Theor Appl Genet (2012) 124:1271–1282
DOI 10.1007/s00122-012-1785-y

REVIEW

Use of the numerator relationship matrix in genetic analysis of autopolyploid species

Richard J. Kerr · Li Li · Bruce Tier ·
Gregory W. Dutkowski · Thomas A. McRae

Received: 26 July 2010 / Accepted: 5 January 2012 / Published online: 5 February 2012

Abstract Mixed models incorporating the inverse of a numerator relationship matrix (NRM) are widely used to estimate genetic parameters and to predict breeding values in animal breeding. A simple and quick method to directly calculate the inverse of the NRM has been historically developed for diploid animal species. Mixed models are less used in plant breeding partly because the existing method for diploids is not applicable to autopolyploid species. This is because of the phenomenon of double reduction and the possibility that gametes carry alleles which are identical by descent. This paper generalises the NRM and its inverse for autopolyploid species, so it can be easily incorporated into their genetic analysis. The technique proposed is to first calculate the kinship coefficient matrix and its inverse as a precursor to calculating the NRM and its inverse. This allows the NRM to be calculated for populations containing individuals of mixed ploidy levels. This generalization can also accommodate uncertain parentage by generating the "average" relationship matrix. The possibility that non-inbred parents can produce inbred progeny (double reduction) is also discussed. Rules are outlined that are applicable for any level of ploidy. Examples of use of the matrix are provided using simulated pedigrees.

Communicated by I. Mackay.

R. J. Kerr (✉) · G. W. Dutkowski
PlantPlan Genetics Pty Ltd, University of Tasmania,
Hobart, TAS 7001, Australia
e-mail: richard.kerr@plantplan.com

L. Li · B. Tier
Animal Genetics and Breeding Unit, University of New England,
Armidale, NSW 2351, Australia

T. A. McRae
PlantPlan Genetics Pty Ltd, PO Box 1811, Mount Gambier,
SA 5290, Australia

Introduction

Restricted maximum likelihood (REML) (Patterson and Thompson 1971) and best linear unbiased prediction (BLUP) (Henderson 1974) are important methodologies in livestock and forest tree breeding for estimating variance components and predicting breeding values, respectively. Dramatic increases in computer capacity and advances in computing techniques have allowed all known ancestral and collateral pedigree relationships to be incorporated into REML and BLUP models for the genetic analysis of data for quantitative traits. The use of such information and selection data increases the accuracy and precision of genetic analysis by removing some inherent biases in data collected in artificial breeding and selection programs. These biases arise because: selection is based on measurements recorded at different times and locations; selection changes trait means and variances (Kennedy and Sorenson 1988); and because mating is non-random (Wiggans and Misztal 1987). These biases can be removed because REML and BLUP models are able to better model covariances among individuals' genetic effects, which arise due to the sharing of genes. These covariances are described statistically using matrices. The matrix describing the covariance structure of additive genetic effects is commonly referred to as the additive relationship ($\boldsymbol{A}$), or numerator relationship matrix (NRM). Variations of this matrix is the focus of the present study.

Numerous algorithms for computing $\boldsymbol{A}$ and its inverse have been developed (Emik and Terril 1949; Tier 1990; Meuwissen and Luo 1992). Rules were formulated for use in diploid species, assuming gametes have no possibility of carrying two or more alleles which are identical by descent (IBD). Such rules are inappropriate for use in autopolyploid species and this is perhaps one reason why genetic analysis methods such as REML and BLUP are not widely used in autopolyploid species. Prevalence of non-additive genetic effects, vastly different breeding systems and unusual cytogenetic phenomena are often cited as reasons why tools which benefit livestock and forest tree breeding have limited application in autopolyploid crops such as potato, sugarcane and strawberries.

For tools such as REML and BLUP to have practical use in autopolyploid crops all that is required is a modification of the rules to construct the $\boldsymbol{A}$ inverse, which conform with the biology of these crops. Our focus has been the cultivated potato (*Solanum tuberosum*), which is a tetraploid ($2n = 4x = 48$). Double reduction is an observed phenomenon in this crop and results in sister chromatids segregating into the same gamete during meiosis. Though two parents are unrelated, their offspring may carry alleles that are IBD. Dihaploids ($2n = 2x = 24$) are often routinely induced through a parthenogenic process and used in breeding programs. Either the dihaploids are intercrossed, or crossed with non-cultivated, diploid species. Dihaploids and dihaploid-wild species hybrids are able to produce unreduced $2n$ gametes. The $2n$ gametes are used to efficiently introduce the products of the $2x$ breeding back into cultivated $4x$ potatoes. Other breeding strategies include crossing diploid species directly with the Tuberosum parent, without first creating a dihaploid, and intercrossing diploid parents, which are both producing $2n$ gametes.

The objective of this study was to generalize the rules to allow the use of the correct $\boldsymbol{A}$ matrix for autopolyploids in the genetic analysis of quantitative traits. Rather than define rules to build the $\boldsymbol{A}$ matrix directly, rules are presented which build a matrix of kinship coefficients, which can then be transformed to an $\boldsymbol{A}$ matrix appropriate for the ploidy level, or levels, of the species under study. This allows the correct $\boldsymbol{A}$ matrix to be derived and used in situations where not all materials have the same ploidy level because of dihaploidisation. The rules also allow for situations when parentage is uncertain. Derivation of inbreeding coefficients and genetic covariances, assuming various potato breeding strategies, has already been presented by Mendoza and Haynes (1973) and Haynes (1990, 1992). The present study differs in that the formulations are written as functions of the elements of the kinship matrix, for any ploidy level and for non-standard definitions of parentage. Computer simulations have been used to check the integrity of the formulae and to examine inbreeding coefficients for different levels of ploidy.

