Consequences of applying Optimum Contribution Selection on conventional and genomic based breeding schemes

Dissertation

zur Erlangung des Doktorgrades
der Agrar- und Ernährungswissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

M.Sc. Jan Körte

aus Greifswald

Dekanin: Prof. Dr. K. Schwarz
Erster Berichterstatter: Prof. Dr. G. Thaller
Zweiter Berichterstatter: Prof. Dr. H. Semianer
Tag der mündlichen Prüfung: 10.05.2011

# Table of Contents

General Introduction ................................................................................................... 1

Chapter 1.................................................................................................................... 5

Analysis of converting a breeding program under BLUP truncation selection to
optimum contribution selection

Chapter 2.................................................................................................................. 21

Impact of applying genomic information on optimum contribution selection

Chapter 3.................................................................................................................. 41

Genomic optimum contribution selection to reduce the effect of hitchhiking

General Discussion .................................................................................................. 59

General Summary .................................................................................................... 73

Zusammenfassung ................................................................................................... 75
General Introduction

The maintenance of the additive genetic variance within a population is one of the most essential aspects in animal breeding although consistently underrated. The genetic variance is the substrate of any genetic progress and allows for reacting on changes in consumers’ habits and markets that directly influence the breeding objectives in livestock. The decrease of additive genetic variance is induced by inbreeding and genetic drift, which are two well-known forces in the genetics of populations. Inbreeding is also known to be connected with aspects of fitness and vitality of a breed.

In the last decade, there have been several efforts to include these considerations in the management of breeding populations where the increase of the genetic progress is of major importance. The approach most commonly applied is the optimum contribution selection first introduced by Meuwissen (1997). This method maximizes the genetic gain under the constraint of a predefined increase of inbreeding. Several studies describe the implementation of optimum contribution selection in practical breeding programs and conclude that optimized selection strategies should be used for the management of inbreeding in practical breeding programs (Avendano et al., 2003; Koenig and Simianer, 2006).

The aim of chapter 1 was to illustrate the consequences of changing a conventional breeding program with truncation selection on best linear unbiased prediction (BLUP) estimated breeding values to optimal genetic contribution selection and two different situations were analyzed. Firstly, how much additional genetic gain could be achieved by changing to optimum contribution selection while keeping the rate of inbreeding constant for the following generations. Secondly, how much the rate of inbreeding could be decreased, if a specific genetic gain is achieved, when changing to optimum contribution selection.

In recent years, advanced techniques in molecular genetics have provided dense marker maps. These can be used to predict genome wide breeding values (Meuwissen et al., 2001). Daetwyler et al. (2007) showed that selection for genome wide breeding values decreased inbreeding compared to selection for BLUP estimated breeding values. Genome wide breeding values had higher prediction accuracies of the Mendelian sampling term and therefore a better differentiation of family members. However, the authors pointed out that despite applying genome
wide breeding values in selection there is a need for managing inbreeding. Furthermore, the marker information allows to predict more accurate relationships and coefficients of inbreeding.

In chapter 2 the optimum contribution selection was extended by replacing the numerator relationship matrix with the genomic relationship matrix and this genomic optimum contribution selection was compared with the original pedigree optimum contribution selection. Thereby, the impact of these breeding schemes on pedigree and genomic rates of inbreeding, responses to selection, changes in QTL gene frequencies, loss of favorable QTL alleles and number of selected animals were evaluated for a low heritable trait by using stochastic computer simulation. Selection methods based on genomic information increase the accuracy of estimated breeding values and, therefore, increase the genetic progress. On the other hand, the selection focuses more directly specific genome regions and the theory of selective neutral loci has to be verified. Roughsedge et al. (2008) showed that marker assisted selection reduces pedigree estimated inbreeding but not true inbreeding at the QTL and the region surrounded. Therefore, the aim of chapter 3 was to investigate to what extend true inbreeding of a QTL and the surrounding region is affected by the optimum contribution selection using a numerator relationship matrix based on pedigree or genomic information.

REFERENCES


Chapter 1

Analysis of converting a breeding program under BLUP truncation selection to optimum contribution selection

Jan Körte
Dirk Hinrichs
Georg Thaller

Institute of Animal Breeding and Husbandry, Christian-Albrechts-University Kiel, D-24098 Kiel, Germany
ABSTRACT

The objective of this study was to analyze the change of an existing breeding program of truncation selection using BLUP estimated breeding values to an optimum contribution selection scheme. It focuses on two major aspects. At first, it was analyzed how much additional genetic gain could be achieved by changing to optimum contribution selection while keeping the rate of inbreeding constant for the following generations. Thereafter, it was analyzed how much the rate of inbreeding could be decreased, if a specific genetic gain was achieved, when changing to optimum contribution selection. To evaluate the benefit of these two scenarios of changed selection, i.e. after 10 generations the selection changed from truncation selection to optimum contribution selection, a comparable truncation selection scheme over 20 generations was simulated. For all scenarios, a nucleus breeding scheme with 20 discrete generation were modeled by stochastic simulation.

For situations where the selection changed to optimum contribution selection the rate of inbreeding was restricted to 0.005, 0.01 or 0.025, respectively. In case of a constant rate of inbreeding after changing the selection the genetic gain of optimum contribution selection was 41% (∆F = 0.01) and 10% (∆F = 0.025) higher compared to 20 generations of truncation selection. For the situations with constant genetic gain inbreeding was decreased up to 31% (∆F = 0.005), when selection changed from truncation selection to optimum contribution selection after 10 generations, compared to 20 generations of truncation selection.

Keywords: Optimum contribution selection, BLUP truncation selection, converting breeding scheme
INTRODUCTION

Today modern breeding programs are characterized by a very accurate estimation of breeding values (EBV) and by an increased use of reproduction techniques. Advances in reproductive technology, such as embryo transfer and in-vitro fertilization, enable an increased selection intensity because less parents are needed to produce the next generation of breeding animals. Improvements of statistical methods, such as using best linear unbiased predictions (BLUP) animal model, have led to higher accuracy of EBV and a greater probability of co-selection of related animals (Quinton et al., 1992).

The optimal balance between genetic gain and the rate of inbreeding (ΔF) is one of the most difficult issues in breeding programs. An aggressive selection strategy based on a few highly ranked individuals will reduce genetic variation and may cause high levels of inbreeding depression, which in turn might decrease the long-term response to selection. On the other hand, a restrictive strategy with larger numbers of individuals selected results in lower genetic gain, especially during the initial generations of selection. Although inbreeding is unavoidable in closed selection programs, increases in inbreeding need to be restricted to alleviate long-term negative effects (Burrow, 1993; Lamberson and Thomas, 1984). Therefore, dynamic selection methods, called optimum contribution (OC) selection, have been proposed (Grundy et al., 1998; Meuwissen, 1997; Meuwissen and Sonesson, 1998). These selection methods maximize genetic gain while constraining the rate of inbreeding in the population by optimizing the contributions of the parents to the next generation. These authors showed via simulation that this method could achieve genetic gains of 21 up to 60 % greater than selection on BLUP EBV after responses were standardized to a constant level of inbreeding. The method has been applied in practical breeding programs, for instance in the UK and the German Holstein dairy population (Kearney et al., 2004; Koenig and Simianer, 2006), in two British livestock populations of sheep and beef cattle (Avendano et al., 2003) and selection in large salomon fish breeding stock (Hinrichs et al., 2006). In these studies the potential of OC selection to increase genetic gain under the same constrained for the increase of average inbreeding was demonstrated.

The aim of this paper was to illustrate the consequences when changing an existing breeding program with a truncation selection for BLUP EBV to the OC selection
method by Meuwissen (1997). Thereby, two major aspects were of interest. Firstly, how much additional genetic gain could be achieved by changing to OC selection while keeping ∆F constant for the following generations. Secondly, how much could ∆F be decreased, if a specific genetic gain is achieved, when changing to OC selection. An empirical approach was applied by using stochastic simulations and comparisons with traditional truncation selection.

MATERIAL AND METHODS

Simulation
Selection over multiple generations was modeled using stochastic computer simulations. The parameters of the simulated breeding schemes are given in Table 1. The general structure of the population is a closed nucleus scheme with discrete generations. The number of selection candidates per generation is 100, 50 male, and 50 female, respectively. An additive infinitesimal model (Bulmer, 1971) was considered for the trait under selection. The true breeding values for animals in the base population were obtained from a normal distribution with mean zero and variance $\sigma_a^2$ equal to the heritability ($h^2$) of the trait. Thus, the phenotypic variance $\sigma_p^2$ was assumed to be equal to one. Later generations are obtained by simulating true breeding values of the offspring drawn from a normal distribution with mean

$$g_i = 1/2g_s + 1/2g_d + m_i,$$

where $s$ is the sire and $d$ the dam of offspring $i$, and $m_i$ is the Mendelian sampling component which is sampled from a normal distribution with mean zero and variance $\sigma_m^2 = 0.5\left[1 - 0.5(S(F_s + F_d))h^2\right]$, where $F_s$ and $F_d$ are the inbreeding coefficients of the sire and dam, respectively. The phenotypic value was obtained by adding an environmental component sampled from a normal distribution with mean zero and variance $1 - h^2$ to the true breeding value. Three traits were assumed with a heritability of 0.1, 0.25 and 0.5, respectively.
Table 1: Parameter of the closed nucleus scheme

<table>
<thead>
<tr>
<th>Size of selection scheme</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of selection candidates</td>
<td>100</td>
</tr>
<tr>
<td>Number male selection candidates</td>
<td>50</td>
</tr>
<tr>
<td>Number of female selection candidates</td>
<td>50</td>
</tr>
<tr>
<td>Number of Generations simulated</td>
<td>20</td>
</tr>
<tr>
<td>Number of generations under BLUP truncation selection</td>
<td>10</td>
</tr>
<tr>
<td>Number of generations under OCS</td>
<td>10</td>
</tr>
<tr>
<td>Number of replicated simulations</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter of traits</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotypic variation</td>
<td>1</td>
</tr>
<tr>
<td>Heritability, h²</td>
<td>0.10, 0.25 or 0.50</td>
</tr>
<tr>
<td>Inbreeding constraint, ∆F</td>
<td>0.005, 0.01, 0.025</td>
</tr>
<tr>
<td>No. of selected sires and dams for equal genetic level</td>
<td>32, 17, 5</td>
</tr>
<tr>
<td>No of selected sires and dams for equal rate of Inbreeding</td>
<td>50, 36, 18</td>
</tr>
</tbody>
</table>

**Breeding schemes**

Populations were evaluated over 20 generations of selection, where the breeding scheme for the first ten generations was that of a standard truncation selection (TS) with a selection of a fixed number of sires and dams with the highest BLUP EBV. Thereafter, the selection scheme was changed to an OC selection and evaluated for another 10 generation. Thereby, two main scenarios were considered. At first, the observed ∆F of the first ten generations of selection should be kept constant in the following OC selection scheme. And secondly, the yearly genetic gain, achieved during the first 10 generations should be maintained when applying the OC selection scheme.

The first scenario can be thought of a breeding program where the actual ∆F is accepted. It addresses the question, whether and to what extent the breeding program would benefit with a change of the selection scheme. The second scenario represents a situation where a breeding program resulting in desired genetic gain but there is an interest to reduce the inbreeding.
To evaluate the benefit of these two scenarios of changing selection (CS), comparable TS over all 20 generations of selection was also simulated.

For the first scenario, three breeding programs with different intensities of selection and thereby different increases in inbreeding were evaluated. The number of selected sires and dams for the TS were set to 50, 36, and 18 to reach a ∆F of 0.005, 0.01 and 0.025, respectively. These numbers are set approximately, following Wright’s (1931) inbreeding formula for random mating: ∆F = 1/8n_s + 1/8n_d, where n_s is the number of sires and n_d is the number of dams selected. In the subsequent OC selection the constraint on ∆F was set equal to the previous observed ∆F.

