Animal models with group-specific additive genetic variances: extending genetic group models

Stefanie Muff[1,2⋆], Alina K. Niskanen[3,4], Dilan Saatoglu[3], Lukas F. Keller[1,5‡], Henrik Jensen[3,‡]

May 28, 2018

[1] Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, Zurich, Switzerland

[2] Epidemiology, Biostatistics and Prevention Institute, Department of Biostatistics, University of Zurich, Hirschengraben 84, Zurich, Switzerland

[3] Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, Høgskoleringen 5, Trondheim, Norway

[4] Department of Ecology and Genetics, University of Oulu, P.O. Box 3000, Oulu, Finland

[5] Zoological Museum, University of Zurich, Karl-Schmid-Strasse 4, Zurich, Switzerland

‡ Joint last authors

⋆ Corresponding author. E-mail: stefanie.muff@uzh.ch, Phone: +41 44 634 5487

Abstract

1. The *animal model* is a key tool in quantitative genetics and has been used extensively to estimate fundamental parameters, such as additive genetic variance, heritability, or inbreeding effects. An implicit assumption of animal models is that all founder individuals derive from a single population. This assumption is commonly violated, for instance in cross-bred livestock breeds, when an observed population receive immigrants, or when a meta-population is split into genetically differentiated subpopulations. Ignoring genetic differences among different source populations of founders may lead to biased parameter estimates, in particular for the additive genetic variance.

2. To avoid such biases, genetic group models, extensions to the animal model that account for the presence of more than one genetic group, have been proposed. As a key limitation, the method to date only allows that the breeding values differ in their means, but not in their variances among the groups. Methodology previously proposed to account for group-specific variances included terms for segregation variance, which rendered the models infeasibly complex for application to most real study systems.

3. Here we explain why segregation variances are often negligible when analyzing the complex polygenic traits that are frequently the focus of evolutionary ecologists and animal breeders. Based on this we suggest an extension of the animal model that permits estimation of group-specific additive genetic variances. This is achieved by employing group-specific relatedness matrices for the breeding value components attributable to different genetic groups. We derive these matrices by decomposing the full relatedness matrix via the generalized Cholesky decomposition, and by scaling the respective matrix components for each group. To this end, we propose a computationally convenient approximation for the matrix component that encodes for the Mendelian sampling variance. Although convenient, this approximation is not critical.

4. Simulations and an example from an insular meta-population of house sparrows in Norway with three genetic groups illustrate that the method is successful in estimating group-specific additive genetic variances and that segregation variances are indeed negligible in the empirical example.

5. Quantifying differences in additive genetic variance within and among populations is of major biological interest in ecology, evolution, and animal and plant breeding. The proposed method allows to estimate such differences for subpopulations that form a connected meta-population, which may also be useful to study temporal or spatial variation of additive genetic variance.

Keywords: Relatedness matrix, immigration, Cholesky decomposition, partial inbreeding coefficient, house sparrows.

1 Introduction

Quantifying the (causal) relationships between genes and observed phenotypic traits is a central task of empirical studies of adaptive evolution (Lynch & Walsh, 1998; Charmantier *et al.*, 2014) and of plant and animal breeding (Falconer & Mackay, 1996). The *animal model* (Henderson, 1984; Kruuk, 2004; Wilson *et al.*, 2010) has become a popular statistical approach to disentangle genetic effects on a phenotype from other factors that may induce phenotypic similarities among relatives, such as shared environmental effects (Stopher *et al.*, 2012), inbreeding (Reid & Keller, 2010), or individual traits such as age or sex (Wilson, 2008; de Villemereuil *et al.*, 2017). Fundamental to the animal model is information on how animals are related to each other, information typically obtained from pedigree data (Wright, 1922; Henderson, 1976) or from single-nucleotide polymorphism (SNP) data (*e. g.* Speed & Balding, 2015). Pedigrees are still the most commonly used source of relatedness information in animal models (*e. g.* Bonnet *et al.*, 2017; Wolak & Reid, 2017), in part because this leads to models that are computationally efficient.

All pedigrees start with a founder generation with unknown parents (Wolak & Reid, 2017, Appendix S1). The animal model assumes that all founder individuals stem from a single, genetically homogeneous baseline population. When this assumption is violated, for example in populations that receive immigrants or in crossbred livestock breeds, estimates of additive genetic variance are biased (Dong *et al.*, 1988; Cantet *et al.*, 2000). To address this problem, animal breeders developed *genetic group models* (*e. g.* Quaas, 1988), models that are now also receiving attention in evolutionary ecology (Wolak & Reid, 2017). The main idea behind genetic group models is that estimates of additive genetic variance may be biased if animals of distinct genetic origin differ in their mean breeding values, and that accounting for such differences in breeding values may reduce or eliminate the bias (Wolak & Reid, 2017). However, current genetic group models have an important key limitation: genetic groups are allowed to differ in mean breeding value but are assumed to have the same additive genetic variance. This homogeneity assumption is violated in some animal breeding applications (*e. g.* Elzo, 1990; Lo *et al.*, 1993; Alfonso & Estany, 1999), and is likely also violated in many natural populations, where source populations of immigrants may differ in additive genetic variance due to differences in effective population size, selection regimes, etc. (*e. g.* Mackay, 1981; Lande, 1988; Hoffmann *et al.*, 2017).

Aiming for better predictions of breeding values in crossbred populations, animal breeders have suggested approaches that account for heterogeneous additive genetic variances across genetic groups (Lo *et al.*, 1993; Cantet & Fernando, 1995; García-Cortés & Toro, 2006). One drawback of these models is that they rapidly become infeasibly complex, mainly due to additional terms that are needed to account for

segregation variance when breeds are mixed. Segregation variance refers to differences in variance when the average proportion of genetic origin from the ancestral genetic groups is the same, as in F_1 and F_2 generations of line crosses (Wright, 1968; Lande, 1981). Segregation variances can be large in crossbreeding applications (Lynch & Walsh, 1998, p. 11). However, as we detail below, we expect the segregation variance from crossing different genetic groups in wild study populations to be small for many traits of interest, so that omitting it from the animal model does not lead to significant bias.

Here we use this simplification to derive genetic group models that allow for group-specific additive genetic variances. Similar to García-Cortés & Toro (2006), we additively split the total breeding value of each individual into group-specific components, which covary according to a group-specific relatedness matrix. The main challenge is to find these matrices, and we derive them here by decomposing the full relatedness matrix via a generalized Cholesky decomposition (as described by Henderson, 1976), and by appropriately scaling the respective matrix components for each group. In the following, we first summarize the current state of the homogeneous genetic group models and then give a detailed description of our extension to heterogeneous group-specific additive genetic variances. We illustrate the performance of our method with a simulation study and an application to a meta-population of house sparrows (*Passer domesticus*) in Norway, where genetic groups are determined by geographical properties of the bird's natal island population. By also fitting a model that includes the segregation variance, we use this empirical study system to illustrate that omitting segregation variances is unproblematic in such applications. Finally, we discuss opportunities and limitations of our extended genetic group model.

2 The animal model for genetic groups with homogeneous variances

The basic animal model for a phenotypic measurement y_i of an individual i ($1 \leq i \leq n$) is

$$y_i = \mu + a_i + e_i \ , \tag{1}$$

with population mean μ, environmental component $e_i \sim \mathsf{N}(0, \sigma_E^2)$ with environmental variance σ_E^2, and breeding values distributed as $\boldsymbol{a}^\top = (a_1, \ldots, a_n)^\top \sim \mathsf{N}(\mathbf{0}, \sigma_A^2 \mathbf{A})$ with additive genetic variance σ_A^2 and additive genetic relatedness matrix $\mathbf{A}$ that represents the relatedness among individuals (Henderson, 1976; Kruuk, 2004). In diploid non-inbred organisms, each element in $\mathbf{A}$ is equal to two times the coefficient of coancestry between a pair of individuals (Lynch and Walsh 1998). Model (1) is often extended

by fixed effects (such as sex or age) and by additional random effects that account for permanent environmental conditions (see *e. g.* Wilson *et al.*, 2010). In all cases, the underlying assumption is that all animals in the analysis derive from the same genetic population, and that the breeding values (a_i) encode for the deviation from the mean of this population and thus have a mean of zero.

As noted by animal breeders a long time ago, this assumption is frequently violated, for example in crossbred populations from genetically differentiated breeds. In such cases it is necessary to allow for differences in mean breeding values among animals with different genetic origin (Dong *et al.*, 1988; Quaas, 1988). Let us denote by a founder population a set of animals with unknown (*i. e.* missing) parents, and assume that r founder populations exist, where each of them corresponds to a different genetic group. When animals from different genetic groups mate, their genetic contributions are propagated through the pedigree following the Mendelian rules of inheritance. Offspring in later generations may thus inherit different proportions of the genome from the genetic groups. Denote by q_{ij} the expected proportional contribution of founder population j to the genome of individual i. The respective values are typically written into a matrix $\mathbf{Q}$ with n rows (n=number of animals) and r columns, such that q_{ij} is the value in the ith row and j^{th} column (Wolak & Reid, 2017, Fig. 3). Following the notation of Wolak & Reid (2017), and denoting by g_j the expected average genetic effect in group j, the basic animal model (1) is then extended to

$$y_i = \mu + \sum_{j=1}^{r} q_{ij} g_j + a_i + e_i \ , \tag{2}$$

with total additive genetic effect of individual i given as $u_i = \sum_{j=1}^{r} q_{ij} g_j + a_i$, that is, the weighted sum of genetic group-mean effects, plus the breeding value a_i of the individual that accounts for deviations from the weighted group mean. This model allows for different mean, but it is actually overparameterized because for each i the contributions from the r groups sum up to 1, that is, $\sum_j q_{ij} = 1$. Similar to ANOVA models or when categorical variables are included in regression models, the parameters become identifiable when one group is set as reference group (*e. g.* assuming $g_1 = 0$), or when additional constraints are added, such as $\sum_j g_j = 0$.

