

Research

Open Access

Inclusion of genetically identical animals to a numerator relationship matrix and modification of its inverse

Takuro Oikawa* and Kazuhiro Yasuda

Address: Graduate School of Natural Science and Technology, Okayama University, Okayama-shi 700-8530, Japan

Email: Takuro Oikawa* - toikawa@cc.okayama-u.ac.jp; Kazuhiro Yasuda - ky_0911@angel.ocn.ne.jp

* Corresponding author

Published: 3 March 2009

Received: 28 January 2009

Genetics Selection Evolution 2009, **41**:25 doi:10.1186/1297-9686-41-25

Accepted: 3 March 2009

This article is available from: <http://www.gsejournal.org/content/41/1/25>

© 2009 Oikawa and Yasuda; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In the field of animal breeding, estimation of genetic parameters and prediction of breeding values are routinely conducted by analyzing quantitative traits. Using an animal model and including the direct inverse of a numerator relationship matrix (NRM) into a mixed model has made these analyses possible. However, a method including a genetically identical animal (GIA) in NRM if genetic relationships between pairs of GIAs are not perfect, is still lacking. Here, we describe a method to incorporate GIAs into NRM using a **K** matrix in which diagonal elements are set to 1.0, off-diagonal elements between pairs of GIAs to (1-x) and the other elements to 0, where x is a constant less than 0.05. The inverse of the **K** matrix is then calculated directly by a simple formula. Thus, the inverse of the NRM is calculated by the products of the lower triangular matrix that identifies the parents of each individual, its transpose matrix, the inverse of the **K** matrix and the inverse of diagonal matrix **D**, in which the diagonal elements comprise a number of known parents and their inbreeding coefficients. The computing method is adaptable to the analysis of a data set including pairs of GIAs with imperfect relationships.

Introduction

Cloning animals is regarded as a means to multiply genetically identical animals (GIAs). In Japan, clones of bulls are routinely produced to test bulls' performance and in some cases to multiply fattening animals. A survey conducted by the Japanese Ministry of Agriculture, Forestry and Fisheries, and published on October 31, 2007, has recorded calves cloned from somatic cells of 535 animals and from embryonic cell nuclei of 716 animals.

In animal breeding, analysis of quantitative traits using a mixed model is essential to predict the breeding value of an individual and to estimate the genetic parameters of the traits. When applying an animal model to perform the genetic analysis, it is necessary to include the inverse of

the numerator relationship matrix (NRM) in order to connect all the animals included in the mixed model; however, calculating the inverse of a large NRM requires exceptionally large computing power. On the one hand, Henderson [1] has developed a method of calculating directly A^{-1} , without calculating the **A** matrix itself in a non-inbred population. This innovation has made it possible to use a model in which the data set includes a large number of animals. On the other hand, Quaas [2] has extended the method for the application to inbred populations by including the inbreeding coefficients in the model. A faster computing method of inbreeding coefficients has been developed by Tier [3] and Meuwissen and Luo [4], where inbreeding coefficients are computed as a subset of the **A** matrix. In addition, Famula [5] has pro-

posed a simplified algorithm for inbred populations, incorporating parental uncertainty to the model.

Inclusion of GIAs in the model raises the problem of a singular **A** matrix because of perfect additive genetic relationships between pairs of GIAs. In the case of an analysis with a singular **A**, Henderson [6] presented a method to solve a mixed model without inversion of the **G** matrix, where $G = A\sigma_a^2$. A few years later, Kennedy and Schaeffer [7] proposed a model in which records on GIAs are treated as repeated records on the same genotype. Their model assumes perfect genetic relationships among GIAs. However in reality, the genetic relationship between pairs of GIAs is not perfect because genetic diversity between such pairs originates from several genetic factors such as the difference in the recipient cytoplasm, mutations in the somatic cell and gene imprinting in the nucleus of the somatic cell [8-10]. Recently a study on human monozygotic twins analysing copy number variation, has revealed that genetic and phenotypic diversities exist even in monozygotic twins within pairs [11].

