrices and variance components.

Theory

We start by reviewing the theory behind models for partitioning the additive genetic value of crossbreds into breed-specific terms [6]. We specify the additive genetic value of individual i as follows (inspired by Lo et al. [4] and the Appendix of Christensen et al. [11]):

$$g_i = \sum_{j=1}^{m} \left(\alpha_{s_i^j} + \alpha_{d_i^j} \right), \quad (1)$$

where $\alpha_{s_i^j}$ is the additive effect for the individual's paternal allele at locus j, $\alpha_{d_i^j}$ is the additive effect for its maternal allele at locus j, and m is the number of loci. The effects, $\alpha_{s_i^j}$ and $\alpha_{d_i^j}$, depend on BOA. We introduce the terms $r^b_{s_i^j}$ and $r^b_{d_i^j}$, which indicate whether the maternal and paternal alleles, respectively, at locus j originate from breed b, with $r^b_{s_i^j} = 1$ ($r^b_{d_i^j} = 1$) when the paternal (maternal) allele is from breed b and zero otherwise. We define the effect of local ancestry as the mean additive effect of $\alpha_{s_i^j}$ and $\alpha_{d_i^j}$ for all alleles at locus j from breed b, i.e. $\epsilon^b_j = E\left(\alpha_{s_i^j} | r^b_{s_i^j} = 1\right) = E\left(\alpha_{d_i^j} | r^b_{d_i^j} = 1\right)$. Furthermore, we define the within-breed additive effects of alleles as $\alpha^b_{s_i^j} = \alpha_{s_i^j} - \epsilon^b_j$ when $r^b_{s_i^j} = 1$ and $\alpha^b_{d_i^j} = \alpha_{d_i^j} - \epsilon^b_j$, when $r^b_{d_i^j} = 1$. The within-breed additive genetic effect of breed b for individual i is obtained by summing over all m loci: $a^b_i = \sum_{j=1}^{m} \left(r^b_{s_i^j} \alpha^b_{s_i^j} + r^b_{d_i^j} \alpha^b_{d_i^j} \right)$. Similarly, the total contribution of the effects of local ancestry of alleles from breed b for individual i is obtained by summing over all m loci: $u^b_i = \sum_{j=1}^{m} \epsilon^b_j (r^b_{s_i^j} + r^b_{d_i^j})$. The additive genetic value of individual i can then be obtained by summing over all breeds:

$$g_i = \sum\nolimits_b a^b_i + \sum\nolimits_b u^b_i. \quad (2)$$

The expectation of the total contribution of effects of local ancestry from breed b to individual i is

$$E\left(u^b_i\right) = E\left[\sum_{j=1}^{m} \epsilon^b_j \left(r^b_{s_i^j} + r^b_{d_i^j} \right)\right] = \sum_{j=1}^{m} \epsilon^b_j \times E\left(r^b_{s_i^j} + r^b_{d_i^j} \right)$$

$$= \delta^b f^b_i,$$

where $\delta^b = \sum_{j=1}^{m} 2\epsilon^b_j$ is the difference between the mean genetic level of breed b and the general mean, and $f^b_i = \left(f^b_{s_i} + f^b_{d_i}\right)/2 = E\left(r^b_{s_i^j} + r^b_{d_i^j} \right)/2$ is the expected proportion of genes in individual i that originate from breed b. We can now modify Eq. (2) to $g_i = \sum_b a^b_i + \sum_b (\delta^b \times f^b_i) + \sum_b \sum_{j=1}^{m} \epsilon^b_j \ddot{z}^b_{ij}$, where $\ddot{z}^b_{ij} = z^b_{s_i^j} + z^b_{d_i^j}$, $z^b_{s_i^j} = r^b_{s_i^j} - f^b_{s_i}$, and $z^b_{d_i^j} = r^b_{d_i^j} - f^b_{d_i}$ are deviations of the contribution of breed b at locus j from $f^b_{s_i}$ and $f^b_{d_i}$, the proportions of genes in s_i and d_i, respectively, that originate from breed b.

Vector $\mathbf{g}$, which contains the additive genetic values for crossbred individuals involving n_b breeds, can then be decomposed as:

$$\mathbf{g} = \sum\nolimits_b \mathbf{a}^b + \sum\nolimits_b (\delta^b \mathbf{f}^b) + \sum\nolimits_b \mathbf{k}^b, \quad (3)$$

where vector $\mathbf{a}^b$ contains the partial additive genetic effects related to breed b as defined in García-Cortés and Toro [6] and Christensen et al. [5], with value $a^b_i = \sum_{j=1}^{m} (r^b_{s_i^j} \alpha^b_{s_i^j} + r^b_{d_i^j} \alpha^b_{d_i^j})$ for individual i, vector $\mathbf{f}^b$ contains f^b_i for all individuals, and vector $\mathbf{k}^b$ contains $k^b_i = \sum_{j=1}^{m} \epsilon^b_j \ddot{z}^b_{ij}$. The effects of breed proportions, δ^b for each breed b, are fixed effects in the model and $\sum_b \mathbf{a}^b$ are random animal effects that are independent from $\sum_b \mathbf{k}^b$ [6, 9]. Therefore, the variance–covariance matrix of the vector of additive genetic values $\mathbf{g}$ for the crossbreds is therefore $Var(\mathbf{g}) = Var\left(\sum_b \mathbf{a}^b\right) + Var\left(\sum_b \mathbf{k}^b\right)$.

García-Cortés and Toro [6] modelled the term $\sum_b \mathbf{k}^b$ in Eq. (3) with BS terms, i.e. between each pair of breeds b and b', instead of summing across breeds,

$$\mathbf{g} = \sum\nolimits_b \mathbf{a}^b + \sum\nolimits_b (\delta^b \mathbf{f}^b) + \sum\nolimits_b \sum\nolimits_{b'>b} \mathbf{w}^{b,b'}, \quad (4)$$

where vector $\mathbf{w}^{b,b'}$ contains the BS effects. In this study, we specify the (co)variance of the $\sum_b \mathbf{k}^b$ term in Eq. (3), i.e. $Var\left(\sum_b \mathbf{k}^b\right)$. Considering $Var\left(\sum_b \sum_{b'>b} \mathbf{w}^{b,b'}\right) = Var\left(\sum_b \mathbf{k}^b\right)$, we develop two alternative models for genetic evaluation of crossbreds, i.e. a model based on LBP terms for each breed and a model based on BS terms (Eq. 4).

Breed-specific partial relationships

Here, we review how $\sum_b Var\left(\mathbf{a}^b\right)$ is inferred from pedigree and genomic information in models where the breeding value of crossbreds is split into breed-specific parts. Pedigree-based breed-specific partial additive genetic (co) variance matrices were derived in García-Cortés and Toro [6] and breed-specific partial genomic relationship matrices for terminal three-way crossbreeding systems were derived in Christensen et al. [11] and Sevillano et al. [16].

Pedigree-based partial relationship matrices are constructed recursively, as described by García-Cortés and Toro [6]. For breed b, the self-relationship for individual i is:

$$\mathbf{A}^b_{i,i} = f^b_i + \frac{1}{2} \mathbf{A}^b_{s,d},$$

and the partial relationship between individuals i and i' is:

$$\mathbf{A}^{b}_{i,i'} = \frac{1}{2}\left(\mathbf{A}^{b,}_{s,i'} + \mathbf{A}^{b,}_{d,i'}\right),$$

where f^b_i is the proportion of breed b, $\mathbf{A}^b_{s,d}$ is the partial relationship between the parents of i, and $\mathbf{A}^{b,}_{s,i'}$ and $\mathbf{A}^{b,}_{d,i'}$ are the partial relationships of individual i' with the sire and dam of i, respectively. Having the pedigree-based partial relationship matrices, the partial genetic effects in Eq. (3) are assumed to be independent and $\mathbf{a}^b \sim N\left(\mathbf{0}, \mathbf{A}^b\sigma^2_{a(b)}\right)$, where $\sigma^2_{a(b)}$ is the additive genetic variance for breed b.

For crossbreds with genome-wide SNP information, we consider a genotype matrix $\mathbf{M}$ of size $n \times m$, where n is the number of individuals and m is the number of SNPs, containing genotype information coded as 0 and 2 for loci that are homozygous for the alternative and reference allele, respectively, and 1 for heterozygous loci. To connect the two alleles to the correct breed origin, we need to have the genotypes phased and split into contributions from maternal and paternal gametes. Therefore, $\mathbf{M}$ is split into paternal and maternal allele matrices, $\mathbf{M} = \mathbf{M}_s + \mathbf{M}_d$, with the $\mathbf{M}_s$ and $\mathbf{M}_d$ matrices containing values equal to 1 for the reference allele and 0 for the alternative allele. Assigned BOA can give information on the origin of alleles throughout the genome and, thus, provides direct estimates of $r^b_{s^j_i}$ and $r^b_{d^j_i}$. From the assigned BOA, we can construct matrices $\mathbf{T}^b_s$ and $\mathbf{T}^b_d$ of size $n \times m$ that contain $r^b_{s^j_i}$ and $r^b_{d^j_i}$, respectively. By connecting the information on BOA and SNP genotypes, breed-specific genotype matrices can be formed: $\mathbf{M}^b = \mathbf{M}_s \circ \mathbf{T}^b_s + \mathbf{M}_d \circ \mathbf{T}^b_d$, where $\circ$ is element-wise multiplication. The genomic partial relationship matrix for breed b then is:

$$\mathbf{G}^b = \frac{(\mathbf{M}^b - \mathbf{1}\left[\mathbf{p}^b\right]^T \circ \mathbf{T}^b)(\mathbf{M}^b - \mathbf{1}\left[\mathbf{p}^b\right]^T \circ \mathbf{T}^b)^T}{c^b}, \tag{5}$$

where vector $\mathbf{p}^b$ contains the allele frequencies of the alternative allele in breed b, $\mathbf{T}^b = \mathbf{T}^b_s + \mathbf{T}^b_d$, and c^b is a scaling parameter that can be chosen such that the partial genomic matrix is compatible with the pedigree partial relationship matrix [5]. Based on the genomic partial relationship matrices, the partial genetic effects in Eq. (3) are assumed to be distributed $\mathbf{a}^b \sim N\left(\mathbf{0}, \mathbf{G}^b\sigma^2_{a(b)}\right)$.

Breed segregation

For the pedigree-based model, García-Cortés and Toro [6] derived the distribution for the $\mathbf{w}^{b,b'}$ term in Eq. (4) to be $\mathbf{w}^{b,b'} \sim N\left(\mathbf{0}, \mathbf{A}^{b,b'}_{(w)}\sigma^2_{w(b,b')}\right)$, where (w) indicates BS, $\mathbf{A}^{b,b'}_{(w)}$ is the pedigree-based partial relationship matrix for BS between breeds b and b', and $\sigma^2_{w(b,b')}$ is the BS variance between the two breeds. Matrix $\mathbf{A}^{b,b'}_{(w)}$ is constructed based on:

$$\mathbf{A}^{b,b'}_{(w)i,i} = 2f^b_{s_i}f^{b'}_{s_i} + 2f^b_{d_i}f^{b'}_{d_i} + \frac{1}{2}\mathbf{A}^{b,b'}_{(w)s,d},$$

and

$$\mathbf{A}^{b,b'}_{(w)i,i'} = \frac{1}{2}\left(\mathbf{A}^{b,b'}_{(w)s,i'} + \mathbf{A}^{b,b'}_{(w)d,i'}\right), \tag{6}$$

where $f^{b'}_{s_i}$ and $f^{b'}_{d_i}$ are the proportions of breed b' for the sire s and dam d of i, respectively [6].

We now consider the BS variance structure based on the BOA information in $\mathbf{T}^b$. We define $\kappa^b_{s^j_i} = \epsilon^b_j z^b_{s^j_i}$, $\kappa^b_{d^j_i} = \epsilon^b_j z^b_{d^j_i}$, $\kappa^{sum}_{s^j_i} = \sum_b \kappa^b_{s^j_i}$, $\kappa^{sum}_{d^j_i} = \sum_b \kappa^b_{d^j_i}$, and $\ddot{\kappa}^b_{ij} = \epsilon^b_j \ddot{z}^b_{ij}$. To derive the formulas for genomic models, we assume that the markers are the QTL in Eq. (1), rather than markers that are in linkage disequilibrium with the QTL. We further assume the local breed origin to be known for the markers and we assume the effect of local ancestry ϵ^b_j to be a random variable (as in Christensen et al. [11]), with mean zero. For a single locus, the total BS variance is: $Var\left(\ddot{\kappa}^{sum}_{ij}\right) = E\left[\left(\sum_b \epsilon^b_j \ddot{z}^b_{ij}\right)^2\right]$.

