

A proximal decomposition of inbreeding, coancestry and contributions

JEAN-JACQUES COLLEAU^{1*} AND MEHDI SARGOLZAEI²

¹INRA, UR337, Station de Génétique Quantitative et Appliquée, 78352 Jouy-en-Josas Cedex, France

²Centre for Genetic Improvement of Livestock, Department of Animal and Poultry Science, University of Guelph, Guelph, ON, N1G 2W1 Canada

(Received 30 April 2007 and in revised form 6 September and 12 December 2007)

Summary

Wright's (1922) coalescence probabilistic method decomposes the average coancestry or the average inbreeding of the population of interest into contributions of ancestors who are directly responsible for coancestry. These ancestors are less numerous and younger than those given by Caballero & Toro's (2000) method. Then, Wright's method yields a more parsimonious representation of the origin of coancestry or inbreeding. Results of Wright's method, called proximal because ancestors are younger, can be obtained from the latter method using a simple matrix transformation. Detailed algorithms are presented. Both methods can be used for assessing the overall efficiency of a breeding scheme in balancing genetic gain and coancestry rate, yielding very similar results when generations are discrete.

1. Introduction

Loss of genetic diversity or increase of inbreeding is of special concern in either captive or selected populations. In order to monitor and manage genetic diversity the genealogical information is usually condensed into a few parameters such as average inbreeding coefficients (\bar{F} ; Wright, 1922), average kinship (Malécot, 1948) or coancestry (\bar{f} ; Cockerham, 1970), effective population size (Wright, 1931), founder equivalents, founder genome equivalent (Lacy, 1989), effective number of ancestors, and founders and non-founders (Boichard *et al.*, 1997; Caballero & Toro, 2000). The most relevant indicator of genetic variability is \bar{f} , including self-coancestries (Ballou & Lacy, 1995; Caballero & Toro, 2000; Fernandez *et al.*, 2003). Selection for minimum coancestry has been shown to reduce the loss of genetic variation or the rate of inbreeding in the long term more effectively than other strategies such as equalizing the founder contributions or minimizing immediate inbreeding (Ballou & Lacy, 1995; Meuwissen, 1997; Caballero *et al.*, 1996; Caballero & Toro, 2000; Sanchez *et al.*, 2003).

Examination of detailed contributions of ancestors gives a better understanding of the development of inbreeding or coancestry over time, in terms of the origin of genes, in general, or the origin of genetic gains in selected populations. The expression 'origin of genes' is sometimes misleading in the literature when it refers to non-founder ancestors. Strictly speaking, unless there are mutations, only founder genes are segregating. Instead, 'origin of breeding values (BVs)' may be preferred, where BVs concern a hypothetical polygenic neutral trait. Throughout the following text, the term 'BV' will refer to the BV for this latent trait.

Some studies (e.g., Woolliams & Thompson, 1994) have insisted on the prominent role played by Mendelian sampling components of BVs. Basically, considering such a trait is quite relevant because covariances between individuals for this trait are equal to twice the coancestry coefficient, when genetic variance is 1. Following this rationale, inbreeding coefficients can be expressed in terms of contributions of parents and Mendelian sampling variances (MSVs) over generations (Woolliams & Thompson, 1994; Woolliams *et al.*, 1999). Conceptually, contribution of an ancestor to an individual quantifies the fraction of ancestral Mendelian sampling components of BV

* Corresponding author. e-mail: ugenccj@dga2.jouy.inra.fr

passed to this individual (James & McBride, 1958). In this context, Caballero & Toro (2000) gave the expression of \bar{F} or \bar{f} in a population in terms of contribution of MSVs from founders and subsequent ancestors. Here contributors are common ancestors (CAs). Later on, this approach will be denoted as the ‘conventional’ approach.

Wright (1922) presented the first method to compute coefficients of inbreeding and relationships. In contrast to the conventional method, this coalescence probabilistic method partitions inbreeding into contributions of genes of ancestors, through a recursive path counting process. This approach can be used for partitioning \bar{F} or \bar{f} in a population, but it is time-consuming. The most striking consequence of using this approach is that the list of contributors is modified, and basically shortened, because the upward search can be stopped as soon as coalescence is obtained (i.e., inbreeding in an intermediate ancestor is encountered). Then, Wright’s approach consists of determining the most recent occurrences of inbreeding, irrespective of the genes being inbred. Remaining contributors can be called ‘nodal common ancestors’ (NCAs). The other CAs can be considered as ‘absorbed CAs’ (ACAs) because their genes contribute to inbreeding or coancestry through the NCAs. Probability of absorption increases when age of ancestors increases, because probability of direct access to inbreeding by older ancestors decreases.

In the present article, the results of Wright’s method can be fully recovered by a non-probabilistic method, analytically linked with Caballero & Toro’s (2000) method, still considering the conventional approach or MSVs. Wright’s decomposition is denoted ‘proximal’ because NCAs are usually younger and less numerous than CAs in the conventional analysis, due to a backward or coalescence decomposition process. Finally, the efficiency of past management of simulated selected populations will be examined by the conventional and the proximal analyses.