The objective of this study was to develop a method to compute directly the inverse of the NRM that includes GIAs with imperfect additive genetic relationships within pairs of GIAs.

Methods

Procedure

The **A** matrix is decomposed according to Famula [5].

$$A = (I - \frac{1}{2}P)^{-1} D (I - \frac{1}{2}P')^{-1}, \tag{1}$$

where **I** is the identity matrix, **P** is a lower triangular matrix which identifies the parents of each individual in the population. The **D** matrix is a diagonal matrix with d_i , i th diagonal element of **D**. Then

$$d_i = \begin{cases} \frac{1}{2} - \frac{1}{4}F_p - \frac{1}{4}F_q & \text{if both parents of } i, \text{ say } p \text{ and } q, \text{ are known,} \\ \frac{3}{4} - \frac{1}{4}F_p & \text{if only one parent, say } p, \text{ is known,} \\ 1 & \text{if neither parent is known,} \end{cases} \tag{2}$$

where F_p and F_q are the inbreeding coefficients of the parents of the i th animal. In (1), $(I - \frac{1}{2}P)$ is the lower triangular matrix, and $(I - \frac{1}{2}P')$ is its transpose matrix. From(1), the inverse of the **A** matrix is as follows:

$$A^{-1} = (I - \frac{1}{2}P') D^{-1} (I - \frac{1}{2}P). \tag{3}$$

Next, we introduce the **K** matrix into (1) so that the **A** matrix includes GIA. Then, **A** is

$$A \approx (I - \frac{1}{2}P)^{-1} DK (I - \frac{1}{2}P')^{-1} \tag{4}$$

and

$$A^{-1} \approx (I - \frac{1}{2}P') K^{-1} D^{-1} (I - \frac{1}{2}P), \tag{5}$$

where **K** is the matrix where a diagonal element is set to 1 and the off-diagonal element is set to $(1 - x)$ when i th and j th animals are genetically identical and x is a constant near 0; **K** is as follows:

$$K = \begin{bmatrix} 1 & 0 & \dots & & \\ 0 & \ddots & & & \\ \vdots & & 1 & 1-x & \\ & & 1-x & 1 & \\ & & & & \ddots \end{bmatrix}$$

Generally, for a matrix with GIAs where the diagonal elements are set to 1, and the off-diagonal elements are $(1 - x)$, the inverse matrix for n GIAs is

$$\begin{bmatrix} 1 & 1-x & 1-x & \dots \\ 1-x & 1 & 1-x & \dots \\ 1-x & 1-x & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}^{-1} = \begin{bmatrix} l_n & m_n & m_n & \dots \\ m_n & l_n & m_n & \dots \\ m_n & m_n & l_n & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where $l_n = \frac{1+(n-2)(1-x)}{x\{1+(n-1)(1-x)\}}$, $m_n = -\frac{1-x}{x\{1+(n-1)(1-x)\}}$

and $l_n + (n - 1)m_n = \frac{1}{1+(n-1)(1-x)}$.

Thus, **K**⁻¹ is calculated, and the diagonal element of **D**⁻¹ is $\frac{1}{d_i}$. Therefore, **A**⁻¹ is calculated directly by the product of the matrices without computing the inverse of **A**.

Algorithm for computation

Provided d_i is calculated by the methods of Quaas [2] and Famula [5], **A**⁻¹ is calculated directly by the following steps:

- i) If both parents of i , say p and q are known, and when i has no GIA,

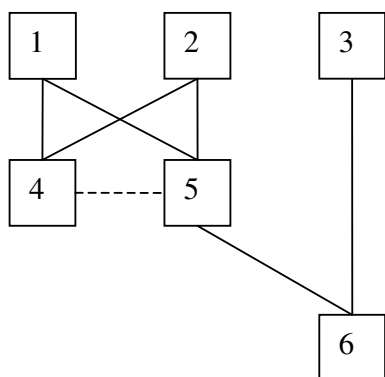


Figure 1
Diagram of the pedigree.

