GSE **G**enetics **S**election **E**volution

RESEARCH ARTICLE

Causal inference for the covariance between breeding values under identity disequilibrium

Rodolfo J. C. Cantet[1,2]*, Belcy K. Angarita-Barajas[2], Natalia S. Forneris[1,2] and Sebastián Munilla[1,2]

Abstract

Background: The covariance matrix of breeding values is at the heart of prediction methods. Prediction of breeding values can be formulated using either an "observed" or a theoretical covariance matrix, and a major argument for choosing one or the other is the reduction of the computational burden for inverting such a matrix. In this regard, covariance matrices that are derived from Markov causal models possess properties that deliver sparse inverses.

Results: By using causal Markov models, we express the breeding value of an individual as a linear regression on ancestral breeding values, plus a residual term, which we call residual breeding value (RBV). The latter is a noise term that accounts for the uncertainty in prediction due to lack of fit of the linear regression. A notable property of these models is the parental Markov condition, through which the multivariate distribution of breeding values is uniquely determined by the distribution of the mutually independent RBV. Animal breeders have long been relying on a causal Markov model, while using the additive relationship matrix as the covariance matrix structure of breeding values, which is calculated assuming gametic equilibrium. However, additional covariances among breeding values arise due to identity disequilibrium, which is defined as the difference between the covariance matrix under the multi-loci probability of identity-by-descent ($\boldsymbol{\Sigma}$) and its *expectation* under gametic phase equilibrium, i.e., **A**. The disequilibrium term $\boldsymbol{\Sigma} - \mathbf{A}$ is considered in the model for predicting breeding values called the "ancestral regression" (AR), a causal Markov model. Here, we introduce the "ancestral regression to parents" (PAR) causal Markov model, which reduces the computational burden of the AR approach. By taking advantage of the conditional independence property of the PAR Markov model, we derive covariances between the breeding values of grandparents and grand-offspring and between parents and offspring. In addition, we obtain analytical expressions for the covariance between collateral relatives under the PAR model, as well as for the inbreeding coefficient.

Conclusions: We introduced the causal PAR Markov model that captures identity disequilibrium in the covariances among breeding values and produces a sparse inverse covariance matrix to build and solve a set of mixed model equations.

Background

In genetic evaluations, the best linear unbiased predictor (BLUP, [1]) is used to predict the breeding values (BV) of selection candidates. Breeding values are a presumed compound random variable that arises from the effect of a large number of genes in gametic equilibrium and follows a multivariate normal distribution across the population [1–3]. Prediction of BV can be formulated using

*Correspondence: rcantet@agro.uba.ar
[1] Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires, 1417 Ciudad Autónoma de Buenos Aires, Argentina
Full list of author information is available at the end of the article

Cantet *et al. Genetics Selection Evolution*   (2022) 54:64

Page 2 of 14

either an "observed" or a theoretical covariance matrix of BV. The latter model takes advantage from the parental Markov condition of the covariance between BV of the parents and their offspring [4] and theorem 1.2.7 in [5]. Throughout this paper, we will refer to these models as recursive causal Markov models or Markov models for short.

In Markov models, the BV of an individual is expressed as a linear regression on parental, or parental and grandparental BV, plus a residual breeding value (RBV) term. RBV is a noise term that accounts for the lack of fit of the regression with respect to the statistical information that the ancestors convey to predict the BV of an individual. The parental Markov condition implies that the multivariate distribution of BV is uniquely determined by the distribution of these RBV, which are mutually independent random variables. This is a familiar property of a Markov model well-known to animal breeders: the regular Animal Model [1] where the RBV are termed "Mendelian sampling terms".

The theoretical covariance matrix of a Markov model describes the sequences of BV across generations as a stochastic process, which is represented by the regression of the BV of an individual on ancestral BV [1, 6]. Starting with the additive relationship matrix as the covariance structure for BV under the regular Animal Model, more informative and realistic Markov models should account for the different proportions of the genomes from the grandparents in the gametes that form the individual, which originate at meiosis due to recombination. We will show that this can be accomplished by extending the concept of identity disequilibrium of Weir and Cockerham [7]. Identity disequilibrium is accounted for in the ancestral regression (AR) Markov model of Cantet et al. [8].

In the original presentation of the AR model, we came short of deriving the covariances between the BV of relatives. Therefore, our objective is threefold. The first goal is to present a causal Markov model that we refer to as the ancestral regression to parents (PAR) and that simplifies inversion of the covariance matrix of BV when the AR model is correct. The second objective is to give explicit expressions for the covariances between the BV of an animal with those of its grandparents and parents under identity disequilibrium. Finally, we derive the covariance of the BV of an individual with its collaterals (full and half-sibs) assuming that PAR is correct. These latter covariances are needed for genomic prediction of animals without an own record and without progeny. However, issues with parameter estimation and how to specifically use the AR and PAR models for genomic prediction will be pursued elsewhere. In the three additional files of this paper, we describe in detail some theoretical arguments and all the derivations.

Methods

Covariance between breeding values under segmental inheritance

The covariance matrix of breeding values (BV) has traditionally been calculated by conditioning on relatedness [1]. In this section, we extend the definition of genetic relationships at one locus to the multilocus situation with segmental inheritance and obtain the parameters that enter into the covariance of BV from the recursive regression of the BV of an animal on the parental and grandparental breeding values [8].

Covariance between BV under gametic phase equilibrium

Under the assumption of gametic phase equilibrium and for any two animals X and Y that share one or more common ancestors, Cockerham [9], Kempthorne [10], and Harris [11] obtained the expression:

$$\text{cov}(a_X, a_Y) = A_{XY}\sigma_A^2, \tag{1}$$

where a_X and a_Y stand for the BV of X and Y, A_{XY} is the additive genetic relationship between X and Y, and σ_A^2 is the additive variance. The additive genetic relationship is related to probabilities of identity-by-descent (IBD, [12]) and thus $\text{cov}(a_X, a_Y)$ in Eq. (1) is always greater or equal to zero. In other words, related individuals X and Y have a positive covariance that is due to a non-zero probability of sharing genes that are IBD. On the other hand, a zero covariance indicates that X and Y are unrelated.

In addition, the fact that $\text{cov}(a_X, a_Y)$ is ≥ 0 explains a joint behavior of the breeding values of X and Y by which they are likely to simultaneously take large (or small) values. As a result, the non-negative covariance property ensures that animals with larger BV than the mean will tend to have progeny with larger BV than the mean, whereas animals with smaller BV than the mean will produce progeny that tend to have smaller BV than the mean. Consequently, $\text{cov}(a_X, a_Y) \geq 0$ ensures the success of selection in either direction when employing predictions of BV calculated under such a premise. This useful stochastic behavior is known as "positive quadrant dependence" [13] and we seek to extend this property to multilocus distributions of IBD.

Under gametic phase equilibrium, the additive relationship A_{XY} in Eq. (1) is equal to twice the coefficient of coancestry between X and Y [12], which in turn is equal to the probability that one random sampled allele in any locus in the genome of X is IBD to a random sampled allele at the same locus but in individual Y; this probability is denoted as $P(X \equiv Y)$. As a measure of relatedness

Cantet *et al. Genetics Selection Evolution*   (2022) 54:64

Page 3 of 14

between two animals, the coefficient of coancestry is then a probability of IBD at a single locus, which is the same for any autosomal loci under the assumption of gametic phase equilibrium. The concept of IBD is always in relation to a "founder population" [14–16].

Covariance between breeding values under segmental inheritance

The original definition of IBD alludes to a discrete random variable at a single locus. However, the genome is inherited in segments produced by recombination during meiosis [15], which introduces dependence among multiple loci. When calculating the covariance matrix of BV, the assumption of independence of causal locus effects is challenged by physical dependence due to closeness, i.e., linkage, and pervasive dependence among non-allelic loci [17] due to population structure [18] and/or selection [19]. Therefore, the genotypic frequencies may deviate from their expectation under a scenario of random union of gametes [20], a phenomenon that is called gametic phase disequilibrium. Recombination that occurs during meiosis breaks these associations among non-allelic loci and produces segments of DNA that are carried by descendant animals after several generations [15, 17].

Donnelly [21] proposed a "continuous IBD" model to calculate the probability that individuals in a given relationship share any part of their genomes IBD. Their proposal takes recombination into account and serves to model inheritance of segments [15]. After [21], Guo [22] defined the proportion of genome that is shared IBD by two individuals as the sum of all the chromosome segments that are IBD at the same chromosome loci. However, for this relationship to comply with Eq. (1) and to produce a proper covariance matrix of BV that is inversible and Markovian, some theoretical considerations should be met.

First, we need a generalization of the definition of pointwise probabilities of IBD to segments. Nolan [23] defined the probability of IBD for shared segments through a multivariate binary random variable as follows. Let j be a position on chromosome k such that $j = 1, \ldots, n$. Also let i_j be a binary random variable that is equal to 1 if a recombination occurred between positions j and $j+1$ on the chromosome, and equal to 0 otherwise. A given sequence $(i_1, \ldots, i_n)$ of 0s and 1s on chromosome k is called a "recombination pattern" and has a probability $P_k(i_1, \ldots, i_n)$ attached to it. The continuous IBD random variable $X(t)$ is a binary function equal to 1 if at position t the genetic material on chromosome k of a pair of gametes, from either the same or from different individuals, originates in the same grandparent, whereas it is 0 otherwise. The positions in the recombination pattern $i_1, \ldots, i_n$ where $X(t)$ changes from 0 to 1 or from 1 to 0

are the precise locations where a crossover has occurred [23]. The multilocus probabilities of IBD are then equal to:

$$
\begin{aligned}
&P(X_1 \equiv Y_1, .., X_n \equiv Y_n) \\
&= P(X(t_1) = i_1, X(t_2) - X(t_1) \\
&= |i_2 - i_1|, .., X(t_{n+1}) - X(t_n) \\
&= |i_{n+1} - i_n|).
\end{aligned}
\tag{2}
$$

Next, we use definition (2) to write a covariance matrix of BV with elements equal to the product of the multilocus probabilities of IBD, summed over chromosomes, times the additive variance:

$$
\text{cov}(a_X, a_Y) = \sum_{k=1}^{K} P_k(X_1 \equiv Y_1, .., X_n \equiv Y_n)\sigma_A^2 = \Sigma_{XY}\sigma_A^2,
\tag{3}
$$

where K is the number of chromosomes of the diploid species. The scalar Σ_{XY} is termed "multilocus additive relationship", and it is defined to be equal to:

$$
\Sigma_{XY} = \sum_{k=1}^{K} P_k(X_1 \equiv Y_1, .., X_n \equiv Y_n).
\tag{4}
$$

Notice that the covariance in Eq. (3) is greater or equal to zero because it is a product of a sum of probabilities times a positive number, and this guarantees that BV are positive quadrant dependent [13] random variables.

