

A Bayesian Model for the Analysis of Transgenerational Epigenetic Variation

Luis Varona,^{*,†,1} Sebastián Munilla,^{*,‡} Elena Flavia Mouresan,^{*} Aldemar González-Rodríguez,^{*}

Carlos Moreno,^{*,†} and Juan Altarriba^{*,†}

^{*}Unidad de Genética Cuantitativa y Mejora Animal, Universidad de Zaragoza, 50013, Zaragoza, Spain, [†]Instituto de Biocomputación y Física de los Sistemas Complejos (BIFI), Universidad de Zaragoza, 50018, Zaragoza, Spain, and

[‡]Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires, 1417, Ciudad Autónoma de Buenos Aires, Argentina

ORCID ID: 0000-0001-6256-5478 (L.V.)

ABSTRACT Epigenetics has become one of the major areas of biological research. However, the degree of phenotypic variability that is explained by epigenetic processes still remains unclear. From a quantitative genetics perspective, the estimation of variance components is achieved by means of the information provided by the resemblance between relatives. In a previous study, this resemblance was described as a function of the epigenetic variance component and a reset coefficient that indicates the rate of dissipation of epigenetic marks across generations. Given these assumptions, we propose a Bayesian mixed model methodology that allows the estimation of epigenetic variance from a genealogical and phenotypic database. The methodology is based on the development of a \mathbf{T} matrix of epigenetic relationships that depends on the reset coefficient. In addition, we present a simple procedure for the calculation of the inverse of this matrix (\mathbf{T}^{-1}) and a Gibbs sampler algorithm that obtains posterior estimates of all the unknowns in the model. The new procedure was used with two simulated data sets and with a beef cattle database. In the simulated populations, the results of the analysis provided marginal posterior distributions that included the population parameters in the regions of highest posterior density. In the case of the beef cattle dataset, the posterior estimate of transgenerational epigenetic variability was very low and a model comparison test indicated that a model that did not include it was the most plausible.

KEYWORDS

epigenetics
Bayesian analysis
genetic variance
resemblance
between
relatives

Epigenetics studies variations in gene expression that are not caused by modifications of the DNA sequence (Eccleston *et al.* 2007) and is now considered as one of the most important fields of biological research. Several biochemical mechanisms that alter gene activity (DNA methylation, histone modifications, etc.) underpin epigenetic processes (Jablonka and Raz 2009; Law and Jacobsen 2010). A number of studies have emphasized the importance of epigenetics in cancer (Esteller

2008) and other human illnesses (Feinberg 2007) as well as in other mammal (Reik 2007) and plant traits (Henderson and Jacobsen 2007). Gene activity modifications may occasionally occur in a sperm or an egg cell and can be transferred to the next generation, denoted as transgenerational epigenetic inheritance (Youngson and Whitelaw 2008), a phenomenon that has been reported in a wide range of organisms (Jablonka and Raz 2009). Despite this, the magnitude of the phenotypic variation that is explained by epigenetic processes still remains unclear (Grossniklaus *et al.* 2013; Heard and Martienssen 2014).

From the perspective of quantitative genetics, the presence of transgenerational epigenetic inheritance involves a redefinition of the covariance between relatives. Tal *et al.* (2010) developed a model for the calculation of the covariance between relatives for asexual and sexual reproduction as a function of epigenetic heritability (γ^2), the reset coefficient (ν), and its complement, the epigenetic transmission coefficient ($1 - \nu$). According to these authors, the covariance between relatives is reduced as the number of opportunities to dissipate (or reset) the epigenetic marks increase. Therefore, for sexual diploid organisms, the covariance between parent and offspring is greater

Copyright © 2015 Varona *et al.*

doi: 10.1534/g3.115.016725

Manuscript received December 1, 2014; accepted for publication January 19, 2015; published Early Online January 23, 2015.

This is an open-access article distributed under the terms of the Creative Commons Attribution Unported License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supporting information is available online at <http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.115.016725/-/DC1>

¹Corresponding author: Unidad de Genética Cuantitativa y Mejora Animal, Facultad de Veterinaria, Universidad de Zaragoza, c/ Miguel Servet 177, 50013, Zaragoza, Spain. E-mail: lvarona@unizar.es

than the covariance between full sibs, and the covariance between half sibs is greater than the covariance between an uncle and its nephew, despite the fact that the additive numerator relationship between them is identical (0.5 and 0.25, respectively).

In animal and plant breeding, the standard procedure for estimating variance components is by means of a linear mixed model (Henderson 1984) that includes systematic or random environmental effects, one or more random genetic effects, and a residual. The variance components are estimated by using likelihood-based procedures (Patterson and Thompson 1971) or Bayesian approaches (Gianola and Fernando 1986). Under the Bayesian paradigm, the standard procedure for determining the posterior distribution of the variance components uses the Gibbs Sampler algorithm (Gelfand and Smith 1990) that involves an updated iterative sampling scheme from the full conditional distributions of all the unknowns in the model.

The aim of this study is to present a Bayesian linear mixed model that allows one to estimate the heritable epigenetic variability, the reset coefficient, and the epigenetic transmission coefficient based on genealogical and phenotypic information. The model is illustrated with two simulated datasets and with one example that considers a beef cattle database.

MATERIAL AND METHODS

Statistical model

The standard mixed linear model is described by the following equation:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

where \mathbf{y} is the vector of phenotypic records, \mathbf{b} is the vector of systematic effects, \mathbf{u} is the vector of random additive genetic effects, and \mathbf{e} is the vector of residuals. Then, \mathbf{X} and \mathbf{Z} are the matrices that link the systematic and additive genetic effects with the data. The usual assumption for the prior distribution of \mathbf{u} and \mathbf{e} are the following multivariate Gaussian distributions (MVN):

$$\mathbf{u} \sim MVN(\mathbf{0}, \mathbf{A}\sigma_u^2) \quad \mathbf{e} \sim MVN(\mathbf{0}, \mathbf{I}\sigma_e^2)$$

where σ_u^2 and σ_e^2 are the additive genetic and residual variances, respectively, and \mathbf{A} is the numerator relationship matrix. Further, the prior distribution of \mathbf{b} is commonly assumed to be a uniform distribution. Conjugate priors for the variance components are the following inverted chi-square distributions:

$$\sigma_u^2 \sim \chi^{-1}(s_u^2, n_u) \quad \sigma_e^2 \sim \chi^{-1}(s_e^2, n_e)$$

where s_u^2 and s_e^2 are the prior values of the variances and n_u and n_e are their corresponding prior “degrees of belief”.