Let us illustrate the idea for two genetic groups. When using the convenient constraint $g_1 = 0$, phenotypes of animals that have ancestors either only from group 1 (*i. e.* $q_{i1} = 1, q_{i2} = 0$) or only from group 2 (*i. e.* $q_{i1} = 0, q_{i2} = 1$) can be described by the following models

$$\begin{aligned} y_i &= \mu + a_i + e_i && \text{for } i \text{ in group 1 ,} \\ y_i &= \mu + g_2 + a_i + e_i && \text{for } i \text{ in group 2 .} \end{aligned}$$

Thus, members of genetic group 1 have a total additive genetic effect of $u_i = a_i$ and members of genetic group 2 have $u_i = g_2 + a_i$, where g_2 estimates the difference between the mean breeding values of the groups. Therefore, the respective values are distributed around an overall mean of 0 and g_2, respectively, but with the same additive genetic variance σ_A^2 and relatedness matrix $\mathbf{A}$. Note that, while the main benefit of including group-specific means is that bias in the estimates of σ_A^2 is reduced, the estimated values (*e. g.* of g_2) may sometimes be of interest themselves (see the Conclusions in Wolak & Reid, 2017).

3 Genetic group models with heterogeneous additive genetic variances

3.1 Segregation variance for polygenic traits

The key limitation of the genetic group model (2) is that all genetic groups are assumed to have the same additive genetic variance σ_A^2, and animal breeders have therefore suggested extensions that allow σ_A^2 to differ among groups (*e. g.* Lo *et al.*, 1993; Cantet & Fernando, 1995; García-Cortés & Toro, 2006). However, these methods quickly become computationally demanding because the respective models include terms to account for the segregation variance between any two genetic groups, thus $g(g-1)/2$ segregation variances in the presence of g genetic groups. The magnitude of these variances may be considerable in artificial breeding scenarios, *e. g.* when crossing genetically differentiated pure-bred lines (see *e. g.* Lande, 1981, Table 3), and if a trait is determined by one or only a few loci. To understand why, let us start by looking at a hypothetical trait that is determined by m loci. Following Lynch & Walsh (1998, equation 9.15), the segregation variance between two genetic groups (*e. g.* breeds) can be computed as

$$\sigma_S^2 = \frac{1}{2}\sum_{i=1}^{m}(\alpha_i^c)^2 \tag{3}$$

where α_i^c denotes the mean additive genetic difference between the groups due to locus i. Let us first look at an extreme example where only one locus determines a trait value, and where two populations or breeds differ in their mean breeding value by $\alpha^c = 1$ (which is the range of what we will find in the house sparrows example of Section 5). The segregation variance expected in a cross between these breeds is then given as $\sigma_S^2 = 1/2 \cdot 1 = 0.5$. Genome-wide association studies (GWAS) suggest however that complex (continuous) traits are mostly polygenic, thus the additive genetic component is not determined by a single locus (*e. g.* Goddard & Hayes, 2009; Flint & Mackay, 2009; Hill & Kirpatrick, 2010; Yang *et al.*, 2011b; Robinson

et al., 2013; Santure *et al.*, 2015; Silva *et al.*, 2017; Bouwman *et al.*, 2018). In fact, it is a fundamental assumption of quantitative genetics that phenotypic traits are determined by many genes that each contribute a small effect to trait variation, known as the "infinitesimal model" (Fisher, 1918; Bulmer, 1971). If, for example, 100 loci each contribute the same proportion of 1/100 to the overall group difference of 1 in the above example, the segregation variance reduces to

$$\sigma_S^2 = \frac{1}{2} \cdot 100 \cdot (1/100)^2 = 0.005 \ ,$$

which is exactly 1/100 of the segregational variance for a single-locus trait. Thus for any number of loci m, the segregation variance is $1/m$ of what it would be for a single locus, given that each locus contributes the same proportion of the effect. Even when considering that the locus-specific effect sizes are typically heterogeneous, the most influential loci often only explain a small proportion of the phenotypic variance (Yang *et al.*, 2011b; Wood *et al.*, 2014; Husby *et al.*, 2015). Thus, the segregation variance σ_S^2 for complex continuous traits is expected to be small compared to the total phenotypic variance in many study systems. In the following extension of the animal model we therefore ignore the segregation variance.

3.2 Animal model for heterogeneous additive genetic variances

To allow for heterogeneous additive genetic variances in the animal model, we extend model (2) by splitting the breeding value a_i into group-specific contributions (similar to García-Cortés & Toro, 2006), such that

$$y_i \ = \ \mu + \sum_{j=1}^{r} q_{ij} g_j + \sum_{j=1}^{r} a_{ij} + e_i \ , \tag{4}$$

with $\boldsymbol{a}_j^\top = (a_{1j}, \ldots, a_{nj})^\top \sim \mathsf{N}(\mathbf{0}, \sigma_{A_j}^2 \mathbf{A}_j)$ for all groups $j = 1, \ldots, r$, where $\sigma_{A_j}^2$ is the additive genetic variance in group j, and $\mathbf{A}_j$ is a group-specific relatedness matrix. The contribution a_{ij} to the breeding value of individual i can be interpreted as the part that is inherited from group j. We assume that the contributions a_{ij} are independent of each other, because they differ in genetic origin.

We illustrate the idea again for the case of two genetic groups. As above, we constrain the mean breeding value of group 1 (the reference group) to $g_1 = 0$ for identifiability reasons. In addition, the model must ensure that a breeding value component is zero ($a_{ij} = 0$) if the animal's genome obtains no contribution from the respective group j (thus if $q_{ij} = 0$), which will be ensured by appropriate choice of the covariance matrix $\mathbf{A}_j$ (details will follow in Section 3.3). Animals in genetic group 1 (*i. e.* $q_{i1} = 1$) then have a genetic effect $u_i = a_{i1}$ and animals in genetic group 2 (*i. e.*

$q_{i2} = 1$) receive $u_i = g_2 + a_{i2}$, and the breeding values are distributed as

$$(a_{11}, \ldots, a_{n1})^\top \sim \mathsf{N}(\mathbf{0}, \sigma^2_{A_1}\mathbf{A}_1) \ ,$$
$$(a_{12}, \ldots, a_{n2})^\top \sim \mathsf{N}(\mathbf{0}, \sigma^2_{A_2}\mathbf{A}_2) \ .$$

While it is relatively straightforward to formulate such a model, it is less obvious what the group-specific relatedness matrices $\mathbf{A}_j$ are. It is, for example, not valid to use $\mathbf{A}$ for $\mathbf{A}_j$, because the within-group relatedness structure is different from the overall relatedness. In addition, the total breeding value is now split into the sum $\sum_j a_{ij}$, with a_{ij} equal to the total breeding value only if an animal has $q_{ij} = 1$ for group j. The entries of $\mathbf{A}_j$ must therefore be scaled accordingly. We now turn to this issue.

3.3 Group-specific relatedness matrices

3.3.1 Decomposition of the relatedness matrix

To understand how to specify the group-specific relatedness matrices $\mathbf{A}_j$, let us first recall a result by Henderson (1976) for efficient calculation of inverse relatedness matrices. He suggested to decompose $\mathbf{A}$ by a *generalized Cholesky decomposition* into

$$\mathbf{A} = \mathbf{TDT}' \ , \tag{5}$$

where $\mathbf{T}$ is lower triangular matrix with transposed $\mathbf{T}'$, and $\mathbf{D} = \mathrm{Diag}(d_{11}, \ldots, d_{nn})$ is a diagonal matrix with entries $d_{11}, \ldots, d_{nn}$. This is equivalent to the Cholesky decomposition $\mathbf{A} = \mathbf{LL}'$ for a lower triangular matrix $\mathbf{L} = \mathbf{T}\sqrt{\mathbf{D}}$. A useful property of the decomposition (5) is that the matrices $\mathbf{T}$ and $\mathbf{D}$ have elegant interpretations: $\mathbf{T}$ traces the flow of alleles from one generation to the other, and the diagonal entries of $\mathbf{D}$ scale the Mendelian sampling variance (Mrode, 2005, p. 27).

Let us illustrate these properties with an example that we adapted from Mrode (2005, Table 2.1), without genetic groups. The pedigree is given in Table 1, with a corresponding graphical representation of parent-offspring relations (Figure 1a) and matrices $\mathbf{A}$, $\mathbf{T}$ and $\mathbf{D}$ (Figures 1 b-d), where the generalized Cholesky decomposition to obtain $\mathbf{T}$ and $\mathbf{D}$ was calculated with the function `gchol()` that is available in the `bdsmatrix` package (Therneau, 2014) in R (R Core Team, 2017), as pointed out *e. g.* by Gorjanc (2011). In this example, animals 1, 2 and 3 have unknown parents and are denoted as *founders* of the population. Each off-diagonal entry in $\mathbf{T}$ corresponds to the relatedness coefficient (expected relatedness) of individuals with their direct descendants (*i. e.* children, grandchildren etc.), where columns represent ancestors and rows descendants. For example, individual 1 is the parent of animals 4 and 5, thus the entries (4,1) and (5,1) in the matrix are 0.5. In addition, animal 6 is the

offspring of animals 4 and 5, thus the relatedness of 1 with 6 is also 0.5. Finally, the relatedness of 1 and 7 is 0.25. These considerations can be repeated for each column in the matrix, where all diagonal elements are 1 and all elements below the diagonal in the respective column correspond to the *expected* proportion of the genome that is transmitted from the respective ancestor to its direct descendants.

On the other hand, the diagonal entry d_{ii} for animal i in $\mathbf{D}$ is calculated as

$$d_{ii} = \begin{cases} 1 \ , & \text{if no parent is known,} \\ 1 - 0.25 - 0.25(F_p) \ , & \text{if one parent } p \text{ is known,} \\ 1 - 0.5 - 0.25(F_s + F_d) \ , & \text{if both parents } s \text{ and } d \text{ are known,} \end{cases}$$

where F_p, F_s and F_d are the pedigree-based inbreeding coefficients of the known parent(s) (Mrode, 2005, p. 28). We do not index the inbreeding coefficients with the animal identity (i) purely for notational simplicity. For later use we note that, in the absence of inbreeding, the diagonal entry is $(1 - 0.5p_i)$, with p_i corresponding to the proportion of i's ancestral genome that is known. Possible values are $p_i = 0, 0.5$ or 1 if no, one or two parents are known, respectively. This can be understood as follows: If, for example, one parent of an animal i is unknown, its predicted breeding value is 0.5 times the breeding value of the known parent, but the other half of its breeding value is unknown. The deviation from the predicted breeding value that could be obtained if both parents were known is absorbed by the Mendelian sampling deviation. The respective variance thus contains the Mendelian sampling variance *plus* a variance that is due to the unknown parent (Kennedy *et al.*, 1988). The more parents are unknown, the larger is this variance.

