GSE Genetics Selection Evolution

Optimum contribution selection for conserved populations with historic migration

Robin Wellmann[*], Sonja Hartwig and Jörn Bennewitz

Abstract

Background: In recent decades, local varieties of domesticated animal species have been frequently crossed with economically superior breeds which has resulted in considerable genetic contributions from migrants. Optimum contribution selection by maximizing gene diversity while constraining breeding values of the offspring or *vice versa* could eventually lead to the extinction of local breeds with historic migration because maximization of gene diversity or breeding values would be achieved by maximization of migrant contributions. Therefore, other objective functions are needed for these breeds.

Results: Different objective functions and side constraints were compared with respect to their ability to reduce migrant contributions, to increase the genome equivalents originating from native founders, and to conserve gene diversity. Additionally, a new method for monitoring the development of effective size for breeds with incomplete pedigree records was applied. Approaches were compared for Vorderwald cattle, Hinterwald cattle, and Limpurg cattle. Migrant contributions could be substantially decreased for these three breeds, but the potential to increase the native genome equivalents is limited.

Conclusions: The most promising approach was constraining migrant contributions while maximizing the conditional probability that two alleles randomly chosen from the offspring population are not identical by descent, given that both descend from native founders.

Background

Many local varieties of domesticated animal species have been established in the last centuries. However, due to agricultural innovations since the beginning of the 19th century and subsequent intensification of production, many landraces are no longer adapted to their changing environments [1,2]. They have been crossed with superior breeds in order to improve the economic value of the breeding stock. Gene flow usually occured from the economically most important breeds to the landraces, but not backwards. Consequently, most historic breeds are now extinct and the remaining ones have considerable genetic contributions from a small number of economically superior breeds. Efforts are needed to prevent the remaining historic breeds and their gene pools to become extinct. Conservation efforts can have different objectives. Objectives of breeding programs can be to breed back

the historic breeds by removing genetic contributions of migrants, to conserve the breeds in their present appearance, or to increase the economic values. In any case, genetic contributions arising from more frequent breeds are not subject to conservation efforts since their genes are widespread.

Meuwissen [3] proposed to maximize the expected mean breeding value of the offspring while constraining its gene diversity to a predefined value. A related but not equivalent approach is to maximize the gene diversity in the offspring with or without constraining its expected mean breeding value to a predefined value. In this paper, the latter approach is applied and generalized. This approach seems more appropriate for conserved populations because for these populations the focus is on conservation. In general, the method consists of calculating an optimum contribution c_a (or the desired number of offspring) for each breeding individual a such that the offspring population maximizes an appropriate objective function ϕ under some side conditions.

*Correspondence: r.wellmann@uni-hohenheim.de
Institute of Animal Husbandry and Animal Breeding, University of Hohenheim, D-70599 Stuttgart, Germany

In the classical approach [4] (Approach A) the gene diversity *GD* in the offspring $O(\mathbf{c})$ is maximized, where the vector $\mathbf{c}$ contains the genetic contribution of each breeding individual to the offspring population. Thus, $\phi_A(\mathbf{c}) = GD(O(\mathbf{c}))$. Gene diversity of a population is the probability that two alleles randomly chosen from the population are not identical by descent (IBD). However, this objective function may be not appropriate for conserved populations because maximization of gene diversity could be achieved by maximization of genetic contributions of migrants. Thus, this approach could eventually lead to extinction of the native breeds. Gene diversity should not fall below a certain level in order to avoid inbreeding depression. Gene diversity is, however, not the parameter that should be maximized in conserved populations. In conserved populations, we are interested in the conservation of alleles that come from native founders, as migrant alleles usually originate from non-endangered breeds. That is, we want to maximize the probability ϕ_B that both alleles are not IBD and descended from native founders (Approach B), or the probability ϕ_C that both alleles are not IBD and at least one of them descended from native founders (Approach C). We also considered the possibility of maximizing the conditional probability ϕ_D that both alleles are not IBD, given that both descended from native founders (Approach D). For Approach D, we constrained the mean migrant contribution in the offspring population.

Lacy [5] introduced the concept of founder genome equivalents (FGE). The FGE of a population is the minimum number of unrelated founders that would be needed to establish a population that has the same gene diversity as the population under study. Recall that gene diversity is the probability that two alleles chosen at random are not IBD. However, a more important parameter to characterize the value of a breed for conservation purposes is the conditional probability that two randomly chosen alleles are not IBD, given that both descended from native founders. We call it the conditional gene diversity of the population. Large conditional gene diversity indicates that many native founder alleles have been retained in the population even though they may be at low frequencies. This has led to the following definition of the native genome equivalents (NGE) of a population as the minimum number of unrelated founders that would be needed to establish a population that has the same conditional gene diversity as the population under study. It can be interpreted as the FGE that originate from native founders and that are still present in the population. Besides maintaining the economic value of the breed, the main objective of a conservation program for a population with historic migration is to maximize the NGE and to minimize the genetic contributions of migrants simultaneously.

In this paper, we compare objective functions ϕ_A, ϕ_B, ϕ_C and ϕ_D with respect to their ability to conserve the gene diversity, to increase the FGE originating from native founders (i.e. the NGE), and to decrease the genetic contributions of migrants. Algorithms for solving these optimization problems are also derived and implemented in the R package *PedAnalysis*. Methods were applied and effective sizes were calculated for three German cattle breeds: Vorderwald, Hinterwald and Limpurg.

Methods
Definitions
Since the methods were applied to populations with overlapping generations, all definitions are based on birth cohorts rather than generations. A birth cohort J is a set of individuals born in a particular time interval, e.g. the individuals B_t born in year t, or the population P_t at time t [6]. Since the date of death is unknown in most cases, the population P_t consists of all individuals up to a particular age T. This age T could be the average age of individuals when their last offspring was born, or, for simplicity, it could be the generation interval I. Thus, population P_t consists of all individuals born in the time interval $[t - T, t]$.

The **gene diversity** $GD(J)$ of birth cohort J is the probability that two alleles chosen at random from the birth cohort are not IBD. We can write

$$GD(J) = P(X_J \neq Y_J),$$

where alleles X_J and Y_J are randomly chosen with replacement from birth cohort J, and founder alleles are assumed to be pairwise different. An equivalent representation is $GD(J) = 1 - \bar{f}_J$, where $\bar{f}_J$ is the average coancestry in birth cohort J.