To estimate transgenerational epigenetic variability, this standard model can be expanded to:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{Z}\mathbf{w} + \mathbf{e}$$

where \mathbf{w} is the vector of individual epigenetic effects. Note that the incidence matrix (\mathbf{Z}) is the same for \mathbf{u} and \mathbf{w} and that both genetic and epigenetic effects are assumed to be independent. The prior distribution of \mathbf{w} is defined as:

$$\mathbf{w} \sim MVN(\mathbf{0}, \mathbf{T}\sigma_w^2)$$

where \mathbf{T} is the matrix of epigenetic relationships between individuals and σ_w^2 is the transgenerational epigenetic variance. As before, the prior distribution of σ_w^2 is defined as:

$$\sigma_w^2 \sim \chi^{-1}(s_w^2, n_w)$$

with hyperparameters s_w^2 and n_w .

The structure of the \mathbf{T} matrix is defined by the recursive relationship between the epigenetic effect of one individual (w_i) with respect to the epigenetic effects of its father (w_{fi}) and mother (w_{mi}):

$$w_i = \lambda w_{fi} + \lambda w_{mi} + \varepsilon_i$$

where

$$\lambda = \frac{1}{2}(1 - \nu)$$

As defined by Tal *et al.* (2010), ν is the reset coefficient and $(1 - \nu)$ is the epigenetic transmission coefficient. The reset coefficient represents the proportion of epigenetic marks across the parental genome that are expected to be erased, whereas its opposite, the epigenetic transmission coefficient, indicates the proportion that are transmitted. Furthermore, ε_i is the residual epigenetic effect of the i th individual, independent from w_i , and whose distribution is:

$$\varepsilon_i \sim N(0, (1 - 2\lambda^2)\sigma_w^2) \quad \text{if both parents are known;}$$

$$\varepsilon_i \sim N(0, (1 - \lambda^2)\sigma_w^2) \quad \text{if only one ancestor is known;}$$

assuming that the variance of transgenerational epigenetics effects ($V()$) is constant across generations:

$$V(w_i) = V(w_{fi}) = V(w_{mi}) = \sigma_w^2$$

In matrix notation:

$$\mathbf{w} = \mathbf{P}\mathbf{w} + \boldsymbol{\varepsilon} \quad (1)$$

where the \mathbf{P} matrix defines a recurrent relationship with the epigenetic effects of the father and mother. For nonbase individuals, the i th row of the \mathbf{P} matrix contains a parameter λ in the column pertaining to the father and mother of the i th individual. The rest of elements are null.

Furthermore if:

$$\mathbf{w} = [\mathbf{I} - \mathbf{P}]^{-1}\boldsymbol{\varepsilon}$$

then

$$\mathbf{V}(\mathbf{w}) = \mathbf{T}\sigma_w^2 = [\mathbf{I} - \mathbf{P}]^{-1}\mathbf{V}(\boldsymbol{\varepsilon})[\mathbf{I} - \mathbf{P}']^{-1}$$

where $\mathbf{V}(\boldsymbol{\varepsilon})$ is a diagonal matrix with entries equal to σ_w^2 for base individuals, $(1 - \lambda^2)\sigma_w^2$ for individuals with one known ancestor, and $(1 - 2\lambda^2)\sigma_w^2$ for individuals whose father and mother are known. The prior distribution for λ will be assumed to be uniform, between 0 and 0.5.

Parameter estimation through a Gibbs sampler

The Gibbs sampler algorithm is an iterative, updating sampling scheme that obtains samples from the marginal posterior distributions of all the unknowns in a model. It requires samples from the full conditional distributions of all the parameters. In the proposed model, the set of parameters can be classified into three main groups: i) the location parameters (\mathbf{b} , \mathbf{u} , and \mathbf{w}); ii) the parameter λ associated with matrix \mathbf{T} ; iii) the variance components (σ_u^2 , σ_w^2 , and σ_e^2).

The full conditional distributions for these groups are:

Sampling of the location parameters (\mathbf{b} , \mathbf{u} , \mathbf{w}): The conditional distributions of the location parameters are univariate Gaussian (Wang *et al.* 1994), with parameters drawn from the following mixed model equations:

$$\mathbf{C}\mathbf{s} = \mathbf{r}$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\boldsymbol{\alpha} & \mathbf{Z}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z} + \mathbf{T}^{-1}\boldsymbol{\psi} \end{bmatrix} \quad \mathbf{s} = \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \\ \mathbf{w} \end{bmatrix}$$

$$\mathbf{r} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

and $\boldsymbol{\alpha} = \sigma_e^2/\sigma_w^2$ and $\boldsymbol{\psi} = \sigma_e^2/\sigma_u^2$.

Specifically, the full conditional distribution for the i th location parameter (s_i) is:

$$s_i \sim N\left(\frac{r_i - \sum_{i \neq j} c_{ij}s_j}{c_{ii}}, \frac{\sigma_e^2}{c_{ii}}\right)$$

given the multivariate Gaussian nature of the conditional distribution of the location parameters.

A key limitation of the procedure is the calculation of the \mathbf{A}^{-1} and \mathbf{T}^{-1} matrices. Whereas \mathbf{A}^{-1} is calculated by the standard Henderson's rules (Henderson 1976) only once, the \mathbf{T}^{-1} matrix needs to be calculated afresh in each cycle as the parameter λ is updated. An algorithm to set-up the \mathbf{T}^{-1} matrix is presented in the appendix of this work (see Appendix).

