Theor Appl Genet (2001) 102:1056–1064

J. Fernández · M.A. Toro

Controlling genetic variability by mathematical programming in a selection scheme on an open-pollinated population in *Eucalyptus globulus*

Received: 14 October 1999 / Accepted: 26 July 2000

Abstract A model using integer quadratic mathematical programming has been developed to control the inbreeding level (or genetic diversity) through group coancestry in a selection programme for a forestry population structured in terms of maternal families coming from different locations. A method to calculate the average group coancestry between- and within-families for these open-pollinated populations is also proposed. This model has been applied to data from a breeding programme of Australian *Eucalyptus globulus*. The strategy proved to be effective as reductions of up to 50% for the group coancestry of the selected individuals were reached with a loss of only 5% of the maximum attainable selection differential (corresponding to truncation selection).

Keywords Restricted coancestry selection · Genetic diversity · Mathematical programming · Open-pollinated populations

Introduction

As was first pointed out by Robertson (1961) any selection programme implies an additional rise in the level of inbreeding in a population and an increase in the genetic drift. This is specially important when dealing with low census numbers, when selection pressure is high or when both heritability of the selected trait is low, and methods including information on relatives (e.g. family selection indices or BLUP) are used to estimate the breeding values of individuals. High inbreeding levels imply a threat to selection aims, as most characters of interest show inbreeding depression that counteract the response. Fitness-related traits also suffer from inbreeding depression, compromising the survival of the breeding population. On the other hand, the loss of genetic variability associated with genetic drift will reduce the long-term response for the character selected and for other traits that could be of interest in the future. Furthermore, the variance of the expected response to selection, and therefore the risk to a breeding programme, will also be increased (Woolliams and Meuwissen 1993).

All these reasons point to the need for maintaining genetic diversity in selection programmes by simultaneously optimising the genetic gain and the inbreeding borne by the population. Following Toro and Pérez-Enciso (1990) the different strategies can be classified according to the factor on which they act: (1) the selection criterion used (see Grundy et al. 1998 for the latest development of this idea); (2) the mating system imposed (see review by Caballero et al. 1996); and (3) the number of selected individuals and their contribution to the next generation (see Fernández and Toro 1999; and references therein).

To the third group of strategies belong the classical within-family selection and its modifications. It is well known that this type of selection results in the lowest variance for family size. However, as pointed out by Toro and Pérez-Enciso (1990), it has often been overlooked that there is a wide range of intermediate selection methods, which differ in the magnitude of the variance for family sizes. If, for example, only full-sib families were considered, all possible distributions of family sizes are equivalent to all possible forms of arranging marbles (the selected individuals) among boxes (families), each of fixed capacity (maximum family size). These authors show that some of the distributions of family sizes would lead to important reductions of inbreeding with small reductions of response compared to truncation selection. A natural extension of the modified within-family selection, taking also into account the possibility of families being related, is the restricted coancestry selection (RCS) where restriction is imposed on the overall coancestry of the selected individuals (including reciprocals and self-coancestries), as was first proposed by Brisbane and Gibson (1995a, b) in the context of animal breeding

Communicated by H. Becker

J. Fernández (✉) · M.A. Toro
Departamento de Mejora Genética y Biotecnología,
I.N.I.A., Madrid, Spain
e-mail: jmj@inia.es

and by Lacy (1995) in the context of conservation programmes.

Recently, Fernandez and Toro (1999) have summarised the advantages of using mathematical programming as a powerful tool for balancing gain and diversity, and they show how RCS and other types of selection, such as weighted selection (Toro and Nieto 1984; Lindgren 1991) and restricted coancestry weighted selection (Wray and Goddard 1994; Meuwissen 1997), could be implemented. Andersson et al. (1999) have applied linear programming to perform RCS in a tree-breeding program of *Pinus sylvestris* L.

Typical open-pollinated forestry populations are structured in respect of maternal families and the paternal lineage is unknown. Most simulation or experimental studies on this material assume that families are unrelated, i.e. no coancestry exists between individuals from different families. In the present work a method has been developed for the simultaneous control of genetic gain and diversity in an open-pollinated population with related families. The results obtained in the implementation of this method on data from a real population of *Eucalyptus globulus* are also included for comparison with the empirical method used in the current breeding plan.

Materials and methods

Selection models

A population structured into maternal families coming from different geographic locations is assumed hereafter.

Empirical method

Current strategy, as implemented in the breeding programme used as an example, consists of ranking the trees according to their genetic merit and then selecting the top *N*, with the restriction that no more than three trees can be taken from the same family. Therefore, selecting too-many close relatives is avoided and inbreeding is limited in some way.

Optimisation method

The proposed mathematical programme is aimed to minimise the average group coancestry of the selected trees so that the target function looks like

$$\min \begin{array}{l} \sum_{p=1}^{NP} \sum_{s=1}^{NF_p} \sum_{n=0}^{NI_{ps}} x_{psn} \cdot n \cdot f_{ps,ii} + \sum_{p=1}^{NP} \sum_{s=1}^{NF_p} \sum_{n=0}^{NI_{ps}} x_{psn} \cdot n \cdot (n-1) \cdot f_{ps,ij} \\ + \sum_{p=1}^{NP} \sum_{s=1}^{NF_p} \sum_{n=0}^{NI_{ps}} \sum_{s' \neq s}^{NF_p} \sum_{n'=0}^{NI_{ps'}} x_{psn} \cdot x_{ps'n'} \cdot n \cdot n' \cdot f_{p,ij} \end{array} \tag{1}$$

x_{psn} being a variable with value 1 when the *n* best individuals from family *s* from location *p* are to be selected, and 0 otherwise. *NP* is the number of locations, NF_p is the number of families from location *p*, and NI_{ps} is the number of trees belonging to family *s* from location *p*. The first part of the expression accounts for the self-coancestry of individuals ($f_{ps,\ ii}$), the second measures the coancestry between individuals from the same family ($f_{ps,\ if}$), and the third estimates this parameter for trees from different families of the same location ($f_{p,\ if}$). Notice that the correct expression to calculate the group coancestry should be divided by the squared number of selected trees; but, as this is a constant, the solution that minimises the group coancestry is the one that minimises the numerator.