Each allele descends from a particular founder. Take $\mathcal{A}$ to be the set of founder alleles. We distinguish between native founders and migrants, whereby a native founder is a founder that is not a migrant. A native founder is typically an individual with unknown pedigree that belongs to the population and was born before a certain date t_s. A migrant is typically an individual that either comes from an other population (breed), or an individual with unknown pedigree that was born after the date. The date t_s could be chosen shortly after establishment of the stud book when a sufficient portion of the population was recorded. We can write

$$\mathcal{A} = \mathcal{F} \cup \mathcal{M},$$

where $\mathcal{F}$ is the set of alleles that come from native founders and $\mathcal{M}$ is the set of alleles that come from migrants.

We define the **conditional gene diversity** $condGD(J)$ of birth cohort J as the conditional probability that two

alleles randomly chosen from the birth cohort are not IBD, given that both descend from native founders. That is,

$$condGD(J) = P(X_J \neq Y_J | X_J \in \mathcal{F}, Y_J \in \mathcal{F}).$$

The **founder genome equivalents** *FGE(J)* of birth cohort *J* is defined as the minimum number of founders that would be needed to establish a population that has the same gene diversity as the individuals in birth cohort *J*. It can be computed as

$$FGE(J) = \frac{1}{2\overline{f}_J} = \frac{1}{2(1 - GD(J))},$$

see [4]. Analogously, we define the **native genome equivalents** *NGE(J)* of birth cohort *J* as the minimum number of founders that would be needed to create a population that has the same conditional gene diversity as the individuals in birth cohort *J*. We have

$$NGE(J) = \frac{1}{2(1 - condGD(J))}.$$

However, a problem with this definition is that native founders of the population are assumed to be unrelated, which is not true. As a consequence, in the first generation the NGE would be almost as large as the total population size. However, due to the invalid assumption of unrelated founders, the limited effective size causes the NGE to decrease tremendously shortly after the last native founders have entered the population. In order to avoid this artifact, we extrapolate the history of the breed back in time and use as the reference population not the founders listed in the stud book, but the population at an earlier time t_0. That is, all individuals are assumed to be unrelated in year t_0. In the applications, the base year was $t_0 = 1800$. We define the conditional gene diversity of an age cohort J_t at time $t \geq t_s$ with respect to base year t_0 as

$$condGD_{t_0}(J_t) = \left(1 - \frac{1}{2\,histN_e}\right)^{\frac{t_s - t_0}{I}} \frac{condGD(J_t)}{condGD(P_{t_s})},$$

where P_{t_s} is the population at time t_s, I is the generation interval, and $histN_e$ is the historic effective size of the population. The historic effective size can be estimated from marker data [7]. The term that defines the conditional gene diversity is the product of two factors. The first is the estimated gene diversity in the population at time t_s, and the second is the factor by which the conditional gene diversity decreased between t_s and t. Consequently, the **NGE with respect to base year** t_0 can be calculated as

$$NGE_{t_0}(J_t) = \frac{1}{2(1 - condGD_{t_0}(J_t))}.$$

A further parameter that can be of interest is the **effective size** of the population. The effective size $N_e([t_1, t_2])$ of a population within a time interval $[t_1, t_2]$ is the size of an idealized random mating population of constant size that causes the same decrease of gene diversity as the true population within $\frac{t_2 - t_1}{I}$ generations. However, in breeds with steady gene flow from other populations, the gene diversity does not decrease below a certain level, so this definition of the effective size does not make much sense for populations with migration. Therefore, we use a slightly different definition. We define the **native effective size** $N_{eN}([t_1, t_2])$ as the size of an idealized random mating population of constant size that causes the same decrease of the *conditional* gene diversity $condGD(P_t)$ as the true population within $\frac{t_2 - t_1}{I}$ generations. The effective population size at time t, defined as $N_{eN}(t) = \lim_{\epsilon \to 0} N_{eN}([t - \epsilon, t + \epsilon])$, was calculated as described in [8], except that it was calculated from the conditional gene diversity. The native effective size quantifies the decrease of genome equivalents originating from native founders because the NGE depend only on the conditional gene diversity, as can be seen from the previous two equations. In a population without migration, N_e and N_{eN} are equal. However, in a population with steady gene flow from other populations, N_{eN} is smaller than N_e because the gene diversity approaches a plateau level, so $N_e(t)$ goes to infinity.

The population P_t at time t, which consists of all individuals up to an age of T years, has gene diversity $GD(P_t)$, native genome equivalents $NGE_{t_0}(P_t)$, and genetic contribution $C_{\mathcal{F}}(P_t)$ from native founders. Note that $C_{\mathcal{F}}(J) = P(X_J \in \mathcal{F})$, so $C_{\mathcal{F}}(J)$ is the probability that a randomly chosen allele from age cohort J descends from a native founder. Besides monitoring of these quantities, a major task for a conservation program is the calculation of optimal genetic contributions for the breeding individuals that maximize the conditional gene diversity in the offspring and simultaneously maximize the genetic contribution from native founders in the offspring. Moreover, a sufficient level of gene diversity must be maintained in order to avoid inbreeding depression. In general, however, the quantities $NGE_{t_0}(J)$ and $C_{\mathcal{F}}(J)$ cannot be maximized simultaneously, so an objective function is needed that considers each appropriately.

