

Using semidefinite programming to optimize unequal deployment of genotypes to a clonal seed orchard

J. Ahlinder · T. J. Mullin · M. Yamashita

Received: 31 May 2013 / Revised: 6 September 2013 / Accepted: 10 September 2013 / Published online: 28 September 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract Tree breeders must often consider the conservation of genetic diversity, while at the same time, maximizing response to selection. In the case of seed orchards, the buyer of seed wants maximum performance, while satisfying a restriction, sometimes legislated, on the diversity deployed to the forest. Optimal selection will not completely avoid kinship but rather maximize gain while imposing a constraint on average relatedness. Here, we present the application of semidefinite programming (SDP) as a flexible approach to optimize the deployment of genotypes to a clonal seed orchard. We formulate the selection problem as an SDP, where average breeding value is to be maximized, while imposing constraints on relatedness, as well as maximum and minimum contributions from each candidate. An open-source solver, SDPA, was embedded into a tool designed to make the optimization of seed orchards by SDP simple and flexible. Case studies optimizing seed orchards for Scots pine and loblolly pine illustrate how this flexibility can be used to impose

additional constraints on the scion material available from some candidate genotypes and optimize selection even when related candidates have varying degrees of coancestry among them. Additional situations where SDP can be employed are discussed.

Keywords Semidefinite programming · Optimization · Genetic gain · Relatedness · Clonal seed orchards · Optimal contributions · Selection · Orchard management · *Pinus sylvestris* · *Pinus taeda*

Introduction

It is common in tree breeding that selection of populations must consider conservation of genetic diversity, while at the same time attempting to maximize response to selection. It is generally recognized that one cannot simply select the “best” trees, without also taking into account the degree of relatedness among them. To ignore relatedness would result in a rapid loss of diversity in breeding populations, reducing the long-term gains possible through recurrent selection and accumulating high levels of inbreeding (Robertson 1960, 1961). Managing relatedness among selections becomes complicated as early as the first cycle of breeding, when parents, siblings, and other close relatives have similar ranks. Historically, tree breeders have often imposed rules of thumb on how many selections can be accepted from a family or other group of relatives (e.g., Jarvis et al. 1995), as well as subdivision or “sublining” of breeding populations to avoid relatedness among sublines, even as it accumulates rapidly within (e.g., McKeand and Bridgwater 1998). Such rules may be successful in regulating the accumulation of relatedness but are unlikely to result in optimal solutions that maximize gain.

The issue of how to truly optimize the balance between gain and relatedness can be approached in several ways. An

Communicated by D. Neale

J. Ahlinder
Swedish Defence Research Agency, Division of CBRN Protection,
Cementvägen Umeå 906 21, Sweden

T. J. Mullin (✉)
BioSylve, PO Box 30709, Nairobi 00100, Kenya
e-mail: tim.mullin@biosylve.com

T. J. Mullin
The Swedish Forestry Research Institute (Skogforsk), Box 3,
Sävar 918 21, Sweden
e-mail: tim.mullin@skogforsk.se

M. Yamashita
Department of Mathematical and Computing Sciences, Tokyo
Institute of Technology, 2-12-1-W8-29 Ookayama, Meguro-ku,
Tokyo 152-8552, Japan

optimal solution will not completely avoid kinship but rather find the set of selections that maximizes gain while imposing an overall constraint on average relatedness. In the context of tree breeding, the problem was formulated by Lindgren and Mullin (1997) who expressed “group merit” of a selected population as a function of average genetic value, penalized by a weight on relatedness among individuals, as expressed by their “group coancestry” *sensu* Cockerham (1967). While they provided a way to maximize group merit for a fixed number of selections with equal representation (“Group-Merit Selection”), the constraint on relatedness is applied indirectly so that achieving a specified level when selecting breeding populations requires a trial-and-error approach, typically with many iterations, and optimizing unequal numbers of ramets deployed to seed orchards is not possible.

The special case of optimizing deployment of genotypes to a seed orchard was explored by Lindgren and Matheson (1986) and Lindgren et al. (1989) who suggested that the optimum relationship between candidate breeding value and their contribution to the population would be linear. The application of “linear deployment” is only possible when the candidates are unrelated as might be the case when selecting “backward” on progeny tested plus tree candidates. If the candidate pool includes relatives, Bondesson and Lindgren (1993) suggested that a more complex formulation would be required, perhaps using a LaGrange function. In more recent work, Lindgren's group used the Microsoft Excel add-in tool, “Solver”, to maximize gain by linear programming under a LaGrange equality constraint on relatedness (Danusevičius and Lindgren 2008; Lindgren et al. 2009).