The mathematical programme should include some constraints to make sure that only a solution per family is set, i.e.,

$$\begin{cases} \sum_{n=0}^{NI_{11}} x_{11n} = 1 \\ \quad \vdots \\ \quad \vdots \\ \sum_{n=0}^{NI_{NPNF}} x_{NPNFn} = 1 \end{cases} \tag{2}$$

and the total number of selected trees (*N*) is the desired one, i.e.,

$$\sum_{p=1}^{NP} \sum_{s=1}^{NF_p} \sum_{n=0}^{NI_{ps}} x_{psn} \cdot n = N. \tag{3}$$

One more condition must be added to the programme if the real purpose of the breeding scheme is to minimise the group coancestry but guaranteeing a determined selection differential (or expected response). The expression would be

$$\sum_{p=1}^{NP} \sum_{s=1}^{NF_p} \sum_{n=0}^{NI_{ps}} x_{psn} \cdot EBV_{psn} \geq N \cdot \overline{EBV}_{t+1} \tag{4}$$

where EBV_{psn} is the sum of the EBV of the *n* best trees of family *s* from location *p*, and EBV_{t+1} the minimum mean EBV desired for the group of selected individuals. By varying this level, a set of solutions can be found balancing the relatedness of the group and its mean genetic value.

If some kind of restriction on the maximum number of trees selected from families (all of them or just a group) is desired, this can be achieved by limiting NI_{ps} and discarding the rest of trees from the beginning. Actually no more than *N* trees (if available) from each family are to be included in the analysis as this is the maximum number of trees to select from the whole population. This practice reduces the number of variables and equations to be considered in the model.

If the aim of the breeding scheme is to maximise the expected response, but keeping the group coancestry below a certain level, then a maximisation mathematical programme can be constructed with the formula for the mean EBV of selected trees as an objective function and the group coancestry as a restriction. The rest of the restrictions would remain as they are. The trade-off obtained in such a programme would be essentially the same.

A simultaneous control of coancestry and response could be made by including both items (with different signs) into a combined objective function such as

$$\min \begin{array}{l} \sum_{p=1}^{NP} \sum_{s=1}^{NF_p} \sum_{n=0}^{NI_{ps}} x_{psn} \cdot n \cdot f_{ps,ii} + \sum_{p=1}^{NP} \sum_{s=1}^{NF_p} \sum_{n=0}^{NI_{ps}} x_{psn} \cdot n \cdot (n-1) \cdot f_{ps,ij} \\ + \sum_{p=1}^{NP} \sum_{s=1}^{NF_p} \sum_{n=0}^{NI_{ps}} \sum_{s' \neq s}^{NF_p} \sum_{n'=0}^{NI_{ps'}} x_{psn} \cdot x_{ps'n'} \cdot n \cdot n' \cdot f_{p,ij} \\ - \lambda \cdot \left(\sum_{p=1}^{NP} \sum_{s=1}^{NF_p} \sum_{n=0}^{NI_{ps}} x_{psn} \cdot EBV_{psn} \right) \end{array} \tag{5}$$

where λ is a scaling factor. Varying this value (i.e. giving more or less importance to the genetic level of the selected trees) a trade-off between coancestry and the expected response can be obtained.

As described, the mathematical programme is an integer quadratic one. Although there are several computer programmes dealing with such problems (as an example the OSL routines of IBM used by the authors) the high number of trees and families makes the programme too large to be managed efficiently by most computers. Hence, the simulated annealing algorithm was also used to solve the programmes, giving equal solutions with low computational (time and memory) cost. The simulated annealing is a minimisation process that uses the following procedure. (1) It starts with a random feasible solution. (2) An alternative solution is generated by a random small change; if the new solution is better than the previous one then the new solution is accepted; if worse, then the solution will be accepted with a probability of $\Omega = exp(-\Delta/T)$, where Δ isthe difference between the values of the alternative and actual solutions and T is a ”cooling” factor or temperature. (3) This step is repeated several times; so that the value of T is reduced by a factor and the process then continues. At the beginning many alternatives are accepted but as T decreases it becomes more difficult to accept an alternative that is worse than the actual solution. The nature of the simulated annealing algorithm makes it possible to avoid local minima in the first steps, and therefore is superior to other algorithms based on following the direction with the highest slope.

A combined objective function was used to make an easy simultaneous control of the expected response and coancestry. The routine started with a value of T=1 and was reduced by a factor of 0.9 for each step. The initial value of T must be proportional to the magnitude of the difference between alternative solutions (Δ) as too-low Ts may cause the algorithm to get stuck in local minima from the beginning (Press et al. 1989). Up to 100 steps were allowed with 10000 alternative solutions generated in each of them. The optimal solution was assumed to be found when the solution was not changed during the 10000 suggested modifications of one step. If the solution does not change further, the current solution is very unlikely to be improved upon, and the temperature has reached a sufficiently low value. The alternative solutions were obtained by choosing one family at random. If its contribution wasn’t zero it was decreased by one unit, then choosing at random another family and increasing its contribution by one unit.