After some algebra, we get (see derivation in Appendix 1):

$$Var\left(\ddot{\kappa}^{sum}_{ij}\right) = \sum_b \sum_{b'>b} -\ddot{z}^b_{ij}\ddot{z}^{b'}_{ij} E\left[\left(\epsilon^b_j - \epsilon^{b'}_j\right)^2\right]. \tag{7}$$

Similarly, the breed-segregation covariance between individuals i and i' for one locus is (see derivation in Appendix 1):

$$Cov\left(\ddot{\kappa}^{sum}_{ij}, \ddot{\kappa}^{sum}_{i'^j}\right) = \sum_b \sum_{b'>b} \left(-\ddot{z}^b_{ij}\ddot{z}^{b'}_{i'^j} - \ddot{z}^{b'}_{ij}\ddot{z}^b_{i'^j}\right)\frac{1}{2}E\left[\left(\epsilon^b_j - \epsilon^{b'}_j\right)^2\right]. \tag{8}$$

Assuming ϵ^b_j and $\epsilon^{b'}_{j'}$ are independent when $j \neq j'$, both for $b = b'$ and for $b \neq b'$, and that each locus contributes equally to the total BS variance for each pair of breeds, we extend the result from the single locus in Eqs. (7) and (8) to the whole genome. We then have $Var\left(k^{sum}_i\right) = \sum_b \sum_{b'>b} -\frac{1}{m}\mathbf{z}^{b\,T}_i \mathbf{z}^{b'}_i \sum^m_{j=1} E\left[\left(\epsilon^b_j - \epsilon^{b'}_j\right)^2\right]$ and $Cov\left(k^{sum}_i, k^{sum}_{i'}\right) = \sum_b \sum_{b'>b} -\frac{1}{2m}\left(\mathbf{z}^{b\,T}_i \mathbf{z}^{b'}_{i'} + \mathbf{z}^{b'\,T}_i \mathbf{z}^b_{i'}\right) \sum^m_{j=1} E\left[\left(\epsilon^b_j - \epsilon^{b'}_j\right)^2\right]$, where $\mathbf{z}^b_i$ is a vector of $\ddot{z}^b_{ij}$ for all loci for individual i, and similarly for breed b' and individual i'.

Realizing that $\mathbf{z}_i^{b\,T}\mathbf{z}_i^{b'} = \frac{1}{2}(\mathbf{z}_i^{b\,T}\mathbf{z}_i^{b'} + \mathbf{z}_i^{b'\,T}\mathbf{z}_i^{b})$, we write the BS (co)variance as $Var\left(\sum_b \sum_{b'>b} \mathbf{w}^{b,b'}\right) = \sum_b \sum_{b'>b} \mathbf{W}^{b,b'}\sigma^2_{w(b,b')}$, where

$$\mathbf{W}^{b,b'} = -\frac{\mathbf{Z}^b\mathbf{Z}^{b'\,T} + \mathbf{Z}^{b'}\mathbf{Z}^{b\,T}}{2m} \tag{9}$$

is a BS similarity matrix for BS between breeds b and b', with $\mathbf{Z}^b = \mathbf{T}^b - 2\mathbf{f}^b\mathbf{1}^T$, and $\mathbf{1}$ is a vector of 1s, and

$$\sigma^2_{w(b,b')} = \frac{1}{2}\sum_{j=1}^{m} E\left[\left(\epsilon_j^b - \epsilon_j^{b'}\right)^2\right] \tag{10}$$

is the variance for BS between breeds b and b'. Here, the definition of the breed-segregation variance follows, e.g., Lo et al. [4], and is defined as the extra genetic variance in the F2 population compared to the F1.

However, matrix $\mathbf{W}^{b,b'}$ does not have a quadratic form and is, thus, not necessarily positive semi-definite. Thus, the BOA-derived BS similarity matrix is not a proper relationship matrix. In contrast, the pedigree-based partial BS relationship matrix given by García-Cortés and Toro [6] is always positive semi-definite.

Genomic local breed proportion (co)variance

Eiríksson et al. [14] fitted three independent LBP effects in a genomic model for a population of crossbreds from three breeds. The effects were either fitted with random regression on LBP (similar to SNP-best linear unbiased prediction (BLUP) for markers) or using LBP similarity matrices. In Appendix 2, we show that the variance for LBP effects from fitting three independent LBP terms is not guaranteed to be non-negative and, therefore, not a valid variance term. Here, we present an alternative LBP model, where one of the breeds is set as a reference breed and is left out of the model, but the other LBP effects are assumed to be correlated. We start by presenting a genomic model based on BOA, followed by the pedigree-based model. As before, we assume that the marker loci are the QTL, the origin of alleles is known, and the ϵ_j^b terms are random unknown variables with mean zero.

The variance for LBP effects over all breeds b for locus j is $Var\left(\ddot{\kappa}_{ij}^{sum}\right) = E\left[\left(\sum_b \ddot{z}_{ij}^b \epsilon_j^b\right)^2\right]$. We set b^* as the reference breed. For breeds $b \neq b^*$, and $b' \neq b^*$, we substitute $\epsilon_j^b = \epsilon_j^{b^*} + \epsilon_j^b - \epsilon_j^{b^*}$. After algebra detailed in Appendix 3, we have:

$$Var\left(\ddot{\kappa}_{ij}^{sum}\right) = \sum_{b\neq b^*}\sum_{b'\neq b^*} \ddot{z}_{ij}^b\ddot{z}_{ij}^{b'} E\left[\left(\epsilon_j^b - \epsilon_j^{b^*}\right)\left(\epsilon_j^{b'} - \epsilon_j^{b^*}\right)\right]. \tag{11}$$

Now, we have $n_b - 1$ LBP variance terms and their covariances between breeds [the $b \neq b'$ case in Eq. (11)]. The covariance between the LBP effects of individuals i and i' comes from the deviation in shared ancestry of loci from the expectation based on breed proportions, which is derived in the same manner as for the LBP variance above (see Appendix 3). Therefore,

$$Cov\left(\ddot{\kappa}_{ij}^{sum}, \ddot{\kappa}_{i'j}^{sum}\right) = \sum_{b\neq b^*}\sum_{b'\neq b^*} \ddot{z}_{ij}^b\ddot{z}_{i'j}^{b'} E\left[\left(\epsilon_j^b - \epsilon_j^{b^*}\right)\left(\epsilon_j^{b'} - \epsilon_j^{b^*}\right)\right]. \tag{12}$$

Assuming that ϵ_j^b and $\epsilon_{j'}^{b'}$ are independent for $j \neq j'$, both when $b = b'$ and when $b \neq b'$, the total variance of LBP effects related to individual i is obtained by summing the contributions of each locus from Eq. (11):

$$Var\left(k_i^{sum}\right) = \sum_{b\neq b^*}\sum_{b'\neq b^*}\sum_{j=1}^{m} \ddot{z}_{ij}^b\ddot{z}_{ij}^{b'} E\left[\left(\epsilon_j^b - \epsilon_j^{b^*}\right)\left(\epsilon_j^{b'} - \epsilon_j^{b^*}\right)\right]. \tag{13}$$

Similar to the SNP-BLUP and genomic BLUP models, we assume that the contribution of the (co)variance of LBP effects is the same for each locus for each breed and we denote this by $\sigma_{\theta(bb')}$. That is, for all loci j and j',

$$\sigma_{\theta(bb')} = E\left[\left(\epsilon_j^b - \epsilon_j^{b^*}\right)\left(\epsilon_j^{b'} - \epsilon_j^{b^*}\right)\right] = E\left[\left(\epsilon_{j'}^b - \epsilon_{j'}^{b^*}\right)\left(\epsilon_{j'}^{b'} - \epsilon_{j'}^{b^*}\right)\right]. \tag{14}$$

Therefore, we can write Eq. (13) as:

$$Var\left(k_i^{sum}\right) = \sum_{b\neq b^*}\sum_{b'\neq b^*} \mathbf{z}_i^{b\,T}\mathbf{z}_i^{b'}\sigma_{\theta(bb')}. \tag{15}$$

Similarly, the covariance of LBP effects between individuals i and i' from summing over loci based on Eqs. (12) and (14) is:

$$Cov\left(k_i^{sum}, k_{i'}^{sum}\right) = \sum_{b\neq b^*}\sum_{b'\neq b^*} \mathbf{z}_i^{b\,T}\mathbf{z}_{i'}^{b'}\sigma_{\theta(bb')}. \tag{16}$$

Combining Eqs. (15) and (16), and extending to multiple individuals, we write the total (co)variance of LBP effects as $Var\left(\sum_b \mathbf{k}^b\right) = \sum_{b\neq b^*}\sum_{b'\neq b^*} \mathbf{Q}^{bb'}\sigma_{k(bb')}$, where

$$\mathbf{Q}^{bb'} = \frac{\mathbf{Z}^b\mathbf{Z}^{b'\,T}}{m} \tag{17}$$

is a partial relationship matrix for LBP effects within breed when $b = b'$ and a matrix describing relationships of LBP effects between breeds b and b' when $b \neq b'$. Furthermore,

$$\sigma_{k(bb')} = m\sigma_{\theta(bb')} = \sum_{j=1}^{m} E\left[\left(\epsilon_j^b - \epsilon_j^{b^*}\right)\left(\epsilon_j^{b'} - \epsilon_j^{b^*}\right)\right] \tag{18}$$

is the variance of LBP effects for breed b when $b = b'$ and the covariance between LBP effects for breeds b and b' when $b \neq b'$.

When $b = b'$, both the partial relationship matrix ($\mathbf{Q}^{bb}$) and the variance component ($\sigma_{k(bb)}$) have a quadratic form and are therefore a valid relationship matrix and variance component, respectively. Defined in this way, the diagonal elements of $\mathbf{Q}^{bb}$ depend on deviations from the global breed proportion of individuals from their assignment to breed b of individual SNPs. The off-diagonal element between two individuals measures the degree to which these individuals share local deviations from the global breed proportions for breed b. For $b \neq b'$, the $\mathbf{Q}^{bb'}\sigma_{k(bb')}$ and $\mathbf{Q}^{b'b}\sigma_{k(bb')}$ terms contain covariances between the LBP effects for breeds b and b' for all pairs of individuals (where the diagonal elements are covariances between breeds within individuals). Note that $\mathbf{Q}^{bb'} = \mathbf{Q}^{b'b^T}$, while $\sigma_{k(bb')} = \sigma_{k(b'b)}$.