The quadratic object function was introduced by Meuwissen (1997) as the basis for seeking an optimal balance between genetic merit and relatedness by simultaneous selection of parents and calculation of their respective mating proportions. Meuwissen's algorithm is based on LaGrangian multipliers (LM) and has been used in both theoretical and practical applications to optimize breeding programs, mainly in an animal-breeding framework (e.g., Grundy et al. 1998; Avendaño et al. 2004; Hinrichs et al. 2006; Villanueva et al. 2006; Woolliams 2007; Hinrichs and Meuwissen 2011). The LM method was recommended for application to forest trees by Kerr et al. (1998) and subsequently demonstrated in the management of breeding (Hallander and Waldmann 2009a) and in the optimal selection of Scots pine (*Pinus sylvestris* L.) parents for a seed orchard with unequal numbers of ramets (Hallander and Waldmann 2009b). A general assumption made by LM is that the optimal solution should occur when the restriction on relatedness is exactly achieved, i.e., the optimal solution is found at the boundary of all possible solutions. After a primary solution is obtained, some candidates will obtain negative contributions as there is no restriction on the minimum (or maximum) contribution of the selection candidates. By removing negatively contributing

candidates from the selection process, either all simultaneously or one-by-one, and resolving the optimization, a final set of candidates and their respective contributions is obtained.

There are, as pointed out by Pong-Wong and Woolliams (2007), some serious drawbacks with the LM method. First, by removing candidates or fixing their contributions to zero and re-optimizing with a new subset of candidates, it is possible that the true optimum solution is bypassed in the iterative procedure. Second, there is no restriction on the maximum allowed contribution of any particular candidate. This means that ad hoc manipulations of the final solution may be required to satisfy other operational constraints on its implementation. For example, in forest tree breeding, one major constraint for the establishment of grafted seed orchards is the number of scions that can be collected from a given genotype.

In this paper, we examine the application of semidefinite programming (SDP) to optimize selection for gain with a quadratic constraint on relatedness as applied to typical situations in establishment and management of forest tree seed orchards. SDP was introduced by Pong-Wong and Woolliams (2007) as an alternative method for finding a solution to a convex optimization problem in an animal-breeding framework. They reported that in several examples, SDP found a more optimal solution than did the LM approach. To our knowledge, theirs is the only published study with a focus on breeding that has utilized SDP to obtain optimal contributions although their comparisons of the SDP and LM approaches were limited to small “toy” examples. Here, we apply SDP to two illustrative case studies: (1) a real Scots pine (*P. sylvestris* L.) pedigree and associated breeding value estimates to illustrate the performance and flexibility of the method, while imposing additional operational constraints of interest to orchard managers and (2) selection for an elite orchard among clonally replicated loblolly pine (*P. taeda* L.) in a publically available pedigree representing varying degrees of relatedness among candidates.

Theoretical development and methods

Semidefinite programming

Semidefinite programming is an optimization method to minimize a linear objective function subject to the constraint that a linear combination of symmetric matrices is positively semidefinite (Vandenberghe and Boyd 1996). The constraint is convex, which means that if two points satisfy the constraint, then any point on the segment between the two points also satisfies the constraint. It is the convex nature of constraints that allows SDPs to be solved efficiently. The objective to find a minimum for a linear function over a linear matrix inequality can be described by the general form:

$$\text{Minimize : } \sum_{i=1}^Z f_i x_i \quad \text{Subject to : } \mathbf{Y}_0 - \sum_{i=1}^Z \mathbf{Y}_i x_i \geq \mathbf{0}$$

The input data are $f_1, f_2, \dots, f_Z \in R$ and $\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_Z$ being matrices of the same dimension, while the decision variable is $\mathbf{x} \in R^Z$. The notation $\mathbf{Y} \geq \mathbf{0}$ indicates that matrix \mathbf{Y} is positive semidefinite, i.e., all the eigenvalues of \mathbf{Y} are non-negative. SDP can be considered as an extension of linear programming to the space of matrices.

The main advantage of SDP is that optimization theory guarantees the optimum solution which is found in an efficient and smooth way (Vandenberghe and Boyd 1996). SDP can be efficiently solved using interior-point methods, which are well understood and perform well in practice (Nesterov and Todd 1997; Kojima et al. 1997; Alizadeh et al. 1998). In addition, interior-point methods have the ability to exploit the structure of the optimization problem, such as the usage of sparse equation solvers. Many computer programs based on interior-point methods are available for solving SDPs, including semidefinite programming algorithm (SDPA) (Yamashita et al. 2003, 2010, 2012).

Object function formulation

To formulate the selection of genotypes with unequal numbers of ramets to a grafted seed orchard as an SDP, we first consider selecting a cohort from a complex pedigree, containing totally Z genotypes that are to contribute their genes in optimal proportions. The object is to maximize the expected genetic merit of contributions from the selected cohort, given by $\mathbf{c}^T \mathbf{g}$, where the estimated breeding values (EBV) for all pedigree members are found in vector \mathbf{g} , of size $Z \times 1$, and the contribution of genes as a proportion is denoted \mathbf{c} , also of size $Z \times 1$. In our problem, the decision variable is \mathbf{c} , where $1 \geq c_i \geq 0$, and the sum of all contributions from the selected cohort equals unity ($\sum_{i=1}^Z c_i = 1$).