Methods

Defining additive genetic covariances

Nomenclature used in this paper generally follows that used by Gallais (2003). Consider an individual X which belongs to an autopolyploid population with a gametic ploidy level denoted by v. It has an additive genetic value at a single locus represented by

$$A_X = \sum_{i=1}^{2v_X} \alpha_i$$

where α_i corresponds to the mean value of allele i in combination with the gene pool of the population (see Kempthorne 1957, or Gallais 2003, page 164). Because our definition of the population extends to the inclusion of individuals of the same species or closely related species, but with differing ploidy levels, it is necessary to specify ploidy level at the individual level. Hence, the use of v_X rather than just v. Another individual, Y, with not necessarily the same ploidy level as X has an additive genetic value

$$A_Y = \sum_{i'=1}^{2v_Y} \alpha_{i'}$$

The additive genetic covariance between these individuals is the expectation of the product of A_X and A_Y, which can be writtenas

$$Cov(X, Y) = 4v_X v_Y E(\alpha_i \alpha_{i'})$$

Authors such as Kempthorne (1957) and Gallais (2003) explain that

$$E(\alpha\alpha') = k_{X,Y} E(\alpha_i^2)$$

because genes can only be in two states: either they are identical by descent with probability $k_{X,Y}$, which is the coefficient of kinship between X and Y, or they are non-identical, in which case the expectation is zero. In a population containing individuals of the same ploidy level, the expectation of a squared mean allele effect is the additive variance divided by $2v$. Future research will examine more closely what is the appropriate divisor when a population contains individuals of varying ploidy levels. A general form of the covariance can be written as

$$Cov(X, Y) = 4v_X v_Y k_{X,Y} \frac{\sigma_A^2}{2v_H} \qquad (1)$$

where v_H can equal v_X or v_Y or some intermediate value. With independent assortment Eq. 1 also defines a total covariance with regard to all segregating loci. When $v_X = v_Y = 1$ then Eq. 1 reduces to $Cov(X, Y) = 2k_{X,Y}\sigma_A^2$ and when $v_X = v_Y = 2$ reduces to $Cov(X, Y) = 4k_{X,Y}\sigma_A^2$.

The $\boldsymbol{A}$ matrix was initially defined for diploids as a matrix containing kinship coefficients among all individuals in the population, multiplied by two. For tetraploids, the coefficients are multiplied by four, hexaploids by six, and so forth. A uniform ploidy level is a reasonable assumption to make with respect to animal populations. It therefore makes sense to construct the $\boldsymbol{A}$ matrix directly. However, in plants this assumption cannot always be made. Therefore we propose using the $\boldsymbol{K}$, or kinship matrix, as the basis for deriving the additive genetic relationship matrix. Once the elements of $\boldsymbol{K}$ are obtained they are then individually manipulated with values for v_X, v_Y and v_H to derive the $\boldsymbol{A}$ matrix. The task is to develop recursive rules for building $\boldsymbol{K}$ and its inverse, rather than $\boldsymbol{A}$ and its inverse.

Deriving $\boldsymbol{K}$ for a population with uniform ploidy level

A diagonal element of $\boldsymbol{K}$ is denoted k_{ii}, where i assigns the position number of an individual in a chronologically ordered pedigree and describes the kinship coefficient of an individual with itself. A general formula for its derivation is (Gallais 2003)

$$k_{ii} = \frac{[1 + (2v - 1)F_i]}{2v} \tag{2}$$

where F_i is the inbreeding coefficient of individual i. A recursive system for building $\boldsymbol{K}$ requires k_{ii} to be a function of previously derived elements of the matrix. To proceed, we use the general formula for computing an inbreeding coefficient F_i under any ploidy level (Gallais 2003)

$$F_i = \frac{C_v^2 \Pr(b_j \equiv b_k) + C_v^2 \Pr(c_j \equiv c_k) + C_v^1 C_v^1 \Pr(b_j \equiv c_k)}{C_v^2 + C_v^2 + C_v^1 C_v^1} \tag{3}$$

$$= \frac{[v(v-1)\mathrm{Pr}(\mathrm{b_j} \equiv \mathrm{b_k}) + \mathrm{v}(\mathrm{v}-1)\mathrm{Pr}(\mathrm{c_j} \equiv \mathrm{c_k}) + 2\mathrm{v}^2\mathrm{Pr}(\mathrm{b_j} \equiv \mathrm{c_k})]}{2v(2v-1)} \tag{4}$$

where one parent p has $2v$ genes b_j (j varying from 1 to $2v$) and the other parent q has $2v$ genes c_j. Hence, $\Pr(b_j \equiv b_k)$ is the probability that 2 genes drawn from parent p are IBD either as a result of double reduction at meiosis with probability φ (for $v > 1$), or because they are drawn from different chromosomes with probability $1-\varphi$, in which case they are IBD with probability F_p. Similarly, $Pr(c_j \equiv c_k)$ represents the probability that two genes drawn from parent q are IBD, which equals $\varphi + (1-\varphi)F_q$. $Pr(b_j \equiv c_k)$ is the probability that a random gene from p is IBD to a random gene from q which by definition is k_{pq}. The binomial coefficient, C_v^2 (C_v^1) quantifies the number of combinations of 2 (1) genes from a gamete with v genes. Thus, after removing the common terms we have

$$k_{ii} = \frac{1 + \frac{1}{2}\left\{(v-1)\left[\varphi + (1-\varphi)F_p\right] + (v-1)\left[\varphi + (1-\varphi)F_q\right] + 2vk_{pq}\right\}}{2v} \tag{5}$$

which after further simplification and expressing parental inbreeding coefficients, F_p and F_q, in terms of their diagonal elements in $\boldsymbol{K}$ reduces to

$$k_{ii} = \frac{1 + (v-1)\varphi + \frac{(v-1)(1-\varphi)(vk_{pp} + vk_{qq} - 1)}{2v-1}}{2v} + \frac{k_{pq}}{2} \tag{6}$$

It can be shown that for diploids (v = 1), the diagonal element for any individual is purely a function of the kinship coefficient between its parents. Off-diagonal elements in $\boldsymbol{K}$ are computed recursively

$$k_{ij} = 0.5(k_{ip} + k_{iq}), \quad i<j \tag{7}$$

where p and q are parents of j. The $\boldsymbol{K}$ matrix is symmetric so $k_{ji} = k_{ij}$. A matrix representation of building $\boldsymbol{K}$ row by row is:

$$\boldsymbol{K}_i = \begin{bmatrix} \boldsymbol{K}_{i-1} & \boldsymbol{K}_{i-1}\boldsymbol{s}_i \\ \boldsymbol{s}_i'\boldsymbol{K}_{i-1} & k_{ii} \end{bmatrix} \tag{8}$$

where $\boldsymbol{s}_i$ is a vector containing two elements $\frac{1}{2}$ corresponding to the female and male parents (if known) and zeroes elsewhere. In applications such as REML for estimating variance components and BLUP for predicting genetic values, the inverse of $\boldsymbol{A}$ is required. A matrix representation of building $\boldsymbol{K}^{-1}$ row by row is:

$$\boldsymbol{K}_i^{-1} = \begin{bmatrix} \boldsymbol{K}_{i-1}^{-1} & \boldsymbol{0} \\ \boldsymbol{0}' & 0 \end{bmatrix} + (k_{ii} - \boldsymbol{s}_i'\boldsymbol{K}_{i-1}\boldsymbol{s}_i)^{-1} \begin{bmatrix} \boldsymbol{s}_i\boldsymbol{s}_i' & -\boldsymbol{s}_i \\ -\boldsymbol{s}_i' & 1 \end{bmatrix} \tag{9}$$