2. Materials and methods

(i) Recovering Wright’s decomposition from Caballero & Toro’s decomposition

Caballero & Toro’s decomposition method can be shown to be obtained exactly after multiplication of Wright’s decomposition by an upper triangular matrix, and then that the reverse can be carried out easily and very fast, due to this simplicity.

Statistics of interest are average pairwise coancestry (\bar{f}_g) or average inbreeding (\bar{F}_g) for a group of individuals. The relationship matrix (**A**) can be decomposed as:

$$\mathbf{A} = \mathbf{CDC}'$$

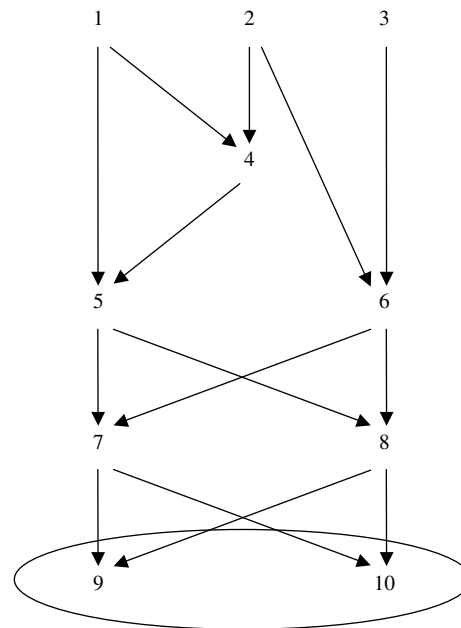


Fig. 1. Example pedigree.

where $\mathbf{C} = \{c_{ij}\} = (\mathbf{I} - \mathbf{T})^{-1}$ is the contribution matrix, a unit lower triangular matrix containing the fraction of genes that individual j passed to descendant i (contribution of j to i in the sense of James & McBride, 1958), \mathbf{T} is a unit lower triangular sparse matrix consisting of only -0.5 at positions where individual i is linked to its sire and dam, $\mathbf{D} = \text{diag}\{d_{ii}\}$ is a diagonal matrix containing the MSVs of individuals. The MSV for individual i is $d_{ii} = 1 - 0.25(a_{s_i s_i} + a_{d_i d_i}) = 0.5 - 0.25(F_{s_i} + F_{d_i})$, where F_{s_i} and F_{d_i} are inbreeding coefficients of the sire and dam of individual i , respectively (Quaas, 1976). Now \bar{f}_g can be decomposed into vector $\mathbf{m} = \{m_j\}$, where m_j is the contribution of MSV of ancestor j . Such a decomposition is referred to as the m -decomposition. Following Caballero & Toro (2000), vector \mathbf{m} is:

$$m_j = 0.5 \left(\sum_{i \in G} c_{ij} \right)^2 d_{jj} n_g^{-2}$$

where n_g is the size of the group. The term $\sum_{i \in G} c_{ij}$ can be readily obtained by the indirect method (Colleau, 2002), which requires tracing back the pedigree only once. Figure 1 shows an example pedigree for illustrating the method.

Let the group in question be individuals 9 and 10. Then vector \mathbf{m} for \bar{f}_g is:

$$\mathbf{m}' = [0.0703 \quad 0.0703 \quad 0.0313 \quad 0.0156 \quad 0.0625 \\ 0.0625 \quad 0.0547 \quad 0.0547 \quad 0.0586 \quad 0.0586]$$

The contribution of MSV of ancestor j to \bar{F}_g can be derived in the same manner:

$$m_j = 0.5 \left(\sum_{k \in G} c_{s_k j} c_{d_k j} d_{jj} \right) n_g^{-1}$$

To obtain vector \mathbf{m} for \bar{F}_g using the indirect method, the pedigree should be traced back twice the number of sires of the group considered, because the sum corresponding to a given sire can be obtained from two back explorations of the population pedigree (Colleau, 2002; Sargolzaei *et al.*, 2005). Vector \mathbf{m} for the \bar{F}_g of the group consisting of individuals 9 and 10 is then:

$$\mathbf{m}' = \begin{bmatrix} 0.0703 & 0.0703 & 0.0313 & 0.0156 & 0.0625 \\ 0.0625 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}.$$

The probabilistic approach of path counting (Wright, 1922) is theoretically well suited to partition the inbreeding coefficients into contributions of genes of NCAs. An NCA can be defined as an ancestor who connects the sire and dam lines that arise from the individual of interest or from its NCA(s). The NCA makes an inbreeding loop in which no individual is present more than once. The details of the path counting method are presented in Appendix A.