Using the variance structure described above, we write the total additive genetic effect in Eq. (3) based on $n_b - 1$ modified LBP effects:

$$\mathbf{g} = \sum_b \left(\mathbf{a}^b + \delta^b \mathbf{f}^b\right) + \sum_{b \neq b^*} \widetilde{\mathbf{k}}^b, \tag{19}$$

where vector $\widetilde{\mathbf{k}}^b$ contains the modified LBP effects of breed b, and we assume that:

$$\begin{bmatrix} \widetilde{\mathbf{k}}^1 \\ \vdots \\ \widetilde{\mathbf{k}}^{n_b-1} \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{Q}^{11}\sigma_{k(11)} & \cdots & \mathbf{Q}^{n_b-1,1}\sigma_{k(1,n_b-1)} \\ \vdots & \ddots & \vdots \\ \mathbf{Q}^{1n_b-1}\sigma_{k(1,n_b-1)} & \cdots & \mathbf{Q}^{n_b-1,n_b-1}\sigma_{k(n_b-1,n_b-1)} \end{bmatrix}\right). \tag{20}$$

In Eq. (20), the combined (co)variance structure for the LBP effects has a quadratic form and is therefore a valid (co) variance structure. Furthermore, the combined relationship matrix for LBP effects, $\mathbf{Q} = \begin{bmatrix} \mathbf{Q}^{11} & \cdots & \mathbf{Q}^{n_b-1,1} \\ \vdots & \ddots & \vdots \\ \mathbf{Q}^{1n_b-1} & \cdots & \mathbf{Q}^{n_b-1,n_b-1} \end{bmatrix}$, also has a quadratic form, similar to a genomic relationship matrix [1]. However, the variance and covariance components cannot be factored directly out of the combined LBP (co)variance matrices. The variance structure in Eq. (20) can alternatively be expressed as:

$$Var\begin{bmatrix} \widetilde{\mathbf{k}}^1 \\ \vdots \\ \widetilde{\mathbf{k}}_\star^{n_b-1,1} \\ \vdots \\ \widetilde{\mathbf{k}}_\star^{1,n_b-1} \\ \vdots \\ \widetilde{\mathbf{k}}^{n_b-1} \end{bmatrix} = \mathbf{\Sigma} \otimes \mathbf{Q}, \tag{21}$$

where $\mathbf{k}_\star^{b,b'}$, for b and b' in (1,..., $n_b - 1$), are vectors with artificial variables, i.e. variables that are not meaningful in themselves but included only to make the expression as a Kronecker product possible, matrix $\mathbf{\Sigma} = \begin{bmatrix} \sigma_{k(11)} & \cdots & \sigma_{k(1,n_b-1)} \\ \vdots & \ddots & \vdots \\ \sigma_{k(1,n_b-1)} & \cdots & \sigma_{k(n_b-1,n_b-1)} \end{bmatrix}$ is positive definite, and $\otimes$ is Kronecker product. Based on this formulation, the variance components for LBP effects can be estimated using standard software.

Similar to usual genomic BLUP, an equivalent SNP-BLUP model can be formed from the model in Eq. (19):

$$\mathbf{g} = \sum_b \left(\mathbf{M}^b\boldsymbol{\alpha}^b + \delta^b\mathbf{f}^b\right) + \sum_{b \neq b^*} \mathbf{Z}^b\boldsymbol{\theta}^b,$$

where vector $\boldsymbol{\alpha}^b$ contains marker effects in breed b and vector $\boldsymbol{\theta}^b$ contains the modified effects of LBP from breed b for each marker. It is assumed that $\boldsymbol{\alpha}^b \sim N(\mathbf{0}, \mathbf{I}\frac{1}{c^b}\sigma^2_{a(b)})$ (c^b is the scaling factor used in building $\mathbf{G}^b$) and $\begin{bmatrix} \boldsymbol{\theta}_x^1 \\ \vdots \\ \boldsymbol{\theta}_x^{n_b-1} \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{I}\sigma_{\theta(11)} & \cdots & \mathbf{I}\sigma_{\theta(1n_b-1)} \\ \vdots & \ddots & \vdots \\ \mathbf{I}\sigma_{\theta(1n_b-1)} & \cdots & \mathbf{I}\sigma_{\theta(n_b-1,n_b-1)} \end{bmatrix}\right)$.

Pedigree-based local breed proportion (co)variance

Here, we present how the (co)variance structure for LBP effects with one breed as a reference breed can be constructed using pedigree information. First, consider the contribution to the variance of the LBP effect from the paternal allele at a single locus. The expectation that the paternal allele at locus j is from breed b equals the sire's breed proportion, i.e. $E\left(r^b_{j\,s_i}\right) = f^b_{s_i}$. Therefore, the expectation of $\kappa^{sum}_{j\,s_i}$ is $E\left(\kappa^{sum}_{j\,s_i}\right) = \sum_b \epsilon^b_j f^b_{s_i}$. For the variance related to the LBP effect of the paternal allele, $r^b_{j\,s_i} r^{b'}_{j\,s_i} = 0$ for all $b \neq b'$, because the paternal allele at locus j originates only from one of the breeds. Furthermore, $\left(r^b_{j\,s_i}\right)^2 = r^b_{j\,s_i}$ for all b, because $r^b_{j\,s_i}$ is either 0 or 1. Therefore, $E\left[Var\left(\kappa^{sum}_{j\,s_i}\right)\right] = \sum_b f^b_{s_i}\epsilon^{b2}_j - \left(\sum_b f^b_{s_i}\epsilon^b_j\right)^2$. By setting one breed as the reference breed, b^*, and with algebra as detailed in Appendix 4, we have $n_b - 1$ LBP terms for

expected variance of the LBP effect of the allele (for $b = b'$),

$$E\left[Var\left(\kappa^{b}_{s^j_i}\right)\right] = \sum_{b\neq b^*} f^b_{s_i}\left(1-f^b_{s_i}\right)\left(\epsilon^b_j - \epsilon^{b^*}_j\right)^2. \quad (22)$$

Furthermore, based on algebra described in Appendix 4, the covariance between the LBP effects for breeds b and b' is:

$$E\left[Cov\left(\kappa^{b}_{s^j_i}, \kappa^{b'}_{s^j_i}\right)\right] = \sum_{b\neq b^*}\sum_{b'\neq\{b,b^*\}}\left[-f^b_{s_i}f^{b'}_{s_i}\left(\epsilon^{b'}_j - \epsilon^{b^*}_j\right)\left(\epsilon^b_j - \epsilon^{b^*}_j\right)\right]. \quad (23)$$

Similarly, for the maternal allele,

$$E\left[Var\left(\kappa^{b}_{d^j_i}\right)\right] = \sum_{b\neq b^*} f^b_{d_i}(1-f^b_{d_i})(\epsilon^b_j - \epsilon^{b^*}_j)^2, \quad (24)$$

and

$$E\left[Cov\left(\kappa^{b}_{d^j_i}, \kappa^{b'}_{d^j_i}\right)\right] = \sum_{b\neq b^*}\sum_{b'\neq\{b,b^*\}}\left[-f^b_{d_i}f^{b'}_{d_i}\left(\epsilon^{b'}_j - \epsilon^{b^*}_j\right)\left(\epsilon^b_j - \epsilon^{b^*}_j\right)\right]. \quad (25)$$

For the expectation of the variance of LBP effects related to both alleles, we need the covariance between the maternal and paternal LBP effects for locus j in individual i. The probability that the paternal allele of individual i comes from the paternal grandsire (or granddam) is $\frac{1}{2}$, and thus,

$$E\left(\kappa^{sum}_{s^j_i}\right) = \frac{1}{2}\left(\kappa^{sum}_{s^j_s} + \kappa^{sum}_{d^j_s}\right) = \frac{1}{2}\ddot{\kappa}^{sum}_{s^j}. \quad (26)$$

Similarly, $E(\kappa^{sum}_{d^j_i}) = \frac{1}{2}(\kappa^{sum}_{s^j_d} + \kappa^{sum}_{d^j_d}) = \frac{1}{2}\ddot{\kappa}^{sum}_{d^j}$. Therefore, the expectation of the covariance between the LBP effects of the maternal and paternal alleles is:

$$\mathrm{E}\left[Cov\left(\kappa^{sum}_{s^j_i}, \kappa^{sum}_{d^j_i}\right)\right] = \frac{1}{4}Cov\left(\ddot{\kappa}^{sum}_{s^j}, \ddot{\kappa}^{sum}_{d^j}\right), \quad (27)$$

where $Cov(\ddot{\kappa}^{sum}_{s^j}, \ddot{\kappa}^{sum}_{d^j})$ is the covariance between LBP effects at locus j of the two parents of i.

The expectation of the covariance of the LBP effects of the maternal and paternal alleles is thus 1/4 of the covariance between LBP effects of the parents, which is a known result for pedigree and BS relationships [6].

Now, we consider the LBP covariance between individuals i and i'. First, we look at the covariance of paternal allele of individual i at locus j with $\ddot{\kappa}^{sum}_{i'^j}$. Using the same argument as for Eq. (26), we have $E\left[Cov\left(\kappa^{sum}_{s^j_i}, \ddot{\kappa}^{sum}_{i'^j}\right)\right] = \frac{1}{2}Cov\left(\ddot{\kappa}^{sum}_{s^j}, \ddot{\kappa}^{sum}_{i'^j}\right)$. Similar to Eq. (11), we can write $Cov\left(\ddot{\kappa}^{sum}_{s^j}, \ddot{\kappa}^{sum}_{i'^j}\right) = \sum_{b\neq b^*}\sum_{b'\neq b^*}\ddot{z}^b_{s^j}\ddot{z}^{b'}_{i'^j}\left(\epsilon^b_j - \epsilon^{b^*}_j\right)\left(\epsilon^{b'}_j - \epsilon^{b^*}_j\right)$, i.e. $Cov\left(\ddot{\kappa}^{sum}_{s^j}, \ddot{\kappa}^{sum}_{i'^j}\right) = \sum_{b\neq b^*}\sum_{b'\neq b^*} Cov\left(\ddot{\kappa}^{b}_{s^j}, \ddot{\kappa}^{b'}_{i'^j}\right)$. Therefore, $E\left[Cov\left(\kappa^{sum}_{s^j_i}, \ddot{\kappa}^{sum}_{i'^j}\right)\right] = \sum_{b\neq b^*}\sum_{b'\neq b^*}\frac{1}{2}\left[Cov\left(\ddot{\kappa}^{b}_{s^j}, \ddot{\kappa}^{b'}_{i'^j}\right)\right]$ and, similarly, $E\left[Cov\left(\kappa^{sum}_{d^j_i}, \ddot{\kappa}^{sum}_{i'^j}\right)\right] = \sum_{b\neq b^*}\sum_{b'\neq b^*}\frac{1}{2}\left[Cov\left(\ddot{\kappa}^{b}_{d^j}, \ddot{\kappa}^{b'}_{i'^j}\right)\right]$. Therefore,

$$\begin{aligned} E\left[Cov\left(\ddot{\kappa}^{sum}_{ij}, \ddot{\kappa}^{sum}_{i'^j}\right)\right] &= E\left[Cov\left(\kappa^{sum}_{s^j_i}, \ddot{\kappa}^{sum}_{i'^j}\right)\right] + E\left[Cov\left(\kappa^{sum}_{d^j_i}, \ddot{\kappa}^{sum}_{i'^j}\right)\right] \\ &= \sum_{b\neq b^*}\sum_{b'\neq b^*}\frac{1}{2}\left[Cov\left(\ddot{\kappa}^{b}_{s^j}, \ddot{\kappa}^{b'}_{i'^j}\right) + Cov\left(\ddot{\kappa}^{b}_{d^j}, \ddot{\kappa}^{b'}_{i'^j}\right)\right]. \end{aligned}$$

Summing over all loci and the paternal and maternal alleles in Eqs. (22) to (25) and Eq. (27) gives the expected variance of LBP effects for individual i:

$$\begin{aligned} E\left[Var\left(k^{sum}_i\right)\right] = &\sum_{b\neq b^*}\left(\left[f^b_{s_i}\left(1-f^b_{s_i}\right) + f^b_{d_i}\left(1-f^b_{d_i}\right)\right]\sum_{j=1}^{m}\left(\epsilon^b_j - \epsilon^{b^*}_j\right)^2 + \frac{1}{2}Cov\left(k^b_{s_i}, k^b_{d_i}\right)\right) \\ &+ \sum_{b\neq b^*}\sum_{b'\neq\{b,b^*\}}\left[\left(-f^b_{s_i}f^{b'}_{s_i} - f^b_{s_i}f^{b'}_{s_i}\right)\sum_{j=1}^{m}\left(\epsilon^{b'}_j - \epsilon^{b^*}_j\right)\left(\epsilon^b_j - \epsilon^{b^*}_j\right)\right] + \frac{1}{4}Cov\left(k^b_{s_i}, k^{b'}_{d_i}\right) + \frac{1}{4}Cov\left(k^{b'}_{s_i}, k^b_{d_i}\right) \end{aligned} \quad (28)$$

and the expected covariance between individuals becomes:

$$E\left[Cov\left(k_i^{sum}, k_{i'}^{sum}\right)\right] = \sum_{b\neq b^*}\sum_{b'\neq b^*}\frac{1}{2}\left[Cov\left(k_{s_i}^{sum}, k_{i'}^{sum}\right) + Cov\left(k_{d_i}^{sum}, k_{i'}^{sum}\right)\right]. \quad (29)$$