If only relations within families were to be considered then the mathematical programme would be almost equal by just removing the third part of the expression in the objective function (see formula 1). In this case the problem would be linear, as in Andersson et al. (1999).

Coancestry in open-pollinated populations

As previously stated, in open-pollinated designs the pedigree is not completely known since there is only information on the maternal lineage. This imposes a degree of uncertainty on the coancestry or relatedness of the individuals, and only an estimation of the mean kinship within and between different groups can be used. Depending on the degree of self-fertilisation the species bears (s), i.e. the proportion of offspring generated by self-fertilisation, then within a particular family there will be a mixture of relations between individuals ranging from self full-sibs to half-sibs. Assuming that parental trees were not inbred, the coancestry between two individuals generated by self-fertilisation would be 1/2; 1/4 if only one of them arose from self-pollination and 1/8 if both came from outcrossing. Thus, multiplying each case by its probability of occurrence, the average coancestry between individuals of the same family would be

$$f_{ps,ij} = s^2 \frac{1}{2} + 2s(1-s)\frac{1}{4} + (1-s)^2 \frac{1}{8} \tag{6}$$

and thus the self-coancestry of any individual is

$$f_{ps,ii} = \frac{1}{2}\left(1 + f_{ps,ij}\right). \tag{7}$$

If the population structure considers families from several locations it seems reasonable to suppose that some relatedness exists between families growing in the same area. Following classical quantitative genetics theory (Falconer and Mackay 1996), in a large population with mixed self- and cross-fertilisation the average inbreeding of an individual at the equilibrium stage (and thus the coancestry between any two individuals) will be:

$$f_{p,ij} = \frac{1-C}{1+C} = \frac{s}{2-s} \tag{8}$$

where C and s are, respectively, the proportion of individuals generated by outcrossing and self-fertilisation. Now, families within the same location are not unrelated and the coancestry between two trees from different families (but from the same location) will be estimated as the one for a population at equilibrium. Using the same arguments as above, and assuming that parents are inbred, the average coancestry between individuals from the same family will now be

$$\begin{aligned} f_{ps,ij} = {} & s^2\left(\frac{1}{2} + \frac{1}{2} f_{p,ij}\right) + 2s(1-s)\left(\frac{1}{4} + \frac{3}{4} f_{p,ij}\right) \\ & + (1-s)^2\left(\frac{1}{8} + \frac{7}{2} f_{p,ij}\right) \end{aligned} \tag{9}$$

and the self-coancestry can be calculated as in formula (7).

Studied population

The trees included in the present study belong to an ongoing experiment by the company Empresa Nacional de Celulosas (ENCE) designed to evaluate traits of interest, under diverse environmental conditions, on open-pollinated progenies from 260 parents of Australian *E. globulus*. These families originated from 46 seedlots of the collection made by the CSIRO in 1988 and were assigned to ten different provenances by Jordan et al. (1994). In this way a hierarchical design is constructed with every tree belonging to a particular family and every family coming from a particular location (or provenance). Two separate trials were carried out, respectively, in northern (seven sites) and southern (eight sites) regions of Spain characterised by different rainfall regimes and temperatures. The experimental layout was in a block design, the 260 Australian families being distributed in 20 blocks per site. One lot of commercial seed and two clones were used as control populations. Height and diameter at breast height (1.3 m) were measured on all the surviving trees at 7 years of age, and heterogeneity of variances across blocks were corrected by a Bayesian estimate of the within-block standard deviation. The following bivariate model was employed (Soria et al. 1998):

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

where y is a matrix of observations (height and diameter) of the order n×2, β is a p×2 matrix of fixed effect including mean and block effects, **u** is a random matrix of $n_u \times 2$ random additive genetic effects and **e** is a n×2 matrix of random residuals. The conditional assumptions in a model fitting provenances as genetic groups were

$$\mathbf{u}|\Sigma_\mathbf{u}, \mathbf{A}, \mathbf{Q}, \mathbf{g} \sim \mathbf{N}(\mathbf{Qg}, \mathbf{A} \otimes \Sigma_\mathbf{u}) \text{ and,}$$

$$\mathbf{e}|\Sigma_\mathbf{e} \sim \mathbf{N}(\mathbf{0}, \mathbf{I} \otimes \Sigma_\mathbf{e})$$

where **A** is the additive genetic relationship matrix assuming a degree of self-fertilization of 0.3 (Potts et al. 1995), **Q** is an $u_n \times 13$ matrix whose elements are the expected proportion of genes of each tree deriving from each provenance, and g is the 13×2 matrix of the provenance effects for both traits. $\Sigma_\mathbf{u}$ and $\Sigma_\mathbf{e}$ are the (co)variance matrices for the additive genetic effects and residuals. In the southern region a genetic analysis of survival to drought was also implemented adjusting the same model on a vector of binary records (with the codes: 0=dead; 1=alive) (Toro et al. 1998). In order to obtain inferences about genetic parameters and breeding values, both the bivariate and the univariate models were analysed with a Bayesian procedure coupled with Gibbs sampling (see Soria et al. 1998 and Toro et al. 1998 for details).