$$\Psi = \begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.2500 & 0.3125 & 0.0000 & 0.1406 & 0.1406 & 0.0703 & 0.0703 \\ 0.0000 & 1.0000 & 0.0000 & 0.2500 & 0.0625 & 0.2500 & 0.0781 & 0.0781 & 0.0703 & 0.0703 \\ 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.2500 & 0.0625 & 0.0625 & 0.0313 & 0.0313 \\ 0.0000 & 0.0000 & 0.0000 & 0.5000 & 0.1250 & 0.0000 & 0.0313 & 0.0313 & 0.0156 & 0.0156 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5000 & 0.0000 & 0.1250 & 0.1250 & 0.0625 & 0.0625 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5000 & 0.1250 & 0.1250 & 0.0625 & 0.0625 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4375 & 0.0000 & 0.1094 & 0.1094 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4375 & 0.1094 & 0.1094 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4688 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4688 \end{bmatrix}.$$

Now, let the results of the path counting method be represented by the known vector $\mathbf{w} = \{w_i\}$, where w_i indicates the contribution of genes of NCA i to the statistics of interest to be partitioned. For simplicity, in the following development, \bar{F}_g is used as the statistic of interest. However, the derivation holds for \bar{f}_g because \bar{f}_g can be considered as \bar{F}_g resulting from a factorial mating design between members of the group investigated.

The relation $1 = 0.25(\text{var}(g_{s_i}) + \text{var}(g_{d_i})) + 0.5 - 0.25(F_{s_i} + F_{d_i})$ is well known. Note that $0.5 - 0.25(F_{s_i} + F_{d_i})$ is the MSV of individual i . Then, recursive decomposition of the genetic variances of parents of i across generations, leads one to find that 1 is the sum of a series involving MSVs of all the ancestors of individual i . Appendix B shows that this development is also the development of one certain occurrence of inbreeding in individual i in terms of MSV.

The decomposition of parental variances is carried out using corresponding rows of matrix \mathbf{C} , which can be computed by tracing back the pedigree of each parent. Let column vector $\boldsymbol{\psi}_i = \{\psi_{ki}\}$ represent the coefficients of this development, where ψ_{ki} indicates

the contribution of ancestor k to inbreeding in i , given the contribution of elder ancestors. Because these weights sum up to 1 ($\mathbf{1}\boldsymbol{\psi}_i = 1$), they can be directly interpreted as the relative probability of origin of the inbreeding for a gene of a given individual, when inbreeding is certain.

If inbreeding through a given individual has probability w_i , then the contribution of individual i to vector \mathbf{m} is $w_i\boldsymbol{\psi}_i$ and finally:

$$\mathbf{m} = \Psi\mathbf{w},$$

where the transition matrix Ψ from \mathbf{w} to \mathbf{m} gathers all the involved coefficients and is an upper triangular matrix. Off-diagonal elements of column i of matrix Ψ are:

$$\psi_{ki} = 0.25(c_{s_i,k}^2 + c_{d_i,k}^2)d_{kk}$$

and diagonal elements of matrix Ψ are simply d_{kk} . Matrix Ψ for the example pedigree is shown below:

A remarkable result is that for finding \mathbf{w} , the equation above does not involve MSVs, but only contributions. This is apparent after dividing each row k of the linear system by the corresponding d_{kk} .

Solve for \mathbf{w} as $\mathbf{w} = \Psi^{-1}\mathbf{m}$. The inverse of matrix Ψ is sparse but cannot be readily obtained because the distribution of non-null terms depends on the unknown situation concerning vector \mathbf{w} , the issue to be solved. However, Ψ is upper triangular and \mathbf{w} can be obtained efficiently from \mathbf{m} by Gaussian elimination because each column of Ψ can be formed separately by tracing back the pedigrees of the two parents involved. Starting from the vector \mathbf{m} pertaining to decomposition of the average inbreeding coefficient of individuals 9 and 10,

$$\mathbf{w}' = \begin{bmatrix} 0.0313 & 0.0313 & 0.0000 & 0.0000 & 0.1250 \\ 0.1250 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix},$$

i.e., exactly the result obtained by the path counting method (see Appendix A). For each individual, distinct computations of contribution vectors of sires and dams are needed, which might be time-consuming

in large populations. However, substantial computation time can be saved if the number of sires is much less than the number of dams (a very frequent situation) and if the population is subdivided into pseudo-generations (Colleau, 2002). In this situation, solutions for \mathbf{w} during Gaussian elimination can be obtained simultaneously by blocks and updating of the right-hand side needs only one trace back per sire involved in this pseudo-generation. The m -decomposition includes the effect of self-coancestries of the individuals of the current population, when the statistic considered is the average relationship. If the weighted average coancestry between parents of

than in the conventional model. Regression coefficients and conditional variances in this context cannot be easily linked to terms of the w -decomposition, because proximal ancestors can have one or even both parents not included in the list of proximal ancestors. Basically, they should be obtained by a numerical method, such as Cholesky decomposition of the relationship matrix of the proximal ancestors.

