

1 **Generalized gametic relationships for flexible analyses of parent-of-**  
2 **origin effects**

3

4

**N. Reinsch, M. Mayer, I. Blunk**

5

6

**Institute of Genetics and Biometry,**

7

**Leibniz-Institute for Farm Animal Biology,**

8

**Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany**

9

## 10 **Abstract**

11 Genomic imprinting causes alleles to influence the phenotype in a parent-of-origin-specific  
12 manner. In attempts to determine the effects of imprinted loci, gametic relationship  
13 matrices have widely been used in pedigree-based parent-of-origin analyses of population  
14 data. One drawback of this is the size of these matrices because they represent each  
15 individual by two gametic effects. Significantly fewer equations are needed if a previously  
16 published reduced imprinting model is used that relates observations from progeny without  
17 its own offspring to the transmitting abilities of their parents. This can be accomplished  
18 using a numerator relationship matrix, with only a single row and column per parent and  
19 ancestors. However, the reduced model is not applicable when the parents have records. To  
20 better handle the curse of dimensionality, we propose a combination of average gametic  
21 effects (transmitting abilities) for individuals without their own records and single gametic  
22 effects for others. The generalized gametic relationship matrix is the covariance of this  
23 mixture of genetic effects that allows for a significant reduction in the number of equations  
24 in gametic models depending on the trait, depth of pedigree, and population structure. It  
25 can also render the reduced model much more flexible by including observations from  
26 parents. Rules for setting-up its inverse from a pedigree are derived and implemented on an  
27 open-source program. The application of the same principles to phased marker data leads to  
28 a genomic version of the generalized gametic relationships. The implementation of  
29 generalized gametic models to the ASReml package is illustrated through worked examples.

30

31 Shortly after its discovery, it was recognized that the gametic relationship matrix (Smith and Allaire,  
32 1985; Schaeffer et al., 1989) can help isolate fractions of the genetic variance in quantitative traits  
33 caused by genomically imprinted loci. Alleles of the latter are expressed in a parent-of-origin-specific  
34 manner. In the early stages of pedigree-based imprinting analysis, animal models were augmented  
35 by an additional vector of paternal (alternatively, maternal) gametic effects, usually modeled as  
36 uncorrelated with any other effect. Its variance was assumed to be the product of a gametic  
37 relationship matrix and a variance component that can be explained by polymorphisms at loci with  
38 only paternal (maternal) gene activity. Pioneered by DeVries et al. (1994), these models were in use  
39 for more than a decade. However, they can account only for a single kind of classical imprinting,  
40 where either maternal or paternal alleles are fully silenced through, e.g., the methylation of DNA. A  
41 proposal (Hill and Keithly, 1988) to consider both kinds of imprinting simultaneously did not  
42 materialize in any pedigree-based analysis of empirical data. Further, there was uncertainty  
43 regarding ways to account for the effects of partially imprinted loci, where both alleles are expressed  
44 but at different strengths depending on their parental origins.

45 A model for parent-of-origin analysis was subsequently developed (Neugebauer et al., 2010a, b) that  
46 is comprehensive in the sense that it accounts for all kinds of imprinting, be it full or partial, maternal  
47 or paternal (Blunk et al., 2014). This so-called reduced imprinting model relates observations from  
48 non-parents (final progeny, e.g., animals used for meat) to transmitting abilities (half of the breeding  
49 values) of their parents. There are two correlated genetic effects per parent, a transmitting ability as  
50 sire and a transmitting ability as dam, which reflect an animal's genetic effect on its offspring under a  
51 paternal or maternal imprinting pattern. In the presence of genomically imprinted loci, these two  
52 genetic effects are different. The variance of these differences has been called the imprinting  
53 variance because it summarizes contributions from all kinds of possible imprinted loci. A numerator  
54 relationship matrix is needed for parents only, as the final progeny with observations but without  
55 offspring do not appear in the underlying pedigree and the resulting relationship matrix. The null  
56 hypothesis of the absence of polymorphic imprinted loci with an effect on the trait under  
57 investigation (i.e., a zero imprinting variance) can be tested by a restricted maximum likelihood  
58 (REML) ratio test.

59 Alternatively to the above, a comprehensive *gametic model* can be used to estimate the same set of  
60 genetic covariances, including the imprinting variance (Tier and Meyer, 2012; Meyer and Tier, 2012).  
61 This requires four gametic effects to be estimated per individual, two as sire and two as dam, where  
62 the relationships include the final progeny with observations. As an advantage over the reduced

63 model, the gametic model allows for records from parents. Moreover, it can be extended to account  
64 for maternal effects (see Appendix A5).

65 The use of measured genotypes in genomic best linear unbiased prediction models (gBLUP) that  
66 include imprinting effects has been outlined by Nishio and Satoh (2015). The first (GBLUP-I1) of the  
67 two variants of the proposed model contains an imprinting effect that is modeled as independent of  
68 the action of un-imprinted Mendelian locus, summarized as an additive genetic effect. The second  
69 model (GBLUP-I2) considers a paternal and a maternal gametic effect with zero mutual correlation.  
70 This clearly could be turned into a comprehensive model by abandoning the assumption of a zero  
71 correlation and replacing pedigree-derived gametic relationships by a genomic counterpart of equal  
72 size and structure. In cases where not all pedigreed individuals are genotyped, this enables a  
73 combined analysis of the genotyped and un-genotyped individuals in a single-step approach (Legarra  
74 et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). The first model (GBLUP-I1), by  
75 contrast, cannot easily be extended to have such a pedigree-derived counterpart.

76 The downside of the gametic model is the large number of equations (Smith and Allaire, 1985) used  
77 to represent the random genetic effects, in particular when variance components are to be  
78 estimated. A pedigree with a size of approximately half a million is a technical barrier for REML  
79 estimation in animal models using currently available software packages (Shor et al., 2019). With a  
80 gametic parent-of-origin model, the same number of equations is reached with only a quarter of  
81 individuals. Therefore, the question arises if there is any option for models that retain the flexibility  
82 of the gametic model while allowing for a considerably smaller number of equations for random  
83 genetic effects, as close as possible to the reduced imprinting model.

84 As a solution, we propose a much smaller re-defined vector of genetic effects obtained by a proper  
85 linear transformation of the gametic effects. This is rendered applicable by introducing a version of a  
86 corresponding relationship matrix, called the generalized gametic relationship matrix, together with  
87 rules for its rapid inversion from the pedigree. As a result, the size of the gametic model can be  
88 reduced to a more manageable one while retaining all of its advantages. We also show how the same  
89 kind of transformation can be applied to measured genotypes to obtain conformable genomic and  
90 pedigree-derived versions of the new relationship matrix.

## 91 **THEORY**

### 92 **Generalized gametic relationships**

93 In gametic models, each individual  $i$  is represented by the additive genetic effects of its paternal  
94 gamete  $g_{i,1}$  and maternal gamete  $g_{i,2}$  (Schaeffer et al., 1989), which usually are arranged in a pair-

95 wise manner in a vector  $\mathbf{g}$  of length  $2t$ , which is twice the number  $t$  of individuals in the pedigree.

96 The model equation for a phenotypic observation  $y_i$  of individual  $i$  then is

97 
$$y_i = \mu_i + g_{i,1} + g_{i,2} + e_i ,$$

98 With  $\mu_i = \mathbf{x}'_i \boldsymbol{\beta}$  as a place-holder for any combination of explanatory variables in vector  $\mathbf{x}'_i$  with fixed

99 effects  $\boldsymbol{\beta}$ , and the residual  $e_i$ . Thus, the gametic model splits the additive genetic value (breeding

100 value)  $b_i$  of individual  $i$  into paternally derived and maternally derived parts,  $b_i = g_{i,1} + g_{i,2}$ .

101 The basic idea of reducing equations in gametic models by a considerable number is to replace the

102 two gametic effects of a subset of  $u$  individuals by their pair-wise average:

103 
$$\frac{1}{2}(g_{i,1} + g_{i,2}) = a_i ,$$

104 which is known as the transmitting-ability (half the breeding value) of individual  $i$ .

105 The vector  $\mathbf{g}$  of gametic effects can be arranged such that the gametic effects of all  $u$  individuals

106 precede the gametic effects of the  $v$  that are bound to retain their distinct gametic effects. The

107 corresponding subdivision of  $\mathbf{g}$  is

108 
$$\mathbf{g} = \begin{bmatrix} \mathbf{g}_u \\ \mathbf{g}_v \end{bmatrix} .$$

109 The sub-vectors  $\mathbf{g}_u$  and  $\mathbf{g}_v$  have respective lengths of  $2u$  and  $2v$ . The covariances of all gametic

110 effects in  $\mathbf{g}$  are the elements of the  $2t \times 2t$  gametic relationship matrix  $\mathbf{G}$  (Schaeffer et al., 1989). It

111 can be partitioned into sections that correspond to the relationships between the gametic effects in

112  $\mathbf{g}_u$  and  $\mathbf{g}_v$ .

113 
$$Var \begin{bmatrix} \mathbf{g}_u \\ \mathbf{g}_v \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{uu} & \mathbf{G}_{uv} \\ \mathbf{G}'_{uv} & \mathbf{G}_{vv} \end{bmatrix} = \mathbf{G}$$

114 The required average gametic effects can be obtained by a linear transformation that is defined by a

115 matrix  $\mathbf{K}'$ , such that

116 
$$\mathbf{K}' \mathbf{g} = \begin{bmatrix} \mathbf{K}'_u & \mathbf{0}_1 \\ \mathbf{0}_2 & \mathbf{I}_v \end{bmatrix} \begin{bmatrix} \mathbf{g}_u \\ \mathbf{g}_v \end{bmatrix} = \begin{bmatrix} \mathbf{a}_u \\ \mathbf{g}_v \end{bmatrix} = \mathbf{a} .$$

117 In effect, all gametic effects of individuals in  $\mathbf{g}_u$  are replaced by their transmitting abilities in  $\mathbf{a}_u$ . The  
 118 upper-left partition  $\mathbf{K}'_u$  of the transformation matrix  $\mathbf{K}'$  has dimensions  $u \times 2u$ , and is defined as  
 119 the Kronecker product of a  $u \times u$  identity matrix  $\mathbf{I}_u$  and a row vector with two elements equal to  $\frac{1}{2}$ :

$$120 \quad \mathbf{K}'_u = \mathbf{I}_u \otimes \left[ \frac{1}{2} \quad \frac{1}{2} \right].$$

121 Further,  $\mathbf{K}'$  comprises a  $2v \times 2v$  identity matrix  $\mathbf{I}_v$  and two null matrices,  $\mathbf{0}_1$  and  $\mathbf{0}_2$ , with  
 122 respective dimensions of  $u \times 2v$  and  $2v \times 2u$ .