For the optimisation method, and assuming also a degree of self-fertilization of 0.3, the average coancestries between individuals from the same family, if no relation exist between families, were 0.21125 (cross-coancestry) and 0.60562 (self-coancestry) according, respectively, to formulae (6) and (7). When related families are assumed, the mean coancestries between individuals from the same family will be 0.35044 and 0.67522 (formulae 9 and 7), and between individuals from the same provenance but

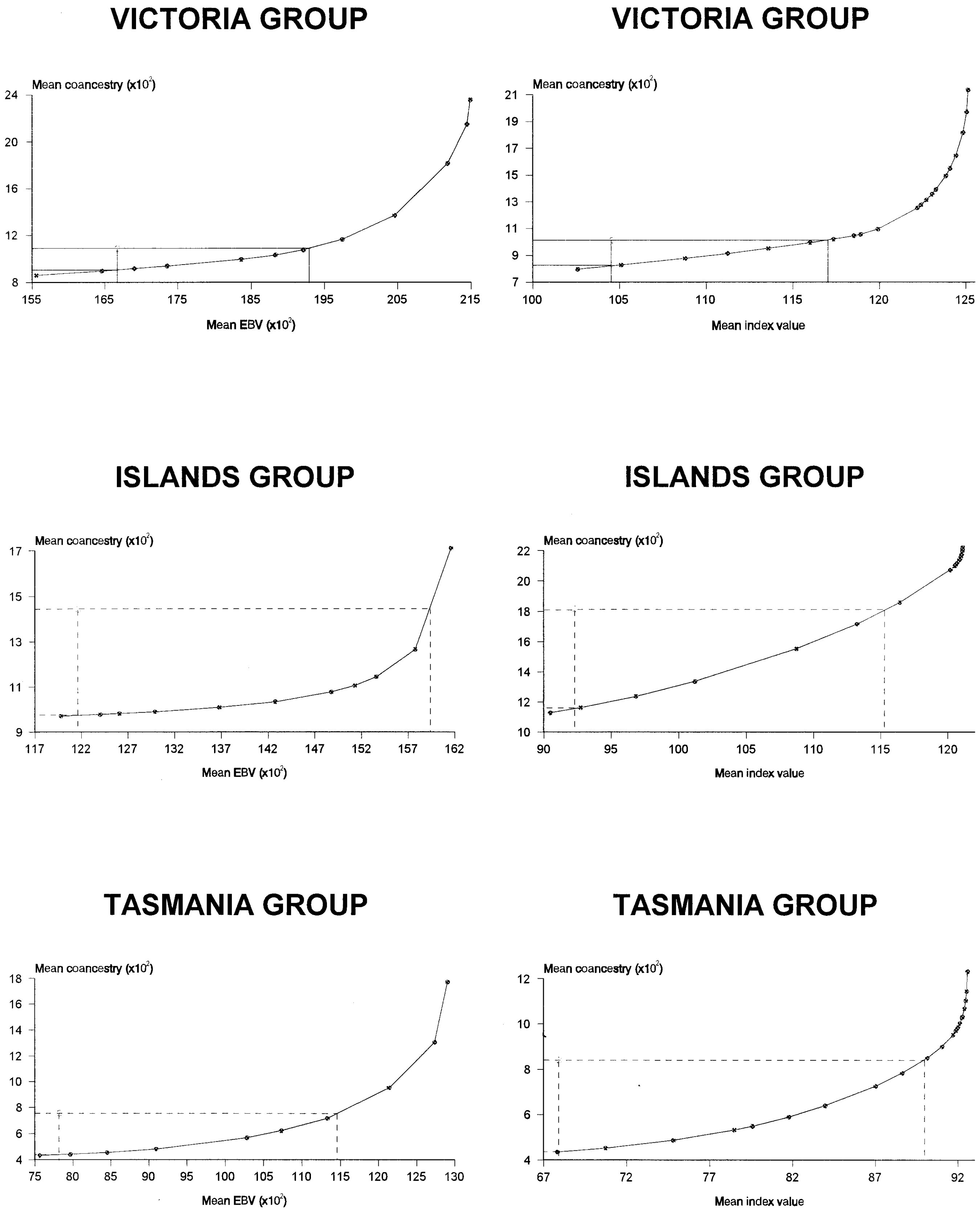

Fig. 1 Trade-offs between mean EBV (mean index value of the selected trees and their group coancestry (*points* on solid line) obtained by mathematical programming. *Open circles* are the solutions yield by the standard current strategy. *Left column* shows data on height on the northern test and *right column* data on productivity index in the southern test for the three groups of provenances considered

Table 1 Mean EBV (height) or mean index value (productivity index) and group coancestry ($\times 10^{-2}$) of the selected trees using truncation selection, standard method (truncation with limit in the number of trees from each family) and mathematical programming (MP, solutions which yield 95%, 90% and 75% of the maximum expected response)

Trait	Method	Group of provenances					
		Victoria		Islands		Tasmania	
		Mean	Coancestry	Mean	Coancestry	Mean	Coancestry
Height	Truncation	2.15	23.57	1.62	16.16	1.29	17.73
	MP 95%	2.04	13.58	1.54	11.43	1.23	10.83
	MP 90%	1.93	10.97	1.45	10.53	1.16	8.00
	MP 75%	1.61	8.27	1.21	9.72	0.97	5.23
	Standard	1.67	10.85	1.22	14.42	0.78	7.55
Productivity index	Truncation	125.12	21.34	121.10	22.21	92.59	12.32
	MP 95%	118.87	10.55	115.04	17.97	87.96	7.58
	MP 90%	112.61	9.35	108.99	15.60	83.33	6.23
	MP 75%	93.84	7.00	90.82	11.31	69.44	4.44
	Standard	104.57	10.09	92.32	18.14	67.92	8.39

different families 0.17647 (formula 8). The number of selected individuals in the present work is N=100. For practical reasons three subsets of provenances are considered, separately grouping those sharing neighbouring areas of distribution or similar environmental conditions: Victoria Group (provenances 1, 2 and 4 in mainland Australia; Islands Group (provenances 3 and 5 in King Island and the Furneaux Group); Tasmania Group (provenances 6, 7, 8, 9 and 10 in Tasmania). All studies were performed in parallel over the three sets.