$$\mathbf{A}^* = \mathbf{C}^* \mathbf{D}^* \mathbf{C}^{*\prime}$$

For the example pedigree, if the cohort of interest contains individuals 9 and 10 then \mathbf{C}^* matrix is:

NCA	1	2	5	6	7	8	9	10
$\mathbf{C}^* =$	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.750	0.250	1.000	0.000	0.000	0.000	0.000	0.000
	0.000	0.500	0.000	1.000	0.000	0.000	0.000	0.000
	0.375	0.375	0.395	0.433	1.000	0.000	0.000	0.000
	0.375	0.375	0.395	0.433	0.000	1.000	0.000	0.000
	0.375	0.375	0.395	0.433	0.331	0.331	1.000	0.000
	0.375	0.375	0.395	0.433	0.331	0.331	0.000	1.000

the current population is preferred because this bias is avoided, the corresponding modified w -decomposition is easy to implement (see Appendix C).

(ii) *Defining a proximal model of breeding values*

In the conventional context, let g_i be the BV of individual i , c_{ik} is the contribution of ancestor k to individual i and δ_{ik} is the Mendelian sampling deviation of ancestor k , then $g_i = \sum_k c_{ik} \delta_k$. The model for BVs is straightforward and leads one naturally to decomposition of variances or covariances, as presented above.

In the proximal context, introduced by Wright's decomposition, quite the reverse occurs: partitioning of variances or covariances is obtained without explicitly defining the model for BVs. A model based on the conventional model is proposed, where the independent genetic variables are deviations of BVs of proximal ancestors (or NCA) from their expectations given BVs of all older proximal ancestors. In contrast, in the conventional model, BVs are only conditioned on parental BVs (two individuals) for providing Mendelian sampling BV. Then, the corresponding regression coefficients are the proximal contributions, i.e., $g_i = \sum_k c_{ik}^* \delta_k^*$, where k is the rank number in the list of proximal ancestors and δ_k^* is the deviation of BV of ancestor k from its expectation given BVs of proximal ancestors 1, ..., $k - 1$. In this model, variances of the random variables can be much more heterogeneous

In selected populations, the origin of genetic gain can be re-examined in a proximal perspective, analogously to Woolliams *et al.* (1999) for the conventional model. The contribution of ancestor k to genetic gain is equal to, $\bar{c}_k^* \delta_k^*$, where \bar{c}_k^* and δ_k^* are obtained based on the Cholesky decomposition.

When BVs of the trait under selection are estimated by best linear unbiased prediction (BLUP), then we also have $\hat{\delta}_k^*$. Finally, the estimated contribution of ancestor k to genetic gain is equal to $\bar{c}_k^* \hat{\delta}_k^*$. Furthermore, the efficiency of the breeding scheme can be examined in retrospect. In the next section, the proximal decomposition is compared to the conventional decomposition of \hat{f} to assess the maximal estimated genetic gain possible while constraining this coancestry to the observed value.

(iii) *Measuring the retrospective efficiency of a breeding scheme*

The inefficiency of a selection breeding scheme for balancing genetic gain and \hat{f} can be measured by the discrepancy between the observed genetic gain and the maximum genetic gain that could have been generated considering the ancestors of the current population and constraining \hat{f} to its observed value. Ideal contributions for producing maximum genetic gains can be determined either by the conventional approach or by the proximal approach. Conventional BLUP selection considers a fixed number of breeding

individuals selected based only on their EBVs without any other consideration. In contrast, optimized schemes such as those defined by Meuwissen (1997) allow contributions of selected breeding individuals to fluctuate permanently according to the balance between their EBVs and the extra coancestry they could generate.

Simulation was carried out to assess the respective potential of the conventional or proximal methods of back-optimization when applied to breeding schemes implementing various degrees of optimization. The founder population consisted of 25 sires and 50 dams, and 15 discrete generations were considered (founder generation being generation 0). Selection was carried out on a single trait with heritability of 0.2 (initial genetic variance set to 1), expressed by both sexes. In BLUP selection schemes, each sire is mated to two dams at random according to a hierarchical design. Each dam produces four progeny. In optimized breeding schemes, a hierarchical design and random mating were maintained, whereas the number of progeny per sire and per dam was allowed to vary. Four intermediate schemes were added starting with BLUP and switching towards an optimized scheme at generations 14, 11, 8 and 5. One hundred replicates of the six schemes were simulated by sextuplets. For increasing the efficiency of comparison between schemes, the same set of founders was considered for each member of a given sextuplet. Furthermore, intermediate schemes were initiated from the data of the BLUP member until the switch. Optimized and intermediate breeding schemes of a given sextuplet were constrained to produce, at each generation, the same coancestry as their BLUP counterpart.