Combining Eqs. (28) and (29), and extending to multiple individuals, we write the (co)variance of LBP effects as $Var\left(\sum_b \mathbf{k}^b\right) = \sum_{b\neq b^*}\sum_{b'\neq b^*}\mathbf{A}_{(k)}^{bb'}\sigma_{k(bb')}$, where $\mathbf{A}_{(k)}^{b,b'}$ is a pedigree-based partial LBP relationship matrix when $b = b'$ and a matrix that describes the relationship between the LBP effects of breeds b and b' when $b \neq b'$. Lo et al. [4] showed that pedigree relationship matrices that include BS can be constructed using the usual recursive rules, which, in turn, García-Cortés and Toro [6] used to construct partial BS relationship matrices. Similarly, the partial LBP relationship matrices $\mathbf{A}_{(k)}^{bb}$ are constructed as:

$$\mathbf{A}_{(k)i,i}^{bb} = f_s^b\left(1-f_s^b\right) + f_d^b\left(1-f_d^b\right) + \frac{1}{2}A_{(k)s,d}^{bb}, \quad (30)$$

and

$$\mathbf{A}_{(k)i,i'}^{bb} = \frac{1}{2}\left(A_{(k)s,i'}^{bb} + A_{(k)d,i'}^{bb}\right).$$

Matrices that describe the relationship of LBP effects between breeds, $\mathbf{A}_{(k)}^{bb'}$, when $b \neq b'$ are constructed as:

$$\mathbf{A}_{(k)i,i}^{bb'} = -f_s^b f_s^{b'} - f_d^b f_d^{b'} + \frac{1}{2}A_{(k)s,d}^{bb'}, \quad (31)$$

and

$$\mathbf{A}_{(k)i,i'}^{bb'} = \frac{1}{2}\left(A_{(k)s,i'}^{bb'} + A_{(k)d,i'}^{bb'}\right).$$

Note that $\mathbf{A}_{(k)}^{bb'}$ is a symmetric matrix and $\mathbf{A}_{(k)}^{bb'} = \mathbf{A}_{(k)}^{b'b}$.

Based on these derivations, we propose a pedigree-based model that partitions the breeding value as an alternative to the model of Garcia-Cortes and Toro [6], in which we consider pedigree LBP relationship matrices, $\mathbf{A}_{(k)}^{bb'}$, for $n_b - 1$ breeds and their covariances instead of BS effects for each pair of breeds. The $\widetilde{\mathbf{k}}^b$ terms in Eq. (19) can be estimated from pedigree by assuming:

$$\begin{bmatrix}\widetilde{\mathbf{k}}^1\\ \vdots \\ \widetilde{\mathbf{k}}^{n_b-1}\end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix}\mathbf{A}_{(k)}^{11}\sigma_{k(11)}^2 & \cdots & \mathbf{A}_{(k)}^{1,n_b-1}\sigma_{k(1,n_b-1)}^2\\ \vdots & \ddots & \vdots \\ \mathbf{A}_{(k)}^{1,n_b-1}\sigma_{k(1,n_b-1)}^2 & \cdots & \mathbf{A}_{(k)}^{n_b-1,n_b-1}\sigma_{k(n_b-1,n_b-1)}^2\end{bmatrix}\right).$$

Table 1 Pedigree and breed proportions for Example I

Animal	Sire	Dam	Breed proportions		
			A	B	C
1			1	0	0
2			1	0	0
3			1	0	0
4			0	1	0
5			0	0	1
6	4	1	0.5	0.5	0
7	5	6	0.25	0.25	0.5
8	2	7	0.625	0.125	0.25
9	3	7	0.625	0.125	0.25

Similar to the genomic LBP (co)variance structure, the combined pedigree LBP (co)variance matrix cannot be directly expressed as a Kronecker product of the relationship matrix and the (co)variance component matrix. However, an alternative model can be formed by including artificial effects, as described for the genomic model in Eq. (21).

Garcia-Cortes and Toro [6] showed that a generalized inverse of the partial BS relationship matrices can be constructed based on a slight modification of the Quaas [17] procedure for constructing the inverse of the numerator relationship matrix. A generalized inverse of $\mathbf{A}_{(k)}^{bb}$ can be found in a similar manner. However, for models with more than two breeds, the inverse of the matrix $\mathbf{A}_{(k)} = \begin{bmatrix}\mathbf{A}_{(k)}^{11} & \cdots & \mathbf{A}_{(k)}^{1,n_b-1}\\ \vdots & \ddots & \vdots \\ \mathbf{A}_{(k)}^{1,n_b-1} & \cdots & \mathbf{A}_{(k)}^{n_b-1,n_b-1}\end{bmatrix}$ is needed, which has a more complicated structure.

Converting (co)variances

The estimated LBP variance components may depend on the choice of the reference breed, b^*, and are not easily interpreted. However, with the derivations presented here, we can convert the estimated LBP (co)variances to represent the better-known concept of BS variance, as will be described in the following.

As detailed in Appendix 5, with some algebra on the expression for BS variance in Eq. (10) and for the LBP

(co)variance in Eq. (18), the BS variance between breeds b and b' is calculated from the LBP (co)variance as:

$$\sigma^2_{w(b,b')} = \frac{1}{2}\sigma_{k(bb)} + \frac{1}{2}\sigma_{k(b'b')} - \sigma_{k(bb')}, \quad (32)$$

and for the BS variance between the reference breed for LBP, b^*, and breed b as:

$$\sigma^2_{w(b,b^*)} = \frac{1}{2}\sum_{j=1}^{m}\left(\epsilon_j^b - \epsilon_j^{b^*}\right)^2 = \frac{1}{2}\sigma_{k(bb)}.$$

It follows that estimates of LBP variances can be obtained from estimated BS variances as $\sigma_{k(bb)} = 2\sigma^2_{w(b,b^*)}$. Furthermore, the covariance between LBP effects for breeds b and b' can be calculated from BS variances as (derivation in Appendix 5) $\sigma_{k(bb')} = \sigma^2_{w(b,b^*)} + \sigma^2_{w(b',b^*)} - \sigma^2_{w(b,b')}$.

In the case of two breeds, b and b', we only have a single BS term and a single LBP term. Furthermore, $\mathbf{Q}^{bb}\sigma_{k(bb)} = \mathbf{W}^{b,b'}\sigma^2_{w(b,b')}$, $\mathbf{Q}^{bb} = 2\mathbf{W}^{b,b'}$, and $\sigma_{k(bb)} = \frac{1}{2}\sigma^2_{w(b,b')}$.

Example I

For a demonstration of the partial BS relationship and similarity matrices and the partial LBP relationship matrices, Table 1 presents the matrices for a small crossbred pedigree for three breeds, A, B, and C, with nine individuals. We constructed the pedigree-based BS relationship matrices for these individuals using Eq. (6). We also obtained genotypes and assigned BOA for three crossbred dairy cows, which had crossbred dams and the same pedigree structure as individuals 7 to 9 in Table 1. The genotypes are a part of the dataset used by Eiríksson et al. [18], where further details can be found. Based on these data, we constructed the $\mathbf{T}^b$ matrices, which were then used to construct the genomic BS similarity matrices using Eq. (9). We calculated the values of a genomic partial LBP relationship matrices for the same cows for breeds A and B and the genomic matrix describing the relationship of LBP effects between breeds A and B using Eq. (17). Furthermore, for individuals 7 to 9 in Table 1, we constructed the pedigree-based partial LBP relationship matrices for breeds A and B (setting breed C as the reference breed) using Eq. (30) and the matrix of relationships between LBP effects of breeds A and B using Eq. (31).

Example II

To demonstrate and test a model with the genomic partial LBP relationship matrices, we simulated genotypes and phenotypes for crossbreds of three breeds using the QMSim software [19] (see Additional file 1 for the QMSim instruction code). We simulated a genome consisting of 10 chromosomes of 50 cM. Initially, each chromosome had 1000 evenly spaced biallelic markers and 50 randomly placed biallelic QTL. The QTL effects were randomly drawn from a gamma distribution with shape parameter 0.5. The simulated trait had a heritability of 0.3. Throughout the simulation, each female had five offspring with an equal probability of being male or female.

The 1000 historical generations consisted of 2000 animals each. From the last historical generation, three populations, A, B and C, were formed by randomly selecting 20 males and 20 females as founders of each population. Each population was then randomly mated for 40 discrete generations with replacements selected at random from the offspring of the previous generation. In the first 10 generations, the number of females in each population increased linearly to 60, but population size was kept stable thereafter. A single generation of F1 crossbred animals was formed by mating 20 randomly selected males from generation 39 of population A to 60 randomly selected females from generation 39 of population B. Subsequently, another crossbred population, D, was formed by mating 20 randomly selected males from generation 40 of population C to 100 randomly selected females from the F1 crossbred population. Population D mated randomly for three generations, resulting in 1500 animals with one or both parents being crossbred.

We used the marker genotypes of generations 38 to 40 of populations A, B and C, and all three generations of population D to test our model. For these genotypes, 7375 SNPs had a minor allele frequency higher than 0.05 and were retained for further analysis. In the output of QMSim, the phase of the genotypes is available in the output. However, for the crossbred individuals, the population of origin of the alleles is not in the output. Therefore, we assigned the marker alleles in the D population to BOA using the AllOr method [10] and using the phased genotypes from generations 38 to 40 of the purebred populations as reference haplotypes for AllOr. Furthermore, we set the window size for the BOA assignment to 50 SNPs with an overlap of 45 SNPs. Based on the assigned BOA and the genotypes of population D, we formed the within-breed genomic relationship matrices based on Eq. (5), with allele frequencies $\mathbf{p}^b$ calculated from generations 38 to 40 for populations $b = \mathrm{A}, \mathrm{B}, \mathrm{C}$, and standardizing factor $c_b = \sum_{j=1}^{m} 2p_j^b(1-p_j^b)$. We set population C as reference breed. Thus, only LBP effects for breeds A and B were included in the model. The LBP relationship matrix was constructed as $\mathbf{Q}^{AB} = \begin{bmatrix}\mathbf{Z}^A \\ \mathbf{Z}^B\end{bmatrix}\begin{bmatrix}\mathbf{Z}^A \\ \mathbf{Z}^B\end{bmatrix}^T \times \frac{1}{m}$, where $\mathbf{Z}^A$ and $\mathbf{Z}^B$ contain the BOA information and were calculated as described for $\mathbf{Z}^b$

Table 2 Pedigree-based breed-segregation relationship matrices for Example 1

Breeds	Animal	7	8	9
A-B	7	0.5	0.25	0.25
	8	0.25	0.125	0.125
	9	0.25	0.125	0.125
A-C	7	0	0	0
	8	0	0.25	0
	9	0	0	0.25
B-C	7	0	0	0
	8	0	0.25	0
	9	0	0	0.25

The coefficients from pedigree-based partial breed-segregation relationship matrices for animals with a crossbred parent in the pedigree in Table 1

Table 3 Genomic breed-segregation similarity matrices for Example 1

Breeds	Animal	7	8	9
A-B	7	0.48	0.25	0.18
	8	0.25	0.12	0.09
	9	0.18	0.09	0.07
A-C	7	0.00	0.03	−0.02
	8	0.03	0.25	0.06
	9	−0.02	0.06	0.19
B-C	7	0.00	−0.03	0.02
	8	−0.03	0.25	−0.06
	9	0.02	−0.06	0.28

The coefficients from genomic breed segregation similarity matrices for animals with a crossbred parent in the pedigree in Table 1

in Eq. (7), and m is the number of SNPs in the analysis (7375).