123 The covariance matrix of the transformed vector of gametic effects  $\mathbf{a}$  then becomes

$$124 \quad \text{Var} \begin{bmatrix} \mathbf{a}_u \\ \mathbf{g}_v \end{bmatrix} = \mathbf{K}'\mathbf{G}\mathbf{K} = \bar{\mathbf{G}},$$

125 which in the following is called a generalized gametic relationship matrix. A natural choice is to retain  
 126 the gametic effects of all individuals with their own phenotypes in vector  $\mathbf{g}_v$  and let all their  
 127 ancestors without records be represented by their transmitting abilities, constituting  $\mathbf{a}_u$ . The  
 128 subdivisions of  $\bar{\mathbf{G}}$  then are

$$129 \quad \bar{\mathbf{G}} = \begin{bmatrix} \mathbf{K}'_u\mathbf{G}_{uu}\mathbf{K}_u & \mathbf{K}'_u\mathbf{G}_{uv}\mathbf{I}_v \\ \mathbf{I}_v\mathbf{G}'_{uv}\mathbf{K}_u & \mathbf{I}_v\mathbf{G}_{vv}\mathbf{I}_v \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\mathbf{A}_u & \mathbf{S}_{uv} \\ \mathbf{S}'_{uv} & \mathbf{G}_{vv} \end{bmatrix}.$$

130 The upper-left part  $\frac{1}{2}\mathbf{A}_u$  is equal to the co-ancestry matrix (half the numerator-relationship matrix)  
 131 of all ancestors without own records, while  $\mathbf{G}_{vv}$  reflects relationships between the gametic effects of  
 132 all individuals with their own observations. Finally,  $\mathbf{S}_{uv}$  contains the covariances between  
 133 transmitting abilities and gametic effects. See the small example involving four individuals (IDs).  
 134 There are three transmitting abilities for individuals 1, 2, and 3, with corresponding pair-wise  
 135 elements of  $\frac{1}{2}$  in the transformation matrix  $\mathbf{K}'$  and two gametic effects, for which the elements in  
 136  $\mathbf{K}'$  are one. The resulting generalized gametic relationship matrix  $\bar{\mathbf{G}}$  has dimensions  $5 \times 5$ .

ID	sire	dam
1	0	0
2	0	0
3	1	2
4	1	3

$$137 \quad \mathbf{K}' = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \bar{\mathbf{G}} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{4} & 1 & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 1 \end{bmatrix}$$

138

### 139 Generalized gametic relationships in a gametic model

140 In light of the above, the model equation for an observation  $y_i$  can be retained as in the gametic  
 141 model, and a mixed model that considers parent-of-origin effects (POEs) and uses the generalized  
 142 relationship matrix becomes

143 
$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_s\mathbf{a}_s + \mathbf{Z}_d\mathbf{a}_d + \mathbf{e} ,$$

144 where  $\mathbf{Y}$  is a vector of observations,  $\boldsymbol{\beta}$  comprises the fixed effects, and  $\mathbf{X}$  is the corresponding  
145 incidence matrix. The covariance of random effects is assumed to be

146 
$$\text{Var} \begin{bmatrix} \mathbf{a}_s \\ \mathbf{a}_d \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{G}}\sigma_s^2 & \bar{\mathbf{G}}\sigma_{sd} & \mathbf{0} \\ \bar{\mathbf{G}}\sigma_{sd} & \bar{\mathbf{G}}\sigma_d^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2 \end{bmatrix} .$$

147 This generalized gametic model contains the gametic effect vectors  $\mathbf{g}_s$  and  $\mathbf{g}_d$  replaced by their  
148 transformed counterparts  $\mathbf{a}_s$  and  $\mathbf{a}_d$ , respectively, and, consequently, uses the corresponding  
149 relationship matrix  $\bar{\mathbf{G}}$  instead of the classical gametic relationships of  $\mathbf{G}$ . Further, incidence  
150 matrices  $\mathbf{Z}_s$  and  $\mathbf{Z}_d$  link observations to the random gametic effects in  $\mathbf{a}_s$  and  $\mathbf{a}_d$ , respectively,  
151 while no observation is linked to any of the transmitting abilities in the latter vectors. As a result, any  
152 incidence matrix  $\mathbf{Z}^a = [\mathbf{0}^u \quad \mathbf{Z}^v]$  that links observations to gametic effects in the generalized vector  
153 of genetic effects  $\mathbf{a}' = [\mathbf{a}'_u \quad \mathbf{g}'_v]$  can be considered a converted incidence matrix  $\mathbf{Z}^g = [\mathbf{0}^{2u} \quad \mathbf{Z}^v]$   
154 from a classical gametic model that links the observations to the gametic effects in  $\mathbf{g}' = [\mathbf{g}'_u \quad \mathbf{g}'_v]$ :

155 
$$\mathbf{Z}^a = \mathbf{Z}^g\mathbf{K}' .$$

156 This transformation retains all columns in the partition  $\mathbf{Z}^v$ , i.e., one per gametic effect of individuals  
157 with records, while the number of null columns in  $\mathbf{0}^u$  of  $\mathbf{Z}^a$  collapses to half of that of  $\mathbf{0}^{2u}$  in  $\mathbf{Z}^g$ . In  
158 the same manner, both incidence matrices  $\mathbf{Z}_s$  and  $\mathbf{Z}_d$  from the previous model equation are  
159 converted versions of their counterparts in the classical gametic imprinting model, which forms the  
160 basis for the proof of equivalence of the classical and the generalized gametic models involving  $\bar{\mathbf{G}}$   
161 (see Appendix A1).

162

### 163 **Reduced gametic model**

164 The reduced imprinting model as initially described by Neugebauer et al. (2010a, b) relates each  
165 observation from the final progeny  $i$  to the transmitting abilities as sire  $a_{si}^s$  and as dam  $a_{di}^d$  of the  
166 parents  $si$  (sire of  $i$ ) and  $di$  (dam of  $i$ ), respectively. For a single observation  $y_i$  we have the  
167 observation equation

168 
$$y_i = \mu_i + a_{si}^s + a_{di}^d + r_i . \tag{1}$$

169 Here, the residual  $r_i$  is a sum of the Mendelian sampling effects of both parents ( $m_{si}$  and  $m_{di}$ ) and  
170 the measurement noise ( $e_i$ ). The latter is identical to the residual of the gametic model. Thus,

171 
$$r_i = m_{si} + m_{di} + e_i .$$

172 Its variance is a function of the inbreeding coefficients  $F_{si}$  and  $F_{di}$  of the parents of  $i$ :

173 
$$\text{var}(r_i) = \frac{1}{2}(1 - F_{si})\sigma_s^2 + \frac{1}{2}(1 - F_{di})\sigma_d^2 + \sigma_e^2.$$

174 By rewriting the transmitting abilities of the parents as the averages of the respective gametic  
175 effects, i.e.,  $a_{si}^s = \frac{1}{2}(g_{1,si}^s + g_{2,si}^s)$  and  $a_{di}^d = \frac{1}{2}(g_{1,di}^d + g_{2,di}^d)$ , we get an observation equation in terms  
176 of gametic effects:

177 
$$y_i = \mu_i + \frac{1}{2}(g_{1,si}^s + g_{2,si}^s) + \frac{1}{2}(g_{1,di}^d + g_{2,di}^d) + r_i. \quad (2)$$

178 The covariance of the gametic effects then is

179 
$$\text{Var} \begin{bmatrix} \mathbf{g}_s \\ \mathbf{g}_d \end{bmatrix} = \begin{bmatrix} \sigma_s^2 & \sigma_{sd} \\ \sigma_{sd} & \sigma_d^2 \end{bmatrix} \otimes \mathbf{G} = \begin{bmatrix} \mathbf{G}\sigma_s^2 & \mathbf{G}\sigma_{sd} \\ \mathbf{G}\sigma_{sd} & \mathbf{G}\sigma_d^2 \end{bmatrix}.$$

180 Here, the relationship matrix  $\mathbf{G}$  of the gametic effects that define the involved transmitting abilities  
181 includes only the parents and their ancestors. The advantage of this gametic version of the reduced  
182 imprinting model over the previously published version that uses only transmitting abilities and their  
183 relationship matrix  $\frac{1}{2}\mathbf{A}$  is that it enables us to easily integrate observations from parents by linking  
184 them to the respective gametic effects. Hence, for observations of any parent  $i$ , the observation  
185 equation becomes

186 
$$y_i = \mu_i + g_{1,i}^s + g_{2,i}^d + e_i. \quad (3)$$

### 187 Generalized reduced gametic model

188 The drawback of the reduced gametic model is that it has twice the number of equations compared  
189 to a version that uses  $\frac{1}{2}\mathbf{A}$ . For all individuals without own records, it is however possible to reduce  
190 the number of equations for random genetic effects by representing the individuals through their  
191 transmitting abilities (average gametic effects) while retaining separate gametic effects for all  
192 parents with records, i.e., vectors of gametic effects  $\mathbf{g}_s$  and  $\mathbf{g}_d$  are replaced by appropriately  
193 transformed counterparts  $\mathbf{a}_s$  and  $\mathbf{a}_d$ , respectively. Consequently the covariances of random genetic  
194 effects in a parsimonious generalized reduced gametic model that allows for parents with records is

195 
$$\text{Var} \begin{bmatrix} \mathbf{a}_s \\ \mathbf{a}_d \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{G}}\sigma_s^2 & \bar{\mathbf{G}}\sigma_{sd} \\ \bar{\mathbf{G}}\sigma_{sd} & \bar{\mathbf{G}}\sigma_d^2 \end{bmatrix}.$$

196 Further, we need a diagonal matrix  $\mathbf{W}$  of weights equal to  $w_i = 1$  for observations from parents, for  
197 which model Equation (3) applies and

198 
$$w_i = \left[ \frac{\frac{1}{2}(1 - F_{si})\sigma_s^2 + \frac{1}{2}(1 - F_{di})\sigma_d^2 + \sigma_e^2}{\sigma_e^2} \right]^{-1}$$

199 for the final progeny, where parents without their own records are represented by transmitting  
 200 abilities or both parents have a record and are represented by gametic effects (the respective  
 201 observation equations are (1) and (2)). The same weight applies to mixed kinds of representation  
 202 that arise from cases where one parent of a final progeny has a record while the other does not. The  
 203 corresponding observation equations for observations  $y_i$  of such final progeny are

204

205 
$$y_i = \mu_i + a_{si}^s + \frac{1}{2}(g_{1,di}^d + g_{2,di}^d) + r_i \quad (4)$$

206 and

207 
$$y_i = \mu_i + \frac{1}{2}(g_{1,si}^s + g_{2,si}^s) + a_{di}^d + r_i. \quad (5)$$

208

209 **A general model for parent-of-origin analyses**

210 A general comprehensive model for parent-of-origin analyses banks on the generalized gametic  
 211 relationship matrix. Special cases of the generalized gametic relationship matrix  $\bar{\mathbf{G}}$  are the classical  
 212 gametic relationship matrix  $\bar{\mathbf{G}} = \mathbf{G}$  in the gametic model and  $\bar{\mathbf{G}} = \frac{1}{2}\mathbf{A}$  as in the reduced imprinting  
 213 model. Correspondingly, the matrix  $\mathbf{W}$  of weights can be an identity matrix that fits the classical  
 214 gametic model, or a matrix all the weights of which are different from one as those in the reduced  
 215 model for records of the final progeny. A general model can be specified for parent-of-origin analyses  
 216 containing these two basic kinds of comprehensive imprinting models as well as models with any  
 217 combination of gametic effects and transmitting abilities that can be obtained using our  
 218 transformation matrix  $\mathbf{K}'$ . In matrix notation, the general model is

219 
$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_s\mathbf{a}_s + \mathbf{Z}_d\mathbf{a}_d + \boldsymbol{\varepsilon},$$

220 where  $\boldsymbol{\varepsilon}$  is a vector of residuals. That is,  $\varepsilon_i = e_i$  for records from individuals represented by two  
 221 gametic effects, or  $\varepsilon_i = r_i$  for observations from final progeny linked to the genetic effects of their  
 222 parents. The respective weights are

223 
$$w_i = 1$$

224 and

$$225 \quad w_i = \left[ \frac{\frac{1}{2}(1-F_{si})\sigma_s^2 + \frac{1}{2}(1-F_{di})\sigma_d^2 + \sigma_e^2}{\sigma_e^2} \right]^{-1}$$

226 Random genetic effects and residuals are assumed to have covariance:

$$227 \quad \text{Var} \begin{bmatrix} \mathbf{a}_s \\ \mathbf{a}_d \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{G}}\sigma_s^2 & \bar{\mathbf{G}}\sigma_{sd} & \mathbf{0} \\ \bar{\mathbf{G}}\sigma_{sd} & \bar{\mathbf{G}}\sigma_d^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{W}\sigma_e^2 \end{bmatrix}.$$