The full conditional distribution of λ : The conditional distribution of λ is developed from the recursive definition of the epigenetic effects (equation 1). It is a truncated univariate Gaussian distribution (TN) between 0 and 0.5 due to the prior distribution of the parameter.

$$p(\lambda|\mathbf{w}) \sim TN_{[0,0.5]}(\mu_\lambda, \sigma_\lambda)$$

$$\mu_\lambda = \frac{\sum_{i=1}^{n_1} \frac{(w_{fi}+w_{mi})w_i}{(\sigma_w^2(1-2\lambda^2))} + \sum_{i=1}^{n_2} \frac{w_{fi}w_i}{(\sigma_w^2(1-\lambda^2))} + \sum_{i=1}^{n_3} \frac{w_{mi}w_i}{(\sigma_w^2(1-\lambda^2))}}{\sum_{i=1}^{n_1} \frac{(w_{fi}+w_{mi})^2}{(\sigma_w^2(1-2\lambda^2))} + \sum_{i=1}^{n_2} \frac{w_{fi}^2}{(\sigma_w^2(1-\lambda^2))} + \sum_{i=1}^{n_3} \frac{w_{mi}^2}{(\sigma_w^2(1-\lambda^2))}}$$

$$\sigma_\lambda = \frac{1}{\sum_{i=1}^{n_1} \frac{(w_{fi}+w_{mi})^2}{(\sigma_w^2(1-2\lambda^2))} + \sum_{i=1}^{n_2} \frac{w_{fi}^2}{(\sigma_w^2(1-\lambda^2))} + \sum_{i=1}^{n_3} \frac{w_{mi}^2}{(\sigma_w^2(1-\lambda^2))}}$$

where n_1 is the number of individuals with known fathers and mothers, n_2 is the number of individuals with only father known, and n_3 is the number of individuals with only mother known.

The full conditional distributions of the variance components ($\sigma_u^2, \sigma_w^2, \sigma_e^2$): The conditional distributions for the variance components are the following inverted χ^2 distributions

$$\begin{aligned} \sigma_u^2 &\sim \chi^{-2}(\mathbf{u}'\mathbf{A}^{-1}\mathbf{u} + s_u^2, nan + n_u) \\ \sigma_w^2 &\sim \chi^{-2}(\mathbf{w}'\mathbf{T}^{-1}\mathbf{w} + s_w^2, nan + n_w) \\ \sigma_e^2 &\sim \chi^{-2}(\mathbf{e}'\mathbf{e} + s_e^2, ndat + n_e) \end{aligned}$$

where nan is the number of individuals in the population, $ndat$ is the number of phenotypic data, and s_x^2 and n_x are the hyperparameters for σ_x^2 .

Simulated datasets

To evaluate the procedure, we simulated two different datasets. The first one had a relative high percentage of variation caused by additive and transgenerational epigenetic effects (35% and 20%, respectively) and an intermediate reset coefficient ($\nu = 0.40$). In the second dataset, a lower percentage of variation was explained by additive genetic and transgenerational epigenetic effects (15% and 10%, respectively) and the reset coefficient was greater ($\nu = 0.80$). Each dataset was composed by a three-generation population. The base populations included 3000 individuals (1500 sires and 1500 dams) and each of the two subsequent generations were composed by 3000 full sib families of 10 individuals each one. The sire and dam for each family were sampled randomly from the individuals of the previous generation. The genetic and epigenetic effects for each individual (u_i and w_i) in the base population were generated from the following Gaussian distributions:

$$\begin{aligned} u_i &\sim N(0, \sigma_u^2) \\ w_i &\sim N(0, \sigma_w^2) \end{aligned}$$

Furthermore, the genetic and epigenetic effects for the individuals (u_j and w_j) of the second and third generation were obtained by sampling from:

$$\begin{aligned} u_j &\sim N\left(\frac{1}{2}u_{fj} + \frac{1}{2}u_{mj}, \frac{\sigma_u^2}{2}\right) \\ w_j &\sim N(\lambda w_{fj} + \lambda w_{mj}, (1-\lambda)^2\sigma_w^2) \end{aligned}$$

where u_{fi} and u_{mj} and w_{fi} and w_{mj} were the additive genetic and transgenerational epigenetic effects of the father and the mother of the j th individual, respectively. In addition, one phenotypic record (y_i) was generated for every individual by:

$$y_i \sim N(\mu + u_i + w_i, \sigma_e^2)$$

where μ was a general mean set to 100 units. The values of the parameters for the two datasets were:

Dataset 1.

$$\begin{aligned} \sigma_u^2 &= 210, \quad \sigma_w^2 = 120, \quad \sigma_e^2 = 270, \quad \lambda = 0.30, \quad h^2 = 0.35, \\ \gamma^2 &= 0.20, \quad \nu = 0.40 \end{aligned}$$

being

$$h^2 = \sigma_u^2/(\sigma_u^2 + \sigma_w^2 + \sigma_e^2) \quad \text{and} \quad \gamma^2 = \sigma_w^2/(\sigma_u^2 + \sigma_w^2 + \sigma_e^2)$$

Dataset 2.

$$\begin{aligned} \sigma_u^2 &= 90, \quad \sigma_w^2 = 60, \quad \sigma_e^2 = 450, \quad \lambda = 0.10, \quad h^2 = 0.15, \\ \gamma^2 &= 0.10, \quad \nu = 0.80 \end{aligned}$$

The Gibbs sampler was implemented using own software written in Fortran 95. The analysis consisted of a single chain of 1,250,000

cycles, and the first 250,000 were discarded. Each analysis took 36 hr using a single thread of an Intel Xeon E5-2650 of 2.00 GHz. The source code is included as Supporting Information, File S1. Convergence was checked by the visual inspection of the chains and with the test of Raftery and Lewis (1992). All samples were stored for calculating summary statistics.

Pirenaica beef cattle data

As an illustrative example, we used a database of phenotypic records from the yield recording system of the Pirenaica beef cattle breed. The Pirenaica breed is a meat-type beef population from northern Spain with an approximate census of 20,000 individuals that are typically reared under extensive conditions (Sánchez *et al.* 2002). The data set was made up of 78,209 records for birth weight with an average value of 41.52 kg and a raw standard deviation of 4.65 kg. In addition, a pedigree file including 125,974 individual-sire-dam records was used. This information was provided by the National Pirenaica Breeders Confederation (*Confederación Nacional de Asociaciones de Ganado Pirenaico*; <http://www.conaspi.net>). Animal Care and Use Committee approval was not required for this study as field data were obtained from the Yield Recording System of the Pirenaica breed; furthermore, data were recorded by the stock-breeders themselves, under standard farm management, with no additional requirements.