For the second scenario, more stringent breeding programs were considered. To reach a higher intensity of selection and with it higher rates of inbreeding, the number of sires and dams selected for the TS were set to 32, 17 and 5.

**Selection methods**

Two selection methods were used to select parents for the next generation. The first was a TS, where an equal number of sires and dams with the highest BLUP EBV were selected. The second was the OC selection as described by Meuwissen (1997). This method maximizes the genetic level of the next generation of animals, 

\[ G_{t+1} = c'_t \text{EBV}_t, \]

where \(c_t\) is a vector of genetic contributions of the selection candidates to generation \(t+1\) and \(\text{EBV}_t\) is a vector of BLUP EBV of the candidates for selection in generation \(t\). Rates of inbreeding are controlled by constraining the average coancestry to 

\[ C_{t+1} = c'_t A_t c_t / 2, \]

where \(A_t\) is the numerator relationship matrix of the selection candidates in generation \(t\).

The constraint was set as 

\[ C_{t+1} = C_t + \Delta F (2 - C_t), \]

where \(C_t\) is the average relationship of the present candidates and ∆F is the targeted rate of inbreeding (Meuwissen, 2002). A second condition constrains the sum of the contributions of male and female to \(\frac{1}{2}\), respectively, with \(Q'c_t = 1/2\). \(Q\) is an \(n \times 2\) incidence matrix of the sex of the selection candidates, where the first column contains ones for males and zeros for females, and the second column ones for females and zeros for males, \(1/2\) is a vector with the value 0.5 of order 2. So the function to maximize is:

\[ H_t = c'_t \text{EBV}_t - \lambda_0 (c'_t A_t c_t - 2C_{t+1}) - \left(c'_t Q - 1/2^t\right)\lambda, \]

where \(\lambda_0\) and \(\lambda\) are LaGrangian multipliers (\(\lambda = \) a vector of two LaGrangian multipliers). A detailed description of the optimization procedure is given by
Meuwissen (1997). The output from the selection method is the vector $c_t$ with genetic contribution of each selection candidate to the next generation.

The conventional BLUP animal model was used to evaluate EBV (Henderson, 1984):

$$ y = 1\mu + Za + e. $$

where $y$ was a vector of the observations for the trait, $\mu$ was the overall mean, $a$ was a vector of random additive genetic effects, and $Z$ was a incidence matrix relating records in $y$ to the random effects. The vector $e$ contained random residuals specific to each individual.

**RESULTS**

**Fixed rate of inbreeding**

Fig. 1 shows the genetic gain over 20 generations of selection for the CS and the comparable TS for an expected $\Delta F = 0.005$, 0.01 and 0.025, respectively. The choice of the number of selected males and females in TS was successful to reach the expected $\Delta F$ in each generation. Some deviations from the expected $\Delta F$ occurred, because the male and female candidates were chosen at random to be parent for the next generation, with probabilities that correspond to the genetic contributions obtained from the OC algorithm.

After changing the breeding scheme to an OC selection, the genetic gain increased for all expected $\Delta F$ and all levels of heritability. As expected, the highest genetic gain could be achieved with the most relaxed $\Delta F$ ($\Delta F = 0.025$) followed by $\Delta F = 0.01$ and 0.005, respectively. However, the difference of relative and absolute genetic gain between CS and TS after 20 generations of selection was conversely (Tab.2). Most advantage of CS was achieved with the lowest expected $\Delta F = 0.005$. This was the most stringent constraint which was reached in TS with the selection of all animals which was de facto a random mating and, as expected, no genetic gain was obtained.
Figure1: Average coefficient of inbreeding and average genetic gain for a converted selection (CS) scheme, with a change from a truncation selection to an optimum contribution (OC) selection in generation 10, and a continued truncation selection (TS) over all 20 generations. The constraint of the ΔF of the OC selection was set to the observed ΔF in generation 1 – 10. [(a) ΔF = 0.025, (b) ΔF = 0.01, (c) ΔF = 0.005]

Nevertheless with the change to an OC selection, the CS scheme was able to generate additional genetic gain from generation 10.
Table 2: Fixed rate of inbreeding. Coefficient of inbreeding (F), genetic gain (G) and number of selected males and females at generation 20 for converting selection (CS) and truncation selection (TS) scheme.

<table>
<thead>
<tr>
<th>$h^2$</th>
<th>CS</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.406</td>
<td>2.057</td>
</tr>
<tr>
<td>0.25</td>
<td>0.393</td>
<td>4.540</td>
</tr>
<tr>
<td>0.50</td>
<td>0.361</td>
<td>8.033</td>
</tr>
</tbody>
</table>

Further relaxation of $\Delta F$ decreased the benefit of OC selection. For an expected $\Delta F = 0.01$ in all generations, the genetic level of CS was 41% higher for all levels of heritability after 20 generations of selection. With an expected $\Delta F = 0.025$, these advantages were around 10%. For the most severe $\Delta F$, the number of selection candidates was 100 for the ten generation of TS (50 males and 50 females) which is similar to a random mating. After converting the breeding program, the number of selection candidates dropped down to in average 84 (42 males and 42 females) for OC selection. The relaxation of $\Delta F$ led to a decreased number of selection candidates. For an expected $\Delta F$ of 0.01 the number of selection candidates dropped down to 72 for TS and average 56 (28 males and 28 females) in OC selection and for $\Delta F = 0.025$ the number of selection candidates was 36 for TS and average 30 (15 males and 15 females) for OC selection.
Fixed genetic gain

Figure 2: Average coefficient of inbreeding and average genetic gain for a converted selection (CS) scheme, with a change from a truncation selection to an optimum contribution (OC) selection in generation 10, and a continued truncation selection (TS) over all 20 generations. The constraint of the $\Delta F$ was chosen such that the same genetic gain as in the TS was realized. [(a) $\Delta F = 0.025$, (b) $\Delta F = 0.01$, (c) $\Delta F = 0.005$]

Figure 2 shows the inbreeding coefficient for the BLUP TS and CS scheme when both achieved equal genetic gains in each generation of selection. Under this
condition the selection of 10 animals (5 males and 5 females) resulted in the highest inbreeding for the TS scheme. Therefore the average $\Delta F$ for the first ten generation of TS was 0.109 and the inbreeding coefficient raised up to 0.603. After changing to the OC selection, the same genetic level as in the comparable BLUP TS could be achieved by restricting $\Delta F$ to 0.025, so that the inbreeding coefficient after 20 generations was 0.703, 26% lower than in the comparable BLUP TS (Tab. 3).

An increase of selection candidates for BLUP TS decreased the inbreeding coefficients but also the genetic gain.

A higher difference with 31 in the inbreeding coefficient between CS and TS after 20 generation was observed when selecting 34 (17 males and 17 females) in the BLUP TS and constrained $\Delta F$ to 0.01 in the following OC selection. Nearly the same difference of the inbreeding coefficient in generation 20 was observed with a constrained $\Delta F$ of 0.005 for OC selection compared to the BLUP TS selecting 64 (32 males and 32 females) animals.

Table 3. Coefficient of inbreeding ($F$), genetic gain ($G$) and number of selected males and females at generation 20 for a converted selection (CS) scheme, with a change from a truncation selection to an optimum contribution (OC) selection in generation 10, and a continued truncation selection (TS) over all 20 generations. The constraint of the $\Delta F$ was chosen such that the same genetic gain as in the TS was realized.

<table>
<thead>
<tr>
<th>$h^2$</th>
<th>CS</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta F_d = 0.025$</td>
<td>$\Delta F_d = 0.01$</td>
</tr>
<tr>
<td></td>
<td>$N_{sid} = 5$</td>
<td>$N_{sid} = 17$</td>
</tr>
<tr>
<td>0.10</td>
<td>0.734 1.826 18 18</td>
<td>0.911 1.811 5 5</td>
</tr>
<tr>
<td>0.25</td>
<td>0.703 4.230 16 16</td>
<td>0.888 4.262 5 5</td>
</tr>
<tr>
<td>0.50</td>
<td>0.653 8.141 13 14</td>
<td>0.841 8.460 5 5</td>
</tr>
</tbody>
</table>
DISCUSSION

In this study, we evaluated the converting of an existing breeding program under BLUP TS to OC selection by stochastic simulation. The change of the selection method had two objectives. Firstly the inbreeding and secondly the genetic gain should increase equally, before and after the changeover. The results demonstrated that the selection of parameters was successful to reach these goals. This study showed that the change to an OC selection method yielded in higher genetic gains or lower rates of inbreeding like a comparable TS schemes.

If the inbreeding increased in all 20 generation of selection equally, the genetic gain increased more after the conversion of the selection method in generation 10 (Fig. 1). The highest increases in genetic gain were obtained by converting TS with selecting 50 males and 50 females to an OC selection with a constraint ∆F of 0.005 for a fixed increase of inbreeding. It should be noted that in the TS scheme all 50 male and 50 female selection candidates were selected, so that this scheme equaled a conservation scheme with no increase in genetic gain. In the CS only the first ten generations of TS had no genetic gain but after changing to OC selection the number of selected parents decreased so that the genetic gain increased. Therefore, an increase in genetic gain could be possible in a conservation breeding scheme. More valuable results were obtained when selecting fewer animals for TS and relaxed constrained in OC selection. Therefore, the advantage of OC selection in generation 20 raised up to 41% higher genetic gains. These results are in agreement with Meuwissen (1997), who described that the advantage of the OC selection method increased when the restriction on ∆F became more stringent. The OC selection method selected fewer animals than TS, at the same levels of inbreeding. This indicated that OC selection achieved its higher genetic gain by realizing a higher selection differential at the same rate of inbreeding.

There was a small trend of selecting more animals with decreasing heritability, because the lower heritability resulted in a higher correlation between EBV of relatives. Hence, the selected animals would be more related, which was compensated in this study by the selection of more animals (Tab 2).

Due to the fact that a relaxation in constraint for relationships resulted in a lower number of selected animals and with this in a higher intensity of selection, the increase in genetic gain was generally associated with an allowed increase of
inbreeding. Hence, higher genetic contributions of genetically favorable animals were possible. The OC selection method would increase the genetic response by combined contributions of unrelated families, by avoiding extreme relationships of offspring and provide reduced inbreeding levels of the parents in the next generation (Sonesson et al., 2000; Sonesson and Meuwissen, 2000). Kearney et al. (2004) applied the OC algorithm of Meuwissen (1997) to the U.S. Holstein population and connected OC and index scores for different levels of constrained inbreeding for selected males. They also found more selected males with a more stringent constrained on rate of inbreeding. Avendano et al. (2003) implemented the OC selection method in beef cattle and sheep livestock populations and found also that OC selection led to higher genetic gain compared with the same level of inbreeding. The highest increase in genetic gain could be obtained when selection was allowed in both males and females. This is in accordance with our results. Similar results were confirmed by Koenig and Simianer (2006) and Weigel and Lin (2002). In these studies, the population size was high and the average relationship of the selection candidates was low. In our study, the population size of 100 animals per generation was small and the average inbreeding up to 0.6 in the generation before OC selection was high. Even in this case, the OC algorithm yielded to an optimal solution and could achieve the constrained ∆F and a higher genetic gain than the comparable TS.

If the genetic gain increased in all 20 generation of selection equally, the inbreeding increased less after the conversion of the selection method in generation 10 (Fig. 2). The lower inbreeding was obtained by selecting fewer selection candidates as parents for the next generation (Tab.3). The reduced inbreeding levels would led to larger Mendelian sampling variance, i.e. the term $\sigma_a^2 = 0.5[1 - 0.5(F_s + F_d)]h^2$ was larger, resulting in more genetic variance, and with it also in more genetic gain. For high heritability, there was a small trend to increase the genetic variance in the generation 11 to 14 for CS. For lower heritability, the genetic variance was proceeding on a higher level than in the comparable TS after converting the breeding program (results not shown). Thus, one of the most discussed risks of selection was minimized, especially for small populations. As described above there was a small trend of selecting more animals with decreasing heritability.
In this work, the possibility chance of a breeding program with discrete generations was shown, but the OC algorithm could be extended to overlapping generations (Grundy et al., 2000; Meuwissen and Sonesson, 1998).