Traits under selection

The main goal in the present forestry breeding programme is to increase the amount of commercial wood yield in a plot. In the northern test the productivity of a tree is measured through its height, as higher trees usually produce more wood. But in the southern area, where drought is an important problem for tree survival, an empirical index including different single traits is used to select the most-productive trees. Basically the index is proportional to the volume of the trunk (as if it was a cone) and the drought resistance of the tree. To calculate the value of the index, the EBV for height, the diameter at breast height and drought resistance (in percentage) are used, all scaled with respect to the values from provenance 10, taken as a reference.

Results

In Fig. 1 the trade-offs between group coancestry and the mean value of the selected trees are shown for the three subsets of provenances, and for height in the northern test (left column) and the productivity index in the southern test (right column). The point at the right end of every graph corresponds to the performance of pure truncation selection not taking into account the group coancestry, and thus the maximum attainable mean genetic value for the selected trees. The rest of the points in the line are solutions reached when lowering the selection differential imposed as a restriction.

A common feature to all graphs is a steep slope on the right side of the lines. This means that a high reduction in the group coancestry can be obtained with little loss in the value for the character of the selected group. Further relaxation on the desired level of the character yields no substantial benefits to coancestry. This can also be seen in Table 1 where the mean EBV or the index value of the selected trees and their group coancestry are presented for mathematical programmes with a restriction on the expected response assuming 95%, 90% or 75% of the maximum attainable value, respectively in addition to the parameters for truncation selection and the current selection method).

In all graphs of Fig. 1 an open circle marks the solution corresponding to the empirical selection strategy of limiting the number of trees from each family to three. A reduction in the group coancestry can be observed when compared with unrestricted truncation selection, but the loss of the mean value of the selected trees is very large (see also Table 1). Dashed lines are drawn to compare this solution with one obtained from mathematical programming with the same level of coancestry or the same mean value for the trait. The latter are substantially better as they give a lower group coancestry or a higher genetic merit, respectively. Moreover, if looking to the trade-offs obtained by mathematical programming, then several solutions can be found which yield both higher levels for the character of interest and lower coancestry values at the same time.

The number of selected trees per family using the empirical strategy, and one obtained by mathematical programming with approximately the same group coancestry, are shown in Fig. 2 (for both types of traits and the three groups of provenances as in Fig. 1). Families are ordered by their mean genetic value (EBV or index value, respectively). Open bars show the results for limited contributions, and filled ones for the unrestricted strategy (right axis). The lines (values on the left axis) show the mean EBV or index value for the best families. In all graphs those families with the greatest mean value tend to contribute more individuals than the worse ones. But this relation is not straight (for any of the strategies) and some discrepancies can be found).