The full model of analysis was:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{m} + \mathbf{Z}_2\mathbf{p} + \mathbf{Z}_3\mathbf{h} + \mathbf{Z}_1\mathbf{w} + \mathbf{e}$$

Where \mathbf{b} is the vector of fixed systematic effects that included sex (2 levels) and age of the mother (16 levels), \mathbf{u} and \mathbf{m} are the vectors of direct and maternal additive genetic effects with 125,974 elements, \mathbf{p} is the vector of permanent environmental maternal effects (21,143 levels), \mathbf{h} is the vector of the random herd-year-season effects (12,925 levels), \mathbf{w} is the vector of transgenerational epigenetic effects (125,974 levels), and \mathbf{e} is the vector of the residuals. Furthermore, \mathbf{X} , \mathbf{Z}_1 , \mathbf{Z}_2 , and \mathbf{Z}_3 are the incidence matrices that link the different effects with the phenotypic data. Appropriate uniform bounded distributions were assumed for the systematic effects and for each variance component ($\sigma_x^2 = \{\sigma_p^2, \sigma_h^2, \sigma_w^2, \sigma_e^2\}$), defined by hyperparameters ($n_x = -2$ and $s_x^2 = 0$). Furthermore, the prior distribution for the (co) variance matrix between direct and maternal genetic effects (\mathbf{G}) was the following inverted Wishart:

$$p(\mathbf{G}|n_G, \mathbf{G}_0) = IW(n_G, \mathbf{G}_0)$$

where

$$\mathbf{G} = \begin{pmatrix} \sigma_u^2 & \sigma_{um} \\ \sigma_{um} & \sigma_m^2 \end{pmatrix}$$

being σ_u^2 and σ_m^2 the direct and maternal genetic variances and σ_{um} the covariance between them. In this case, n_G was set to -3 and \mathbf{G}_0 to a 2×2 matrix of zeroes to define a uniform distribution.

Two statistical models (I and II) were fitted. Model I includes transgenerational epigenetic variance (σ_w^2), whereas Model II does not. For each model, the Gibbs sampler was implemented with a single chain of 3,250,000 cycles, and the first 250,000 were discarded. Convergence was checked by the visual inspection of the chains and with the test of Raftery and Lewis (1992). The models were compared based on the pseudo-log-marginal probability of the data (Gelfand 1996; Varona and Sorensen 2014) by computing the logarithm of the Conditional Predictive Ordinate (LogCPO)

calculated from the Markov chain Monte Carlo (MCMC) samples. It was calculated as:

$$\text{Log CPO} = \sum_i \ln \hat{p}(y_i | \mathbf{y}_{-i}, M_k)$$

where \mathbf{y}_{-i} is the vector of data with the i th datum (y_i) deleted, M_k is the k th model and

$$\hat{p}(y_i | \mathbf{y}_{-i}, M_k) = Ns \left[\sum_{j=1}^{Ns} \frac{1}{p(y_i | \theta_k^j, M_k)} \right]^{-1}$$

being Ns is the number of MCMC draws, θ_k^j is the j th draw from the posterior distribution of the parameters of the k th model.

RESULTS AND DISCUSSION

From the perspective of quantitative genetics, the estimation of variance components is obtained from the statistical information provided by the covariance between the phenotypes of relatives (Falconer and Mckay 1996). Tal *et al.* (2010) proposed a simple model for the resemblance between relatives under transgenerational epigenetic inheritance for asexual and sexual reproduction. This model assumes that transgenerational epigenetic effects are not correlated with the additive genetic ones, and that they are distributed under a Gaussian law. After the central limit theorem, this may be explained by a large number of epigenetic marks randomly distributed across the genome. In addition, the distribution pattern of these marks is assumed independent between individuals unless they are relatives. In that case, some of these marks should have not been erased during the meiosis drawing them apart from their common ancestor. The population rate of erasure of these marks is measured through the reset coefficient (ν).

In this study, we propose a procedure that makes use of the aforementioned authors' definition for sexual diploid organisms to estimate the parameters under a Bayesian mixed model framework. Our approach makes it feasible to estimate transgenerational epigenetic variability from huge datasets of genealogical and phenotypic data. The key to the procedure is the definition of a \mathbf{T} matrix of (co) variance between transgenerational epigenetic effects. This matrix exclusively depends on a single parameter (λ) which is directly related to the reset coefficient (ν) put forward by Tal *et al.* (2010). As an example, for a simple pedigree of seven individuals (Figure 1), the \mathbf{T} matrix is:

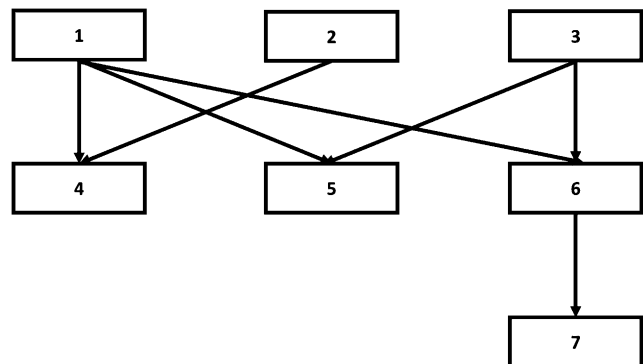


Figure 1 Example of a pedigree.

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & \lambda & \lambda & \lambda & \lambda^2 \\ 0 & 1 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \lambda & \lambda & \lambda^2 \\ \lambda & \lambda & 0 & 1 & \lambda^2 & \lambda^2 & \lambda^3 \\ \lambda & 0 & \lambda & \lambda^2 & 1 & 2\lambda^2 & 2\lambda^3 \\ \lambda & 0 & \lambda & \lambda^2 & 2\lambda^2 & 1 & \lambda \\ \lambda^2 & 0 & \lambda^2 & \lambda^3 & 2\lambda^3 & \lambda & 1 \end{bmatrix}$$

As Tal *et al.* (2010) stated, if we compare the covariance between full sibs (individuals 5 and 6) and the covariance between sire-offspring (individuals 1 and 4), the covariance is greater for the sire offspring pair, λ vs. $2\lambda^2$, as λ is defined as being within the interval (0, 0.5). Note that if the reset coefficient is 0, then λ is 1/2 and the \mathbf{T} matrix becomes equivalent to the numerator relationship matrix (\mathbf{A}), where the covariances between full sibs and sire offspring are equivalent. Similarly, the uncle–nephew covariance (individuals 5 and 7) is $2\lambda^3$, lower than the covariance between half sibs (individuals 4 and 5) which is only λ^2 .

The mixed model implementation of the covariance between relatives requires the inverse of the \mathbf{T} matrix to construct the mixed model equations, in a similar manner as the numerator relationship matrix in the standard model (Henderson 1984). In this study, we describe a simple procedure for the calculation of this \mathbf{T}^{-1} matrix (see the *Appendix* section). The procedure takes into account the recursive nature of the transgenerational epigenetic effects, using an argument equivalent to Quass (1976), for the inverse of numerator relationship matrix (\mathbf{A}^{-1}), and Quintanilla *et al.* (1999), for dam-related permanent maternal environmental effects. The main consequence is that the inverse of the \mathbf{T} matrix can be sequentially constructed by reading the pedigree of the population, in the light of very simple rules. These rules are close to Henderson's (1976) rules for constructing the inverse of the \mathbf{A} matrix.