The left term in the right-hand side of the above equation is the $\boldsymbol{K}^{-1}$ matrix for all individuals in the pedigree list prior to individual i. It has been augmented with an extra row and column containing zeros in order for it to have the same size as the $\boldsymbol{K}^{-1}$ matrix for all individuals in the pedigree list up to and including individual i. A vector of zeros is represented as $\boldsymbol{0}$. The right term is a matrix of the same size multiplied by a scalar. The matrix $\boldsymbol{A}^{-1}$ is obtained by multiplying every element in $\boldsymbol{K}^{-1}$ by $2v_H/(4\ v_X\ v_Y)$. Expressing the vector $\boldsymbol{s}$ as the sum of vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ is useful in situations when parentage is uncertain. Often important cultivars are derived from field-picked berries. These cultivars are either selfed or open pollinated with pollen of neighbouring varieties. If the field trial contains small plots, it may be possible to limit the number of potential pollen parents to a relatively small number (<40). In this case, the vector $\boldsymbol{q}$ contains pollination probabilities in positions corresponding to the candidate male parents. Henderson (1988) developed this method for use in livestock genetic evaluation to cover cases where more than one sire is placed in the paddock for mating. A probability can be assigned in $\boldsymbol{q}$ to the female parent if there is some chance of selfing. The vector $\boldsymbol{p}$ will contain a single $\frac{1}{2}$ corresponding to the female parent. If ever there is a situation when female parentage is uncertain, probabilities can also be assigned in this vector as well. Equation 6 can be more generally written as

$$k_{ii} = \frac{1+(v-1)\varphi + \frac{4(v-1)(1-\varphi)(v\boldsymbol{p}_i'\boldsymbol{K}_{i-1}\boldsymbol{p}_i + v\boldsymbol{q}_i'\boldsymbol{K}_{i-1}\boldsymbol{q}_i - 0.25)}{2v-1}}{2v} + \boldsymbol{p}_i'\boldsymbol{K}_{i-1}\boldsymbol{q}_i + \boldsymbol{q}_i'\boldsymbol{K}_{i-1}\boldsymbol{p}_i$$

(In diploid species and under standard parentage when $\boldsymbol{p}$ and $\boldsymbol{q}$ contain single $\frac{1}{2}$ *s* corresponding to female and male parents, it is a straightforward exercise in demonstrating this formula's equivalence to $1/2(1 + k_{pq})$). The scalar term in Eq. 9, $(k_{ii} - \boldsymbol{s}_i'\boldsymbol{K}_{i-1}\boldsymbol{s}_i)^{-1}$, which will henceforth be denoted d^i, can also be more generally written as

$$d^i = \left(\frac{1+(v-1)\varphi + \frac{4(v-1)(1-\varphi)(v\boldsymbol{p}_i'\boldsymbol{K}_{i-1}\boldsymbol{p}_i + v\boldsymbol{q}_i'\boldsymbol{K}_{i-1}\boldsymbol{q}_i - 0.25)}{2v-1}}{2v} - \boldsymbol{p}_i'\boldsymbol{K}_{i-1}\boldsymbol{p}_i - \boldsymbol{q}_i'\boldsymbol{K}_{i-1}\boldsymbol{q}_i \right)^{-1} \tag{10}$$

Under standard parentage in diploid organisms Eq. 10 reduces to $d^i = 0.5 - 0.25(k_{pp} + k_{qq}))^{-1}$. Adding the *i*th row to $\boldsymbol{K}^{-1}$ is completed using the instructions contained in Table 1. These same instructions, but based on d^i computed using a_{pp} and a_{qq}, rather than k_{pp} and k_{qq}, were orginally reported by Henderson (1976). Instructions for adding the *i*th row when parentage is uncertain is shown in Table 2.

To apply this methodology in a potato improvement program, researchers and breeders will need to build a pedigree database for their populations, from which the complete pedigree, or sections of it, can be extracted. The important point is that for whatever section of the population that is to be analysed with tools such as REML or BLUP, all antecedents to the materials under study are required to be extracted. Algorithms in common use for computing the diagonals and other elements of $\boldsymbol{A}$ necessary for building the inverse can be easily modified to build the inverse of $\boldsymbol{K}$ for any level of ploidy (Tier 1990; Meuwissen and Luo 1992).

Table 1 Contributions to the $\boldsymbol{K}^{-1}$ matrix when adding row and column *i*

Contribution	Position
d^i	(i, i)
$-0.5\ d^i$	(i, f), (i, m), (f, i), (m, i)
$0.25\ d^i$	(f, f), (f, m), (f, f), (m, m)

Parents of individual *i* have positions in a chronologically ordered pedigree denoted by *f* and *m*

Table 2 Contributions to the $\boldsymbol{K}^{-1}$ matrix when adding row and column *i*

Contribution	Position
d^i	(i, i)
$-0.5\ p_j\ d^i$	(i, f_j), (f_j, i) $\quad j = 1, \ldots, n_f$
-0.5 $q_k\ d^i$	(i, m_k), (m_k, i) $\quad m = 1, \ldots, n_m$
$0.25\ p_j p_{j'} d^i$	$(f_j, f_{j'})$ $\quad j = 1, \ldots, n_f;\ j' = 1, \ldots, n_f$
$0.25\ q_k q_{k'} d^i$	$(m_k, m_{k'})$ $\quad k = 1, \ldots, n_m;\ k' = 1, \ldots, n_m$
$0.25\ p_j\ q_k\ d^i$	(f_j, m_k) $\quad j = 1, \ldots, n_f;\ k = 1, \ldots, n_m$

There are n_f and n_m candidate female and male parents, respectively, of individual *i* and they have positions in a chronologically ordered pedigree denoted by f_j and m_k. The probabilities associated with f_j and m_k are denoted by p_j and q_k

Equations 8 and 9 remain sufficiently general for constructing $\boldsymbol{K}$ and $\boldsymbol{K}^{-1}$ when the breeding strategy involves intercrossing materials of different ploidy levels, provided the correct diagonal elements can be obtained. This is the topic of the next section.