The D matrix has elements calculated in (2). Thus,

$$D = \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1/2 & & \\ & & & & 1/2 & \\ & & & & & 1/2 \end{bmatrix}$$

Therefore, A⁻¹ in (3) is as follows:

$$A^{-1} = (I - 1/2P)D^{-1}(I - 1/2P)$$

$$= \begin{bmatrix} 1 & 0 & 0 & -1/2 & -1/2 & 0 \\ & 1 & 0 & -1/2 & -1/2 & 0 \\ & & 1 & 0 & 0 & -1/2 \\ & & & 1 & 0 & 0 \\ & & & & 1 & -1/2 \\ & & & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 2 & & & \\ & & & 2 & & \\ & & & & 2 & \\ & & & & & 2 \end{bmatrix} \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 1 & 0 & -1 & -1 & 0 \\ 1 & 2 & 0 & -1 & -1 & 0 \\ 0 & 0 & 3/2 & 0 & 1/2 & -1 \\ -1 & -1 & 0 & 2 & 0 & 0 \\ -1 & -1 & 1/2 & 0 & 5/2 & -1 \\ 0 & 0 & -1 & 0 & -1 & 2 \end{bmatrix}$$

A from (4) is

$$A = \begin{bmatrix} 1 & 0 & 0 & 1/2 & 1/2 & 1/4 \\ & 1 & 0 & 1/2 & 1/2 & 1/4 \\ & & 1 & 0 & 0 & 1/2 \\ & & & 1 & 1/2 & 1/4 \\ & & & & 1 & 1/2 \\ & & & & & 1 \end{bmatrix}$$

2) When animals 4 and 5 are GIAs,

the K matrix is as follows:

$$K = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 \\ & & 1 & 0 & 0 & 0 \\ & & & 1 & 1-x & 0 \\ & & & & 1 & 0 \\ & & & & & 1 \end{bmatrix}$$

Then the inverse of the sub-matrix of K is

$$\begin{bmatrix} 1 & 1-x \\ 1-x & 1 \end{bmatrix}^{-1} = \begin{bmatrix} l_2 & m_2 \\ m_2 & l_2 \end{bmatrix},$$

where

$$l_2 = \frac{1+(2-2)(1-x)}{x\{1+(2-1)(1-x)\}} = \frac{1}{x(2-x)},$$

$$m_2 = -\frac{1-x}{x\{1+(2-1)(1-x)\}} = -\frac{1-x}{x(2-x)}, \quad \text{and}$$

$$l_2 + m_2 = \frac{1}{1+(1-x)}.$$

Therefore, K⁻¹ is

$$K^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 \\ & & 1 & 0 & 0 & 0 \\ & & & \frac{1}{x(2-x)} & -\frac{1-x}{x(2-x)} & 0 \\ & & & & \frac{1}{x(2-x)} & 0 \\ & & & & & 1 \end{bmatrix}$$

A⁻¹ in (5) is as follows:

$$A^{-1} = \begin{bmatrix} 1 + \frac{1}{2-x} & \frac{1}{2-x} & 0 & -\frac{1}{2-x} & -\frac{1}{2-x} & 0 \\ & 1 + \frac{1}{2-x} & 0 & -\frac{1}{2-x} & -\frac{1}{2-x} & 0 \\ & & 3/2 & 0 & 1/2 & -1 \\ & & & \frac{2}{x(2-x)} & -\frac{2(1-x)}{x(2-x)} & 0 \\ & & & & \frac{1}{2} + \frac{2}{x(2-x)} & -1 \\ & & & & & 2 \end{bmatrix}$$

Then A in (4) is

$$A = \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ & 1 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ & & 1 & 0 & 0 & \frac{1}{2} \\ & & & 1 & \frac{1}{2} + \frac{1}{2}(1-x) & \frac{1}{4} + \frac{1}{4}(1-x) \\ & & & & 1 & \frac{1}{2} \\ & & & & & 1 \end{bmatrix}$$

Here, if $x \rightarrow 0$, A is as expected,

$$A = \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ & 1 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ & & 1 & 0 & 0 & \frac{1}{2} \\ & & & 1 & 1 & \frac{1}{2} \\ & & & & 1 & \frac{1}{2} \\ & & & & & 1 \end{bmatrix}$$

Thus, the elements of animals 4 and 5 are those of the GIAs.