For the variance component estimation, we considered the following model:

$$\mathbf{y} = \mathbf{F}\boldsymbol{\delta} + \mathbf{a}_A + \mathbf{a}_B + \mathbf{a}_C + \mathbf{X}_A\widetilde{\mathbf{k}}^A_\star + \mathbf{X}_B\widetilde{\mathbf{k}}^B_\star + \mathbf{e},$$

where vector $\mathbf{y}$ contains the phenotypes of 1500 individuals from population D, matrix $\mathbf{F}$ contains the proportion of alleles assigned to each breed, vector $\boldsymbol{\delta}$ contains the effects of breed proportions, vectors $\mathbf{a}_A$, $\mathbf{a}_B$, and $\mathbf{a}_C$ contain the partial genetic effects for breeds A, B, and C, respectively, vector $\widetilde{\mathbf{k}}^A_\star = \begin{bmatrix} \widetilde{\mathbf{k}}^A \\ \widetilde{\mathbf{k}}^{B,A}_\star \end{bmatrix}$ contains the modified LBP effects of breed A ($\widetilde{\mathbf{k}}^A$) and artificial effects $\widetilde{\mathbf{k}}^{B,A}_\star$, and matrix $\mathbf{X}_A = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix}$ connects the phenotypes to the first 1500 elements of $\widetilde{\mathbf{k}}^A_\star$. Similarly, $\widetilde{\mathbf{k}}^B_\star = \begin{bmatrix} \widetilde{\mathbf{k}}^{A,B}_\star \\ \widetilde{\mathbf{k}}^B \end{bmatrix}$ contains the modified LBP effects of breed B ($\widetilde{\mathbf{k}}^B$) and artificial effects ($\widetilde{\mathbf{k}}^{A,B}_\star$), and $\mathbf{X}_B = \begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix}$. We assumed the variance structure to be $\mathbf{a}_b \sim N(\mathbf{0}, \sigma^2_{a(b)}\mathbf{G}^b)$ for populations b = A, B, and C, $\begin{bmatrix} \widetilde{\mathbf{k}}^A_\star \\ \widetilde{\mathbf{k}}^B_\star \end{bmatrix} \sim N(\mathbf{0}, \begin{bmatrix} \sigma_{k(AA)} & \sigma_{k(AB)} \\ \sigma_{k(AB)} & \sigma_{k,(BB)} \end{bmatrix} \otimes \mathbf{Q}^{AB})$, and $\mathbf{e} \sim N(\mathbf{0}, \sigma^2_e\mathbf{I})$, where σ^2_e is the residual variance.

We estimated the variance components using the average information REML (AI-REML) algorithm implemented in the DMU package [20]. We calculated the inverses of the relationship matrices $\mathbf{G}^A$, $\mathbf{G}^B$, $\mathbf{G}^C$ and $\mathbf{Q}^{AB}$ using Julia [21] (see Additional file 2). We added a small value, 0.0001, to the diagonal of the relationship matrices to avoid singularity.

For comparison with the estimated values, we calculated the true variance components based on the simulated QTL

Table 4 Pedigree-based local breed proportion relationship matrices for Example 1

Breed	Animal	7	8	9
A	7	0.25	0.125	0.125
	8	0.125	0.1875	0.0625
	9	0.125	0.0625	0.1875
B	7	0.25	0.125	0.125
	8	0.125	0.1875	0.0625
	9	0.125	0.0625	0.1875
A-B	7	−0.25	−0.125	−0.125
	8	−0.125	−0.0625	−0.0625
	9	−0.125	−0.0625	−0.0625

The coefficients from the pedigree-based partial local breed proportion relationship matrices for breeds A and B and the elements from the pedigree-based matrix describing the relationship of local breed proportion effects between breeds A and B. Values are shown for animals with a crossbred parent in the pedigree in Table 1. Breed C is the reference breed

Table 5 Genomic local breed proportion matrices for Example 1

Breed	Animal	7	8	9
A	7	0.24	0.14	0.08
	8	0.14	0.19	0.08
	9	0.08	0.08	0.13
B	7	0.24	0.11	0.10
	8	0.11	0.19	0.02
	9	0.10	0.02	0.17
A-B	7	−0.24	−0.14	−0.08
	8	−0.11	−0.06	−0.04
	9	−0.10	−0.06	−0.03

The coefficients from the genomic local breed proportion partial relationship matrices for breeds A and B and the elements from the genomic matrix describing the relationship of local breed proportion effects between breeds A and B. Values are shown for animals with a crossbred parent in the pedigree in Table 1. Breed C is the reference breed

effects and allele frequencies. The within-population additive genetic variance for each population b was calculated as $\sigma^2_{a(b)_{true}} = 2\sum_j p_{QTL,b_j}(1-p_{QTL,b_j})(\beta_{j,1}-\beta_{j,2})^2$, where the summation is over all segregating QTL, p_{QTL,b_j} is the frequency of allele 1 for QTL j in population b, calculated from generations 38 to 40, and $\beta_{j,1}$ ($\beta_{j,2}$) is the simulated effect of allele 1 (2) for QTL j. The true LBP variances were calculated as $\sigma^2_{k(bb')_{true}} = \sum_j(\epsilon^b_j - \epsilon^C_j)(\epsilon^{b'}_j - \epsilon^C_j)$ for $b = A, B$ and $b' = A, B$, and $\epsilon^b_j = p_{QTL,b_j}\beta_{j,1} + \left(1 - p_{QTL,b_j}\right)\beta_{j,2}$ for $b =$ A, B, C.

Results

Example I

The pedigree-based partial BS relationship matrices for individuals 7, 8, and 9, which are those with non-zero elements in at least one BS matrix, are in Table 2. For individual 7, which is the offspring of a F1 crossbred dam of breeds A and B, and a purebred sire of breed C, BS is only expected between breeds A and B and therefore all elements related to this individual in the pedigree-based partial relationship matrices for BS between A and C, and B and C, are 0. The genomic BS similarity matrices are in Table 3. The values in the genomic BS similarity matrices deviate from those in the pedigree-based matrices but show the same pattern. For segregation between breeds A and C and between breeds B and C, the diagonal element for individual 7 is 0, while the off-diagonal elements that connect individual 7 to individuals 8 and 9 are non-zero. These matrices are, therefore, not valid relationship matrices.

Table 4 presents the pedigree-based partial LBP relationship matrices for breeds A and B for individuals 7, 8, and 9, as well as the matrix describing the relationship between the LBP effects of breeds A and B. In this example, the matrices for breeds A and B are identical. All elements of the between-breed partial LBP relationship matrix (A-B) are negative. The genomic partial LBP relationship matrices are in Table 5. Similar to the comparison between the genomic and pedigree BS matrices, the elements of the genomic LBP partial relationship matrices deviate from those in the pedigree matrices but show the same pattern. The genomic matrix with relationships of LBP effects between breeds is non-symmetric, in contrast to its pedigree-based counterpart.

Example II

The AI-REML algorithm converged in 10 rounds of iterations. The true and estimated variance components from Example II are presented in Table 6. The within-breed additive genetic variances were slightly smaller than the simulated values for the last historical generation. This reduction in genetic variance is expected because of the small population sizes for the 40 generations that separated the populations, resulting in considerable genetic drift. This genetic drift, however, resulted in substantial segregation between the populations, as reflected in the true LBP variance of up to 0.10 (Population B; Table 6). The true BS variances were 0.04, 0.03, and 0.05 for $\sigma^2_{w(A,B)}$, $\sigma^2_{w(A,C)}$, and $\sigma^2_{w(B,C)}$, respectively. The estimated LBP variances were on the same level as their true values, while the estimated population-specific additive genetic variances were slightly smaller than their true values.

Discussion

In this paper, we present the theoretical background for including LBP effects in genetic and genomic evaluation models for crossbred populations with varying breed compositions based on pedigree information or BOA of SNP genotypes. We also show that LBP and BS account for the same extra additive genetic variance in later generations of crossbreeding compared to the first generation. Furthermore, we provide a method for constructing the genomic BS similarity matrices based on the estimated BOA of genotypes for a general crossbreeding population. However, we found that the genomic BS similarity matrices are not necessarily positive semidefinite and present an alternative method that consists of setting one breed as reference breed and fitting LBP effects of all other breeds, while accounting for the correlation between the LBP effects of different breeds. This model was applied to a simulated dataset.

The results from Example I illustrate the properties of the relationship matrices for a small toy example. Interestingly, the genomic BS similarity matrix between breeds A

Table 6 Estimated and true variance components for the simulated data from Example II

	$\sigma^2_{a(A)}$	$\sigma^2_{a(B)}$	$\sigma^2_{a(C)}$	$\sigma_{k(AA)}$	$\sigma_{k(BB)}$	$\sigma_{k(AB)}$	σ^2_e
True	0.26	0.27	0.23	0.07	0.10	0.04	0.70
Estimated	0.23 ± 0.09	0.21 ± 0.08	0.19 ± 0.05	0.06 ± 0.04	0.11 ± 0.06	0.01 ± 0.04	0.75 ± 0.03

Breed specific additive genetic variance for breed A ($\sigma^2_{a(A)}$), B ($\sigma^2_{a(B)}$), and C ($\sigma^2_{a(C)}$), and local breed proportion variance for breed A ($\sigma_{k(AA)}$) and B ($\sigma_{k(BB)}$), local breed proportion covariance between breed A and B ($\sigma_{k(AB)}$), and residual variance (σ^2_e), with standard errors of the estimates

and B (Table 3) was not positive semi-definite, underlining the limitations of modelling BS using BOA information. The results from Example II demonstrate the ability of the genomic LBP model to capture the effects of segregation between breeds in crossbred populations. The simulated scenario was intentionally designed to obtain large LBP variances, with multiple generations since the separation of the breeds, a small size of the purebred populations, and relatively few (500) QTL affecting the trait, and this was evident from the results. Furthermore, the crossbreeding structure with both parents of all individuals in generations 2 and 3 of population D being crossbred, facilitated estimation of LBP effects. In addition, the half-sib and full-sib progeny from crossbred parents are expected to increase the LBP relationships between individuals. Still, the fast convergence of the AI-REML algorithm and variance estimates that were close to their true variances, for a data set with phenotypes of only 1500 individuals, supports that the model will be able to disentangle the contribution of breed-specific and LBP effects to the phenotypes of crossbreds in real data sets also.

The models proposed in this paper are relevant for genetic and genomic prediction based on data from more than one generation of crossbreeding, for which the breeding values of crossbreds are partitioned into breed-specific terms [6]. Previous studies [11–13] have presented BOA models with BS terms for specific cases only [11, 12]. The BS terms presented in the supplementary material of Aase et al. [13] are equivalent to our genomic BOA model with BS, and thus, suffer from the same problem of matrices that may be not positive semi-definite. The way we propose to account for segregation between breeds, i.e. with LBP effects, can be applied to any number of breeds or breed composition and is, therefore, more general than the methods presented by Christensen et al. [11] and Rio et al. [12].

Compared to the pedigree-based model of García-Cortés and Toro [6], for the models with LBP effects, a complication for estimation of variance components is that the LBP (co)variance structure cannot be directly expressed as a Kronecker product. Therefore, the model must be extended with artificial effects (Eq. 21) for variance component estimation. In Example II, we successfully estimated the variance components using AI-REML implemented in a commonly used mixed model software for genetic evaluation, using such an extension. Therefore, the need to extend the models with artificial effects appears to be only a minor complication. When the variance components are assumed to be known and only the effects need to be predicted, the extension can be avoided by multiplying the parts of $\mathbf{Q}^{b,b'}$ relating to each breed, or to each pair of breeds, with the appropriate (co)variance estimates, as shown in Eq. (20).

In a terminal crossbreeding system, the crossbred animals are not potential breeding stock and estimates of the $\widetilde{\mathbf{k}}^{b}$ or $\mathbf{w}^{b,b'}$ terms will therefore not be a part of the predicted breeding values required for selection, but they are used only to reduce the residual error when including records on crossbreds to improve the prediction for purebred selection candidates. Thus, ignoring BS or LBP, as was done in the literature [8, 16, 22, 23], is unlikely to have a considerable impact on selection among purebred selection candidates, even if there was significant BS. When there is selection among crossbred animals, which can occur in, e.g. rotational crossbreeding systems, the BS terms are part of the breeding values of the crossbred individuals. In this case, models with or without a BS term could lead to different selection decisions. However, models for genomic evaluation in a rotational crossbreeding system of three breeds have excluded BS and LBP effects in previous studies on simulated data [12, 13]. The same is true for studies with BOA models applied to real data for admixed populations [24] or populations with crossbreeding involving varying breed combinations [18]. For such scenarios, the LBP model presented here could improve predictions of breeding values if there was significant BS, which would need investigation. Although the focus of this paper is on segregation between breeds in a livestock context, the developed models may also be applicable to model segregation between genetic groups in plant breeding [12].