If we now add and subtract A_{XY} to Σ_{XY} in Eq. (4) we obtain:

$$
\Sigma_{XY} = A_{XY} + (\Sigma_{XY} - A_{XY}).
\tag{5}
$$

Thus, the multilocus additive relationship is equal to the single-locus additive relationship plus a "disequilibrium" term that measures how distant is the realized relationship from A_{XY}. In the next section, we relate $\Sigma_{XY} - A_{XY}$ to the parameters of the 'ancestral regression' (AR) of Cantet et al. [8].

We call the disequilibrium term $\Sigma_{XY} - A_{XY}$ "identity disequilibrium" after ideas that were first discussed by Weir and Cockerham [7] and Cockerham and Weir [24]. When considering the effect of recombination on the probability of IBD of alleles at two linked loci (say 1 and 2) for animals X and Y, Weir and Cockerham [7] attributed to JBS Haldane the following inequality:

$$
P(X_1 \equiv Y_1, X_2 \equiv Y_2) \geq P(X_1 \equiv Y_1)P(X_2 \equiv Y_2). \tag{6}
$$

In words, Eq. (6) says that the probability of IBD at two linked locations on a chromosome is higher or equal than the product of the marginal probabilities at each locus when X and Y are related· The value of $P(X_1 \equiv Y_1, X_2 \equiv Y_2)$ depends on the pedigree and on the recombination rate between loci 1 and 2. Weir and Cockerham [7] called identity disequilibrium the difference

between the joint probability of IBD at two loci and the product of the single locus probabilities of IBD. Formally, identity disequilibrium at two loci is equal to:

$$P(X_1 \equiv Y_1, X_2 \equiv Y_2) - P(X_1 \equiv Y_1)P(X_2 \equiv Y_2). \quad (7)$$

Identity disequilibrium is the main source for variation in actual inbreeding [7] or relationships [25]. Expres-

$$\mathrm{Var}(a_x) = \left\{ \begin{array}{l} 0.25[2 + F_S + F_D + 2\Sigma_{S,D}] + \beta_S^2[2 + F_{SS} + F_{DS} - 2\Sigma_{SS,DS}] \\ +\beta_D^2[2 + F_{SD} + F_{DD} - 2\Sigma_{SD,DD}] + 2\beta_S\beta_D[\Sigma_{SS,SD} - \Sigma_{SS,DD} - \Sigma_{DS,SD} + \Sigma_{DS,DD}] \\ +0.5\beta_S[\Sigma_{SS,S} - \Sigma_{DS,S} + \Sigma_{SS,D} - \Sigma_{DS,D}] + 0.5\beta_D[\Sigma_{SD,S} - \Sigma_{DD,S} + \Sigma_{SD,D} - \Sigma_{DD,D}] \end{array} \right\} \sigma_A^2 + \mathrm{Var}(\phi_X). \quad (10)$$

sion (4) in Weir et al. [26] (page 401) formalizes the link between the variance of the inbreeding coefficient at two linked loci and identity disequilibrium as displayed in Eq. (7).

Ancestral regression

The "ancestral regression" (AR) model is a recursive causal Markov model for BV under identity disequilibrium that was developed by Cantet et al. [8]. For completeness, more details about this model are in Additional file 2. According to this model, the regression equation of the BV of X on the BV of its parents and grandparents is:

$$a_X = 0.5a_S + 0.5a_D + \beta_S(a_{SS} - a_{DS}) + \beta_D(a_{SD} - a_{DD}) + \phi_X, \quad (8)$$

where subscripts refer to parents (S = sire and D = dam) and grandparents (SS = paternal grandsire; DS = paternal granddam; SD = maternal grandsire; DD = maternal granddam) of individual X, and ϕ_X is the residual breeding value (RBV). It is shown in Additional file 2 that the disequilibrium term in the ancestral regression is equal to:

$$0.5(P_S - 0.5)(a_{SS} - a_{DS}) + 0.5(P_D - 0.5)(a_{SD} - a_{DD})$$
$$= \beta_S(a_{SS} - a_{DS}) + \beta_D(a_{SD} - a_{DD}). \quad (9)$$

Either side of the equalities $0.5(P_S - 0.5) = \beta_S$ and $0.5(P_D - 0.5) = \beta_D$ expresses identity disequilibrium. In the ancestral regression, β_S and β_D contain the identity disequilibrium and represent the excess, or the deficiency, in the BV of the paternal and maternal grandsires, i.e., a_{SS} and a_{SD}, with respect to the paternal and maternal granddams, i.e., a_{DS} and a_{DD}, in the paternal and maternal gametes, respectively (see Additional file 2). On the other hand, β_S and β_D may be zero due to the fact that under identity equilibrium $P_S = P_D = 0.5$. For standardized covariances, as used here (i.e., additive genetic variance equal to 1), β_S is the path coefficient for

the *contrast* $a_{SS} - a_{DS}$, while β_D is that for the *contrast* $a_{SD} - a_{DD}$ (see Eq. (2) in Wright [27]). For the more general case in which the additive variance differs from 1, β_S and β_D are path *regressions* (see Eq. (1) in Wright [27]) or, simply, partial regression coefficients [4]. In Appendix 1, the general expression for the variance of a BV under the ancestral regression (i.e., without setting the additive genetic variance to 1) is shown to be equal to:

Inbreeding also increases identity disequilibrium [7, 24] and, thus, the variance of the Σ relationships. Equation (10) is complex because the AR model involves the BV of six individuals that may be related and may also be inbred.

The inbreeding values of the grandparents and parents of X that appear in Eq. (10), i.e., F_i ($i = 1, ..., 6$), should not be confused with the usual measure of inbreeding of Wright [28] under gametic equilibrium. Under identity disequilibrium, the inbreeding coefficient must be redefined, and this is done in Appendix 2. The formula to calculate inbreeding under AR is:

$$F_X = 0.5 \left[\begin{array}{l} \Sigma_{SD} + \beta_S(\Sigma_{SS,D} - \Sigma_{DS,D}) + \beta_D(\Sigma_{SD,S} - \Sigma_{DD,S}) \\ +\beta_S\beta_D(\Sigma_{SS,SD} - \Sigma_{SS,DD} - \Sigma_{DS,SD} + \Sigma_{DS,DD}) \end{array} \right]. \quad (11)$$

If all F_i in Eq. (9) are 0, the expression for $\mathrm{Var}(a_x)$ reduces to:

$$\mathrm{Var}(a_x) = \left[0.5 + 2(\beta_S^2 + \beta_D^2) \right] \sigma_A^2 + \mathrm{Var}(\phi_X). \quad (12)$$

For the purpose of comparison, the variance of the BV of individual X under identity equilibrium is equal to $\mathrm{Var}(a_x) = 0.5\sigma_A^2 + \mathrm{Var}(\phi_X^*)$, where the asterisk on ϕ_X refers to the assumption of identity equilibrium. Since $2(\beta_S^2 + \beta_D^2) \geq 0$, $\mathrm{Var}(\phi_X) \leq \mathrm{Var}(\phi_X^*)$, the equality of $\mathrm{Var}(\phi_X)$ and $\mathrm{Var}(\phi_X^*)$ holds only if $\beta_S = \beta_D = 0$. This result was previously obtained by Cantet and Vitezica [29] for any predictor of BV that incorporates pedigree and genomic information. Pourahmadi [30] found that there is a negative relationship between the diagonal element of the inverse covariance matrix and prediction error variance of the random variable. Therefore, in general AR displays smaller variance of RBV, higher inverse diagonal elements of, reduced prediction error variance of BV, and increased accuracy of BV prediction, when compared to the regular animal model.

Cantet *et al. Genetics Selection Evolution*   (2022) 54:64

Page 5 of 14

For a vector of BV under the AR model ordered such that the BV of any ancestor precedes those of all its descendants, we can write the causal Markov model in vector–matrix form as [31, 32]:

$$\mathbf{a} = \mathbf{Ba} + \boldsymbol{\phi}. \tag{13}$$

Matrix $\mathbf{B}$ is *lower triangular* because genomes always descend from ancestors to descendants. There are at most seven elements different from zero in any row of $\mathbf{B}$: 0.5 for the columns belonging to the parents S and D, β_S for SS, $-\beta_S$ for DS, β_D for SD, $-\beta_D$ for DD and 1 for the diagonal, all other elements are zero. Working algebraically in Eq. (13) results in a triangular system of equations [32]:

$$(\mathbf{I} - \mathbf{B})\mathbf{a} = \boldsymbol{\phi} \Rightarrow \mathbf{a} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\phi}. \tag{14}$$

The expectation and covariance matrix of $\mathbf{a}$ in Eq. (13) are respectively equal to:

$$E(\mathbf{a}) = (\mathbf{I} - \mathbf{B})^{-1}E(\boldsymbol{\phi}) = \mathbf{0}$$

$$\text{and } \mathrm{Var}(\mathbf{a}) = \boldsymbol{\Sigma} = (\mathbf{I} - \mathbf{B})^{-1}\mathrm{Var}(\boldsymbol{\phi})\left(\mathbf{I} - \mathbf{B}'\right)^{-1}. \tag{15}$$

In Eq. (15), $\mathrm{Var}(\boldsymbol{\phi})$ is a diagonal matrix ($\mathbf{D}$), which implies independence among the RBV of different animals due to the parental Markov property. The inverse of the covariance matrix is:

$$\boldsymbol{\Sigma}^{-1} = \left(\mathbf{I} - \mathbf{B}'\right)\mathbf{D}^{-1}(\mathbf{I} - \mathbf{B}). \tag{16}$$

$\mathbf{B}_i$ and $\mathbf{B}_j$, respectively, whereas let D_{ii} be the variance of the RBV of i, each non-zero element of the sparse matrix $\mathbf{B}_i\mathbf{D}_{ii}\mathbf{B}'_j$ is called a 'contribution' to the elements of $\boldsymbol{\Sigma}^{-1}$. For a single trait model, $\boldsymbol{\Sigma}^{-1}$ is written down in symmetric sparse format with 28 contributions per animal, compared to six contributions per animal to calculate $\mathbf{A}^{-1}$ [33]. In the multiple trait case (t traits and q animals), the number of contributions in the animal model where the covariance matrix comes from a Kronecker product of the $t \times t$ covariance matrix of BV from different traits times $\mathbf{A}$, the total number of contributions is equal to $(6t + 4.5t(t - 1))q$, whereas in the AR model, the number of contributions is $(28t + 24.5t(t - 1))q$. For the genetic evaluation of growth in beef cattle where $t = 4$ and assuming that one million animals are evaluated, the number of non-zero elements in the mixed model equations is equal at least to 78 million in the animal model with $\mathbf{A}^{-1}$ and 406 million in the AR model, i.e., more than 5.2 times larger. Thus, with data from several million animals, fitting a multiple trait AR model may be prohibitive. Therefore, we will look for a causal model of genetic evaluation that induces a storage of non-zeros similar to the animal model and that also accounts for identity disequilibrium.