Simulation

Figure 2a presents the results of the estimated genetic variance. Averages of 20 estimates are shown together with their ranges (minimum and maximum). Genetic variances were overestimated at low (1-x) and around the true value at high (1-x) and estimated residual variances were underestimated at low (1-x) and around the true value at high (1-x) (data not shown).

Figure 2b shows averages of log likelihood (-2logL) for various (1-x) values together with their ranges when REML estimates were obtained for the variance components. The log likelihood declined at high (1-x): 0.95 or 0.99, indicating the validity of the model with high (1-x). The difference between the models with 0.95 and 0.99 was statistically insignificant by the likelihood ratio test.

Discussion

The K matrix proposed in this study is directly calculated by a simple formula; thus, calculating the inverse of a large matrix can be avoided as in standard methods [5,1,2]. Using NRM with GIAs results in adding more animal records to a dataset for variance component estimation and in calculating genetic evaluation by the mixed model procedure.

The simulation study showed that the estimated genetic variance is reasonably accurate with a high (1-x); however, a low (1-x) resulted in overestimation of the genetic vari-

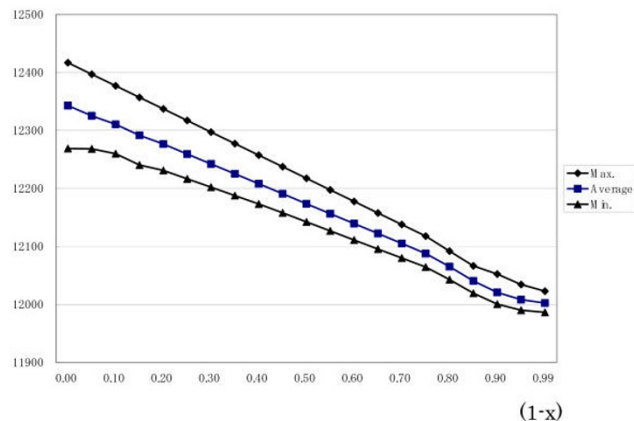
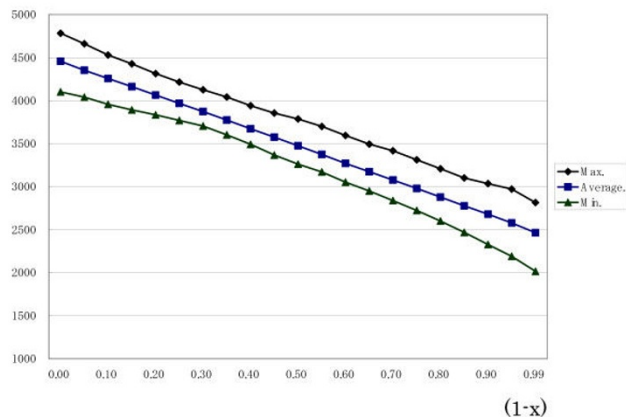


Figure 2
a. Averages and ranges of estimated additive genetic variances for different (1-x) values with 20 simulated data sets. b. Averages and ranges of log likelihood (-2logL) for different (1-x) values with 20 simulated data sets.

ance, which is caused by false genetic relationships between pairs of GIAs. For instance, in the case of (1-x) equal to 0.0, where the GIAs were erroneously treated as full sibs, a large genetic variance and, consequently, a small residual variance were estimated because the variance component within full sibs was far smaller than that expected for full sibs and vice versa for the variance component between full sibs.