As concluded by Avendano et al. (2003) and Kearney et al. (2004) the practical realization of OC selection requires a coordinated policy of the use of selected candidates and also for breeding objectives. A cooperation of pedigree breeders and artificial insamation organizations could make sure that the correct animals and their contributions are identified. For small-scale schemes a reasonable target could be a coordinated use of selection candidates among breeding herds. In large schemes with different breeding objectives this might be difficult. Therefore, a more reasonable approach could be the application of OC selection on individual groups of herds like Koenig and Simianer (2006) did with applying OC selection on the elite mating selection of bull sires and bull dams.

**CONCLUSION**

This work showed the consequences of the change of an existing breeding scheme with TS for BLUP EBV to an OC selection scheme. Implementing OC selection could increase genetic gain for a given rate of inbreeding or attain similar genetic gains at much lower rates of inbreeding compared to previous TS. Compared with TS over the same selection horizon the change of a breeding program to an OC selection could lead to a decrease of rates of inbreeding or increase of genetic gain for a long term. Also in a conservation breeding scheme the change to OC selection could be necessary to increase the genetic gain.

**ACKNOWLEDGEMENTS**

This study is part of the project FUGATO-plus Brain and was financially supported by the German Ministry of Education and Research, BMBF, the Förderverein Biotechnologieforschung e.V. (FBF), Bonn, Lohmann Tierzucht GmbH, Cuxhaven, Verband hannoverscher Warmblützüchter e.V, Verden, Verband der Züchter des Oldenburger Pferdes e.V., Vechta, Verband der Züchter des Holsteiner Pferdes e.V., Elmshorn, Westfälisches Pferdestammbuch e.V., Münster, and Trakehner Verband, Neumünster.
REFERENCES


Chapter 2

Impact of applying genomic information on optimum contribution selection

Jan Körte
Dirk Hinrichs
Georg Thaller

Institute of Animal Breeding and Husbandry, Christian-Albrechts-University Kiel, D-24098 Kiel, Germany
ABSTRACT

The aim of this study was to extend the optimum contribution selection to the genomic relationship matrix and compare this genomic optimum contribution selection with the pedigree optimum contribution selection. The impact of this breeding schemes on pedigree and genomic rates of inbreeding, responses to selection, changes in QTL gene frequencies, loss of favorable QTL alleles and number of selected animals was evaluated for a trait ($h^2 = 0.1$) by using stochastic computer simulation.

Breeding schemes were closed nuclei with 1000 animals from 100 full sib families. True breeding values for a trait were obtained by summing the true effect of the 500 simulated QTL, allocated of a genome consists of 10 chromosomes with a total length of 1000 cM. BLUP breeding values were predicted from phenotypic and pedigree information. Genomic estimated breeding values were maintained from effects of 2000 markers whose effects were estimated using a BayesB model and were updated every generation. Inbreeding followed the expectations on the scale it was constrained. In case of constraining pedigree inbreeding and selecting for genomic estimated breeding values, the genomic inbreeding was around 5 times higher in generation 10 than the constraint. This indicated that the true inbreeding was underestimated when using additive genetic relationship matrix. After 10 generations of selection the cumulative response of the genomic estimated breeding scheme was 44% higher than in the BLUP breeding scheme, when constraining genomic inbreeding and 33% higher when constraining pedigree inbreeding. For all selection strategies, accuracy of the estimated breeding values decreased, whereas the decrease in pedigree optimum contribution strategies was greater. Trends in mean frequencies of favorable QTL alleles followed the response to selection. The pedigree optimum contribution resulted in higher mean frequencies than genomic optimum contribution.

These results indicate that it is useful to restrict genomic inbreeding in genomic improvement programs. Genomic optimum contribution can maintain higher genomic variability in terms of lower frequencies of favorable QTL and lower proportion of QTL that were lost. Furthermore, a smaller decrease of genetic variance led to a higher response of selection on the long-term.

Keywords: Genomic optimum contribution, genomic estimated breeding values, genetic improvement program
INTRODUCTION

Recent advances in the field of molecular genetics provide dense marker data across the whole genome for livestock species. A major application of this information is the so-called genomic selection (GS), which uses genome-wide SNP-marker information to estimate genomic breeding values (GEBV) (Meuwissen et al., 2001). Genomic selection is of increasing importance and has been implemented in dairy cattle breeding schemes in Germany (Habier et al., 2009; Seefried et al., 2009) and North America (VanRaden et al., 2009). Similar work is underway in further countries and the approach is currently evaluated in other species as well (Goddard and Hayes, 2007; Heffner et al., 2009; Nielsen et al., 2009). This new methodology of estimating breeding values promises several benefits. Accuracies of genomic breeding values and therefore response to selection are higher compared to those estimated by using traditional BLUP methods (Muir, 2007). The GS method especially increases the accuracy of the Mendelian sampling component, which allows a better differentiation of family members and thus may reduce inbreeding (Daetwyler et al., 2007; Dekkers, 2007). On the other hand, GS enables the testing of many candidates and the selection of comparatively young animals based on breeding values with a sufficiently high accuracy. Thereby the selection intensity gets higher and the generation interval is reduced especially in cattle breeding schemes (Schaeffer, 2006), resulting in an increase in annually inbreeding. In view of those developments it can be argued that the need to manage inbreeding properly is more pronounced when applying GS in breeding programs.

A well-adopted method to optimize the genetic gain while restricting the increase of the average relationship of selection candidates, and therefore the inbreeding of the next generation, is the optimum contribution (OC) selection introduced by Meuwissen (1997) and Grundy et al. (1998). So far the method uses the additive genetic relationship matrix based on pedigree data. The presence of genomic data now also allows to account for the genomic relationships among selection candidates. It is expected that the genomic relationship matrix represents a more accurate measurement of realized relationships between animals.

The aim of this study was to compare pedigree and genomic OC selection schemes, by evaluating their impact on pedigree and genomic based rates of inbreeding, responses to selection, changes in QTL gene frequencies, loss of favorable QTL
alleles and number of selected animals. Properties of various breeding schemes, either using GEBV or BLUP EBV for selection and using either the pedigree or the genomic relationship matrix were evaluated by computer simulations.

MATERIALS AND METHODS

Genomic relationship matrix

The calculation of the optimal genetic contribution of an animal to future generations requires the knowledge of relationships between selection candidates. Meuwissen (1997) used the additive genetic relationship matrix ($A$) and its inverse. The coefficients of inbreeding in $A$ are expected values derived by probabilistic approaches. The use of genetic markers provides more accurate relationship coefficients compared to $A$. Therefore, the genomic relationship matrix ($G$) as described by VanRaden (2008a) was considered as an alternative and can be obtained as follows:

$$G = \frac{ZZ'}{2\sum p_i(1-p_i)},$$

where $Z$ is a matrix of marker genotypes and $p_i$ is the frequency of the second allele at locus $i$. $Z$ can be calculated as $M \cdot P$, where $M_{n \times m}$ is an incidence matrix with the dimension of the number of individuals ($n$) by the number loci ($m$). Elements of $M$ are set to -1, 0 or 1 for the homozygous, heterozygous and alternative homozygous genotypes, respectively. $P$ is a matrix that contains allele frequencies expressed as a difference from 0.5 and multiplied by 2, such that column $i$ of $P$ is $2(p_i - 0.5)$, which sets mean values of the allele effects to zero. Allele frequencies in $P$ should be taken from the unselected base population rather than those observed after selection or inbreeding, because an earlier or later base population can lead to greater or lower relationships and to more or less inbreeding. When calculating genomic relationships, the subtraction of $P$ from $M$ gives more credit to rare alleles than to common alleles. Likewise, if the individual is homozygous for rare alleles compared to common alleles, the genomic inbreeding coefficient will be higher.

The denominator $2\sum p_i(1-p_i)$ ensures that the scale of $G$ is analogous to $A$. The matrix $G$ is generally positive semi-definite but can also be singular, if the total number of alleles is less than the number of individuals genotyped. Hence, the
number of alleles has to be greater than the number of selection candidates if \( G \) is implemented into the OC algorithm.

**Optimal Genetic Contribution selection**

The optimization algorithm as described by Meuwissen (1997) was used. This algorithm ensures maximum genetic gain whilst constraining \( \Delta F \) to a pre-defined level. The genetic level of the next generation can be written as:

\[
g_{t+1} = c_t'EBV_t,
\]

where \( c_t \) is the vector of genetic contributions of selection candidates to generation \( t+1 \) and \( EBV_t \) is the vector of BLUP EBV or GEBV, respectively. To find the optimal contribution vector \( c_t \), two constraints have to be considered. Firstly, the genetic contributions of all males and females have to sum up to 0.5, respectively:

\[
Q'c_t = [1/2,1/2],
\]

where \( Q \) is a known incidence matrix for the sex. The second constraint controls the future inbreeding, by restricting the average coancestry between selection candidates:

\[
\overline{C}_{t+1} = c_t'A_t c_t / 2,
\]

where \( \overline{C}_{t+1} \) is set to \([1-(1-\Delta F)^{t+1}]\), with \( \Delta F \) being the desired rate of inbreeding and the \( A \) matrix is either the numerator relationship matrix based on pedigree or the genomic relationship matrix based on marker information. These two constraints can be implemented into the initial function by introducing two Lagrangian multipliers and solved to obtain the contribution vectors that satisfy the requirements.

A more detailed description of optimum genetic contribution selection and the derivation of equations for the two Lagrangian multipliers is given by Meuwissen (1997).

**Simulation of the genome and the population**

A closed nucleus program was considered. Simulation started with a base population 1010 generations back from now (Gen1010) consisting of 500 individuals that were randomly mated for 800 discrete generations and then reduced to an effective population size of 100 individuals and randomly mated for another 200 generations. The initial 1000 generations were simulated assuming a mutation rate of \( 2.5*10^{-5} \) and recombination events following the Haldane function to generate mutation-drift equilibrium. Next, the population was increased over the next 10 generations to
obtain a founder population (Gen\textsubscript{1}) of 500 males and 500 females. Habier et al. (2009) showed that these simulation parameters result in populations that are in mutation-drift equilibrium with respect to linkage disequilibrium (LD) and distribution of allele frequencies.

From Gen\textsubscript{1} onwards 100 males and 100 females were selected as parents for the next generation based on the OC method using either the A or the G matrix as described before. The selected males and females were randomly mated according to probabilities that reflect their assigned contribution for the next generation. Each mating pair produced ten offspring resulting in 500 males and 500 females as new selection candidates.

The genome simulated consisted of 10 chromosomes each of length 100 cM and thus covering a total of 1000 cM. In the initial generation Gen\textsubscript{1010}, 5001 equally spaced SNP loci per chromosome were simulated, with alleles sampled from a Bernoulli distribution with frequencies of 0.5. In Gen\textsubscript{1}, when the pedigreed population started, 500 SNPs with a minor allele frequency (MAF) ≥ 0.1 were randomly assigned to represent QTL. Effects for each QTL were sampled from a Gamma distribution with shape parameter 0.4 and inverse scale parameter 1.66, following Hayes and Goddard (2001). In addition, the effect of each QTL was scaled in Gen\textsubscript{1} such that the heritability of the quantitative trait was equal to 0.1. SNPs used in genomic evaluations were also chosen in this generation. Each chromosome was divided into 200 bins with an equal number of SNPs in each bin at first. Afterwards, from each bin the SNP with a frequency closest to 0.5 was selected. This resulted in a total number of 2000 SNP markers for the analyses. Phenotypes were calculated as the sum of genetic effects of an individual plus a residual effect sampled from a standard normal distribution.

