Theor Appl Genet (2009) 118:1133–1142
DOI 10.1007/s00122-009-0968-7

ORIGINAL PAPER

Optimum contribution selection in large general tree breeding populations with an application to Scots pine

Jon Hallander · Patrik Waldmann

Received: 6 August 2008 / Accepted: 8 January 2009 / Published online: 31 January 2009

Abstract Development of selection methods that optimises selection differential subject to a constraint on the increase of inbreeding (or coancestry) in a population is an important part of breeding programmes. One such method that has received much attention in animal breeding is the optimum contribution (OC) dynamic selection method. We implemented the OC algorithm and applied it to a diallel progeny trial of *Pinus sylvestris* L. (Scots pine) focussing on two traits (total tree height and stem diameter). The OC method resulted in a higher increase in genetic gain (8–30%) compared to the genetic gain achieved using standard restricted selection method at the same level of coancestry constraint. Genetic merit obtained at two different levels of restriction on coancestry showed that the benefit of OC was highest when restriction was strict. At the same level of genetic merit, OC decreased coancestry with 56 and 39% for diameter and height, respectively, compared to the level of coancestry obtained using unrestricted truncation selection. Inclusion of a dominance term in the statistical model resulted in changes in contribution rank of trees with 7 and 13% for diameter and height, respectively, compared to results achieved by using a pure additive model. However, the genetic gain was higher for the pure additive model than for the model including dominance for both traits.

Introduction

One of the most difficult issues in breeding programmes is to find optimal balance between increased genetic merit and the rate of inbreeding (Robertson 1961; Toro and Pérez-Enciso 1990; Lindgren and Mullin 1997). An aggressive selection strategy with few highly ranked individuals will reduce genetic variation and may cause high levels of inbreeding depression, which in turn might jeopardise the long-term response to selection. On the other hand, a restrictive strategy with larger number of selected individuals results in lower genetic gain, especially during the initial generations of selection. For species with long generation time, for example most forest tree species, it is crucial to use efficient breeding strategies that yield good initial selection response in a few generations. The efficiency of a breeding strategy depends on many parameters that have been studied in detail; among the most important are genetic parameter and breeding value predictions, selection method and mating design (e.g. Meuwissen 1997; Fernandez and Toro 1999; Waldmann et al. 2008). Nevertheless, non-optimal methods are still used in practical tree breeding programmes despite the advancement of methodological development in the animal breeding literature (Gianola 2001; Woolliams 2007).

Many tree breeding strategies are based on breeding value prediction and selection from family designs. A strict within-family selection procedure is known to maintain highest possible genetic variability in the breeding population if the parental contributions to each generation are equal (Sánchez 2000). However, the short-term response to selection will be low compared to a more intensive selection strategy. Furthermore, a fixed parental contribution results in a static selection scheme which has been shown to produce a non-optimal short term increase in selection response

---

Communicated by M. Sillanpää.

J. Hallander (✉) · P. Waldmann
Department of Forest Genetics and Plant Physiology,
SLU, 901 83 Umeå, Sweden
e-mail: jon.hallander@genfys.slu.se

(Woolliams 2007). A more favourable alternative is a dynamic selection scheme which finds the number of candidates and their respective mating proportions that maximises genetic gain in the breeding population given a restriction on the level of coancestry of each generation (e.g. Meuwissen 1997).

One successful dynamic approach used in animal breeding is the optimum contribution (OC) algorithm which maximises selection response while constraining the level of inbreeding in the population by optimising mating proportions of the parents (Meuwissen 1997; Grundy et al. 1998; Hinrichs et al. 2006). In tests, the OC method has outperformed standard truncation BLUP selection at the same predefined rate of inbreeding and improved the long-term response to selection by 20–60% (Meuwissen 1997). The method has been used in practical animal breeding, for instance selection in UK Holstein dairy cattle population (Kearney et al. 2004), selection in two British livestock populations of sheep and beef cattle (Avendaño et al. 2003) and selection in a large salmon fish breeding stock (Hinrichs et al. 2006). Because of the differences in mating systems between trees and animals, Kerr et al. (1998) extended the OC algorithm to a tree breeding context. However, the OC algorithm has rarely been used in practical applications outside the animal breeding framework.

In the field of forest tree breeding, large emphasis has been devoted to develop selection methods and apply these methods to breeding situations (e.g. Lindgren and Mullin 1997; Rosvall et al. 2003; Lstiburek et al. 2005). Mathematical linear programming has been used to select candidates over one generation in *Eucalyptus globulus* (Fernandez and Toro 2001) and *Pinus sylvestris* (Andersson et al. 1999). Fernandez and Toro (2001) investigated the maximum reduction in coancestry at a given level of genetic gain, whereas Andersson et al. (1999) optimised selection at a restricted level of relatedness. Unfortunately, the method used in Andersson et al. (1999) is limited to family designs and cannot handle complex breeding designs or pedigrees over multiple generations. Lindgren and Mullin (1997) used a weight between group coancestry and genetic gain to maximise a population merit criterion by using an iterative search approach. However, they did not allow for varying number of matings for each generation in the optimisation of the long term response to selection, and they used a static selection approach of candidates (i.e. fixed number of selections for each generation). The method of Lindgren and Mullin (1997) has been further developed to dynamically select individuals in different breeding applications, and applied to both simulated and real data (e.g. Zheng et al. 1997; Olsson et al. 2000; Rosvall et al. 2003; Wei and Lindgren 2006; Stoehr et al. 2007). One limitation with their approach is that even for moderately large data sets, iterative search approaches tend to become computationally very time consuming. Furthermore, it is not straightforward to evaluate the weight between genetic improvement and group coancestry, which is important for the outcome of the approach of Lindgren and Mullin (1997). None of these studies have, however, considered simultaneously selecting candidates and their mating proportions dynamically. Kerr et al. (1998) used OC selection on simulated data, but did not compare the OC selection to any standard strategy at the same level of increase in inbreeding. In addition, due to the need of inverting additive relationship matrix between candidates in each iteration in the selection procedure, they were forced to preselect a number of selection candidates in each generation, and as a result, optimal solution achieved from OC could not be guaranteed. Hence, there is a need to further develop the approach of Kerr et al. (1998) and apply it on real data.