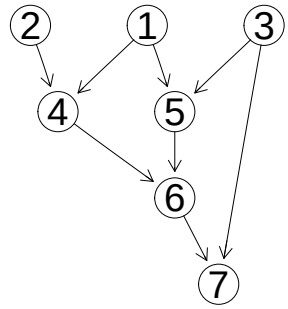
ID	**Dam**	**Sire**
1 (g_1)	NA	NA
2 (g_1)	NA	NA
3 (g_2)	NA	NA
4	1	2
5	1	3
6	5	4
7	6	3

Table 1: Pedigree example, adapted from Mrode (2005, Table 2.1).

3.3.2 T and D for genetic groups

In the presence of genetic groups, each unknown parent of an observed animal is assigned to one of the groups, and expected proportions of individual's genomes that originate from the respective genetic groups can be calculated from the pedigree

(a)



(b)

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0.25 \\ 0 & 1 & 0 & 0.5 & 0 & 0.25 & 0.125 \\ 0 & 0 & 1 & 0 & 0.5 & 0.25 & 0.625 \\ 0.5 & 0.5 & 0 & 1 & 0.25 & 0.625 & 0.3125 \\ 0.5 & 0 & 0.5 & 0.25 & 1 & 0.625 & 0.5625 \\ 0.5 & 0.25 & 0.25 & 0.625 & 0.625 & 1.125 & 0.6875 \\ 0.25 & 0.125 & 0.625 & 0.3125 & 0.5625 & 0.6875 & 1.125 \end{pmatrix}$$

(c)

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 1 & 0 & 0 \\ 0.5 & 0.25 & 0.25 & 0.5 & 0.5 & 1 & 0 \\ 0.25 & 0.125 & 0.625 & 0.25 & 0.25 & 0.5 & 1 \end{pmatrix}$$

(d)

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.46875 \end{pmatrix}$$

Figure 1: Graphical representation of the pedigree example (a) with the corresponding relatedness matrix $\mathbf{A}$ (b). The matrices from the decomposition $\mathbf{A} = \mathbf{TDT}'$ are given in (c) and (d). In the genetic group example, animals 1 and 2 are founders of genetic group 1, and animal 3 of genetic group 2.

(Quaas & Pollak, 1981; Wolak & Reid, 2017). For simplicity, we again consider the case with two groups, g_1 and g_2, and denote by $\mathbf{A}_1$ and $\mathbf{A}_2$ the respective relatedness matrices. These can be decomposed in the same way as $\mathbf{A}$ into

$$\mathbf{A}_1 = \mathbf{T_1 D_1 T_1}' \quad \text{and} \tag{6}$$

$$\mathbf{A}_2 = \mathbf{T_2 D_2 T_2}' \ , \tag{7}$$

with matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ describing the transmission of alleles through the generations, and Mendelian sampling variance matrices $\mathbf{D}_1$ and $\mathbf{D}_2$. The generalization to more than two groups is straightforward.

Let us assume in the pedigree example of Figure 1 that the parents of founder animals 1 and 2 belong to genetic group 1, and the parents of animal 3 to genetic group 2. This leads to proportional contributions of each genetic group to the genomes of the descending individuals as given by the matrix $\mathbf{Q}$ (Figure 2a), with columns $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ that contain the respective proportions of genetic origin from groups 1 and 2 for each individual. The transmission of alleles within each group is represented by the matrices $\mathbf{T}_j$ ($j = 1, 2$). They are designed such that animals with a certain proportion of genetic origin can only pass on the respective fraction of alleles. This means, for example, that an animal i with $q_{i1} = 0.5$ passes only a proportion of 0.25 (and not 0.5) of alleles to its offspring as part of genetic group 1, while another expected proportion of 0.25 is passed on to its offspring within group 2. The matrices $\mathbf{T}_j$ are thus obtained by scaling the respective entries in $\mathbf{T}$ by the respective group-proportions. This is

achieved by multiplying each row of $\mathbf{T}$ by $\boldsymbol{q}_j$ or, equivalently, by

$$\mathbf{T}_j = \mathbf{T} \cdot \mathrm{Diag}(\boldsymbol{q}_j) \ , \tag{8}$$

where $\mathrm{Diag}(\boldsymbol{q}_j)$ denotes a diagonal matrix with diagonal equal to $\boldsymbol{q}_j$. The matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ for our example are given in Figures 2b and c. Note that the diagonal of $\mathbf{T}_j$ corresponds to $\boldsymbol{q}_j$, which is the respective expected fraction of the genome that belongs to group j, and all entries in the respective column are scaled with that same value. Animal 3, for example, that belongs exclusively to group 2 (thus $q_{13} = 0$), has only 0 entries in the third column of $\mathbf{T}_1$, because it cannot transmit any group-1-alleles to its descendants. Animal 5, on the other hand, has $q_{15} = 0.5$, thus the fifth column in $\mathbf{T}$ is multiplied by 0.5 to obtain the respective column in $\mathbf{T}_1$.

(a)

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0.5 & 0.5 \\ 0.75 & 0.25 \\ 0.375 & 0.625 \end{pmatrix} = \begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 \end{pmatrix}$$

(b)

$$\mathbf{T}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0.25 & 0 & 0.5 & 0.25 & 0.75 & 0 \\ 0.25 & 0.125 & 0 & 0.25 & 0.125 & 0.375 & 0.375 \end{pmatrix}$$

(c)

$$\mathbf{T}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.25 & 0 & 0.25 & 0.25 & 0 \\ 0 & 0 & 0.625 & 0 & 0.125 & 0.125 & 0.625 \end{pmatrix}$$

Figure 2: Genetic group matrix $\mathbf{Q}$ and $\mathbf{T}_j$ matrices for the pedigree example. Group-specific proportions of the genome are stored in the $\mathbf{Q}$ matrix (a). Its columns can be used to derive the group-wise matrices $\mathbf{T_1}$ (b) and $\mathbf{T}_2$ (c) in our example.

Next, we need to find appropriate versions of $\mathbf{D}_1$ and $\mathbf{D}_2$. We noted in Section 3.3.1 that, in the absence of inbreeding, $d_{ii} = 1 - 0.5p_i$ with p_i representing the proportion of the ancestral genome that is known. To calculate the respective entries $d_{ii}^{(1)}$ and $d_{ii}^{(2)}$ in the group-specific matrices, we have to multiply p_i by the proportions of genetic origin q_{i1} and q_{i2}, because the respective product then corresponds to the ancestral proportions that are known *within* the respective group. In the case where only one parent is known, multiplication must be with the genetic proportion of the known parent, denoted here as $q_{ij}^{(p)}$, because only this respective part of the ancestral genome within group j is then known. This leads to

$$d_{ii}^{(j)} = \begin{cases} 1 \ , & \text{if no parent is known,} \\ 1 - 0.25 \cdot q_{ij}^{(p)} \ , & \text{if one parent is known,} \\ 1 - 0.5 \cdot q_{ij} \ , & \text{if both parents are known.} \end{cases}$$

Let us now also account for inbreeding, which may influence $d_{ii}^{(j)}$ for individuals with at least one parent known. Starting with animals that have both parents known, and rearranging the entries $d_{ii} = 1 - 0.5 - 0.25(F_s + F_d)$ in the original matrix $\mathbf{D}$, leads to

$$(F_s + F_d) = 2 - 4d_{ii} \ .$$

The knowledge of $(F_s + F_d)$ for each animal is useful to derive the entries $d_{ii}^{(j)}$ for group j in the presence of inbreeding: By scaling the effect of inbreeding with the group-specific proportions q_{ij} of an animal's genome, the entries are given as

$$\begin{aligned} d_{ii}^{(j)} &= 1 - 0.5 \cdot q_{ij} - 0.25 \cdot q_{ij}(F_s + F_d) && (9) \\ &= 1 - 0.5 \cdot q_{ij} - 0.25 \cdot q_{ij}(2 - 4d_{ii}) \\ &= 1 - q_{ij}(1 - d_{ii}) \ , && (10) \end{aligned}$$

where the third line is an algebraic simplification of the second line. The same calculation for animals with only one parent known, using $d_{ii} = 1 - 0.25 - 0.25(F_p)$ and solving for F_p, leads to the same formula with q_{ij} replaced by $q_{ij}^{(p)}$ from the known parent. Finally, if both parents are unknown (*i. e.* for $d_{ii} = 1$), the formula also leads to the correct value of $d_{ii}^{(j)} = 1$. Applying formula (10) to the above pedigree example leads to $\mathbf{D}_1$ and $\mathbf{D}_2$ (Figures 3a and b).

Formula (10) is simple and convenient. However, it provides only an approximation of the correct matrix entries, because in (9) we assumed that parental inbreeding can simply be scaled by the genetic group proportions q_{ij} (for two known parents) or $q_{ij}^{(p)}$ (for one known parent). Instead, the theoretically correct way to deal with parental inbreeding coefficients to derive $d_{ii}^{(j)}$ would be to use the actual *partial* (*i. e.* group-specific) parental inbreeding coefficients, denoted *e. g.* as $F_s^{(j)}$ or $F_d^{(j)}$ for parents s and d. These group-specific inbreeding coefficients contain only the inbreeding that emerge due to inbreeding *within genetic group* j, that is, they measure the probability that an individual is identical by descent for an allele that descended from founders within group j (Lacy *et al.*, 1996). The correct way to calculate $d_{ii}^{(j)}$ is thus given by

$$d_{ii}^{(j)} = \begin{cases} 1 \ , & \text{if no parent is known,} \\ 1 - 0.25 \cdot q_{ij}^{(p)} - 0.25(F_p^{(j)}) \ , & \text{if one parent } p \text{ is known,} \\ 1 - 0.5 \cdot q_{ij} - 0.25(F_s^{(j)} + F_d^{(j)}) \ , & \text{if both parents } s \text{ and } d \text{ are known.} \end{cases} \quad (11)$$

Obviously, this formula requires the calculation of group-specific inbreeding coefficients, which is computationally cumbersome. One way to obtain these coefficients is by first calculating *founder*-specific inbreeding coefficients that partition the total inbreeding coefficient F_i into the additive components F_{ik} from each founder animal k, as proposed by Lacy *et al.* (1996). Because partial contributions for all founders sum up

to F_i (*e. g.* Lacy *et al.*, 1996; Gulisija *et al.*, 2006), we can sum only over founders from genetic group j to obtain group-specific inbreeding coefficients $F_i^{(j)} = \sum_{k \in \text{group } j} F_{ik}$.