There are a number of constraints required to fully formulate the selection problem. First, we wish to impose a quadratic restriction on the relatedness or group coancestry, θ , of the selected cohort, specified as $\theta \geq \mathbf{c}^T \mathbf{A} \mathbf{c} / 2$, where the additive or numerator relationship matrix of the pedigree is denoted \mathbf{A} , of size $Z \times Z$. The maximum and minimum contributions that a particular individual can make are denoted \mathbf{m} and \mathbf{u} , respectively, and both vectors are of size $Z \times 1$. Here, if pedigree member i is not itself a candidate for selection (e.g., not physically available for use), the corresponding maximum contribution is set to zero (i.e., $m_i = 0$). Similarly, while the minimum number of contributions for a genotype might normally be zero, there may be times when prior investments might motivate setting u_i to a specific value greater than zero, provided that $m_i \geq u_i$. In monoecious tree species, where individuals can have reproductive structures for both genders,

there are no limits imposed by gender and the total sum of all contributions in the pedigree should equal unity. For dioecious tree species, restrictions on the contribution of the separate genders are needed: $\mathbf{c}^T \mathbf{d} = 0.5$ and $\mathbf{c}^T \mathbf{s} = 0.5$ for female and male contributions, respectively, where: \mathbf{d} and \mathbf{s} are indicator vectors of size $Z \times 1$, $d_i = 1$ and $s_i = 0$ if tree i is a female and vice versa if i is a male.

In order to find the optimal gene contributions, the problem can be formulated as:

$$\text{Maximize : } \mathbf{c}^T \mathbf{g} \tag{1a}$$

$$\text{Subject to : } \mathbf{c}^T \mathbf{A} \mathbf{c} / 2 \leq \theta \tag{1b}$$

$$\mathbf{c}^T \mathbf{1} = 1 \tag{1c}$$

$$\mathbf{c} \geq \mathbf{u} \tag{1d}$$

$$\mathbf{c} \leq \mathbf{m} \tag{1e}$$

where $\mathbf{1}$ is a vector of size $Z \times 1$ containing 1s. This formulation of the optimization problem is flexible; for example (1c) could easily be replaced with $\mathbf{c}^T \mathbf{s} = 0.5$ and $\mathbf{c}^T \mathbf{d} = 0.5$, if required to account separately for gender.

When reformulating as an SDP problem, the quadratic constraint (1b) is expressed in linear form using its *Shur complement* and the equality constraint (1c) replaced by two inequality constraints so that the selection problem for a monoecious species is reformulated:

$$\text{Minimize : } -\mathbf{c}^T \mathbf{g} \tag{2a}$$

$$\text{Subject to : } \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{c} \\ \mathbf{c}^T & 2\theta \end{bmatrix} \geq \mathbf{0} \tag{2b}$$

$$\mathbf{c}^T \mathbf{1} - 1 \geq 0 \tag{2c}$$

$$-\mathbf{c}^T \mathbf{1} + 1 \geq 0 \tag{2d}$$

$$\mathbf{c} - \mathbf{u} \geq \mathbf{0} \tag{2e}$$

$$\mathbf{m} - \mathbf{c} \geq \mathbf{0} \tag{2f}$$

We can now formulate the results in $\mathbf{Y} = \mathbf{Y}_0 - \sum_{i=1}^Z \mathbf{Y}_i c_i$:

$$\mathbf{Y} = \begin{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{c} \\ \mathbf{c}^T & 2\theta \end{bmatrix} & & & & \\ & \mathbf{c}^T \mathbf{1} - 1 & & & \\ & & -\mathbf{c}^T \mathbf{1} + 1 & & \\ & & & \text{diag}(\mathbf{c} - \mathbf{u}) & \\ & & & & \text{diag}(\mathbf{m} - \mathbf{c}) \end{bmatrix} \geq \mathbf{0} \tag{3a}$$

with the $Z+1$ set of affine matrices being:

$$\mathbf{Y}_0 = \begin{bmatrix} \begin{bmatrix} -\mathbf{A}^{-1} & \mathbf{0}_{(Z \times 1)} \\ \mathbf{0}_{(1 \times Z)} & -2\theta \end{bmatrix} & & & \\ & 1 & & \\ & & -1 & \\ & & & [\mathbf{diag}(\mathbf{u})] \\ & & & & [-\mathbf{diag}(\mathbf{m})] \end{bmatrix} \geq \mathbf{0} \quad (3b)$$