After each round of selection and mating, breeding values were estimated either using a BLUP animal model or GS. For both BLUP and GS, phenotypic and pedigree data were available from all generations from Gen\textsubscript{1} to the present such that the prediction model for GS was updated every generation.

To develop prediction equations for GS, the BayesB method of Meuwissen et al. (2001) and Habier et al. (2007) was used. The basic model to predict SNP effects can be written as:

\[ y_i = \mu + \sum_j X_{ij} g_i + e_i, \]

26
where \( y_i \) is the trait phenotype of individual \( i \), \( \mu \) is the overall mean, \( X_{ij} \) is the number (0,1 or 2) of copies of allele 1 that individual \( i \) carries at SNP \( j \), \( g_i \) is the effect of SNP \( j \), and \( e_i \) is a random residual.

For selection candidates, GEBV were computed based on their SNP genotypes as:

\[
GEBV = \sum_j X_{ij} g_i.
\]

BLUP breeding values were estimated using a standard animal model and the available phenotypic data and a numerator relationship matrix based on the pedigree going back to Gen_1 (Henderson, 1984).

**RESULTS**

**Inbreeding**

The average coefficient of inbreeding either based on \( A \) or \( G \) is given in Figure 1 when either pedigree inbreeding (Figure 1a and c) or genomic inbreeding (Figure 1b and d) was constrained. Results are averages of 80 replicates of each breeding scheme.

In all four scenarios, the course of the inbreeding followed the expectation on the scale it was constrained. In three (Figure 1a – c) situations, the average inbreeding coefficients based on pedigree were lower than the genomic inbreeding coefficients. When genomic inbreeding (Figure 1d) was constraint and animals were selected for GEBV, the average coefficients of genomic inbreeding were lower in all generations of selection. The highest difference could be observed when selection based on GEBV and constraining pedigree inbreeding (Figure 1a). In this case, genomic inbreeding was around 5 times higher than pedigree inbreeding after 10 generations of selection. Additionally, the genomic inbreeding coefficient was higher than zero in generation two. This differed from the course of the pedigree inbreeding coefficient, who was zero for the first two generations of selection starting from a base generation.
**Figure 1:** Pedigree and genomic based inbreeding averaged over 80 replicates of the breeding scheme for traditional BLUP estimated breeding values and genomic estimated breeding values (GEBV), when the rate of inbreeding bases on pedigree (a and c) or genomic data (b and d) is constrained to 1% in each generation. The trait has an heritability of 0.1 and the number of selection candidates is 1000 (500 males and 500 females) in each generation.

**Response to selection**

Figure 2 shows the cumulative response to selection for the various simulated strategies. For all strategies, the responses were the highest for the initial generation but then declined over generations. Responses were greatest when constraining pedigree inbreeding instead of genomic inbreeding for both estimated breeding values whereas selecting for GEBV led to somewhat higher responses especially in later generations.
After 10 generations of selection the cumulative response of the GEBV breeding scheme was 44% higher than in the BLUP breeding scheme, when constraining genomic inbreeding and 33% higher when constraining pedigree inbreeding.

![Graph showing cumulative response to selection over generations for GEBV and BLUP breeding schemes.](image)

**Figure 2:** Cumulative response to selection (in phenotypic standard deviations) based on 80 replicates of the breeding scheme for traditional BLUP estimated breeding values and genomic estimated breeding values (GEBV) when the rate of inbreeding (bases on pedigree or genomic data) is constrained to 1% in each generation. The trait has an heritability of 0.1 and the number of selection candidates is 1000 (500 males and 500 females) in each generation.

**Accuracy of selection**

Figure 3 shows the changes in accuracy over generations and between selection methods, calculated as the correlations between the true and estimated breeding values. Accuracy decreased for all four selection strategies continuously from Gen1 to Gen10. The decrease in pedigree OC strategies was greater than in genomic OC strategies. The two breeding schemes with GEBV yielded higher accuracies over all generations as BLUP breeding schemes.
Figure 3: Accuracy of selection based on 80 replicates of the breeding scheme for traditional BLUP estimated breeding values and genomic estimated breeding values when the rate of inbreeding (based on pedigree or genomic data) is constrained to 1% in each generation. The trait has an heritability of 0.1 and the number of selection candidates is 1000 (500 males and 500 females) in each generation.

**Genetic Variance**

The trends of the genetic variance are presented in Figure 4. Genetic variance was reduced over generations for all scenarios. The reductions in genetic variance were the greatest for the initial generations in all four selection strategies. Constraining genomic inbreeding obtained higher genetic variance, especially in the last six generations of selection compared within BLUP or GEBV strategies. The highest reduction in genetic variance from Gen\textsubscript{1} to Gen\textsubscript{10} was observed for constraining pedigree inbreeding and selecting for GEBV (0.14 to 0.03) whereas the lowest decline could be found for genomic OCS and selection for BLUP EBV (0.14 to 0.7). The reduction in genetic variances in initial generations was primarily the result of selection-induced gametic phase disequilibrium (Bulmer, 1971).
Figure 4: Genetic variance based on 80 replicates of the breeding scheme for traditional BLUP estimated breeding values and genomic estimated breeding values when the rate of inbreeding (based on pedigree or genomic data) is constrained to 1% in each generation. The trait has an heritability of 0.1 and the number of selection candidates is 1000 (500 males and 500 females) in each generation.

QTL frequencies
Changes in QTL frequencies due to selection are given in Figure 5 and 6, which show the average frequency of favorable QTL alleles (Figure 5) and the proportion of favorable QTL alleles that were lost (Figure 6). Trends in mean frequencies of favorable QTL alleles followed the response to selection shown in Figure 2. This was found for both variants of estimated breeding values. However, when comparing the genomic OC and pedigree OC strategies, the pedigree OC resulted in higher mean frequencies of favorable QTL alleles. When comparing GEBV and BLUP strategies, the breeding programs for GEBV resulted in higher mean frequencies in both methods of constraining the inbreeding of the next generation.
Figure 5: Average frequency of favourable QTL alleles based on 80 replicates of the breeding scheme for traditional BLUP estimated breeding values and genomic estimated breeding values (GEBV) when the rate of inbreeding (based on pedigree or genomic data) is constrained to 1% in each generation. The trait has an heritability of 0.1 and the number of selection candidates is 1000 (500 males and 500 females) in each generation.

A slightly different picture emerged for the proportion of QTL for which the favorable allele was lost (Figure 6). For the GEBV strategies, pedigree OC led to a much greater proportion of QTL alleles that were lost over the generations than genomic OC. When using BLUP EBV the results were conversely. Constraining genomic inbreeding led to a higher proportion of QTL alleles that were lost than constraining pedigree inbreeding, but the difference between these two methods was smaller than in the GEBV strategies. However, the lowest proportion of QTL alleles that were lost could be found in BLUP and pedigree OC and GEBV and genomic OC strategies which were approximately on the same level.
Figure 6: Proportion of favourable QTL alleles based on 80 replicates of the breeding scheme for traditional BLUP estimated breeding values and genomic estimated breeding values (GEBV) when the rate of inbreeding (based on pedigree or genomic data) is constrained to 1% in each generation. The trait has an heritability of 0.1 and the number of selection candidates is 1000 (500 males and 500 females) in each generation.

Number of selected animals

The number of selected animals as parents for the next generation was presented in Figure 7 for all selection strategies. For both, GEBV and BLUP EBV, the pedigree OC algorithm selected more animals than the genomic OC algorithm in all generations. However, using BLUP EBV the difference between the two methods of constraining inbreeding was much higher than using GEBV. In the pedigree OC strategy, selection for BLUP EBV led to a higher number of selected animals than selection for GEBV. For genomic OC strategies this ratio was reversed.
**DISCUSSION**

In the recent years, selection tools for maximizing the genetic gain by a predefined rate of inbreeding were developed (Meuwissen, 1997; Grundy et al., 1998; Meuwissen and Sonesson, 1998). These OC methods have been applied in breeding programs of several livestock species, for instance in the UK and the German Holstein dairy population (Kearney et al., 2004; Koenig and Simianer, 2006), in two British livestock populations of sheep and beef cattle (Avendano et al., 2003) and in a large salomon fish breeding stock (Hinrichs et al., 2006). These authors used the OC method of Meuwissen (1997) which constrained inbreeding based on pedigree data.

**Figure 7**: Number of selected animals based on 80 replicates of the breeding scheme for traditional BLUP estimated breeding values and genomic estimated breeding values when the rate of inbreeding (based on pedigree or genomic data) is constrained to 1% in each generation. The trait has an heritability of 0.1 and the number of selection candidates is 1000 (500 males and 500 females) in each generation.
However, false and missing sire information is a well-known problem for setting up accurate additive genetic relationships. Furthermore, these relationships and inbreeding coefficients are only expected and no realized values. With the use of GEBV, thousands of marker data information become available. Therefore, the use of actual genotypes like in genomic relationship matrices can provide more accurate measures of realized relationships between different animals.

The aim of this study was to compare genomic OC and pedigree OC with selection either for GEBV or BLUP EBV. Therefore, we have evaluated the impact of the four breeding programs on pedigree and genomic rates of inbreeding, response to selection, genetic variance, changes in QTL frequencies and loss of favorable QTL alleles and the number of selected animals, when selecting for a trait with low heritability ($h^2 = 0.1$) by using stochastic simulations.

As the results in Figure 1 show, the OC method could restrict the inbreeding of the scale it was constrained (genomic or pedigree). When using GEBV and restricting pedigree inbreeding, the genomic inbreeding was around 5 times higher in Gen$_{10}$. However, when constraining genomic inbreeding the pedigree inbreeding was only 1.3 times lower. Similar results were reported by Sonesson et al. (2010). When using BLUP EBV and constraining pedigree inbreeding, the difference to genomic inbreeding was much lower than in GEBV strategies. This indicated that the inbreeding at the QTL was higher for GEBV than for BLUP EBV when constraining pedigree inbreeding. As pointed out before genomic OC selected animals that were genetically different. That fact had an impact on the accuracy, response to selection, genetic variance and the frequency of favorable QTL. The accuracy of GEBV decreased much slower when using genomic OC, over the generations (0.65 in Gen$_1$ to 0.54 in Gen$_{10}$). This could be expected because the reasons of declining accuracy for GEBV was the erosion of LD by recombination, changes in marker and QTL frequencies and as a result the fixation of marker and loss of marker QTL LD. As shown in Figure 5 and 6, the frequency of favorable QTL increased more when using pedigree OC, especially with GEBV. This resulted also in a higher proportion of QTL that were lost. The same had to be expected on the marker side which led to a greater loss of the marker QTL LD. Sonesson et al. (2010) showed that when constraining genomic inbreeding the genomic IBD was more variable across the genome, than when constraining pedigree inbreeding. Furthermore, the authors concluded that genomic OC resulted in a quite evenly distributed increase of the IBD
across the genome. Pedigree OC, however, seems to increase the frequency of the largest QTL as quickly as possible. This could also conclude from the presented results. Therefore the decline of genomic OC and GEBV was very small over the generation (0.67 in Gen$_1$ to 0.62 in Gen$_{10}$). The two breeding schemes based on BLUP EBV obtained lower accuracies than the breeding schemes based on GEBV. This difference based on the increased accuracy of the Mendelian sampling term from the use of genetic markers and was also reported in several studies (Meuwissen et al., 2001; Habier et al., 2007; Muir, 2007; Nielsen et al., 2009).

Due to this, the response to selection is higher for pedigree OC than for genomic OC when selecting for GEBV. Especially the higher frequency of favorable QTL alleles and the higher proportion of QTL that were lost, led to the greater decrease of genetic variance as shown in Figure 4.