228 The resulting mixed model equations are

$$229 \quad \begin{bmatrix} \mathbf{X}'\mathbf{W}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{W}^{-1}\mathbf{Z}_s & \mathbf{X}'\mathbf{W}^{-1}\mathbf{Z}_d \\ \mathbf{Z}'_s\mathbf{W}^{-1}\mathbf{X} & \mathbf{Z}'_s\mathbf{W}^{-1}\mathbf{Z}_s + \bar{\mathbf{G}}^{-1}\alpha_1 & \mathbf{Z}'_s\mathbf{W}^{-1}\mathbf{Z}_d + \bar{\mathbf{G}}^{-1}\alpha_2 \\ \mathbf{Z}'_d\mathbf{W}^{-1}\mathbf{X} & \mathbf{Z}'_d\mathbf{W}^{-1}\mathbf{Z}_s + \bar{\mathbf{G}}^{-1}\alpha_2 & \mathbf{Z}'_d\mathbf{W}^{-1}\mathbf{Z}_d + \bar{\mathbf{G}}^{-1}\alpha_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{a}_s \\ \mathbf{a}_d \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{W}^{-1}\mathbf{y} \\ \mathbf{Z}'_s\mathbf{W}^{-1}\mathbf{y} \\ \mathbf{Z}'_d\mathbf{W}^{-1}\mathbf{y} \end{bmatrix}.$$

230 with

$$231 \quad \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{bmatrix} = \sigma_e^2 \begin{bmatrix} \sigma_s^2 & \sigma_{sd} \\ \sigma_{sd} & \sigma_d^2 \end{bmatrix}^{-1}.$$

232 The general model comprehends any combination of observation Equations (1) to (5) to provide a  
 233 large degree of flexibility in parent-of-origin analyses. Model variants may be chosen to minimize the  
 234 number of equations for random genetic effect by using as many reduced observation equations as  
 235 possible, which comes at the expense of the need for the recomputation of weights when estimating  
 236 the components of variance. Alternatively, the repeated recomputation of weights may be avoided  
 237 by representing all individuals with an observation using gametic effects. The underlying reason for  
 238 this flexibility is that for the given data (observations, fixed effects, and pedigree), each possible  
 239 general imprinting model has as an equivalent the same classical gametic model (that follows from  
 240 Appendices A2 and A3). Consequently, any two general models that share the same equivalent  
 241 classical model are also equivalent, and can replace each other, especially for the sake of estimating  
 242 the components of variance.

243

244 **Figure 1: about here**

245

246 **Direct inversion of the generalized gametic relationship matrix**

247 Setting-up the inverse generalized gametic relationship matrix is key to any large-scale application.  
248 Rules for direct inversion can be derived by factoring the inverse  $\bar{\mathbf{G}}^{-1}$  into inverses of a matrix  $\mathbf{T}'$   
249 and a diagonal matrix  $\mathbf{D}$  of inverse Mendelian sampling variances

$$250 \quad \bar{\mathbf{G}}^{-1} = (\mathbf{T}')^{-1} \mathbf{D}^{-1} \mathbf{T}'^{-1} .$$

251 The above is known from the direct inversion of the numerator relationship matrix (Henderson,  
252 1976; Quaas, 1976) and the classical gametic relationship matrix (Schaeffer et al., 1989). The matrix  
253  $(\mathbf{T}')^{-1}$  is lower triangular, as shown in Figure 1. The underlying pedigree of this example  
254 (Supplement) comprises 12 individuals. Nine of them are represented by a single transmitting ability  
255 while the remaining three by two gametic effects. The kind of representation is indicated by the  
256 respective values of one and two in the last column of the pedigree file. Consequently, the  
257 dimensions of the inverse of the example are 15×15. Each of the 15 rows of  $(\mathbf{T}')^{-1}$  pertain to a single  
258 genetic effect, which itself may be derived from different kinds of genetic parental predecessor  
259 effects: An individual's transmitting ability may be derived from two unknown parents (a-00) or a  
260 single unknown parent, where the known parent may be represented by a transmitting ability (a-0a,  
261 a-a0) or two gametic effects (a-0gg, a-gg0). Two known parents may show up as any combination of  
262 transmitting abilities or gametic effects (a-aa, a-agg, a-gga, a-gggg). Likewise, a gametic effect may be  
263 derived from an unknown parent (g-0), or a known parent enrolled by either a transmitting ability or  
264 two gametic effects (g-a, g-gg). These 12 cases need to be distinguished for directly inverting the  
265 generalized gametic relationship matrix. The example pedigree was constructed such that each case  
266 appeared at least once. For each effect related to a particular row of the lower-triangular matrix in  
267 Figure 1, the case is indicated in the last column. Note that the six cases a-0gg, a-gg0, a-agg, a-gga, a-  
268 gggg, and g-a are specific to the generalized gametic relationship matrix as they appear neither in the  
269 direct inversion of the numerator relationship matrix—involving only a-00, a-0a, a-a0, and a-aa—nor  
270 the classical gametic relationship matrix, for which only g-0 and g-gg need to be distinguished.

271 Mendelian sampling variances that define the diagonal elements of  $\mathbf{D}$  are different for transmitting  
272 abilities and gametic effects. Further, they depend on the occurrence of unknown parents and the  
273 inbreeding coefficients of the known ones. In particular, this is  $F_{\text{known parent}}$  when an individual with a  
274 transmitting ability in the matrix has only one known parent, or  $F_{\text{sire}}$  and  $F_{\text{dam}}$  in case of full  
275 parentage information. For gametes, we need to account for the inbreeding coefficient  $F_{\text{parent}}$  of the  
276 known parent from which a gamete is derived. Accordingly, the 12 cases (a-00, a-0a, ... , g-00) are  
277 grouped into five classes with distinct formulae for the inverse Mendelian sampling variance  $\delta$  :

278	a-00	$\delta = 2$
279	a-0a, a-a0, a-0gg, a-gg0	$\delta = \left[ \frac{1}{2} (1 - F_{\text{known parent}}) \right]^{-1}$
280	a-aa, a-agg, a-gga, a-gggg	$\delta = 2 \left( \frac{1}{2} \left[ \frac{1}{2} (1 - F_{\text{sire}}) + \frac{1}{2} (1 - F_{\text{dam}}) \right] \right)^{-1}$
281	g-a, g-gg	$\delta = 2 \left[ \frac{1}{2} (1 - F_{\text{parent}}) \right]^{-1}$
282	g-0	$\delta = 1$

283 For any arbitrary order of genetic effects, the inverse generalized gametic relationship matrix can be  
 284 constructed step by step from the pedigree. In each step, a matrix contribution  $\mathbf{U}_i$  is added for  
 285 genetic effect  $i$  to a matrix composed of an inverse  $\bar{\mathbf{G}}_{i-1}^{-1}$  that already covers the preceding  $i-1$   
 286 effects and zeroes:

$$287 \quad \bar{\mathbf{G}}_i^{-1} = \begin{bmatrix} \bar{\mathbf{G}}_{i-1}^{-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} + \mathbf{U}_i,$$

288 where  $\mathbf{0}$  is a column vector of  $i-1$  zeroes and

$$289 \quad \mathbf{U}_i = \mathbf{u}_i \mathbf{u}_i' \delta_i$$

290 is the contribution made for each genetic effect  $i$ . The row-vector  $\mathbf{u}_i'$  consists of all zeros, except for  
 291 those elements with indices indicating the genetic effects of the respective parent(s). At minimum,  
 292 the  $i$ -th element is always equal to unity as a non-zero element in this vector. All other non-zero  
 293 elements are negative, with values of either  $-\frac{1}{2}$  or  $-\frac{1}{4}$ . Thus, the number of non-zero entries varies  
 294 from one to five, as can be derived from the rows of the example triangular matrix  $(\mathbf{T}')^{-1}$  in Figure 1.  
 295 For all 12 possible cases, the non-zero coefficients in  $\mathbf{u}_i'$  and their indices are summarized in Table 1.  
 296 The non-zero elements of the resulting matrix  $\mathbf{U}_i = \mathbf{u}_i \mathbf{u}_i' \delta_i$  correspond to the (scaled) cross-products  
 297 of the elements of the non-zero vector, and their coordinates in the matrix are the respective  
 298 combinations of indices.

299 **Table 1: about here**

300

### 301 **Transforming measured genotypes in a generalized genomic gametic relationship matrix**

302 Parent-of-origin analyses may also use genomic relationships, or combined genomic and pedigree  
 303 relationships. A specific feature of this is that ordinary marker genotypes (AA, AB, BB) are not  
 304 sufficient for this purpose, and the parental origin of the marker alleles at each locus has to be

305 inferred instead (Lawson et al., 2013, and references therein) and summarized as ordered genotypes  
 306 AA, AB, BA, and BB, where the first allele is paternally derived. This is, however, not always possible  
 307 for all members of a genealogy. In such a case, the principles above are beneficial for integrating  
 308 ordered and unordered genomic information into a single genomic version of the generalized  
 309 gametic relationship matrix.

310 Let us assume that all  $t$  individuals are genotyped with  $p$  markers and all genotypes are phased into  
 311  $2t$  haplotypes. Information on the number (zero or one at each locus before centering) of minor  
 312 alleles for all marker loci on each haplotype can be summarized in a column-wise mean-centered  
 313  $2t \times p$  matrix  $\mathbf{C}$ . To this matrix, each individual  $i$  contributes two p-row-vectors  $\mathbf{c}'_{i1}$  and  $\mathbf{c}'_{i2}$ , where  
 314 the centered allele accounts for its first and second haplotype. Matrix  $\mathbf{C}$  can then be split into two  
 315 submatrices  $\mathbf{C}_v$  and  $\mathbf{C}_u$  :

$$316 \quad \mathbf{C} = \begin{bmatrix} \mathbf{C}_v \\ \mathbf{C}_u \end{bmatrix}.$$

317 For imprinting analyses, at least all  $u$  individuals with records need to have their paternal and  
 318 maternal haplotypes identified in  $\mathbf{C}_u$ . This can be achieved by adding at least one preceding  
 319 generation without records but with genotypes. In case of only a single generation, all their  $2v$   
 320 haplotypes in partition  $\mathbf{C}_v$  would be left unordered. If the additional  $v$  genotyped individuals contain  
 321 more than a single successive generation, only a part of their genotypes may qualify as ordered, with  
 322 the exceptions coming from the founders.

323 From  $\mathbf{C}$ , a genomic gametic relationship matrix can be derived:

$$324 \quad \mathbf{G}_g = \frac{\mathbf{C}\mathbf{C}'}{s} = \begin{bmatrix} \mathbf{C}_u\mathbf{C}'_u & \mathbf{C}_u\mathbf{C}'_v \\ \mathbf{C}_v\mathbf{C}'_u & \mathbf{C}_v\mathbf{C}'_v \end{bmatrix} \frac{1}{s} = \begin{bmatrix} \mathbf{G}_{guu} & \mathbf{G}_{guv} \\ \mathbf{G}_{gvu} & \mathbf{G}_{gvv} \end{bmatrix},$$

325 ,

326 where  $s$  is a scaling factor,  $s = \sum p_j(1 - p_j)$ , and  $p_j$  is the frequency of the allele at marker  $j$ .