For the example pedigree, the \mathbf{T}^{-1} matrix is:

$$\mathbf{T}^{-1} = \begin{bmatrix} 1 + \frac{3\lambda^2}{(1-2\lambda^2)} & 0 & 0 & \frac{-\lambda}{(1-2\lambda^2)} & \frac{-\lambda}{(1-2\lambda^2)} & \frac{-\lambda}{(1-2\lambda^2)} & 0 \\ 0 & 1 + \frac{\lambda^2}{(1-2\lambda^2)} & 0 & \frac{-\lambda}{(1-2\lambda^2)} & 0 & 0 & 0 \\ 0 & 0 & 1 + \frac{2\lambda^2}{(1-2\lambda^2)} & 0 & \frac{-\lambda}{(1-2\lambda^2)} & \frac{-\lambda}{(1-2\lambda^2)} & 0 \\ \frac{-\lambda}{(1-2\lambda^2)} & \frac{-\lambda}{(1-2\lambda^2)} & 0 & \frac{1}{(1-2\lambda^2)} & 0 & 0 & 0 \\ \frac{-\lambda}{(1-2\lambda^2)} & 0 & \frac{-\lambda}{(1-2\lambda^2)} & 0 & \frac{1}{(1-2\lambda^2)} & 0 & 0 \\ \frac{-\lambda}{(1-2\lambda^2)} & 0 & \frac{-\lambda}{(1-2\lambda^2)} & 0 & 0 & \frac{1}{(1-2\lambda^2)} + \frac{\lambda^2}{(1-\lambda^2)} & \frac{-\lambda}{(1-\lambda^2)} \\ 0 & 0 & 0 & 0 & 0 & \frac{-\lambda}{(1-\lambda^2)} & \frac{1}{(1-\lambda^2)} \end{bmatrix}$$

Given this algorithm to set-up the \mathbf{T}^{-1} matrix, the implementation of a Gibbs sampling approach becomes straightforward; it merely requires sequential iterative sampling from Gaussian (\mathbf{b} , \mathbf{u} , and \mathbf{w}),

truncated Gaussian (λ) and inverted χ^2 distributions (σ_u^2 , σ_w^2 , σ_e^2). The only minor complication is that the \mathbf{T}^{-1} matrix depends on parameter λ and consequently must be calculated in each Gibbs sampler's iteration. One possible alternative to avoid this matrix inversion, although not to avoid its updating in each cycle, is to implement an approach similar to the one proposed by Rodríguez-Ramilo *et al.* (2014) for the genomic relationship matrix. However, setting up the \mathbf{T} matrix is computationally more demanding than the calculation of its inverse with the algorithm proposed in this study.

The proposed procedure was first checked with simulated data. The summary of the marginal posterior distributions for the variance components, additive genetic (h^2), and epigenetic (γ^2) heritability, reset coefficient (ν), and epigenetic transmission coefficient ($1 - \nu$) for both cases of simulation are presented in Table 1. The greatest posterior density at 95% for all parameters in the model included the simulated values. However, some details of the results should be highlighted. In the second case of simulation, with lower σ_w^2 and λ , the posterior standard deviation (and the HPD95) for the σ_w^2 and σ_e^2 were remarkable wider (66.19 and 67.42, respectively). The cause of this phenomenon is that there is a statistical confounding between the epigenetic and residual variance components when the reset coefficient (ν) is very high, because the \mathbf{T} and \mathbf{I} matrices becomes very similar. To illustrate this fact, in Figure 2 and Figure 3, we present the joint posterior densities of γ^2 and ν for both cases of simulation. The results of the first case of simulation showed posterior independence between both parameters, whereas in the second, the marginal posterior density presented a half-moon shape. This indicates that for large values of ν , γ^2 (and σ_w^2) may take any value on its support with a fairly equal probability. In other words, in the first simulated case the model showed very good ability to discriminate between both parameters, whereas in the second it did not. In addition, mixing of the MCMC procedure in the second case of simulation was clearly worst, and adaptive MCMC algorithms, such as the proposed by Mathew *et al.* (2012) may represent an interesting alternative for its implementation in large data sets.

It must be noted that the sources of information for the estimation of σ_w^2 and ν (or λ) comes from the comparison between the covariance between different categories of relatives (see Table 2). Following

■ Table 1 PM, PSD, and HPD95 for simulation cases I and II

Parameter	Case I			Case II		
	$\sigma_u^2 = 210 \quad \sigma_w^2 = 120 \quad \sigma_e^2 = 270 \quad \lambda = 0.30$			$\sigma_u^2 = 90 \quad \sigma_w^2 = 60 \quad \sigma_e^2 = 450 \quad \lambda = 0.10$		
	PM	PSD	HPD95	PM	PSD	HPD95
σ_u^2	217.87	18.21	174.66–247.49	90.91	4.63	81.42–99.25
σ_w^2	132.92	14.42	106.30–164.53	111.99	66.19	37.45–288.61
σ_e^2	251.00	19.71	205.07–282.31	396.30	67.42	217.79–475.16
h^2	0.362	0.029	0.291–0.408	0.152	0.007	0.136–0.164
γ^2	0.221	0.024	0.176–0.274	0.187	0.110	0.062–0.481
λ	0.256	0.055	0.148–0.365	0.091	0.059	0.020–0.233
ν	0.488	0.111	0.270–0.704	0.818	0.117	0.534–0.960
$1 - \nu$	0.512	0.111	0.296–0.730	0.182	0.117	0.040–0.466

PM, posterior mean estimate; PSD, posterior standard deviation; HPD95, highest posterior density at 95%; σ_u^2 , additive genetic variance, σ_w^2 , transgenerational epigenetic variance; σ_e^2 , residual variance. Moreover, h^2 , heritability; γ^2 , transgenerational epigenetic heritability; λ , autoregressive parameter, ν , the reset coefficient; $1 - \nu$, epigenetic transmission coefficient.