For the pedigree-based model, we have not found a simple algorithm for the construction of the generalized inverses of the partial LBP relationship matrices when covariances between LBP effects for different breeds are included. Therefore, inclusion of BS effects, as in García-Cortés and Toro [6], is more attractive for pedigree-based models than for the pedigree-based LBP model. However, for single-step genomic models [2, 3], phenotypes for both genotyped and non-genotyped individuals are included. For such models, compatible pedigree and genomic relationship matrices are important [25]. Christensen et al. [11] presented a model for three-way terminal crossbreeding with a combined genomic and pedigree-based partial BS relationship matrix. However, for more complex crossbreeding scenarios, the genomic BS similarity matrix may not be positive semi-definite. Therefore, single-step models for complex crossbreeding should include LBP effects rather than BS and, thus, pedigree-based LBP matrices are needed. Values for the pedigree and genomic partial LBP relationship matrices presented here have the same expectation if pedigree information is complete, i.e. when the pedigree of all crossbred individuals can be traced back to purebred ancestors and BOA assignment is complete, i.e. all alleles are assigned breed origin. Compatibility of the partial LBP matrices is demonstrated in the results of

Example I, Tables 4 and 5. The genomic and pedigree partial LBP relationship matrices can therefore be combined in a similar manner as in Christensen et al. [11], but single-step models with LBP effects need further investigation.

For a simulated data set in a two-breed rotational crossbreeding system, Poulsen et al. [26] tested different relationship matrices and found that partitioning the breeding value as García-Cortés and Toro [6] performed similar to the metafounder approach [27] and outperformed other options. Both these approaches include BS. No genotypes of crossbreds were included in the study by Poulsen et al. [26]. Further comparison of models for rotational crossbreeding could include genotyped crossbreds and account for BS or LBP effects using the methods presented in this paper.

For practical genetic and genomic evaluation, the relevance of LBP or BS terms depends on the magnitude of the LBP or BS variance in the population analysed. Such estimates are rare in the literature. Eiríksson et al. [14] found small but statistically significant LBP variance for milk production traits in crossbred dairy cows. Munilla Leguizamón and Cantet [28] did not find significant BS variance for weaning weight in Angus×Hereford crossbred beef cattle. However, Birchmeier et al. [29] found significant BS variance for birth weight in crossbred beef cattle. Here, we have shown that LBP effects can be modelled based on assigned BOA in genomic models for data with complicated crossbreeding structures. Although BOA assignment is generally not complete and contains errors [10, 18, 30], BOA information should provide a more precise estimate of LBP or BS relationships than pedigree-based models. Therefore, the models developed here should facilitate more accurate estimation of BS or LBP variances than the previous pedigree-based models and open the possibility of accounting for BS or LBP in genomic predictions for crossbred populations with BOA models, regardless of the crossbreeding structure.

Based on the results presented here, combined with previous work on the inclusion of BS effects in genetic models, we recommend that for pedigree-based models, BS terms are included following García-Cortés and Toro [6]. For genomic BOA models applied to data that include segregation for two breeds only, either LBP or BS effects can be included. However, for genomic BOA models applied to data that includes segregation for more than one pair of breeds, LBP terms should be used, as presented here, rather than BS terms.

Conclusions

Models for the genetic evaluation using crossbred data that partition the genetic value into breed-specific terms should, in theory, account for the effects of LBP or BS. The (co)variance structure for BS effects in genomic BOA models for crossbred data involving more than two breeds are not guaranteed to lead to relationship matrices that are positive semi-definite. The LBP (co)variance structure can be constructed from the pedigree or BOA information by including LBP effects for each breed except for a reference breed, given that the covariance between LBP effects of each pair of breeds is included in the model.

Appendices

Appendix 1

Here, we derive the formulas for breed-segregation (BS) variance and covariance for a single locus in Eqs. (7) and (8), respectively. The contribution of locus j to the BS variance of individual i, a crossbred of n_b breeds, is $Var\left(\sum_b \epsilon_j^b \ddot{z}_{ij}^b\right) = E\left[\left(\sum_b \epsilon_j^b \ddot{z}_{ij}^b\right)^2\right]$. Furthermore,

$$\left(\sum_b \epsilon_j^b \ddot{z}_{ij}^b\right)^2 = \sum_b \epsilon_j^{b2} \ddot{z}_{ij}^{b2} + \sum_b \sum_{b'>b} 2\epsilon_j^b \epsilon_j^{b'} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'}.$$

Using $\sum_b \ddot{z}_{ij}^b = 0$, we get $\left(\sum_b \epsilon_j^b \ddot{z}_{ij}^b\right)^2 = \sum_b \sum_{b' \neq b} -\epsilon_j^{b2} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} + \sum_b \sum_{b'>b} 2\epsilon_j^b \epsilon_j^{b'} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'}$.

Rearranging, gives:

$$\left(\sum_b \epsilon_j^b \ddot{z}_{ij}^b\right)^2 = \sum_b \left(\sum_{b'>b} -\epsilon_j^{b2} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} + \sum_{b'>b} -\epsilon_j^{b'2} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} \sum_{b'>b} 2\epsilon_j^b \epsilon_j^{b'} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'}\right)$$
$$= \sum_b \sum_{b'>b} -\ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} \left(\epsilon_j^{b2} - 2\epsilon_j^b \epsilon_j^{b'} + \epsilon_j^{b'2}\right) = \sum_b \sum_{b'>b} -\ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} \left(\epsilon_j^b - \epsilon_j^{b'}\right)^2.$$

Therefore, $Var\left(\sum_b \epsilon_j^b \ddot{z}_{ij}^b\right) = \sum_b \sum_{b'>b} -\ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} E\left[\left(\epsilon_j^b - \epsilon_j^{b'}\right)^2\right]$.

The contribution of locus j to the BS covariance between animals i and i' is: $Cov\left(\sum_b \epsilon_j^b \ddot{z}_{ij}^b, \sum_b \epsilon_j^b \ddot{z}_{i'j}^b\right) = E\left[\left(\sum_b \epsilon_j^b \ddot{z}_{ij}^b\right)\left(\sum_b \epsilon_j^b \ddot{z}_{i'j}^b\right)\right]$. Furthermore, $\left(\sum_b \epsilon_j^b \ddot{z}_{ij}^b\right)\left(\sum_b \epsilon_j^b \ddot{z}_{i'j}^b\right) = \sum_b \epsilon_j^{b2} \ddot{z}_{ij}^b \ddot{z}_{i'j}^b + \sum_b \sum_{b'>b} \epsilon_j^b \epsilon_j^{b'} \ddot{z}_{ij}^b \ddot{z}_{i'j}^{b'} + \sum_b \sum_{b'>b} \epsilon_j^b \epsilon_j^{b'} \ddot{z}_{i'j}^b \ddot{z}_{ij}^{b'}$.

Substituting $\epsilon_j^{b2} \ddot{z}_{ij}^b \ddot{z}_{i'j}^b$ with $\sum_{b' \neq b} -\epsilon_j^{b2} \ddot{z}_{ij}^b \ddot{z}_{i'j}^{b'}$ and rearranging,

$$\left(\sum_b \epsilon_j^b \ddot{z}_{ij}^b\right)\left(\sum_b \epsilon_j^b \ddot{z}_{i'j}^b\right) = \sum_b \sum_{b' \neq b} -\epsilon_j^{b2} \ddot{z}_{ij}^b \ddot{z}_{i'j}^{b'} + \sum_b \sum_{b'>b} \epsilon_j^b \epsilon_j^{b'} \ddot{z}_{ij}^b \ddot{z}_{i'j}^{b'} + \sum_b \sum_{b'>b} \epsilon_j^b \epsilon_j^{b'} \ddot{z}_{i'j}^b \ddot{z}_{ij}^{b'}$$
$$= \sum_b \left(\sum_{b'>b} -\epsilon_j^{b2} \ddot{z}_{ij}^b \ddot{z}_{i'j}^{b'} + \sum_{b'>b} -\epsilon_j^{b'2} \ddot{z}_{ij}^b \ddot{z}_{i'j}^{b'} + \sum_{b'>b} \epsilon_j^b \epsilon_j^{b'} \ddot{z}_{ij}^{b'} \ddot{z}_{i'j}^{b'} + \sum_{b'>b} \epsilon_j^b \epsilon_j^{b'} \ddot{z}_{ij}^{b'} \ddot{z}_{i'j}^b \right).$$

Furthermore,

$$\left(\sum_b \epsilon_j^b \ddot{z}_{ij}^b\right)\left(\sum_b \epsilon_j^b \ddot{z}_{i'}^b\right) = \sum_b \sum_{b'>b} -\frac{1}{2} \ddot{z}_{ij}^b \ddot{z}_{i'}^{b'} \left(\epsilon_j^{b2} - 2\epsilon_j^b \epsilon_j^{b'} + \epsilon_j^{b'2}\right) + \sum_b \sum_{b'>b} -\frac{1}{2} \ddot{z}_{ij}^{b'} \ddot{z}_{i'}^b \left(\epsilon_j^{b2} - 2\epsilon_j^b \epsilon_j^{b'} + \epsilon_j^{b'2}\right)$$
$$= \sum_b \sum_{b'>b} \left(-\ddot{z}_{ij}^b \ddot{z}_{i'}^{b'} - \ddot{z}_{ij}^{b'} \ddot{z}_{i'}^b\right) \frac{1}{2} \left(\epsilon_j^b - \epsilon_j^{b'}\right)^2.$$

Therefore, $Cov\left(\sum_b \epsilon_j^b \ddot{z}_{ij}^b, \sum_b \epsilon_j^b \ddot{z}_{i'j}^b\right) = \sum_b \sum_{b'>b} \left(-\ddot{z}_{ij}^b \ddot{z}_{i'}^{b'} - \ddot{z}_{ij}^{b'} \ddot{z}_{i'j}^b\right) \frac{1}{2} E\left[\left(\epsilon_j^b - \epsilon_j^{b'}\right)^2\right]$, which gives the expression in Eq. (8) in the main text.

Appendix 2

In Appendix 2, we investigate the modelling of three independent LBP effects in relation to the presented theory of BS. We assume that the marker loci are the QTL for simplicity, the origin of the alleles is known, and ϵ_j^b are random unknown variables with mean zero.

We name the three breeds as A, B and C. The genomic LBP variance for three breeds related to locus *j* for animal *i* is:

$$\mathrm{Var}\left(\ddot{\kappa}_{ij}^{sum}\right) = E\left[\left(\epsilon_j^A \ddot{z}_{ij}^A + \epsilon_j^B \ddot{z}_{ij}^B + \epsilon_j^C \ddot{z}_{ij}^C\right)^2\right].$$

Furthermore,

$$\left(\varepsilon_j^A \ddot{z}_{ij}^A + \varepsilon_j^B \ddot{z}_{ij}^B + \varepsilon_j^C \ddot{z}_{ij}^C\right)^2 = \varepsilon_j^{A2} \ddot{z}_{ij}^{A2} + \varepsilon_j^{B2} \ddot{z}_{ij}^{B2} + \varepsilon_j^{C2} \ddot{z}_{ij}^{C2} + 2\varepsilon_j^A \varepsilon_j^B \ddot{z}_{ij}^A \ddot{z}_{ij}^B + 2\varepsilon_j^A \varepsilon_j^C \ddot{z}_{ij}^A \ddot{z}_{ij}^C + 2\varepsilon_j^B \varepsilon_j^C \ddot{z}_{ij}^B \ddot{z}_{ij}^C. \quad (33)$$

First, we show that when $z^A + z^B + z^C = 0$,

$$2z^A z^B = z^{C2} - z^{A2} - z^{B2}. \quad (34)$$

The proof is as follows. Note that $z^A = -z^B - z^C$ and $z^B = -z^A - z^C$. This leads to $2z^A z^B = z^A\left(-z^A - z^C\right) + z^B\left(-z^B - z^C\right)$. Rearranging gives $2z^A z^B = -z^{A2} - z^{B2} - z^C\left(z^A + z^B\right) = z^{C2} - z^{A2} - z^{B2}$ which completes the proof.