Furthermore, this study showed, that when constraining pedigree inbreeding and using GEBV much fewer animals were selected compared to BLUP EBV (63 vs. 91 in Gen$_{10}$). This was in agreement with the findings of Nielsen et al. (2010) who also compared breeding programs with the use of GEBV and BLUP EBV and constraining pedigree inbreeding. It indicated that the desired rate of inbreeding was obtained more easily with GEBV than with BLUP EBV due to the higher estimation accuracy of the Mendelian sampling term (Daetwyler et al., 2007; Sonesson and Meuwissen, 2009). However, when constraining genomic inbreeding the BLUP breeding scheme selected much fewer animals than the GEBV scheme (31 vs. 60 in Gen$_{10}$). The number of selected animals differed for GEBV (59 vs. 63 in Gen$_{10}$) not as much as for BLUP EBV (31 vs. 91 in Gen$_{10}$) breeding schemes when constraining different kinds of inbreeding. It indicated also that with the use of GEBV the desired rate of inbreeding could be better obtained because the variation of GEBV was bigger within families. In contrast, BLUP EBV was more equal within families and therefore the pedigree OC selected more animals to obtain the desired rate of inbreeding.

CONCLUSION

All in all, this study showed that it is useful to restrict genomic inbreeding in genetic improvement programs. The inbreeding on the QTL level was lower and the decrease of the genetic variance was smaller.
The advantage of genomic OC was greater in GS schemes, because the selection pressure is higher on favorable QTL with high effects in GS. Genomic OC can maintain higher genomic variability and therefore, a higher response of selection on the long term. Furthermore, constraining pedigree inbreeding leads to an underestimation of the true inbreeding, especially when using GEBV.

ACKNOWLEDGEMENTS

This study is part of the project FUGATO-plus Brain and was financially supported by the German Ministry of Education and Research, BMBF, the Förderverein Biotechnologieforschung e.V. (FBF), Bonn, Lohmann Tierzucht GmbH, Cuxhaven, Verband hannoverscher Warmblutzüchter e.V, Verden, Verband der Züchter des Oldenburger Pferdes e.V., Vechta, Verband der Züchter des Holsteiner Pferdes e.V., Elmshorn, Westfälisches Pferdestammbuch e.V., Münster, and Trakehner Verband, Neumünster.

REFERENCES


Chapter 3

Genomic optimum contribution selection to reduce the effect of hitchhiking

Jan Körte.
Dirk Hinrichs
Georg Thaller

Institute of Animal Breeding and Husbandry, Christian-Albrechts-University Kiel, D-24098 Kiel, Germany
ABSTRACT

In order to investigate the true inbreeding, i.e. identity by descent (IBD), of a QTL and its surrounding region, stochastic computer simulations were performed. Furthermore, the impact of the use of optimized selection using a numerator relationship matrix based on pedigree or genomic information on different kinds of inbreeding was investigated. The effect of both OC methods was studied by varying the heritability ($h^2 = 0.1$ and $0.3$) and the favorable QTL allele effect ($\alpha$ explained 5 or 20% of the total genetic variance). The simulated genome consisted of 10 chromosomes of 100 cM each. The trait under selection was assumed to be controlled by 501 biallelic QTL, where 500 QTL were distributed on chromosome 2 – 10. Chromosome 1 was the chromosome examined and consisted of one QTL placed in the middle, e.g. 50 cM, and 100 evenly distributed IBD tags.

It was shown that constraining pedigree inbreeding underestimated the true inbreeding in the QTL position and the surrounding region, as a result of hitchhiking. This resulted in a higher mean frequency of the favorable QTL allele. However, constraining genomic inbreeding led to a more equal increase of all kinds of inbreeding measures. The true inbreeding in the QTL was slightly lower than the constrained genomic inbreeding.

Therefore, pedigree inbreeding was not a good measure to restrict inbreeding in GS programs obtaining long term-variability and long-term response to selection.

Keywords: genomic optimum contribution, identical by descent, genetic sweep
INTRODUCTION

Genome-wide breeding values (GEBV) are of increasing interest and can be used for genome-wide selection (GS) (Meuwissen and Goddard, 2001; Schaeffer, 2006). Daetwyler et al. (2007) discussed possible changes in rates of inbreeding and genetic gain when applying GS: The accuracy of breeding values predicted on the basis of genomic information will be higher than for BLUP EBV, due to an increased estimation accuracy of the Mendelian sampling component, which leads to a better differentiation, and therefore, to lower co-selection of sibs. This reduces the rate of inbreeding per generation. On the other hand, with applying genome-wide selection the generation interval can be reduced, especially in dairy breeding designs. This may increase the rate of inbreeding per year. Therefore, there still is a need to manage inbreeding and genetic diversity.

Methods have been proposed for selection and mating strategies which can maximize genetic gain while constraining rates of inbreeding to a predefined rate of inbreeding (Meuwissen, 1997; Grundy et al., 1998; Sonesson et al., 2000). Petersen et al. (2010) showed that marker assisted selection reduce pedigree estimated inbreeding but not true inbreeding at the QTL and the region surrounding the QTL. This is a result of hitchhiking (Maynard Smith and Haigh, 1974) which can affect the overall genomic inbreeding. Roughsedge et al. (2008) demonstrated that OC selection restricts inbreeding across the genome but does not restrict inbreeding in the region around the QTL under positive selection. However, these authors used OC method with an additive numerator relationship matrix based on pedigree information. Due to the availability of genome-wide dense marker maps, methods have been applied to restrict inbreeding on a genomic scale (Sonesson et al., 2010). A genomic relationship matrix should be used in combination with optimum contribution selection for a better control of genome wide inbreeding.

The objective of this study was to evaluate the use of genomic OC selection and pedigree OC selection with genomic estimated breeding values (GEBV), on genetic gain, pedigree and genomic estimated inbreeding. Additionally, the identical by descent coefficients of a known QTL and in the surrounding region was estimated.
MATERIALS AND METHODS

Optimum contribution selection
The optimization algorithm for obtaining maximum genetic gain whilst constraining the rate of inbreeding (ΔF) to a specific level, described by Meuwissen (1997), was used. To constrain ΔF to a specific level the additive numerator relationship matrix based on pedigree information (A) or the relationship matrix based on genetic information (G), described by vanRaden (2008) was used and can be written as:

\[ G = \frac{ZZ'}{2 \sum p_i (1 - p_i)} \]

where \( Z \) is a matrix of marker genotypes and \( p_i \) is the frequency of the second allele at locus \( i \).

Simulation of the population and the genome
A closed nucleus program with selection over ten discrete generations was considered. Simulation started 1010 generations before a base population (Gen.1010) consisting of 500 individuals that was randomly mated for 800 discrete generations and then reduced to an effective population size of 100 individuals and randomly mated for another 200 generations. To generate mutation drift equilibrium, the initial generations were simulated with a mutation rate of 2.5*10^{-5} and recombination. To obtain a founder population (Gen1) of 500 males and 500 females, the population was increased over the next 10 generations. Habier et al. (2009) showed that these simulation parameters resulted in populations that were in mutation-drift equilibrium in levels of linkage disequilibrium (LD) and distribution of allele frequencies.

From Gen1, 100 males and 100 females as parents for the next generation were selected with pedigree or genomic OC method, as described above. The selected males and females were randomly mated according to probabilities that reflected their assigned contribution for the next generation. Each mating pair produced ten offsprings such that 500 males and 500 females were produced as selection candidates for the next generation.

The simulated genome consisted of 10 chromosomes of a length 100 cM each, resulting in a total of 1000 cM. The trait was under control of in total 501 bi-allelic QTL. To analyze the effect of optimized selection (using a pedigree or genomic based numerator relationship matrix) on a known QTL and the region around that
QTL, on chromosome 1 only one QTL was positioned at 50 cM. All other 500 QTL were equally sampled on chromosome 2 – 10 so that an infinitesimal model was assumed. Each QTL was assigned an effect sampled from a Gamma distribution with shape parameter 0.4 and inverse scale parameter 1.66, following Hayes and Goddard (2001). In addition, the effect of each QTL was scaled in Gen1 so that the heritability of the quantitative trait was set to 0.1. Thereby, the effect of the QTL on chromosome 1 was assumed to contribute 5 or 20% of the total genetic variance. At first, the initial variance of the QTL was obtained by \( \sigma_{\text{estimated}}^2 = 2p(1-p)\alpha^2 \) (Falconer and Mackay, 1996), where \( p \) is the initial frequency of the favorable QTL allele and \( \alpha \) is the gene substitution effect of the favorable QTL allele. Therefore, \( \alpha \) was rescaled by \( \alpha_{\text{sampled}} \times \frac{\sqrt{\sigma_{\text{desired}}^2}}{\sqrt{\sigma_{\text{estimated}}^2}} \), where \( \alpha_{\text{sampled}} \) is the sampled QTL effect from the Gamma distribution, as described above, and \( \alpha_{\text{desired}} \) is the desired variance of the QTL when explaining 5 or 20% of the total genetic variance.

In the initial generation Gen\(_{1010}\), 5001 equally spaced SNP loci per chromosome were simulated for the first 1010 generations, with alleles sampled from a Bernoulli distribution with frequencies of 0.5.

In Gen\(_1\) SNPs for the analysis were chosen by dividing each chromosome into 200 bins with an equal number of SNPs in each bin. Afterwards, out of each bin the SNP with a frequency closest to 0.5 was selected. This resulted in a total number of 2000 SNP markers for the analyses. Phenotypes were calculated as the sum of genetic effects of an individual plus a residual effect sampled from a standard normal distribution.

In addition, a set of 100 tags were evenly spaced across each chromosome. The tags were set each 1 cM to monitor the IBD status. Each base animal was given a unique genotype for all tags to trace individual tag alleles through the population.

After each round of selection and mating, breeding values were estimated using GS. For GS, genomic data were available from all generations from Gen\(_1\) to the present such that the prediction model for GS was updated every generation.

To develop prediction equations for GS, the BayesB method of Meuwissen et al. (2001) and Habier et al. (2007) was used. The basic model to predict SNP effects can be written as:

\[
y_i = \mu + \sum_j X_{ij} g_j + e_i,
\]
where $y_i$ is the trait phenotype of individual $i$, $\mu$ is the overall mean, $X_{ij}$ is the number (0,1 or 2) of copies of allele 1 that individual $i$ carries at SNP $j$, $g_i$ is the effect of SNP $j$, and $e_i$ is a random residual.

For selection candidates, GEBV were computed based on their SNP genotypes as:

$$GEBV = \sum_j X_{ij} g_i.$$

**Measure of inbreeding and IBD**

Expected inbreeding coefficients based on pedigree information were calculated by using the additive numerator relationship matrix. Inbreeding coefficients based on genomic information were estimated by using the genomic relationship matrix of vanRaden (2008) as described above. The IBD, which represented the true inbreeding in the QTL and over the entire genome, was calculated based on the tags. Therefore, the homozygosity of each tag was measured. When an animal had identical alleles at a given tag, both alleles were derived from the same allele in the same base animal and the animal was defined as IBD for that position. The proportion of animals being IBD at the tag in a particular locus was the true level of inbreeding in the population at a given locus.

**RESULTS**

In Figure 1 and 2 the IBD in each tag across chromosome 1 for animals born in the final Gen10 are presented for a trait with heritability of 0.3 (Figure 1) and 0.1 (Figure 2). For all scenarios the IBD in each tag over the entire chromosome 1 was higher for pedigree OC method than for genomic OC. For all pedigree OC scenarios there was a clear peak in the position of the QTL which represented the IBD of that QTL. For both heritabilities, the level of IBD in each tag across the chromosome 1 was higher with an increasing effect of the QTL. However, for genomic OC there was only a slightly peak in the position of the QTL.
Figure 1: Average IBD on each tag across chromosome 1 of animals borne in generation 10, for a trait under selection with heritability of 0.3. One QTL is placed on 50 cM whose effect explains 5 (a) or 20 (b) % of the total genetic variance.
Figure 2: Average identity by descent (IBD) on each tag across chromosome 1 of animals borne in generation 10, for a trait under selection with heritability of 0.1. One QTL is placed on 50 cM whose effect explains 5 (a) or 20 (b) % of the total genetic variance.