Computing diagonal elements of $\boldsymbol{K}$ for derived autopolyploids

Though our focus will mainly be restricted to the tetraploid potato, presented formulae will still assume a general ploidy level. As Haynes (1992) has noted, there are three main potato breeding strategies involving derived tetraploids and which involve using unreduced 2*n* gametes:

1. Producing a haploid from a tetraploid parent and crossing it with a diploid parent to create a haploid-species hybrid, which is then bred. Improved hybrids, which are able to produce unreduced gametes, are then crossed back to a tetraploid parent.
2. Crossing a tetraploid parent directly with a diploid parent, where the latter is producing 2*n* gametes
3. Crossing diploid parents which are both producing 2*n* gametes

Firstly, consider the implications of producing the haploid parent. In potatoes, haploids are routinely induced by crossing *S. tuberosum* 4*x* seed parents with 2*x* selections from known cultivar groups, which are known for their propensity to trigger parthenogenesis in the female, e.g. Phureja (Ortiz and Peloquin 1994). Haploids induced this way are often referred to as dihaploids in order to distinquish them from haploids of diploid species. Dihaploid progeny are therefore random gametic samples of the seed parent and their inbreeding is a function solely of the parent's inbreeding. If p denotes the seed parent and v_p its ploidy level, Eq. 4 reduces to

$$F_i = \frac{C^2_{v_p} Pr(b_j \equiv b_k)}{C^2_{v_p}} = F_p$$

and Eq. 5 reduces to

$$k_{ii} = \frac{1 + (2v_i - 1)F_p}{2v_i}$$
$$= \frac{1 + (2v_i - 1)\left[\varphi + \frac{(1-\varphi)(2v_p k_{pp} - 1)}{2v_p - 1}\right]}{2v_i}$$

Rules for building either the $\boldsymbol{K}$ matrix or its inverse, as represented by Eqs. 8 or 9 should still be followed. However, whenever adding a row corresponding to an induced dihaploid the vector $\boldsymbol{s}$ will contain a single element with a value 1 corresponding to the seed parent.

Next consider the restoration of the improved haploid-species hybrid germplasm to the 4x level. Returning to Eq. (4) assume that parent p is the tetraploid parent and q is the haploid-species hybrid parent. The derivation of the probability Pr($b_j \equiv b_k$) remains as before, but the derivation of Pr($c_j \equiv c_k$) now becomes dependent on the mechanism for 2n gamete production and the parameter, β, which is the average cross-over frequency between the centromere and loci influencing the trait under study. The explanation of the differences between first (FDR) and second division restitution (SDR) and how β affects the configuation of genotypes in the 2n gametes in each case is discussed by other authors (Tai 1992; Carputo et al. 2003). In summary under FDR Pr($c_j \equiv c_k$) $= (1 - \beta/2)F_q + \beta/2$, while under SDR Pr($c_j \equiv c_k$) $= \beta F_q + 1 - \beta$. In both cases, we assume there is the potential for the haploid-species hybrid parent to be inbred. It is unlikely the tetraploid parent from which the dihaploid is extracted is related to a wild species diploid parent. However, there is potential for inbreeding to appear in the haploid-species hybrid population if more than one generation of breeding ensues after hybridisation, and before restoration to the 4x level. Researchers need to keep in mind that if derived tetraploids are to be analysed together with standard tetraploids, then the pedigrees of the materials that gave rise to the derived tetraploids have to be included in the analysis. The algorithm that computes the diagonals of the $\boldsymbol{K}$ matrix has to be sufficiently flexible to take into account the changing ploidy levels and cytological mechanisms among antecedents and current materials. The formula for computing the diagonal element for the derived autopolyploid, i, assuming FDR occurs in parent q, is

$$k_{ii} = \frac{1 + \frac{(v_p-1)}{2}\left[\varphi + \frac{(1-\varphi)(2v_p k_{pp}-1)}{2v_p-1} + \frac{\beta}{2} + \frac{(1-\frac{\beta}{2})(2v_q k_{qq}-1)}{2v_q-1}\right]}{2v_p} + \frac{k_{pq}}{2}$$

where v_q is the ploidy level of a haploid-species hybrid parent q and v_p is the ploidy level of parent p and the derived autopolyploid. This results because when sampling genes from the gamete of q there are v_p, not v_q genes to sample from and common terms can still be factored out. The offdiagonal element k_{pq} will be non null if both parents descend from the individual from which the dihaploid was extracted. Assuming $v_p = 2$ and $v_q = 1$ because β used here is specific to the tetraploid potato, we have

$$k_{ii} = \frac{1}{4} + \frac{\varphi}{8} + \frac{\beta}{16} + \frac{(1-\varphi)(4k_{pp}-1)}{24} + \frac{(1-\frac{\beta}{2})(2k_{qq}-1)}{8} + \frac{k_{pq}}{2} \tag{11}$$

If it is assumed SDR occurs in parent q a different formula is used

$$k_{ii} = \frac{1}{4} + \frac{\varphi}{8} + \frac{1-\beta}{8} + \frac{(1-\varphi)(4k_{pp}-1)}{24} + \frac{\beta(2k_{qq}-1)}{8} + \frac{k_{pq}}{2} \tag{12}$$

Equations 11 and 12 are also applicable if a tetraploid parent is crossed directly with a diploid producing unreduced gametes (strategy 2 above). Different formulae are needed under strategy 3 where two diploid parents are crossed to produce a derived tetraploid. Generally, FDR and SDR are mechanisms that occur in the pollen parent. The most common mechanism that leads to 2n egg formation is ommission of the second meiotic division, which though not strictly SDR is equivalent to it (Carputo et al. 2003). The following formula is relevant to SDR in egg formation and FDR in pollen formation.

$$k_{ii} = \frac{1}{4} + \frac{2-\beta}{16} + \frac{\beta(2k_{pp}-1) + (1-\frac{\beta}{2})(2k_{qq}-1)}{8} + \frac{k_{pq}}{2} \tag{13}$$

Most diploid potato breeding programs take a parental line approach. That is, diploid species are screened for some particular trait. A parent or parents, once identified, are crossed with either dihaploid, tetraploid or diploid parents. It is unlikely representations of Eqs. 11, 12 and 13 for scenarios of uncertain parentage would be necessary. However, there are examples of breeding programs that take a population improvement approach with diploids. In such cases, open-pollinated seed may be collected from the field and occasionally 4x seedlings are recovered; presumably the result of 2x − 2x pollinations where both parents produced 2n gametes. If a set of candidate male parents can be identified, Eq. 13 can be more generally written to accommodate uncertain paternity

$$\begin{aligned} k_{ii} = {} & \frac{1}{4} + \frac{2-\beta}{16} \\ & + \frac{4\beta(2\boldsymbol{p}_i'\boldsymbol{K}_{i-1}\boldsymbol{p}_i - 0.25) + 4(1-\frac{\beta}{2})(2\boldsymbol{q}_i'\boldsymbol{K}_{i-1}\boldsymbol{q}_i - 0.25)}{8} \\ & + \boldsymbol{p}_i'\boldsymbol{K}_{i-1}\boldsymbol{q}_i + \boldsymbol{q}_i'\boldsymbol{K}_{i-1}\boldsymbol{p}_i. \end{aligned} \tag{14}$$

Pedigree examples

Four pedigree examples were designed to check the validity of our formulae and to investigate the effects of different pedigrees on kinship coefficients. In the first three pedigrees (see Table 3), all individuals are assumed to have the sample ploidy level and individuals 1 to 3 constitute the base generation. Parentage is known in Pedigree 1. Individuals 7 and 8 have uncertain parentage in Pedigrees 2 and 3. Pedigree 3 differs in that there is a probability individuals 7 and 8 are selfed. The candidate female parents of offspring 7 are 3 and 5 with probabilities of 0.3 and 0.7, respectively, and the candidate male parents of offspring 7 are the same candidates (3 and 5) but with different probabilities of 0.4 and 0.6, respectively. A similar situation was created for offspring 8. Pedigree 4 (see Fig. 1) was designed to check the validity of our rules for breeding strategies involving derived autotetraploids and unreduced gametes.