The usual approach (Approach A) for populations without migration is the calculation of genetic contributions $\mathbf{c}_t^A$ for the breeding individuals of population P_t such that the gene diversity

$$\phi_A(J) = GD(J) = P(X_J \neq Y_J)$$

is maximized by a hypothetical (infinitely large) offspring population $O_t(\mathbf{c}_t^A)$. This approach is called minimum kinship selection [9]. Note that the gene diversity $GD(O_t(\mathbf{c}_t^A)) = \phi_A(O_t(\mathbf{c}_t^A))$ of the hypothetical offspring is known as the potential diversity of the population at time t [6]. A more appealing approach for populations

with migration is to use genetic contributions $\mathbf{c}_t^B$ for the breeding individuals such that the probability

$$\phi_B(J) = P(X_J \neq Y_J \text{ and } X_J \in \mathcal{F} \text{ and } Y_J \in \mathcal{F})$$

is maximized by the resulting offspring population $O_t(\mathbf{c}_t^B)$. This is the probability that two randomly chosen alleles from the offspring are not IBD and are from native founders (Approach B). As a third approach, we consider maximization of the probability that two randomly chosen alleles from the offspring are not IBD and at least one of them descends from a native founder (Approach C). In this case, genetic contributions $\mathbf{c}_t^C$ for the breeding individuals are calculated such that the offspring population $O_t(\mathbf{c}_t^C)$ maximizes

$$\phi_C(J) = P(X_J \neq Y_J \text{ and } (X_J \in \mathcal{F} \text{ or } Y_J \in \mathcal{F})).$$

Finally, we consider maximizing the conditional gene diversity in the offspring population. That is, genetic contributions $\mathbf{c}_t^D$ for the breeding individuals were calculated such that the conditional probability

$$\phi_D(J) = P(X_J \neq Y_J \mid X_J \in \mathcal{F} \text{ and } Y_J \in \mathcal{F})$$

is maximized. This approach is intuitively appealing because it maximizes NGE. It has, however, the disadvantage that the conditional gene diversity can be large even for offspring populations with very large migrant contributions. This is due to conditioning on the event that the randomly chosen alleles X_J and Y_J originate from native founders. This can be seen as follows. Take a solution $\mathbf{c}_t^D$ of the optimization problem and suppose that at least one migrant is a potential breeding individual. Then it can be shown mathematically that the genetic contribution of this migrant to the offspring population can be arbitrarily increased without changing the value of the objective function. Thus, the solution of the optimization problem may be not unique, and one solution maximizes migrant contributions. In order to avoid this, we put an additional constraint on the maximum permissible value for the genetic contribution from migrants to the offspring population.

Computations

To calculate the parameters defined in the previous section, the following quantities are needed. First, the coancestry $f_{i,j}$ is needed for each pair of individuals i,j. It is the probability that two alleles randomly chosen from the individuals are IBD. That is,

$$f_{i,j} = P(X_i = X_j),$$

where allele X_i is randomly chosen from the two alleles of individual i at a particular locus.

Now we define an equivalence relation on the set of founder alleles. Two alleles x_i, x_j are equivalent ($x_i \equiv_M x_j$)

if they are IBD or if both are migrant alleles. For two alleles randomly chosen from individuals i,j, the probability for this to occur is

$$
\begin{aligned}
f_{i,j}^M &= P(X_i \equiv_M X_j) \\
&= P\left(X_i = X_j \text{ or } (X_i \in \mathcal{M}, X_j \in \mathcal{M})\right).
\end{aligned}
$$

A second equivalence relation is defined as follows. Two alleles x_i, x_j are equivalent ($x_i \equiv_{FM} x_j$) if both are native founder alleles or if both are migrant alleles. For two alleles randomly chosen from individuals i,j, the probability for this to occur is

$$
\begin{aligned}
f_{i,j}^{FM} &= P(X_i \equiv_{FM} X_j) \\
&= P\left((X_i \in \mathcal{F}, X_j \in \mathcal{F}) \text{ or } (X_i \in \mathcal{M}, X_j \in \mathcal{M})\right).
\end{aligned}
$$

These probabilities have the advantage that they can easily be computed with existing software, e.g. with function *kinship()* from the R-package *kinship*. For calculation of $f_{i,j}^M$, the parents of all migrants were identified with the same dummy individual and for this individual a pedigree with several generations of selfing was added. The coancestry of individuals i,j, computed from this extended pedigree is equal to $f_{i,j}^M$. Equality holds only approximately because only a finite number of generations of selfing was added. For calculation of $f_{i,j}^{FM}$, the parents of all migrants were identified with one single dummy individual, the parents of all native founders were identified with another single dummy individual, and for both individuals pedigrees with several generations of selfing were added. The coancestry of individuals i,j, computed from this extended pedigree, is equal to $f_{i,j}^{FM}$. For example, consider two full sibs i,j whose sire is a migrant and whose dam is a native founder. Their coancestry is $f_{ij} = \frac{1}{4}$, but $f_{ij}^M = \frac{3}{8}$, and $f_{ij}^{FM} = \frac{1}{2}$.

Let $\mathbf{f}_{P_t}$ be the $N_t \times N_t$ coancestry submatrix for the N_t individuals from population P_t that is obtained from the true pedigree (i.e., $\mathbf{f}_{P_t} = (f_{ij})_{i,j \in P_t}$). The $N_t \times N_t$ matrix that contains the probabilities $f_{i,j}^M$ for each pair of individuals i,j from population P_t is denoted as $\mathbf{f}_{P_t}^M = (f_{ij}^M)_{i,j \in P_t}$, and the $N_t \times N_t$ matrix that contains the probabilities $f_{i,j}^{FM}$ is denoted as $\mathbf{f}_{P_t}^{FM} = (f_{ij}^{FM})_{i,j \in P_t}$. That is, rows and columns that correspond to individuals not born in time interval $[t-T, t]$ and dummy individuals were excluded from the matrix.

Additionally, the N_t-dimensional vector $\mathbf{C}_t = (C_{t1}, \ldots, C_{tN_t})^T$ is needed and contains the genetic contribution of native founders for each individual of population P_t. Note that $C_{\mathcal{F}}(P_t)$ is the mean of vector $\mathbf{C}_t$. Let $\bar{f}_{P_t}, \bar{f}_{P_t}^M$, and $\bar{f}_{P_t}^M$ be the means of the respective matrices. It is well known that the gene diversity can be computed as [4]

$$GD(P_t) = 1 - \bar{f}_{P_t}.$$

Proofs of all numbered equations are presented in Additional file 1, in which it is shown that the conditional gene diversity satisfies

$$condGD(P_t) = \frac{\bar{f}_{P_t}^{FM} - \bar{f}_{P_t}^{M}}{C_{\mathcal{F}}(P_t) - \frac{1 - \bar{f}_{P_t}^{FM}}{2}}. \tag{1}$$