In breeding most attention has been focussed on the additive genetic variation since it determines the heritability and the possible response to selection. Non-additive variances have generally been ignored because non-additive effects are difficult to utilise in breeding programmes and require high-quality pedigrees as well as good computational resources (Misztal 1997). However, the rank of selection candidates might be considerably altered if non-additive genetic interaction terms are included in the statistical model because the breeding values will be more accurately predicted (e.g. Wall et al. 2005; Serenius et al. 2006; Waldmann et al. 2008). Additionally, if the breeding population contains large full-sib families, which often is the case in forest tree breeding populations, it is erroneous to assume a purely additive covariance structure (Lynch and Walsh 1998).

The main objective of the current study is to implement the flexible and powerful pedigree OC approach of Kerr et al. (1998) and show how the method can be used to select breeding candidates in a large tree breeding population. The OC dynamic selection algorithm maximises the genetic gain at a predefined level of coancestry increase and the obtained mating proportions are converted into actual number of crosses. Application of the method is illustrated using a large unbalanced diallel progeny trial of Scots pine from the Swedish tree breeding program. We evaluate the performance of the OC method by comparing it with an existing restricted selection breeding strategy and strict truncation selection. To obtain predictions of breeding values in the breeding population, the individual tree model is used in a mixed model equation (MME) framework including a dominance term. Because the method is employed in a large data set, the relationship between included candidates (i.e. the additive relationship matrix) is computed using parental contributions to save computational time. In addition, we will examine how inclusion of a dominance term into the genetic evaluation procedure affects the outcome of the OC method.

Methods

Selection procedure

Meuwissen (1997) introduced the optimum contribution (OC) algorithm for simultaneous selection of candidates and calculation of their respective mating proportion, given a restriction on relatedness in the breeding population. First, the additive relationship matrix, $\mathbf{A_t}$ (of order $n_s \times n_s$), is obtained between all selection candidates from the additive relationship matrix between all individuals in the pedigree, where the number of selected candidates is n_s. Additionally, a vector $\mathbf{a_t}$ of order n_s containing the estimated breeding values (EBV) of the selection candidates is constructed based on the BLUP evaluation (see below). To maximise the genetic merit of the offspring, G_{t+1}, the following linear relationship is used

$$G_{t+1} = \mathbf{c}_\mathbf{t}'\mathbf{a_t}, \quad (1)$$

where $\mathbf{c}_t{}'$ is a vector containing the mating proportion of each candidate in the current breeding population. The restriction on group coancestry is quadratic and gives

$$C_{t+1} = \mathbf{c}_\mathbf{t}'\mathbf{A_t}\mathbf{c_t}/2, \quad (2)$$

which holds if the increase in group coancestry is small between generations. Additional constraints are: (1) the sum of all contributions is one (i.e. $\mathbf{1}'\boldsymbol{c}_t = 1$), where $\mathbf{1}$ is a $n_s \times 1$ vector of ones (2) one particular individual, i, cannot transfer a negative contribution to the next generation ($c_{t,i} \geq 0$). The optimisation problem is solved by introducing two LaGrangian multipliers, λ_0 and λ_1, and by maximising the corresponding objective function

$$f(\mathbf{c_t}) = \mathbf{c}_\mathbf{t}'\mathbf{a_t} - \lambda_0\left(\mathbf{c}_\mathbf{t}'\mathbf{A_t}\mathbf{c_t} - 2C_{t+1}\right) - \lambda_1\left(\mathbf{c}_\mathbf{t}'\mathbf{1} - 1\right). \quad (3)$$

The individual having the most negative mating proportion after finding the maximum in (3) is removed from the process and the system is solved again. This iteration procedure continues until all remaining individuals have non-negative contributions which imply that the final solution is obtained. For further details regarding how to maximise (3), see Meuwissen (1997) and Kerr et al. (1998).

If the above procedure is employed in pedigrees with large number of selection candidates, the required inversion of $\mathbf{A_t}$ in (3) might be very time consuming and memory demanding. To overcome this hurdle, Hinrichs et al. (2006) derived an algorithm to compute $\mathbf{A}_\mathbf{t}^{-1}$ by expressing the additive relationship between two candidates based on their parents' relationships, which would significantly speed up the iteration procedure when solving (3).

Since LaGrangian multipliers are used to solve the object function, recommended values of the mating proportions ($\mathbf{c_t}'$) are not integer values. Hence, $\mathbf{c_t}'$ needs to be converted into actual number of crosses for each tree. The number of crosses for each tree was obtained by first multiplying $\mathbf{c_t}'$ to twice the number of desired offspring N_{off}, and then $2N_{\text{off}}\mathbf{c_t}'$ was rounded off to an integer value below the real value. The tree having the highest deviation between the integer and real value of $2N_{\text{off}}\mathbf{c_t}'$ had its number of crosses increased by one until the number of offspring sum to N_{off}. This procedure resulted in that some trees with a very small contribution did not get selected at all. However, if all trees selected by OC were allowed to contribute, over 8,000 matings in total were suggested, which might be well out of range for most practical breeding situations.

Scots pine data

The OC method was applied to a real data set of a Scots pine (*P. sylvestris* L.) progeny trial, which has previously been studied by Olsson et al. (2000), Waldmann and Ericsson (2006) and Waldmann et al. (2008). The Forestry Research Institute of Sweden, Skogforsk, selected 52 parent trees from natural stands and crossed them according to a partial diallel design in 1971. The 4,970 surviving offspring represent 202 families. Furthermore, all parent trees were assumed to be unrelated and non-inbred. To better handle the environmental effects in the statistical analysis, the plantation was divided into 70 squared blocks which represent a fixed effect. Two traits were analysed in the current study: trunk diameter at 130 cm (DBH) and total tree height (*H*), both traits were measured at the age of 16 years. The mean values of DBH and *H* were 114 mm and 70.5 cm, respectively. For a detailed description of the crossing design, see Waldmann and Ericsson (2006).