Let us illustrate the difference between the approximate method suggested in equation (10) and the correct formula for $d_{ii}^{(j)}$ given in (11) for our example from Figure 1. Animal 6 is the only parent in the pedigree with a non-zero inbreeding coefficient, which is $F_6 = 0.125$. However, because animals 1 and 2 are founders of group 1 and animal 3 is a founder of group 2, the pedigree reveals that inbreeding originates only from matings within group 1. Therefore, F_6 is split into group-specific inbreeding coefficients as $F_6^{(1)} = 0.125$ and $F_6^{(2)} = 0$. By plugging these values into (11) to estimate the respective values for animal 7 (which is the only animal that is affected by this change), we obtain $d_{7,7}^{(1)} = 0.812$ and $d_{7,7}^{(2)} = 0.656$, which are quite close to the approximate values $d_{7,7}^{(1)} = 0.801$ and $d_{7,7}^{(2)} = 0.668$ from Figures 3a and b. Note that in this paper we will continue to use the convenient and computationally efficient approximation (10) to scale the entries in $\mathbf{D}_j$, but we will illustrate the consequences of this approximation with simulations and for our application to the house sparrow example below and in Appendix 1.

(a)
$$\mathbf{D}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.75 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.625 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.801 \end{pmatrix}$$
(b)
$$\mathbf{D}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.75 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.875 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.668 \end{pmatrix}$$

Figure 3: Group-specific matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ for the example pedigree.

3.3.3 Properties of group-specific relatedness matrices

Once the components $\mathbf{T}_j$ and $\mathbf{D}_j$ for each group j are known, a simple matrix multiplication yields the group-specific relatedness matrices $\mathbf{A}_j = \mathbf{T}_j \mathbf{D}_j \mathbf{T}_j'$. These are given in Figure 4 for the two groups considered in our example. An important aspect is that both $\mathbf{A}_1$ and $\mathbf{A}_2$ contain columns and rows with all variances and covariances equal to zero, namely for animals i without a contribution from group j ($q_{ij} = 0$). While this is theoretically correct, because the respective breeding value is then $a_{ij} = 0$, the resulting matrices are singular. When it comes to implementation, the problem can be solved by replacing zeros on the diagonal by very small values, for example 10^{-6} or even 10^{-12}. The choice is not critical in our experience, but it would be prudent to check the robustness of the results to different values.

3.3.4 Scaling the inverse relatedness matrix

In practice it is usually the inverse relatedness matrix $\mathbf{A}^{-1}$, not $\mathbf{A}$, that is calculated and stored directly from the pedigree. Consequently, it is computationally more

(a)

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0.25 \\ 0 & 1 & 0 & 0.5 & 0 & 0.25 & 0.12 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 1 & 0.25 & 0.62 & 0.31 \\ 0.5 & 0 & 0 & 0.25 & 0.44 & 0.34 & 0.17 \\ 0.5 & 0.25 & 0 & 0.62 & 0.34 & 0.84 & 0.42 \\ 0.25 & 0.12 & 0 & 0.31 & 0.17 & 0.42 & 0.32 \end{pmatrix}$$

(b)

$$\mathbf{A}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0.5 & 0.25 & 0.62 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.44 & 0.22 & 0.36 \\ 0 & 0 & 0.25 & 0 & 0.22 & 0.16 & 0.21 \\ 0 & 0 & 0.62 & 0 & 0.36 & 0.21 & 0.68 \end{pmatrix}$$

Figure 4: Group-specific relatedness matrices for the example pedigree with entries rounded to two digits.

convenient, and in most cases much more efficient, to apply the appropriate scaling and transformation operations directly to $\mathbf{A}^{-1}$ to obtain group-specific inverses $\mathbf{A}_j^{-1}$. From (5) it follows that the inverse of $\mathbf{A}$ can be decomposed into

$$\mathbf{A}^{-1} = (\mathbf{T}^{-1})'\mathbf{D}^{-1}\mathbf{T}^{-1} \ , \tag{12}$$

where $\mathbf{D}^{-1} = \mathrm{Diag}(1/d_{11}, \ldots, 1/d_{nn})$. Using that $\mathbf{T}_j = \mathbf{T} \cdot \mathrm{Diag}(\boldsymbol{q}_j)$, and the decomposition of $\mathbf{A}_j$ as given *e. g.* in equation (6), some matrix algebra shows that

$$\begin{aligned} \mathbf{A}_j^{-1} &= (\mathbf{T}_j^{-1})'\mathbf{D}_j^{-1}\mathbf{T}_j^{-1} \\ &= (\mathbf{T}^{-1})' \underbrace{\mathrm{Diag}(1/\boldsymbol{q}_j)\mathrm{Diag}(1/\boldsymbol{d}_j)\mathrm{Diag}(1/\boldsymbol{q}_j)}_{\tilde{\mathbf{D}}_j^{-1}} \mathbf{T}^{-1} \ , \end{aligned} \tag{13}$$

where $\boldsymbol{d}_j = (d_{11}^{(j)}, \ldots, d_{nn}^{(j)})$ are the diagonal entries of $\mathbf{D}_j$ as derived from equation (10), and $\boldsymbol{q}_j$ is again the vector of genetic proportions inherited from group j. This illustrates that it is sufficient to calculate the (computationally relatively expensive) generalized Cholesky decomposition $\mathbf{A}^{-1}$ into $\mathbf{T}^{-1}$ and $\mathbf{D}^{-1}$ only once, and to derive $\tilde{\mathbf{D}}_j^{-1}$ with diagonal entries $1/(d_{ii}^{(j)} \cdot q_{ij}^2)$ to obtain $\mathbf{A}_j^{-1} = (\mathbf{T}^{-1})'\tilde{\mathbf{D}}_j^{-1}(\mathbf{T})^{-1}$ as in (13) for each group. Again, entries with $q_{ij} = 0$ are replaced by very small values, *e. g.* 10^{-12}, to avoid singularities.

4 Simulation

4.1 Generating data

To illustrate the performance of genetic group models with group-specific additive genetic variances we simulated data using the `simGG()` function from the R package `nadiv` (Wolak, 2012). The function allows generating pedigrees and phenotypes for a focal population (group 1) that receives a specified number of immigrants from another population in each generation (group 2). Group-specific mean breeding values and additive genetic variances can be determined by the user, and breeding values for the founder animals of both genetic groups are sampled from the respective dis-

tributions. Offspring breeding values are calculated from the parental mean, plus a Mendelian sampling deviation that depends on the additive genetic variance of the resident population, but there is no additional term that introduces a segregation variance (for more details, see Wolak, 2012). The simulation assumes random mating among individuals that currently live in the same population, thus offspring may inherit genetic components from both genetic groups due to immigration. The contributions q_{ij} from group j for animal i were calculated with the `ggcontrib()` function from the `nadiv` package. For data generated with group-specific mean breeding values and additive genetic variances, the appropriate underlying model for the analysis is

$$\begin{aligned} y_i &= \mu + q_{i2}g_2 + a_{i1} + a_{i2} + e_i \ , \\ \boldsymbol{a}_1^\top &\sim \mathsf{N}(\mathbf{0}, \sigma^2_{A_1}\mathbf{A_1}) \ , \quad \boldsymbol{a}_2^\top \sim \mathsf{N}(\mathbf{0}, \sigma^2_{A_2}\mathbf{A_2}) \ , \\ e_i &\sim \mathsf{N}(0, \sigma^2_E) \ , \end{aligned} \tag{14}$$

where we used the same notation as in Section 3, and fixed the focal group mean $g_1 = 0$ for identifiability reasons (thus the term $q_{i1}g_1$ is omitted from equation (14)). We simulated data according to three different scenarios, but always setting the population mean $\mu = 10$, group means $g_1 = 0$, $g_2 = 2$, group-specific variances $\sigma^2_{A_1} = 2$ and $\sigma^2_{A_2} = 3$ and residual variance $\sigma^2_E = 1$. Each scenario encompassed 10 non-overlapping generations.

Scenario 1 The carrying capacity of the population was set to 300 individuals. In each generation, 100 mating pairs were created by random sampling with replacement from the adults, and each pair contributed 4 offspring to the next generation. In addition, 30 immigrants were added to the population in each (except the first) generation. Finally, a subset of the offspring was randomly selected such that the population size always corresponded exactly to its carrying capacity.

Scenario 2 This scenario was the same as scenario 1, except that only 5 (instead of 30) immigrants were allowed in each of the non-overlapping generations, so that animals of the immigrant group were rare.

Scenario 3 In this scenario we used the same carrying capacity as above, 20 immigrants per generation, but we only allowed for 5 breeding pairs per generation that produced 60 offspring each. While this scenario has an immigration rate that lies between scenarios 1 and 2, the low number of breeding pairs induces higher inbreeding levels. This scenario is therefore suitable to illustrate the consequence of using approximation (10) to scale the group-specific Mendelian sampling variance matrices $\mathbf{D}_j$, which is only (potentially) critical in the presence of inbreeding, because the approximation affects only the scaling of parental inbreeding coefficients.

For each scenario, we generated 100 datasets, and each of them was analyzed with the genetic group animal model that accounted for heterogeneous additive genetic variances $\sigma^2_{A_1}$ and $\sigma^2_{A_2}$, as given in (14), and compared to the outcome of the standard genetic groups model that allowed for different mean breeding values, but only a single (homogeneous) variance in both groups, given as

$$\begin{aligned} y_i &= \mu + q_{i2}g_2 + a_i + e_i \ , \\ \boldsymbol{a}^\top &\sim \mathsf{N}(\mathbf{0}, \sigma^2_A \mathbf{A}) \ , \quad e_i \sim \mathsf{N}(0, \sigma^2_E) \ . \end{aligned} \tag{15}$$

For scenario 3 we also investigated how close the group-specific Mendelian sampling variance approximations for $d^{(j)}_{ii}$ from equation (10) are in comparison to the correct version given in (11). The group-specific inbreeding coefficients present in the correct formula were calculated with the software `GRain` (Baumung *et al.*, 2015, details are given in Appendix 1). Correlation coefficients ρ between the (correct) $d^{(j)}_{ii}$ values from (11) and the approximated values from formula (10) were calculated, and all simulations were analyzed with both versions for comparison.