Let $O_t^N(\mathbf{c})$ be an arbitrary (hypothetical) offspring population of size N that is obtained from population P_t such that each breeding individual $a \in P_t$ has genetic contribution c_a to the offspring population. The probability that an allele randomly chosen from the offspring population descends from a native founder is

$$C_{\mathcal{F}}(O_t(\mathbf{c})) = C_{\mathcal{F}}(O_t^N(\mathbf{c})) = \mathbf{c}^T \mathbf{C}_t, \tag{2}$$

and the conditional gene diversity in the offspring population is

$$condGD(O_t(\mathbf{c})) = \lim_{N \to \infty} condGD(O_t^N(\mathbf{c}))$$

$$= \frac{\mathbf{c}^T (\mathbf{f}_{P_t}^{FM} - \mathbf{f}_{P_t}^{M}) \mathbf{c}}{\mathbf{c}^T \mathbf{C}_t - \frac{1 - \mathbf{c}^T \mathbf{f}_{P_t}^{FM} \mathbf{c}}{2}}. \tag{3}$$

It is well known that

$$\lim_{N \to \infty} \phi_A(O_t^N(\mathbf{c})) = 1 - \mathbf{c}^T \mathbf{f}_{P_t} \mathbf{c},$$

so the optimum contributions $\mathbf{c}_t^A$ for the breeding individuals with respect to objective function ϕ_A minimize $\mathbf{c}^T \mathbf{f}_{P_t} \mathbf{c}$ under side conditions $c_a \geq 0$ and $\sum_a c_a = 1$. Additional side conditions can be added to fulfil biological and practical requirements. Moreover, we have

$$\lim_{N \to \infty} \phi_B(O_t^N(\mathbf{c})) = \mathbf{c}^T (\mathbf{f}_{P_t}^{FM} - \mathbf{f}_{P_t}^{M}) \mathbf{c}, \tag{4}$$

so the optimum contributions $\mathbf{c}_t^B$ for the breeding individuals with respect to objective function ϕ_B minimize $\mathbf{c}^T (\mathbf{11}^T - (\mathbf{f}_{P_t}^{FM} - \mathbf{f}_{P_t}^{M})) \mathbf{c}$ under the side conditions described above, where $\mathbf{1}$ is a vector with ones. Since

$$\lim_{N \to \infty} \phi_C(O_t^N(\mathbf{c})) = 1 - \mathbf{c}^T \mathbf{f}_{P_t}^{M} \mathbf{c}, \tag{5}$$

the optimum contributions $\mathbf{c}_t^C$ for the breeding individuals with respect to objective function ϕ_C minimize $\mathbf{c}^T \mathbf{f}_{P_t}^{M} \mathbf{c}$ under the side conditions. Finally, we have

$$\lim_{N \to \infty} \phi_D(O_t^N(\mathbf{c})) = \frac{\mathbf{c}^T (\mathbf{f}_{P_t}^{FM} - \mathbf{f}_{P_t}^{M}) \mathbf{c}}{\mathbf{c}^T \mathbf{Q}_t \mathbf{c}}, \tag{6}$$

where $\mathbf{Q}_t = \frac{1}{2} \left(\mathbf{C}_t \mathbf{1}^T + \mathbf{1} \mathbf{C}_t^T - \mathbf{11}^T + \mathbf{f}_{P_t}^{FM} \right)$ is a $N_t \times N_t$ matrix. This function was maximized under the side conditions described above. Moreover, the additional side constraint $\mathbf{c}^T \mathbf{C}_t \geq c_{\mathcal{F}}$ was applied, where $c_{\mathcal{F}}$ is the minimum permissible contribution of native founders to the offspring population. This is a quadratic fractional programming problem with linear constraints, so the objective function could have multiple local maxima. As

mentioned in the previous section, one solution of the optimization problem maximizes migrant contributions, so the inequality constraint could be replaced by the equality constraint $\mathbf{c}^T \mathbf{C}_t = c_{\mathcal{F}}$. For each offspring population J that satisfies this equality constraint, the objective function (i.e. the conditional gene diversity) satisfies

$$\phi_D(J) \approx \frac{\phi_B(J)}{P(X_J \in \mathcal{F})^2} = \frac{\phi_B(J)}{c_{\mathcal{F}}^2} \propto \phi_B(J),$$

where the approximation is exact if the events $X_J \in \mathcal{F}$ and $Y_J \in \mathcal{F}$ are independent. Therefore, an approximate solution was obtained by maximizing objective function ϕ_B under the additional constraint $\mathbf{c}^T \mathbf{C}_t = c_{\mathcal{F}}$. The resulting contributions for the breeding individuals were used as starting values for general nonlinear optimization in order to obtain the exact solution. In the applications, the threshold value $c_{\mathcal{F}}$ was quite arbitrarily chosen as the 75% quantile of the genetic contributions from native founders to individuals in the population. The same quantile was used for all breeds and years in order to make the results comparable. Results could be improved by choosing breed dependent threshold values.

We used the interior point method *ipop* in R-package *kernlab* (see [10]) for objective functions ϕ_B and ϕ_D, whereas for objective functions ϕ_A and ϕ_C with positive definite matrices we used *solve.QP* from R-package *quadprog*. It implements the dual method of Goldfarb and Idnani [11,12].

Materials

Only three local cattle varieties of Baden and Württemberg in the south-west of Germany have been preserved from extinction. These are the Vorderwald cattle, Hinterwald cattle, and Limpurg cattle. Other local breeds were replaced by Simmentaler Fleckvieh after their introduction at the beginning of the 19th century because the small landraces were not suitable for tillage [1].

The small *Hinterwald cattle* could be preserved as an almost pure breed until the beginning of the 20th century [13,14] because the poor soil quality in its region of origin was not suitable for larger breeds. Nevertheless, this breed adopted the colour of the Simmentaler Fleckvieh during the 19th century [15]. The Hinterwald cattle were occasionally crossed with the Vorderwald cattle [16] and with Fleckvieh.

The red-and-white marked, colour-sided [17] *Vorderwald cattle* were frequently crossed with Simmentaler cattle. Consequently, the white stripe along the back became rare already around 1900 [16]. After the Second World War, Vorderwald cattle were also crossed with Ayrshire, Red Holstein and Montbéliard cattle in order to improve milk yield. These crosses were registered as Vorderwald cattle. Extinction probabilities for Vorderwald and Hinterwald cattle were estimated by [18].