327 In all cases where the parental origin of the two haplotypes can be traced back, the first haplotype of  
 328 each individual is assumed to be paternal and the second maternal ( $\mathbf{c}'_{i1} = \mathbf{c}'_{ip}$  and  $\mathbf{c}'_{i2} = \mathbf{c}'_{im}$ );  
 329 otherwise, the ordering of haplotypes is arbitrary. This is where the concept of generalization from  
 330 above is used. A transformation matrix  $\mathbf{K}'$  can be defined such that for all individuals  $i$  with  
 331 unordered genomic information, the two row vectors  $\mathbf{c}'_{i1}$  and  $\mathbf{c}'_{i2}$  are replaced by their averages:

332 
$$\bar{\mathbf{c}}_i = \frac{1}{2}(c_{i1} + c_{i2}).$$

333  $\bar{\mathbf{c}}_i$  does not depend on the order or the parental origin of the haplotypes of an individual:

334 
$$\bar{\mathbf{c}}_i = \frac{1}{2}(c_{i1} + c_{i2}) = \frac{1}{2}(c_{ip} + c_{im}).$$

335 That is,  $\bar{\mathbf{c}}_i$  is also the vector of average paternal- and maternal-centered number of gene counts.

336 Consequently, a generalized genomic gametic relationship matrix can be defined as

337 
$$\bar{\mathbf{G}}_g = \mathbf{K}'\mathbf{C}\mathbf{C}'\mathbf{K} \frac{1}{s} = \mathbf{K}'\mathbf{G}_g \mathbf{K} ,$$

338 with  $\mathbf{K}'$  defined as before. The partition  $\mathbf{C}_u\mathbf{C}'_u$  of  $\mathbf{G}_g$  can be used to determine only the ordered  
339 genomic information of all individuals with records and, as such, is sufficient to estimate the  
340 components of genetic variance in a parent-of-origin analysis. All respective gametic effects of these  
341 individuals can also be estimated. The entire matrix  $\mathbf{G}_g$  delivers gametic effects (as sire and dam) for  
342 all individuals, including those with no phenotypes. The generalized variant  $\bar{\mathbf{G}}_g$  by design is also  
343 appropriate for parent-of-origin analyses, with no other requirements for  $\mathbf{K}'$  as for the pedigree-  
344 derived counterpart. Thus, the general model for parent-of-origin analyses is also applicable to  
345 genomic relationships, provided the marker haplotypes of individuals with observations can be  
346 ordered.

347

#### 348 **Software and data availability**

349 A detailed *guide to practical implementation* is available on the RADAR repository  
350 ([https://www.radarservice.eu/radar/dataset/get/IGjshsdpCzWftGAQ?lang=en&token=DpsQIXcXRuD](https://www.radarservice.eu/radar/dataset/get/IGjshsdpCzWftGAQ?lang=en&token=DpsQIXcXRuDkLmbwmzB)  
351 [kLmbwmzB](https://www.radarservice.eu/radar/dataset/get/IGjshsdpCzWftGAQ?lang=en&token=DpsQIXcXRuDkLmbwmzB) – this is a temporary link for the purpose of review only and will later be replaced by a  
352 permanent DOI). It includes the source code of a program to directly set-up the inverse of the  
353 generalized gametic relationship matrix from a pedigree file, a detailed program description and  
354 example input and output files. There we also provide a collection of six worked toy examples  
355 demonstrating in very detail how various mixed models with generalized gametic relationships can  
356 be implemented using the ASReml package. Each example is also accompanied with R-code to check  
357 details and the correctness of the ASReml results.

358

## 359 DISCUSSION

360 The outlined generalization introduces elements of the reduced imprinting model to the gametic  
361 model and vice versa, accompanied by gains in flexibility and substantial savings in terms of the  
362 number of equations used. The latter is important especially for estimating the components of  
363 variance (Shor et al., 2019). The matrix  $\bar{\mathbf{G}}$  contains two limiting cases that set the boundaries for the  
364 ratio of equations that can be eliminated. The first is the classical gametic relationship matrix itself  
365 (dimensions  $2t \times 2t$ ), when  $\mathbf{K}'$  is an identity matrix. The other limiting case is  $\mathbf{K}' = \mathbf{I} \otimes \begin{bmatrix} \frac{1}{2} & \\ & \frac{1}{2} \end{bmatrix}$ , such  
366 that  $\bar{\mathbf{G}} = \frac{1}{2} \mathbf{A}$  with dimensions  $t \times t$ . Therefore, the reduction in the number of equations for  
367 genetic effects can take a range of 0%–50%, compared with a classical gametic model. However, the  
368 actual savings depend on the specifics of each dataset. As examples, two animal datasets were  
369 considered: the first was from an analysis of daily net gain in Brown Swiss fattening bulls (Blunk et al.,  
370 2018; Blunk et al., 2019), with a pedigree of 663,515 individuals (173,051 non-parents with records),  
371 whereas the second dealt with litter size in an experimental line of mice (2,137 females with an  
372 observation for first-parity litter size; necessary pedigree size for variance component estimation:  
373 4544; total pedigree size 15222; unpublished data). In the Brown Swiss, the number of gametic  
374 effects for all animals was 1,327,030, compared with 836,566 with gametic effects for animals with  
375 observations only. The relative saving was 37% in terms of the number of equations and 32% in  
376 terms of the number of non-zero elements of the half-stored inverse. The respective numbers of  
377 equations in the mice example were 9088 versus 6681, with relative savings of 26% and 25%. In  
378 particular, small proportions of individuals that have records cause large reductions as all ancestors  
379 without a record are assigned only one equation. This applied to the mice example, as only females  
380 that had reproduced had records of litter size. If all available animals from the same number of  
381 generations were included (no “pruning” performed; 15222 animals), as one would prefer e.g. for the  
382 estimation of the genetic trend, there were 30,444 gametic effects versus 17,359 effects with the  
383 generalized relationships, with relative savings of 42% of equations and 56% of non-zero elements.  
384 The vast majority of this pedigree included males, females from older generations with no data, and  
385 non-reproducing females of younger generations.

386 Sex-specific traits such as litter size, number of eggs, or milk yield provide the opportunity to  
387 represent all males by their transmitting abilities. Thus, the resulting number of equations is  
388 considerably smaller in comparison with a trait recorded in both sexes. The family structure also has  
389 an effect: More equations are saved in the presence of typically small paternal groups of offspring,  
390 given that sires without own phenotypes are represented by their transmitting abilities. Further, it  
391 makes sense in imprinting analyses to add a high ratio of ancestors without phenotypes to better  
392 reflect inbreeding, and the relationships between genetic effects as sire and dam. Including their

393 transmitting abilities rather than gametic effects in the model therefore also leads to a large number  
394 of saved equations.

395 A certain fraction of individuals with records might have either not reproduced at all or not yet  
396 reproduced at the time of data recall (i.e. they appear as final progeny), which provides the  
397 opportunity for representing them by reduced observational equations rather than having their own  
398 gametic effects in the model. In the Brown Swiss example, where all observations were from final  
399 progeny, this leads to a fully reduced model with relationships of 490,464 ancestors, a reduction of  
400 63 %. In the small mouse example more equations for 634 final progeny can be saved (3910 animals  
401 and 5413 equations left), forcing the relative savings up to 40% of equations.

402 In certain cases, one could, however, abstain from reduced observational equations, which has the  
403 advantage that no weights are required that depend on as-yet undetermined components of  
404 variance. That has not proven to be a particular problem in the REML estimation of the components  
405 of variance (Neugebauer et al., 2010a, b; Blunk et al., 2017a,b), but may be beneficial to avoid in  
406 Bayesian approaches that employ Markov chain Monte Carlo methods, where the values of the  
407 components of variance change from iteration to iteration. By capitalizing on the flexibility of the  
408 generalized approach, weights become obsolete by representing all individuals with records—be  
409 they final progeny or not—by two gametic effects, which helps offset the computational burden  
410 resulting from repeated reweighting. At the same time, individuals without observations can be  
411 integrated by single equations.

412 For reasons of principle, a maternal genetic component of variance provides a special challenge as it  
413 is difficult to separate from the imprinting variance. Okamoto et al. (2019) showed that when  
414 estimated with a model variant that uses information only on the sire and maternal grandsire (Blunk  
415 et al., 2017; Okamoto et al., 2019), the imprinting variance may also be interpreted as maternal  
416 genetic. Similarly, for the reduced imprinting model, it can be shown that the imprinting variance and  
417 maternal genetic variance cannot be disentangled when both are present, and instead only a  
418 composite component of variance can be inferred (Appendix A4). A way out of this is to avoid  
419 reduced model equations and, instead, to represent individuals with records explicitly by their  
420 gametic effects in a model that includes maternal genetic effects. Then, gametic variances as sire and  
421 dam can, at least in principle, be separated from the maternal genetic variance (Appendix A5). In  
422 practice, however, this may be hampered by limitations in the amount and structure of the data, as  
423 has been reported for Mendelian models (Heydarpour et al, 2008). Like maternal effects models,  
424 other kinds of imprinting models may also comprise more than a single genetic effect as sire and dam  
425 per individual—e.g., random regression models or multitrait models. As they all suffer from a large  
426 number of gametic equations, they benefit even more from generalized relationships.

427 In applications where all  $\nu$  individuals with records plus at least one preceding generation have  
428 measured genotypes and variance components are to be estimated, it is sufficient to include only the  
429 subset of these  $\nu$  individuals with their genomic covariance  $\mathbf{G}_{g\nu\nu}$ . If there is interest in the genetic  
430 effects of the  $u$  founders as sire and dam, either  $\mathbf{G}_g$  or  $\bar{\mathbf{G}}_g$  is the choice. An example is an  $F_2$  line-  
431 cross experiment with phenotypes recorded only in the  $F_2$  generation, and the genotypes of  $F_1$  and  $P_0$   
432 generations needed only for phasing and determining line origins of the markers.

433 Often in animal breeding, large pedigrees are combined with smaller cohorts of genotyped  
434 individuals. Then, certain individuals are the first in their genealogy to be genotyped while the  
435 pedigree can be traced further back. In contrast to their own descendants, haplotypes of such a  
436 candidate cannot be ordered, which renders uncertain whether the first of two unordered marker  
437 haplotypes matches the paternal gametic effect in a pedigree-derived gametic relationship matrix or  
438 the maternal one. Consequently, a combined relationship matrix that is suitable for parent-of-origin  
439 analyses cannot be constructed. This problem can be solved by collapsing gametic effects into  
440 transmitting abilities both in the genomic relationships and the pedigree-derived ones. Then,  
441 generalized pedigree relationships for all animals can be combined with their matching generalized  
442 genomic counterparts  $\bar{\mathbf{G}}_g$  for the genotyped cohort in a way that allows for the easy integration of  
443 unordered genomic information. To this end, the available theory (Legarra *et al.* 2009; Christensen  
444 and Lund 2010; Aguilar *et al.* 2010) can be used to combine pedigree-derived relationships (here,  $\bar{\mathbf{G}}$ )  
445 and genomic relationships ( $\bar{\mathbf{G}}_g$ ) into a joint matrix, at least in the many cases where candidates with  
446 unordered genotypes have no record, such as dairy bulls.

447 In conclusion the generalized gametic relationship matrix provides the necessary flexibility to adapt  
448 imprinting analyses to specific computational and analytical needs in a large variety of situations  
449 through tailored versions of the general imprinting model. The most important aspects are the  
450 effective estimation of the imprinting variance in REML and Bayesian approaches in case the parents  
451 have records and the inclusion of maternal genetic effects and genomic relationships that integrate  
452 ordered and unordered genomic information. All things considered, these new possibilities are  
453 expected to stimulate systematic research on the importance of parent-of-origin effects for the  
454 genetic variation of quantitative traits in farm animals and other species.

455

456

457

458 **Acknowledgements**

459 Inga Blunk was was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research  
460 Foundation) project 418890112 “Making unused data resources available for imprinting  
461 analyses by using new methods to uncover parent-of-origin effects in human and livestock”.

