FORWARDS VERSUS BACKWARDS SELECTION: TRADE-OFFS BETWEEN EXPECTED GENETIC GAIN AND RISK AVOIDANCE

R. D. BURDON and S. KUMAR

New Zealand Forest Research Institute,
Private Bag 3020, Rotorua, New Zealand

ABSTRACT

Genetic gains were simulated for backwards and forwards selection (1.5- and second-generation respectively), using offspring of 300 plus-trees from a base population of 30 000, and making 15 backwards selections throughout. Options were: (1) backwards selection from half-sib progeny test; (2) forwards selection within same; (3) forwards selection within select-pollen polycrosses; (4) forwards selection within single-pair crosses. Parameters varied were: heritability (h^2) (0.1–0.6 in progeny, 0.05–0.3 in base population); dominance : additive genetic variance (d) (0–1); total offspring (N) (9000 or 30 000); and forwards selections made (n) (15–120). Each set of conditions involved 200 replicate runs. For mean gain across runs (Mean) Option 4 was consistently best, except with high d; Option 3 excelled Option 2 at higher heritabilities and lower n; Option 2 excelled Option 1 only at h^2 = 0.6 and $n \leq 30$. Standard deviations of breeding values among selections (SD) and among-run coefficients of variation for genetic gain (CV) were predictably lower for forwards than for backwards selection, with superimposed influences of both h^2 and d. Increasing n, while reducing Mean and CV, minimally affected SD. Varying N did not greatly affect comparisons. Crucially, among-run percentile distributions of gains indicated that, for $n \geq 15$, realisation of any theoretical superiority of forwards selection is almost guaranteed despite imprecise genetic information on individuals. Thus, increased protection against selection errors with higher n should seldom compensate for reduced selection intensity.

Keywords: genetic gain; progeny testing; second-generation selection; stochastic simulation.

INTRODUCTION

Forest tree breeding programmes typically employ both forwards and backwards selection. Forwards selection refers to selection among the offspring of the previous generation of select parents. As such, it is an essential part of a long-term recurrent selection scheme which involves achieving cumulative genetic gain by raising the frequencies of desirable alleles through repeated cycles of selection, intermating and selection, and so on. Backwards selection typically amounts to reselection of provisional selections on the basis of progeny-test performance, although it can also be done on the basis of clonal-test performance.

While the need for forwards selection in a breeding population is beyond doubt, its desirability for selecting for a production population (typically a seed orchard) has been subject to debate (Ruotsalainen & Lindgren 1998). Backwards selection usually has the great advantage of being generally based on a high effective heritability, but the effective selection intensity will typically be quite low. By contrast, forwards selection can often allow much higher selection intensities (through the sheer numbers of offspring in progeny trials), but the effective heritability may be much lower. Under extreme heritabilities the relative merits of these two forms of selection, in terms of genetic gain, may be clear, with very high heritabilities (say, ≥0.7) pointing to forwards selection and very low heritabilities (say, <0.05) pointing to backwards selection. However, economically important traits can often show low to intermediate heritabilities (say, 0.1–0.3) which will make the relative merits less obvious. And even if the expected gain from forwards selection is greater, there is the risk that forwards selection could include some bad mistakes in some individuals that are well below the expected genetic merit. This risk has undoubtedly tended to inhibit tree breeders from using forwards selection for seed orchards, since they want to guarantee the genetic merit of all genotypes used therein, and the risk could be significant if only a few parents are chosen for the seed orchard(s). In principle, it should be possible to avert such risks by using so many parents in a seed orchard that expected gains are virtually assured, in effect minimising the sampling error about the deterministic gain expectations. This would entail some sacrifice of expected gain through lowering selection intensity. However, in very large progeny trials it should be possible to select enough offspring genotypes to give very small sampling errors without severely compromising selection intensity.

This paper covers a set of stochastic simulations which address the trade-offs concerned, for several selection scenarios, and ranges of heritabilities, ratios of dominance to additive genetic variance, numbers of forwards selections, and population sizes. The backwards selection corresponds to 1.5-generation selection and the forwards selection to three forms of second-generation selection.