Following the recommendation by He & Hodges (2008), we stored posterior modes (and not posterior means) of the variance components in each iteration. All models were fitted with integrated nested Laplace approximations (INLA, version from June 20, 2017), which provides a fast and accurate alternative to MCMC (Rue *et al.*, 2009), although it has so far only rarely been used for animals models (but see *e. g.* Holand *et al.*, 2013; Steinsland *et al.*, 2014; Roulin & Jensen, 2015). All variance components were given penalized complexity (PC) priors, which were suggested as valid alternatives to the (less recommended) gamma priors (Simpson *et al.*, 2017). The PC(u, α) prior has an intuitive parameterization: The prior probability for the standard deviation σ is given as $\mathsf{Pr}(\sigma > u) = \alpha$ (with $0 < \alpha < 1$). Here we used PC$(1, 0.05)$ priors for all variances (thus $\mathsf{Pr}(\sigma > 1) = 0.05$ *a priori*), but results were insensitive to this choice. All fixed effect parameters were assigned independent $\mathsf{N}(0, 10^4)$ priors. A short tutorial including R code to generate and analyze data for the models used here can be found in Appendix 4.

4.2 Simulation results

Using the appropriate model (14) with heterogeneous group variances resulted in estimates that were close to the variances used to generate the data in scenarios 1 and 3, while the model estimates in scenario 2 suffer from large uncertainty (Figure 5, left). In particular the variance $\sigma^2_{A_2}$ of the underrepresented immigrant population in scenario 2 was difficult to identify and biased towards $\sigma^2_{A_1}$ (Figure 5c). The results thus indicate that the genetic group model (14) is able to isolate approximately correct group-specific additive genetic variances, but that some caution is required if

representatives of a genetic group are rare in the data set: Group-specific additive genetic variances are then not completely identifiable.

When the simulated data were fitted using the genetic group model with a single, homoscedastic variance σ_A^2 (equation (15)), the resulting estimates were generally between the two simulated group-specific variances (Figure 5, right). In the presence of only few immigrants, the estimate tended to be close to $\sigma_{A_1}^2$ (scenario 2), while it tended towards $\sigma_{A_2}^2$ when there was more immigration (scenarios 1 and 3). This is as expected, and illustrates that genetic group models with a single, homogeneous variance will estimate a value in between the true group-specific additive genetic variances, with a tendency towards the variance of the more numerous group. These patterns were qualitatively similar when the additive genetic variances of the immigrant and resident population were switched, such that residents had larger additive genetic variance than immigrants (results not shown).

Simulation scenario 3 led to datasets with a mean inbreeding coefficient of $F = 0.10$. Interestingly, the comparison between the approximate versus the correct values in the $\mathbf{D}_1$ and $\mathbf{D}_2$ matrices shows that the approximation suggested in equation (10) leads to $d_{ii}^{(j)}$ values that are highly correlated with the correct values from equation (11). As an example, we found correlation coefficients $\rho \geq 0.988$ in three randomly selected simulation runs. In addition, using the correct $\mathbf{D}_j$ matrices led to distributions of estimated variances that are indistinguishable from the results when the approximations were used (see Figures S1 and S2 in Appendix 1).

5 Application to house sparrow data

5.1 Study population

As a proof of concept, we applied our method to empirical data from a long-term study of an insular house sparrow meta-population off the Helgeland coast in northern Norway. The study has been running continuously since 1993, and is used as a model system to examine ecological and evolutionary processes in fragmented vertebrate populations (*e. g.* Ringsby *et al.*, 2002; Jensen *et al.*, 2008; Pärn *et al.*, 2012; Baalsrud *et al.*, 2014). The islands in the meta-population differ in characteristics related to environmental conditions, habitat type and population size, with considerably larger and more stable populations on the five islands that are located closer to the mainland (denoted as *inner* islands) compared to the three islands located further away (denoted as *outer* islands). The ten remaining islands are summarized as *other* islands (see Figure S4 in Appendix 2 for an overview of the island system).

Small blood samples were collected from all captured birds on the eight inner and outer islands to provide DNA for Single Nucleotide Polymorphism (SNP) genotyping

on a 200K SNP array (see Lundregan *et al.*, 2018). Only successfully genotyped birds that were also measured for body mass and/or wing length were included in our animal model analyses. For inner and other islands, the dataset included phenotypic measurements taken during the breeding seasons since 1993 and 1995, respectively. Due to strong population bottlenecks on the outer islands in 2000 (Baalsrud *et al.*, 2014), only measurements taken since 2002 were used for the populations in the outer group. Details on how morphological measurements of wing length and body mass were taken on adult birds are given by *e. g.* Jensen *et al.* (2004, 2008).

Parentage analyses for the eight island populations in this study were carried out with the R package `SEQUOIA` (Huisman, 2017). Briefly, SNP genotype data of all adults recorded as present on any of the eight inner or outer islands during the years 1998-2013 (the inner group) or 2003-2013 (the outer group) were used in the parentage analyses. This resulted in a "meta-population pedigree" ($N = 3116$) spanning up to 14 generations, where both parents were known for 52.7%, one parent was known for 25.0%, and no parents were known for 22.3% of the individuals. Since `SEQUOIA` introduces *dummy parents* to preserve known relationships, *e. g.* sibling relationships, even when parents are not genotyped, a higher percentage of individuals had "known" parentages (81.0%, 5.5% and 13.5% with two, one or no parents known, respectively).

The genetic group analysis that we carry out here requires that each unknown parent (*i. e.* each founder) must be assigned to one of the genetic groups. That is, we must attribute unknown parents to the *inner*, *outer* or *other* island group to determine their genetic origin. This was done here by first identifying the natal (hatch) island of all individuals, either from ecological data or, if unavailable, by using genetic assignment procedures based on the SNP genotype data (see Saatoglu *et al.*, 2018). Of all individuals in our dataset, 1436, 481, and 64 individuals were assigned to a natal island in the inner, outer and other group, respectively (Saatoglu *et al.*, 2018). This information was then used to assign missing parents to the genetic group to which the hatch island of the respective individual belonged. As an example, if an individual with missing parents is known to be born on one of the inner islands, the respective missing parents were assumed to be founder individuals of the inner island group, and so on. Finally, because inbreeding depression is known to occur in our study system (Jensen *et al.*, 2007; Billing *et al.*, 2012), we accounted for any inbreeding effects on body mass and wing length (Reid *et al.*, 2006; Reid & Keller, 2010) by including each individual's genomic inbreeding coefficient F_{GRM} (Yang *et al.*, 2011a) as a covariate in all models fitted here, where F_{GRM} was estimated as described by Niskanen *et al.* (2018).

5.2 Analysis of wing length and body mass

With all unknown parents assigned to one of the genetic groups (*inner*, *outer*, *other*), each individual i obtained expected proportions of genetic origin q_{ij} from the three groups j by propagating the founder individual's genome through the pedigree using the `ggcontrib()` function. Note that the different groups are very unequal in sampling size, which can be seen from summation of genetic proportions over all animals within a group j, that is, $n_j = \sum_i q_{ij}$, which corresponds to equivalents of full animal genomes (Table 2). This is not surprising given the smaller population sizes (Baalsrud *et al.*, 2014), lower recapture rates (Holand *et al.*, 2016), and shorter time-series for populations in the outer islands group, and considering that there were no genotyping efforts on the other islands, thus we only see genome from the other islands if it was introduced via an immigration event to one of the inner or outer islands.

For the two traits investigated here, mass (in g) and wing length (in mm), we fitted separate models that accounted for sex (0=males, 1=females), inbreeding F_{GRM}, month of measurement (numeric with values 5, 6, 7, and 8) and age as fixed effects that were stored in matrix $\boldsymbol{X}$, and current island of residence where the measurement was taken (*island*), hatchyear (*year*), animal (*id*) and an independent residual term (e) as random factors. An individual i was included in the model if it had at least one recorded observation k of the respective trait. The model with group-specific mean and variances of the breeding values is thus given as

$$y_{ik} = \mu + \boldsymbol{X}_i\boldsymbol{\beta} + \underbrace{\sum_{j=2}^{3} q_{ij}g_j + \sum_{j=1}^{3} a_{ij}}_{u_i} + island_{ik} + year_i + id_i + e_{ik} \ , \tag{16}$$

where the total genetic contribution u_i is the sum of the weighted means g_j and the variability from the different genetic groups, as introduced in equation (4). The three genetic groups *inner*, *outer* and *other* are encoded as groups 1, 2 and 3, respectively, where the mean of the inner group was set to $g_1 = 0$ for identifiability reasons. The estimates g_2 and g_3 thus reflect differences in group means with respect to the inner population. The components a_{i1}, a_{i2} and a_{i3} are distributed with mean zero and heterogeneous additive genetic variances $\sigma^2_{A_{inner}}$, $\sigma^2_{A_{outer}}$ and $\sigma^2_{A_{other}}$, with dependency structures given by the group-specific relatedness matrices $\mathbf{A}_{inner}$, $\mathbf{A}_{outer}$ and $\mathbf{A}_{other}$ that were calculated as explained in Section 3. The results from fitting model (16) were compared to the standard genetic group model that only accounts for differences in group means, but with homogeneous additive genetic variance σ^2_A and dependency structure defined through the relatedness matrix $\mathbf{A}$. Both models were fitted to the data using INLA. All variances were given $\mathrm{PC}(1, 0.05)$ priors, and fixed effects parameters were assigned independent $\mathsf{N}(0, 10^4)$ priors.

Interestingly, the results indicate that the outer group has a somewhat smaller estimated additive genetic variance than the inner and other groups in the case of body mass (Table 2, left), while the estimated additive genetic variance is larger for the outer group for wing length (Table 2, right), although for both traits the respective 95% credible intervals (CI) are relatively large and overlap. The differences in additive genetic variance estimates could indicate that animals living on the three island groups show different levels of variation in the genes that affect the phenotypic traits, *e. g.* due to different allele frequencies, although we cannot rule out that additional shared environmental effects are confounded with additive genetic variances (Stopher *et al.*, 2012). In both cases, the variance estimate $\hat{\sigma}_A^2$ from the homogeneous model lies between the group-specific variance estimates, with a tendency towards the estimates for the inner population, which is by far the largest group. These results were also compared to the observed phenotypic variance $\hat{\sigma}_P^2$ of the two traits, where the respective group-specific phenotypic variances were only calculated using the pure-bred animals (*i. e.* those with $q_{ij} = 1$) in each group (Table 2). It is also interesting to note that the estimates for g_2 indicate that animals on the outer islands are lighter and have somewhat shorter wings than animals on the inner islands (Table 3), which is in agreement with earlier findings (Jensen *et al.*, 2004, 2008; Holand *et al.*, 2011). The remaining variance estimates of the model are not of primary interest and are thus given in Table S3 of Appendix 2.