and

$$\mathbf{Y}_i = \begin{bmatrix} \begin{bmatrix} \mathbf{0}_{(Z \times Z)} & \mathbf{I}_i \\ \mathbf{I}_i^T & 0 \end{bmatrix} & & & \\ & 1 & & \\ & & -1 & \\ & & & [\mathbf{diag}(\mathbf{I}_i)] \\ & & & & [-\mathbf{diag}(\mathbf{I}_i)] \end{bmatrix} \geq \mathbf{0}, i = 1, 2, \dots, Z \quad (3c)$$

where the size of the first block is $(Z+1) \times (Z+1)$, the next two are 1×1 , and the last two blocks are of size $Z \times Z$. \mathbf{I}_i is the i column of the identity matrix of size $Z \times Z$, and $\mathbf{diag}(\mathbf{I}_i)$ is a diagonal matrix with diagonal equal to \mathbf{I}_i . All other matrix elements are zero.

Solving the SDP

SDPA¹ is an open-source solver that can be applied to many types of SDPs. SDPA is very flexible, but with flexibility comes complexity and room for error, both in setting up the SDP properly and interpreting the output. We have simplified this task by embedding the SDPA solver in a user-friendly open-source tool, known as OPSEL² (Mullin et al. 2013), which is designed specifically for such selection optimization applications in tree breeding. OPSEL receives input regarding the total number of ramets to be established in the orchard, the constraint on relatedness (as group coancestry or Status Number), whether a minimum is to be imposed on individual genotype contributions (i.e., $\mathbf{u} \neq 0$), and the name of a text file containing the EBVs of all candidate genotypes, their maximum (and minimum) frequency in the selected group, as well as the complete pedigree including ancestors. These data are used to prepare the SDP for solving by SDPA. Once SDPA has completed its work, OPSEL then reads the SDPA output and generates a file with the original data, as submitted in the pedigree file, and with additional columns specifying the optimum contribution as a proportion and as an integer number of ramets for each genotype.

Case study 1: establishing a Scots pine orchard in northern Sweden

To illustrate the application of SDP to optimize a seed orchard, we use the actual pedigree and breeding value data from the

Scots pine breeding program in northern Sweden. The program has access to many plus tree founders (F_0 generation) that have been progeny tested by open pollination and/or polycross and to the progeny of many pair crosses between the F_0 parents that have been established in F_1 family field tests. Comparable BLUP EBVs for the target orchard deployment region were available from the TREEPLAN[®] system (McRae et al. 2004), using all available field test data. The client's specification for the orchard is that it should contain a total of $N=2,800$ ramets, having a status number $N_s \geq 14$ (*sensu* Lindgren et al. 1996 and equivalent to group coancestry $\theta \leq 0.03571$).

The first step was to perform a truncation of the candidate list to a total of 2,000 F_1 genotypes, including not more than 15 per full-sib family, and their F_0 parents. The complete pedigree list contained 2,045 genotypes. We then satisfied the orchard specification using three selection approaches:

1. *Forwards selection of the best 14 unrelated F_1 candidates.* By establishing 200 ramets for each of the best 14 unrelated F_1 genotypes, we can establish a 2,800-tree orchard that satisfies the genetic diversity specification $N_s=14$. Performing the selection is as simple as preparing a ranked list of the single best candidate from each F_1 family and selecting from the top down, such that each candidate added to the selected group is unrelated to all previous selections.
2. *Optimum selection with unequal contributions and no maximum.* Here, we apply selection across all F_0 and F_1 candidates, while constraining on group coancestry $\theta \leq 0.03571$ but with no maximum contribution from any given candidate. This optimization can be solved by Meuwissen's LM algorithm or by OPSEL based on an SDP solved by SDPA.
3. *Optimum selection with constraint on maximum contribution from F_1 genotypes.* Whereas the F_0 candidates are large mature trees that can contribute very large quantities of scions for grafting, the younger F_1 candidates are much smaller. Practically speaking, the F_1 trees are not likely capable of contributing more than 50 scions each. OPSEL can include this additional constraint in the SDP.

Case study 2: optimizing selection in a clonally replicated test of loblolly pine

Resende et al. (2012) recently made a standard set of data available online³ from a clonally replicated population of loblolly pine in the southeastern USA. The population was derived by controlled crossing among 32 selected parents from the Lower Gulf Elite Population, consisting of 22 field-selected F_0 plus trees and 10 selected F_1 progeny. These

¹ The current version of SDPA software for various platforms is maintained at <http://sdpa.sourceforge.net/>.

² OPSEL is open-source software. The current version can be requested from tim.mullin@skogforsk.se.

³ Pedigree data and estimated breeding values for this standard loblolly pine dataset are available in *File S2* and *File S3*, respectively, within the supporting information for the paper by Resende et al. (2012) and found at <http://www.genetics.org/content/190/4/1503/suppl/DC1>.