Substituting $2z^A z^B$ with $z^{C2} - z^{A2} - z^{B2}$ and similarly with the other pairs of breeds in Eq. (33) gives:

$$\left(\epsilon_j^A \ddot{z}_{ij}^A + \epsilon_j^B \ddot{z}_{ij}^B + \epsilon_j^C \ddot{z}_{ij}^C\right)^2 = \epsilon_j^{A2} \ddot{z}_{ij}^{A2} + \epsilon_j^{B2} \ddot{z}_{ij}^{B2} + \epsilon_j^{C2} \ddot{z}_{ij}^{C2} + \epsilon_j^A \epsilon_j^B \ddot{z}_{ij}^{C2} + \epsilon_j^A \epsilon_j^C \ddot{z}_{ij}^{B2} + \epsilon_j^B \epsilon_j^C \ddot{z}_{ij}^{A2} - \epsilon_j^A \epsilon_j^B \ddot{z}_{ij}^{A2} - \epsilon_j^A \epsilon_j^C \ddot{z}_{ij}^{A2} - \epsilon_j^B \epsilon_j^A \ddot{z}_{ij}^{B2} - \epsilon_j^B \epsilon_j^C \ddot{z}_{ij}^{B2} - \epsilon_j^C \epsilon_j^A \ddot{z}_{ij}^{C2} - \epsilon_j^C \epsilon_j^B \ddot{z}_{ij}^{C2}.$$

Some rearranging gives:

$$\left(\epsilon_j^A \ddot{z}_{ij}^A + \epsilon_j^B \ddot{z}_{ij}^B + \epsilon_j^C \ddot{z}_{ij}^C\right)^2 = \ddot{z}_{ij}^{A2} \left(\epsilon_j^A - \epsilon_j^B\right)\left(\epsilon_j^A - \epsilon_j^C\right) + \ddot{z}_{ij}^{B2} \left(\epsilon_j^B - \epsilon_j^A\right)\left(\epsilon_j^B - \epsilon_j^C\right) + \ddot{z}_{ij}^{C2} \left(\epsilon_j^C - \epsilon_j^A\right)\left(\epsilon_j^C - \epsilon_j^B\right).$$

Therefore,

$$\begin{aligned}\mathrm{Var}\left(\ddot{\kappa}_{ij}^{sum}\right) = \ddot{z}_{ij}^{A^2} E\left[\left(\epsilon_j^A - \epsilon_j^B\right)\left(\epsilon_j^A - \epsilon_j^C\right)\right] \\ + \ddot{z}_{ij}^{B^2} E\left[\left(\epsilon_j^B - \epsilon_j^A\right)\left(\epsilon_j^B - \epsilon_j^C\right)\right] + \ddot{z}_{ij}^{C^2} E\left[\left(\epsilon_j^C - \epsilon_j^A\right)\left(\epsilon_j^C - \epsilon_j^B\right)\right].\end{aligned}$$

This decomposition of $\mathrm{Var}\left(\ddot{\kappa}_{ij}^{sum}\right)$ gives a sum over three independent LBP terms and could be extended to a relationship matrix across multiple individuals. However, the $E\left[\left(\epsilon_j^A - \epsilon_j^B\right)\left(\epsilon_j^A - \epsilon_j^C\right)\right]$ term is not guaranteed to be positive. Therefore, the variance for the whole genome, $\sum_j E\left[\left(\epsilon_j^A - \epsilon_j^B\right)\left(\epsilon_j^A - \epsilon_j^C\right)\right]$ is not necessarily non-negative, and therefore not a valid variance term, and the same holds for other combinations of breeds.

Appendix 3

Here, we present the derivation of Eq. (11). The variance for LBP over all breeds b for locus j is:

$$Var\left(\ddot{\kappa}_{ij}^{sum}\right) = E\left[\left(\sum_b \ddot{z}_{ij}^b \epsilon_j^b\right)^2\right].$$

We set one breed as reference breed, b^*. Next, for breeds $b \neq b^*$, and $b' \neq b^*$, we substitute $\epsilon_j^b = \epsilon_j^{b^*} + \epsilon_j^b - \epsilon_j^{b^*}$ and $\epsilon_j^{b'} = \epsilon_j^{b^*} + \epsilon_j^{b'} - \epsilon_j^{b^*}$, respectively, and thus get:

$$\begin{aligned}\left(\sum_b \ddot{z}_{ij}^b \epsilon_j^b\right)^2 &= \sum_b \sum_{b'} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} \epsilon_j^b \epsilon_j^{b'} \\ &= \sum_{b\neq b^*} \sum_{b'\neq b^*} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} \left(\epsilon_j^{b^*} + \epsilon_j^b - \epsilon_j^{b^*}\right)\left(\epsilon_j^{b^*} + \epsilon_j^{b'} - \epsilon_j^{b^*}\right) \\ &\quad + 2\sum_{b\neq b^*} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b^*} \epsilon_j^b \epsilon_j^{b^*} + \ddot{z}_{ij}^{b^{*2}} \epsilon_j^{b^{*2}}.\end{aligned}$$

Since:

$$\begin{aligned}\left(\epsilon_j^{b^*} + \epsilon_j^b - \epsilon_j^{b^*}\right)\left(\epsilon_j^{b^*} + \epsilon_j^{b'} - \epsilon_j^{b^*}\right) = -\epsilon_j^{b^{*2}} \\ + \epsilon_j^b \epsilon_j^{b^*} + \epsilon_j^{b'} \epsilon_j^{b^*} + \left(\epsilon_j^{b'} - \epsilon_j^{b^*}\right)\left(\epsilon_j^b - \epsilon_j^{b^*}\right),\end{aligned} \tag{35}$$

Then:

$$\begin{aligned}\left(\sum_b \ddot{z}_{ij}^b \epsilon_j^b\right)^2 = \sum_{b\neq b^*} \sum_{b'\neq b^*} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} \left(\epsilon_j^{b'} - \epsilon_j^{b^*}\right)\left(\epsilon_j^b - \epsilon_j^{b^*}\right) - \sum_{b\neq b^*} \sum_{b'\neq b^*} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} \epsilon_j^{b^{*2}} \\ + 2\sum_{b\neq b^*} \sum_{b'\neq b^*} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} \epsilon_j^{b'} \epsilon_j^{b^*} + 2\sum_{b\neq b^*} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b^*} \epsilon_j^b \epsilon_j^{b^*} + \ddot{z}_{ij}^{b^*2} \epsilon_j^{b^{*2}}.\end{aligned}$$

Using that $\sum_{b\neq b^*} \ddot{z}_{ij}^b = -\ddot{z}_{ij}^{b^*}$ and $\sum_{b'\neq b^*} \ddot{z}_{ij}^{b'} = -\ddot{z}_{ij}^{b^*}$ on the second term, we see that the second and the fifth terms cancel out, and using that $\sum_{b\neq b^*} \ddot{z}_{ij}^b = -\ddot{z}_{ij}^{b^*}$ on the third term, we see that the third and the fourth terms cancel out. Therefore, $\left(\sum_b \ddot{z}_{ij}^b \epsilon_j^b\right)^2 = \sum_{b\neq b^*} \sum_{b'\neq b^*} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} \left(\epsilon_j^b - \epsilon_j^{b^*}\right)\left(\epsilon_j^{b'} - \epsilon_j^{b^*}\right)$, and finally, $Var\left(\ddot{\kappa}_{ij}^{sum}\right) = \sum_{b\neq b^*} \sum_{b'\neq b^*} \ddot{z}_{ij}^b \ddot{z}_{ij}^{b'} E\left[\left(\epsilon_j^b - \epsilon_j^{b^*}\right)\left(\epsilon_j^{b'} - \epsilon_j^{b^*}\right)\right]$, which is Eq. (11).

Equation (12) about the covariance between the LBP effects of individuals i and i' over all breeds b for locus j is derived similarly to the above derivation of Eq. (11), and we obtain:

$$Cov\left(\ddot{\kappa}_{ij}^{sum}, \ddot{\kappa}_{i'j}^{sum}\right) = \sum_{b\neq b^*} \sum_{b'\neq b^*} \ddot{z}_{ij}^b \ddot{z}_{i'j}^{b'} E\left[\left(\epsilon_j^b - \epsilon_j^{b^*}\right)\left(\epsilon_j^{b'} - \epsilon_j^{b^*}\right)\right].$$

Appendix 4

Here, we derive Eqs. (22) and (23), with the expectation of pedigree-based LBP variance and covariance for single loci, respectively. In a pedigree-based LBP model, we have the expectation of LBP variance as:

$$\begin{aligned}E\left[\mathrm{Var}\left(\kappa_{s_i^j}^{sum}\right)\right] &= \sum_b f_{s_i}^b \epsilon_j^{b^2} - \left(\sum_b f_{s_i}^b \epsilon_j^b\right)^2 \\ &= \sum_b f_{s_i}^b \epsilon_j^{b^2} - \sum_b \sum_{b'} f_{s_i}^b f_{s_i}^{b'} \epsilon_j^b \epsilon_j^{b'}.\end{aligned}$$

We set one breed as reference breed, b^*. Then for breeds $b \neq b^*$ and $b' \neq b^*$, we do the substitutions $\epsilon_j^b = \epsilon_j^{b^*} + \epsilon_j^b - \epsilon_j^{b^*}$, and $\epsilon_j^{b'} = \epsilon_j^{b^*} + \epsilon_j^{b'} - \epsilon_j^{b^*}$, and we get:

$$
\begin{aligned}
E\left[\mathrm{Var}\left(\kappa_{s_i}^{j\,sum}\right)\right] = & \sum_{b\neq b^*} f_{s_i}^{b}\left(\epsilon_j^{b^*}+\epsilon_j^{b}-\epsilon_j^{b^*}\right)^2 + f_{s_i}^{b^*}\epsilon_j^{b^{*2}} \\
& - \sum_{b\neq b^*}\sum_{b'\neq b^*} f_{s_i}^{b} f_{s_i}^{b'}\left(\epsilon_j^{b^*}+\epsilon_j^{b}-\epsilon_j^{b^*}\right)\left(\epsilon_j^{b^*}+\epsilon_j^{b'}-\epsilon_j^{b^*}\right) \\
& - 2\sum_{b\neq b^*} f_{s_i}^{b} f_{s_i}^{b^*}\epsilon_j^{b}\epsilon_j^{b^*} - f_{s_i}^{b^{*2}}\epsilon_j^{b^{*2}}.
\end{aligned}
$$

Using Eq. (35), then:

$$
\begin{aligned}
E\left[\mathrm{Var}\left(\kappa_{s_i}^{j\,sum}\right)\right] = & 2\sum_{b\neq b^*} f_{s_i}^{b}\epsilon_j^{b}\epsilon_j^{b^*} - \sum_{b\neq b^*} f_{s_i}^{b}\epsilon_j^{b^{*2}} \\
& + \sum_{b\neq b^*}\left[f_{s_i}^{b}\left(\epsilon_j^{b}-\epsilon_j^{b^*}\right)^2\right] + f_{s_i}^{b^*}\epsilon_j^{b^{*2}} + \sum_{b\neq b^*}\sum_{b'\neq b^*} f_{s_i}^{b} f_{s_i}^{b'}\epsilon_j^{b^{*2}} \\
& - 2\sum_{b\neq b^*}\sum_{b'\neq b^*} f_{s_i}^{b} f_{s_i}^{b'}\epsilon_j^{b}\epsilon_j^{b^*} - \sum_{b\neq b^*}\sum_{b'\neq b^*}\left[f_{s_i}^{b} f_{s_i}^{b'}\left(\epsilon_j^{b'}-\epsilon_j^{b^*}\right)\left(\epsilon_j^{b}-\epsilon_j^{b^*}\right)\right] \\
& - 2\sum_{b\neq b^*} f_{s_i}^{b} f_{s_i}^{b^*}\epsilon_j^{b}\epsilon_j^{b^*} - f_{s_i}^{b^{*2}}\epsilon_j^{b^{*2}}.
\end{aligned}
$$