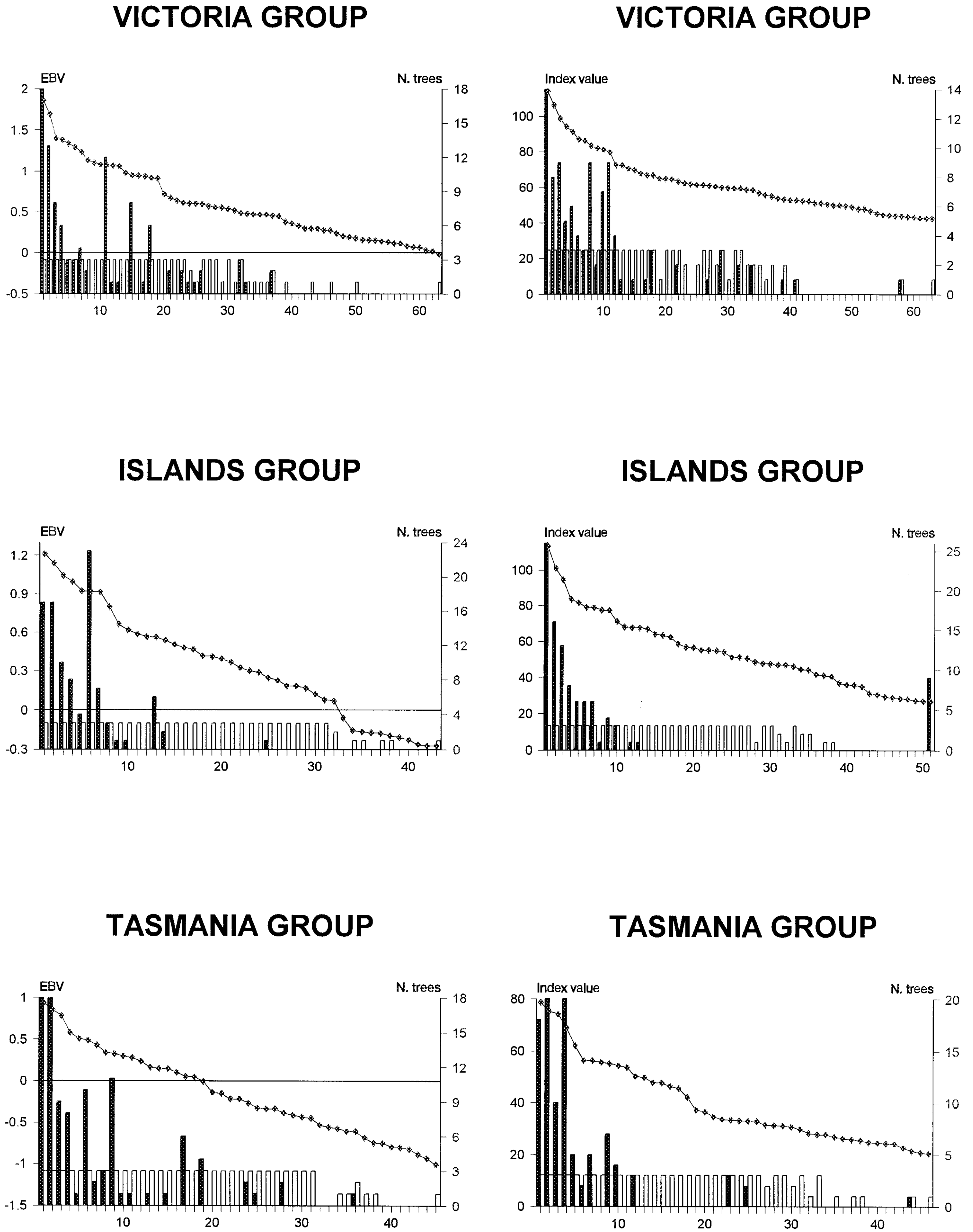

Fig. 2 Number of trees selected from each of the families with the best performance. *Open bars* are results from empirical current breeding strategy. *Solid bars* are results from solutions with a similar group coancestry as previous ones but obtained by quadratic mathematical programming. *Lines* show the mean value for the trait of each family. The *left column* shows data on height in the northern test and the *right column* data on productivity index in the southern test for the three groups of provenances considered

Discussion

There is no doubt about the final aim of a breeding program: namely, to reach higher performance levels of the trait of interest in the population. The most direct way, according to classical response to selection theory, is to rank the individuals by their estimated breeding value and use the best ones as parents for the next generation, thus maximising the gain in the short-term. But selecting few individuals will lead to a rapid exhaustion of genetic variability because of the loss of information from non-selected individuals, and because the top individuals probably would be relatives and would carry the same alleles in the loci affecting the trait. The lack of genetic variability will compromise genetic improvement in the long-term response. Moreover, the increased genetic drift will lead to a rise of inbreeding, and the appearance of inbreeding depression because of recessives deleterious in a homozygous state being more probable both in the trait of interest (counteracting the effect of selection) and in the fitness traits (compromising the survival of the population). Thus, any efficient breeding program should also take into account the genetic variability maintained in the population.

Different methods or parameters have been suggested to measure the genetic diversity of a population (Ballou and Lacy 1995; Lacy 1995; Caballero and Toro 2000). According to the latter (and references therein), the best way of maintaining genetic variability seems to be by minimising the group coancestry (defined as the average pairwise coancestry of a given group of individuals including reciprocals and self-coancestries) of the selected individuals. The benefits from using this method can be summarised as follows: (1) Minimising group coancestry maximises the diversity in terms of the expected heterozygosity. (2) If families are unrelated, or relations among them are uniform, the technique reduces to equalise family sizes and, therefore, maximises the effective population size in a single generation. (3) If individuals have no uniform relation this criterion will maximise the effective population number by equalising not only contributions from one generation but from individuals in all previous generations. This is equivalent to minimising the number of individuals with common ancestors. Zeng et al. (1997) and Lindgren and Mullin (1997) have pointed out another advantage of controlling coancestry instead of inbreeding itself. If starting from unrelated non-inbred individuals, in the first cycle of breeding (where most forestry programmes actually are, due to the long generation interval) all possible combinations would yield the same level of inbreeding in the next generation. Thus selection would not account for genetic diversity leading to many paremts being selected from a few families and the build-up of undesirables kinships.

The calculation of the group coancestry is rather trivial through the use of the additive relationship matrix when the pedigree is known. But, in open-pollinated populations this is not possible, as only the maternal line is traceable. One solution is to assume that all families are not related with parents which are not inbred. When the area of distribution of the trees is very large this does not seem reasonable as trees growing near each other have a greater probability of being relatives than those grown far apart. The specific structure of the population considered in the present work is even clearer in this respect with families coming from different geographic locations (what have been called provenances). As shown above, using classical population-genetics theory and accounting for the possibility of a certain degree of self-fertilisation, the mean coancestry between individuals from the same family or provenance can be estimated.

From all the points discussed above it can be concluded that a more sensible goal of a breeding programme should be to find the optimum balance between the gain obtained for the trait and the genetic diversity maintained in the population (measured through the group coancestry). As presented, the problem is combinatorial with a huge number of solutions corresponding to all forms of taking the desired number of selected individuals from the available families. An exhaustive search for the solution with the best performance is impossible, even in small populations, so different methods and algorithms have been proposed to find it. Basically, all strategies deal with maximising a function where the values for expected gain and group coancestry are combined, the latter with a negative sign and weighted by a factor indicating the relative importance given to maintaining variability in the breeding programme. Then, iterative approximated methods are implemented adding to, and removing individuals from, the selected group until no improvement in the solution can be made (Lindgren and Mullin 1997; Zheng et al. 1997). Other authors look for the optimal solution for the combined function by means of Lagrange multipliers and first derivatives (Lindgren et al. 1989; Brisbane and Gibson 1995a, b) or quadratic interpolation (Lindgren et al. 1993; Wei and Lindgren 1995). These approaches have the disadvantage of yielding non-integer solutions that could include negative contributions for some families. Hence, iterative algorithms have to be used to eliminate negative contributions and the final results rounded to account for the fact that contributions must be integers.