For each breeding scheme and each replicate, the maximum genetic gain obtained in retrospect was determined analytically by maximizing a Lagrange function where unknowns were optimal contributions as the main constraint, so the coancestry between parents of generation 15 should be equal to that observed. For the optimization in retrospect based on conventional contributions, additional constraints were introduced. Firstly, sums of contributions per generation \times sex combination should obviously equal 0.5. Secondly, contributions of CA with a single progeny in the list of CAs were constrained to be equal to 0.5 contributions of respective progeny: in our simulations, because these CAs were always found to be the ACAs. Detailed algorithms can be found in Appendix D.

3. Results

Table 1 shows the results obtained in generation 15 for BLUP and optimized schemes. Coancestry coefficients were similar in both schemes as expected, 27.5 and 27.6%, respectively. Genetic gain (average of

Table 1. Results of simulations (generation 15)

	BLUP schemes	Optimized schemes
Average breeding value	6.31	8.80
Average coancestry (%)	27.5	27.6
Average inbreeding (%)	26.6	26.0
Number of ancestors (founders)		
Conventional	498 (28)	350 (14)
Proximal	296 (14)	200 (10)
Age of ancestors		
Conventional	7.2	7.7
Proximal	6.2	6.8

true breeding values) was 39% higher (8.8 vs 6.3) in the optimized breeding scheme in comparison to BLUP. In both schemes, the number of proximal ancestors of the last generation was substantially lower than the number of ancestors given by the conventional approach, by about 42%. As expected, ancestors found by the proximal approach were also younger (by about 1 generation). Estimated genetic gains were directly obtained based on the EBV for the selected trait. They were 6.28 for BLUP schemes and 8.82 for optimized schemes, i.e., very close to the true genetic gains. Maximum estimated genetic gains obtained by optimizing in retrospect based on conventional contributions were 7.61 (+21%) for BLUP and 9.21 (+5%) for optimized. Using the proximal approach would have yielded 7.79 (+24%) for BLUP and 9.02 (+2%) for optimized. Results given by both approaches were similar. The optimal genetic gain obtained in retrospect was far from the gain exhibited by permanent optimization, thus illustrating the impact of drift and the loss of valuable ancestors before the optimization in retrospect.

Intermediate schemes showed an interesting pattern. At the generation corresponding to the switch, a lift was observed (approximately doubling the instantaneous genetic gain) and, afterwards, rates of gain were quite similar to those of the optimized schemes. The estimated genetic gains for these four schemes (very close to the true genetic gains), of increasing efficiency, were 6.7, 7.1, 7.7 and 8.3. Corresponding maximal genetic gains were 7.7 (+14%), 7.9 (+12%), 8.3 (+9%) and 8.7 (+5%) according to the conventional decomposition and 7.7 (+14%), 7.9 (+12%), 8.2 (+8%) and 8.6 (+4%) according to the proximal decomposition. Once again, both decompositions gave similar results. The overall comparison of the six schemes confirmed that the efficiency of a breeding programme can be safely assessed by the discrepancy between the estimated genetic gain and the one of back optimization: when efficiency increased, discrepancy decreased.

4. Conclusion

Wright’s (1922) method for decomposing inbreeding and coancestry could be linked exactly with the conventional approach summarized by Caballero & Toro (2000). Wright’s decomposition led exactly to the well-established conventional decomposition. Analytical calculations were relatively straightforward. However, Wright’s coalescence approach yielded fewer and younger ancestors, and hence was called a ‘proximal’ approach. If the same information was given by a smaller number of ancestors, then this meant that information given by the absorbed ancestors was totally dependent on the information given by the remaining ancestors, as a special consequence of drift. In the general situation, involving overlapping generations, identification of absorbed ancestors required the full analytical calculations presented above. Prediction of status (ACA vs NCA) by simple methods would be highly desirable for understanding the mechanisms involved and also to skip useless calculations. However, early attempts failed and further research is needed in this area.

Simulation showed that the discrepancy between both approaches in terms of number of ancestors could be substantial. Inbreeding and coancestry could be explained parsimoniously by fewer ancestors capturing the whole information due to genetic drift. For selected populations, contributions of ancestors to genetic gains should be considered too. Simulation on discrete generations showed that evaluation of the past efficiency of breeding schemes through back-optimization was quite similar using both decompositions. However, in the context of discrete generations, constraints occurring on conventional contributions could be identified clearly. For overlapping generations, in contrast, the proximal approach might be more robust in such a situation. Research is needed in this area.