The estimation of breeding values was based on the individual tree model under a multivariate mixed model framework with Gaussian assumptions (Henderson and Quaas 1976; Henderson 1985)

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}_2\end{bmatrix}\begin{bmatrix}\mathbf{a}_1\\ \mathbf{a}_2\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}_2\end{bmatrix}\begin{bmatrix}\mathbf{d}_1\\ \mathbf{d}_2\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix}, \quad (6)$$

where $i = 1, 2$ refers to the trait being analysed, $\mathbf{y}_i$ contains the individual phenotypic records, $\mathbf{a}_i$ and $\mathbf{d}_i$ are location vectors of individual additive and dominance genetic values, respectively. $\boldsymbol{\beta}_i$ is a location vector containing systematic environmental block effects. $\mathbf{X}_i$ and $\mathbf{Z}_i$ are known incidence matrices relating $\boldsymbol{\beta}_i$, $\mathbf{a}_i$ and $\mathbf{d}_i$ to $\mathbf{y}_i$, and $\mathbf{e}_i$ is a vector containing the individual residual errors. In addition, the incidence matrices are identical for each trait (i.e. $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}$ and $\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}$). The matrices $\mathbf{a}' = [\mathbf{a}_1{}'\mathbf{a}_2{}']'$, $\mathbf{d}' = [\mathbf{d}_1{}'\mathbf{d}_2{}']$ and $\mathbf{e} = [\mathbf{e}_1{}'\mathbf{e}_2{}']$ follow joint multivariate normal distributions with mean null matrix and covariance structure

$$Var\begin{bmatrix}\mathbf{a}\\ \mathbf{d}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{G_A}\otimes\mathbf{A} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G_D}\otimes\mathbf{D} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{R}_0\otimes\mathbf{I}\end{bmatrix}, \tag{7}$$

where **A** and **D** are the additive and dominance relationship matrices, respectively. **I** is the identity matrix and **0** is the null matrix, $\mathbf{G_A}$ and $\mathbf{G_D}$ are 2×2 matrices including additive and dominance (co)variance components, respectively, and $\mathbf{R}_0$ is a 2×2 matrix including residual (co)variance components. See Waldmann et al. (2008) for details on how to calculate **D**. The software used in the genetic evaluation procedure was ASReml (Gilmour et al. 2006). To obtain reliable starting parameter values for the bivariate evaluation, univariate analyses were first conducted for both traits considered here.

Current breeding strategy

In general, the Swedish Scots pine breeding strategy is based on within-family selection and progeny trial of selection candidates to increase accuracy of genetic evaluations (Wilhelmsson and Andersson 1993). In the breeding population, selected trees are crossed using single pair mating where each tree contributes equally to the breeding population of the next generation. Because completion of one breeding cycle of Scots pine is very time consuming in Sweden, the strategy to perform progeny tests of selection candidates has been questioned (e.g. Olsson et al. 2000; Hannrup et al. 2007). To speed up the time of one rotation, Skogforsk performed selections in one sub-population of Scots pine without using progeny trial to obtain EBV, which was based on the individual performances instead.

For our comparative study, we defined the current breeding strategy by first ranking families according to their average EBV. Then the two highest ranked individuals within each of the 25 best families were selected for further breeding, resulting in a future breeding population of totally 50 trees. We refer to this selection procedure as the RES strategy. To maintain a reasonable level of genetic variability, 2/3 of the 52 founder trees were represented as parent trees of the selected to keep a large number of founder alleles within the breeding population. Additionally, to avoid selecting individuals from half-sib families, parents were not allowed to contribute more than twice. Applying this selection method to the Scots pine data resulted in a level of group coancestry of 2.1% for both traits. This level was used as a restriction in the OC scheme (i.e. $C_2 = 0.021$). Furthermore, a more aggressive breeding strategy was also considered where three selections were made from the best ranked one-third of the families and only one selection was made from the worst ranked one-third of the families. Two selections were made from the intermediate ranked families, which are in line with the standard approach. This resulted in an expected group coancestry of 2.6%. In both strategies, all selected individuals were expected to contribute equally to the next generation, i.e. equal mating proportion of 1/50 for each of the selected trees. We additionally performed standard unrestricted truncation selection on the same pedigree, where all trees were ranked according to EBVs and the top 50 candidates were selected. No restrictions on the relatedness of the selected cohort were imposed on these trees. The truncation selection is referred to as TRU throughout the study.

Results

Comparison of selection methods

The results obtained from the genetic parameter evaluations of the Scots pine data can be found in Table 1. Table 2 shows the performances of the different selection strategies on the expected increase in genetic merit (G_2) and group coancestry in next generation (C_2). The optimum contribution algorithm (OC) outperformed the restricted selection approach (RES) at the same coancestry restriction for both traits. For H, OC gave a relative increase in G_2 of 16 and 8% compared to the results obtained by RES when C_2 was 2.1 and 2.6%, respectively. The corresponding results for DBH showed that the attained G_2 by OC was 30 and 18% higher than RES at the same $C_{2.}$ Hence, the improvement of OC was greatest at the lower level of coancestry, which agrees well with the conclusion made by Meuwissen (1997) and Avendaño et al. (2004).