This peak was also higher for a higher effect of the QTL. In addition, the level of the IBD in each tag across the chromosome 1 was only slightly affected by the effect of
the QTL and the heritability. The figure 1 and 2 shows clearly that the pedigree OC method resulted in higher levels of IBD in the tags around the QTL which means in fact higher IBD of loci, assumed to be selectively neutral. However, for genomic OC the increase in the regions around the QTL was much lower.

Figure 3: Mean QTL frequency of the QTL on chromosome 1 which effect explains 5 (a) or 20 (b) % of the total genetic variance of a trait with a heritability of 0.3 as averaged over replicates.
Figure 4: Mean QTL frequency of the QTL on chromosome 1 which effect explains 5 (a) or 20 (b) % of the total genetic variance of a trait with a heritability of 0.1 as averaged over replicates.

The mean frequency of the QTL on chromosome 1 can be seen in Figure 3 for heritability of 0.3 and in Figure 4 for heritability of 0.1. As expected, the frequency of the favorable QTL allele was higher for the pedigree OC for all scenarios. The highest mean frequency was found for a heritability of 0.3 and a QTL effect, explaining 20% of the total genetic variance. The QTL reached 99% fixation in Gen8, which was the only observed scenario where the QTL reached fixation. However, in
no genomic OC scenario the QTL was fixed in the 10 generations. The observed frequency was higher, the higher the effect of the QTL for both levels of heritability was.

**Table1:** Genotypic value (GEN_Val), rate of inbreeding based on pedigree (ΔF_ped) and on genomic (ΔF_genomic) relationship matrices, identity by decent coefficient over entire template (IBD_T) and in QTL position (IBD_Q) at Generation 10 for different selection methods when the QTL on chromosome one explains 5% or 20% of the total genetic variance of a trait with a heritability of 0.3. (standard errors in brackets)

<table>
<thead>
<tr>
<th></th>
<th>GEN_Val</th>
<th>ΔF_ped</th>
<th>ΔF_genomic</th>
<th>ΔIBD_T</th>
<th>ΔIBD_Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>QTL explains 5%</td>
<td>GG</td>
<td>GP</td>
<td>GG</td>
<td>GP</td>
<td></td>
</tr>
<tr>
<td>GEN_Val</td>
<td>4.668 (0.480)</td>
<td>5.883 (0.603)</td>
<td>4.382 (0.464)</td>
<td>5.506 (0.566)</td>
<td></td>
</tr>
<tr>
<td>ΔF_ped</td>
<td>0.009 (0.001)</td>
<td><strong>0.010</strong> (0.001)</td>
<td>0.009 (0.001)</td>
<td><strong>0.010</strong> (0.001)</td>
<td></td>
</tr>
<tr>
<td>ΔF_genomic</td>
<td><strong>0.011</strong> (0.001)</td>
<td>0.043 (0.002)</td>
<td><strong>0.011</strong> (0.001)</td>
<td>0.043 (0.001)</td>
<td></td>
</tr>
<tr>
<td>ΔIBD_T</td>
<td>0.006 (0.001)</td>
<td>0.016 (0.001)</td>
<td>0.009 (0.001)</td>
<td>0.018 (0.001)</td>
<td></td>
</tr>
<tr>
<td>ΔIBD_Q</td>
<td>0.007 (0.002)</td>
<td>0.030 (0.001)</td>
<td>0.009 (0.001)</td>
<td>0.027 (0.002)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>GEN_Val</th>
<th>ΔF_ped</th>
<th>ΔF_genomic</th>
<th>ΔIBD_T</th>
<th>ΔIBD_Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>QTL explains 20%</td>
<td>GG</td>
<td>GP</td>
<td>GG</td>
<td>GP</td>
<td></td>
</tr>
<tr>
<td>GEN_Val</td>
<td>4.668 (0.480)</td>
<td>5.883 (0.603)</td>
<td>4.382 (0.464)</td>
<td>5.506 (0.566)</td>
<td></td>
</tr>
<tr>
<td>ΔF_ped</td>
<td>0.009 (0.001)</td>
<td><strong>0.010</strong> (0.001)</td>
<td>0.009 (0.001)</td>
<td><strong>0.010</strong> (0.001)</td>
<td></td>
</tr>
<tr>
<td>ΔF_genomic</td>
<td><strong>0.011</strong> (0.001)</td>
<td>0.043 (0.002)</td>
<td><strong>0.011</strong> (0.001)</td>
<td>0.043 (0.001)</td>
<td></td>
</tr>
<tr>
<td>ΔIBD_T</td>
<td>0.006 (0.001)</td>
<td>0.016 (0.001)</td>
<td>0.009 (0.001)</td>
<td>0.018 (0.001)</td>
<td></td>
</tr>
<tr>
<td>ΔIBD_Q</td>
<td>0.007 (0.002)</td>
<td>0.030 (0.001)</td>
<td>0.009 (0.001)</td>
<td>0.027 (0.002)</td>
<td></td>
</tr>
</tbody>
</table>

1 GG = GEBV / genomic optimum contribution, GP = GEBV / pedigree Optimum contribution

In all scenarios, ΔF followed the scale it was constrained (Table 1 and 2). When constraining the rate of genomic inbreeding (ΔF_genomic), the rate of pedigree inbreeding (ΔF_ped) was only slightly lower in Gen10. Also the rate of IBD over the entire chromosome 1 (ΔIBD_C) and the rate of IBD in the QTL position (ΔIBD_Q) differed not much, except for a heritability of 0.1 and an effect of the QTL explaining 20% of the total genetic variance. The IBD_Q was then 88% higher than the ΔIBD_C. However, constraining ΔF_ped resulted in an around 330% higher ΔF_genomic in all genomic OC scenarios in Gen10. Interestingly, the ΔIBD_Q and ΔIBD_C was lower than ΔF_genomic, but much higher than ΔF_ped. ΔIBD_Q was between 18% (which was in the range of the scenarios constraining pedigree inbreeding) and 87% higher than the IBD over the entire chromosome 1.

In general, restricting pedigree inbreeding resulted in higher genotypic values than restricting genomic inbreeding when using GEBV for both levels of heritability (Table 1 and 2). Thereby, the difference for a low heritable trait (h²=0.1) was bigger with 33% as for a trait with higher heritability with 26%.
Table 2: Genotypic value (GEN_Val), rate of inbreeding based on pedigree (ΔF_ped) and on genomic (ΔF_genomic) relationship matrices, identity by descent coefficient over entire template (IBD_T) and in QTL position (IBD_Q) at Generation 10 for different selection methods when the QTL on chromosome one explains 5% or 20% of the total genetic variance of a trait with a heritability of 0.1. (standard errors in brackets)

<table>
<thead>
<tr>
<th></th>
<th>QTL explains 5%</th>
<th>QTL explains 20%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GG</td>
<td>GP</td>
</tr>
<tr>
<td>GEN_Val</td>
<td>1.936 (0.199)</td>
<td>2.566 (0.263)</td>
</tr>
<tr>
<td>ΔF_ped</td>
<td>0.009 (0.001)</td>
<td><strong>0.010</strong> (0.001)</td>
</tr>
<tr>
<td>ΔF_genomic</td>
<td><strong>0.011</strong> (0.001)</td>
<td>0.045 (0.002)</td>
</tr>
<tr>
<td>ΔIBD_T</td>
<td>0.008 (0.001)</td>
<td>0.023 (0.001)</td>
</tr>
<tr>
<td>ΔIBD_Q</td>
<td>0.010 (0.001)</td>
<td>0.031 (0.001)</td>
</tr>
</tbody>
</table>

*GG = GEBV / genomic optimum contribution, GP = GEBV / pedigree Optimum contribution

The comparison of different inbreeding levels and measures of IBD coefficients in the last generation of selection for scenarios with heritability 0.1 and 0.3 is presented in Figures 5a and 5b, respectively. Similar trends were found for both heritabilities. For all scenarios the F_genomic was the highest. For genomic OC scenarios, inbreeding and IBD coefficients were close together. Here, F_ped was a good measure of true inbreeding over the genome. For pedigree OC scenarios, F_genomic was around 400% higher than F_ped. The IBD_Q and IBD_C were higher than F_ped, whereas IBD_Q was higher than IBD_C. The effect of the QTL had only a slight influence on IBD_Q and IBD_C (Figure 5a) for heritability 0.3. However, for heritability 0.1, the IBD_Q increased with a higher effect of the QTL. In addition, the IBD_Q position and IBD_C were higher for heritability 0.1 than for heritability 0.3.
**DISCUSSION**

This study has quantified inbreeding in a QTL, with different sizes of the QTL effect and the surrounding region, under selection when $\Delta F$ was constrained using a numerator relationship matrix based on pedigree or genomic information. The constraint of $\Delta F$ was achieved using optimum contribution selection.
The results clearly demonstrated the positive effect of constraining $\Delta F_{\text{genomic}}$ rather than $\Delta F_{\text{ped}}$. As shown in Figure 5, using $F_{\text{ped}}$ underestimated the true inbreeding in the QTL position. The true inbreeding in the QTL position can be up to 138% higher than expected from $F_{\text{ped}}$. Furthermore, $\text{IBD}_{\text{C}}$ was also higher as expected from $F_{\text{ped}}$. Interestingly, the QTL effect only influenced the $\text{IBD}_{\text{QTL}}$ whereas $\text{IBD}_{\text{C}}$ was the same within each heritability. Pedersen et al. (2009) showed that MAS resulted in higher inbreeding in the QTL position than expected from $F_{\text{ped}}$. Pedersen et al. (2010) reported also that in the region around a favorable QTL was extra inbreeding which could affect the overall genomic inbreeding. This increase in overall genomic inbreeding is demonstrated in Figure 5. Because of these negative effects it would be better to use a genomic relationship matrix to control the overall genomic inbreeding. The results showed that the use of a genomic relationship matrix led to a more equal increase of all kinds of inbreeding measures. The inbreeding in the QTL region was lower than the constrained genomic inbreeding and was in the range of the pedigree estimated inbreeding which was much lower than the inbreeding in the QTL region found for pedigree OC scenarios. Pedigree estimated inbreeding based on the assumption that all loci are selectively neutral and, therefore, the two alleles of the same neutral locus on two homologous chromosomes had an equal chance of being selected. This assumption is clearly violated at the QTL. A locus under selection was exposed to a greater rate of inbreeding. But also the region around that QTL had to be taken into account due to linkage. The locus linked to a QTL under selection cannot be assumed to be selectively neutral and, therefore, these loci had a higher true inbreeding than expected from pedigree estimated inbreeding (Figure 1 and 2). This effect of hitchhiking (Maynard Smith and Haigh, 1974) was observed for all scenarios of constraining pedigree estimated inbreeding. As shown in Figure 3 and 4, the frequency of the favorable QTL alleles was higher for pedigree OC rather than for genomic OC methods. This accumulation of the favorable QTL allele may be desirable in a breeding program leading to a higher response to selection (Table 1 and 2). However, with respect to that observed linkage drag, the accumulation of the inbreeding in the regions around a selected QTL had to be taken into account. Also undesirable loci linked to a selected QTL could accumulate or reach fixation. Furthermore, the faster fixation of the favorable QTL alleles increased the short term
selection response, but reduced selection response on the long-term (Villanueva et al., 2002).

Due to the progress in MAS and GAS in the recent years, methods have been proposed to restrict $\Delta F$ in the QTL and in the surrounding region (Meuwissen and Sonesson, 2004; Roughsedge et al., 2008). Roughsedge et al. (2008) constrained group coancestry at a specific position around the QTL by using a relationship matrix computed from the pedigree and genetic markers. They reported that the rate of inbreeding realized at the position of constraint was lower than the $\Delta F$ expected from $F_{\text{ped}}$. However, at the positions around the restriction, the observed inbreeding was 2.5 times greater than the desired value. This approach implied the increase in allele frequencies of unknown genome regions, due to the control of one large QTL. This assumption was not realistic and led to fixation of the next QTL. Therefore the constraint should place over the whole genome.