Matrices $\boldsymbol{K}$ and $\boldsymbol{K}^{-1}$ were computed for Pedigree 1 assuming three ploidy levels: diploid ($v = 1$); tetraploid ($v = 2$); and octoploid ($v = 4$). Matrices $\boldsymbol{K}$ and $\boldsymbol{K}^{-1}$ were computed for Pedigrees 2 and 3 assuming a tetraploid species only. In the case of Pedigree 4, it was assumed individuals 1 and 8 are tetraploid; individuals 3 and 4 are dihaploid, and individuals 2, 5, 6 and 7 are diploid.

Simulation

A single locus gene drop simulation (MacCluer et al. 1986) was written in order to examine the integrity of our formulae for computing kinship coefficients. In a gene drop simulation unique alleles at a single locus are assigned to each founder and a genotype is created for each descendent by Mendelian segregation of parental alleles. Gene drop simulations were completed only for pedigree 1, assuming a tetraploid species, and for pedigree 4. Individuals 1, 2 and 3 are the founders in the case of Pedigree 1 and individuals 1 and 2 are the founders in the case of Pedigree 4. Pseudocode is presented in Appendix demonstrating how the simulation works in the case of pedigree 4. Pedigree 4 is more complicated in terms of unusual cytological events: individual 1 gives rise to dihaploid individuals 3 and 4 via parthenogenesis; and individual 7 produces unreduced gametes via first division restitution. The pseudocode reveals how double reduction occurs with probability φ in the simulation of gametes produced by individual 1. When creating individual 8, first division restitution occurs with probability β when sampling alleles from individual 7. The main loop in the pseudo-code repeats the gene drop 50,000 times (FOR DROP:=1 to 50,000 in Appendix). For each drop, and for every pairing of individuals, i, j, the number of instances where a gene sampled from individual i is IBD to a gene sampled from individual j, is counted and divided by 4 v_i v_j to derive the coefficient k_{ij} for this drop. It is irrelevant whether the main loop is seen as repeating the drop at the same locus 50,000 times or completing drops at 50,000 different loci. By repeating the drop 50,000 times, we are manually computing a $\boldsymbol{K}$ matrix for use in an infinitesimal model (Bulmer 1980). That is, the gene drop simulation should in theory produce the same $\boldsymbol{K}$ matrix as that derived from rules which implicilty assume an averaging over a very large number of loci. In this sense, the $\boldsymbol{K}$ matrix computed using one method (simulation or rules) is not the true or correct matrix, while the matrix computed using the other method is an estimate. If our rules are correct, both methods should produce the same $\boldsymbol{K}$ matrix.

Results

Table 4 shows the diagonal elements of the $\boldsymbol{A}$ and $\boldsymbol{A}^{-1}$ matrices, computed via the formulae for diploids, tetraploids and octoploids. The diagonal elements of $\boldsymbol{A}$ and

Table 3 Three pedigree examples, where all individuals are assumed to have the same ploidy level

Individual	Pedigree 1		Pedigree 2		Pedigree 3	
	Female parent	Male parent	Female parent	Male parent	Female parent	Male parent
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	1	2	1	2	1	2
5	1	4	1	4	1	4
6	3	5	3	5	3	5
7	5	6	1 (0.3)	3 (0.4)	3 (0.3)	3 (0.4)
			2 (0.7)	4 (0.6)	5 (0.7)	5 (0.6)
8	6	7			1 (0.1)	4 (0.2)
			3 (0.3)	6 (0.2)	6 (0.3)	6 (0.3)
			5 (0.7)	7 (0.8)	7 (0.6)	7 (0.5)

There are two or more candidate female or male parents (Pedigrees 2 and 3), the probability of parentage is shown in parentheses