The yellow coloured *Limpurg cattle* were not only frequently crossed with Simmentaler cattle [19], but also occasionally with Braunvieh and Gelbvieh cattle [15] in order to increase body size. Nevertheless, the population size decreased dramatically. Only 17 Limpurg cows were registered in 1967, so the breeding association was dissolved. Several Limpurg cattle, however, were rediscovered in 1986 and a new stud book was established. Not only Limpurg cattle were registered, but also Fleckvieh crosses, and some Gelbvieh and Glan-Donnersberger bulls [16].

The data consisted of the pedigrees and additional information on 25 412 Hinterwald cattle, 185 315 Vorderwald cattle, and 4 150 Limpurg cattle. Vorderwald cattle without offspring were removed from the data in order to reduce the data set. Pedigrees of Hinterwald and Vorderwald cattle trace back only to 1948 because the stud books were renewed after the Second World War. Pedigrees of Limpurg cattle trace back only to 1970. Cattle from other breeds were considered to be migrants. Additionally, Hinterwald and Vorderwald cattle with unknown pedigree born after $t_s = 1970$ were also considered migrants, although some may have purebred ancestors. Limpurg cattle with unknown pedigree were considered to be migrants if they were born after $t_s = 1988$. The generation intervals were similar for the three breeds (unpublished results). Here, we assumed a generation interval of $I = 5.3$ years for all breeds.

Results

The left hand side of Figure 1 shows the development of the native effective size N_{eN} for the three breeds. Around 1990, the effective size of Limpurg cattle was only about 20, which was due to the small population size. However in most cases, the effective size was above 50 for all three breeds. In 2011, 7952 Vorderwald cows, 2328 Hinterwald cows, and 471 Limpurg cows were registered. Interestingly, there appeared to be no relationship between the effective size and the total population size when the number of individuals exceeds the minimum number required to reach an N_e of approximately 50.