METHODS

Underlying Conditions

The following basic conditions were assumed, primarily for simplicity: base population mean arbitrarily set at zero; all effects normally distributed about zero; all effects, genotypic and environmental, orthogonal to each other, hence no interactions; no epistasis; no maternal effects; all selected parents contributing equally to offspring generation; additive genetic variance fixed arbitrarily at unity; with other variances set to satisfy specified heritabilities, etc.

Thus the basic genetic model was

$$P = \mu + A + D + E$$

where P = effect of individual phenotype

μ = population mean

A = additive genetic value, i.e., breeding value, with an arbitrary variance of 1

D = genetic dominance deviation

E = environmental effect.

Additional assumptions that were implicit in the modelling were: (1) no inbreeding depression, (2) effective heritability independent of progeny-test size.

Thus, expressing the phenotypic variance, σ^2_P, in terms of its component variances

$$\sigma^2_P = \sigma^2_A + \sigma^2_D + \sigma^2_E = 1/h^2$$

and $\quad \sigma^2_E = 1/h^2 - \sigma^2_A - \sigma^2_D$

where $\quad h^2 = 1/\sigma^2_P$, (given $\sigma^2_A = 1$) = narrow-sense heritability.

Selection Scenarios

All simulations were started with a base population of 30 000 which was generated (Step 1) on the basis of assumed genetic parameters, from which 300 plus-trees would be selected as the best phenotypes (Step 2). From there several options were addressed for comparison: (1) backwards selection using maternal half-sib families (of non-select pollen parentage) of the 300 select parents, (2) forwards selection within the same families ("base half-sibs"), (3) forwards selection within polycross progenies (of select pollen parentage) of the 300 select parents ("select polymix"), (4) forwards selection within random single-pair crosses between the select parents ("select pair-cross"). It is assumed that a fifth option, the use of polycrosses using select pollen parents for backwards selection, based on the material for Option 3, would give the same gain as that from backwards selection using non-select pollen (Option 1).

Special conditions were: (i) narrow-sense heritability for making the initial selections (h^2_0) one-half the assumed h^2 for progeny trials, (ii) family sizes for Option 4 double those for the other options, in order to maintain a constant total number of offspring (N).

For backwards selection 15 parents were always reselected on the basis of progeny-test performance, this being deemed to be a prudent minimum for a seed orchard. For forwards selection, the number of selections made (n) was varied: 15, 30, 60, and 120.

Superimposed on these were factorial variations (with some common-sense omissions) of:

- heritability (h^2) in field trials (0.1, 0.2, 0.4, 0.6),
- ratio (d) of dominance to additive genetic variance (σ^2_D/σ^2_A) (0, 0.3, 1), this being relevant only for Option 4,
- total number (N) of offspring (9000 and 30 000).

Combinations of conditions that were omitted were: $d = 1$ for $h^2 \geq 0.4$, $d = 0.3$ for $h^2 = 0.6$.

Forwards selection was subject to the restriction that no more than $n/10$ individuals could come from any one family in Options 2 and 3 and no more than $n/5$ in Option 4.

Simulation Processes

The main steps involved in simulation were: (1) creating the initial base population of 30 000; (2) making 300 "phenotypic" selections therefrom; (3) generating offspring according to prescribed crossing design, and recording within- and among-family phenotypic and additive genetic effects; (4) for forwards selection only, estimating genetic parameters from simulated data for combined among- and within-family index selection; (5) conducting

the selection, according to the restriction of number of individuals from any one family, and calculating genetic gain. Details of the steps, for the various breeding options, are given in Appendix 1.

For each set of conditions, 200 replicate runs were made from the outset to give reliable mean responses to selection and to characterise the distributions about those means. "Empirical" percentile ranges were calculated for the mean breeding values of selections made in the replicate runs for each case.