Tal *et al.* (2010), one estimate of ν can be achieved from the difference between the estimates of covariance between half-sib and uncle–nephew relationships, a difference that becomes almost null for low values of λ (or high ν), as seen in the second case of simulation (23.10 vs. 22.62). On the contrary, with greater values of λ (or lower ν), as in the first case of simulation, this difference is higher (60.60 vs. 57.36), and, thus, the amount of information available for estimation purposes increases. This uncertainty in the estimation of λ (or ν) is also reflected in the estimation of σ_w^2 (γ^2) and σ_e^2 and implies wider posterior marginal densities. Nevertheless, it is important to emphasize that the joint posterior mode ($\nu = 0.80$, $\gamma^2 = 0.13$) in the second case of simulation was very close to the simulated parameters (0.80 and 0.10, respectively).

We have also checked our model with a real dataset. In Table 3, summary statistics of the marginal posterior distributions for the parameters and LogCPO values are presented for models I and II in the analysis of birth weight from the Pirenaica beef cattle population. The posterior mean estimates for heritability ranged between 0.38 (Model I) to 0.41 (Model II), and the posterior mean estimates for

transgenerational epigenetic heritability was only 0.04 (Model I), with a greatest posterior density at 95% that ranged between 0.00 and 0.11. These results are coherent with the output of the model comparison test (LogCPO), which pointed to Model II as the more plausible. If we focus in this latter model, the estimates of direct and maternal heritability were within the range of estimates in the literature for these traits (Meyer 1992; Varona *et al.* 1999; Jamrozik and Miller 2014), and the absence of the epigenetic transgenerational heritability is coherent with the fact that most of the epigenetic marks are erased during the meiosis in mammals (Jablonka and Raz 2009, Schmitz 2014). However, it should be highlighted that the posterior mean estimate of the reset coefficient under Model I was surprisingly low (0.20), although its posterior standard deviation was very high (0.20), and that the HPD95 region covered almost all the parametric space (0.01–0.89). This wide range reflects the absence of information to properly estimate the parameter, given the low magnitude of σ_w^2 .

It is worth noting that the procedure also can provide epigenetic breeding values in the fields of animal and plant breeding. These are calculated by weighting the phenotypic information of relatives

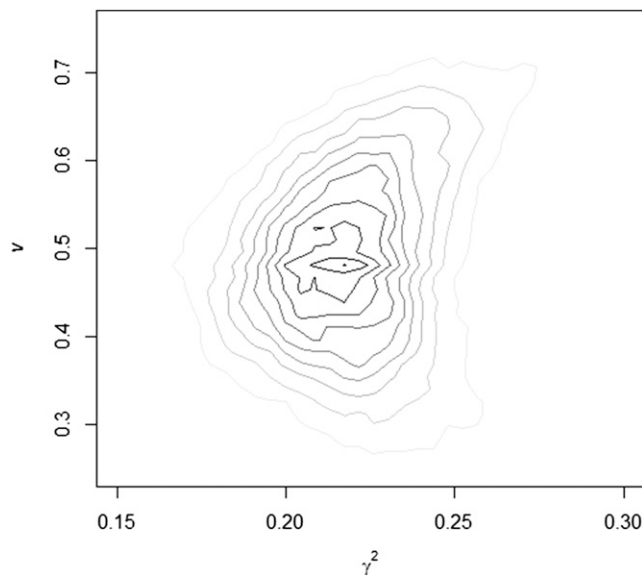


Figure 2 Joint posterior distribution of the transgenerational epigenetic heritability (γ^2) and the reset coefficient (ν) in the first case of simulation.

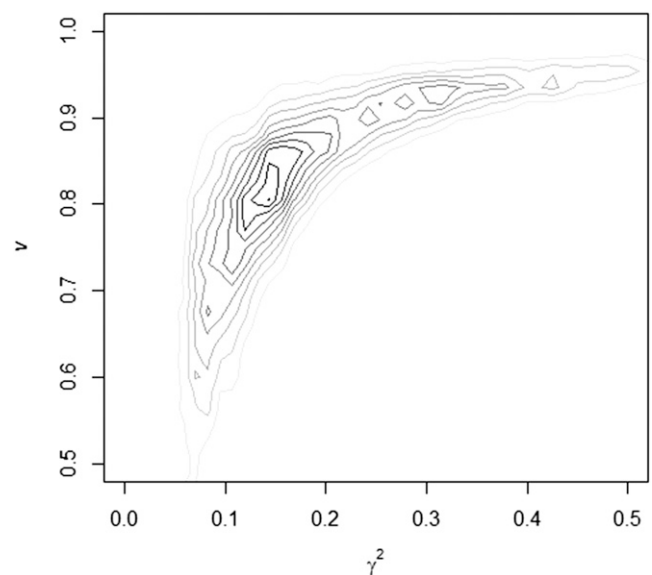


Figure 3 Joint posterior distribution of the transgenerational epigenetic heritability (γ^2) and the reset coefficient (ν) in the second case of simulation.

■ **Table 2** Expected covariance between relatives in cases of simulation I and II

Relatives	Expected Covariance	Case I		Case II	
		$\sigma_u^2 = 210$	$\sigma_w^2 = 120$ $\lambda = 0.30$	$\sigma_u^2 = 90$	$\sigma_w^2 = 60$ $\lambda = 0.10$
Offspring–Progeny	$\frac{1}{2}\sigma_u^2 + \lambda\sigma_w^2$		132		51
Full-Sibs	$\frac{1}{2}\sigma_u^2 + 2\lambda^2\sigma_w^2$		121.2		46.2
Half-Sibs	$\frac{1}{4}\sigma_u^2 + \lambda^2\sigma_w^2$		60.6		23.1
Uncle–Nephew	$\frac{1}{4}\sigma_u^2 + 2\lambda^3\sigma_w^2$		57.36		22.62

σ_u^2 , additive genetic variance; σ_w^2 , transgenerational epigenetic variance; λ , autorecursive parameter.

according to the magnitude of the elements of the T matrix. In fact, for epigenetic effects, the weight of distantly related individuals is lowered as the number of opportunities to reset the epigenetic marks increases. Both additive genetic and epigenetic effects can be used for prediction of the future performance of the individual: the expected genetic response (R) after one cycle of mass selection should be $R = i[h^2 + (1 - v)\gamma^2]\sigma_p^2$, where i is the intensity of selection and σ_p^2 is the phenotypic variance. However, further research must be undertaken to develop adequate indices of selection that consider genetic and epigenetic effects. Although both of them affect the immediate future performance of the offspring, epigenetic effects are diluted in future generations as the epigenetic marks are cleaned. It is important to note that if this selection pressure is relaxed, the average epigenetic effect declines to zero with a rate of $(1 - v)^n$, where n is the number of generations without selection.