Appendix A. Tabular implementation of Wright’s method

The basic formula for computing inbreeding coefficients, which is based on computation of contributions of genes of NCAs, is:

$$F_i = \sum_{j \in NCA_i} \left(\sum_{k \in L_j} 0.5^{n1_k + n2_k + 1} (1 + F_j) \right) \Rightarrow$$

$$F_i = \sum_{j \in NCA_i} \left(\sum_{k \in L_j} 0.5^{n1_k + n2_k + 1} + \sum_{k \in L_j} 0.5^{n1_k + n2_k + 1} F_j \right), \tag{A1}$$

where F_i and F_j are inbreeding coefficients of individuals i and j , respectively, NCA_i is a set of identification numbers of NCAs of individual i , and L_j is a

set of inbreeding loops for NCA j of individual i ; $n1_k$ and $n2_k$ are the number of generations from sire and dam of individual i , respectively, to NCA j in loop k . The first term is the contribution of genes of individual j regardless of its inbreeding and the second term is the contribution of genes of NCAs of individual j to inbreeding of individual i . Inbreeding coefficients of the NCA can be expressed in the same way and therefore (A1) can be expanded recursively. After expansion, each term would be the contribution of genes of the corresponding NCA (regardless of its inbreeding) to the inbreeding of individual i .

As an example, take the pedigree of individual 10 (Fig. 1). There are three NCAs for individual 10, namely 5, 6 and 2. From these three, only individual 5 is inbred. So, in addition, we need to consider individual 1, who is the only NCA of individual 5. The NCAs 1, 2, 5 and 6 contribute 0.03125, 0.03125, 0.125 and 0.125 to the inbreeding coefficient of individual 10 ($F_{10} = 0.3125$). Contributions of individuals 1 and 2 are marginal.

The same results as those by Wright’s method can be recovered by using an upward exploration method, faster but still storage-demanding. The method is tabular and adapted to partition \bar{F}_g or \bar{f}_g for a specified group of individuals. The outline of the algorithm is as follows:

The identifications (ID) should be consecutive integers and parents precede progeny.

1. Define a table of three columns and initialize all elements to 0.
2. Set the first and the second columns for the m first rows to the ID of two parents of m individuals in question. Add the minimum ID in the first column.
3. Set the third column of the m rows to 0.5.
4. Process from the largest generation to 0.

For overlapping generations, consider pseudo-generations.

- (i) For the 1st and 2nd columns, in turn, expand the table by adding two parents of individuals belonging to the current generation to the end of table while duplicating the corresponding ID in other column and adding half of the corresponding contribution to the 3rd column. Always place the minimum ID in the 1st column and avoid repeated rows by summing their contributions.
- (ii) If IDs in the 1st and the 2nd columns are identical, then this is a NCA; therefore add its ID and contribution to the result table.

The resulting table will contain the ID and contribution of NCAs to the inbreeding of the individuals of interest. Table A1 shows the steps of the tabular algorithm applied to the example pedigree for decomposing the average inbreeding of individuals 9 and 10.

Table A1. Implementation of the tabular algorithm to decompose the average inbreeding of individuals 9 and 10 of the example pedigree

Generation processed	Identification		Contribution (3rd column)	NCA	Contribution
	1st column	2nd column			
4	7	8	0.5+0.5		
3	5	5	0.25	5	0.25/2 ^a
	5	6	0.5		
	6	6	0.25	6	0.25/2
2	1	4	0.125		
	6	6	0.25		
	1	6	0.25		
	4	6	0.25		
1	2	3	0.1875		
	1	1	0.0625	1	0.0625/2
	1	2	0.25		
	1	3	0.1875		
	2	2	0.0625	2	0.0625/2

^a 2 is the number of individuals in the group considered.

Appendix B. Detailed proof that *m*-decomposition can be recovered from *w*-decomposition

We partition one occurrence of inbreeding through individual *i* at locus *k* into link probabilities with inbred states of ancestors, given that the polygenic neutral trait involves *L* unlinked loci.

Transform the probability problems into problems of genetic variance decomposition. Sire of individual *i* is *s_i* and its dam is *d_i* with genetic values at locus *k* equal to *g_{s_i}^k* and *g_{d_i}^k*, respectively. Genetic and gametic variances at locus *k* are 1/*L* and 0.5/*L*, respectively. A gene sampled from the sire has an expected genetic value of 0.5*g_{s_i}^k* and of sampling variance equal to 0.25(1 - *F_{s_i}*)/*L*.

Then, conditioning on inbreeding occurring, the same gene from the sire is sampled twice so that the expectation of the sum of the genetic values is *g_{s_i}^k* and its sampling variance is equal to (1 - *F_{s_i}*)/*L*. Because overall variance of the sum of the genetic values corresponding to these genes is equal to 2/*L*, we have:

$$2/L = \text{var}(g_{s_i}^k) + (1 - F_{s_i})/L.$$

The same is true when the gene comes from the dam. When the same gene is sampled twice, the probability for paternal or maternal origin is 0.5. Then, the conditioned variance of the BV at locus *k* is equal to 2/*L* and to

$$0.5(\text{var}(g_{s_i}^k) + \text{var}(g_{d_i}^k)) + 1/L - 0.5(F_{s_i} + F_{d_i})/L.$$

If *w_i* is the probability that the gene sampled twice comes from ancestor *i*, then participation of ancestor *i* to drift at locus *k* is

$$w_i(0.5(\text{var}(g_{s_i}^k) + \text{var}(g_{d_i}^k)) + 1/L - 0.5(F_{s_i} + F_{d_i})/L).$$