Figure 1 shows the number of matings of the selected trees and their respective EBV obtained by OC, whereas Fig. 2 shows the total achieved number of crosses of the selected families and their respective average EBVs for H (results similar for DBH). The family contribution for the RES strategy was equal among the represented families when $C_2 = 2.1\%$ and unequal with proportions 3-2-1 when $C_2 = 2.6\%$ as described in the section "Current breeding strategy". The explanation for the marked differences in G_2 between the strategies is partly revealed in Figs. 1 and 2, where it is shown that OC utilises the higher ranking individuals more heavily while selecting trees in lower ranked families to reach the predefined level of C_2. Additionally, this difference can be seen in Table 2, where OC selected 254 and 188 trees from 77 to 44 families, respectively, to keep within the predefined level of group coancestry for DBH. On the other hand, with RES, 50

Table 1 Summary of results of the Scots pine progeny trial for total tree height (*H*) and trunk diameter at 130 cm (DBH)

Parameter	Estimate	SE[a]
σ_a^2 (*H*)	30.6	7.2
σ_a^2 (DBH)	55.3	17.2
r_a (*H*, DBH)	0.644	0.111
σ_d^2 (*H*)	16.3	4.4
σ_d^2 (DBH)	83.9	25.3
r_d (*H*, DBH)	0.831	0.065
σ_e^2 (*H*)	104.4	5.6
σ_e^2 (DBH)	728.1	27.0
r_e (*H*, DBH)	0.830	0.039
h^2 (*H*)	0.202	0.0433
h^2 (DBH)	0.0637	0.0194
d^2 (*H*)	0.108	0.029
d^2 (DBH)	0.0968	0.0289

Fifty two parents were crossed according to a diallel design producing 202 families with a total of 4,970 surviving trees

r_a, r_d and r_e are additive, dominance and environmental correlation between traits, respectively. *H* were measured in cm and DBH in mm

[a] Standard error

Table 2 Level of coancestry in the second generation (C_2), expected increase in genetic merit (G_2), total number of selected trees (N), number of families represented in the selected cohort (N_{fam}), number of founders represented via their offspring (N_f), total number of crosses (N_c) for restricted selection (RES), optimum contribution selection (OC), and truncation selection (TRU) for the *H* and DBH traits

	H						DBH					
	C_2 (%)	G_2	N	N_{fam}	N_f	N_c	C_2 (%)	G_2	N	N_{fam}	N_f	N_c
RES	2.1	6.7	50	25	35	50	2.1	6.1	50	25	34	50
RES	2.6	7.8	50	25	35	50	2.6	7.6	50	25	34	50
OC	2.1	7.8	141	74	41	200	2.1	7.9	254	77	41	450
OC	2.6	8.4	105	56	36	150	2.6	9.0	188	44	31	500
TRU	7.3	10.0	50	10	10	50	15.8	11.9	50	6	5	50

Results are based on estimated breeding values achieved from the AD model

individuals from 25 families were always selected. Furthermore, Table 2 shows the total number of founders contributing to the selected cohort of trees, where OC picked out more contributing founders than RES and TRU when $C_2 = 2.1\%$. This result suggests that given the restrictions in coancestry, OC better utilises the genetic variation to give higher genetic improvement in the breeding population. However, when $C_2 = 2.6\%$, the number of founders selected by OC was similar to RES for *H* and even lower for DBH. For TRU, selections were made from very few families and very few founders were allowed to contribute to future generations, which will result in a quick loss of genetic variation in the breeding population.

Effect of different levels of group coancestry on OC

Figures 1 and 2 show the impact of the predefined level of coancestry on the number of selected individuals and their respective number of crosses produced by OC. More individuals were selected and the average EBV of the selected cohort was lower for $C_2 = 2.1\%$ than for $C_2 = 2.6\%$. Due to the high restrictions in the optimisation scheme, OC selected individuals from families having negative average EBV to maximise G_2. Furthermore, the number of matings that maximised G_2 increased for DBH compared to the optimal number of matings for *H* (Fig. 1; Table 2: N_c). The explanation for this might be the low heritability for DBH (Table 1). Table 2 shows that 35% more trees were selected when $C_2 = 2.1\%$ than when $C_2 = 2.6\%$ for both traits, but similar pattern for optimal number of crosses could not be found. Hence, when $C_2 = 2.1\%$, the number of crosses per tree was lower than when $C_2 = 2.6\%$ (Fig. 1). A decrease of N_c, so that only a subset of the suggested cohort of trees was selected, will result in departure from the optimal solution for OC. Figure 3 shows the relationship between obtained G_2 at different levels of C_2 with $N_c = 300$ for *H* and $N_c = 400$ for DBH. At lower levels of C_2 ($1\% < C_2 < 4\%$), G_2 follows the expected log linear relationship. At higher levels of C ($> 4\%$), the output of OC seems to differ from the expected optimal levels of G_2 probably due to noise (Fig. 3) induced by the rounding off procedure when converting obtained mating proportions to crosses. For high values of C_2, G_2 eventually reaches a plateau, which corresponds to the theoretical upper limit of possible obtained level of G_2 in the current pedigree. Additionally, Table 2 and Fig. 3 show that when the output of OC is compared to TRU (at the same level of G_2), C_2 is reduced by 39 and 56% for *H* and DBH, respectively.

Effect of dominance on OC

Generally the rank of the selected trees based on mating proportion achieved by OC differed between the combined additive and dominance (AD) and the pure additive (A) statistical models. The obtained standard deviation of the EBVs was higher for model A as compared to model AD, which results in a greater increase in genetic merit for the pure A model. The achieved G_2 for the A model at $C_2 = 2.6\%$ was 8.8 and 10.3 for H and DBH, respectively, which is higher than the corresponding values for AD (8.4 and 9.0; Table 2). For *H*, 139 trees were selected, which is considerably more than for AD (105; Table 2). For DBH on the other hand, a similar number of selections were