As discussed in the section of Additional file 3 entitled *A recursive calculation of the rows and columns of* $\boldsymbol{\Sigma}$, the BV of an animal in the AR model is independent of the BV of any other animal when conditioning on those of its grandparents and parents. The random vector $\mathbf{a}'_A = [a_{SS}a_{DS}a_{SD}a_{DD}a_Sa_D]$ contains a Markov blanket [34] for estimating β_S and β_d, i.e., a set of locally sufficient statistics for both parameters [5]. The covariance matrix of $\mathbf{a}_A$ is equal to:

$$\boldsymbol{\Sigma}_A = \begin{bmatrix} \Sigma_{SS,SS} & \Sigma_{SS,DS} & \Sigma_{SS,SD} & \Sigma_{SS,DD} & \Sigma_{SS,S} & \Sigma_{SS,D} \\ \Sigma_{DS,SS} & \Sigma_{DS,DS} & \Sigma_{DS,SD} & \Sigma_{DS,DD} & \Sigma_{DS,S} & \Sigma_{DS,D} \\ \Sigma_{SD,SS} & \Sigma_{SD,DS} & \Sigma_{SD,SD} & \Sigma_{SD,DD} & \Sigma_{SD,S} & \Sigma_{SD,D} \\ \Sigma_{DD,SS} & \Sigma_{DD,DS} & \Sigma_{DD,SD} & \Sigma_{DD,DD} & \Sigma_{DD,S} & \Sigma_{DD,D} \\ \hline \Sigma_{S,SS} & \Sigma_{S,DS} & \Sigma_{S,SD} & \Sigma_{S,DD} & \Sigma_{S,S} & \Sigma_{S,D} \\ \Sigma_{D,SS} & \Sigma_{D,DS} & \Sigma_{D,SD} & \Sigma_{D,DD} & \Sigma_{D,S} & \Sigma_{D,D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_G & \boldsymbol{\Sigma}_{G,P} \\ \boldsymbol{\Sigma}_{P,G} & \boldsymbol{\Sigma}_P \end{bmatrix}. \tag{17}$$

Cantet et al. [8] proposed an algorithm to calculate $\boldsymbol{\Sigma}^{-1}$ that was derived from the method to compute $\mathbf{A}^{-1}$ [33], by inferring the pattern of non-zero elements in $\boldsymbol{\Sigma}^{-1}$ from the sparsity pattern of $\mathbf{B}$ and the diagonal elements of $\mathbf{D}$.

Ancestral regression to parents (PAR)

Whereas the algorithm to easily invert $\boldsymbol{\Sigma}$ [8] makes it possible to build the mixed model equations to fit the AR without direct inversion of the covariance matrix of BV, the amount of memory needed is greater than that to calculate $\mathbf{A}^{-1}$. Let the rows of $\mathbf{B}$ related to animals i and j be

Matrix $\boldsymbol{\Sigma}_G$ is 4×4 and contains the covariances among the BV of the grandparents, and $\boldsymbol{\Sigma}_{G,P}$ is a 4×2 matrix of the covariances between the grandparents and parents. They are respectively equal to:

$$\boldsymbol{\Sigma}_G = \begin{bmatrix} 1 + F_{SS} & \boldsymbol{\Sigma}_{SS,DS} & \boldsymbol{\Sigma}_{SS,SD} & \boldsymbol{\Sigma}_{SS,DD} \\ \boldsymbol{\Sigma}_{SS,DS} & 1 + F_{DS} & \boldsymbol{\Sigma}_{DS,SD} & \boldsymbol{\Sigma}_{DS,DD} \\ \boldsymbol{\Sigma}_{SS,SD} & \boldsymbol{\Sigma}_{DS,SD} & 1 + F_{SD} & \boldsymbol{\Sigma}_{SD,DD} \\ \boldsymbol{\Sigma}_{SS,DD} & \boldsymbol{\Sigma}_{DS,DD} & \boldsymbol{\Sigma}_{SD,DD} & 1 + F_{DD} \end{bmatrix},$$

and $\boldsymbol{\Sigma}_{P,G} = \begin{bmatrix} \boldsymbol{\Sigma}_{G,S} \\ \boldsymbol{\Sigma}_{G,D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{SS,S} & \boldsymbol{\Sigma}_{DS,S} & \boldsymbol{\Sigma}_{SD,S} & \boldsymbol{\Sigma}_{DD,S} \\ \boldsymbol{\Sigma}_{SS,D} & \boldsymbol{\Sigma}_{DS,D} & \boldsymbol{\Sigma}_{SD,D} & \boldsymbol{\Sigma}_{DD,D} \end{bmatrix}$ (18)

The vectors $\boldsymbol{\Sigma}_{G,S}$ and $\boldsymbol{\Sigma}_{G,D}$ in Eq. (18) are the rows of $\boldsymbol{\Sigma}_{P,G}$, and the covariance matrix of parental BV is equal to:

$$\boldsymbol{\Sigma}_P = \begin{bmatrix} 1 + F_S & \boldsymbol{\Sigma}_{SD} \\ \boldsymbol{\Sigma}_{SD} & 1 + F_D \end{bmatrix}. \quad (19)$$

We now write the 6×1 vector of covariances between the grandparents and parents of individual X as follows:

$$\begin{aligned} \boldsymbol{\Sigma}'_{AX} &= \left[\boldsymbol{\Sigma}'_{G,X} \middle| \boldsymbol{\Sigma}'_{P,X} \right] \\ &= \left[\Sigma_{SS,X} \Sigma_{DS,X} \Sigma_{SD,X} \Sigma_{DD,X} \middle| \Sigma_{S,X} \Sigma_{D,X} \right]'. \end{aligned} \quad (20)$$

The vector $\boldsymbol{\Sigma}_{G,X}$ includes the covariances between the BV of X and those of the four grandparents, whereas $\boldsymbol{\Sigma}_{P,X}$ includes the covariances between the BV of X and those of its parents. Finally, the vector $\mathbf{B}_{X,r}$ of path coefficients in Eq. (17) is defined in the following manner:

$$\mathbf{B}_{X,r} = \left[\beta_S, -\beta_S, \beta_D, -\beta_D, 0.5, 0.5 \right]'. \quad (21)$$

From Expression (20) in [8], we can write vector $\boldsymbol{\Sigma}_{AX}$ as a function of $\mathbf{B}_X$ as: $\boldsymbol{\Sigma}_A \mathbf{B}_{X,r} = \boldsymbol{\Sigma}_{A,X}$, and in more detail, as:

$$\begin{bmatrix} \boldsymbol{\Sigma}_G & \boldsymbol{\Sigma}_{G,P} \\ \boldsymbol{\Sigma}_{P,G} & \boldsymbol{\Sigma}_P \end{bmatrix} \mathbf{B}_{X,r} = \begin{bmatrix} \boldsymbol{\Sigma}_{G,X} \\ \boldsymbol{\Sigma}_{P,X} \end{bmatrix} \quad (22)$$

As first done in [35] to obtain the reduced animal model, we now absorb the first four equations in Eq. (22), which correspond to the grandparents, into the last two parental equations. The solution of the resulting system is the vector $\boldsymbol{\rho}$ ($\boldsymbol{\rho}' = [\rho_S, \rho_D]$), which contains the regression coefficients to regress the BV of X on those of its parents, that is:

$$\left[\boldsymbol{\Sigma}_P - \boldsymbol{\Sigma}_{P,G} \boldsymbol{\Sigma}_G^{-1} \boldsymbol{\Sigma}_{G,P} \right] \boldsymbol{\rho} = \left[\boldsymbol{\Sigma}_{P,X} - \boldsymbol{\Sigma}_{P,G} \boldsymbol{\Sigma}_G^{-1} \boldsymbol{\Sigma}_{G,X} \right]. \quad (23)$$

In scalar notation system, Eq. (23) is equal to:

$$\begin{bmatrix} 1 + F_S - \Sigma_{S,G} \, \Sigma_G^{-1} \Sigma_{G,S} & \Sigma_{S,D} - \Sigma_{S,G} \, \Sigma_G^{-1} \, \Sigma_{G,D} \\ \Sigma_{D,S} - \Sigma_{D,G} \, \Sigma_G^{-1} \Sigma_{G,S} & 1 + F_D - \Sigma_{D,G} \, \Sigma_G^{-1} \Sigma_{G,D} \end{bmatrix} \begin{bmatrix} \rho_S \\ \rho_D \end{bmatrix} = \begin{bmatrix} \Sigma_{S,X} - \Sigma_{S,G} \, \Sigma_G^{-1} \, \Sigma_{G,X} \\ \Sigma_{D,X} - \Sigma_{D,G} \, \Sigma_G^{-1} \, \Sigma_{G,X} \end{bmatrix} \quad (24)$$

Thus, the ancestral regression to the parents parallels the recursion equation for BV in the animal model (i.e., a "parental regression"), but with unequal parental contributions due to distinct contributions of the grandparental BV. The equation for the PAR model is then equal to:

$$a_X = \rho_S a_S + \rho_D a_D + \phi_{P,X}. \quad (25)$$

The random variable $\phi_{P,X}$ is the RBV in the PAR model. The parameter spaces of ρ_S and ρ_D are $0.25 \leq \rho_S, \rho_D \leq 0.75$, and almost always ρ_S and ρ_D will be different from each other. Rather than the proportion of the genome of the individual that originates from a given parent, the difference between ρ_S or ρ_D with 0.50 is not a genome proportion but the amount of statistical information on their BV that the sire and the dam pass to a particular progeny under identity disequilibrium, instead of 0.5.