Response variables considered for each set of conditions were: mean gain across runs (Mean); pooled standard deviation for additive genetic values among selections within runs (SD); and coefficient of variation among runs (CV) in respect of genetic gain from individual runs.

Supplementary conditions studied involved $h^2 = 0.05$ for Option 4, with $d = 0$, 1, and 2, with and without restrictions on numbers of individuals selected per family. Results are covered briefly, without detailed tabulation.

RESULTS
Mean Gains

Mean gains ("Mean") are summarised in Tables 1 and 2, for N (total offspring) = 9000 and 30 000 respectively. In these Tables the components of gain resulting from initial plus-tree selection are also shown for the respective sets of conditions. Values differed markedly among the options. Forwards selection in select pair-crosses gave consistently the highest values except for some situations at the highest ratio of dominance : additive genetic variance (d). Forwards selection within the select polymix generally gave higher gain than backwards selection at the two higher heritabilities, while forwards selection within base half-sibs did so only at the highest heritability and 15 or 30 selections.

While increasing n depressed Mean gain appreciably, it did not greatly affect the relative differences between options. Nevertheless, the effects of increasing n did vary among options and according to heritability. Increasing n from 15 to 120 had different effects on gain from different options (Table 3). The consequent percentage drop in gain with increasing n was consistently least for select pair-crosses, even with the variations in d, ($d > 0$ tending to cause a slight relative increase in the effect of n). Not surprisingly, the relative drop in gain with increasing n was consistently less for $N = 30 000$ than for $N = 9000$. The relative effect of n generally increased with heritability.

Similarly, the variations in d did not radically affect relative differences between options. However, the impact of $d > 0$ on Mean did not appear to be fully consistent, but this represented relatively small differentials between tabulated values that were based on numbers of runs that, despite being large, were still finite.

Taking the extreme situation (beyond the range of the main study), with $h^2 = 0.05$ and $d = 1$ and 2, Mean was consistently reduced by around 0.13 and 0.23 for the respective values of d (details not tabulated). This was relative to a range of Mean gain of 1.71 to 1.83 for $N = 9000$. Considering the impacts on Mean of the restriction on related selections, lifting the restriction increased gain by around 17–22% for $n \leq 60$, this increase dropping markedly to around 10% for $n = 120$ (details not shown).

TABLE 1–Comparisons of genetic-gain means among options for different h^2 (heritabilities), d (σ^2_D/σ^2_A) (where relevant), and n (selections made), for N (total offspring) = 9000. Boldface denotes values for forwards selection exceeding those from backwards selection. Gains from forwards and backwards selection are cumulative, including plus-tree gains**.**

h^2	Category	Forwards option	d	n			
				15	30	60	120
0.1	Plus-trees	–	–	0.59	0.59	0.59	0.59
	Backwards	–	–	1.90	–	–	–
	Forwards	Base half-sibs	–	1.54	1.47	1.38	1.31
		Select polymix*	–	1.79	1.69	1.64	1.56
		Select pair-cross	0	**2.27**	**2.23**	**2.18**	**2.11**
			0.3	**2.15**	**2.10**	**2.07**	**2.00**
			1.0	**2.06**	**2.02**	**1.96**	1.90
0.2	Plus-trees	–	–	0.84	0.84	0.84	0.84
	Backwards	–	–	2.32	–	–	–
	Forwards	Base half-sibs	–	1.98	1.89	1.79	1.69
		Select polymix*	–	**2.34**	2.23	2.13	2.02
		Select pair-cross	0	**2.72**	**2.67**	**2.58**	**2.49**
			0.3	**2.59**	**2.53**	**2.47**	**2.38**
			1.0	**2.50**	**2.44**	**2.35**	2.26
0.4	Plus-trees	–	–	1.18	1.18	1.18	1.18
	Backwards	–	–	2.81	–	–	–
	Forwards	Base half-sibs	–	2.65	2.51	2.38	2.23
		Select polymix*	–	**3.06**	**2.95**	**2.82**	2.67
		Select pair-cross	0	**3.31**	**3.21**	**3.12**	**3.00**
			0.3	**3.24**	**3.15**	**3.05**	**2.92**
0.6	Plus-trees	–	–	1.46	1.46	1.46	1.46
	Backwards	–	–	3.02	–	–	–
	Forwards	Base half-sibs	–	**3.06**	2.91	2.76	2.59
		Select polymix*	–	**3.60**	**3.45**	**3.30**	**3.14**
		Select pair-cross	0	**3.79**	**3.65**	**3.51**	**3.35**