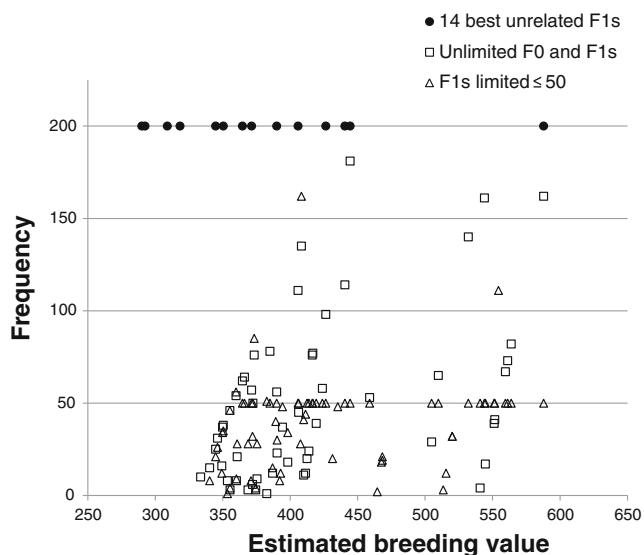


Fig. 1 Distribution of numbers of ramets established versus estimated breeding value for genotypes selected by each method for case study 1 on Scots pine

parents were crossed in a partial diallel mating design, and the progeny propagated for field testing as rooted cuttings (Baltunis et al. 2007a, b). For this case study, we developed the SDP to optimize the selection of a grafted elite seed orchard with a total of $N=2,000$ ramets, based on the published 6-year height EBVs for 861 of the candidate genotypes. The candidates together had a status number $N_s=23.3$, reflecting the considerable relatedness among the clones. Contributions to the seed orchard were constrained to status number $N_s \geq 10$ (group coancestry $\theta \leq 0.05$). There were no other limitations placed on the contributions from any given genotype. The selection optimized by means of an SDP is compared with selecting the ten top-ranked, unrelated genotypes.

Results and discussion

The Scots pine case study

The case study illustrates several of the computational and operational issues faced when attempting to optimize a real orchard (Fig. 1 and Table 1). The simplest approach, (1) Forwards the selection of best unrelated F_1 candidates, successfully achieves the target status number, but requires that each of the 14 selected genotypes produce 200 successful grafts. Most of the young F_1 selections could not produce that many scions in a single collection, and it would take several years to completely establish the orchard by returning to the ortets in future years or collecting scion material from the earlier grafts. We also see in Fig. 1 the not unusual situation where the very best genotype is far better than average. The two parents of this genotype themselves would be good candidates for selection, but since we apply a strict restriction on inclusion of half-sib or other relatives, we are forced to go well down the candidate list to find the next eligible, unrelated genotype.

In this particular example, optimization by LM or by solving the SDP gives identical results, provided that there is no constraint on the numbers of ramets contributed by any given genotype. Given that there are various degrees of coancestry between candidates, the relationship between contribution and EBV was very weak (Fig. 1), in contrast to the strong linear relationship that would be expected had the candidates been unrelated (Lindgren et al. 1989). The optimum solution in this example, where no constraints were placed on maximum contributions, utilized 56 genotypes, many of them related to each other, but still producing the required status number for the orchard. The average EBV was over 17 % greater than that from selecting 14 unrelated candidates (Table 1).

While the improvement in gain is impressive, we are still left with many F_1 genotypes having to contribute very large numbers of scions that are simply not available on these

Table 1 Comparison of selection methods for case study 1

	Top 14 unrelated F_1 s used equally	Optimized with unlimited F_0 s and F_1 s	Optimized with unlimited F_0 s and F_1 s ≤ 50
Number of ramets	2,800	2,801	2,796
Number of genotypes	14	56	71
Status number	14	14	14
Group coancestry	0.0357	0.0357	0.0357
Average EBV	381	447	439
Difference in EBV	–	+17.3 %	+13.0 %

Genotype deployment options for a grafted Scots pine seed orchard in northern Sweden with a total $N=2,800$ ramets, constrained to a status number $N_s \geq 14$ (equivalent to a group coancestry $\theta \leq 0.03571$), including the average estimated breeding value (EBV) of the orchard and the difference in EBV compared with using the top 14 unrelated genotypes in equal numbers

smaller trees. The ability to constrain on maximum numbers of ramets is available through SDP, and the final approach applied a constraint of 50 ramets from F_1 candidates, while no restriction was imposed on the numbers of ramets from F_0 candidates.