Summing over all independent loci, with the same probability of inbreeding, shows that the overall contribution of ancestor *i* to drift is equal to

$$w_i(0.5(\text{var}(g_{s_i}) + \text{var}(g_{d_i})) + 1 - 0.5(F_{s_i} + F_{d_i})),$$

and after division by 2, because relationship coefficient is twice coancestry coefficient, we have another expression of the contribution of ancestor *i* to the average coancestry, i.e.,

$$w_i(0.25(\text{var}(g_{s_i}) + \text{var}(g_{d_i})) + 0.5 - 0.25(F_{s_i} + F_{d_i})).$$

Appendix C. Removing the effect of self-contributions from vector *m*

Let \bar{g} be the average BV of the current population, \bar{g}_p and \bar{g}_m the average BVs of sires and dams. Then, $\bar{g} = 0.5(\bar{g}_p + \bar{g}_m) + \bar{s} = \bar{g}^* + \bar{\delta}$, where $\bar{\delta}$ is the average Mendelian sampling BV. Drift due to parents is equal to the variance of the parental mean, i.e., the average weighted relationship coefficient between parents, where weights are proportional to progeny size and sum to 0.5 per sex. If we express drift in terms of coancestry (*f*), then this value should be multiplied by 0.5. A direct adaptation of the *m*-decomposition for \bar{f} is possible but still complicates calculation.

Then, an indirect approach is proposed that is much simpler, where the *m*-decomposition of \bar{f} is still carried out, but where the effect of $\bar{\delta}$ is easily removed. If \bar{c}_k is the average contribution of ancestor *k* and δ_k is his Mendelian sampling BV, then $\bar{g} = \sum_k \bar{c}_k \delta_k$. Index *k* includes the *N* indices *l* corresponding to the current population. For these indices, the direct impact of *l* is through *c_l* equal to 1/*N* for any

l (self-contribution). Then,

$$\bar{g}^* = \bar{g} - (1/N) \sum_l \delta_l = \sum_{k \neq l} \delta_k + \sum_l (\bar{c}_l - 1/N) \delta_l.$$

The new m -decomposition is the decomposition of $0.5\text{var}(\bar{g}^*)$ for l , and the new value is $m_l^* = 0.5(\bar{c}_l - 1/N)^2 d_k^*$. Given that $m_l = 0.5\bar{c}_l^2 d_l$, then $m_l^* = (\sqrt{m_l} - (1/N)\sqrt{2d_l^*})^2$.

Appendix D. Back-optimization of contributions

(i) Conventional contributions

The average coancestry between parents of the last generation is constrained to be φ . This coancestry is equal to $0.5\mathbf{x}'\mathbf{D}\mathbf{x}$, where \mathbf{x} is the vector of contributions of CA and where \mathbf{D} is the diagonal matrix involving their MSVs. The second constraint on the sums per sex \times generation can be written as $\mathbf{P}'\mathbf{x} = 0.5\mathbf{1}$, where \mathbf{P} is the incidence matrix relating the ancestors to the combined factor sex \times generation, and the third constraint on CAs with one progeny can be written as $\mathbf{K}'\mathbf{x} = \mathbf{0}$. For simplicity, we do not indicate the sizes of matrices and vectors involved. Given these constraints, the objective is to maximize the estimated genetic gain, i.e., $\hat{\boldsymbol{\delta}}'\mathbf{x}$, where $\hat{\boldsymbol{\delta}}$ is the vector of the estimated Mendelian deviations for the trait under selection. Then, the Lagrange function to optimize is

$$\hat{\boldsymbol{\delta}}'\mathbf{x} - \lambda(0.5\mathbf{x}'\mathbf{D}\mathbf{x} - \varphi) - \boldsymbol{\theta}'(\mathbf{P}'\mathbf{x} - 0.5\mathbf{1}) - \boldsymbol{\eta}'\mathbf{K}'\mathbf{x}.$$