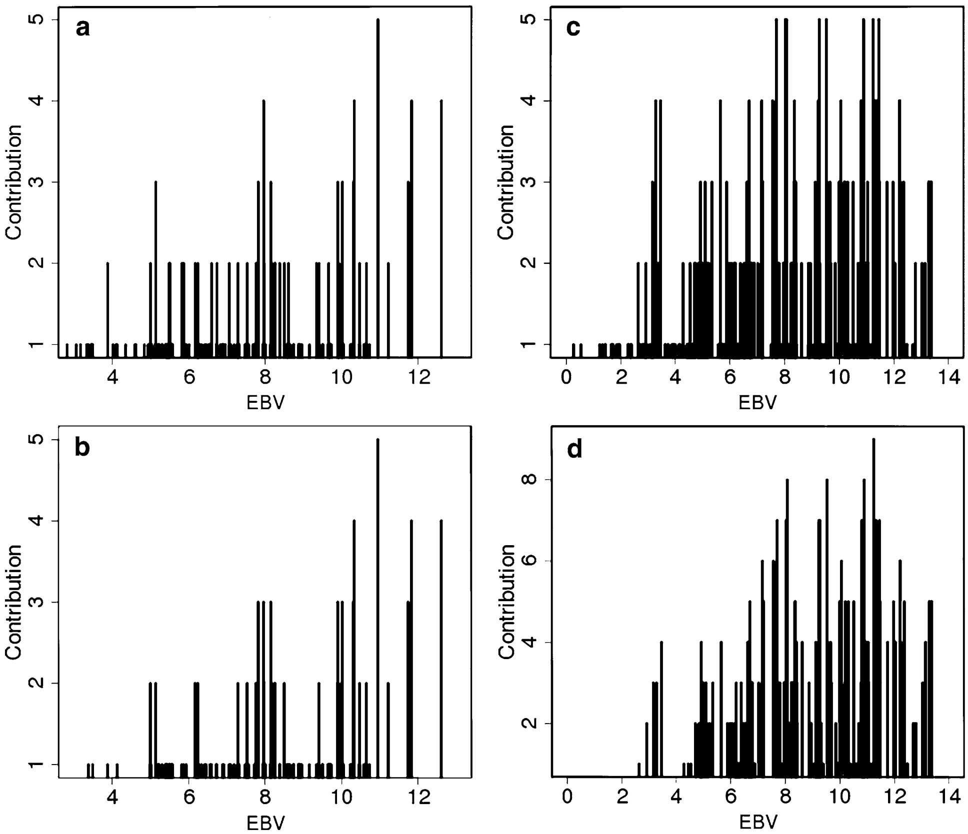

Fig. 1 The obtained number of crosses of the selected trees of the Scots pine data set
a $C_2 = 2.1\%$ for H,
b $C_2 = 2.6\%$ for H,
c $C_2 = 2.1\%$ for DBH,
d $C_2 = 2.6\%$ for DBH

made for the A model (180) compared to the number of selections for AD (188; Table 2). In addition, when ranking selected trees according to their mating proportion for H, 13 of the top ranked 100 selected trees from AD were not included in the top ranked 100 trees from A. The corresponding result for DBH was 7 out of 100 selections. In Table 3, the results of the model comparison between AD and A models using Akaike's information criterion (AIC) is shown. AIC was lower for the AD model than for the A model, suggesting that the AD model provides the best fit to the data given the number of parameters. Waldmann et al. (2008) came to the same conclusion when performing univariate analysis on the same Scots pine data. When the EBVs obtained by the AD model were used as the true breeding values, G_2 became 8.3 and 8.7 for H and DBH, respectively. Hence, the decrease in G_2 (0.1 for H and 0.3 for DBH) compared to the A model EBVs was relatively small.

Discussion

In the current study, we have implemented a powerful optimum contribution (OC) method for selection of breeding individuals and optimisation of their contributions to the next generation under pre-defined levels of group coancestry that can be used in large general tree pedigrees. The OC method outperformed a traditional restricted selection (RES) approach when applied to two different Scots pine traits. The largest improvement for OC was for trunk diameter (DBH) at a group coancestry increase of 2.1%, which resulted in an increased genetic gain of as much as 30% more than that obtained with the RES approach. The OC method can handle large, general pedigrees and uses the genetic relationship between parents and offspring to compute the inverse of the additive relationship matrix between selection candidates. Therefore, there is no need to pre-select individuals (i.e. based on EBV) when entering the OC procedure. The inclusion of a dominance term into the statistical model of the current study will give more accurate EBVs, but the cause of the lower genetic gain of the AD models is probably because of the lower estimate of the additive genetic variance.

Selection methods in tree breeding

Olsson et al. (2000) compared the group-merit selection approach introduced by Lindgren and Mullin (1997) with restricted selection, using the same data as in the current study. The maximum difference between the methods regarding genetic gain occurred at intermediate levels of status number, i.e. when $N_s = 14$ ($C = 3.6\%$), where they

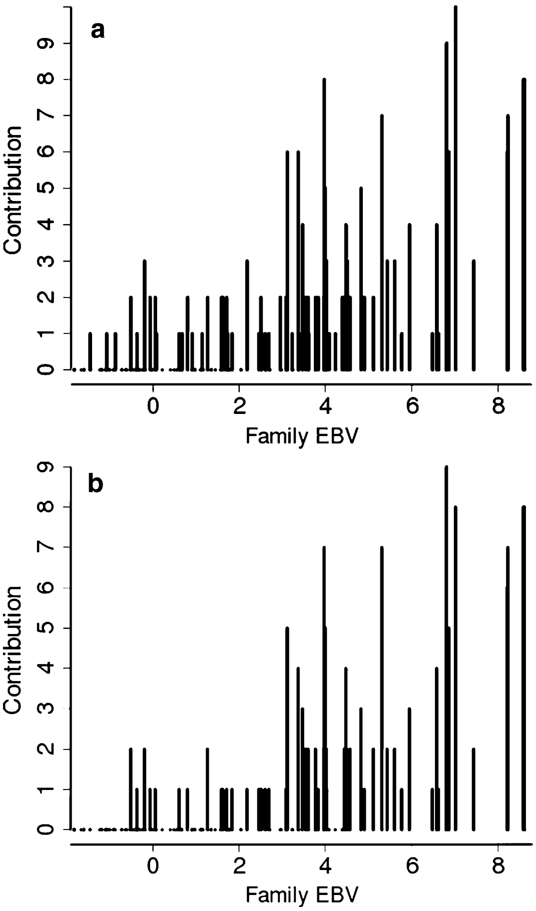

Fig. 2 Number of total matings from each family ranked according to their average EBV attained using OC for *H*, **a** $C_2 = 2.1\%$, **b** $C_2 = 2.6\%$

obtained an increase in genetic merit of 5.2%. However, when N_s was set to 20 ($C = 2.5\%$), they achieved similar genetic gain by both methods. These results differ considerably from the results in the current study of the OC method, where the greatest difference appears at low increase of coancestry (or at low decrease of status number). This conclusion was also arrived at by Meuwissen (1997) and Avedaño et al. (2004).