462

463 **References**

464 Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta et al., 2010: Hot topic: A unified approach to  
465 utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final  
466 score. *J. Dairy Sci.* 93: 743–752.

467 Blunk, I., und N. Reinsch, 2014: Genetic variance components when fluctuating imprinting patterns  
468 are present. Proceedings of the 10<sup>th</sup> World Congress Applied to Livestock Genetics, 17-22. August  
469 2014, Vancouver, BC, Canada, pp. 17-22

470 Blunk, I., M. Mayer, H. Hamann, and N. Reinsch, 2017 a: Parsimonious model for analyzing parent-of-  
471 origin effects related to beef traits in dual-purpose Simmental. *J. Anim. Sci.* **95**, 559-571

472 Blunk, I., Mayer, M., Hamann, H. and Reinsch, N., 2017 b: A new model for parent-of-origin effect  
473 analyses applied to Brown Swiss cattle slaughterhouse data. *Animal* **11**, 1096-1106.

474

475 Blunk, I., M. Mayer, H. Hamann, and N.Reinsch, 2019: Scanning the genomes of parents for imprinted  
476 loci acting in their un-genotyped progeny. *Sci. Rep.* **9**, 654

477 Christensen, O. F., and M. S. Lund, 2010: Genomic prediction when some animals are not genotyped.  
478 *Genet. Sel. Evol.* 42: 2.

479 De Vries, A. G., R. Kerr, B. Tier, T. Long, and T. H. E. Meuwissen, 1994: Gametic imprinting effects on  
480 rate and composition of piggrowth. *Theor. Appl. Genet.* 88:1037–1042.

481 Gibson, J.P., B.W. Kennedy, L.R. Schaeffer, and O.I. Southwood, 1998: Gametic models for  
482 estimation of autosomally inherited genetic effects that are expressed only when received from  
483 either a male or female parent. *J. Dairy Sci.* **71**, 143 (Abstr.)

484 Henderson, C.R. 1985: Equivalent Linear Models to Reduce Computations. *J. Dairy Sci.* 68(9): 2267–  
485 2277

486 Henderson, C.R, 1976: A simple method for computing the inverse of a numerator relationship matrix  
487 used in prediction of breeding values. *Biometrics* **32**, 69-83

488 Heydarpour, M., L.R. Schaeffer, and M.H. Yazdi, 2008: Influence of population structure on estimates  
489 of direct and maternal parameters. *J. Anim. Breed. Genet.* 125: 89–99

490 Hill, W.G., and P.D. Keightly, 1988: Interaction between molecular and quantitative genetics. In  
491 *Proceedings of the World Symposium in Honour of Professor R.D. Politiek, Wageningen* (eds Korver, S.  
492 et al.) 45-55 (Advances in Animal Breeding, Wageningen, The Netherlands, 1988).

493  
494 Lawson, H. A., J. M. Cheverud, and J. B. Wolf, 2013: Genomic imprinting and parent-of-origin effects  
495 on complex traits. *Nat. Rev. Genet.* 14: 609-617  
496  
497 Legarra, A., I. Aguilar, and I. Misztal, 2009: A relationship matrix including full pedigree and genomic  
498 information. *J. Dairy Sci.* 92: 4656–4663.  
499  
500 Meyer, K., and B. Tier (2012): Estimates of variances due to parent of origin effects for weights of  
501 Australian beef cattle. *Animal Production Science* 52(4) 215-224  
502  
503 Neugebauer, N., I. Räder, H.J. Schild, D. Zimmer, and N. Reinsch, 2010 a: Evidence for parent-of-origin  
504 effects on genetic variability of beef traits. *J. Anim. Sci.* **88**, 523-532.  
505  
506 Neugebauer, N., Luther, H. & Reinsch, N., 2010 b: Parent-of-origin effects cause genetic variation in  
507 pig performance traits. *Animal* **4**, 672-681  
508  
509 Nishio, M., and M. Satoh, 2015: Genomic linear best unbiased prediction method including  
510 imprinting effects for geneomic evaluation. *Genet. Sel. Evol.* 47:32  
511  
512 Okamoto, K., Oishi, K., Nakamura, L., Abe, A., Inoue, K., Kumagi, H., and H. Hirooka, 2019: Parent-of-  
513 origin effects on carcass traits in Japanese Black cattle. *J. Anim. Breed. Genet.* 136(3): 190-198  
514  
515 Quaas, R.L., 1976: Computing the diagonal elements and inverse of a large numerator relationship  
516 matrix. *Biometrics* **32**, 949-953  
517  
518 Schaeffer, L.R., B.W. Kennedy, and J.P. Gibson, 1989: The inverse of the gametic relationship matrix. *J.*  
*Dairy Sci.* **72**, 1266-1272  
519  
520 Shor, T., Kalka, I., Geiger, D., Erlich, Y., and O. Weissbrod, 2019: Estimating variance components in  
population scale family trees. *PLoS Genet* 15(5): e1008124  
521  
522 Smith S-P. and F.R. Allaire, 1985: Efficient selection rules to increase non-linear merit: application to  
mate selection. *Genet. Sel. Evol.* 17(3): 387-406  
523  
524 Tier, B., and K.Meyer, 2012: Analysing quantitative parent-of-origin effects with examples from  
525 ultrasonic measures of body composition In Australian beef cattle. *J. Anim. Breed. Genet.* **129**, 359-  
368  
526  
527

528 **APPENDIX**

529 **Appendix A1: Equivalence of the classical gametic model and the generalized gametic model in**  
 530 **which all individuals with records have two gametic effects.**

531 Both models have the same expectation  $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$  of the vector of observations  $\mathbf{y}$ .

532 The variance of observations in the classical gametic model is  $Var(\mathbf{y}) = \mathbf{Q}_c + \mathbf{I}\sigma_e^2$ ,

533 where

$$534 \quad \mathbf{Q}_c = [\mathbf{Z}_s \quad \mathbf{Z}_d] \begin{bmatrix} \mathbf{G}\sigma_s^2 & \mathbf{G}\sigma_{sd} \\ \mathbf{G}\sigma_{sd} & \mathbf{G}\sigma_d^2 \end{bmatrix} \begin{bmatrix} \mathbf{Z}'_s \\ \mathbf{Z}'_d \end{bmatrix} =$$

$$535 \quad = \mathbf{Z}_s \mathbf{G} \mathbf{Z}'_s \sigma_s^2 + \mathbf{Z}_s \mathbf{G} \mathbf{Z}'_d \sigma_{sd} + \mathbf{Z}_d \mathbf{G} \mathbf{Z}'_s \sigma_{sd} + \mathbf{Z}_d \mathbf{G} \mathbf{Z}'_d \sigma_d^2.$$

536 The first term can be rewritten as

$$537 \quad \mathbf{Z}_s \mathbf{G} \mathbf{Z}'_s = \begin{bmatrix} \mathbf{0}^{2u} & \mathbf{Z}_s^v \end{bmatrix} \begin{bmatrix} \mathbf{G}_{uu} & \mathbf{G}_{uv} \\ \mathbf{G}_{uv}' & \mathbf{G}_{vv} \end{bmatrix} \begin{bmatrix} (\mathbf{0}^{2u})' \\ (\mathbf{Z}_s^v)' \end{bmatrix} = \mathbf{Z}_s^v \mathbf{G}_{vv} \mathbf{Z}_s^{v'}.$$

538 Likewise,

$$539 \quad \mathbf{Z}_s \mathbf{G} \mathbf{Z}'_d = \mathbf{Z}_s^v \mathbf{G}_{vv} \mathbf{Z}_d^{v'}, \quad \mathbf{Z}_d \mathbf{G} \mathbf{Z}'_s = \mathbf{Z}_d^v \mathbf{G}_{vv} \mathbf{Z}_s^{v'} \quad \text{and} \quad \mathbf{Z}_d \mathbf{G} \mathbf{Z}'_d = \mathbf{Z}_d^v \mathbf{G}_{vv} \mathbf{Z}_d^{v'}.$$

540 Finally,

$$541 \quad \mathbf{Q}_c = \mathbf{Z}_s^v \mathbf{G}_{vv} \mathbf{Z}_s^{v'} \sigma_s^2 + \mathbf{Z}_s^v \mathbf{G}_{vv} \mathbf{Z}_d^{v'} \sigma_{sd} + \mathbf{Z}_d^v \mathbf{G}_{vv} \mathbf{Z}_s^{v'} \sigma_{sd} + \mathbf{Z}_d^v \mathbf{G}_{vv} \mathbf{Z}_d^{v'} \sigma_d^2.$$

542 In the generalized case, the variance of observations is

$$543 \quad Var(\mathbf{y}) = \mathbf{Q}_g + \mathbf{I}\sigma_e^2$$

544 with

$$545 \quad \mathbf{Q}_g = [\mathbf{Z}_s \mathbf{K} \quad \mathbf{Z}_d \mathbf{K}] \begin{bmatrix} \bar{\mathbf{G}}\sigma_s^2 & \bar{\mathbf{G}}\sigma_{sd} \\ \bar{\mathbf{G}}\sigma_{sd} & \bar{\mathbf{G}}\sigma_d^2 \end{bmatrix} \begin{bmatrix} \mathbf{K}'\mathbf{Z}'_s \\ \mathbf{K}'\mathbf{Z}'_d \end{bmatrix}.$$

546 We make use of  $\mathbf{Z}_s \mathbf{K} = \begin{bmatrix} \mathbf{0}^u & \mathbf{Z}_s^v \end{bmatrix}$  and  $\mathbf{Z}_d \mathbf{K} = \begin{bmatrix} \mathbf{0}^u & \mathbf{Z}_d^v \end{bmatrix}$ , and rewrite

$$547 \quad \mathbf{Z}_s \mathbf{K} \bar{\mathbf{G}} \mathbf{K}' \mathbf{Z}_s' = \begin{bmatrix} \mathbf{0}^u & \mathbf{Z}_s^v \end{bmatrix} \begin{bmatrix} \frac{1}{2} \mathbf{A}_u & \mathbf{S}_{uv} \\ \mathbf{S}_{uv}' & \mathbf{G}_{vv} \end{bmatrix} \begin{bmatrix} (\mathbf{0}^u)' \\ (\mathbf{Z}_s^v)' \end{bmatrix} = \mathbf{Z}_s^v \mathbf{G}_{vv} \mathbf{Z}_s^{v'}.$$

548 In the same manner,

$$549 \quad \mathbf{Z}_s \mathbf{K} \bar{\mathbf{G}} \mathbf{K}' \mathbf{Z}_d' = \mathbf{Z}_s^v \mathbf{G}_{vv} \mathbf{Z}_d^{v'} , \quad \mathbf{Z}_d \mathbf{K} \bar{\mathbf{G}} \mathbf{K}' \mathbf{Z}_s' = \mathbf{Z}_d^v \mathbf{G}_{vv} \mathbf{Z}_s^{v'} \text{ and } \mathbf{Z}_d \mathbf{K} \bar{\mathbf{G}} \mathbf{K}' \mathbf{Z}_d' = \mathbf{Z}_d^v \mathbf{G}_{vv} \mathbf{Z}_d^{v'} .$$

550 From this, we get

$$551 \quad \mathbf{Q}_g = \mathbf{Z}_s^v \mathbf{G}_{vv} \mathbf{Z}_s^{v'} \sigma_s^2 + \mathbf{Z}_s^v \mathbf{G}_{vv} \mathbf{Z}_d^{v'} \sigma_{sd} + \mathbf{Z}_d^v \mathbf{G}_{vv} \mathbf{Z}_s^{v'} \sigma_{sd} + \mathbf{Z}_d^v \mathbf{G}_{vv} \mathbf{Z}_d^{v'} \sigma_d^2 = \mathbf{Q}_c .$$

552 From  $\mathbf{Q}_g = \mathbf{Q}_c$ , it follows that  $Var(\mathbf{y})$  is the same in both models and that they are equivalent.