The solution is $\mathbf{x} = \lambda^{-1}\mathbf{D}^{-1}(\hat{\boldsymbol{\delta}} - \mathbf{P}\boldsymbol{\theta} - \mathbf{K}\boldsymbol{\eta})$. Constraint 3 leads one to find that $\boldsymbol{\eta} = (\mathbf{K}'\mathbf{D}^{-1}\mathbf{K})^{-1}(\mathbf{K}'\mathbf{D}^{-1}\hat{\boldsymbol{\delta}} - \mathbf{P}\boldsymbol{\theta})$ and then that, $\boldsymbol{\eta}'\mathbf{K}'\mathbf{x} = \bar{\boldsymbol{\delta}}'\mathbf{x} - \boldsymbol{\theta}'\bar{\mathbf{P}}'\mathbf{x}$, where $\bar{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}} - \mathbf{K}(\mathbf{K}'\mathbf{D}^{-1}\mathbf{K})^{-1}\mathbf{K}'\mathbf{D}^{-1}\mathbf{P}\hat{\boldsymbol{\delta}}$ and $\bar{\mathbf{P}} = \mathbf{P} - \mathbf{K}(\mathbf{K}'\mathbf{D}^{-1}\mathbf{K})^{-1}\mathbf{K}'\mathbf{D}^{-1}\mathbf{P}$. The solution for $\boldsymbol{\theta}$ is $\boldsymbol{\theta} = (\bar{\mathbf{P}}'\mathbf{D}^{-1}\bar{\mathbf{P}})^{-1}(\bar{\mathbf{P}}'\mathbf{D}^{-1}\hat{\boldsymbol{\delta}} - 0.5\lambda\mathbf{1})$ and the solution for λ is +square root of the ratio $(\bar{\boldsymbol{\delta}}'\mathbf{D}^{-1}\hat{\boldsymbol{\delta}} - \mathbf{y}'\mathbf{H}\mathbf{y}) / (2\varphi - 0.25\mathbf{1}'\mathbf{H}\mathbf{1})$, where $\mathbf{H} = (\bar{\mathbf{P}}'\mathbf{D}^{-1}\bar{\mathbf{P}})^{-1}$ and $\mathbf{y} = \bar{\mathbf{P}}'\mathbf{D}^{-1}\hat{\boldsymbol{\delta}}$.

(ii) Proximal contributions

The Lagrange function to optimize is $\hat{\boldsymbol{\delta}}^*\mathbf{x}^* - \lambda(0.5\mathbf{x}^*\mathbf{D}^*\mathbf{x}^* - \varphi)$, with only a single constraint because other constraints on the solutions cannot be established analytically in the proximal model. Solutions are $\mathbf{x}^* = \lambda^{-1}\mathbf{D}^{*-1}\hat{\boldsymbol{\delta}}^*$ with $\lambda = \sqrt{(\hat{\boldsymbol{\delta}}^*\mathbf{D}^{*-1}\hat{\boldsymbol{\delta}}^*) / (2\varphi)}$ and maximum genetic gain equals $\sqrt{2\varphi\hat{\boldsymbol{\delta}}^*\mathbf{D}^{*-1}\hat{\boldsymbol{\delta}}^*}$.

References

- Ballou, J. D. & Lacy, R. C. (1995). Identifying genetically important individuals for management of genetic diversity in pedigreed populations. In *Population Management for Survival and Recovery* (ed. Ballou, J. D., Gilpin, M. & Foose, T. J.), pp. 76–111. New York: Columbia University Press.
- Boichard, D., Maignel, L. & Verrier, E. (1997). The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics, Selection, Evolution* **29**, 5–23.
- Caballero, A. & Toro, M. A. (2000). Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetical Research* **75**, 331–343.
- Caballero, A., Santiago, E. & Toro, M. A. (1996). Systems of mating to reduce inbreeding in selected populations. *Animal Science* **62**, 431–442.
- Cockerham, C. C. (1970). Avoidance and rates of inbreeding. In *Mathematical Topics in Population Genetics* (ed. K. Kojima), pp. 104–127. Berlin: Springer-Verlag.
- Colleau, J. J. (2002). An indirect approach to the extensive calculation of relationship coefficients. *Genetics, Selection, Evolution* **34**, 409–421.
- Fernandez, J., Toro, M. A. & Caballero, A. (2003). Fixed contributions designs vs. minimization of global coancestry to control inbreeding in small populations. *Genetics* **165**, 885–894.
- James, J. W. & McBride, G. (1958). The spread of genes by natural and artificial selection in a closed poultry flock. *Journal of Genetics* **56**, 55–62.
- Lacy, R. C. (1989). Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalence. *Zoo Biology* **8**, 111–124.
- Malécot, G. (1948). *Les Mathématiques de l'hérédité*. Paris, France: Masson et Cie.
- Meuwissen, T. H. E. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science* **75**, 934–940.
- Quaas, R. L. (1976). Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* **32**, 949–953.
- Sanchez, L., Bijma, P. & Woolliams, J. A. (2003). Minimizing inbreeding by managing genetic contributions across generations. *Genetics* **164**, 1589–1595.
- Sargolzaei, M., Iwaisaki, H. & Colleau, J. J. (2005). A fast algorithm for computing inbreeding coefficients in large populations. *Journal of Animal Breeding and Genetics* **122**, 325–331.
- Woolliams, J. A. & Thompson, R. (1994). A theory of genetic contributions. *Proceedings of the 5th Congress on Genetics Applied to Livestock Production* **19**, 127–136.
- Woolliams, J. A., Bijma, P. & Villanueva, B. (1999). The expected development of pedigree and its impact on genetic gain. *Genetics* **153**, 1009–1020.
- Wright, S. (1922). Coefficients of inbreeding and relationship. *American Naturalist* **15**, 330–338.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.