The loblolly pine case study

The second case study illustrates the use of SDP to optimize selection when there is more relatedness among the candidates. Having a status number of $N_s=23.3$, the relationships among the tested clones vary considerably, with coancestry between related clones from 0.0313 to 0.25. Selection of the top ten unrelated genotypes, each to be deployed in equal numbers, requires that we go well down the candidate list, and the orchard-wide average EBV for 6-year height is 89.0 (Table 2). Solving the SDP gives an optimum solution with 30 genotypes, each contributing from 1 to 289 ramets, distributed as shown in Fig. 2, and producing an average EBV of 114.0, over 28 % higher than deploying the top ten unrelated clones.

While the number of candidates is relatively small, this example illustrates the difficulty of avoiding relatedness after only 2 or 3 cycles of breeding. Solving the problem with an SDP maximizes the genetic value of the orchard, while satisfying the constraint on relatedness.

Resource requirements

A comparison of computing efficiency between the LM algorithm and OPSEL's solution by SDPA is not really possible as there exists no public access to LM software that is truly optimized. It can be noted that the time on a typical office computer to optimize selection by SDPA for the case study examples presented here is a matter of minutes. The solution

for a longer candidate list of 12,000 genotypes plus ancestors used all available memory in a 16-Gb machine under Windows 7 but was completed in just over 5 h. Practically speaking, breeders would want to truncate their candidate list to avoid exceedingly long execution times.

When to optimize or re-optimize?

There are several points in the process of establishing and managing a seed orchard when a manager might wish to optimize:

1. Planning the initial makeup of an orchard. This is an obvious time when one would want to prepare a list of genotypes and contributions to plan the establishment of N grafted plants in an orchard.
2. After collection of scions, to prepare the nursery's grafting list. Typically, the scion collection operation will require a crew to visit the ortets in the field or to make collections from ramets established in clone banks or other orchards. Some donor plants will produce more than enough scion material, whereas others may fall well short of the number prescribed by the initial optimization. Some donors may be dead or too remote to allow access during the collection period. Furthermore, the grafting operation would normally prepare excess numbers of rootstock, and the grafting list should be optimized for this larger number of plants. Once the scion collection has been completed, the optimization can be rerun, using the actual numbers of scions available as the new constraint on maximum contribution and with N set to the total number of rootstock to be grafted in the nursery.
3. Before shipping surviving grafts to the orchard site. The nursery inventory of surviving grafts supplies the

Table 2 Comparison of selection methods for case study 2

	Top 10 unrelated genotypes used equally	Optimized with selection across entire pedigree
Number of ramets	2,000	1,999
Number of genotypes	10	30
Status number	10	10
Group coancestry	0.05	0.05
Average EBV	89.0	114.0
Difference in EBV	–	+28.1 %

Genotype deployment options for a grafted loblolly pine seed orchard using candidates from the CCLONES test population, having a total $N=2,000$ ramets, constrained to a status number $N_s \geq 10$ (equivalent to a group coancestry $\theta \leq 0.05$), including the average estimated breeding value (EBV) for 6-year height of the orchard and the difference in EBV compared with using the top ten unrelated genotypes in equal numbers

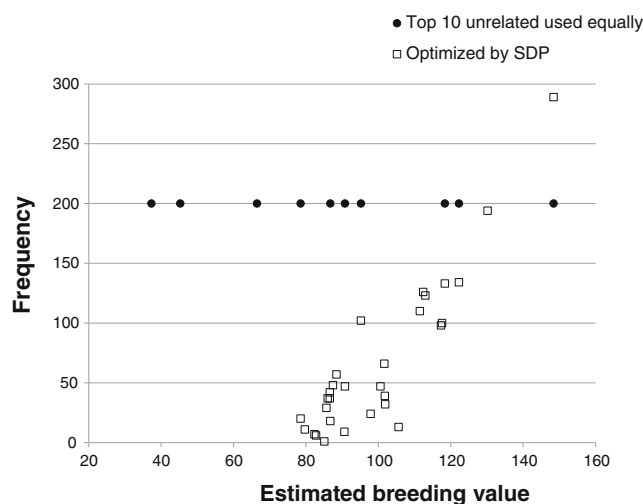


Fig. 2 Distribution of numbers of ramets versus estimated breeding value for case study 2, comparing simple selection of the top ten unrelated genotypes deployed equally with selection by SDP to optimize contributions to a 2,000-ramet loblolly pine seed orchard

information for the maximum possible contribution of each genotype as the grafts are shipped for establishment in the orchard.