553

## 554 **Appendix A2: Equivalence of classical and generalized gametic relationships in reduced models.**

555 We consider a reduced model with classical gametic relationships. With classical gametic  
 556 relationships all parents of final progeny and their ancestors have two gametic effects in the model  
 557 with covariance  $\mathbf{G}$ . In the generalized case the gametic effects of  $u$  of them are collapsed into  
 558 transmitting abilities, while the remaining  $v$  individuals retain their gametic effects. For the sake of  
 559 generality the latter group, among an arbitrary choice of others, includes all parents who may have  
 560 records. Parents with records need to have gametic effects, while all other individuals may be  
 561 modelled by gametic effects or by transmitting abilities. As final progeny have no genetic effects of  
 562 their own in the reduced model the variance of residuals  $\mathbf{W} \sigma_e^2$  is not affected by relationships. With  
 563 classical gametic relationships the variance of observations is

$$564 \quad Var(\mathbf{y}) = \mathbf{Q}_c + \mathbf{W} \sigma_e^2 ,$$

565 where

$$566 \quad \mathbf{Q}_c = \begin{bmatrix} \mathbf{Z}_s & \mathbf{Z}_d \end{bmatrix} \begin{bmatrix} \mathbf{G} \sigma_s^2 & \mathbf{G} \sigma_{sd} \\ \mathbf{G} \sigma_{sd} & \mathbf{G} \sigma_d^2 \end{bmatrix} \begin{bmatrix} \mathbf{Z}_s' \\ \mathbf{Z}_d' \end{bmatrix} .$$

567 The incidence matrix  $\mathbf{Z}_s$  for genetic effects can be partitioned as  $\mathbf{Z}_s = \begin{bmatrix} \mathbf{Z}_s^{2u} & \mathbf{Z}_s^v \end{bmatrix}$ . In the first

568 partition are two adjacent columns per individual, i.e.  $\mathbf{Z}_s^{2u} = \mathbf{Z}_s^u \otimes \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} = \mathbf{Z}_s^u \left[ \mathbf{I}_u \otimes \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \right]$ ,

569 where  $\mathbf{Z}_s^u$  is the corresponding partition from the same kind of incidence matrix in the model with  
 570 generalized gametic relationships and  $\mathbf{I}_u$  is an  $u \times u$  identity matrix. Note that a multiplication with

571  $\mathbf{K}$  cannot be applied here for the conversion of the matrix  $\begin{bmatrix} \mathbf{Z}_s^{2u} & \mathbf{Z}_s^v \end{bmatrix}$  into  $\begin{bmatrix} \mathbf{Z}_s^u & \mathbf{Z}_s^v \end{bmatrix}$ , because  $\mathbf{Z}_s^v$

572 may both have entries of single ones for records of parents and of pairs of one half for records from

573 final progeny. All of that also applies in an analogous manner to  $\mathbf{Z}_d^{2u}$  and  $\mathbf{Z}_d^u$ .

574 The first component of  $\mathbf{Q}_c$  is

$$\begin{aligned}
 575 \quad \mathbf{Z}_s \mathbf{G} \mathbf{Z}'_s &= \begin{bmatrix} \mathbf{Z}_s^{2u} & \mathbf{Z}_s^v \end{bmatrix} \begin{bmatrix} \mathbf{G}_{uu} \sigma_s^2 & \mathbf{G}_{uv} \sigma_s^2 \\ \mathbf{G}_{uv} \sigma_s^2 & \mathbf{G}_{vv} \sigma_s^2 \end{bmatrix} \begin{bmatrix} (\mathbf{Z}_s^{2u})' \\ (\mathbf{Z}_s^v)' \end{bmatrix} = \\
 576 \quad &\mathbf{Z}_s^{2u} \mathbf{G}_{uu} (\mathbf{Z}_s^{2u})' \sigma_s^2 + \mathbf{Z}_s^{2u} \mathbf{G}_{uv} (\mathbf{Z}_s^v)' \sigma_s^2 + \mathbf{Z}_s^v \mathbf{G}'_{uv} (\mathbf{Z}_s^{2u})' \sigma_s^2 + \mathbf{Z}_s^v \mathbf{G}_{vv} (\mathbf{Z}_s^v)' \sigma_s^2 = \\
 577 \quad &\mathbf{Z}_s^u \left[ \mathbf{I} \otimes \begin{pmatrix} \frac{1}{2} & \\ & \frac{1}{2} \end{pmatrix} \right] \mathbf{G}_{uu} \left[ \mathbf{I} \otimes \begin{pmatrix} \frac{1}{2} & \\ & \frac{1}{2} \end{pmatrix} \right]' (\mathbf{Z}_s^u)' \sigma_s^2 + \mathbf{Z}_s^u \left[ \mathbf{I} \otimes \begin{pmatrix} \frac{1}{2} & \\ & \frac{1}{2} \end{pmatrix} \right] \mathbf{G}_{uv} (\mathbf{Z}_s^v)' \sigma_s^2 + \\
 578 \quad &\mathbf{Z}_s^v \mathbf{G}'_{uv} \left[ \mathbf{I} \otimes \begin{pmatrix} \frac{1}{2} & \\ & \frac{1}{2} \end{pmatrix} \right]' (\mathbf{Z}_s^u)' \sigma_s^2 + \mathbf{Z}_s^v \mathbf{G}_{vv} (\mathbf{Z}_s^v)' \sigma_s^2 \\
 579 \quad &= \mathbf{Z}_s^u \frac{1}{2} \mathbf{A} (\mathbf{Z}_s^u)' \sigma_s^2 + \mathbf{Z}_s^u \mathbf{S}_{uv} (\mathbf{Z}_s^v)' \sigma_s^2 + \mathbf{Z}_s^v \mathbf{S}'_{uv} (\mathbf{Z}_s^u)' \sigma_s^2 + \mathbf{Z}_s^v \mathbf{G}_{vv} (\mathbf{Z}_s^v)' \sigma_s^2.
 \end{aligned}$$

580 Similarly,

$$\begin{aligned}
 581 \quad \mathbf{Z}_s \mathbf{G} \mathbf{Z}'_d &= \begin{bmatrix} \mathbf{Z}_s^{2u} & \mathbf{Z}_s^v \end{bmatrix} \begin{bmatrix} \mathbf{G}_{uu} \sigma_{sd} & \mathbf{G}_{uv} \sigma_{sd} \\ \mathbf{G}_{uv} \sigma_{sd} & \mathbf{G}_{vv} \sigma_{sd} \end{bmatrix} \begin{bmatrix} (\mathbf{Z}_d^{2u})' \\ (\mathbf{Z}_d^v)' \end{bmatrix} = \\
 582 \quad &\mathbf{Z}_s^u \frac{1}{2} \mathbf{A} (\mathbf{Z}_s^u)' \sigma_{sd} + \mathbf{Z}_s^u \mathbf{S}_{uv} (\mathbf{Z}_s^v)' \sigma_{sd} + \mathbf{Z}_s^v \mathbf{S}'_{uv} (\mathbf{Z}_s^u)' \sigma_{sd} + \mathbf{Z}_s^v \mathbf{G}_{vv} (\mathbf{Z}_s^v)' \sigma_{sd},
 \end{aligned}$$

583 and

$$\begin{aligned}
 584 \quad \mathbf{Z}_d \mathbf{G} \mathbf{Z}'_s &= \begin{bmatrix} \mathbf{Z}_d^{2u} & \mathbf{Z}_d^v \end{bmatrix} \begin{bmatrix} \mathbf{G}_{uu} \sigma_{sd} & \mathbf{G}_{uv} \sigma_{sd} \\ \mathbf{G}_{uv} \sigma_{sd} & \mathbf{G}_{vv} \sigma_{sd} \end{bmatrix} \begin{bmatrix} (\mathbf{Z}_s^{2u})' \\ (\mathbf{Z}_s^v)' \end{bmatrix} = \\
 585 \quad &\mathbf{Z}_d^u \frac{1}{2} \mathbf{A} (\mathbf{Z}_s^u)' \sigma_{sd} + \mathbf{Z}_d^u \mathbf{S}_{uv} (\mathbf{Z}_s^v)' \sigma_{sd} + \mathbf{Z}_d^v \mathbf{S}'_{uv} (\mathbf{Z}_s^u)' \sigma_{sd} + \mathbf{Z}_d^v \mathbf{G}_{vv} (\mathbf{Z}_s^v)' \sigma_{sd},
 \end{aligned}$$

586 finally

$$\begin{aligned}
 587 \quad \mathbf{Z}_d \mathbf{G} \mathbf{Z}'_d &= \begin{bmatrix} \mathbf{Z}_d^{2u} & \mathbf{Z}_d^v \end{bmatrix} \begin{bmatrix} \mathbf{G}_{uu} \sigma_d^2 & \mathbf{G}_{uv} \sigma_d^2 \\ \mathbf{G}_{uv} \sigma_d^2 & \mathbf{G}_{vv} \sigma_d^2 \end{bmatrix} \begin{bmatrix} (\mathbf{Z}_d^{2u})' \\ (\mathbf{Z}_d^v)' \end{bmatrix} = \\
 588 \quad &\mathbf{Z}_d^u \frac{1}{2} \mathbf{A} (\mathbf{Z}_d^u)' \sigma_d^2 + \mathbf{Z}_d^u \mathbf{S}_{uv} (\mathbf{Z}_d^v)' \sigma_d^2 + \mathbf{Z}_d^v \mathbf{S}'_{uv} (\mathbf{Z}_d^u)' \sigma_d^2 + \mathbf{Z}_d^v \mathbf{G}_{vv} (\mathbf{Z}_d^v)' \sigma_d^2.
 \end{aligned}$$

589 From this,  $\mathbf{Q}_c$  can be summarized as

$$590 \quad \mathbf{Q}_c = \begin{bmatrix} (\mathbf{Z}_s^u & \mathbf{Z}_s^v) & (\mathbf{Z}_d^u & \mathbf{Z}_d^v) \end{bmatrix} \begin{pmatrix} \sigma_s^2 & \sigma_{sd} \\ \sigma_{sd} & \sigma_s^2 \end{pmatrix} \otimes \begin{bmatrix} \frac{1}{2} \mathbf{A}_{uu} & \mathbf{S}_{uv} \\ \mathbf{S}_{uv} & \mathbf{G}_{vv} \end{bmatrix} \begin{bmatrix} (\mathbf{Z}_s^u & \mathbf{Z}_s^v)' \\ (\mathbf{Z}_d^u & \mathbf{Z}_d^v)' \end{bmatrix},$$

591 which is equal to the equivalent quantity  $\mathbf{Q}_r$  using generalized gametic relationships

$$592 \quad \mathbf{Q}_r = \begin{bmatrix} (\mathbf{Z}_s^u & \mathbf{Z}_s^v) & (\mathbf{Z}_d^u & \mathbf{Z}_d^v) \end{bmatrix} \begin{bmatrix} \bar{\mathbf{G}}\sigma_s^2 & \bar{\mathbf{G}}\sigma_{sd} \\ \bar{\mathbf{G}}\sigma_{sd} & \bar{\mathbf{G}}\sigma_d^2 \end{bmatrix} \begin{bmatrix} (\mathbf{Z}_s^u & \mathbf{Z}_s^v)' \\ (\mathbf{Z}_d^u & \mathbf{Z}_d^v)' \end{bmatrix}.$$