The variance of the BV in PAR is as follows:

$$\mathrm{Var}_\rho(a_X) = \mathrm{Var}_\rho(\rho_S a_S + \rho_D a_d) + \mathrm{Var}_\rho(\phi_{Pi}), \quad (26)$$

where ϕ_{Pi} is the RBV under the PAR model for animal i. The subscript ρ in the variance operator or in $\boldsymbol{\Sigma}_\rho$ emphasizes the fact that the variance is calculated under the PAR. Elements of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_\rho$ are different because AR and PAR are not equivalent models in the sense of [36] (see Appendix 4). Using variance operator rules, Eq. (26) results in being equal to:

$$\begin{aligned} \mathrm{Var}_\rho(a_X) &= \rho_S^2 \mathrm{Var}_\rho[a_S] + \rho_D^2 \, \mathrm{Var}_\rho[a_D] \\ &\quad + 2\rho_S \rho_D \mathrm{cov}_\rho[a_S, a_D] + \mathrm{Var}_\rho(\phi_{Pi}) \\ &= \rho_S^2(1 + F_S) + \rho_D^2(1 + F_D) \\ &\quad + 2\rho_S \rho_D \Sigma_{\rho S,D} + \mathrm{Var}_\rho(\phi_{Pi}). \end{aligned} \quad (27)$$

Finally, we obtain the variance of the RBV by solving Eq. (27) for $\mathrm{Var}_\rho(\phi_{Pi})$ as:

$$\begin{aligned} \mathrm{Var}_\rho(\phi_{Pi}) =& 1 + F_X - \Big[\rho_S^2(1 + F_S) \\ &+ \rho_D^2(1 + F_D) + 2\rho_S \rho_D \Sigma_{\rho S,D} \Big], \end{aligned} \quad (28)$$

where $\Sigma_{\rho S,D}$ is the covariance between the BV of the parents of X under the PAR model.

As the PAR model displays the same set of conditional independencies as the animal model [1], the number of non-zeros in the coefficient matrix of the mixed model equations under PAR is equal to that in the animal model that is generated with pedigree to calculate $\mathbf{A}^{-1}$. To obtain predictions of BV using Henderson's mixed model equations [1], the inverse covariance structure of BV under PAR, $\boldsymbol{\Sigma}_\rho^{-1}$, can be calculated by simply modifying the algorithm of Henderson [33] for $\mathbf{A}^{-1}$ without ever calculating $\boldsymbol{\Sigma}_\rho$, by replacing the 0.5 values in the contributions to $\mathbf{A}^{-1}$ by ρ_S or ρ_D and 0.25 values by the product of ρ_S and ρ_D, or the squared values of these path coefficients.

Covariance between the breeding values of an ancestor and a descendant in causal Markov models

For any causal Markov model, the covariance between the BV of an ancestor (a_A) and the BV of a descendant (a_X) results from the recursion of a_X to a linear combination of BV of direct ancestors, until reaching the BV of the ancestor. At that point, $cov(a_A, a_A) = V\,ar(a_A)$ and the recursion ends. Calculating this covariance is then similar to calculating additive relationships using the tabular method [37], for which individuals in a pedigree are ordered by date of birth to ensure that a descendant is never placed before a direct ancestor. The correctness of the algebraic procedure for the tabular method was demonstrated in a paper from Rohan Fernando's group back in the 1990s [38].

The ancestor–descendant covariance of breeding values in the ancestral regression model

As explained in the above paragraph, the ancestor–descendant covariance of BV results from the recursion of the BV of the descendant only, because of the parental Markov property. By definition:

$$\Sigma_{A,X} = cov(a_A, a_X).$$

Now we replace the BV of the descendant with the AR expression for the BV of X in Eq. (8), as follows:

$$
\begin{aligned}
cov(a_A, a_X) = {} & cov(a_A, 0.5a_S + 0.5a_D \\
& + \beta_S(a_{SS} - a_{DS}) + \beta_D(a_{SD} - a_{DD})) \\
= {} & 0.5cov(a_A, a_S) + 0.5cov(a_A, a_D) \\
& + \beta_S[cov(a_A, a_{SS}) - cov(a_A, a_{DS})] \\
& + \beta_D[cov(a_A, a_{SD}) - cov(a_A, a_{DD})].
\end{aligned}
$$

Because the covariance between the BV of animals j and k in AR is defined to be equal to $cov(a_j, a_k) = \Sigma_{j,k}$, i.e., the elements of Σ, the above equation results in being equal to:

$$\Sigma_{A,X} = 0.5(\Sigma_{A,S} + \Sigma_{A,D}) + \beta_S(\Sigma_{A,SS} - \Sigma_{A,DS}) + \beta_D(\Sigma_{A,SD} - \Sigma_{A,DD}). \quad (29)$$

The covariance between the BV of a sire and an offspring (say X) is obtained after setting the ancestor to the sire in Eq. (29), i.e., A = S, such that:

$$\Sigma_{S,X} = 0.5(\Sigma_{S,S} + \Sigma_{S,D}) + \beta_S(\Sigma_{S,SS} - \Sigma_{S,DS}) + \beta_D(\Sigma_{S,SD} - \Sigma_{S,DD}).$$

Now, since $\Sigma_{S,S}$ is the variance of the BV of the sire and is equal to $\Sigma_{S,S} = 1 + F_S$, the sire – offspring covariance of BV is equal to:

$$\Sigma_{S,X} = 0.5(1 + F_S + \Sigma_{S,D}) + \beta_S(\Sigma_{SS,S} - \Sigma_{DS,S}) + \beta_D(\Sigma_{SD,S} - \Sigma_{DD,S}). \quad (30)$$

In case of gametic equilibrium, $\beta_S = \beta_D = 0$ and $\Sigma_{S,X} = 0.5(1 + F_S + 2F_X) = A_{S,X}$, as for Expression (7.4) in [38].

By a similar reasoning, the covariance between the BV of X and of its dam is equal to:

$$
\begin{aligned}
\Sigma_{D,X} = {} & 0.5(1 + F_D + \Sigma_{S,D}) \\
& + \beta_S(\Sigma_{SS,D} - \Sigma_{DS,D}) \\
& + \beta_D(\Sigma_{SD,D} - \Sigma_{DD,D}).
\end{aligned} \quad (31)
$$

If the paternal grandsire is the ancestor in Eq. (31), $cov(a_{SS}, a_X)$ under the AR model becomes equal to:

$$
\begin{aligned}
\Sigma_{SS,X} = {} & 0.5\Sigma_{SS,S} + 0.5\Sigma_{SS,D} \\
& + \beta_S(1 + F_{SS} - \Sigma_{SS,DS}) \\
& + \beta_D(\Sigma_{SS,SD} - \Sigma_{SS,DD}).
\end{aligned} \quad (32)
$$

In the absence of inbreeding, $\Sigma_{SS,D} = F_{SS} = \Sigma_{SS,DS} = \Sigma_{SS,SD} = \Sigma_{SS,DD} = 0$, and with $\Sigma_{SS,S} = 0.5$, the covariance between the BV of a grandsire and grand-offspring is equal to:

$$\Sigma_{SS,X} = 0.25 + \beta_S. \quad (33)$$

Under gametic equilibrium, $\beta_S = 0$ and $\Sigma_{SS,X} = 0.25$.

Similarly, the covariance between the BV of the paternal granddam and of X is:

$$
\begin{aligned}
\Sigma_{DS,X} = {} & 0.5\Sigma_{DS,S} + 0.5\Sigma_{DS,D} \\
& - \beta_S(1 + F_{DS} - \Sigma_{SS,DS}) \\
& + \beta_D(\Sigma_{DS,SD} - \Sigma_{DS,DD}).
\end{aligned} \quad (34)
$$

If there is no inbreeding, Eq. (34) becomes equal to:

$$\Sigma_{DS,X} = 0.25 - \beta_S. \quad (35)$$

From Eqs. (33) and (35), it is deduced that $-0.25 \leq \beta_S \leq 0.25$. As with the regular animal model and in the absence of inbreeding, the sum of the covariances between the grandsire and X and the granddam and X in AR give the parent–offspring covariance. To see this, add Eq. (33) to Eq. (35) to obtain $\Sigma_{SS,X} + \Sigma_{DS,X} = 0.25 + \beta_S + 0.25 - \beta_S = 0.5$, which is precisely the value of the covariance between the BV of parent and offspring in the absence of inbreeding. As the values of β_S and β_D are set arbitrarily to the positive sign for the grandsire and the covariance parent–offspring is the same under identity equilibrium or gametic disequilibrium, it follows that the resemblance gained (or lost) by the grandsire with respect to the expected value of

Cantet *et al. Genetics Selection Evolution*   (2022) 54:64

Page 8 of 14

0.25 under gametic equilibrium is lost (or gained) by the granddam.

The covariances of the BV of the maternal grandsire and the maternal granddam with the BV of X are, respectively, equal to:

$$\Sigma_{SD,X} = 0.5\Sigma_{SD,S} + 0.5\Sigma_{SD,D}$$
$$+\beta_S(\Sigma_{SD,SS} - \Sigma_{SD,DS}) \qquad (36)$$
$$+ \beta_D(1 + F_{SD} - \Sigma_{SD,DD}),$$

and

$$\Sigma_{DD,X} = 0.5\Sigma_{DD,S} + 0.5\Sigma_{DD,D}$$
$$+\beta_S(\Sigma_{DD,SS} - \Sigma_{DD,DS}) \qquad (37)$$
$$- \beta_D(1 + F_{DD} - \Sigma_{SD,DD}).$$

Similar expressions to Eqs. (36) and (37) but in the absence of inbreeding are:

$$\Sigma_{SD,X} = 0.25 + \beta_D,$$

$$\text{and } \Sigma_{DD,X} = 0.25 - \beta_D. \qquad (38)$$

Covariances between breeding values when relatives are not ancestors of each other in Markov causal models

Two related individuals X and Y that do not descend from each other have one or more common ancestors $CA_1, \ldots, CA_I$, where I is the number of common ancestors. A *trek* is the path between the BV of X and Y passing through the BV of the common ancestor CA_i and is represented by $a_X \leftarrow \cdots \leftarrow a_{CA_i} \rightarrow \cdots \rightarrow a_Y$ (see page 12 in [4] and Expression (3.1) in [39]). Each trek contributes a value that is equal to the product of all single path coefficients involved in the trek times the variance of the BV of the common ancestor. Then, the covariance between the BV of X and Y can be calculated by summing contributions over all possible treks [40]:

$$\text{cov}(a_X, a_Y) = \sum_{i=1}^{I} (b_j \ldots b_{j'}) \text{Var}(a_{CA_i}), \qquad (39)$$

where b_j stands for the single path coefficient of the j-th ancestor in the trek. Note that these path coefficients take either the value 0.5 or β. In the parental regression model, all path coefficients are equal to 0.5.