One potential problem of using group merit selection is to set the weight that adjusts the importance of genetic diversity compared to genetic gain. Finding an optimum weight is difficult because of the need for extensive evaluations that depends on many different parameters, i.e. economic parameters, biological characteristics of the trait under consideration, evaluation of the cost of inbreeding depression etc. Furthermore, the objective function in group merit selection is not quadratic which contrasts to the corresponding objective function in OC (see Eq. 3 in "Methods" section). Grundy et al. (1998) showed that quadratic optimisation also yields the optimal long-term contribution of ancestors, particularly for levels

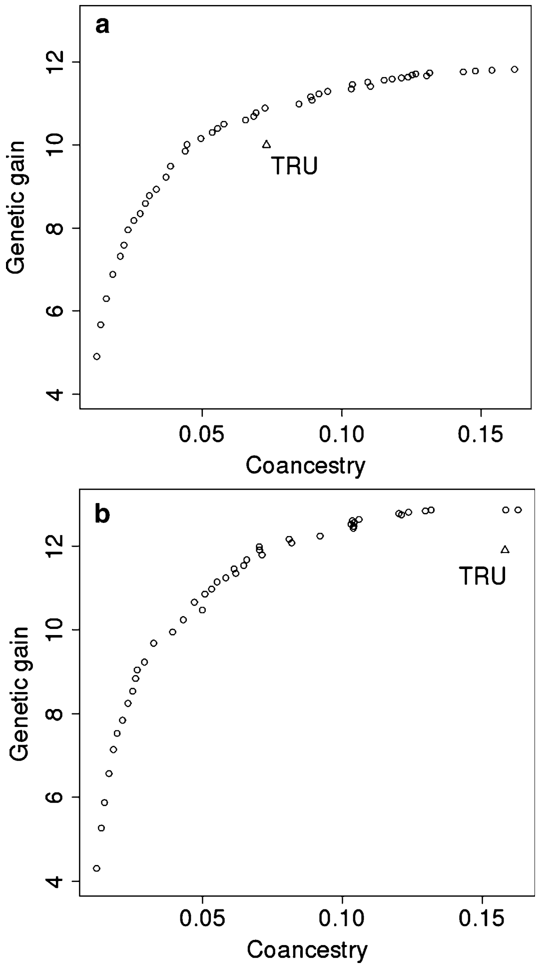

Fig. 3 Maximum obtainable genetic merit given the level of group coancestry obtained both by OC using constraints on group coancestry ranging from 1.2 to 16.2% (*dots*) and by TRU (*triangle*), **a** *H*, $N_c = 300$, **b** DBH, $N_c = 400$

Table 3 Model comparison parameters for A and AD models

Model	Model parameters	Log likelihood	AIC
$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Za} + \mathbf{e}$	143	−1,219.28	2,724.56
$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Za} + \mathbf{Zd} + \mathbf{e}$	146	−1,197.16	2,686.32

AIC Akaike's information criterion

of heritability and constraint on coancestry considered here. Andersson et al. (1999) used optimised selection in an open pollinated field trial of *P. sylvestris,* which resulted in an increased genetic gain between 2.7 and 8.3% as compared to standard within-family selection. The higher genetic gain obtained in the present study (8–30%) compared to standard methodology could partly be explained by the optimisation of the contributions by the OC. Both Olsson et al. (2000) and Andersson et al. (1999) only considered the selected set of trees, not their contributions to following generations. A similar conclusion has been arrived at in the forest tree breeding literature by Ruotsalainen and Lindgren (2001), and Rosvall et al. (2003) based on simulated data. Although none of these studies combined unequal selections and contributions simultaneously. Lstiburek et al. (2005) combined unequal selections and family size (number of offspring) and achieved additional increase in genetic improvement of the breeding population, but they used a static selection procedure (i.e. equal number of selections in each generation) which gives suboptimal results. Using similar methodology to Andersson et al. (1999), Fernandez and Toro (2001) showed that by minimising group coancestry of the selected set of trees in an open pollinated breeding population of *E. globulus*, group coancestry could be reduced by as much as 50% at a loss of 5% of the mean EBV compared to truncation selection. We found that OC selection reduced coancestry to a similar degree when compared with TRU selection.

Simulation studies in animal breeding literature have shown the superiority of the OC method as compared to truncation selection (e.g. Meuwissen 1997; Grundy et al. 1998; Fernandez and Toro 1999). Moreover, when OC had been implemented and tested on various animal breeding pedigrees, increased level of genetic merit was achieved when compared to merit obtained when performing traditional selection methods at the same increase in average inbreeding (e.g. Avendaño et al. 2003; Kearney et al. 2004; Hinrichs et al. 2006). All these results show the importance of optimising selections in breeding programmes, either to increase genetic merit or to decrease group coancestry depending on the goal of the breeding/management programme, particularly if the restriction on keeping genetic variation is strict.

Optimal solutions cannot always be guaranteed by the OC method when using LaGrangian multipliers since mating proportions have to be converted into actual number of matings. Moreover, if fewer matings are required in the breeding programme, less optimal solutions are achieved compared to the case where many matings are possible. A quadratic integer optimisation method would be a better solution than the OC algorithm, so that the rounding off problem could be avoided. Another important advantage of using a quadratic integer programming method would be that inversion of the additive relationship matrix between selection candidates can be avoided, which would speed up the optimisation procedure. Another possibility is the use of stochastic optimisation methods, such as simulated annealing algorithm, which has been successfully implemented and tested by Fernandez and Toro (1999, 2001).