The right hand side of Figure 1 shows for each breed how the genetic contributions of migrants changed over time. Migrant contributions are shown for the true population P_t and for the hypothetical offspring populations that would be obtained if optimum contribution selection were applied to population P_t. The solid lines show that migrant contributions increased steadily for all three breeds. The dashed line for offspring A shows that all three breeds would become extinct if optimum contribution selection were used to maximize the gene diversity in the offspring. In contrast, objective functions ϕ_B and ϕ_C would reduce migrant contributions substantially by more than 50% in all three breeds. According to the constraint
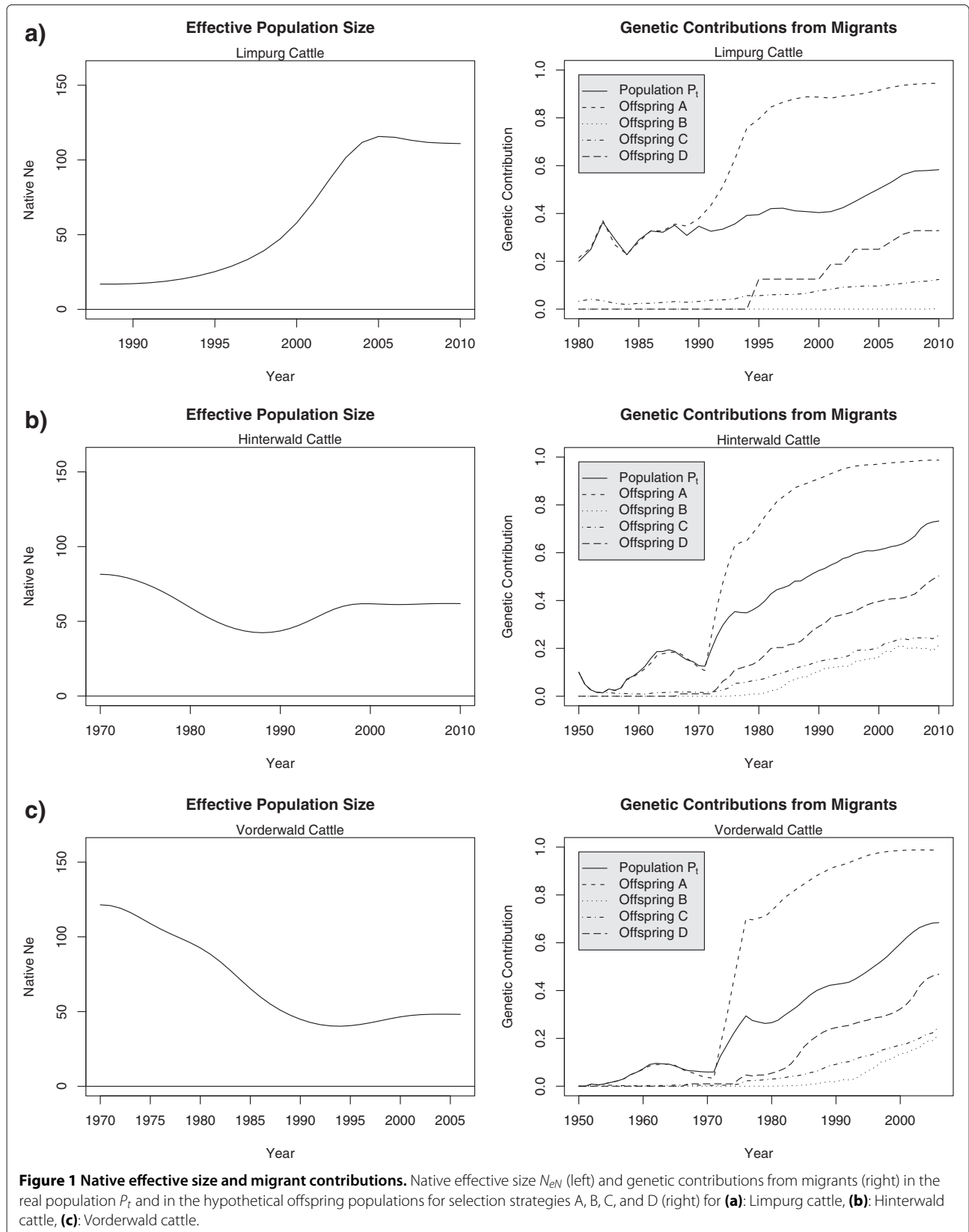
applied for objective function ϕ_D, the corresponding line shows the 25% quantile of the migrant contributions in the population.
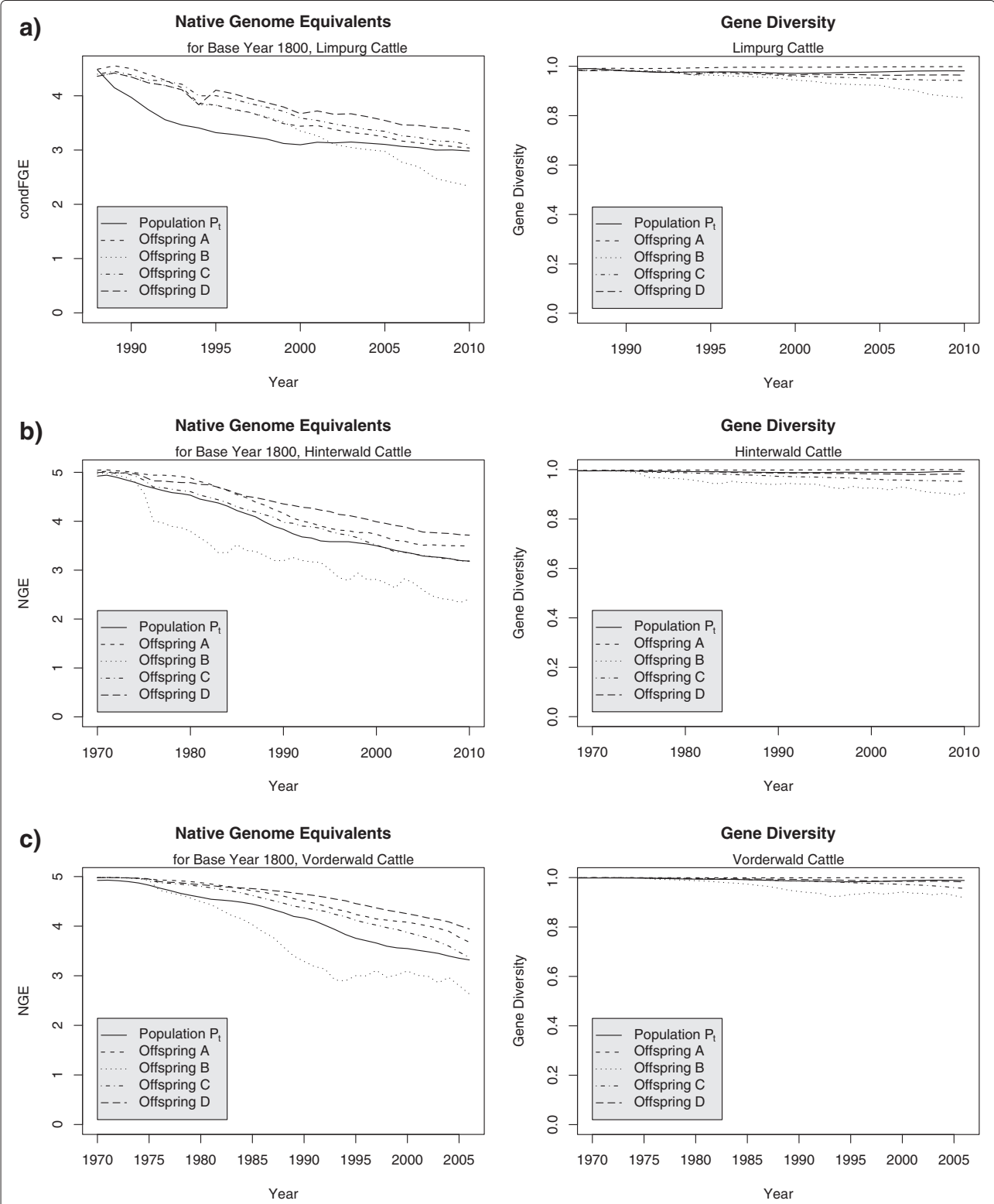
The left hand side of Figure 2 shows the development of NGE for the true population and for the hypothetical offspring populations. We used the year $t_0 = 1800$ as the base year. The historic N_e is not known for these breeds. In the figure, we assumed a historic N_e of 150 for each breed, which is in good accordance with the results obtained by [20] for various cattle breeds during this period of time. For the population in 2005, the computed NGE with respect to base year t_0 was 3.1 for Limpurg cattle, 3.3 for Hinterwald cattle, and 3.2 for Vorderwald cattle. For comparison, the NGE computed under the assumption of unrelated founders was 7.3 for Limpurg cattle, 8.3 for Hinterwald cattle, and 8.0 for Vorderwald cattle. Figure 2 shows that the NGE would decrease by using objective function ϕ_B. This suggests that the individuals with the smallest migrant contributions are closely related, so they share the same founder alleles. The use of objective function ϕ_C would cause a small increase of the NGE for all three breeds. If the constraint on migrant contributions is not too serious, then objective function ϕ_D would cause the largest increase in NGE. However, the potential to increase NGE is limited.

The right hand side of Figure 2 shows the changes in gene diversity. It can be seen that the gene diversity is high for all three breeds. This is caused by migration. Note that the native effective population size quantifies the decrease of genome equivalents arising from native founders, so the gene diversity can be constant (or increase due to migration) even if the native effective population size is small. As expected, optimum contribution selection with objective functions ϕ_B, ϕ_C, or ϕ_D would cause a moderate but an acceptable loss of gene diversity.

Discussion

Most of the time, the native effective size N_{eN} was above 50 for the three breeds and due to migration, N_e was larger than N_{eN}. An effective size of at least 50 is considered acceptable, although an N_e of 100 is recommended to be on the safe side [21]. Many cattle breeds have effective sizes between 50 and 100 regardless of the total population size. Therefore, in order to conserve the overall gene diversity, it is generally recommended to conserve a large number of breeds with small population sizes rather than a small number of breeds with large population sizes. In this case, different alleles would be preserved in different subpopulations. These populations can be used as resources to identify advantageous genes that can be introgressed into commercial populations. Conserved populations must be sufficiently large to allow for this. However, breeds that are close to the economic viability threshold and populations that are expected to occupy

Figure 1 Native effective size and migrant contributions. Native effective size N_{eN} (left) and genetic contributions from migrants (right) in the real population P_t and in the hypothetical offspring populations for selection strategies A, B, C, and D (right) for **(a)**: Limpurg cattle, **(b)**: Hinterwald cattle, **(c)**: Vorderwald cattle.