This simulation study assumed perfect genetic relationships between pairs of GIAs. Then, a (1-x) value higher than 0.95 can result in unbiased estimates of additive genetic variance for the simulated data sets. Therefore, a (1-x) value of 0.95 is adequate for the data set of this simulation; however, the choice of the x value may depend on

the size of the data set and the genetic constituents of a population: a higher (1-x) may be adequate for a large data set and/or a data set containing monozygotic twins.

In the analysis of real records, a lower (1-x) value is expected because perfect genetic relationships between pairs of GIAs are no longer attainable. The highest genetic relationship is found among monozygotic twins; however, diversity within pairs of twins was found to be larger than expected, according to the human study by Bruder *et al.* [11]. A similar diversity is observed in embryo splitting studies, but manipulating embryos may constitute a potential source of increased diversity. Clones obtained from embryonic cell nuclear transfer may show a higher diversity caused by the recipient cytoplasm [9]. In the case of clones obtained from somatic cell nuclear transfer, an additional source for genetic diversity can originate from mutations in the somatic cell and gene imprinting in the nucleus of the somatic cell [8,10]. Different types of GIAs with various degrees of genetic diversity do exist. Thus, GIAs can be regarded as highly related animals rather than identical animals. Although, (1-x) values can range from 0.0 to 1.0, it is probably nearer to 1.0. To resolve this question, statistical studies such as REML on the estimation of x are needed with a large data set including various types of GIA. The methodology presented here provides an analytical tool to analyse GIAs with an imperfect genetic relationship within pairs of GIAs.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TO provided the idea of this study, prepared the research plan, initiated the statistical analysis and revised the draft manuscript. KY derived the equation, conducted the simulation study and wrote the draft of this manuscript.

References

- Henderson CR: **A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values.** *Biometrics* 1976, **32**:69-83.
- Quaas RL: **Computing the diagonal elements and inverse of a large numerator relationship matrix.** *Biometrics* 1976, **32**:949-953.
- Tier B: **Computing inbreeding coefficients quickly.** *Genet Sel Evol* 1990, **22**:419-430.
- Meuwissen THE, Luo Z: **Computing inbreeding coefficients in large populations.** *Genet Sel Evol* 1992, **24**:305-313.
- Famula TR: **Simple and rapid inversion of additive relationship matrices incorporating parental uncertainty.** *J Anim Sci* 1992, **70**:1045-1048.
- Henderson CR: **Prediction when G is singular.** In *Application of Linear Models in Animal Breeding* Guelph: Univiversity of Guelph Press; 1984:48-53.
- Kennedy BW, Schaeffer LR: **Genetic evaluation under an animal model when identical genotypes are represented in a population.** *J Anim Sci* 1989, **67**:1946-1955.
- Rideout VM III, Eggan K, Jaenisch R: **Nuclear cloning and epigenetic reprogramming of the genome.** *Science* 2001, **293**:1093-1098.
- Takeda K, Takahashi S, Onishi A, Goto Y, Miyazawa A, Imai H: **Dominant distribution of mitochondrial DNA from recipient oocyte in the bovine embryos and offspring after nuclear transfer.** *J Reprod Fertil* 1999, **116**:253-259.
- Yang X, Smith SL, Tian XC, Lewin HA, Renard J-P, Wakayama T: **Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning.** *Nat Genet* 2007, **39**:295-302.
- Bruder CEG, Piotrowski A, Gijsbers AACJ, Andersson R, Erickson S, de Ståhl TD, Menzel U, Sandgren J, von Tell D, Poplawski A, Crowley M, Crasto C, Partridge EC, Tiwari H, Allison DB, Komorowski J, van Ommen G-JB, Boomsma DI, Pedersen NL, den Dunnen JT, Wirdefeldt K, Dumanski JP: **Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles.** *Am J Hum Genet* 2008, **82**:763-771.
- Brown BW, Lovato J, Russell K: **RANLIB Library of fortran routines for random number generation.** 1991 [<http://www.netlib.org/random/index.html>].
- Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH: **BLUPF90 and related programs (BGF90).** *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production: 19-23 August 2002; Montpellier 2002.*

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