4. Thinning an orchard. Most orchards will require thinning at some point in their development. Updated EBVs will likely be available, while the current inventory of surviving ramets provides the constraint on the maximum numbers of ramets to be left. The desired census number after thinning provides the total number of contributions, N .
5. Adding material to an existing orchard. It is not uncommon for the establishment of an orchard to begin with only a portion of the total number of trees, leaving gaps for future planting. To optimize the filling in of this orchard, the existing inventory of established grafts is used as the constraint on minimum contributions, while the maximum for these genotypes must be at least as large as the minimum. N is declared as the total size of the orchard, including the preexisting material. The same approach can be used to optimize the replacement of mortality.
6. Optimizing an orchard when “standard” genotypes are included. There are situations when an orchard manager will want to ensure that certain genotypes are included in the orchard. These “standards” may represent genotypes whose performance or response to orchard management is well known, or there may be a known market for their seeds. Whatever is the reason for their inclusion, a constraint on the minimum contribution from these genotypes can be declared when the orchard is optimized.

SDP for optimizing seed mixtures from orchards

Of course, the optimizing of seed orchard establishment around a constraint on relatedness assumes that each ramet of the various candidate genotypes will contribute the same number of gametes to the seed produced in the orchard. This may be a reasonable assumption for many open-pollinated orchard species but will certainly not be the case when fertility varies greatly among genotypes or when seed is produced by controlled crossing. SDP can still be useful in such situations by optimizing the mixing of seedlots collected in the orchard, providing high genetic value while satisfying a constraint on status number.

Acknowledgments We are grateful to Ola Rosvall (formerly Skogforsk, Sävar, Sweden) for his strong support in pursuing the optimization of selection in tree breeding. We also acknowledge the important contributions of Torgny Persson (Skogforsk, Sävar, Sweden) who supplied pedigree and breeding value data for the Scots pine case study, Ricardo Pong-Wong (The Roslin Institute, Edinburgh, UK) who clarified some issues in preparing of the SDP for input into SDPA, and both Greg Dutkowski and Richard Kerr (PlantPlan Genetics Pty Ltd., Hobart, Australia) for their valuable feedback on an early version of OPSEL. We are also grateful to Associate Editor Rowland Burdon and two anonymous referees whose comments and suggestions greatly improved the

manuscript. The establishment of a seed orchard in northern Sweden honoring the pioneering work in this area of Professor Dag Lindgren (SLU, Umeå, Sweden) provided the motivation for the Scots pine case study. Funding for this study was provided by Föreningen Skogsträdsförädling (The Swedish Tree Breeding Foundation), the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 211868 (Project Noveltree), and Skogforsk (The Swedish Forestry Research Institute). All trademark terms are the property of their respective owners.

Data archiving statement We followed the standard Tree Genetics and Genomes policy. The pedigree and estimated breeding values used for the case studies are deposited in the Dryad Digital Repository at <http://dx.doi.org/10.5061/dryad.9pn5m>.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Alizadeh F, Haeberly J, Overton M (1998) Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. *SIAM J Optim* 8(3):746–768. doi:10.1137/S1052623496304700
- Avendaño S, Woolliams JA, Villanueva B (2004) Mendelian sampling terms as a selective advantage in optimum breeding schemes with restrictions on the rate of inbreeding. *Genet Res* 83(1):55–64. doi:10.1017/S0016672303006566
- Baltunis B, Huber D, White T, Goldfarb B, Stelzer H (2007a) Genetic gain from selection for rooting ability and early growth in vegetatively propagated clones of loblolly pine. *Tree Genet Genom* 3(3):227–238. doi:10.1007/s11295-006-0058-9
- Baltunis BS, Huber DA, White TL, Goldfarb B, Stelzer HE (2007b) Genetic analysis of early field growth of loblolly pine clones and seedlings from the same full-sib families. *Can J For Res* 37(1):195–205. doi:10.1139/x06-203
- Bondesson L, Lindgren D (1993) Optimal utilization of clones and genetic thinning of seed orchards. *Silvae Genet* 42(4–5):157–163
- Cockerham CC (1967) Group inbreeding and coancestry. *Genetics* 56:89–104
- Danusevičius D, Lindgren D (2008) Strategies for optimal deployment of related clones into seed orchards. *Silvae Genet* 57(3):119–127
- Grundy B, Villanueva B, Woolliams JA (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genet Res* 72(2):159–168
- Hallander J, Waldmann P (2009a) Optimization of selection contribution and mate allocations in monoecious tree breeding populations. *BMC Genetics* 10(1):1–17. doi:10.1186/1471-2156-10-70
- Hallander J, Waldmann P (2009b) Optimum contribution selection in large general tree breeding populations with an application to Scots pine. *Theor Appl Genet* 118(6):1133–1142. doi:10.1007/s00122-009-0968-7
- Hinrichs D, Meuwissen THE (2011) Analyzing the effect of different approaches of penalized relationship in multistage selection schemes. *J Anim Sci* 89(11):3426–3432. doi:10.2527/jas.2010-3621
- 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(12):3212–3218. doi:10.2527/jas.2006-145
- Jarvis SF, Borralho NMG, Potts BM (1995) Implementation of a multivariate BLUP model for genetic evaluation of *Eucalyptus globulus* in Australia. In: Potts BM, Borralho NMG, Reid JB, Cromer RN, Tibbits WN, Raymond CA (eds) *Eucalypt Plantations: Improving*