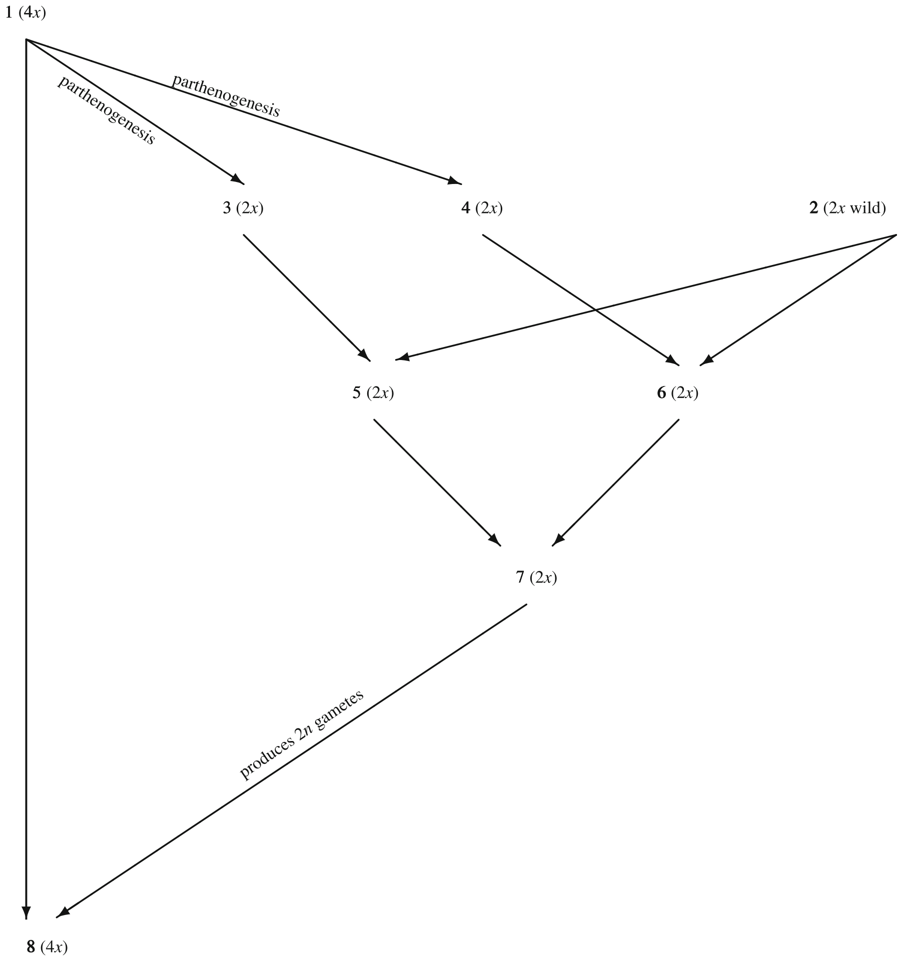

Fig. 1 Pedigree 4. Individual *1* is a cultivated tetraploid and produces two dihaploid offsprings via parthenogenesis (*3* and *4*), which in turn are mated with a 2x individual (*2*) to produce two offspring (*5* and *6*). Individual *7* produces unreduced gametes so that it can be crossed back to individual *1*

computed via the simulation for tetraploids are also shown. (The $\boldsymbol{K}$ and $\boldsymbol{K}^{-1}$ matrices have been multiplied by 2v and 1/2v, respectively, to obtain $\boldsymbol{A}$ and $\boldsymbol{A}^{-1}$). The value for φ was set at either 0 or 0.1 (Table 5). Diagonal elements of $\boldsymbol{A}$, denoted as a_{ii} are equivalent when computed using the formulae and when using simulation. The $\boldsymbol{K}$ matrices computed from formulae and from simulation, for Pedigree 4, are equivalent to the 3rd decimal place (see Table 6). These results demonstrate our formulae for computing an individual's kinship coefficient with itself and the tabular method for constructing $\boldsymbol{K}$ and $\boldsymbol{K}^{-1}$ (hence $\boldsymbol{A}$ and $\boldsymbol{A}^{-1}$) are correct.

Table 4 also shows that when assuming the same pedigree structures, species with higher ploidy levels will exhibit greater inbreeding. For example, without double reduction, the diagonal element a_{88}, which is equal to $1 + (2v - 1)F_8$, where F_8 is the inbreeding coefficient for individual 8, increased from 1.41 for diploids to 1.48 for tetraploids and 1.51 for octoploids. When the double reduction rate was assumed to be 0.1 for tetraploids and octoploids, the diagonal element became even higher: 1.68 for tetraploids; and 1.73 for octoploids.

In diploids, when parents of an individual are unrelated, the progeny are not inbred (for example, a_{66} = 1.0 for diploids, Table 4). However, assuming no double reduction ($\varphi = 0$), a_{66} became 1.04 and 1.05 for the tetraploid and octoploid situations, respectively. This occurs because one of the parents (individual 5) is partly inbred, and its gametes can carry alleles IBD. With double reduction ($\varphi = 0.1$), there is even greater chance gametes from parent 5 are identical by descent. Parent 3, though not inbred itself, now has some probabililty of producing gametes which carry alleles IBD. This translates to higher values for a_{66}: 1.18 and 1.20.

Discussion

Generally a kinship coefficient is defined as the probability of drawing a set of genes in one zygote, and another set in another zygote, with identity of descent relationships between the two sets. The simplest type of kinship coefficient considers drawing only a single gene from each zygote. Wright (1922) in defining what he called the correlation of genic values, with regard to a quantitative trait, between two diploid individuals,

Table 4 Diagonal elements of $\boldsymbol{A}$ and $\boldsymbol{A}^{-1}$ for Pedigree 1 for different ploidy levels (v = 1, 2 and 4) and with ($\varphi = 0.1$) and without ($\varphi = 0$) double reduction, derived using rules (a_{ii} and a^{ii} are the diagonal elements of $\boldsymbol{A}$ and $\boldsymbol{A}^{-1}$ for individual i, respectively)

Individual	Female parent	Male parent	Simulation		Diploid (v = 1)		Tetraploid (v = 2)				Octoploid (v = 4)			
			$\varphi = 0$	$\varphi = 0.1$	$\varphi = 0$		$\varphi = 0$		$\varphi = 0.1$		$\varphi = 0$		$\varphi = 0.1$	
			a_{ii}	a_{ii}	a_{ii}	a^{ii}	a_{ii}	a^{ii}	a_{ii}	a^{ii}	a_{ii}	a^{ii}	a_{ii}	a^{ii}
1	0	0	1.00	1.11	1.00	2.00	1.00	2.00	1.10	1.77	1.00	2.00	1.10	1.76
2	0	0	1.00	1.11	1.00	1.50	1.00	1.50	1.10	1.34	1.00	1.50	1.10	1.33
3	0	0	1.00	1.11	1.00	1.57	1.00	1.52	1.10	1.36	1.00	1.51	1.10	1.35
4	1	2	1.00	1.13	1.00	2.50	1.00	2.50	1.13	2.16	1.00	2.50	1.14	2.13
5	1	4	1.25	1.41	1.25	3.14	1.25	3.05	1.41	2.65	1.25	3.02	1.42	2.59
6	3	5	1.04	1.18	1.00	3.45	1.04	3.15	1.18	2.76	1.05	3.06	1.20	2.65
7	5	6	1.36	1.54	1.31	2.88	1.36	2.64	1.54	2.32	1.38	2.56	1.58	2.22
8	6	7	1.48	1.68	1.41	2.37	1.48	2.14	1.68	1.89	1.51	2.06	1.73	1.80

Diagonal elements of A when v = 2 are also shown when derived from simulation

Table 5 Diagonal elements of $\boldsymbol{A}$ and $\boldsymbol{A}^{-1}$ for Pedigrees 2 and 3 for ploidy level v = 2 and with ($\varphi = 0.1$) and without ($\varphi = 0$) double reduction, derived using rules

Individual	Pedigree 2				Pedigree 3			
	$\varphi = 0$		$\varphi = 0.1$		$\varphi = 0$		$\varphi = 0.1$	
	a_{ii}	a^{ii}	a_{ii}	a^{ii}	a_{ii}	a^{ii}	a_{ii}	a^{ii}
1	1.00	2.04	1.10	2.04	1.00	2.01	1.10	1.78
2	1.00	1.71	1.10	1.71	1.00	1.50	1.10	1.34
3	1.00	1.63	1.10	1.63	1.00	1.74	1.10	1.55
4	1.00	2.66	1.13	2.66	1.00	2.52	1.13	2.17
5	1.25	2.75	1.41	2.75	1.25	3.28	1.41	2.84
6	1.04	2.05	1.18	2.11	1.04	2.26	1.18	1.97
7	1.00	2.03	1.15	2.03	1.21	2.36	1.38	2.02
8	1.12	1.84	1.28	1.84	1.32	1.87	1.51	1.61