* Values shown are for zero genetic dominance ($d = 0$). Increasing d even to the highest values gave only very marginal drops in gain, so details not shown here or in Tables 2 and 3 and the Figures.

Increasing N from 9000 to 30 000 predictably increased gains (Tables 1–3), but again did not greatly affect the comparisons among options.

The Mean gains were overall in close agreement with deterministic expectations (details not shown), noting that there was no exact straightforward formulation for the latter under the restrictions in selecting relatives.

Standard Deviations

Standard deviations (SDs) among individual breeding values of selections differed very little according to n (number selected), although this could be a fortuitous result of elements of non-normality resulting from the degree of truncation selection. Accordingly, results shown (Table 4) are means averaged across all four values of n for forwards selection. The

TABLE 2–Comparisons of genetic-gain means among options for different h^2 (heritabilities), d (σ^2_D/σ^2_A) (where relevant), and n (selections made), for N (total offspring) = 30 000. Boldface denotes values for forwards selection exceeding those from backwards selection. Gains from forwards and backwards selection are cumulative, including plus-trees gains.

h^2	Category	Forwards option	d	n			
				15	30	60	120
0.1	Plus-trees	–	–	0.59	0.59	0.59	0.59
	Backwards	–	–	2.24	–	–	–
	Forwards	Base half-sibs	–	1.82	1.75	1.68	1.61
		Select polymix	–	2.09	1.99	1.94	1.88
		Select pair-cross	0	**2.39**	**2.36**	**2.32**	**2.28**
			0.3	**2.39**	**2.36**	**2.31**	**2.26**
			1.0	2.20	2.18	2.13	2.08
0.2	Plus-trees	–	–	0.84	0.84	0.84	0.84
	Backwards	–	–	2.60	–	–	–
	Forwards	Base half-sibs	–	2.29	2.20	2.11	2.02
		Select polymix	–	2.58	2.52	2.45	2.36
		Select pair-cross	0	**2.88**	**2.83**	**2.77**	**2.70**
			0.3	**2.84**	**2.78**	**2.69**	**2.62**
			1.0	**2.64**	2.58	2.50	2.43
0.4	Plus-trees	–	–	1.18	1.18	1.18	1.18
	Backwards	–	–	2.96	–	–	–
	Forwards	Base half-sibs	–	2.90	2.79	2.68	2.55
		Select polymix	–	**3.35**	**3.24**	**3.13**	**3.01**
		Select pair-cross	0	**3.51**	**3.41**	**3.32**	**3.22**
			0.3	**3.38**	**3.29**	**3.19**	**3.09**
0.6	Plus-trees	–	–	1.46	1.46	1.46	1.46
	Backwards	–	–	3.14	–	–	–
	Forwards	Base half-sibs	–	**3.36**	**3.23**	3.09	2.93
		Select polymix	–	**3.84**	**3.71**	**3.58**	**3.44**
		Select pair-cross	0	**3.98**	**3.87**	**3.76**	**3.63**

SDs were the lowest throughout for backwards selection, clearly reflecting the more precise evaluation than in forwards selection. Among the forwards-selection options, base half-sibs gave the highest SDs followed closely by select polymix, with select pair-cross somewhat lower although this difference was reduced by increasing d. While SDs fell appreciably with increasing heritability, the relative differences among options varied little according to heritability. At the higher N (total offspring), SDs were marginally less, except for being appreciably less in relative terms for base half-sibs.