593 Thus

$$594 \quad \text{Var}(\mathbf{y}) = \mathbf{Q}_c + \mathbf{W}\sigma_e^2 = \mathbf{Q}_r + \mathbf{W}\sigma_e^2, \text{ q.e.d. .}$$

595

596 **Appendix A3: Equivalence between the model with gametic effects for all individuals and and the**  
 597 **reduced model with gametic effects for parents**

598 We consider a classical gametic model that includes a number  $f$  of final progeny. The vector  $\mathbf{g}'$  is  
 599 partitioned into two components; in  $\mathbf{g}'_f$  are the  $2f$  gametic effects of the final  $f$  progeny, and  
 600 other gametic effects are in  $\mathbf{g}'_g$ . The covariance of  $\mathbf{g}'$  then is

$$601 \quad \text{Var}(\mathbf{g}) = \text{Var} \begin{bmatrix} \mathbf{g}_g \\ \mathbf{g}_f \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{gg} & \mathbf{G}_{gf} \\ \mathbf{G}'_{gf} & \mathbf{G}_{ff} \end{bmatrix} = \mathbf{G}.$$

602 The incidence matrices for gametic effects are

$$603 \quad \mathbf{Z}_s^{all} = \begin{bmatrix} \mathbf{Z}_{s1}^g & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_s^f \end{bmatrix} \text{ and } \mathbf{Z}_d^{all} = \begin{bmatrix} \mathbf{Z}_{d1}^g & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_d^f \end{bmatrix},$$

604 where  $\mathbf{Z}_{s1}^g$  ( $\mathbf{Z}_{d1}^g$ ) relates the observations to the gametic effects as sire (as dam) of individuals that  
 605 are not in the set of the  $f$  final progeny. Accordingly  $\mathbf{Z}_s^f$  ( $\mathbf{Z}_d^f$ ) relates observations of the  $f$  final  
 606 progeny to their respective gametic effects as sire (as dam).

607 By contrast, in a reduced model, all observations of the  $f$  final progeny are to be related to the  
 608 gametic effects as sire (as dam) of their parents. The respective incidence matrices are

$$609 \quad \mathbf{Z}_s^{red} = \begin{bmatrix} \mathbf{Z}_{s1}^g & \mathbf{0} \\ \mathbf{Z}_{s2}^g & \mathbf{0} \end{bmatrix} \text{ and } \mathbf{Z}_d^{red} = \begin{bmatrix} \mathbf{Z}_{d1}^g & \mathbf{0} \\ \mathbf{Z}_{d2}^g & \mathbf{0} \end{bmatrix},$$

610 where  $\mathbf{Z}_s^{red}$  and  $\mathbf{Z}_d^{red}$  have only zero entries in columns for gametic effects of final progeny.

611 The relationships between the incidence matrices of the two types of models are

$$612 \quad \mathbf{Z}_s^{red} = \begin{bmatrix} \mathbf{Z}_{s1}^g & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_s^f \end{bmatrix} - \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ -\mathbf{Z}_{s2}^g & \mathbf{Z}_s^f \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{s1}^g & \mathbf{0} \\ \mathbf{Z}_{s2}^g & \mathbf{0} \end{bmatrix} = \mathbf{Z}_s^{all} - \mathbf{Z}_s^\delta,$$

613 The matrix  $\mathbf{Z}_s^\delta$  is the difference between  $\mathbf{Z}_s^{all}$  and  $\mathbf{Z}_s^{red}$  :

$$614 \quad \mathbf{Z}_s^\delta = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ -\mathbf{Z}_{s2}^g & \mathbf{Z}_s^f \end{bmatrix}, \text{ and analogously } \mathbf{Z}_d^\delta = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ -\mathbf{Z}_{d2}^g & \mathbf{Z}_d^f \end{bmatrix},$$

$$615 \quad \mathbf{Z}_s^{all} = \mathbf{Z}_s^{red} + \mathbf{Z}_s^\delta, \text{ and}$$

$$616 \quad \mathbf{Z}_d^{all} = \mathbf{Z}_d^{red} + \mathbf{Z}_d^\delta.$$

617 For the proof of equivalence of the two models, we express the variance of observations  $Var(\mathbf{y})$  in  
618 terms of model-specific incidence matrices and show their equality by making use of the last two  
619 identities.

620 For any reduced observation equation the variances of the relevant Mendelian sampling effects are  
621 part of the residual. For each paternal gamete as sire of a final progeny, the Mendelian sampling  
622 effect is the difference between the effect of the paternal gamete and the transmitting ability of the  
623 individual's sire as sire. The respective vector is

$$624 \quad \mathbf{m}_s = \mathbf{Z}_s^\delta \mathbf{g}_f,$$

625 and the maternal counterpart as dam is

$$626 \quad \mathbf{m}_d = \mathbf{Z}_d^\delta \mathbf{g}_f.$$

627 The common covariance matrix is

$$628 \quad Var \begin{bmatrix} \mathbf{m}_s \\ \mathbf{m}_d \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_s^\delta \mathbf{G} (\mathbf{Z}_s^\delta)' \sigma_s^2 & \mathbf{Z}_s^\delta \mathbf{G} (\mathbf{Z}_d^\delta)' \sigma_{sd} \\ \mathbf{Z}_s^\delta \mathbf{G} (\mathbf{Z}_d^\delta)' \sigma_{sd} & \mathbf{Z}_d^\delta \mathbf{G} (\mathbf{Z}_d^\delta)' \sigma_d^2 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_s^\delta \mathbf{G} (\mathbf{Z}_s^\delta)' \sigma_s^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_d^\delta \mathbf{G} (\mathbf{Z}_d^\delta)' \sigma_d^2 \end{bmatrix}.$$

629 The covariance between  $\mathbf{m}_s$  and  $\mathbf{m}_d$  is zero as all rows of  $\mathbf{Z}_s^\delta$  have their non-zero entries at places  
630 other than the rows of  $\mathbf{Z}_d^\delta$ , causing  $\mathbf{Z}_s^\delta \mathbf{G} (\mathbf{Z}_d^\delta)'$  to be a matrix of zeroes.

631 In detail the product is

$$632 \quad \mathbf{Z}_s^\delta \mathbf{G} (\mathbf{Z}_d^\delta)' = \left\{ \text{sum} \left( \mathbf{z}_{d,i}^\delta (\mathbf{z}_{s,j}^\delta)' \square \mathbf{G} \right) \right\}_{ij} = \{0\}_{ij},$$

633 where  $\square$  denotes element-wise multiplication,  $\sum ()$  is the sum of all matrix elements in  $()$ , and

634  $(\mathbf{z}_{d,i}^\delta)'$  and  $(\mathbf{z}_{s,j}^\delta)'$  are the  $i$ th and  $j$ th rows of the two involved incidence matrices.

635 The total  $Var(\mathbf{y})$  in the reduced model is

$$636 \quad Var(\mathbf{y}) = \mathbf{Q}_{red} + Var(\mathbf{m}_s) + Var(\mathbf{m}_d) + \mathbf{I}\sigma_e^2,$$

637 with

$$638 \quad \mathbf{Q}_{red} = \begin{bmatrix} \mathbf{Z}_s^{red} & \mathbf{Z}_d^{red} \end{bmatrix} \begin{bmatrix} \sigma_s^2 & \sigma_{sd} \\ \sigma_{sd} & \sigma_d^2 \end{bmatrix} \otimes \mathbf{G} \begin{bmatrix} (\mathbf{Z}_s^{red})' \\ (\mathbf{Z}_d^{red})' \end{bmatrix}$$

$$639 \quad = \mathbf{Z}_s^{red} \mathbf{G} (\mathbf{Z}_s^{red})' \sigma_s^2 + \mathbf{Z}_s^{red} \mathbf{G} (\mathbf{Z}_d^{red})' \sigma_{sd} + \mathbf{Z}_d^{red} \mathbf{G} (\mathbf{Z}_s^{red})' \sigma_{sd} + \mathbf{Z}_d^{red} \mathbf{G} (\mathbf{Z}_d^{red})' \sigma_d^2.$$

640 In the classical gametic model the variance of observations  $Var(\mathbf{y})$  is

$$641 \quad \mathbf{Q}_{all} + \mathbf{I}\sigma_e^2.$$

642 The first component is

$$643 \quad \mathbf{Q}_{all} = \begin{bmatrix} \mathbf{Z}_s^{all} & \mathbf{Z}_d^{all} \end{bmatrix} \begin{bmatrix} \sigma_s^2 & \sigma_{sd} \\ \sigma_{sd} & \sigma_d^2 \end{bmatrix} \otimes \mathbf{G} \begin{bmatrix} (\mathbf{Z}_s^{all})' \\ (\mathbf{Z}_d^{all})' \end{bmatrix}$$

$$644 \quad = \begin{bmatrix} \mathbf{Z}_s^{red} + \mathbf{Z}_s^\delta & \mathbf{Z}_d^{red} + \mathbf{Z}_d^\delta \end{bmatrix} \begin{bmatrix} \sigma_s^2 & \sigma_{sd} \\ \sigma_{sd} & \sigma_d^2 \end{bmatrix} \otimes \mathbf{G} \begin{bmatrix} (\mathbf{Z}_s^{red} + \mathbf{Z}_s^\delta)' \\ (\mathbf{Z}_d^{red} + \mathbf{Z}_d^\delta)' \end{bmatrix}.$$

645 This results in a sum of 16 terms, of which the first four are

$$646 \quad = \mathbf{Z}_s^{red} \mathbf{G} (\mathbf{Z}_s^{red})' \sigma_s^2 + \mathbf{Z}_s^{red} \mathbf{G} (\mathbf{Z}_d^{red})' \sigma_{sd} + \mathbf{Z}_d^{red} \mathbf{G} (\mathbf{Z}_s^{red})' \sigma_{sd} + \mathbf{Z}_d^{red} \mathbf{G} (\mathbf{Z}_d^{red})' \sigma_d^2.$$

647 This is equal to  $\mathbf{Q}_{red}$  in the reduced model. Further, we have two more terms

$$648 \quad \mathbf{Z}_s^\delta \mathbf{G} (\mathbf{Z}_s^\delta)' \sigma_s^2 + \mathbf{Z}_d^\delta \mathbf{G} (\mathbf{Z}_d^\delta)' \sigma_d^2,$$

649 equivalent to  $Var(\mathbf{m}_s) + Var(\mathbf{m}_d)$ . The remaining 10 terms in

$$\begin{aligned}
 & \mathbf{Z}_s^{red} \mathbf{G} (\mathbf{Z}_s^{red})' \sigma_s^2 + \mathbf{Z}_s^\delta \mathbf{G} (\mathbf{Z}_s^{red})' \sigma_s^2 + \\
 & \mathbf{Z}_s^{red} \mathbf{G} (\mathbf{Z}_d^\delta)' \sigma_{sd} + \mathbf{Z}_s^\delta \mathbf{G} (\mathbf{Z}_d^{red})' \sigma_{sd} + \mathbf{Z}_s^\delta \mathbf{G} (\mathbf{Z}_d^\delta)' \sigma_{sd} + \\
 & \mathbf{Z}_d^{red} \mathbf{G} (\mathbf{Z}_s^\delta)' \sigma_{sd} + \mathbf{Z}_d^\delta \mathbf{G} (\mathbf{Z}_s^{red})' \sigma_{sd} + \mathbf{Z}_d^\delta \mathbf{G} (\mathbf{Z}_s^\delta)' \sigma_{sd} + \\
 & \mathbf{Z}_d^{red} \mathbf{G} (\mathbf{Z}_d^{red})' \sigma_s^2 + \mathbf{Z}_d^\delta \mathbf{G} (\mathbf{Z}_d^{red})' \sigma_s^2
 \end{aligned}$$

all are zero matrices. Thus,  $\mathbf{Q}_{all} = \mathbf{Q}_{red} + Var(\mathbf{m}_s) + Var(\mathbf{m}_d)$  and, therefore, the variance  $Var(\mathbf{y})$  in the classical gametic model and the reduced model with gametic relationships are identical. As both models also have identical expectations of  $\mathbf{y}$ , they are equivalent; **q.e.d.**

#### Appendix A4: Maternal genetic variance in a reduced model.

We consider a reduced model equation for a single observation:

$$y_i = \mu + m_d + a_d^d + a_s^s + r_i.$$

This equation comprises the maternal breeding value  $m_d$  of the dam  $d$  of individual  $i$ , together with the transmitting ability as dam  $a_d^d$  of the dam  $d$  of  $i$ , the transmitting ability as sire  $a_s^s$  of the sire  $s$  of  $i$ , and the residual  $r_i$ .