It should be further noted that the proposed model uses a very basic definition of transgenerational epigenetic inheritance, as it assumes equal epigenetic variance for all individuals in the population. However, when phenotypic records are recorded through the life of the individual, transgenerational epigenetic variance can be modeled as age dependent, based on the consideration that the number of epigenetic marks was accumulated in the genome of the individuals. Moreover, in the proposed model, it is assumed that the λ parameter (or the reset coefficient) is equal for fathers and mothers, and, in future research, it seems reasonable to assume a different transmission coefficient for males and females, allowing for the presence of sex differential genomic imprinting (Barlow 1997; Reik and Walker

1998). The model can be refined by the inclusion of a hierarchical Bayesian paradigm that includes systematic effects for any environmental factors that may influence the epigenetic transmission coefficient. In addition, a more profound approach could even allow for the consideration of a genetic determinism of the reset coefficient with a model similar to that proposed by Varona *et al.* (2008) for the degree of asymmetry of a skewed Gaussian distribution.

Finally, and as mentioned by Jablonka and Raz (2009) and Tal *et al.* (2010), epigenetics may be viewed as a more wide ranging concept that could include several types of cultural transmission (Avital and Jablonka 2000; Richerson and Boyd 2005). The procedure suggested in this work also could be applied to the analysis of human or animal datasets that involve any kind of transgenerational transmission, even those not directly related to gene expression.

By way of a conclusion, we can say that this paper presents an original procedure for the estimation of transgenerational epigenetic variability based on some generalized assumptions. It is hoped that the proposal will lead to future research on variations of epigenetic transmission abilities caused by environmental and/or genetic factors.

ACKNOWLEDGMENTS

We thank the Pirenaica Breeders Association (Confederación Nacional de Asociaciones de Ganado Pirenaico) for the provision of data used in this work. We also thank the useful comments of the editor and the reviewers. This research was funded by grant AGL2010-15903 (Ministerio de Ciencia e Innovación, Spain).

■ **Table 3** PM, PSD, and HPD95 for models I and II

	Model I			Model II		
	PM	PSD	HPD95	PM	PSD	HPD95
σ_u^2	9.033	0.914	7.059, 10.449	9.897	0.408	9.123, 10.727
σ_m^2	3.574	0.212	3.168, 3.998	3.554	0.217	3.135, 3.981
σ_{um}	-4.473	0.253	-4.979, -3.990	-4.508	0.259	-5.032, -4.013
σ_p^2	0.747	0.081	0.590, 0.906	0.765	0.079	0.611, 0.922
σ_w^2	0.876	0.820	0.000, 2.701	–	–	–
σ_h^2	2.634	0.069	2.501, 2.771	2.633	0.069	2.499, 2.770
σ_e^2	7.055	0.223	6.606, 7.478	7.156	0.209	6.735, 7.556
λ	0.398	0.103	0.056, 0.494	–	–	–
v	0.204	0.207	0.012, 0.888	–	–	–
$1 - v$	0.796	0.207	0.112, 0.988	–	–	–
h^2	0.377	0.035	0.299, 0.427	0.412	0.012	0.389, 0.436
m^2	0.149	0.007	0.135, 0.164	0.148	0.007	0.133, 0.162
γ^2	0.036	0.034	0.000, 0.113	–	–	–
LogCPO		-287542.3			-287475.1	

PM, posterior mean estimate; PSD, posterior standard deviation; HPD95, highest posterior density at 95%; σ_u^2 , additive genetic variance, σ_m^2 , maternal environmental variance, σ_{um} , covariance between them, σ_p^2 , permanent maternal environmental variance, σ_h^2 , herd-year-season variance, σ_w^2 , transgenerational epigenetic variance, and σ_e^2 , residual variance. Moreover, h^2 , heritability, m^2 , maternal heritability, γ^2 , transgenerational epigenetic heritability, λ , autorecursive parameter, v , reset coefficient, and $1 - v$, epigenetic transmission coefficient. LogCPO, logarithm of the conditional predictive ordinate.