Mathematical programming is a widely used tool in different fields that allows one to find the combination of values which maximise or minimise a function (called the objective function), subject to a number of restrictions or constraints. The advantage of such a method relies on the fact that it is an exact method (i.e. it ensures that the solution is optimal) and that it allows obtaining only integer solutions, if desired. This technique has been used in animal breeding to implement minimum coancestry matings (Toro et al. 1988), to perform mate selection (Jansen and Wilton 1985; Kinghorn 1987; Sanchez et al. 1999), or to carry out restricted coancestry selection, weighted selection or restricted coancestry weighted selection (Toro and Nieto 1984; Toro et al. 1988; Toro and Pérez-Enciso 1990; Fernández and Toro 1999). Recently, Andersson et al. (1999) applied integer

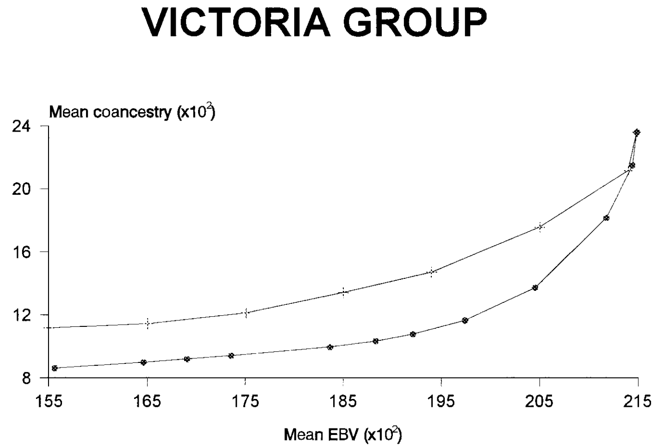

Fig. 3 Trade-offs between mean EBV of the selected trees and their group coancestry obtained by mathematical programming assuming related families (*solid line with points*) or unrelated families (*dashed line with pulses*). Data on height for the group of provenances from Victoria

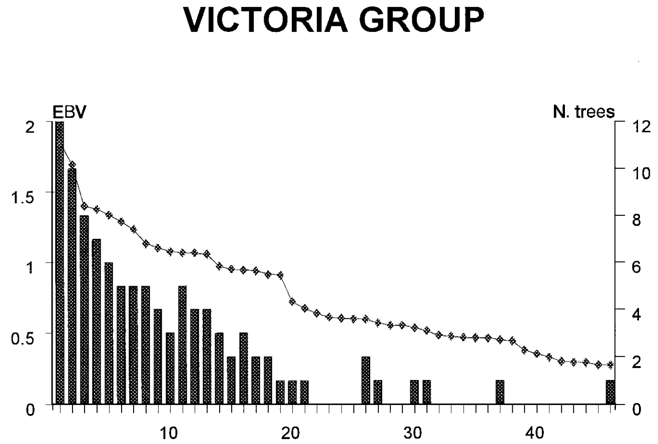

Fig. 4 Number of trees selected from each of the families with the best performance in a design assuming unrelated families (*solid bars*). The *line* shows the mean value for the trait of each family. Data on height for the group of provenances from Victoria

linear programming to a selection programme with controlled group coancestry. The logic underlying the construction of this program is similar to the one used by the authors of the present paper: one objective function measuring the expected performance of the group for the trait under selection, and a restricted function measuring the group coancestry. The difference is in the relation structure of the population studied here that makes the function of group coancestry to have a quadratic term (see Material and methods).

The present work has shown how to implement restrictions in the group coancestry through quadratic programming. This method has proven to be very effective at balancing the expected gain and the genetic diversity. Furthermore, a high slope in the right side of trade-offs for all combinations of provenance and trait means that a substantial reduction in coancestry can be attained with a small loss in gain. Further relaxation in the expected gain gives little advantage to the observed decrease of group coancestry (see Fig. 1). In Table 1 it can be seen that reductions of up to 50% of the group coancestry can be obtained with only 5% loss from the maximum expected gain with truncation selection. Therefore, the maintainance of genetic diversity should not be an obstacle for reaching acceptable rates of response.

Current empirical strategy fails to efficiently balance gain and diversity, since a reduction in group coancestry is only possible with low levels of the trait in the selected trees (Fig. 1 and Table 1). By limiting the number of selectable trees from a family one tries to maximise the effective population size by minimising the variation of family size, making some form of within-family selection. But this criterion forces one to choose trees from families with a poor performance and to discard individuals with much-better breeding values. Mathematical programming facilitates reaching a compromise solution between the two forces and makes the solutions yielded by this method to be much better than those obtained by the empirical one, either increasing genetic gain for the same level of coancestry or decreasing coancestry for the same level of genetic gain (Fig. 1).

A model with unrelated families can be invoked for simplicity and the mathematical programme will then become a linear one. But if this assumption is not true then the loss in efficiency could be great as can be seen in Fig. 3 for data of the Victoria group in the northern test. Comparing both lines it is clear that, for the same level of selection differential, the unrelated families-design yield more group coancestry. This is true for all groups of provenances and for the two traits considered (data not shown). Maybe the model presented is too conservative, but as no great losses in response are found it is worth avoiding as much inbreeding as possible. It is also advisable to account for the kinship between- and within-families when selecting individuals, although the estimation of genetic value does not account for it (as is the case for the EBVs in the present work), even when only the phenotypic values are available. As software actually exists, and computer capabilities are being rapidly increased, the model with related families seems to be the one to choose in the kind of problems presented in this paper.