A difficulty with the AR model is that the covariance between the BV of two relatives, none of which is an ancestor of the other, involves the calculation of up to 36 covariances among possibly 12 different animals (see the derivation in Appendix 3). The resulting expression is formidable and suggests the need for a simpler method to calculate covariances between collateral individuals, which is what we present in the next section.

Covariance between breeding values under the ancestral regression model to parents when neither animal is an ancestor of each other

Once parameters ρ are estimated using Eq. (24), the genetic covariances under the PAR model are calculated by a recursion to the youngest individual, in the following manner:

$$\Sigma_{\rho A,X} = \rho_S \Sigma_{\rho A,S} + \rho_D \Sigma_{\rho A,D}. \qquad (40)$$

The subscript ρ in the elements of the covariance matrix indicates that those elements are defined under PAR. In Appendix 4, we explain the reason why AR and PAR are not Markov equivalent models and display an example in which the covariances between the BV of parent and off-spring under AR and PAR are numerically different.

When neither individual is an ancestor of the other, the covariance of the BV of relatives under PAR results from replacing BV by the PAR model for breeding values:

$$\text{cov}_\rho(a_X, a_Y)$$
$$= \text{cov}_\rho(\rho_{S_X} a_{S_X} + \rho_{D_X} a_{D_X}, \rho_{S_Y} a_{S_Y} + \rho_{D_Y} a_{D_Y})$$
$$= \text{cov}_\rho(\rho_{S_X} a_{S_X}, \rho_{S_Y} a_{S_Y})$$
$$+\text{cov}_\rho(\rho_{S_X} a_{S_X}, \rho_{D_Y} a_{D_Y})$$
$$+\text{cov}_\rho(\rho_{D_X} a_{D_X}, \rho_{S_Y} a_{S_Y})$$
$$+ \text{cov}_\rho(\rho_{D_X} a_{D_X}, \rho_{D_Y} a_{D_Y}).$$

Then, the resulting covariance between BV is:

$$\text{cov}_\rho(a_X, a_Y) = \rho_{S_X} \rho_{S_Y} \text{cov}_\rho(a_{S_X}, a_{S_Y})$$
$$+ \rho_{S_X} \rho_{D_Y} \text{cov}_\rho(a_{S_X}, a_{D_Y})$$
$$+ \rho_{D_X} \rho_{S_Y} \text{cov}_\rho(a_{D_X}, a_{S_Y}) \qquad (41)$$
$$+ \rho_{D_X} \rho_{D_Y} \text{cov}_\rho(a_{D_X}, a_{D_Y}).$$

In general, we can write the covariance terms in Eq. (41) using Eq. (39) as follows:

$$\text{cov}_\rho(a_{S_X}, a_{S_Y}) = (1 + F_{\rho S})\sigma_A^2,$$

$$\text{cov}_\rho(a_{S_X}, a_{D_Y}) = \text{cov}_\rho(a_{D_X}, a_{S_Y}) = \Sigma_{\rho SD}\sigma_A^2,$$

$$\text{cov}_\rho(a_{D_X}, a_{D_Y}) = (1 + F_{\rho D})\sigma_A^2.$$

Specific expressions for the covariance between BV of full-sibs (*FS*), paternal half-sibs (*PHS*), and maternal half-sibs (*MHS*) are calculated by replacing in Eq. (41) with appropriate elements of $\boldsymbol{\Sigma}$. The respective covariances between BV of *FS*, *PHS*, and *MHS* are:

$$\text{cov}_\rho(FS) = [\rho_{S_X} \rho_{S_Y}(1 + F_{\rho S})$$
$$+ [\rho_{S_X} \rho_{D_Y} + \rho_{D_X} \rho_{S_Y}] \Sigma_{\rho SD} \qquad (42)$$
$$+ \rho_{D_X} \rho_{D_Y}(1 + F_{\rho D})]\sigma_A^2$$

$$\text{cov}_\rho(PHS) = \left[\rho_{S_X}\rho_{S_Y}(1 + F_{\rho S}) + \rho_{S_X}\rho_{D_Y}\Sigma_{\rho S_X D_Y} + \rho_{S_Y}\rho_{D_X}\Sigma_{\rho S_Y D_X}\right]\sigma_A^2 \qquad (43)$$

$$\text{cov}_{P\rho}(MHS) = \left[\rho_{S_X}\rho_{D_Y}\Sigma_{\rho S_X D_Y} + \rho_{S_Y}\rho_{D_X}\Sigma_{\rho S_Y D_X} + \rho_{D_X}\rho_{D_Y}(1 + F_{\rho D})\right]\sigma_A^2 \qquad (44)$$

Discussion

In this paper, we introduce the ancestral regression to parents (PAR) model, which derives from the ancestral regression model (AR) introduced by Cantet et al. [8] and show how to calculate additive covariances between the breeding values of related individuals under identity disequilibrium. Both, the AR and PAR models, and the standard parental regression model, are recursive causal Markov models [4, 5] under a multivariate Gaussian distribution of BV. The expressions for the covariances between the BV of relatives are derived using the recursive causal framework. In particular, the covariances between the BV of grandparents and grand-offspring, between parents and offspring, and the covariance between collaterals for the AR and PAR models. These covariances are different from the usual additive covariances under gametic equilibrium because they account for identity disequilibrium [7]: the difference between the realized relationship and its expected value under gametic equilibrium, which is originated in the individual processes of whole genome recombination during the formation of the gametes that originated both individuals in the relationship. These recombinations involve the genomes of the grandsire and the granddam during meiosis, and result in unique gametes that form each individual.

Causal Markov models for BV possess the property that their residual terms (referred to as RBV) are independent, thus they allow for a recursive calculation of the covariance between parent and offspring. In Markov causal models, this covariance is always positive and, in general, the covariance between the BV of any two animals in a pedigree is non-negative and, therefore, BV are positive quadrant dependent [13] random variables. Moreover, the covariance matrix of BV under PAR can be inverted with a computational effort that is proportional to the number of animals and, particularly, the inversion of the covariance matrix under PAR is similar to inverting $\mathbf{A}$ [33], but with the paternal and maternal path coefficients ρ of PAR replacing the 0.5 in the parental contributions.

While in the pre-genomic era covariances between BV of relatives were calculated on the basis of the degree of *relatedness* between animals [1] using only information from the pedigree, in recent times, animal breeders have also used dense sets of SNPs. The use of pedigree relationships is compatible with a stochastic process that mimics a "regression" to the BV of equally contributing parents, whereas genomic relationships result from using sample statistics, specifically marker frequencies, to build an observed "empirical genetic relatedness matrix" (see VanRaden [41]). Although BLUP of BV can be obtained by using either a theoretically-derived or an observed covariance matrix, the variance of the elements of the latter may be quite large. For example, Garcia-Baccino et al. [42] found that the variance of half-sib relationships among crossbred pigs was half as large when relatedness was measured in IBD segments compared to using single-loci relatedness based on the method of Han and Abney [43], and almost one fourth of the variance of "identity-by-state" calculated with the method described in [41]. In addition, and for the same crossbred population and for either ancestor–descendant relationships or for relationships of collateral relatives, Forneris et al. [44] found that the variance of relatedness measures was smaller when conditioning on pedigree and SNPs than when conditioning on markers only. More importantly, observed probabilities of IBD result in non-zero covariances between RBV [45] and, neither the parent–offspring covariance of BV does not have the parental Markov property, and the resulting covariance matrix of BV does not admit a Cholesky-root-free decomposition such as in Eq. (15).

When considering the transmission of DNA across generations in a lineage or a pedigree (see Additional file 1), Matsen and Evans [46] differentiated the genealogical or ancestry process from the differential recombination of the grandparental genomes when producing the parental gamete that produces their grand progeny. Accounting for recombination, the fraction of the additive variance that is explained by the regression of the BV of a progeny on the BV of parents and grandparents exceeds the genetic variance explained by the "parental regression" by $\beta_S(a_{SS} - a_{DS}) + \beta_D(a_{SD} - a_{DD})$, such that the difference of variances for the RBV in the animal model with the one from the AR is non-negative, i.e., $\text{Var}(\phi^*) - \text{Var}(\phi_{AR}) \geq 0$, as proven in [29]. In general, identity disequilibrium can be expressed as the matrix difference $(\mathbf{\Sigma} - \mathbf{A})$. When the matrix $(\mathbf{\Sigma} - \mathbf{A})$ is positive definite, it is possible to recover more additive variance and thus, heritability, by using our covariance matrices $\mathbf{\Sigma}$ (of AR) or $\mathbf{\Sigma}_\rho$ (of PAR) in the resulting estimating

Cantet *et al. Genetics Selection Evolution*   (2022) 54:64

Page 10 of 14

equations, than by using the standard **A** matrix calculated with pedigree information solely.

Conclusions

The ancestral regression (AR) and ancestral regression to parents (PAR) are causal Markov models that guarantee a positive dependence between breeding values of relatives which, in turn, result in sparse inverses of the respective covariance matrices. We have derived the covariances between the breeding values of relatives that are needed to fit the AR and PAR models to phenotypes, pedigree, and genotypes to obtain BLUP of breeding values. The inbreeding coefficient was also derived under gametic disequilibrium due to differential recombination of grandparental genomes in the gametes.