Table 6 The $\boldsymbol{K}$ matrix for pedigree 4 under the assumption that φ = 1/6 and $\beta = 0.0822e^{15.4635\varphi} - 1$

	1	2	3	4	5	6	7	8
1	0.250	0.000	0.250	0.250	0.125	0.125	0.125	0.188
2	0.000	0.500	0.000	0.000	0.250	0.250	0.250	0.125
3	0.250	0.000	0.583	0.250	0.292	0.125	0.208	0.229
4	0.250	0.000	0.250	0.583	0.125	0.292	0.208	0.229
5	0.125	0.250	0.292	0.125	0.500	0.188	0.344	0.234
6	0.125	0.250	0.125	0.292	0.188	0.500	0.344	0.234
7	0.125	0.250	0.208	0.208	0.344	0.344	0.594	0.359
8	0.187	0.125	0.229	0.229	0.234	0.234	0.360	0.408

The matrix compared when derived from rules or derived using simulation are equivalent to the third decimal place

used an expression in which the numerator was twice this type of kinship coefficient. According to Hill (1995), it was Jay Lush who first introduced the concept of a matrix in which cells contain the numerator term in Wright's expression and could be applied in describing the additive genetic variance-covariance structure of a population. It is animal breeding custom to denote this matrix as the numerator relationship matrix or $\boldsymbol{A}$ matrix. The matrix is important in animal breeding because it is central to the estimation of variances using REML and the prediction of breeding values using BLUP.

In plant breeding, it has been less of a custom to use this matrix. This may be partly due to the fact that plants are often polyploid and exhibit phenomena such as double reduction, making current rules for computing the diagonal elements of $\boldsymbol{A}$ no longer applicable. Another reason may be that variety selection is often emphasised in plant breeding, while population improvement appears to receive less attention. Hence, breeding values are deemed less important and there is less urgency in adopting techniques such as BLUP. Presently clonal replication is not an option in animal breeding and incorporating information from relatives is an important issue, hence the motivation for using BLUP. Finally, for many species plants can survive when inbred, making the prediction of genetic values and predicting the outcome of crosses much easier.

Nevertheless, the authors believe BLUP and REML do have applicability in plant breeding and have sought to generalise the rules for computing the diagonal elements of $\boldsymbol{A}$. Given the correct diagonal elements, the tabular method for constructing the inverse remains unchanged. Breeders of outcrossing plants generally take a broad approach when looking for new varieties. Each cycle of variety selection begins with a large seedling generation grown from true seeds. A pool of elite parents is maintained for generating this seed. It would be useful in our opinion to integrate all data collected across the subsequent clonal generations, and collected across all cycles of variety selection, into a single, meta-type analysis. BLUP would be used to predict breeding values for all genotypes. By comparing genotypes used across generations and across testing sites, the breeder is aided in identifying the best set of elite parents, not just within their own program, but potentially on a national or regional basis.

A correct $\boldsymbol{A}$ matrix for auto-polyploid species would also have applicability in mixed models of the type pioneered by Yu et al. (2006) for use in association mapping. Malosetti et al. (2007) have used such a model to study quantitative trait loci (QTL) affecting variation in late blight resistance in potato cultivars. They partitioned the total genetic variation into a fixed part associated with a QTL and a random part due to polygenes. They demonstrated a clear improvement in statistical accuracy when using a pedigree derived $\boldsymbol{A}$ matrix to model the variance-covariance structure of the polygenic component, relative to a structure that only had diagonal elements. Because they were using only available statistical packages and not special purpose software, we assume the $\boldsymbol{A}$ matrix was constructed assuming a diploid species. Hence, additional gain in accuracy could be achieved using rules applicable for an auto-tetraploid species.

A feature of potato breeding is the intense selection that occurs in the first clonal generation. All seedling progenies (the number often exceeds 100,000) are replicated once into single hill plots. It is not uncommon for only 1–3% of the progeny to remain after single hill selection. For these progenies to be accurately assessed for future use as parents, it is important that their breeding values have a distribution that resembles the actual population distribution. Consider an analysis that includes only individuals from a small, intensely selected group. A breeding value gained from such an analysis may falsely indicate eliteness, because it is not expressed on a true population scale. To obtain a distribution that better reflects reality, the analysis must include all founders as well as progeny that are not selected. An $\boldsymbol{A}$ matrix used in the analysis helps to properly position an individual's breeding value in a population context.

First stage selection is often based on agronomic traits, because they are easy to assess. Traits, which are the actual focus of the breeding effort, such as processing characteristics and eating quality, only get recorded in advanced selection stages because they are harder to measure. Use of the $\boldsymbol{A}$ matrix linking descendents to their antecedents in a REML multivariate analysis will help to recover the population variance for traits recorded only on small, intensely selected groups; and to establish the correlation between traits recorded in first stages and traits recorded in advanced stages of selection. These correlations and recovered variances for non-agronomic traits should be be used in subsequent, multi-variate BLUP analyses for the prediction of breeding values.

If the meta-type analysis covers all current and historical materials used in breeding and selection, a complication will arise if some genotypes used are outside the normal range of cultivar groups. Examples are: the direct use of a wild species which has a different ploidy level; or the use of dihaploids derived from the cultivated species. If this is the case then the factor which is multiplied by the kinship coefficient to obtain Wright's numerator expression will have different values. The recursive rules for computing diagonal elements of $\boldsymbol{A}$ assume the value for this factor to be constant. A constant value can only be used if a uniform ploidy level is assumed; for example, the factor is 2 if all individuals are diploid and 4 if all individuals are tetraploid. The value this factor takes stems from our assumptions on what is the nature of the additive genetic variance and is dependent on how we define the gene pool and average effects of alleles.

Because of the potential for materials of different ploidy levels to be used, we have decided to focus on deriving rules for constructing the kinship coefficient matrix, denoted $\boldsymbol{K}$, and its inverse, $\boldsymbol{K}^{-1}$. These rules are independent from any assumptions on what is the nature of an additive genetic covariance. Once constructed $\boldsymbol{K}$, or its inverse, $\boldsymbol{K}^{-1}$ can be translated to $\boldsymbol{A}$ or $\boldsymbol{A}^{-1}$ by multiplying each cell individually by its appropriate factor. The value of this factor for a cell which defines the genetic covariance between two individuals of different ploidy levels is the subject of ongoing research.