Different authors (for example Lindgren et al. 1989; Lindgren 1991; Lindgren et al. 1993; Andersson et al. 1999) have reached the same conclusion: when implementing RCS, families with the best trait performance also contribute more to the selected group. Similar results can be seen in Fig. 2 of the present paper, but some discrepancies are found with the better-ranked families contributing less than others below them. This result is different from the one found by Andersson et al. (1999) in which contributions were strictly proportional to family performance. But the Andersson et al. (1999) population structure (as most studies) assumes unrelated families, whereas in the present one different degrees of relatedness are supposed for different couples of families; therefore, it would be advisable to select more trees from non-in-top families but from different provenances, thus and not related. This factor also explains the differences between the number of families that effectively contribute to the present work and this number in other studies. If families are not related, then using trees from more families than in truncation selection is the only way of reducing the coancestry created by selecting more trees from the top families. In the current study, the levels of coancestry can also be lowered by choosing families from different provenances. In Fig. 2 it can be seen how a mathematical-programming solution takes individuals from fewer families than truncation selection, in opposition to what would be expected. If families are assumed to be unrelated then the correlation between family merit and contrubutions is tighter (see Fig. 4 for a particular case) although there are some cases where, just by chance, some individuals of a given family could have higher estimated breeding values than all the individuals of a better-ranked family.

There is some uncertainty in the values of the true coancestries among families because they have been calculated assuming an equal value for all the families coming from the same provenance. However, this parameter seems to be very influential when trying to optimise jointly the selection response and the maintenance of genetic variability; therefore, methods based on molecular markers that could refine these coancestries could be very valuable.

Acknowledgements The authors thank L. Silió M.C. Rodríguez, F. Soria and F. Basurco for providing the data, and two anonymous referees for helpful comments on the manuscript. This work was supported by a I N I A – E N C E contract (CC96-104).

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

Ballou JD, Lacy RC (1995) Identifying genetically important individuals for management of genetic variation in pedigreed populations. In: Ballou JD, Gilpin M, Foose TJ (eds) Population management of survival and recovery. Analytical methods and strategies in small-population conservation. Columbia University Press, New York, pp 76-111

Brisbane JR, Gibson JP (1995a) Balancing selection response and inbreeding by including predicted stabilised genetic contributions in selection decisions. Genet Sel Evol 27:541–549

Brisbane JR, Gibson JP (1995b) Balancing selection response and rate of inbreeding by including genetic relationships in selection decisions. Theor Appl Genet 91:421–431

Caballero A, Toro MA (2000) Interrelations between effective population size and other pedigree tools for the management of conserved populations. Genet Res 75:331–343

Caballero A, Santiago E, Toro MA (1996) Systems of mating to reduce inbreeding in selected populations. Anim Sci 62: 431–442

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Longman, Essex, England

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

Grundy B, Luo ZW, Villanueva B, Woolliams JA (1998) The use of Mendelian indices to reduce the rate of inbreeding in selection programmes. J Anim Breed Genet 115:39–51

Jansen GB, Wilton JW (1985) Selecting mating pairs with linear programming techniques. J Dairy Sci 62:1302–1306

Jordan GJ, Borralho NMG, Tilyard P, Potts BM (1994) Identification of races in *Eucalyptus globulus* spp. *globulus* based on growth traits in Tasmania and geographic distribution. Silvae Genet 43:292–298

Kinghorn B (1987) On comparing strategies for mate allocation. J Anim Breed Genet 111:81–88

Lacy RC (1995) Clarification of genetic terms and their use in the management of captive populations. Zoo Biol 14:565–578

Lindgren D (1991) Optimal utilization of genetic resources. For Tree Improv 23:49–67

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

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 77:825–831

Lindgren D, Wei RP, Bondesson FL (1993) Optimal selection from families. Heredity 70:619–621

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

Potts B, Volker P, Hodge G, Borralho N, Hardner C, Owen J (1995) Genetic limitations in the explotation of base populations of *Eucalyptus globulus* spp. *globulus*. In: Potts BM, Borralho NMG, Reid J, Cromer R, Tibbits W, Raymond C (eds) Proc CRC-IUFRO Conf on Eucalypt Plantations: Improving Fibre Yield and Quality, 19–24 February 1995, Hobart, Australia. CRC Temperate Hardwood Forestry, Hobart, Australia, pp 217–221

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1989) Numerical recipes. Cambridge University Press, Cambridge, UK

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

Sánchez L, Toro MA, García C (1999) Improving the efficiency of artificial selection: more selection pressure with less inbreeding. Genetics 151:1103–1114

Soria F, Basurco F, Toval G, Silió L, Rodríguez MC, Toro M (1998) An application of Bayesian techniques to the genetic evaluation of growth traits in *Eucalyptus globulus.* Can J For Res 28:1286–1294

Toro MA, Nieto B (1984) A simple method for increasing the response to artificial selection. Genet Res 44:347–349

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

Toro MA, Nieto B, Salgado C (1988) A note on minimization of inbreeding in small-scale selection programmes. Livest Prod Sci 20:317–323

Toro MA, Silió L, Rodríguez MC, Soria F, Toval G (1998) Genetic analysis of survival to drought in Eucalyptus globulus. Proc 6th World Congr Genetics Applied to Livestock Production, 11–16 January 1998, Armidale, Australia, pp 499–502

Wei RP, Lindgren D (1995) Optimal family contributions and a linear approximation. Theor Pop Biol 48:318–332

Woolliams JA, Meuwissen THE (1993) Decision rules and variance of response in breeding schemes. Anim Prod 56:179–186

Wray NR, Goddard ME (1994) Increasing long-term response to selection. Genet Sel Evol 26:431–451

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