Appendix 1

Variance of breeding values under the ancestral regression model (AR) and under the ancestral regression to parents (PAR)

By using the variance operator in the ancestral regression of Eq. ([8]) we obtain:

$$
\begin{aligned}
\mathrm{Var}(a_X) =\,& \mathrm{Var}(0.5a_S + 0.5a_D \\
& + \beta_S(a_{SS} - a_{DS}) \\
& + \beta_D(a_{SD} - a_{DD}) + \phi_X)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}_\rho(a_X) &= \mathrm{Var}_\rho(\rho_S a_S + \rho_D a_D) + \mathrm{Var}_\rho(\phi_i) \\
&= \rho_S^2 \mathrm{Var}_\rho(a_S) + \rho_D^2 \mathrm{Var}_\rho(a_D) + 2\rho_S \rho_D\, \mathrm{cov}_\rho(a_S, a_D) + \mathrm{Var}_\rho(\phi_i) \\
&= \left[\rho_S^2(1 + F_S) + \rho_D^2(1 + F_D) + 2\rho_S \rho_D\, \Sigma_{\rho S,D}\right]\Sigma_A^2 + \mathrm{Var}(\phi_i),
\end{aligned}
\tag{46}
$$

$$
\begin{aligned}
=\,& \mathrm{Var}(0.5a_S + 0.5a_D) + \mathrm{Var}(\beta_S(a_{SS} - a_{DS}) \\
& + \beta_D(a_{SD} - a_{DD})) \\
& + 2\mathrm{cov}(0.5a_S + 0.5a_D, \beta_S(a_{SS} - a_{DS}) \\
& + \beta_D(a_{SD} - a_{DD}) + \mathrm{Var}(\phi_X)
\end{aligned}
$$

$$
\begin{aligned}
=\,& 0.25\mathrm{Var}(a_S + a_D) + \beta_S^2 \mathrm{Var}(a_{SS} - a_{DS}) \\
& + \beta_D^2 \mathrm{Var}(a_{SD} - a_{DD}) \\
& + 2\beta_S\beta_D \mathrm{cov}(a_{SS} - a_{DS}, a_{SD} - a_{DD}) \\
& + 0.5\mathrm{cov}(a_S + a_D, \beta_S(a_{SS} - a_{DS})) \\
& + 0.5\mathrm{cov}(a_S + a_D, \beta_D(a_{SD} - a_{DD}) + \mathrm{Var}(\phi_X)
\end{aligned}
$$

$$
\begin{aligned}
=\,& 0.25[\mathrm{Var}(a_S) + \mathrm{Var}(a_D) + 2\mathrm{cov}(a_S, a_D)] \\
& + \beta_S^2[\mathrm{Var}(a_{SS}) + \mathrm{Var}(a_{DS}) - 2\mathrm{cov}(a_{SS}, a_{DS})] \\
& + \beta_D^2[\mathrm{Var}(a_{SD}) \\
& + \mathrm{Var}(a_{DD}) - 2\mathrm{cov}(a_{SD}, a_{DD})] \\
& + 2\beta_S\beta_D \mathrm{cov}(a_{SS}, a_{DS}) - \mathrm{cov}(a_{SS}, a_{DD}) \\
& - \mathrm{cov}(a_{DS}, a_{SD}) + \mathrm{cov}(a_{DS}, a_{DD}) \\
& + 0.5\beta_S[\mathrm{cov}(a_S, a_{SS}) - \mathrm{cov}(a_S, a_{DS}) \\
& + \mathrm{cov}(a_D, a_{SS}) - \mathrm{cov}(a_D, a_{DS})] \\
& + 0.5\beta_D[\mathrm{cov}(a_S, a_{SD}) - \mathrm{cov}(a_S, a_{DD}) \\
& + \mathrm{cov}(a_D, a_{SD}) - \mathrm{cov}(a_D, a_{DD})] + \mathrm{Var}(\phi_X),
\end{aligned}
$$

Finally, writing all the terms as a function of elements of Σ results in the following expression:

$$
\begin{aligned}
\mathrm{Var}(a_X) =\,& \big\{ 0.25[2 + F_S + F_D + 2\Sigma_{S,D}] \\
& + \beta_S^2[2 + F_{SS} + F_{DS} - 2\Sigma_{SS,DS}] \\
& + \beta_D^2[2 + F_{SD} + F_{DD} - 2\Sigma_{SD,DD}] \\
& + 2\beta_S\beta_D[\Sigma_{SS,SD} - \Sigma_{SS,DD} - \Sigma_{DS,SD} + \Sigma_{DS,DD}] \\
& + 0.5\beta_S[\Sigma_{SS,S} - \Sigma_{DS,S} + \Sigma_{SS,D} - \Sigma_{DS,D}] \\
& + 0.5\beta_D[\Sigma_{SD,S} - \Sigma_{DD,S} + \Sigma_{SD,D} - \Sigma_{DD,D}]\big\}\sigma_A^2 \\
& + \mathrm{Var}(\phi_X).
\end{aligned}
\tag{45}
$$

Application of the variance operator for the ancestral regression to parents (PAR) model, i.e., Eq. ([25]) results in being equal to:

Appendix 2

Inbreeding under the models of ancestral regression and the ancestral regression to parents

Wright [28] "defined" the coefficient of inbreeding of individual X $(F_{(W)})$ as $F_{X(W)} = 0.5 r_{SD}\sqrt{(1 + F_{S(W)})(1 + F_{D(W)})}$, where $F_{S(W)}$ and $F_{D(W)}$ are the inbreeding of the sire and dam of X, respectively, and r_{SD} is the correlation between the parental breeding values $r_{SD} = \mathrm{cov}(a_S, a_D)/\sqrt{\mathrm{Var}(a_S)\mathrm{Var}(a_D)}$. By replacing with this correlation in

$F_{X(W)}$, we obtain the value of the inbreeding of X under the AR model (F_X) in the following manner:

$$
\begin{aligned}
F_X &= 0.5\left[\frac{\text{cov}(a_S, a_D)}{\sqrt{\text{Var}(a_S)\text{Var}(a_D)}}\right]\sqrt{(1+F_S)(1+F_D)}\\
&= 0.5\text{cov}(a_S, a_D)\\
&= 0.5\text{cov}(a_S + \beta_S(a_{SS} - a_{DS}), a_D + \beta_D(a_{SD} - a_{DD}))\\
&= 0.5[\text{cov}(a_S, a_D)\\
&\quad +\text{cov}(a_S, \beta_D(a_{SD} - a_{DD})) + \text{cov}(\beta_S(a_{SS} - a_{DS}), a_D)\\
&\quad +\text{cov}(\beta_S(a_{SS} - a_{DS}), \beta_D(a_{SD} - a_{DD}))]\\
&= 0.5\left[\Sigma_{S,D} + \beta_D\left(\Sigma_{SD,S} - \Sigma_{DD,S}\right) + \beta_S\left(\Sigma_{SS,D} - \Sigma_{DS,D}\right)\right]\\
&\quad + \text{cov}(\beta_S(a_{SS} - a_{DS}), \beta_D(a_{SD} - a_{DD})).
\end{aligned}
\tag{47}
$$

Consider now the cross-covariance between the paternal and maternal identity disequilibrium terms:

$$
\begin{aligned}
&\text{cov}(\beta_S(a_{SS} - a_{DS}), \beta_D(a_{SD} - a_{DD}))\\
&= \beta_S\beta_D\text{cov}(a_{SS} - a_{DS}, a_{SD} - a_{DD})\\
&= \beta_S\beta_D[\text{cov}(a_{SS}, a_{SD}) - \text{cov}(a_{SS}, a_{DD})\\
&\quad -\text{cov}(a_{DS}, a_{SD}) + \text{cov}(a_{DS}, a_{DD})]\\
&= \beta_S\beta_D[\Sigma_{SS,SD} - \Sigma_{SS,DD} - \Sigma_{DS,SD} + \Sigma_{DS,DD}].
\end{aligned}
\tag{48}
$$

If we replace with Eq. (48) in Eq. (47), inbreeding under AR is equal to:

$$
\begin{aligned}
F_X = 0.5\big[&\Sigma_{SD} + \beta_S\left(\Sigma_{SS,D} - \Sigma_{DS,D}\right)\\
&+\beta_D\left(\Sigma_{SD,S} - \Sigma_{DD,S}\right)\\
&+\beta_S\beta_D(\Sigma_{SS,SD} - \Sigma_{SS,DD} - \Sigma_{DS,SD} + \Sigma_{DS,DD})\big].
\end{aligned}
\tag{49}
$$

For the PAR model inbreeding is defined to be equal to:

$$
F_{\rho_X} = 2\rho_S\rho_D\Sigma_{\rho_S,D}.
\tag{50}
$$

Hence, $\text{Var}_\rho(a_X)$ is:

$$
\text{Var}_\rho(a_X) = \left(1 + F_{\rho X}\right)\sigma_A^2,
\tag{51}
$$

such that $\rho_S^2\left(1 + F_{\rho S}\right) + \rho_D^2\left(1 + F_{\rho D}\right) + \text{Var}_\rho(\phi_i)\left(\sigma_A^2\right)^{-1} = 1$.

Appendix 3

Covariance between breeding values under the ancestral regression model when neither animal is an ancestor of each other

Let the BV of the related individuals X and Y be a_X and a_Y. The notation, for example, will be that DDY is the maternal granddam of Y, SSX is the paternal grandsire of X, SX is the sire of X and DY the dam of Y. By using the covariance operator in the AR model when neither individual is an ancestor of each other, the covariance between the BV of X and Y is equal to:

$$
\begin{aligned}
\Sigma_{X,Y} &= \text{cov}[0.5(a_{SX} + a_{DX}) + \beta_{SX}(a_{SSX} - a_{DSX})\\
&\quad +\beta_{DX}(a_{SDX} - a_{DDX}), 0.5(a_{SY} + a_{DY})\\
&\quad +\beta_{SY}(a_{SSY} - a_{DSY}) + \beta_{DY}(a_{SDY} - a_{DDY})]\\
&= 0.25\text{cov}[a_{SX} + a_{DX}, a_{SY} + a_{DY}]\\
&\quad + 0.5\beta_{SY}\text{cov}[a_{SX} + a_{DX}, a_{SSY} - a_{DSY}]\\
&\quad + 0.5\beta_{DY}\text{cov}[a_{SX} + a_{DX}, a_{SDY} - a_{DDY}]\\
&\quad + 0.5\beta_{SX}\text{cov}[a_{SSX} - a_{DSX}, a_{SY} + a_{DY}]\\
&\quad + \beta_{SX}\beta_{SY}\text{cov}[a_{SSX} - a_{DSX}, a_{SSY} - a_{DSY}]\\
&\quad + \beta_{SX}\beta_{DY}\text{cov}[a_{SSX} - a_{DSX}, a_{SDY} - a_{DDY}]\\
&\quad + 0.5\beta_{DX}\text{cov}[a_{SDX} - a_{DDX}, a_{SY} + a_{DY}]\\
&\quad + \beta_{DX}\beta_{SY}\text{cov}[a_{SDX} - a_{DDX}, a_{SSY} - a_{DSY}]\\
&\quad + \beta_{DX}\beta_{DY}\text{cov}[a_{SDX} - a_{DDX}, a_{SDY} - a_{DDY}],
\end{aligned}
$$

$$
\begin{aligned}
&=0.25[\Sigma_{SX,SY} + \Sigma_{SX,DY} + \Sigma_{DX,SY} + \Sigma_{DX,DY}]\\
&\quad + 0.5\beta_{SY}[\Sigma_{SX,SSY} - \Sigma_{SX,DSY} + \Sigma_{DX,SSY} - \Sigma_{DX,DSY}]\\
&\quad + 0.5\beta_{DY}[\Sigma_{SX,SDY} - \Sigma_{SX,DDY} + \Sigma_{DX,SDY} - \Sigma_{DX,DDY}]\\
&\quad + 0.5\beta_{SX}[\Sigma_{SSX,SY} + \Sigma_{SSX,DY} - \Sigma_{DSX,SY} - \Sigma_{DSX,DY}]\\
&\quad + 0.5\beta_{DX}[\Sigma_{SDX,SY} + \Sigma_{SDX,DY} - \Sigma_{DDX,SY} - \Sigma_{DDX,DY}]\\
&\quad + \beta_{SX}\beta_{SY}[\Sigma_{SSX,SSY} - \Sigma_{SSX,DSY} - \Sigma_{DSX,SSY} + \Sigma_{DSX,DSY}]\\
&\quad + \beta_{SX}\beta_{DY}[\Sigma_{SSX,SDY} - \Sigma_{SSX,DDY} - \Sigma_{DSX,SDY} + \Sigma_{DSX,DDY}]\\
&\quad + \beta_{DX}\beta_{SY}[\Sigma_{SDX,SSY} - \Sigma_{SDX,DSY} - \Sigma_{DDX,SSY} + \Sigma_{DDX,DSY}]\\
&\quad + \beta_{DX}\beta_{DY}[\Sigma_{SDX,SDY} - \Sigma_{SDX,DDY} - \Sigma_{DDX,SDY} + \Sigma_{DDX,DDY}],
\end{aligned}
$$