LITERATURE CITED

- Avital, E., and E. Jablonka, 2000 *Animal Traditions: Behavioural Inheritance in Evolution*. Cambridge University Press, Cambridge.
- Barlow, D. P., 1997 Competition—a common motif for the imprinting mechanism? *EMBO J.* 16: 6899–6905.
- Eccleston, A., N. DeWitt, C. Gunter, B. Marte, and D. Nath, 2007 Introduction epigenetics. *Nature* 447: 395–400.
- Esteller, M., 2008 Epigenetics of cancer. *N. Engl. J. Med.* 358: 1148–1159.
- Falconer, D. S., and T. F. C. McKay, 1996 *Introduction to Quantitative Genetics*. Ed. 4. Longmans Green, Harlow, Essex.
- Feinberg, A. P., 2007 Phenotypic plasticity and the epigenetics of human disease. *Nature* 447: 433–440.
- Gelfand, A. E., 1996 Model determination using sampling-based methods, pp. 145–161 in *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. Chapman & Hall, London.
- Gelfand, A. E., and A. F. M. Smith, 1990 Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 85: 398–409.
- Gianola, D., and R. L. Fernando, 1986 Bayesian methods in animal breeding theory. *J. Anim. Sci.* 63: 217–244.
- Grossniklaus, U., W. G. Kelly, A. C. Ferguson-Smith, M. Pembrey, and S. Lindquist, 2013 Transgenerational epigenetic inheritance: how important is it? *Nat. Rev. Genet.* 14: 228–235.
- Heard, E., and R. A. Martienssen, 2014 Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* 157: 95–109.
- Henderson, C. R., 1976 A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. *Biometrics* 32: 69–83.
- Henderson, C. R., 1984 *Application of Linear Models in Animal Breeding*. University of Guelph, Guelph, ON.
- Henderson, I. R., and S. E. Jacobsen, 2007 Epigenetic inheritance in plants. *Nature* 447: 418–424.
- Jablonka, E., and G. Raz, 2009 Transgenerational epigenetic inheritance: prevalence, mechanisms and implications for the study of heredity and evolution. *Q. Rev. Biol.* 84: 131–176.
- Jamrozik, J., and S. P. Miller, 2014 Genetic evaluation of calving ease in Canadian Simmentals using birth weight and gestation length as correlated traits. *Livest. Sci.* 162: 42–49.
- Law, J. A., and S. E. Jacobsen, 2010 Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11: 204–220.
- Mathew, B., A. M. Bauer, P. Koistinen, T. C. Reetz, J. Léon *et al.*, 2012 Bayesian adaptive Markov chain Monte Carlo estimation of genetic parameters. *Heredity* 109: 235–234.
- Meyer, K., 1992 Variance components due to direct and maternal effects for growth traits of Australian beef cattle. *Livest. Prod. Sci.* 11: 143–177.
- Patterson, H. D., and R. Thompson, 1971 Recovery of inter-block information when block sizes are unequal. *Biometrika* 58: 545–554.
- Quass, R. L., 1976 Computing the diagonal elements of the inverse of a large numerator relationship matrix. *Biometrics* 32: 949–953.
- Quintanilla, R., L. Varona, M. R. Pujol, and J. Piedrafita, 1999 Maternal animal model with correlation between maternal environmental effects of related dams. *J. Anim. Sci.* 77: 2904–2917.
- Raftery, A. E., and S. M. Lewis, 1992 How many iterations in the Gibbs sampler, pp. 763–774 in *Bayesian Statistics IV*, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith. Oxford University Press, New York.
- Reik, W., 2007 Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447: 425–432.
- Reik, W., and J. Walker, 1998 Imprinting mechanisms in mammals. *Curr. Opin. Genet. Dev.* 8: 154–164.
- Richerson, P. J., and R. Boyd, 2005 *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press, Chicago.
- Rodríguez-Ramilo, S. T., L. A. García-Cortés, and O. González-Recio, 2014 Combining genomic and genealogical information in a reproducing kernel Hilbert spaces regression model for genome-enabled predictions in dairy cattle. *PLoS One* 9: e93424.
- Sánchez, A., J. Ambrona, and L. Sánchez, 2002 *Razas Ganaderas Españolas Bovinas*. MAPA-FEAGAS. Spain, Madrid.
- Schmitz, R. J., 2014 The secret garden—epigenetic alleles underlies complex traits. *Science* 343: 1082–1083.
- Tal, O., E. Kisdí, and E. Jablonka, 2010 Epigenetic contribution to covariance between relatives. *Genetics* 184: 1037–1050.
- Varona, L., and D. Sorensen, 2014 Joint analysis of binomial and continuous traits with recursive model: a case study using mortality and litter size in pigs. *Genetics* 196: 643–651.
- Varona, L., I. Misztal, and J. K. Bertrand, 1999 Threshold-linear versus linear-linear analysis of birth weight and calving ease using an animal model: I. Variance component estimation. *J. Anim. Sci.* 77: 1994–2002.
- Varona, L., N. Ibañez-Escriche, R. Quintanilla, J. L. Noguera, and J. Casellas, 2008 Bayesian analysis of quantitative traits using skewed distributions. *Genet. Res.* 90: 179–190.
- Wang, C. S., J. J. Rutledge, and D. Gianola, 1994 Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genet. Sel. Evol.* 26: 91–115.
- Youngson, N. A., and E. Whitelaw, 2008 Transgenerational epigenetic effects. *Annu. Rev. Genomics Hum. Genet.* 9: 233–257.

Communicating editor: D. J. de Koning

APPENDIX

Derivation of the \mathbf{T}^{-1} matrix

The vector of epigenetic effects (\mathbf{w}) can be represented as:

$$\mathbf{w} = \mathbf{P}\mathbf{w} + \varepsilon$$

where the \mathbf{P} matrix defines a recurrent relationship with the epigenetic effects of the father and mother. For nonbase individuals, the i th row of the \mathbf{P} matrix contains a parameter λ in the column pertaining to the father and mother of the i th individual. The rest of the elements are null.

Furthermore, if:

$$\mathbf{w} = [\mathbf{I} - \mathbf{P}]^{-1}\varepsilon$$

then

$$\mathbf{V}(\mathbf{w}) = [\mathbf{I} - \mathbf{P}]^{-1}\mathbf{V}(\varepsilon)[\mathbf{I} - \mathbf{P}']^{-1}$$

given that $\mathbf{V}(\mathbf{w}) = \mathbf{T}\sigma_w^2$

$$[\mathbf{V}(\mathbf{w})]^{-1} = \mathbf{T}^{-1}\frac{1}{\sigma_w^2} = [\mathbf{I} - \mathbf{P}'][\mathbf{V}(\varepsilon)]^{-1}[\mathbf{I} - \mathbf{P}] \quad (\text{A1})$$

From $\varepsilon = [\mathbf{I} - \mathbf{P}]\mathbf{w}$, it is possible to define a \mathbf{Q}_T matrix as:

$$\mathbf{Q}_T = \mathbf{V}(\varepsilon)\frac{1}{\sigma_w^2} = [\mathbf{I} - \mathbf{P}']\mathbf{T}[\mathbf{I} - \mathbf{P}]' \quad (\text{A2})$$

If we consider the structure of $[\mathbf{I} - \mathbf{P}]$ and $[\mathbf{I} - \mathbf{P}']$, the \mathbf{Q}_T matrix has a diagonal structure with 1s for base individuals and $1 - \lambda^2$ and $1 - 2\lambda^2$ for those with one or two known parents, respectively.

Finally, replacing (A2) into (A1):

$$\mathbf{T}^{-1} = [\mathbf{I} - \mathbf{P}']\mathbf{Q}_T^{-1}[\mathbf{I} - \mathbf{P}]$$

Taking advantage of this structure, it is possible to define the following algorithm for constructing \mathbf{T}^{-1} by reading the pedigree information and calculating for each individual (i) in the pedigree:

- If the father and mother are unknown;
 - add 1 to the diagonal (i,i)
- If only one parent (p) is known;
 - add $\frac{1}{(1-\lambda^2)}$ to the diagonal (i,i)
 - add $\frac{-\lambda}{(1-\lambda^2)}$ to the elements (i,p), (p,i)
 - add $\frac{\lambda^2}{(1-\lambda^2)}$ to the element (p,p)
- If both ancestors (p and m) are known;
 - add $\frac{1}{(1-2\lambda^2)}$ to the diagonal (i,i)
 - add $\frac{-\lambda}{(1-2\lambda^2)}$ to the elements (i,p), (i,m), (m,i), (p,i)
 - add $\frac{\lambda^2}{(1-2\lambda^2)}$ to the elements (p,p), (p,m), (m,p), (m,m)