However, for the present the rules for deriving $\boldsymbol{K}$ and $\boldsymbol{K}^{-1}$ will be beneficial to researchers working in crops which are assumed to have a constant ploidy level. Kerr et al. (2009) recently performed an integrated analysis on potato data. The analysis used BLUP incorporating the correct $\boldsymbol{A}$ for an auto-tetraploid species. The analysis was somewhat limited in that only recent ancestors of test materials were included in the analysis and any genotype with a differing ploidy level was excluded. There will also be benefit for those researchers keen to ascertain what are the kinship coefficients among materials of different ploidy levels.

Acknowledgments The authors would like to thank AUSIndustry for providing funding to conduct this research and to two anonymous reviewers for their helpful comments.

Appendix A

```
# Set up a vector denoting the number of alleles at the locus, for each of the 8 individuals
N = [(4,2,2,2,2,2,2,4)]                 # e.g. $v$ for individual 1 is 2 hence has $2v = 4$ alleles
# Set up parental arrays p1 and p2
p1 = [(0,0,1,1,3,4,5,1)]                # e.g. the first parent for individual 7 is individual 5
p2 = [(0,0,1,1,2,2,6,7)]                # e.g. the second parent for individual 7 is individual 6
# assign unique alleles to founders
# Gtype is a is 2-dimensional array storing genotypes, i.e. Gtype[individual number][allele number]
Gtype[1] = [(1,2,3,4)]                  # e.g the 4th allele of individual 1 is “4”
Gtype[2] = [(5,6)]                      # e.g.the 2nd allele of individual 2 is “6”
FOR DROP:=1 TO 50000 DO
    # sample genotypes for individuals 3 and 4 which are derived as a result of parthenogenesis
    FOR I:= 3 TO 4 DO
        # random(low,high) is a function that generates an integer uniformly distributed between low and high (inclusive)
        a1 = random(1,4)
        IF rand() < $\varphi$ THEN          # rand()is a function that generates a real from a uniform (0-1) distribution
            a2 = a1
        ELSE
            a2 = random(1,4)
            WHILE (a2 == a1)
                a2 = random(1,4)
            ENDWHILE
        ENDIF
        Gtype[I] = [(a1,a2)]
    ENDFOR
    # sample genotypes for diploid individuals 5,6,7
    FOR I = 5 TO 7 DO
        a1 = random(1,2); a2 = random(1,2)
        Gtype[I] = [ (Gtype[p1[I]][a1], Gtype[p2[I]][a2]) ]
    ENDFOR
    # sample genotype for tetraploid individual 8
    # first sample alleles from parent 1
    a1 = random(1,4);
    if(rand() < $\varphi$) THEN
        a2 = a1
    ELSE
```

```
            a2 = random(1,4);
            WHILE a2 == a1
                a2 = random(1,4);
            ENDWHILE
        END
        # now sample alleles from parent 7
        a3 = random(1,2);
        if(rand < $\beta/2$) THEN
            a4 = a3
        ELSE
            a4 = random(1,2);
            WHILE a4 == a3
                a4 = random(1,4);
            ENDWHILE
        ENDIF
        Gtype[8] = [ (Gtype[1][a1], Gtype[1][a2], Gtype[7][a3], Gtype[7][a4]) ]
        # determine $k_{i,j}$ for $i = 1,\ldots,8$ and $j = i,\ldots,8$, for this drop (K matrix)
        FOR I = 1 TO 8 DO
            FOR J = I TO 8 DO
                IBD = 0;
                FOR K = 1 TO N[I] DO
                    a1 = Gtype[I][K]
                    FOR L = 1 TO N[J] DO
                        a2 = Gtype[J][L]
                        IBD++ if a1 == a2
                    ENDFOR
                ENDFOR
                IBD /= (N[I]*N[J])        # the probability of IBD, when drawing alleles from I and J
                K[I][J] += IBD;      # sum up over drops
            ENDFOR
        ENDFOR
    ENDFOR
    K /= 50000          # compute the average over 50000 drops
```

References

Bulmer M (1980) The mathematical theory of quantitative genetics. Oxford University Press, Oxford

Carputo D, Frusciante L, Peloquin SJ (2003) The role of 2n gametes and endosperm balance number in the origin and evolution of polyploids in the tuber-bearing solanums. Genetics 163(1): 287–94

Emik L, Terril C (1949) Systematic procedures for calculating inbreeding coefficients. J Hered 40:51–55

Gallais A (2003) Quantitative genetics and breeding methods in autopolyploid plants. INRA, Paris

Haynes KG (1990) Covariances between diploid parent and tetraploid offspring in tetraploid × diploid crosses of *solanum-tuberosum* l. J Hered 81(3):208–210

Haynes KG (1992) Some aspects of inbreeding in derived tetraploids of potatoes. J Hered 83(1):67–70

Henderson CR (1974) General flexibility of linear model techniques for sire evaluation. J Dairy Sci 57:963–972

Henderson CR (1976) Simple method for computing inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32(1):69–83

Henderson CR (1988) Use of an average numerator relationship matrix for multiple-sire joining. J Anim Sci 66(7):1614–1621

Hill WG (1995) Sewall wright's "systems of mating". Genetics 143:1499–1506

Kempthorne O (1957) An introduction to genetic statistics. Wiley, NY

Kennedy B, Sorenson D (1988) Properties of mixed-model methods for prediction of genetic merit. In: Second international conference on quantitative genetics, Sinauer Associates, Sunderland, pp 91–103

Kerr R, Dutkowski G, Li L, McRae T, Novy R, Schneider B, Tier B (2009) Integrated genetic analysis for potato improvement. In: 14th Australasian plant breeding and 11th SABRO, SABRO, Cairns

MacCluer JW, Vandeberg JL, Read B, Ryder OA (1986) Pedigree analysis by computer-simulation. Zoo Biol 5(2):147–160

Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to phytophthora infestans in potato. Genetics 175(2):879–89 (Epub 2006 Dec 6)

Mendoza H, Haynes F (1973) Some aspects of breeding and inbreeding in potatoes. Am Potato J 50:216–222

Meuwissen THE, Luo Z (1992) Computing inbreeding coefficients in large populations. Genet Sel Evol 24(4):305–313

Ortiz R, Peloquin SJ (1994) Use of 24-chromosome potatoes (diploids and dihaploids) for genetical analysis. In: Bradshaw J, Mackay G (eds.) Potato genetics, CAB International, Wallingford, pp 133–154

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554

Tai G (1992) Use of 2n gametes. In: Bradshaw J, Mackay G (eds.) Potato genetics, CAB International, Wallingford, pp 109–132

Tier B (1990) Computing inbreeding coefficients quickly. Genet Sel Evol 22:419–430

Wiggans G, Misztal I (1987) Supercomputer for animal model evaluation of ayrshire milk yield. J Dairy Sci 70:1906–1912

Wright S (1922) Coefficients of inbreeding and relationship. Am Nat 56:330

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat genet 38(2):203–238