The final expression is equal to:

$$
\begin{aligned}
\Sigma_{X,Y} = 0.5\big\{&0.5[\Sigma_{SX,SY} + \Sigma_{SX,DY} + \Sigma_{DX,SY} + \Sigma_{DX,DY}]\\
&+\beta_{SY}[\Sigma_{SX,SSY} - \Sigma_{SX,DSY} + \Sigma_{DX,SSY} - \Sigma_{DX,DSY}]\\
&+\beta_{DY}[\Sigma_{SX,SDY} - \Sigma_{SX,DDY} + \Sigma_{DX,SDY} - \Sigma_{DX,DDY}]\\
&+\beta_{SX}[\Sigma_{SSX,SY} + \Sigma_{SSX,DY} - \Sigma_{DSX,SY} - \Sigma_{DSX,DY}]\\
&+\beta_{DX}[\Sigma_{SDX,SY} + \Sigma_{SDX,DY} - \Sigma_{DDX,SY} - \Sigma_{DDX,DY}]\big\}\\
&+ \beta_{SX}\big\{\beta_{SY}[\Sigma_{SSX,SSY} - \Sigma_{SSX,DSY} - \Sigma_{DSX,SSY} + \Sigma_{DSX,DSY}]\\
&+\beta_{DY}[\Sigma_{SSX,SDY} - \Sigma_{SSX,DDY} - \Sigma_{DSX,SDY} + \Sigma_{DSX,DDY}]\big\}\\
&+ \beta_{DX}\big\{\beta_{SY}[\Sigma_{SDX,SSY} - \Sigma_{SDX,DSY} - \Sigma_{DDX,SSY} + \Sigma_{DDX,DSY}]\\
&+\beta_{DY}[\Sigma_{SDX,SDY} - \Sigma_{SDX,DDY} - \Sigma_{DDX,SDY} + \Sigma_{DDX,DDY}]\big\}.
\end{aligned}
\tag{52}
$$

This is a formidable expression but will be generally much less involved because many of the elements of Σ will be zero, especially if inbreeding is absent.

As an example, consider the covariances among full-sibs and half-sibs in the absence of inbreeding. For full-sibs $S_X = S_Y$, $D_X = D_Y$, $SS_X = SS_Y$, $DS_X = DS_Y$, $SD_X = S_{DY}$, and $DD_X = DD_Y$. Then, order the individuals SS, DS, SD, DD, S, and D in a matrix array that

contains the relationships for the ancestors of X (in the rows) with the ancestors of Y (in the columns), which results in:

$$\Sigma_{FS} = 0.5 + 2(\beta_{SX}\beta_{SY} + \beta_{DX}\beta_{DY})$$

Using these elements in Eq. (52) results in:

$$\Sigma_{FS} = 0.5 + 2(\beta_{SX}\beta_{SY} + \beta_{DX}\beta_{DY}) \tag{53}$$

By a similar reasoning, we obtain the following expressions for the relationships between paternal and maternal half-sibs, respectively:

$$\Sigma_{PHS} = 0.25 + 2\beta_{SX}\beta_{SY}$$

$$\Sigma_{MHS} = 0.25 + 2\beta_{DX}\beta_{DY} \tag{54}$$

It can be inferred that twice the product of the β parameters of the two sibs from either the paternal or maternal (half-sibs), or for both (full-sibs), results in the value of identity disequilibrium. Also, please note that the magnitude of a half-sib relationship can be greater than 0.25 whenever X and Y possess β with the same sign, meaning that both half-sibs are more alike the same grandparent. Conversely, when X and Y differ in the signs of their β, the covariance between BV is smaller than 0.25. The resemblance is less clear for full-sibs because their β need to agree in sign *simultaneously* with respect to the paternal and maternal sources of resemblance.

Appendix 4

The AR and PAR models are not equivalent

Two causal models are "Markov equivalent" whenever their associated direct acyclic graph (DAG) captures the same dependence–independence structure and the same

Fig. 1 The AR and the PAR models for the breeding values of two half-sibs (X and Y), the common parent (P) and the parents of P (GS and GD)

set of conditional independencies [47, 48]. As an example, Appendix Fig. 1 displays the DAG of half-sibs in the absence of any inbreeding of the grandparents (GS and GD), the common parent (P) and both half-sibs, for the AR and the PAR models. Please note that there are extra treks in the AR model going from a_X to a_Y through each grandparent without passing by the BV of the parent.

Reading out the DAG in Appendix Fig. 1, the covariance between the BV of X and Y involves two parameters in both the AR (β_X and β_Y) and PAR ($\rho_{P,X}$ and $\rho_{P,Y}$) models. However, in the PAR model a_X and a_Y are conditionally independent on the BV of their grandparents GS and GD given that of their common parent P, thus the edges are not represented in the DAG. As a result, a_X is conditionally independent of a_Y given a_P, a_{GS} and a_{GD} under the AR model, while under the PAR model a_X is conditionally independent of a_Y given a_P only.

Please note that, in Appendix Fig. 1, the regression coefficients β_X, $-\beta_X$, β_Y and $-\beta_Y$ directly connect a_{GS} and a_{GD} to a_X and a_Y. These path (or regression) coefficients that are worth plus or minus beta are called *chords* and the graph that contains them is named a *chordal graph* [46], whereas the graphs that do not contain chords are called *chordless graphs*. Consequently, AR is a chordal graph whereas PAR is a chordless graph. Theorem 2 in [48] says that two graphs are Markov equivalent models if and only if they do not contain any chordless *collider* structure in four nodes. A collider dependence structure (see page 10 in [4]) is that for "progeny in common":

$$a_{GP1} - 0.5 \rightarrow a_{Sire} - 0.5 \rightarrow X \leftarrow 0.5 - a_{Dam} \leftarrow 0.5 - a_{GP2}$$

or for "grand-progeny in common":

$$a_{GS} - \beta_X, \rightarrow X \leftarrow -\beta_X - a_{GD}$$
$$\text{and } a_{GS} - \beta_Y, \rightarrow Y \leftarrow -\beta_Y - a_{GD},$$

in the same DAG. Colliders can occur in chordal and chordless graphs, whereas non-collider structures such as the *lineage* $(1 \rightarrow 2 \rightarrow 3 \rightarrow 4)$ or the "common ancestor" (CA in $4 \leftarrow 2 \leftarrow CA \rightarrow 1 \rightarrow 3$) can only occur in chordless graphs. Hence, for pedigrees with four or more animals, the chordless PAR has a different conditional dependence structure than the chordal AR, and thus they are not Markov equivalent models.

Also, the AR and PAR models are not equivalent in the sense of Henderson [36], who defined two models as equivalent when they have the same covariance matrix. We will show this by a numerical example. Let $\beta_S = 0.1166$ and $\beta_D = 0.1226$, and the matrix in Eq. (17) be equal to:
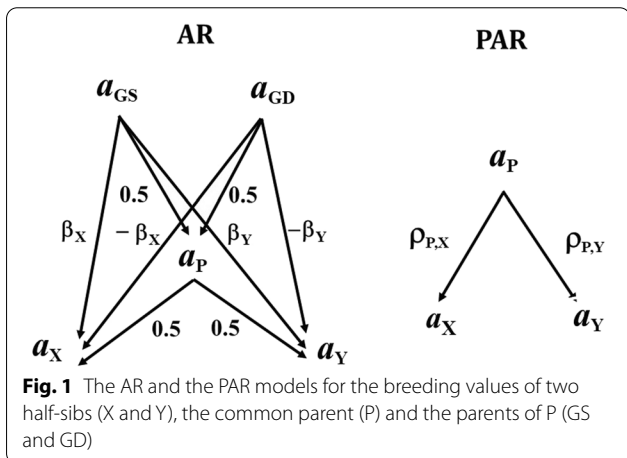
$$\Sigma_{\mathrm{A}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 1 & 0 & 0.5 \\ 0.5 & 0.5 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 1 \end{bmatrix}$$

Use of Eq. (30) results in the covariance between the sire and X under the AR model being equal to:

$$\Sigma_{\mathrm{S,X}} = 0.5(1.0148 + 0) + 0.1166(0.50 - 0.5365)$$
$$+ 0.1226(0 - 0.0443) = 0.5085$$

In turn, we can derive the covariance between the BV of the sire and X under the PAR model as follows:

$$\begin{aligned} \mathrm{cov}_\rho(a_\mathrm{S}, a_\mathrm{X}) &= \mathrm{cov}_\rho(a_\mathrm{S}, \rho_\mathrm{S}, a_\mathrm{S} + \rho_\mathrm{D} a_\mathrm{D}) \\ &= \mathrm{cov}_\rho(a_\mathrm{S}, \rho_\mathrm{S}, a_\mathrm{S}) + \mathrm{cov}_\rho(a_\mathrm{S}, \rho_\mathrm{D} a_\mathrm{D}) \\ &= \rho_\mathrm{S}(1 + F_{\rho\mathrm{S}}) + \rho_\mathrm{D} \Sigma_{\rho\mathrm{DS}} \end{aligned}$$

Now, to calculate the numerical value of this covariance, we need to know the values of the ρ coefficients. These are obtained by solving the system of Eq. (23) or Eq. (24). In this example:

$$\begin{bmatrix} 0.4917 & -0.0088 \\ -0.0088 & 0.4864 \end{bmatrix} \begin{bmatrix} \rho_\mathrm{S} \\ \rho_\mathrm{D} \end{bmatrix} = \begin{bmatrix} 0.2547 \\ 0.2478 \end{bmatrix}$$

with solution $\rho_\mathrm{S} = 0.528$ and $\rho_\mathrm{D} = 0.519$. Then, the covariance is equal to:

$$\mathrm{cov}_{\mathrm{PAR}}(a_\mathrm{S}, a_\mathrm{X}) = 0.5274(1.0148) + 0.519(0) = 0.5352$$

Since $0.5352 \neq 0.5085$, the AR and PAR models are not covariance equivalent in the sense of [36].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12711-022-00750-6.