Daetwyler et al. (2007) and Pedersen et al. (2010) raised the question whether pedigree based estimated inbreeding are valid inbreeding measurements in a future involving genomic selection. They, and also the results presented here, showed that pedigree based estimated inbreeding seriously underestimate true inbreeding when a single QTL is selected by using MAS or GS. However, all scenarios using genomic OC method showed a lower variation of inbreeding across the chromosome 1. This is in agreement with Sonesson et al. (2010). They also combined OC method with a genomic relationship matrix, and concluded that for a trait that has many QTL contributing to the variance the concept of inbreeding at a neutral locus is no longer tenable.

**CONCLUSION**

The results of this study showed that it is not useful to manage pedigree in combination with GS. To prevent the loss of genetic variation of the QTL and especially in positions around the QTL, a genomic based relationship matrix should be used to restrict genomic inbreeding.
ACKNOWLEDGEMENTS

This study is part of the project FUGATO-plus Brain and was financially supported by the German Ministry of Education and Research, BMBF, the Förderverein Biotechnologieforschung e.V. (FBF), Bonn, Lohmann Tierzucht GmbH, Cuxhaven, Verband hannoverscher Warmblutzüchter e.V, Verden, Verband der Züchter des Oldenburger Pferdes e.V., Vechta, Verband der Züchter des Holsteiner Pferdes e.V., Elmshorn, Westfälisches Pferdestammbuch e.V., Münster, and Trakehner Verband, Neumünster.

REFERENCES


General Discussion

Genetic improvements of livestock involve a struggle to achieve the proper balance between intense selection of a small number of parents in the current generation and to maintain sufficient quantitative genetic variation to ensure substantial response in future generations. Recent advances in genetic selection programs have strongly increased the annual response to selection, but rates of inbreeding have likewise increased substantially. Inbreeding reduces the within population genetic variation by making the population more homozygous. The genetic variance decreases linearly as inbreeding increases (Falconer and Mackay, 1996):

\[ V_g(t) = (1 - F_t) V_g(0) , \]

where \( V_g(t) \) is the genetic variance in generation \( t \), and \( F_t \) is the average inbreeding coefficient in generation \( t \).

Furthermore, the frequency of heterozygote animals decreases. A decrease of heterozygosity leads to inbreeding depression for traits affected by directional dominance. Genes that express dominance effects are often positively related to fitness traits. Therefore, a lower fitness is expected when inbreeding increases. In addition, the accumulation of lethal recessive defect genes and inbreeding depressions are of practical economic importance. Several studies quantified the effect of inbreeding on performance and productive traits. In US Holstein population, Smith et al. (1998) reported a regression of -26.7kg for 305-day milk yield and -5.9 days in productive life for an increase of 1% inbreeding. In German Holstein population, Hinrichs and Thaller (2011) found that a one percent increase in inbreeding results in a 0.22% higher risk of stillbirth. For Iberian pigs, a weight reduction of up to 5.37% at day 120 and lower daily gain of 6.49% were reported by Fernandez et al. (2002). Even for wild populations, e.g. red deer, inbreeding depression could be detected (Slate et al., 2000).

These results indicate a possible decrease of the long term response to selection and emphasize the need to manage inbreeding in genetic improvement programs.

Therefore, the aim of the current study was to analyze the optimum contribution procedure in genetic improvement programs. Special emphasis was put on methodologies that make benefit from the marker information now available. Initially, a change of the breeding scheme of a small closed nucleus program was studied. In a second step, the optimum contribution theory described by Meuwissen (1997),
extended to a relationship matrix based on genomic information, was compared to methods using pedigree data. Finally, the effect of this new genomic optimum contribution selection was investigated on the genomic scale.

Managing $\Delta F$ with optimum contributions

Several methods have been proposed to reduce inbreeding. Wray and Goddard (1994) and Brisbane and Gibson (1995) suggested a cost factor approach. The cost of a unit of inbreeding was derived based on the negative effects like inbreeding depression and reduction of genetic variance. The objective function to be maximized is $c'\hat{g} - 1/2kc'Ac$, where $c$ is a vector of genetic contributions of animals, $\hat{g}$ is a vector of estimated breeding values (EBV) of the candidates that were available for selection, $A$ is an matrix with additive genetic relationships of the selection candidates, and $k$ reflects the costs of inbreeding. This approach relies on a correct calculation of the cost factor. If some costs of inbreeding are omitted from the cost factor calculation, the inbreeding of the resulting breeding scheme will be too high. As it is extremely difficult to quantify all inbreeding effects it got evident that the use of the cost factor approach in breeding schemes is not optimal. Meuwissen (1997) replaced the cost factor $k$ by a LaGrangian multiplier to restrict $\Delta F$ and to optimize $c$. The optimum contributions of an individual depend not only on its own EBV and inbreeding coefficient but also on the relationships of that individual to the other selected animals. Grundy et al. (1998) mentioned that the method of Meuwissen (1997) does not achieve a constant rate of inbreeding over several generations, especially because the mean level of inbreeding increases. They modified the existing algorithm using a augmented numerator relationship matrix ($A^*$) and applied the constraint $c' \cdot A^* \cdot c \leq t\Delta C$, where $\Delta C$ is set to $2\Delta F[1 - 3\Delta F + 12(\Delta F)^2]$. However, these methods were developed for discrete generations that were investigated in this study. Grundy et al. (2000) mentioned that optimum contribution methods perform better with discrete generations, because BLUP selection is very efficient with overlapping generations. However, Meuwissen and Sonesson (1998) extended the method of Meuwissen (1997) to overlapping generations for dealing more realistic scenarios. This method optimized the contributions of each animal from many reproductive age classes. Therefore, the expected long term contributions of the age classes were predicted using the gene flow method (Hill, 1974). Grundy et al. (2000) extended the method of Grundy et al. (1998) using the gene flow method as well.
They optimized the long term genetic contributions of each sex-age class with a simulated annealing algorithm (Press et al., 1989).

Sonesson et al. (2000) compared the algorithms of Meuwissen and Sonesson (1998) and Grundy et al. (2000) to determine the optimum contribution of each animal in the current generation to genetic response and inbreeding in the future generation. The iterative algorithm of Meuwissen and Sonesson (1998) gave similar results to the method of Grundy et al. (2000) in a simulation study, and it was suitable for the application of much larger data sets. For practical implementation the algorithm had to handle a large number of selection candidates. Otherwise, candidates for the optimization process had to be preselected which might reduce the benefit of the optimum contribution selection. Nevertheless, several studies reported about the implementation of optimum contribution selection in practical breeding programs and concluded that optimized selection should be used for the management of inbreeding in this context (Avendano et al., 2003; Kearney et al., 2004; Hinrichs et al., 2006; Koenig and Simianer, 2006).

The optimum contribution selection methods were only approximate methods and did not guarantee to find the best solution. Due to the use of LaGrangian multipliers, some contributions could be negative. These negative values, and therefore the contribution of an animal, were set to zero by eliminating those animals from the optimization process and the process was repeated until all contributions were positive. In the method of Meuwissen (1997), all negative contributions in iteration were eliminated together. A better alternative is to eliminate only the candidates with the most negative values. This would provide solutions which are closer to the optimal solution but increase the computational time needed.

The problem with pedigree Information

The estimation of inbreeding coefficients based on extended historical data is not feasible, because early pedigree data does not exist. Therefore, most estimates of inbreeding are calculated in relation to a base year of around 1960, which is the time breed associations started to record ancestry data routinely on computers (Wiggans et al., 1995). Furthermore, setting up an accurate pedigree based numerator relationship matrix requires correct pedigree information. A well-documented problem are wrong and missing sire information. Several studies quantified the proportion of wrong sire information up to 23% in the Holstein Friesian breed (Weller et al., 2004).
The reasons for this finding can originate from failures in artificial insemination companies, wrong recording by the farmer or genotyping errors and can arise from human or technical error (Christensen et al., 1982; Weller et al., 2004). Christensen et al. (1982) mentioned different reasons for paternity errors:

- the artificial insemination center or the technician mistaking the semen of one bull for the semen of another
- the further insemination of a pregnant cow
- the last insemination incorrectly registered due to a misprinting of the bull’s herdbook number or name
- the use of natural service bulls leading to pregnancy of cows which have beforehand been served by artificial insemination bulls and were supposed to be pregnant
- a higher frequency of the interchange of calves due to increasing herd sizes
- the trade in calves and heifers relating to incorrect pedigree information.

As a seventh factor, Weller et al. (2004) included mistakes in the paternity made by the laboratory, which result in rejection of paternity of cows with correct paternity identification.

Inbreeding based on pedigree data involves the theory of measuring inbreeding at neutral loci, i.e. loci that are not under selection and are not linked to loci under selection. However, selection can increase inbreeding even for neutral loci, because it tends to select animals from the same families. As found in chapter 3, loci, e.g. Quantitative Trait Loci (QTL), under selection change allele frequencies and thus inbreeding, but also those at linked loci that are not directly under selection. This phenomenon is called selective sweep and can be seen as a reduced variation at linked sites of loci under selection. A synonym is genetic hitchhiking (Maynard Smith and Haigh, 1974), i.e. the change in the allele frequency of an allele due to selection on a closely linked locus with a positive allele.

**Use of molecular marker information**

As mentioned before, in most standard genetic evaluations founder animals are the first generation recorded. It is assumed that they share no genes from older ancestors. Due to this assumption, relationship and pedigree coefficients from later generations are estimated based on the relatedness to the founders and common ancestors. Genomic analysis showed that this assumption is not true and indeed,
founder animals share genes identical by descent. This shifts the relationship and inbreeding coefficient up or down down also.

In the last decade, the costs of genotype information have been reduced by orders of magnitude, such that this information has become more affordable for science and for commercial applications. Primarily, the advances in both SNP discovery and high density SNP genotyping open new opportunities to quantify inbreeding. SNP information helps for a better estimation of the true situation. Furthermore, marker data enable the estimation among populations (Eding and Meuwissen, 2001). Therefore, the pedigree based relationship matrix was replaced by a genomic based relationship matrix on chapter 2 and chapter 3. A genomic relationship matrix can be calculated by different methods. In chapter 2 and chapter 3 the method of vanRaden (2008) was used. Another often used method is the similarity index proposed by Eding and Meuwissen (2001). The authors used microsatellite markers but dense SNP marker information can be implemented as well (Schierenbeck, unpublished data). For both methods, only molecular data can be used which are not present for all animals. To handle these problems, methods have been proposed to combine pedigree and genomic relationship matrices (Legarra et al., 2009; Aguilar et al., 2010; Bömcke et al., 2010). To build up the genomic relationship matrix and scale them equally to pedigree based numerator relationship matrices, all these methods uses the allele frequencies from the unselected base populations. This information can be rarely extracted from historical data and approximations have to be used. One simple and often used method was presented by Gengler et al. (2007). However, many current studies deal with the development of these so called imputation methods (e.g. Daetwyler et al., 2010; Hickey et al., 2010).
Figure 1: Relationship based on marker versus relationship based on pedigree. The solid line represents relationship based on marker = relationship based on pedigree.

The approximations from the imputation methods can lead to incorrect results in biased relationships and inbreeding coefficients especially for young animals (Aguilar et al., 2009). Forni et al (2011) studied different options of creating genomic relationship matrices in order to find the optimal one. They obtained a genomic numerator relationship matrix following vanRaden (2008), as it was done in chapter 2 and chapter 3, for which the frequencies from the second allele should be from the unselected base population. Instead of frequencies from the base generation they used frequencies of 0.5 for all markers, the average minor allele frequency and the observed frequency of each SNP marker. VanRaden (2008) also treated the problem of the correct approximation of the frequencies. He computed genomic relationships using different allele frequencies:

- the true allele frequencies from the base population
- an estimation of the allele frequencies of the base population obtained with a linear model that solves for gene content of non-genotyped ancestors and descendants using pedigrees (Gengler et al., 2007)
- a simple estimation as the mean of only the known genotypes.