All results presented here involved the approximate approach from formula (10) to scale the group-specific $\mathbf{D}_j$ matrices. To illustrate that this approximation is unproblematic, we repeated all calculations with the correct versions as given in equation (11), which again involved the gene dropping method provided in `GRain`. Details are given in Appendix 1, and all results remain essentially unchanged (see Figure S3 and Tables S1 and S2 in Appendix 1). Finally, it is worth reiterating that model (16) does not account for the three segregation variances that would occur between any two groups, because these are expected to be negligibly small. In addition, estimating these three additional variances would impose unrealistic requirements on these data. To illustrate that ignoring these variances is not critical, we also fitted a model with a segregation term for the sparrow example with only two genetic groups (inner and outer). Details on how to estimate segregation variances are given in Appendix 3. Importantly, the results in Table S4 confirm that segregation terms are indeed very small, and that their inclusion only leads to irrelevant changes in the results.

6 Discussion

We have introduced an extension of the animal model that allows for unequal additive genetic variances in the presence of multiple interbreeding genetic groups. Our method

	Body mass			Wing length		
	$\hat{\sigma}^2_A$	n_j or n	$\hat{\sigma}^2_P$	$\hat{\sigma}^2_A$	n_j or n	$\hat{\sigma}^2_P$
inner	1.56 (1.15, 2.06)	1572.5	5.72	1.77 (1.53, 2.13)	1571.0	5.34
outer	0.93 (0.42, 2.24)	267.3	5.65	2.29 (1.78, 3.44)	262.6	6.48
other	0.70 (0.26, 2.85)	131.3	4.87	1.00 (0.59, 2.69)	129.4	6.02
total	1.54 (1.19, 2.02)	1971	5.53	1.79 (1.60, 2.21)	1963	5.41

Table 2: Estimates (posterior modes) and 95% CIs (in brackets) of the three group-specific additive genetic variances ($\hat{\sigma}^2_A$) for inner, outer and other genetic groups, as well as for a single homogeneous variance across groups for the two traits body mass and wing length. The sample sizes denote the equivalent of full animal genomes that are present in the three genetic groups (n_j, for the model with heterogeneous variances) or in the total dataset for the respective trait (n, for the homogeneous model). For comparison, the phenotypic variances ($\hat{\sigma}^2_P$) in the three groups and for the total population, calculated only from the 1113, 148 and 33 pure-bred animals for mass and from the 1117, 146 and 33 pure-bred animals for wing length in the inner, outer and other groups.

	Body mass		Wing length	
	heterogeneous	homogeneous	heterogeneous	homogeneous
sex	0.46	0.48	−2.76	−2.77
	(0.29, 0.64)	(0.30, 0.65)	(−2.89, −2.63)	(−2.90, −2.64)
F_{GRM}	−1.23	−1.24	−1.32	−1.36
	(−3.09, 0.63)	(−3.09, 0.61)	(−2.69, 0.06)	(−2.74, 0.01)
month	−0.29	−0.30	−0.18	−0.19
	(−0.35, −0.24)	(−0.36, −0.24)	(−0.22, −0.15)	(−0.22, −0.15)
age	0.08	0.08	0.47	0.47
	(0.02, 0.14)	(0.02, 0.14)	(0.43, 0.50)	(0.43, 0.50)
g_2 (outer)	−0.83	−0.84	−0.44	−0.62
	(−1.29, −0.39)	(−1.38, −0.29)	(−0.82, −0.01)	(−1.09, −0.12)
g_3 (other)	−0.28	−0.18	0.08	−0.07
	(−0.84, 0.27)	(−0.85, 0.50)	(−0.34, 0.49)	(−0.64, 0.51)

Table 3: Posterior means and 95% CIs of the fixed effects for the animal models used for body mass and wing length. The estimates were extracted from models with either group-specific (heterogeneous) additive genetic variances, or a single (homogeneous) additive genetic variance.

requires that group-specific relatedness matrices $\mathbf{A}_j$ are derived. To this end, the full relatedness matrix $\mathbf{A}$ is decomposed into matrices $\mathbf{T}$ and $\mathbf{D}$, and simple algebraic scaling operations are used to derive the group-specific versions $\mathbf{T}_j$ and $\mathbf{D}_j$ that are then again multiplied to obtain $\mathbf{A}_j$. The method is computationally efficient, in particular when an (accurate) approximation for the group-specific Mendelian sampling variance

matrices $\mathbf{D}_j$ is used. Although genetic group animal models have been used before, in particular in animal and plant breeding setups, modeling heterogeneous additive genetic variances has so far been considered unfeasible (Wolak & Reid, 2017). Therefore, natural populations have only been analyzed with genetic group models that account for mean differences in additive genetic effects, but not for heterogeneous variances. Here we overcome this limitation by assuming that segregation variances are often negligible for polygenic traits. How often will this assumption hold? The segregation variance from formula (3) approximates zero whenever the *infinitesimal model*, which underlies the animal model and which posits that traits are determined by a large number of genes with small effects, holds approximately. This is likely the case for many polygenic, complex traits (Barton *et al.*, 2017). Our estimates of the segregation variance in the empirical house sparrow data set support the view that segregation variance may often be negligible. This result is mirrored in GWAS of the genetic architectures of body mass and wing length in house sparrows and other passerines, which revealed a polygenic basis for these traits, where any significant genomic region explains only a very small proportion of the phenotypic variance (Santure *et al.*, 2015; Silva *et al.*, 2017). Taken together, these results suggest that segregation variances can often be neglected in genetic group models, provided the traits of interest are truly very polygenic. When focal traits have a genetic architecture with only few causal genes with a large effect, omitting the segregational variance may however introduce a non-negligible bias in the estimated additive genetic variances. In such a case it is still possible to formulate a model that accounts for segregation variances, as explained in Appendix 3, although such models may quickly impose unrealistic demands on the data.

Estimating and disentangling variance components is generally known to be difficult, and it is particularly challenging for genetic group models with group-specific additive genetic variances. The problem is that the group-specific covariance matrices $\mathbf{A}_j$ are the sole sources of information that allow to discriminate the variance components, yet these matrices may be similar in the presence of many multi-bred animals. The results from the house sparrow example in Table 2 illustrate that group-specific variance estimates may suffer from larger uncertainty than a single homogeneous variance, especially when group sizes are small, as it is the case for the *outer* and *other* groups. Allowing for heterogeneous variances is therefore only advocated if there is an actual need for them. While this is often not evident in advance, the user may want to fit the homogeneous and the heterogeneous models to determine if the more complex model is needed. A possible "objective" way to find this out might be via the use of information criteria, such as the deviance information criterion (DIC) for Bayesian models (Spiegelhalter *et al.*, 2002), although we do not generally recommend to rely on it (for some DIC criticism see *e. g.* Plummer, 2008; Hodges, 2014). Interestingly,

in the sparrow example the DIC for the heterogeneous model for mass was 4 units larger than for the model with a single variance, while it was −15 units smaller for wing length. Thus there is some weak indication that the more complex model is not needed when body mass is analyzed, while there appears to be some improvement for wing length, which is consistent with the observation that the estimated variances show less overall differences for body mass than for wing length (Table 2). In principle, the variance estimates can also be used to assess group-specific versions of heritability (Falconer & Mackay, 1996) or evolvability (Houle, 1992), where the additive genetic variance is replaced by the group-specific version $\sigma^2_{A_j}$.

A general limitation of genetic group models is that parent-offspring relations are needed to propagate genetic contributions from founder individuals through the pedigree to determine $\mathbf{Q}$. Genomic relatedness matrices alone are therefore not sufficient to fit genetic group models, but a combination of genomic and pedigree information, the latter possibly inferred from genetic data (Ko & Nielsen, 2017), may provide a powerful basis for genetic group models.

The proposed extension of genetic group models will be useful for any study population that is structured into subpopulations, given that sufficient information on migration or crossbreeding events is available. In particular, the fact that group-specific additive genetic variances can be estimated for subpopulations that are not completely isolated might also be useful when interest centers around the dependency of additive genetic variance on the effective population size, a relation that is of pivotal interest in evolutionary and conservation biology. Finally, the method may provide a starting point for the estimation of temporal or spatial variation of additive genetic variance.

Acknowledgements

We are grateful to Dr. Timothée Bonnet for comments to the manuscript. This work was mainly funded by the Faculty of Science of the University of Zurich (S.M., L.F.K.). We thank the many researchers, students, and fieldworkers for their contributions to collecting the empirical data on house sparrows, and laboratory technicians for assistance with laboratory analyses. This study was supported by grants from the Norwegian Research Council (programmes STORFORSK, Strategic University Program in Conservation Biology, project 221956 to H.J.), the Norwegian Directorate for Nature Management, the EU-commission (project METABIRD), and the Academy of Finland (project 295204 to A.N.). This work was also partly supported by the Research Council of Norway through its Centres of Excellence funding scheme, project 223257).