Additional file 1. Lineages, pedigree, graphs, and breeding values [49–55]. **Figure S1.** Example of a DAG for a small pedigree. **Figure S2.** Total, direct, and indirect effects of breeding values in an ancestral regression model.

Additional file 2. Breeding values under gametic equilibrium and disequilibrium [56–65]. **Figure S3.** DAG representation of ancestral regression (AR). **Figure S4.** Acyclic mixed graph representation of ancestral regression to parents (PAR).

Additional file 3. Covariance matrix of breeding values under identity disequilibrium [66–68].

Author details
[1] Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires, 1417 Ciudad Autónoma de Buenos Aires, Argentina. [2] Instituto de Investigaciones en Producción Animal (INPA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina.

References
1. Henderson CR. Applications of linear models in animal breeding. University of Guelph; 1984.
2. Bulmer MG. The effect of selection on genetic variability. Am Nat. 1971;105:201–11.
3. Bulmer MG. The mathematical theory of quantitative genetics. Oxford: Oxford University Press; 1980.
4. Pearl J. Causality: models, reasoning, and inference. New York: Cambridge University Press; 2000.
5. Kiiveri H, Speed TP, Carlin JB. Recursive causal models. J Aust Math Soc. 1984;36:30–52.
6. Misztal I, Legarra A, Aguilar I. Using recursion to compute the inverse of the genomic relationship matrix. J Dairy Sci. 2014;97:3943–52.
7. Weir BS, Cockerham CC. Pedigree mating with two linked loci. Genetics. 1969;61:923–40.
8. Cantet RJC, García-Baccino CA, Rogberg-Muñoz A, Forneris NS, Munilla S. Beyond genomic selection: the animal model strikes back (one generation)! J Anim Breed Genet. 2017;134:224–31.
9. Cockerham CC. An extension of the concept for partitioning hereditary variance for analysis of covariances among relatives, when epistasis is present. Genetics. 1954;39:859–82.
10. Kempthorne O. An introduction to genetic statistics. New York: Wiley; 1957.
11. Harris DL. Genotypic covariances between inbred relatives. Genetics. 1964;50:1319–48.
12. Malecot G. The mathematics of heredity. San Francisco: W. H. Freeman & Co.; 1969.
13. Lehmann EL. Some concepts of dependence. Ann Math Statist. 1966;35:1137–53.
14. Tiret M, Hospital F. Blocks of chromosomes identical by descent in a population: models and predictions. PLoS One. 2017;12:e0187416.
15. Thompson EA. Identity by descent: variation in meiosis, across genomes, and in populations. Genetics. 2013;194:301–26.
16. Vigeland MD. Relatedness coefficients in pedigrees with inbred founders. J Math Biol. 2020;81:185–207.
17. Sved JA. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor Popul Biol. 1971;2:125–41.

Cantet *et al. Genetics Selection Evolution*   (2022) 54:64

Page 14 of 14

18. Calus MPL, Vandenplas J. SNPrune: an efficient algorithm to prune large SNP array and sequence datasets based on high linkage disequilibrium. Genet Sel Evol. 2018;50:34.

19. Banos G, Coffey MP. Short communication: characterization of the genome-wide linkage disequilibrium in 2 divergent selection lines of dairy cows. J Dairy Sci. 2010;93:2775–8.

20. Watterson GA. On the equivalence of random mating and random union of gametes models in finite, monoecious populations. Theor Popul Biol. 1970;1:233–50.

21. Donnelly KP. The probability that related individuals share some section of genome identical by descent. Theor Popul Biol. 1983;23:34–63.

22. Guo SW. Proportion of genome shared identical by descent by relatives: concept, computation, and applications. Am J Hum Genet. 1995;56:1468–76.

23. Nolan JP. Erlang renewal models for genetic recombination. J Stat Distrib Appl. 2017;4:1–17.

24. Cockerham CC, Weir BS. Descent measures for two loci with some applications. Theor Popul Biol. 1973;4:300–30.

25. Laurie C, Weir BS. Dependency effects in multi-locus match probabilities. Theor Popul Biol. 2003;63:207–19.

26. Weir BS, Avery PJ, Hill WG. Effect of mating structure on variation in inbreeding. Theor Popul Biol. 1980;18:396–429.

27. Wright S. Path coefficients and path regressions. Alternative or complementary concepts? Biometrics. 1960;16:189–202.

28. Wright S. Coefficients of inbreeding and relationship. Am Nat. 1922;56:330–8.

29. Cantet RJC, Vitezica ZG. Properties of Mendelian residuals when regressing breeding values using a genomic covariance matrix. In: Proceedings of the 10th World Congress on Genetics Applied to Livestock Production: 17–22 August 2014; Vancouver; 2014.

30. Pourahmadi M. Covariance estimation: the GLM and regularization perspectives. Stat Sci. 2011;26:369–87.

31. Quaas RL. Additive genetic model with groups and relationships. J Dairy Sci. 1988;71:91–8.

32. Wermuth N, Cox DR. Joint response graphs and separation induced by triangular systems. J R Statist Soc B. 2004;66:687–717.

33. Henderson CR. Rapid method for computing the inverse of a relationship matrix. J Dairy Sci. 1975;58:1727–30.

34. Kaufmann D, Parbhoo S, Wieczorek A, Keller S, Adametz D, Roth V. Bayesian Markov blanket estimation. In: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics: 7–11 May 2016; Cadiz; 2016.

35. Quaas RL, Pollak EJ. Mixed model methodology for farm and ranch beef cattle testing programs. J Anim Sci. 1980;51:1277–87.

36. Henderson CR. Equivalent models to reduce computations. J Dairy Sci. 1985;68:2267–77.

37. Emik LO, Terril CE. Systematic procedures for calculating inbreeding coefficients. J Hered. 1949;40:51–5.

38. Chang HL, Fernando RL, Grossman M. On the principle underlying the tabular method to compute coancestry. Theor Appl Genet. 1991;81:233–8.

39. Fox CJ, Käufl A, Drton M. On the causal interpretation of acyclic mixed graphs under multivariate normality. Linear Algebra Appl. 2015;473:93–113.

40. Wright S. The method of path coefficients. Ann Math Stat. 1934;5:161–215.

41. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91:4414–23.

42. García-Baccino CA, Munilla S, Legarra A, Vitezica ZG, Forneris NS, Bates RO, et al. Estimates of the actual relationship between half-sibs in a pig population. J Anim Breed Genet. 2017;134:109–18.

43. Han L, Abney M. Identity by Descent estimation with dense genome-wide genotype data. Genet Epidemiol. 2011;35:557–67.

44. Forneris NS, Steibel JP, Legarra A, Vitezica ZG, Bates RO, Ernst CW, et al. A comparison of methods to estimate genomic relationships using pedigree and markers in livestock populations. J Anim Breed Genet. 2016;133:452–62.

45. Bonk S, Reichelt M, Teuscher F, Segelke D, Reinsch N. Mendelian sampling covariability of marker effects and genetic values. Genet Sel Evol. 2016;48:36.

46. Matsen FA, Evans SN. To what extent does genealogical ancestry imply genetic ancestry? Theor Popul Biol. 2008;74:182–90.

47. Pearl J, Verma T. The logic of representing dependencies by directed graphs. In: Proceedings of the 6th National Conference on Artificial Intelligence: 13–17 July 1987; Seattle; 1987.

48. Wermuth N, Sadeghi K. Sequences of regressions and their independences. TEST. 2012;21:215–52.

49. Li CC. Population genetics. Chicago: University of Chicago Press; 1955.

50. Arjas E, Eerola M. On predictive causality in longitudinal studies. J Stat Plan Inference. 1993;34:361–86.

51. Richardson T, Spirtes P. Ancestral graph Markov models. Ann Statist. 2002;30:962–1030.

52. Wermuth N, Cox DR. On association models defined over independence graphs. Bernoulli. 1998;4:477–95.

53. Baba K, Shibata R, Sibuya M. Partial correlation and conditional correlation as measures of conditional independence. Aust NZ J Stat. 2004;46:657–64.

54. Cochran WG. Recent work on the analysis of variance. J R Stat Soc. 1938;101:434–49.

55. Kenny DA. Correlation and causality. New York: Wiley; 1979.

56. Kennedy BW, Schaeffer LR, Sorensen DA. Genetic properties of animal models. J Dairy Sci. 1988;71:17–26.

57. Schaeffer LR, Kennedy BW, Gibson JP. The inverse of the gametic relationship matrix. J Dairy Sci. 1989;72:1266–72.

58. Searle SR. Average relationship within a herd. J Dairy Sci. 1963;46:447–51.

59. Foulley JL, Chevalet C. Méthode de prise en compte de la consanguinité dans un modèle simple de simulation de performances. Ann Génét Sél Anim. 1981;13:189–96.

60. Tachida H, Cockerham CC. Effects of identity disequilibrium and linkage on quantitative variation in finite populations. Genet Res. 1989;53:63–70.

61. Esary JD, Proschan F, Walkup DW. Association of random variables, with applications. Ann Math Stat. 1967;38:1466–74.

62. Veller C, Edelman NB, Muralidhar P, Nowak MA. Variation in genetic relatedness is determined by the aggregate recombination process. Genetics. 2020;216:985–94.

63. Wang M, Xu S. Statistics of Mendelian segregation—a mixture model. J Anim Breed Genet. 2019;136:341–50.

64. Cox JT, Grimmett G. Central limit theorems for associated random variables and the percolation model. Ann Probab. 1984;12:514–28.

65. Tallis GM. On the joint asymptotic distribution of additive genotype for polygenic characters. J Math Biol. 1986;24:233–5.

66. Bickel P, Doksum K. Mathematical statistics: basic ideas and selected topics. 1st ed. Upper Saddle River: Prentice-Hall; 1977.

67. Andersson SA, Perlman MD. Normal linear regression models with recursive graphical Markov structure. J Multivar Anal. 1998;66:133–87.

68. Rasmussen CE, Williams CKI. Gaussian processes for machine learning. Boston: MIT Press; 2006.

Publisher's Note