- Fibre Yield and Quality, CRCTHF-IUFRO Conference 19–24 February 1995, Hobart, Tasmania, Australia. Hobart, Tasmania, Australia, pp 212–216
- Kerr RJ, Goddard ME, Jarvis SF (1998) Maximising genetic response in tree breeding with constraints on group coancestry. *Silvae Genet* 47(2–3):165–173
- Kojima M, Shindoh S, Hara S (1997) Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM J Optim* 7(1):86–125. doi:10.1137/S1052623494269035
- Lindgren D, Danusevičius D, Rosvall O (2009) Unequal deployment of clones to seed orchards by considering genetic gain, relatedness and gene diversity. *Forestry* 82(1):17–28. doi:10.1093/forestry/cpn033
- Lindgren D, Libby WS, Bondesson FL (1989) Deployment to plantations of numbers and proportions of clones with special emphasis on maximizing gain at a constant diversity. *Theor Appl Genet* 77(6):825–831. doi:10.1007/bf00268334
- Lindgren D, Matheson AC (1986) An algorithm for increasing the genetic quality of seed from seed orchards by using the better clones in higher proportions. *Silvae Genet* 35(5–6):173–177
- Lindgren D, Mullin TJ (1997) Balancing gain and relatedness in selection. *Silvae Genet* 46(2–3):124–129
- McKeand SE, Bridgwater FE (1998) A strategy for the third breeding cycle of loblolly pine in the Southeastern U.S. *Silvae Genet* 47(4):223–234
- McRae TA, Dutkowski GW, Pilbeam DJ, Powell MB, Tier B (2004) Genetic evaluation using the TREEPLAN® system. In: Li B, McKeand SE (eds) Forest genetics and tree breeding in the age of genomics: progress and future. IUFRO Joint Conference of Division 2, 1–5 November 2004, Charleston, SC, USA. pp 388–399
- Meuwissen THE (1997) Maximizing the response of selection with a predefined rate of inbreeding. *J Anim Sci* 75(4):934–940
- Mullin TJ, Ahlinder J, Yamashita M, Rosvall O (2013) OPSEL 1.0: A computer program for optimal selection in forest tree breeding by mathematical programming. Arbetsrapport från Skogforsk, Uppsala, Sweden. (*in press*),
- Nesterov YE, Todd MJ (1997) Self-scaled barriers and interior-point methods for convex programming. *Math Oper Res* 22(1):1–42. doi:10.1287/moor.22.1.1
- Pong-Wong R, Woolliams JA (2007) Optimisation of contribution of candidate parents to maximise genetic gain and restricting inbreeding using semidefinite programming. *Genet Sel Evol* 39(1):3–25. doi:10.1186/1297-9686-39-1-3
- Resende MFR, Muñoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M (2012) Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* 190(4):1503–1510. doi:10.1534/genetics.111.137026
- Robertson A (1960) A theory of limits in artificial selection. Proceedings of the Royal Society of London Series B Biological Sciences 153(951):234–249. doi:10.1098/rspb.1960.0099
- Robertson A (1961) Inbreeding in artificial selection programmes. *Genet Res* 2:189–194. doi:10.1017/S0016672300000690
- Vandenberghe L, Boyd S (1996) Semidefinite programming. *SIAM Review* 38(1):49–95. doi:10.1137/1038003
- Villanueva B, Avendaño S, Woolliams JA (2006) Prediction of genetic gain from quadratic optimisation with constrained rates of inbreeding. *Genet Sel Evol* 38(2):127–146. doi:10.1051/gse:2005032
- Woolliams JA (2007) Genetic contributions and inbreeding. In: Oldenbroek K (ed) Utilisation and conservation of farm animal genetic resources. Wageningen Academic Publishers, Netherlands, pp 147–165
- Yamashita M, Fujisawa K, Fukuda M, Kobayashi K, Nakata K, Nakata M (2012) Latest developments in the SDPA family for solving large-scale SDPs. In: Anjos MF, Lasserre JB (eds) Handbook on semidefinite, cone and polynomial optimization: theory, algorithms, software and applications international series in operations research and management Science 166. Springer Science + Business Media, New York, NY, pp 687–713. doi:10.1007/978-1-4614-0769-0_24
- Yamashita M, Fujisawa K, Kojima M (2003) Implementation and evaluation of SDPA 6.0 (semidefinite programming algorithm 6.0). *Optimization Methods and Software* 18(4):491–505. doi:10.1080/1055678031000118482
- Yamashita M, Fujisawa K, Nakata K, Nakata M, Fukuda M, Kobayashi K, Goto K (2010) A high-performance software package for semidefinite programs: SDPA 7. Tokyo, Japan