From a practical tree breeding point of view, a large number of controlled pollinated crosses increases financial costs and management risks (e.g. miss-identification of parents when performing crosses). The method suggested here gives the breeder the possibility to decide how many families (and offspring) to progress with. Hence, our approach allows the breeder to use the available resources in an optimal way, for example by considering the land area available for plantation and the number of matings that are possible to perform. In addition, we used the approach of Schneeberger et al. (1992) to make a selection index of height and diameter to create a joint volume measure. However, the higher h^2 in H and the large additive genetic correlation ($r_a = 0.64$) between the traits resulted in that more than 90% of the index was controlled by H for reasonable levels of correlations between the measured traits and the goal trait (per hectare volume production). Consequently, selection based on H alone might be a good predictor of volume production for the levels of h^2 reported here.

Excluding dominance when predicting breeding values using full-sib family pedigree structure

Due to the relatively large pedigree analysed here, even small changes in EBV could affect both which trees that become selected and their respective mating proportions. The accuracy of EBVs can, for instance, be improved by inclusion of a dominance term in the statistical model, which will alter the solution obtained by OC. Waldmann et al. (2008) showed that addition of a dominance term altered the EBV ranking of the 100 best candidates compared to ranking based on a pure additive model using the same pedigree as in the current study. They found that as much as 21 and 13% of the 100 top ranked trees were not included among the 100 best ranked trees using AD compared to A for H and DBH, respectively. In addition, they used a Bayesian model selection criterion to prove that the dominance term provided a better fit to the data, which corresponds well with the results of the multivariate analysis in this study.

Similar population structures are also common in fish breeding populations, where Pante et al. (2002) concluded that heritability and additive variance decreased with as much as 70% when including a dominance component compared to a pure additive model when analysing harvest body weight in different populations of rainbow trout. Additionally, they found that fitting a dominance term increased the log likelihood value when compared to the log likelihood obtained from a pure additive model. The importance of dominant genetic effects has also been reported in sow longevity in the Finnish pig breeding population, where Serenius et al. (2006) showed that including a dominance effect changed estimates of the heritability. The analysed pedigree in their study contained large full-sib families. Wall et al. (2005) concluded that the rank of EBV of fertility traits in dairy cattle changed if a non-additive effect was included in the genetic parameter evaluation, although not as pronounced as in Waldmann et al. (2008).

In our study, we found that the dominance term influenced the OC method and resulted both in different individuals being selected and their respective number of crossings changed compared to a pure A model, particularly for DBH. For AD, 13 and 7% among the 100 top ranked trees, based on mating proportion, were not included among the top 100 trees using A for *H* and DBH, respectively. These values are lower than corresponding values obtained by Waldmann et al. (2008) where trees were ranked based on EBVs. The reason for this difference is partly that OC takes relationships into account when computing mating proportion, compared to the pure truncation of the top 100 ranked trees (Waldmann et al. 2008). Nonetheless, if the ratio of V_d/V_a for the analysed trait is considerable in the population, inclusion of dominance term might be important, particularly for fitness related traits which often exhibits large portion of non-additive variance (Lynch and Walsh 1998).

The higher merit achieved using the pure A model might partly be a result due to the higher standard deviation of the EBVs, which gives a wider distribution compared to EBVs obtained when using AD. This suggests that the highest EBVs might be overestimated using A instead of AD. However, the differences between genetic gain attained from mating proportion for A and AD using EBV obtained from AD model was small. For DBH, the obtained difference in gain between A and AD was higher than for H, which imply that if a pure A model is used, traits sustaining a large level of dominance variance give less increase in genetic merit as otherwise would have been possible. Hence, inclusion of dominance when performing genetic evaluations might be important; otherwise the increase in genetic merit might be overestimated.

Model comparison between the A and AD models was based on Akiakes information criterion (AIC). The obtained value of the likelihood is penalised by twice the number of parameters (*K*) that are to be estimated in the model. For simpler models (i.e. only one random effect), it is straightforward to determine the number of parameters. However, if fixed and multiple random effects are included in the model, one does not in general take correlation between effects into account when computing the penalty term. Vaida and Blanchard (2005) showed how to derive an alternative penalty parameter, referred to as the effective number of parameters. Unfortunately, the effective number of parameters is difficult to calculate and the difference in AIC for A and AD achieved here might be overestimated.

Status effective number and rise of inbreeding in the breeding population

When trees are selected among the available candidates in a closed breeding population, the coancestry will increase. Consequently, if random mating is performed among the selected individuals, group coancestry becomes the expected average inbreeding in the next generation of the breeding population. Even though group coancestry is associated with inbreeding, the mating design determines the level of inbreeding in the upcoming generation. Hence, group coancestry can be seen as the potential rise in inbreeding, not a measure of current inbreeding. One alternative to group coancestry is status effective number (Lindgren et al. 1996), which describes the census number for unrelated and non-inbred trees in the breeding population. The OC algorithm can easily restrict the status number of the population instead of group coancestry. Kerr et al. (1998) showed by simulation that using status number as a restriction to OC, the degree of group coancestry change differed compared to using group coancestry as a restriction. Hence, the breeder can choose what measure he or she wants to use: group coancestry or status number. However, further extensive simulation studies of the respective merits of group coancestry and status number in the OC framework are needed.

Acknowledgments Financial support was provided by the Research School in Forest Genetics and Breeding at the Swedish University of Agricultural Sciences (SLU). The Scots pine data was provided by the The Forestry Research Institute of Sweden, Skogforsk. We would like to acknowledge the associate editor and three anonymous reviewers for comments that improved the paper. In addition, we would also like to thank Chunkao Wang, Johan Kroon, Mats Berlin and Bengt Andersson for helpful discussions.