Then, the covariance of the respective vectors of random genetic effects is

$$Var \begin{bmatrix} \mathbf{m} \\ \mathbf{a}^s \\ \mathbf{a}^d \end{bmatrix} = \begin{bmatrix} \sigma_m^2 & \sigma_{ms} & \sigma_{md} \\ \sigma_{ms} & \sigma_s^2 & \sigma_{sd} \\ \sigma_{md} & \sigma_{sd} & \sigma_d^2 \end{bmatrix} \otimes \frac{1}{2} \mathbf{A},$$

where  $\sigma_m^2$  is the maternal gametic variance. As we use  $\frac{1}{2} \mathbf{A}$  as relationship matrix (assuming all records are from final progeny) the incidence matrix for maternal breeding values needs to have non-zero entries of two to match this set of covariances.

The variance of observations has the non-residual component

$$\mathbf{Q}_{rm} = [\mathbf{Z}_m \quad \mathbf{Z}_s \quad \mathbf{Z}_d] \begin{bmatrix} \sigma_m^2 & \sigma_{ms} & \sigma_{md} \\ \sigma_{ms} & \sigma_s^2 & \sigma_{sd} \\ \sigma_{md} & \sigma_{sd} & \sigma_d^2 \end{bmatrix} \otimes \frac{1}{2} \mathbf{A} \begin{bmatrix} \mathbf{Z}_m' \\ \mathbf{Z}_s' \\ \mathbf{Z}_d' \end{bmatrix},$$

involving the incidence matrices  $\mathbf{Z}_m$ ,  $\mathbf{Z}_s$ , and  $\mathbf{Z}_d$  that link observations to maternal genetic effects, transmitting abilities as sire, and transmitting abilities as dam, respectively.  $\mathbf{Q}_{rm}$  is a sum of nine

670 matrices; of them, the following matrix equalities can be found by dropping the respective  
671 components of variance:

$$672 \quad \mathbf{Z}_m \mathbf{A} \mathbf{Z}'_m = \mathbf{Z}_d \mathbf{A} \mathbf{Z}'_d = \mathbf{Z}_m \mathbf{A} \mathbf{Z}'_d = \mathbf{Z}_d \mathbf{A} \mathbf{Z}'_m$$

$$673 \quad \mathbf{Z}_m \mathbf{A} \mathbf{Z}'_s = \mathbf{Z}_d \mathbf{A} \mathbf{Z}'_s$$

$$674 \quad \mathbf{Z}_s \mathbf{A} \mathbf{Z}'_m = \mathbf{Z}_s \mathbf{A} \mathbf{Z}'_d.$$

675 The underlying fact is that the incidence matrices  $\mathbf{Z}_m$  and  $\mathbf{Z}_d$  link all observations to genetic effects  
676 of the same animals, i.e., of the dam of each final progeny. Thus, the incidence matrices  $\mathbf{Z}_m = \mathbf{Z}_d$   
677 are equal, and constitute equalities from above. Consequently,  $\mathbf{Q}_{rm}$  can be rewritten as

$$678 \quad \mathbf{Q}_{rm} = \frac{1}{2} \begin{bmatrix} \mathbf{Z}_m \mathbf{A} \mathbf{Z}'_m \sigma_m^2 + \mathbf{Z}_d \mathbf{A} \mathbf{Z}'_d \sigma_d^2 + \mathbf{Z}_m \mathbf{A} \mathbf{Z}'_d \sigma_{md} + \mathbf{Z}_d \mathbf{A} \mathbf{Z}'_m \sigma_{md} & \mathbf{Z}_m \mathbf{A} \mathbf{Z}'_s \sigma_{ms} + \mathbf{Z}_d \mathbf{A} \mathbf{Z}'_s \sigma_{sd} \\ \mathbf{Z}_s \mathbf{A} \mathbf{Z}'_m \sigma_{ms} + \mathbf{Z}_s \mathbf{A} \mathbf{Z}'_d \sigma_{sd} & \mathbf{Z}_s \mathbf{A} \mathbf{Z}'_s \sigma_s^2 \end{bmatrix},$$

679 which, in terms of the incidence matrices of the reduced model without maternal genetic effects, is

$$680 \quad \mathbf{Q}_{rm} = \frac{1}{2} \begin{bmatrix} \mathbf{Z}_d \mathbf{A} \mathbf{Z}'_d (\sigma_d^2 + \sigma_m^2 + 2\sigma_{md}) & \mathbf{Z}_d \mathbf{A} \mathbf{Z}'_s (\sigma_{sd} + \sigma_{ms}) \\ \mathbf{Z}_s \mathbf{A} \mathbf{Z}'_d (\sigma_{sd} + \sigma_{ms}) & \mathbf{Z}_s \mathbf{A} \mathbf{Z}'_s \sigma_s^2 \end{bmatrix}.$$

681 The variance in the transmitting ability as dam and the covariance with the transmitting ability as sire  
682 are therefore contaminated by components of the maternal genetic (co-)variances. This shows that  
683 in the presence of maternal genetic effects,  $\sigma_d^2$  and  $\sigma_s^2$  cannot be inferred from the reduced model.  
684 Moreover, we cannot correctly calculate the weights of the observations as this would require that  
685 we know these two components of variance.

686 Interestingly, we can assume the absence of genomic imprinting and make use of  
687  $\sigma_g^2 = \sigma_s^2 = \sigma_d^2 = \sigma_{sd}$ , from which the residual variance of observation  $i$  becomes

$$688 \quad \frac{1}{2}(1 - F_{s,i})\sigma_g^2 + \frac{1}{2}(1 - F_{d,i})\sigma_g^2 + \sigma_e^2.$$

689 Consequently, the imprinting variance becomes

$$690 \quad \sigma_i^2 = (\sigma_d^2 + \sigma_m^2 + 2\sigma_{md} + \sigma_s^2) - 2(\sigma_{sd} + \sigma_{ms}) = \sigma_m^2.$$

691

## 692 **Appendix A5: Maternal variance in a classical gametic model.**

693 The model equation for a single observation  $y_i$  in a gametic model with maternal effects is

694 
$$y_i = \mu + g_{d,1}^m + g_{d,2}^m + g_{i,1}^s + g_{i,2}^d + e_i.$$

695 In this, we have the maternal effect (superscript  $m$ ) of the paternal (1) gamete  $g_{d,1}^m$  and the  
 696 maternal (2) gamete  $g_{d,2}^m$  of the dam  $d$  of individual.  $g_{i,1}^s$  is the effect of the paternal gamete of  
 697 individual  $i$  as sire,  $g_{i,2}^d$  is the effect of the maternal allele of individual  $i$  as dam, and  $e_i$  is the  
 698 residual. The covariance of random gametic effects is

699 
$$\text{Var} \begin{bmatrix} \mathbf{g}^m \\ \mathbf{g}^s \\ \mathbf{g}^d \end{bmatrix} = \begin{bmatrix} \sigma_m^2 & \sigma_{ms} & \sigma_{md} \\ \sigma_{ms} & \sigma_s^2 & \sigma_{sd} \\ \sigma_{md} & \sigma_{sd} & \sigma_d^2 \end{bmatrix} \otimes \mathbf{G}.$$

700 The vector of observations  $\mathbf{y}$  has covariance  $\mathbf{V} = \mathbf{Q}_{mm} + \mathbf{I}\sigma_e^2$ , with

701 
$$\mathbf{Q}_{mm} = \begin{bmatrix} \mathbf{Z}_m & \mathbf{Z}_s & \mathbf{Z}_d \end{bmatrix} \begin{bmatrix} \sigma_m^2 & \sigma_{ms} & \sigma_{md} \\ \sigma_{ms} & \sigma_s^2 & \sigma_{sd} \\ \sigma_{md} & \sigma_{sd} & \sigma_d^2 \end{bmatrix} \otimes \mathbf{G} \begin{bmatrix} \mathbf{Z}_m' \\ \mathbf{Z}_s' \\ \mathbf{Z}_d' \end{bmatrix}.$$

702 All other components of covariance are defined as before in appendix A4.

703  $\mathbf{Q}_{mm}$  has three components— $\mathbf{Z}_m \mathbf{G} \mathbf{Z}_m' \sigma_m^2$ ,  $\mathbf{Z}_d \mathbf{G} \mathbf{Z}_d' \sigma_d^2$ , and —related to components of variance,  
 704 and another three— $\mathbf{Z}_m \mathbf{G} \mathbf{Z}_s' \sigma_{ms} + \mathbf{Z}_s \mathbf{G} \mathbf{Z}_m' \sigma_{ms}$ ,  $\mathbf{Z}_m \mathbf{G} \mathbf{Z}_d' \sigma_{md} + \mathbf{Z}_d \mathbf{G} \mathbf{Z}_m' \sigma_{md}$ , and—that are connected  
 705 to the covariances. In contrast to the reduced model, the incidence matrices  $\mathbf{Z}_m$  and  $\mathbf{Z}_d$  relate the  
 706 records to different gametic effects and, therefore, are not equal. As a result, all six addends of  $\mathbf{Q}_{mm}$   
 707 are linearly independent and all components of variance can be separated.

708

709

710

711

712

713

714

715



**Table 1:** Size and indices of non-zero elements of vectors  $\mathbf{u}'_i$  by kind of genetic effect (a: transmitting ability; g: gametic effect). The cases indicate unique combinations of kind of genetic effect and kind of indices. The latter consist of i: number of genetic effects; d: transmission ability of dam; s: transmission ability of sire; u: paternal gamete of dam; v: maternal gamete of dam; p: paternal gamete of sire; q: maternal gamete of sire. For gametic effects (cases g-a and g-gg), the respective effects of the known parent are indexed as for a sire.

Kind of effect	Case	Non-zero-elements in $\mathbf{u}'_i$	Indices of non-zero elements
a	a-00	1	i
a	a-0a	$-\frac{1}{2}$ 1	d, i
a	a-0gg	$-\frac{1}{4}$ $-\frac{1}{4}$ 1	u, v, i
a	a-a0	$-\frac{1}{2}$ 1	s, i
a	a-gg0	$-\frac{1}{4}$ $-\frac{1}{4}$ 1	p, q, i
a	a-aa	$-\frac{1}{2}$ $-\frac{1}{2}$ 1	s, d, i
a	a-agg	$-\frac{1}{2}$ $-\frac{1}{4}$ $-\frac{1}{4}$ 1	s, u, v, i
a	a-gga	$-\frac{1}{4}$ $-\frac{1}{4}$ $-\frac{1}{2}$ 1	p, q, d, i
a	a-gggg	$-\frac{1}{4}$ $-\frac{1}{4}$ $-\frac{1}{4}$ $-\frac{1}{4}$ 1	p, q, u, v, i
g	g-a	-1 1	s, i
g	g-gg	$-\frac{1}{2}$ $-\frac{1}{2}$ 1	p, q, i
g	g-0	1	i