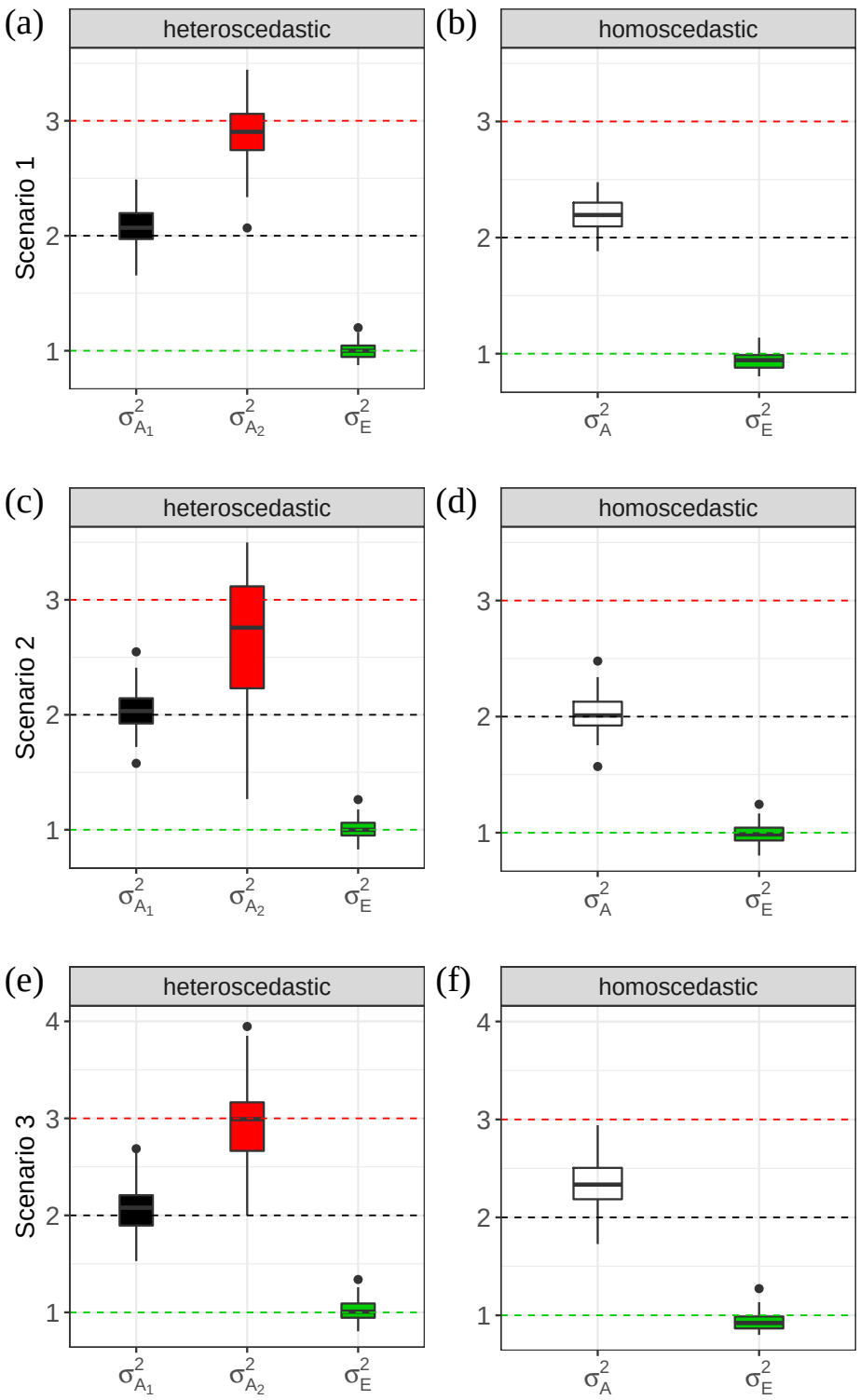


Figure 5: Results from 100 iterations for simulation scenarios 1–3. The boxplots represent the distributions of estimated variances (posterior modes) from a model with genetic groups and heteroscedastic additive genetic variances $\sigma^2_{A_1}$ and $\sigma^2_{A_2}$ (left panel), compared to the results from a model that only allowed for a single homogeneous variance σ^2_A (right panel). Dashed lines indicate the references that were used to generate the data (black: $\sigma^2_{A_1}$, red: $\sigma^2_{A_2}$, green: σ^2_E).

References

Alfonso, L. & Estany, J. (1999) An expression of mixed model equations to account for different means and variances in the base population. *Genetics Selection Evolution*, **31**, 105–113. https://doi.org/10.1051/gse:19990201.

Baalsrud, H.T., Sæther, B.E., Hagen, I.J., Myhre, A.M., Ringsby, T.H., Pärn, H. & Jensen, H. (2014) Effects of population characteristics and structure on estimates of effective population size in a house sparrow metapopulation. *Molecular Ecology*, **23**, 2653–2668. https://doi.org/10.1111/mec.12770.

Barton, N., Etheridge, A. & Véber, A. (2017) The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, **118**, 50 – 73. https://doi.org/10.1016/j.tpb.2017.06.001.

Baumung, R., Farkas, J., Boichard, D., Mészáros, G., Sölkner, J. & Curik, I. (2015) GRAIN: a computer program to calculate ancestral and partial inbreeding coefficients using a gene dropping approach. *Journal of Animal Breeding and Genetics*, **132**, 100–108. https://doi.org/10.1111/jbg.12145.

Billing, A.M., Lee, A.M., Skjelseth, S., Borg, A.A., Hale, M.C., Slate, J., Pärn, H., Ringsby, T.H., Sæther, B.E. & Jensen, H. (2012) Evidence of inbreeding depression but not inbreeding avoidance in a natural house sparrow population. *Molecular Ecology*, **21**, 1487–1499. https://doi.org/10.1111/j.1365-294X.2012.05490.x.

Bonnet, T., Wandeler, P., Camenisch, G. & Postma, E. (2017) Bigger is fitter? Quantitative genetic decomposition of selection reveals an adaptive evolution decline of body mass in a wild rodent population. *PLOS Biology*, **15**, e1002592. https://doi.org/10.1371/journal.pbio.1002592.

Bouwman, A.C., Daetwyler, H.D., Chamberlain, A.J. & et al. (2018) Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nature Genetics*, **50**, 362–367. https://doi.org/10.1038/s41588-018-0056-5.

Bulmer, M.G. (1971) The effect of selection on genetic variability. *American Naturalist*, **105**, 201–211.

Cantet, R.J.C., Birchmeier, A.N., Santos-Cristal, M.G. & de Avila, V.S. (2000) Comparison of restricted maximum likelihood and method R for estimating heritability and predicting breeding value under selection. *Journal of Animal Science*, **78**, 2554 – 2560.

Cantet, R.J.C. & Fernando, R.L. (1995) Prediction of breeding values with additive animal models for crosses from 2 populations. *Geneticis Selection Evolution*, **17**, 323 – 334.

Charmantier, A., Kruuk, L.E.B. & Garant, D. (2014) *Quantitative Genetics in the Wild*. Oxford University Press, Oxford.

de Villemereuil, P., Morrissey, M.B., Nakagawa, S. & Schielzeth, H. (2017) Fixed effect variance and the estimation of the heritability: Issues and solutions. *Journal of Evolutionary Biology*. In press. https://doi.org/10.1111/jeb.13232.

Dong, M.C., Van Vleck, L.D. & Wiggans, G.R. (1988) Effect of relationships on estimation of variance components with an animal model and restricted maximum likelihood. *Journal of Dairy Science*, **71**, 3047 – 3052. https://doi.org/10.1111/jeb.13232.

Elzo, M.A. (1990) Recursive procedures to compute the inverse of multiple trait additive genetic covariance matrix in inbred and non inbred multibreed populations. *Journal of Animal Science*, **68**, 1215–1228.

Falconer, D.S. & Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*. Longman Group Ltd, Harlow, 4 edition.

Fisher, R.A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of The Royal Society of Edinburgh*, **52**, 399–433. https://doi.org/10.1017/S0080456800012163.

Flint, J. & Mackay, T.F. (2009) Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Research*, **19**, 723–733. https://doi.org/10.1101/gr.086660.108.

García-Cortés, L.A. & Toro, M.Á. (2006) Multibreed analysis by splitting the breeding values. *Genetics Selection Evolution*, **38**, 601. https://doi.org/10.1051/gse:2006024.

Goddard, M.E. & Hayes, B.J. (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Review Genetics*, **10**, 381 – 391. https://doi.org/10.1038/nrg2575.

Gorjanc, G. (2011) Setup up the inverse of additive relationship matrix in R [Blog post]. Retrieved from http://ggorjan.blogspot.no/2011/08/setup-up-inverse-of-additive.html.

Gulisija, D., Gianola, D., Weigel, K. & Toro, M. (2006) Between-founder heterogeneity in inbreeding depression for production in Jersey cows. *Livestock Science*, **104**, 244 – 253. https://doi.org/10.1016/j.livsci.2006.04.007.

He, Y. & Hodges, J.S. (2008) Point estimates for variance-structure parameters in Bayesian analysis of hierarchical models. *Computational Statistics & Data Analysis*, **52**, 2560 – 2577. https://doi.org/10.1016/j.csda.2007.08.021.

Henderson, C.R. (1984) *Applications of Linear Models in Animal Breeding.* University of Guelph Press, Guelph, Canada.

Henderson, C. (1976) Simple method for computing inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*, **32**, 69–83. https://doi.org/10.2307/2529339.

Hill, W.G. & Kirpatrick, M. (2010) What animal breeding has taught us about evolution. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 1–19. https://doi.org/10.1146/annurev-ecolsys-102209-144728.

Hodges, J.S. (2014) *Richly Parameterized Linear Models: Additive, Time Series, and Spatial Models Using Random Effects.* Chapman & Hall/CRC, Boca Raton.

Hoffmann, A.A., Sgrò, C.M. & Kristensen, T.N. (2017) Revisiting adaptive potential, population size, and conservation. *Trends in Ecology & Evolution*, **32**, 506 – 517. https://doi.org/10.1016/j.tree.2017.03.012.

Holand, A.M., Jensen, H., Tufto, J. & Moe, R. (2011) Does selection or genetic drift explain geographic differentiation of morphological characters in house sparrows *Passer domesticus*? *Genetics Research*, **93**, 367–379. https://doi.org/10.1017/S0016672311000267.

Holand, A.M., Steinsland, I., Martino, S. & Jensen, H. (2013) Animal models and integrated nested Laplace approximations. *G3*, **3**, 1241–1251.

Holand, H., Kvalnes, T., Gamelon, M., Tufto, J., Jensen, H., Pärn, H., Ringsby, T.H. & Sæther, B.E. (2016) Spatial variation in senescence rates in a bird metapopulation. *Oecologia*, **181**, 865–871. https://doi.org/10.1007/s00442-016-3615-4.

Houle, D. (1992) Comparing evolvability and variability of quantitative traits. *Genetics*, **130**, 195–204.

Huisman, J. (2017) Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Molecular Ecology Resources*, **17**, 1009–1024. http://dx.doi.org/10.1111/1755-0998.12665.

Husby, A., Kawakami, T., Rönnegård, L., Smeds, L., Ellegren, H. & Qvarnström, A. (2015) Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life-history trait. *Pro-*

ceedings of the Royal Society of London B: Biological Sciences, **282**, 20150156. https://doi.org/10.1098/rspb.2015.0156.