Figure 2 Native genome equivalents and gene diversities. Genome equivalents originating from native founders NGE (left) and gene diversity (right) in the real population P_t and in the hypothetical offspring populations for selection strategies A, B, C, and D for **(a)**: Limpurg cattle, **(b)**: Hinterwald cattle, **(c)**: Vorderwald cattle.

niches that are different from that of established commercial breeds, should have larger population sizes in order to enable a sufficient selection response. Examples of the importance of farm animal genetic resources are the introgression of the polled gene into economically important cattle breeds, the introduction of indicine cattle breeds to South America because of their adaption to extreme environments, and introgression of genes for disease resistance into highly productive susceptible breeds [22].

The current N_{eN} of the Vorderwald cattle was smaller than the estimates of the effective size obtained by [23] with other methods. The reason is probably that other methods do not distinguish between migrants and native founders. Genome equivalents arising from native founders are likely to decline faster than those arising from migrants because migrants are usually from economically superior breeds. The sufficiently large N_{eN} show that for all three breeds, migration from other breeds was much larger than it was needed to avoid unacceptably high inbreeding depression. As a consequence, these breeds share only a small portion of their genes with the corresponding historic breeds of the same name. We showed that it is still possible to substantially increase the genetic contribution from the historic breeds by optimum contribution selection.

For optimum contribution selection, the choice of the objective function was crucial. Maximization of gene diversity (Approach A) turned out to substantially increase the migrant contributions and thus would lead to the extinction of these breeds. Approach B has the desired effect to substantially decrease the migrant contributions but does not put enough weight on the conservation of gene diversity. It is not recommended because it would reduce NGE and cause the largest loss of gene diversity. Approach C is recommended for conserved populations because for all three breeds the use of this objective function substantially decreased the migrant contributions, increased the NGE, and caused only a moderate decrease of gene diversity. Approach D can also be recommended, although it requires choice of a threshold for the migrant contributions. If the threshold is chosen appropriately, then this approach causes the largest increase in NGE. However, the potential to increase the NGE was small for the breeds considered. Interestingly, for the current populations, optimum contributions for Approach A were slightly negatively correlated with the optimum contributions obtained for the other approaches, whereas the optimum contributions for the approaches B, C, and D were pairwise positively correlated (not shown).

Amador *et al.* [24] proposed two other approaches to reduce migrant contributions. Their first approach was to minimize migrant contributions in the offspring population. Their second approach was to minimize the probability that two alleles randomly chosen from the offspring population are IBD and descend from migrants. This objective function was computed from partial coancestry coefficients [25], but could also be computed by the methodology introduced in this paper. For both approaches, the maximum rate of inbreeding was restricted. However, provided that an acceptable rate of inbreeding can be achieved, it is not obvious why it is desirable that alleles originating from migrants should be not IBD in the offspring population. In contrast, all approaches proposed in this paper aim at increasing the probability that alleles originating from native founders are not IBD. Amador *et al.* [24] concluded that even with only a few generations without management, a small amount of introgression can spread into the population and it may be almost impossible to recover. This was not observed in our study. The reason is probably that the total population sizes of the cattle breeds were much larger than their effective sizes, which increased the probability to find individuals with small migrant contributions. Moreover, the cattle populations may deviate from random mating populations because some breeders avoid the use of bulls with high migrant contributions.

Another approach could be to minimize the effective number of non-founders N_{enf}, as defined by Caballero and Toro [4], in the offspring population. This approach would be equivalent to maximization of $\bar{f}_O - \frac{1}{2N_{ef}(O)}$ which would be achieved by increasing the average relationship $\bar{f}_O$ in the offspring population O and by increasing the effective number N_{ef} of founders in the offspring generation. Thus, the rate of inbreeding would have to be restricted by this alternative approach. This approach, however, would by definition not be optimal with respect to the objective functions introduced in this paper.

Our results show that migrant contributions can be substantially decreased for all three breeds, but the potential to increase the NGE is limited. The reduction of migrant contributions would be largely achieved in the first generation of management. In subsequent generations, some further improvement would be possible due to biological restrictions in previous generations. However, thereafter the management method becomes equivalent to an equalization of family sizes and no further reduction of migrant contributions could be achieved. Moreover, pedigree-based optimum contribution selection cannot remove genetic contributions of migrants that arose before recording of pedigrees started. However, removal of migrant contributions that arose earlier can be done subsequent to pedigree-based optimum contribution selection by identification of chromosome segments that are also present in the migrant breeds and by removal of those individuals with large migrant contributions from the breeding pool. Since migrants are usually males, haplotype variants of the Y-chromosome

can be used as markers for paternal lineage [26] to identify the migrant breeds. For individuals that are not removed from the breeding pool (i.e individuals with small migrant contributions), optimum contributions can be calculated based on genomic relationships. In order to avoid that this approach causes the frequencies of migrant alleles to increase, the set of breeding individuals could be enlarged with individuals of the migrant breeds. After the optimum contributions have been computed, the contributions of these additional migrant individuals are set to zero, and the optimum contributions for individuals of the breed of interest are rescaled, so that they add up to one. Thereafter, it would be beneficial to combine closely related breeds with low gene diversity in order to reduce extinction probabilities [27], and to split breeds with a high gene diversity into several subpopulations in order to reduce the decrease of overall gene diversity [28]. Breeds with highest value for conservation should be given priority [29]. These breeds are likely found near the domestication center (since genetic diversity declines with increasing distance from the domestication centre [30]), far from the native areas of economically superior breeds, or live in harsh environmental conditions. Candidates are also breeds that are used for uncommon purposes (e.g. fighting cattle, cattle breeds used for cow racing).

Conclusions

The usual recommendation to optimize contributions for breeding individuals by maximizing gene diversity in the offspring is not suitable for populations with historic migration because maximization of gene diversity would be achieved by maximization of migrant contributions. Thus, this approach, applied to populations with migration, would rapidly lead to their extinction. Two approaches can be recommended. The first is to maximize the probability that two alleles randomly chosen from the offspring population are not IBD and that at least one of them descended from a native founder (Approach C). The other approach is to constrain migrant contributions while maximizing the conditional probability that two alleles randomly chosen from the offspring population are not IBD, given that both descended from native founders (Approach D). Migrant contributions could be substantially decreased for the three breeds investigated here, but the potential to increase the NGE is limited.