VanRaden (2008) found correlations between true and estimated frequencies in the base population of 0.98, but only of 0.94 with simple frequency estimates from the current population. The two estimates were correlated with each other by 0.97. They found a downward bias in the estimation of genomic inbreeding coefficients if using estimated frequencies. The mean was 7% using true frequencies, but was -4% using base frequency estimates and -2% with simple frequency estimates compared with 5% for traditional inbreeding coefficients of a simulated Holstein pedigree. Leutenegger et al. (2003) also reported that frequency estimation can bias genomic inbreeding coefficients.

In chapter 2 and chapter 3 stochastic simulations were used. The base allele frequencies were calculated as the mean of the true genotypes of the base animals. However, the knowledge of these frequencies is important for practical implementation. Further research on this topic is needed.

A note on the simulation scheme
Stochastic computer simulations were used to investigate the impact of genomic OC compared to pedigree OC with BLUP EBV and GEBV on responses to selection and rates of inbreeding. We have evaluated the characteristics of this response in terms of variance of response, changes in QTL frequencies and loss of favorable QTL alleles in chapter 2. Similar simulations were used in chapter 3 to study the true inbreeding in the QTL position and the region surrounded for both selection strategies using GEBV.

The accuracy of GEBV that can be obtained with GS and the resulting responses to selection depends on several factors. The main factors were summarized by Goddard (2008) and Hayes et al. (2009):
- the number of animals with phenotypes and genotypes in the reference population from which SNP effects were estimated
- the heritability of the trait
- the structure and reliability of the breeding values used
- the level of linkage disequilibrium (LD) between marker and QTL
• the density of the SNP panel
• the distribution of the QTL effects and their allele frequencies.

The extent of LD that is present in a population and that can be captured by a genome-wide SNP panel, is inversely proportional to the product of historical effective population size (\(N_e\)) and the average genetic distance between neighboring SNPs (d) in cM (Solberg et al., 2006). In the present study, effective population size was assumed to be 500 in the distant past and 100 in the past 200 generations. The marker density used in the simulations, e.g. 2,000 SNPs on a 1000 cM genome has been shown to be effective to generate accurate GEBV in other simulation studies (Meuwissen and Goddard, 2001; Meuwissen et al., 2001; Solberg et al., 2006; Habier et al., 2007). The chosen density is lower than what is currently state of the art in livestock populations, i.e. 50,000 SNPs on a 3,000 cM genome. However, it should be recognized that only 75% of the 50,000 SNPs may be segregating in the population, that halving the number of SNPs did not have a large impact on the accuracy of GEBV in dairy cattle (vanRaden et al., 2009), and that historical effective population sizes are larger for most livestock populations than assumed here. The parameters used here are within the range of \(N_e \times d\) that apply to commercial livestock breeding populations. The number of segregating QTL simulated was 500, which implies de facto a polygenic genetic model. The number of simulated QTL was much higher than that used in other studies (Meuwissen et al., 2001; Habier et al., 2007; Solberg et al., 2008). Results from recent analyses using genome-wide SNP analyses suggest that the number of QTL may even be much greater. For the training population, we assumed a data set of 1,000 individuals added in every generation. This is similar to what has been used in other simulation studies (Meuwissen et al., 2001; Solberg et al., 2008; Habier et al., 2007). However, Goddard (2008) showed that much larger training data sets may be required if the number of QTL is large and QTL effects are small. The need for much larger training data has also been demonstrated in the recent application of GS in dairy cattle populations (Van Raden et al. 2009) but may again be the result of the population structure rather than the result of a large number of QTL. However, a larger training size would increase the computational time substantially and would prevent higher number of replicates for the various scenarios in this simulation study.

The Bayes-B method of Meuwissen et al. (2001) was used for the development of prediction models for GS. This method has been found to give greater accuracies
than other methods of many simulation studies (Meuwissen et al., 2001, Habier et al., 2007). The method was, however, not optimized with respect to the prior probability that a SNP has zero effect. Optimizing this prior may result in some increases in accuracy.

REFERENCES


Hinrichs, D., and G. Thaller. 2011. Pedigree analysis an inbreeding effects on calving traits in large dairy herds in Germany. Submitted for publication.


contribution selection. In: Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany


General Summary

In the last decade management methods have been proposed, e.g. optimum contribution selection, for maximizing the genetic gain whilst constraining the increase in inbreeding. The decrease of genetic variance influenced by inbreeding and drift has to be taken into account when designing breeding programs. Therefore, this thesis investigated the use of optimum contribution selection in genetic improvement programs via simulation studies. Special emphasis was led on the integration of genomic marker information.

The aim of chapter 1 was to analyze the change of a conventional breeding program with truncation selection on BLUP estimated breeding values to two different breeding programs based on optimum contribution selection. Two major aspects were of interest. At first, how much additional genetic gain could be achieved while keeping the rate of inbreeding constant. Thereafter, it was analyzed how much the inbreeding could be decreased, if a specific genetic gain was achieved, when changing to optimum contribution selection. In case of a constant rate of inbreeding after changing the selection the genetic gain of optimum contribution selection was 41% ($\Delta F = 0.01$) and 10% ($\Delta F = 0.025$) higher compared to 20 generations of truncation selection. For the situations with constant genetic gain inbreeding was decreased up to 31% ($\Delta F = 0.005$), when selection changed from truncation selection to optimum contribution selection after 10 generations, compared to 20 generations of truncation selection.

In chapter 2 the optimum contribution selection method was extended to the genomic relationship matrix. This genomic optimum contribution selection was compared with the pedigree optimum contribution selection. The impact of these breeding schemes on pedigree and genomic rates of inbreeding, responses to selection, changes in QTL gene frequencies, loss of favorable QTL alleles and number of selected animals was evaluated for a trait with a heritability of 0.1. Inbreeding followed on the scale it was constrained, either pedigree or genomic. In case of constraining pedigree inbreeding and selecting for genomic estimated breeding true inbreeding was underestimated. After 10 generations of selection the cumulative response of the genomic estimated breeding values breeding scheme
was 44% higher than in the BLUP breeding scheme, when constraining genomic inbreeding and 33% higher when constraining pedigree inbreeding. For all selection strategies, accuracy of the estimated breeding values decreases, whereas the decrease in pedigree optimum contribution strategies is greater. Trends in mean frequencies of favorable QTL alleles followed the response to selection whereby the pedigree optimum contribution resulted in higher mean frequencies than genomic optimum contribution.

The presented results indicated that it is useful to restrict genomic inbreeding in genomic improvement programs. The genomic variability in terms of lower frequencies of favorable QTL and lower proportion of QTL that were lost and a smaller decrease of genetic variance was maintain higher in genomic optimum contribution. This could lead to higher response of selection on the long term.

The objective of chapter 3 was to investigate true inbreeding, e.g. identity by descent (IBD), of a QTL and the surrounding region using genomic optimum contribution or pedigree optimum contribution with genomic estimated breeding values. The true inbreeding in the QTL position and the surrounding region was underestimated using pedigree optimum contribution. This was also reflected in higher mean frequency of the favorable QTL allele. However, using genomic optimum contribution leads to a more equal increase of all kinds of inbreeding. It was concluded that pedigree inbreeding is not a good measure to restrict inbreeding in breeding programs selecting for genomic estimated breeding values to obtain long-term variability and therefore, long-term response to selection.
Zusammenfassung

In den vergangenen Jahren wurden Methoden entwickelt, die den genetischen Fortschritt bei einer zuvor festgelegten Inzuchtrate maximieren. Mit Hilfe dieser „Optimum Contribution“ Methoden lässt sich die Abnahme der additiv genetischen Varianz, die einen wichtigen Aspekt in Zuchtprogrammen darstellt, verringern.

In Kapitel 1 wurde die Umstellung einer konventionellen Abschnittsselektion nach BLUP Zuchtwerten auf die „Optimum Contribution“ Selektion analysiert. Die Untersuchung verfolgte zwei Schwerpunkte. Zum einen, wie viel genetischer Fortschritt erzielt werden kann, wenn die Rate der Inzucht konstant gehalten wird. Zum anderen, um welchen Anteil die Rate der Inzucht gesenkt werden kann, wenn ein bestimmter genetischer Fortschritt erzielt werden soll. Nach 20 Generationen des Selektionsschemas konnte mit der Umstellung der Selektionsmethode ein um 41% höherer Genetischer Fortschritt erzielt werden, wenn die Rate der Inzucht auf 0.01 beschränkt wurde. Bei einer Beschränkung der Rate der Inzucht auf 0.025 lag der genetische Fortschritt um 10% höher als bei der vergleichbaren Abschnittsselektion.


Daher lässt sich sagen, dass in Zuchtprogrammen, in denen genomische Daten zur Zuchtwertschätzung herangezogen werden auch die genomische Inzucht in Selektionsentscheidungen berücksichtigt werden sollte, da die genomische „Optimum Contribution“ Selektion zu einer höheren genetischen Variabilität in Form von niedrigeren Frequenzen der günstigen QTL Variante, einem niedrigeren Anteil an verlorenen QTL Allelen und eine geringere Abnahme der genetischen Varianz führt.

Danksagung

An dieser Stelle möchte ich mich bei den Personen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

Mein besonderer Dank gilt Herrn Prof. Dr. Georg Thaller für die Überlassung des Themas, die wissenschaftliche Betreuung und die gewährten Freiräume bei der Erstellung der Arbeit sowie für die Möglichkeit, meine Ergebnisse auf Tagungen im In- und Ausland vorstellen zu dürfen.

Für die Übernahme des Koreferats bedanke ich mich bei Herrn Prof. Dr. Henner Semianer.

Ein großes Dankeschön geht an Herrn Dr. Dirk Hinrichs für wertvollen Diskussionen und die fortwährende Hilfsbereitschaft und Unterstützung bei der Erstellung der Arbeit.

Für die finanzielle Unterstützung im Rahmen des FUGATO-plus Projektes Brain danke ich dem Bundesministerium für Bildung und Forschung, dem Förderverein Biotechnologieforschung e.V., der Lohmann Tierzucht GmbH, dem Verband hannoverscher Warmblutzüchter e.V., dem Verband der Züchter des Oldenburger Pferdes e.V., dem Verband der Züchter des Holsteiner Pferdes e.V., dem Westfälischen Pferdestammbuch e.V. und dem Trakehner Verband.


Mein größter Dank gilt jedoch meiner Familie und Verena. Ihr habt immer an mich geglaubt, mich unterstützt und mir Rückhalt gegeben. Vielen Dank dafür.
Lebeslauf

Name Jan Körte
Geburtsdatum 06.04.1984
Geburtsort Greifswald
Familienstand ledig
Staatsangehörigkeit deutsch

Schulausbildung
1990 – 1992 Maxim Gorki Grundschule Grammendorf
1994 – 2003 Gymnasium Grimmen
Abschluss: Allgemeine Hochschulreife

Studium
2003 – 2006 Studium der Agrarwissenschaften,
Fachrichtung Nutztierwissenschaften,
Christian-Albrechts-Universität zu Kiel
Abschluss: Bachelor of Science

2006 – 2008 Studium der Agrarwissenschaften,
Fachrichtung Nutztierwissenschaften,
Christian-Albrechts-Universität zu Kiel
Abschluss: Master of Science

Berufliche Tätigkeiten
2008 – 2011 Wissenschaftlicher Mitarbeiter am Institut für Tierzucht
und –haltung der Christian-Albrechts-Universität zu Kiel
bei Herrn Prof. Dr. Georg Thaller