Using that $\sum_{b\neq b^*} f_{s_i}^{b} = 1 - f_{s_i}^{b^*}$ on the fifth, sixth and last terms, we have:

$$
\begin{aligned}
E\left[\mathrm{Var}\left(\kappa_{s_i}^{j\,sum}\right)\right] = & 2\sum_{b\neq b^*} f_{s_i}^{b}\epsilon_j^{b}\epsilon_j^{b^*} - \sum_{b\neq b^*} f_{s_i}^{b}\epsilon_j^{b^{*2}} \\
& + \sum_{b\neq b^*}\left[f_{s_i}^{b}\left(\epsilon_j^{b}-\epsilon_j^{b^*}\right)^2\right] + f_{s_i}^{b^*}\epsilon_j^{b^{*2}} + \sum_{b\neq b^*} f_{s_i}^{b}\epsilon_j^{b^{*2}} \\
& - \sum_{b\neq b^*} f_{s_i}^{b} f_{s_i}^{b^*}\epsilon_j^{b^{*2}} - 2\sum_{b\neq b^*} f_{s_i}^{b}\epsilon_j^{b}\epsilon_j^{b^*} + 2\sum_{b\neq b^*} f_{s_i}^{b} f_{s_i}^{b^*}\epsilon_j^{b}\epsilon_j^{b^*} \\
& - \sum_{b\neq b^*}\sum_{b'\neq b^*}\left[f_{s_i}^{b} f_{s_i}^{b'}\left(\epsilon_j^{b'}-\epsilon_j^{b^*}\right)\left(\epsilon_j^{b}-\epsilon_j^{b^*}\right)\right] \\
& - 2\sum_{b\neq b^*} f_{s_i}^{b} f_{s_i}^{b^*}\epsilon_j^{b}\epsilon_j^{b^*} - f_{s_i}^{b^*}\epsilon_j^{b^{*2}} + \sum_{b\neq b^*} f_{s_i}^{b} f_{s_i}^{b^*}\epsilon_j^{b^{*2}}.
\end{aligned}
$$

Removing the terms that cancel out,

$$
E\left[\mathrm{Var}\left(\kappa_{s_i}^{j\,sum}\right)\right] = \sum_{b\neq b^*}\left[f_{s_i}^{b}\left(\epsilon_j^{b}-\epsilon_j^{b^*}\right)^2\right] - \sum_{b\neq b^*}\sum_{b'\neq b^*}\left[f_{s_i}^{b} f_{s_i}^{b'}\left(\epsilon_j^{b'}-\epsilon_j^{b^*}\right)\left(\epsilon_j^{b}-\epsilon_j^{b^*}\right)\right].
$$

Therefore, we have $n_b - 1$ LBP variance terms (for $b = b'$),

$$
\sum_{b\neq b^*}\left[f_{s_i}^{b}\left(\epsilon_j^{b}-\epsilon_j^{b^*}\right)^2\right] - \sum_{b\neq b^*}\left[f_{s_i}^{b^2}\left(\epsilon_j^{b}-\epsilon_j^{b^*}\right)^2\right] = \sum_{b\neq b^*}\left[f_{s_i}^{b}\left(1-f_{s_i}^{b}\right)\left(\epsilon_j^{b}-\epsilon_j^{b^*}\right)^2\right],
$$

and $(n_b - 1)^2$ covariances between breeds b and b',

$$\sum_{b \neq b^*} \sum_{b' \neq \{b, b^*\}} \left[-f_{s_i}^b f_{s_i}^{b'} \left(\epsilon_j^{b'} - \epsilon_j^{b^*} \right) \left(\epsilon_j^b - \epsilon_j^{b^*} \right) \right].$$

Appendix 5

Here, we show the relationship between BS variances and LBP variances in Eq. (32) for the variance of BS between breeds b and b' when neither of the breeds is the reference breed, b^*. The BS variance is $\sigma^2_{w(b,b')} = \frac{1}{2} \sum_{j=1}^{m} \left(\epsilon_j^b - \epsilon_j^{b'} \right)^2$. Substituting $\epsilon_j^b - \epsilon_j^{b'}$ with $(\epsilon_j^b - \epsilon_j^{b^*}) - (\epsilon_j^{b'} - \epsilon_j^{b^*})$ and rearranging gives:

$$\sigma^2_{w(b,b')} = \frac{1}{2} \sum_{j=1}^{m} \left(\epsilon_j^b - \epsilon_j^{b^*} \right)^2 + \frac{1}{2} \sum_{j=1}^{m} \left(\epsilon_j^{b'} - \epsilon_j^{b^*} \right)^2 - \sum_{j=1}^{m} \left(\epsilon_j^b - \epsilon_j^{b^*} \right) \left(\epsilon_j^{b'} - \epsilon_j^{b^*} \right) = \frac{1}{2} \sigma_{k(bb)} + \frac{1}{2} \sigma_{k(b'b')} - \sigma_{k(bb')}$$

Furthermore, here we derive how the covariance between LBP effects of breeds b and b' can be calculated from BS variance components. We have the covariance between LBP effects across breeds: $\sigma_{k(bb')} = \sum_{j=1}^{m} \left(\epsilon_j^b - \epsilon_j^{b^*} \right) \left(\epsilon_j^{b'} - \epsilon_j^{b^*} \right)$. Rearranging gives:

$$\begin{aligned} \sigma_{k(bb')} &= \sum_{j=1}^{m} \left(\epsilon_j^b \epsilon_j^{b'} - \epsilon_j^b \epsilon_j^{b^*} - \epsilon_j^{b'} \epsilon_j^{b^*} + \epsilon_j^{b^{*2}} + \frac{1}{2} \epsilon_j^{b^2} + \frac{1}{2} \epsilon_j^{b'^2} - \frac{1}{2} \epsilon_j^{b^2} - \frac{1}{2} \epsilon_j^{b'^2} \right) \\ &= \frac{1}{2} \sum_{j=1}^{m} \left(\epsilon_j^b - \epsilon_j^{b^*} \right)^2 + \frac{1}{2} \sum_{j=1}^{m} \left(\epsilon_j^{b'} - \epsilon_j^{b^*} \right)^2 - \frac{1}{2} \sum_{j=1}^{m} \left(\epsilon_j^b - \epsilon_j^{b'} \right)^2. \end{aligned}$$

Therefore,

$$\sigma_{k(bb')} = \sigma^2_{w(b,b^*)} + \sigma^2_{w(b',b^*)} - \sigma^2_{w(b,b')}.$$

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12711-023-00810-5.

Additional file 1. It contains the QMSim instruction file for the simulation in Example II.

Additional file 2. It contains the Julia code for the construction and inversion of the relationship matrices used in Example II.

Author contributions
JHE derived the models with significant input from OFC. JHE designed and analysed the two examples and wrote the first draft of the manuscript. OFC supervised the study at all stages. JHE, GS, IS and OFC discussed the results and revised the manuscript. All authors read and approved the final manuscript.

Funding
JHE, GS and OFC acknowledge funding from the green development and demonstration program (GUDP) from the Danish ministry of food, agriculture and fisheries (J. nr. 34009-18-1365). JHE and OFC acknowledge funding from the Danish pig levy fund (Svineafgiftfonden).

Availability of data and materials
Data and scripts from Example II are available from the first author upon reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 26 October 2022 Accepted: 4 May 2023
Published online: 05 July 2023

References

1. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91:4414–23.
2. Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. Genet Sel Evol. 2010;42:2.
3. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J Dairy Sci. 2010;93:743–52.
4. Lo LL, Fernando RL, Grossman M. Covariance between relatives in multibreed populations: additive model. Theor Appl Genet. 1993;87:423–30.
5. Christensen OF, Madsen P, Nielsen B, Su G. Genomic evaluation of both purebred and crossbred performances. Genet Sel Evol. 2014;46:23.
6. García-Cortés LA, Toro MÁ. Multibreed analysis by splitting the breeding values. Genet Sel Evol. 2006;38:601.
7. Strandén I, Mäntysaari EA. Use of random regression model as an alternative for multibreed relationship matrix. J Anim Breed Genet. 2013;130:4–9.
8. Ibáñez-Escriche N, Fernando RL, Toosi A, Dekkers JC. Genomic selection of purebreds for crossbred performance. Genet Sel Evol. 2009;41:12.
9. Karaman E, Su G, Croue I, Lund MS. Genomic prediction using a reference population of multiple pure breeds and admixed individuals. Genet Sel Evol. 2021;53:46.
10. Eiríksson JH, Karaman E, Su G, Christensen OF. Breed of origin of alleles and genomic predictions for crossbred dairy cows. Genet Sel Evol. 2021;53:84.
11. Christensen OF, Legarra A, Lund MS, Su G. Genetic evaluation for three-way crossbreeding. Genet Sel Evol. 2015;47:98.
12. Rio S, Moreau L, Charcosset A, Mary-Huard T. Accounting for group-specific allele effects and admixture in genomic predictions: theory and experimental evaluation in maize. Genetics. 2020;216:27–41.
13. Aase K, Jensen H, Muff S. Genomic estimation of quantitative genetic parameters in wild admixed populations. Methods Eco Evol. 2022;13:1014–26.

14. Eiríksson JH, Strandén I, Su G, Mäntysaari EA, Christensen OF. Local breed proportions and local breed heterozygosity in genomic predictions for crossbred dairy cows. J Dairy Sci. 2022;105:9822–36.
15. Bolormaa S, Hayes BJ, Hawken RJ, Zhang Y, Reverter A, Goddard ME. Detection of chromosome segments of zebu and taurine origin and their effect on beef production and growth. J Anim Sci. 2011;89:2050–60.
16. Sevillano CA, Vandenplas J, Bastiaansen JWM, Bergsma R, Calus MPL. Genomic evaluation for a three-way crossbreeding system considering breed-of-origin of alleles. Genet Sel Evol. 2017;49:75.
17. Quaas RL. Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics. 1976;32:949–53.
18. Eiríksson JH, Byskov K, Su G, Thomasen JR, Christensen OF. Genomic predictions for crossbred dairy cows by combining solutions from purebred evaluation based on breed origin of alleles. J Dairy Sci. 2022;105:5178–91.
19. Sargolzaei M, Schenkel FS. QMSim: a large-scale genome simulator for livestock. Bioinformatics. 2009;25:680–1.
20. Madsen P, Jensen J, Labouriau R, Christensen OF, Sahana G. DMU—a package for analyzing multivariate mixed models in quantitative genetics and genomics. In: Proceedings of the 10th world congress of genetics applied to livestock production: 17–22 August 2014, Vancouver; 2014. p. 18–22.
21. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: a fresh approach to numerical computing. SIAM Rev. 2017;59:65–98.
22. Duenk P, Calus MPL, Wientjes YCJ, Breen VP, Henshall JM, Hawken R, et al. Validation of genomic predictions for body weight in broilers using crossbred information and considering breed-of-origin of alleles. Genet Sel Evol. 2019;51:38.
23. Sevillano CA, Bovenhuis H, Calus MPL. Genomic evaluation for a cross-breeding system implementing breed-of-origin for targeted markers. Front Genet. 2019;10:418.
24. Guillenea A, Su G, Lund MS, Karaman E. Genomic prediction in Nordic Red dairy cattle considering breed origin of alleles. J Dairy Sci. 2022;105:2426–38.
25. Christensen OF. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. Genet Sel Evol. 2012;44:37.
26. Poulsen BG, Ostersen T, Nielsen B, Christensen OF. Predictive performances of animal models using different multibreed relationship matrices in systems with rotational crossbreeding. Genet Sel Evol. 2022;54:25.
27. Legarra A, Christensen OF, Vitezica ZG, Aguilar I, Misztal I. Ancestral relationships using metafounders: finite ancestral populations and across population relationships. Genetics. 2015;200:455–68.
28. Munilla Leguizamón S, Cantet RJ. Equivalence of multibreed animal models and hierarchical Bayes analysis for maternally influenced traits. Genet Sel Evol. 2010;42:20.
29. Birchmeier AN, Cantet RJC, Fernando RL, Morris CA, Holgado F, Jara A, Santos Cristal M. Estimation of segregation variance for birth weight in beef cattle. Livest Prod Sci. 2002;76:27–35.
30. Vandenplas J, Calus MPL, Sevillano CA, Windig JJ, Bastiaansen JWM. Assigning breed origin to alleles in crossbred animals. Genet Sel Evol. 2016;48:61.

Publisher's Note