Programs for pedigree-based optimum contribution selection and for the analyses presented in this paper are available in R package *PedAnalysis* from the first author. Since migrants are usually from genetically superior breeds, optimum contribution selection is likely to reduce breeding values if there is no constraint on the expected breeding value of the offspring. The program for optimum contribution selection allows adding the constraint that the expected mean breeding value of the offspring does not fall below a certain value. Moreover it is possible to put a constraint on the maximum number of offspring per male and female.

Additional file

Additional file 1: Proofs. The file provides the proofs of all numbered equations.

References
1. Brüggemann H: *Viehhaltung im Bauernbetrieb*. DLG-Verlag: Frankfurt am Main; 1954.
2. Oldenbroek K: **Introduction.** In *Utilisation and conservation of farm animal genetic resources*. Edited by Oldenbroek K. Wageningen: Wageningen Academic Publishers; 2007:13–27.
3. Meuwissen THE: **Maximising the response of selection with a pre-defined rate of inbreeding.** *J Anim Sci* 1997, **75:**934–940.
4. Caballero A, Toro MA: **Interrelations between effective population size and other pedigree tools for the management of conserved populations.** *Genet Res* 2000, **75:**331–343.
5. Lacy RC: **Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents.** *Zoo Biol* 1989, **8:**111–123.
6. Wellmann R, Pfeiffer I: **Pedigree analysis for conservation of genetic diversity and purging.** *Genet Res* 2009, **91:**209–219.
7. Meuwissen THE, Goddard ME: **Multipoint Identity-by-Descent prediction using dense markers to map QTL and estimate effective population size.** *Genetics* 2007, **176:**2551–2560.
8. Wellmann R, Bennewitz J: **Identification and characterization of hierarchical structures in dog breeding schemes, a novel method applied to the Norfolk Terrier.** *J Anim Sci* 2011, **89:**3846–3858.
9. Meuwissen THE: **Operation of conservation schemes.** In *Utilisation and conservation of farm animal genetic resources*. Edited by Oldenbroek K. Wageningen: Wageningen Academic Publishers; 2007:167–193.
10. Vanderbei RJ: **LOQO: An interior point code for quadratic programming.** *Optim Method Softw* 1999, **11:**451–484.
11. Goldfarb D, Idnani A: **Dual and primal-dual methods for solving strictly convex quadratic programs.** In *Numerical Analysis*. Edited by Hennart JP. Berlin: Springer-Verlag; 1982:226–239.
12. Goldfarb D, Idnani A: **A numerically stable dual method for solving strictly convex quadratic programs.** *Math Program* 1983, **27:**1–33.
13. Martin W, Zeeb R: *Martin-Zeeb Handbuch der Landwirtschaft*. Stuttgart: Verlag Eugen Ulmer; 1922.
14. Schlipf JA: *Schlipf's Handbuch der Landwirtschaft*. Berlin: Verlag Paul Parey; 1914.
15. Hansen J: *Lehrbuch der Rinderzucht*. Berlin: Verlag Paul Parey; 1921.
16. Sambraus HH: *Gefährdete Nutztierrassen*. Stuttgart: Ulmer Eugen Verlag; 2010.
17. Mangold P, Reicherter E: *Neues illustriertes Haustierarzneibuch*. Reutlingen: Verlag Enklin und Laiblins; 1925.

18. Bennewitz J, Meuwissen THE: **Estimation of extinction probabilities of five German cattle breeds by population viability analysis.** *J Dairy Sci* 2005, **88:**2949–2961.
19. Wilckens M: *Die Rinderrassen Mittel-Europas*. Berlin: Verlag von Paul Parey; 1885.
20. Villa-Angulo R, Matukumalli LK, Gill CA, Choi J, Van Tassell CP, Grefenstette JJ: **High-resolution haplotype block structure in the cattle genome.** *BMC Genet* 2009, **10:**19.
21. Meuwissen THE: **Genetic management of small populations: A review.** *Acta Agr Scand A* 2009, **59:**71–79.
22. Toro M, Mäki-Tanila A: **Genomics reveals domestication history and facilitates breed development.** In *Utilisation and conservation of farm animal genetic resources*. Edited by Oldenbroek K. Wageningen: Wageningen Academic Publishers; 2007:13–27.
23. vit LfL, ADR: *Erfassungsprojekt Erhebung von Populationsdaten tiergenetischer Ressourcen in Deutschland: Tierart Rind*. Verden/Grub/Bonn: Bundesministerium für Landwirtschaft und Ernährung; 2010.
24. Amador C, Toro MA, Fernández J: **Removing exogeneous information using pedigree data.** *Conserv Genet* 2011, **12:**1565–1573.
25. García-Cortés LA, Martínez-Ávila JC, Toro MA: **Fine decomposition of the inbreeding and the coancestry coefficients by using the tabular method.** *Conserv Genet* 2010, **11:**1945–1952.
26. Svensson E, Götherström A: **Temporal fluctuations of Y-chromosomal variation in Bos taurus.** *Biol Lett* 2008, **4:**752–754.
27. Bennewitz J, Simianer H, Meuwissen THE: **Investigations on merging breeds in genetic conservation schemes.** *J Dairy Sci* 2008, **91:**2512–2519.
28. Kimura M, Crow JF: **On the maximum avoidance of inbreeding.** *Genet Res* 1963, **4:**399–415.
29. Bennewitz J, Kantanen J, Tapio I, Li MH, Kalm E, Vilkki J, Ammosov I, Ivanova Z, Kiselyova T, Popov R, Meuwissen TH: **Estimation of breed contributions to present and future genetic diversity of 44 North Eurasian cattle breeds using core set diversity measures.** *Genet Sel Evol* 2006, **38:**201–220.
30. Groeneveld LF, Lenstra JA, Eding H, Toro MA, Scherf B, Pilling D, Negrini R, Finlay EK, Jianlin H, Groeneveld E, Weigend S: **GLOBALDIV Consortium: Genetic diversity in farm animals - a review.** *Anim Genet* 2010, **41:**6–31.