Jensen, H., Bremset, E.M., Ringsby, T.H. & Sæther, B.E. (2007) Multilocus heterozygosity and inbreeding depression in an insular house sparrow metapopulation. *Molecular Ecology*, **16**, 4066–4078. https://doi.org/10.1111/j.1365-294X.2007.03452.x.

Jensen, H., Sæther, B.E., Ringsby, T.H., Tufto, J., Griffith, S.C. & Ellegren, H. (2004) Lifetime reproductive success in relation to morphology in the house sparrow Passer domesticus. *Journal of Animal Ecology*, **73**, 599–611. https://doi.org/10.1111/j.0021-8790.2004.00837.x.

Jensen, H., Steinsland, I., Ringsby, T.H. & Sæther, B.E. (2008) Evolutionary dynamics of a sexual ornament in the house sparrow (Passer domesticus): The role of indirect selection within and between sexes. *Evolution*, **62**, 1275–1293. https://doi.org/10.1111/j.1558-5646.2008.00395.x.

Kennedy, B.W., Schaeffer, L.R. & Sorensen, D.A. (1988) Genetic properties of animal models. *Journal of Dairy Science*, **71**, 17–26. https://doi.org/10.1016/S0022-0302(88)79975-0.

Ko, A. & Nielsen, R. (2017) Composite likelihood method for inferring local pedigrees. *PLoS Genetics*, **13**, e1006963. https://doi.org/10.1371/journal.pgen.1006963.

Kruuk, L. (2004) Estimating genetic parameters in natural populations using the "animal model". *Philosophical Transactions of the Royal Society of London*, **359**, 873–890. https://doi.org/10.1098/rstb.2003.1437.

Lacy, R.C., Alaks, G. & Walsh, A. (1996) Hierarchical analysis of inbreeding depression in Peromyscus polionotus. *Evolution*, **50**, 2187–2200. https://doi.org/10.1111/j.1558-5646.1996.tb03609.x.

Lande, R. (1988) Genetics and demography in biological conservation. *Science*, **241**, 1455–1460.

Lande, R. (1981) The minimum number of genes contributing to quantiative variation between and within populations. *Genetics*, **99**, 541–553.

Lo, L.L., Fernando, R.L. & Grossman, M. (1993) Covariance between relatives in multibreed populations: additive model. *Theoretical and Applied Genetics*, **87**, 423–430. https://doi.org/10.1007/BF00215087.

Lundregan, S.L., Hagen, I.J., Gohli, J., Niskanen, A.K., Kemppainen, P., Ringsby, T.H., Kvalnes, T., Pärn, H., Rønning, B., Holand, H., Ranke, P.S., Båtnes, A.S., Selvik, L.K., Husby, A., Sæther, B.E. & Jensen, H. (2018) Investigating the genetic architecture of bill morphology in a free-living house sparrow metapopulation using a 200K SNP array. In review.

Lynch, M. & Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

Mackay, T.F.C. (1981) Genetic variation in varying environments. *Genetical Research*, **37**, 79–93. https://doi.org/10.1017/S0016672300020036.

Mrode, R.A. (2005) *Linear Models for the Prediction of Animal Breeding Values*. CABI Publishing, Oxfordshire, 2 edition.

Niskanen, A.K., Billing, A., Hagen, I.J., Muff, S., Huisman, J., Holand, H., Rønning, B., Kvalnes, T., Pärn, H., Ringsby, T.H., Lien, S., Sæther, B.E., Husby, A. & Jensen, H. (2018) A genome-wide study of inbreeding depression in a wild house sparrow metapopulation. In preparation.

Pärn, H., Ringsby, T.H., Jensen, H. & Sæther, B.E. (2012) Spatial heterogeneity in the effects of climate and density dependence on dispersal in a house sparrow metapopulation. *Proceedings of the Royal Society B*, **279**, 144–152. https://doi.org/10.1098/rspb.2011.0673.

Plummer, M. (2008) Penalized loss functions for Bayesian model comparison. *Biostatistics*, **9**, 523–539. https://doi.org/10.1093/biostatistics/kxm049.

Quaas, R.L. (1988) Additive genetic model with groups and relationships. *Journal of Dairy Science*, **71**, 91 – 98. https://doi.org/10.1016/S0022-0302(88)79986-5.

Quaas, R.L. & Pollak, E.J. (1981) Modified equations for sire models with groups. *Journal of Dairy Science*, **64**, 1868–1872. https://doi.org/10.3168/jds.S0022-0302(81)82778-6.

R Core Team (2017) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reid, J.M., Arcese, P. & Keller, L.F. (2006) Intrinsic parent-offspring correlation in inbreeding level in a song sparrow (*Melospiza melodia*) population open to immigration. *The American Naturalist*, **168**, 1–13. https://doi.org/10.1086/504852.

Reid, J.M. & Keller, L.F. (2010) Correlated inbreeding among relatives: occurrence, magnitude and implications. *Evolution*, **64**, 973–985. https://doi.org/10.1111/j.1558-5646.2009.00865.x.

Ringsby, T.H., Sæther, B.E., Tufto, J., Jensen, H. & Solberg, E.S. (2002) Asynchronous spatiotemporal demography of a house sparrow metapopulation in a correlated environment. *Ecology*, **83**, 561–569. https://doi.org/10.1890/0012-9658(2002)083[0561:asdoah]2.0.co;2.

Robinson, M.R., Santure, A.W., Decauwer, I., Sheldon, B.C. & Slate, J. (2013) Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Molecular Ecology*, **22**, 3963–3980. https://doi.org/10.1111/mec.12375.

Roulin, A. & Jensen, H. (2015) Sex-linked inheritance, genetic correlations and sexual dimorphism in three melanin-based colour traits in the barn owl. *Journal of Evolutionary Biology*, **28**, 655–666. https://doi.org/10.1111/jeb.12596.

Rue, H., Martino, S. & Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society Series B (Methodological)*, **71**, 319–392.

Saatoglu, D., Niskanen, A.K., Kuismin, M.O., Hagen, I.J., Ranke, P.S., Araya-Ajoy, Y., Myhre, A.M., Holand, H., Kvalnes, T., Rønning, B., Pärn, H., Sommerli, S., Lien, S., Ringsby, T.H., Sæther, B.E., Husby, A., Sillanpää, M.J. & Jensen, H. (2018) Identification of dispersers in a natural bird metapopulation by using high-density SNP genotype data and genetic assignment methods. In preparation.

Santure, A.W., Poissant, J., De Cauwer, I., Van Oers, K., Robinson, M.R., Quinn, J.L., Groenen, M.A.M., Visser, M.E. & Sheldon, B.C. (2015) Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. *Molecular Ecology*, **24**, 6148–6162. https://doi.org/10.1111/mec.13452.

Silva, C.N.S., McFarlane, S.E., Hagen, I.J., Rönnegård, L., Billing, A.M., Kvalnes, T., Kemppainen, P., Rønning, B., Ringsby, T.H., Sæther, B.E., Qvarnström, A., Ellegren, H., Jensen, H. & Husby, A. (2017) Insights into the genetic architecture of morphological traits in two passerine bird species. *Heredity*, **119**, 197 – 205. https://doi.org/10.1038/hdy.2017.29.

Simpson, D., Rue, H., Riebler, A., Martins, T.G. & Sørbye, S.H. (2017) Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, **32**, 1–28. https://doi.org/10.1214/16-STS576.

Speed, D. & Balding, D.J. (2015) Relatedness in the post-genomic era: Is it still useful? *Nature Review Genetics*, **16**, 33–44. https://doi.org/10.1038/nrg3821.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, **64**, 583–639. https://doi.org/10.1111/1467-9868.02022.

Steinsland, I., Larsen, C.T., Roulin, A. & Jensen, H. (2014) Quantitative genetic modeling and inference in the presence of nonignorable missing data. *Evolution*, **68**, 1735–1747. https://doi.org/10.1111/evo.12380.

Stopher, K.V., Walling, C.A., Morris, A., Guinness, F.E., Clutton-Brock, T.H., Pemberton, J.M. & Nussey, D.H. (2012) Shared spatial effects on quantitative genetic parameters: Accounting for spatial autocorrelation and home range overlap reduces estimates of heritability in wild red deer. *Evolution*, **66**, 2411–2426. https://doi.org/10.1111/j.1558-5646.2012.01620.x.

Therneau, T. (2014) *bdsmatrix: Routines for Block Diagonal Symmetric Matrices*. R package version 1.3-2.

Wilson, A.J. (2008) Why h^2 does not always equal $\mathrm{V}_A/\mathrm{V}_P$? *Journal of Evolutionary Biology*, **21**, 647 – 650. https://doi.org/10.1111/j.1420-9101.2008.01500.x.

Wilson, A., Reale, D., Clements, M., Morrissey, M., Postma, E., Walling, C., Kruuk, L. & Nussey, D. (2010) An ecologist's guide to the animal model. *Journal of Animal Ecology*, **79**, 13–26. https://doi.org/10.1111/j.1365-2656.2009.01639.x.

Wolak, M.E. (2012) nadiv: an R package to create relatedness matrices for estimating non-additive genetic variances in animal models. *Methods in Ecology and Evolution*, **3**, 792–796. https://doi.org/10.1111/j.2041-210X.2012.00213.x.

Wolak, M.E. & Reid, J.M. (2017) Accounting for genetic differences among unknown parents in microevolutionary studies: How to include genetic groups in quantitative genetic animal models. *Journal of Animal Ecology*, **86**, 7–20. https://doi.org/10.1111/1365-2656.12597.

Wood, A.R., Esko, T., Yang, J. & et al (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, **46**, 1173 – 1186.

Wright, S. (1922) Coefficients of inbreeding and relationship. *The American Naturalist*, **56**, 330–338.

Wright, S. (1968) *Genetic and Biometric Foundations*, volume 1 of *Evolution and the Genetics of Populations*. The University of Chicago Press, Chicago.

Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. (2011a) GCTA: A tool for genome-wide complex trait analysis. *Americal Journal of Human Genetics*, **88**, 76–82.

Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G., Hill, W.G., Landi, M.T., Alonso, A., Lettre, G., Lin, P., Ling, H., Lowe, W., Mathias, R.A., Melbye, M., Pugh, E., Cornelis, M.C., Weir, B.S., Goddard, M.E. & Visscher, P.M. (2011b) Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics*, **43**, 519 – 525. https://doi.org/10.1038/ng.823.