References

Andersson EW, Sánchez L, Andersson B (1999) Group coancestry-controlled selection in a *Pinus sylvestris* L. breeding program. Theor Appl Genet 99:73–80

Avendaño S, Villanueva B, Woolliams JA (2003) Expected increases in genetic merit from using optimized contributions in two livestock populations of beef cattle and sheep. J Anim Sci 81:2964–2975

Avendaño S, Woolliams JA, Villanueva B (2004) Mendelian sampling terms as a selective advantage in optimum breeding schemes with restriction on the rate of inbreeding. Genet Res 83:55–64

Fernandez J, Toro MA (1999) The use of mathematical programming to control inbreeding in selection schemes. J Anim Breed Genet 116:447–466

Fernandez J, Toro MA (2001) Controlling genetic variability by mathematical programming in a selection scheme on an open-pollinated population in *Eucalyptus globulus*. Theor Appl Genet 102:1056–1064

Gianola D (2001) Inferences about breeding values. In: Balding DJ, Bishop M, Cannings C (eds) Handbook of statistical genetics. Wiley, Buffins Lane, pp 645–672

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2006) ASReml User Guide Release 2.0 VSN International Ltd. Hemel Hempstead

Grundy B, Villanueva B, Woolliams JA (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genet Res 72:159–168

Hannrup B, Jansson G, Danell O (2007) Comparing gain and optimum test size from progeny testing and phenotypic selection in *Pinus sylvestris*. Can J For Res 37:1227–1235

Henderson CR (1985) Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. J Anim Sci 60:111–117

Henderson CR, Quaas RL (1976) Multiple trait evaluation using relatives' records. J Anim Sci 43:1188–1197

Hinrichs D, Wetten M, Meuwissen THE (2006) An algorithm to compute optimal genetic contributions in selection programs with large numbers of candidates. J Anim Sci 84:3212–3218

Kearney JF, Wall E, Villanueva B, Coffey MP (2004) Inbreeding trends and application of optimized selection in the UK Holstein population. J Dairy Sci 87:3503–3509

Kerr RJ, Goddard ME, Jarvis SF (1998) Maximising genetic response in tree breeding with constraints on group coancestry. Silvae Genet 47:165–173

Lindgren D, Mullin TJ (1997) Balancing gain and relatedness in selection. Silvae Genet 46:124–129

Lindgren D, Gea LD, Jefferson PA (1996) Loss of genetic diversity monitored by status number. Silvae Genet 45:52–59

Lstiburek M, Mullin TJ, Mackay TFC, Huber D, Li B (2005) Positive assortative mating with family size as a function of predicted parental breeding values. Genetics 171:1311–1320

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Assoc, Sunderland

Meuwissen THE (1997) Maximizing the response of selection with a predefined rate of inbreeding. J Anim Sci 75:934–940

Misztal I (1997) Estimation of variance components with large-scale dominance models. J Dairy Sci 80:965–974

Olsson T, Lindgren D, Ericsson T (2000) A comparison of group merit selection and restricted selection among full-sib progenies of Scots pine. For Genet 7:137–144

Pante MJR, Gjerde B, McMillan I, Misztal I (2002) Estimation of additive and dominance genetic variances for body weight at harvest in rainbow trout, *Oncorhynchus mykiss*. Aquaculture 204:383–392

Robertson A (1961) Inbreeding in artificial selection programs. Genet Res 2:189–194

Rosvall O, Mullin TJ, Lindgren D (2003) Controlling parent contributions during positive assortative mating and selection increases gain in long-term forest tree breeding. For Genet 10:35–53

Ruotsalainen S, Lindgren D (2001) Number of founders for a breeding population using variable parental contribution. For Genet 8:57–67

Sánchez L (2000) Balanced verus slightly unbalanced selection. Heredity 84:685–691

Schneeberger M, Barwick SA, Crow GH, Hammond K (1992) Economic indices using breeding values predicted by BLUP. J Anim Breed Genet 109:180–187

Serenius T, Stalder KJ, Pounti M (2006) Impact of dominance effects on sow longevity. J Anim Breed Genet 123:355–361

Stoehr M, Yanchuk A, Xie C-Y, Sánchez L (2007) Gain and diversity in advanced generation coastal Douglas-fir selections for seed production populations. Tree Genet Genomes 4:193–200

Toro M, Pérez-Enciso M (1990) Optimization of selection response under restricted inbreeding. Genet Sel Evol 22:93–107

Vaida F, Blanchard S (2005) Conditional Akaike information for mixed-effects models. Biometrika 92:351–370

Waldmann P, Ericsson T (2006) Comparison of REML and Gibbs sampling estimates of multi-trait breeding values in Scots pine. Theor Appl Genet 112:1441–1451

Waldmann P, Hallander J, Hoti F, Sillanpää MJ (2008) Efficient MCMC implementation of Bayesian analysis of additive and dominance genetic variances in non-inbred pedigrees. Genetics 179:1101–1112

Wall E, Brotherstone S, Kearney JF, Woolliams JA, Coffey MP (2005) Impact of nonadditive genetic effects in the estimation of breeding values for fertility and correlated traits. J Dairy Sci 88:376–385

Wei RP, Lindgren D (2006) Stepwise penalty index selection from populations with a hierarchical structure. Silvae Genet 55:62–70

Wilhelmsson L, Andersson B (1993) Breeding programs in Sweden-Scots pine and Lodgepole pine. 1 Progeny testing and breeding strategies. In: Lee SJ (ed) Proceedings of the Nordic group for tree breeding. Edinburgh, Scotland, pp 135–145

Woolliams JA (2007) Genetic contributions and inbreeding. In: Oldenbroek K (ed) Utilisation and conservation of farm animal genetic resources. Wageningen Academic Publisher, Netherlands, pp 147–165

Zheng YQ, Lindgren D, Rosvall O, Westin J (1997) Combining genetic gain and diversity by considering average coancestry in clonal selection of Norway spruce. Theor Appl Genet 95:1312–1319