Coefficients of Variation

Coefficients of variation (CVs) among gains from individual runs are shown in Tables 5 and 6 for N = 9000 and 30 000 respectively.

Generally, CVs dropped with increasing n (selections made), h^2 (heritability), and d (ratio of dominance : additive genetic variance). There were, however, some minor

TABLE 3–Effects of increasing n (selections made) from 15 to 120 in terms of percentage decrease in mean gain from the forwards-selection options, for different h^2 (heritabilities), d (dominance : additive genetic variance), and N (total offspring).

h^2	Option	d	N	
			9000	30 000
0.1	Base half-sibs	–	15	12
	Select polymix	–	13	10
	Select pair-cross	0	7.0	4.6
		0.3	7.0	5.4
		1.0	7.8	5.5
0.2	Base half-sibs	–	15	12
	Select polymix	–	14	8.5
	Select pair-cross	0	8.4	6.3
		0.3	8.1	7.7
		1.0	9.6	8.0
0.4	Base half-sibs	–	16	12
	Select polymix	–	13	10
	Select pair-cross	0	9.4	8.3
		0.3	9.9	8.6
0.6	Base half-sibs	–	15	13
	Select polymix	–	13	10
	Select pair-cross	0	12	8.8

exceptions to these trends, mostly for $N = 9000$. For $N = 9000$ these exceptions included: increasing n beyond 30 for select pair-cross and $h^2 = 0.1$; increasing d for select pair-cross and $n < 60$; and backwards selection for $h^2 > 0.4$. For $N = 30 000$ these exceptions were confined to: $n > 60$ for select pair-cross, $h^2 = 0.4$, $d = 0.3$; and $h^2 = 0.6$, $d = 0$.

With N increased from 9000 to 30 000, the CVs were generally reduced, but there were a few minor exceptions for some sets of conditions with select polymix and select pair-cross.

Interestingly, the impacts of variations in n, h^2, and d varied among options such that there were essentially no across-the-board differences in CVs between options.

Percentile Distributions for Genetic Gain

Percentile plots of gains from individual runs are plotted in Fig. 1–3 for forwards selection and backwards selection, for only those sets of conditions where mean gain from forwards selection exceeded that from backwards selection in Fig. 1–3 for the respective forwards-selection options.

In the comparison of forwards and backwards selection, there were very limited crossovers in the percentile distributions (Fig. 1–3). The most notable exception was for select polymix with heritability of 0.2 (Fig. 2A), where mean gain differed minimally between the two options (Table 1).

TABLE 4–Comparisons of mean SD (standard deviations among breeding values of individual selections), across all four *n* values (numbers of selections) in forwards selection, for different h^2 (heritabilities), $d(\sigma^2_D/\sigma^2_A)$ (where relevant) for the various selection scenarios, with *N* (total offspring) = 9000 and 30 000.

h^2	Category	Forwards option	d	N 9000	N 30 000
0.1	Backwards	–	–	0.78	0.60
	Forwards	Base half-sibs	–	0.92	0.89
		Select polymix*	–	0.90	0.86
		Select pair-cross	0	0.78	0.75
			0.3	0.82	0.77
			1.0	0.83	0.80
0.2	Backwards	–	–	0.64	0.52
	Forwards	Base half-sibs	–	0.85	0.82
		Select polymix*	–	0.84	0.81
		Select pair-cross	0	0.72	0.72
			0.3	0.75	0.74
			1.0	0.78	0.78
0.4	Backwards	–	–	0.58	0.42
	Forwards	Base half-sibs	–	0.77	0.75
		Select polymix*	–	0.74	0.72
		Select pair-cross	0	0.67	0.66
			0.3	0.69	0.68
0.6	Backwards	–	–	0.46	0.38
	Forwards	Base half-sibs	–	0.66	0.64
		Select polymix*	–	0.63	0.61
		Select pair-cross	0	0.60	0.59

* Values shown are for zero genetic dominance (*d* = 0). Increasing *d* even to the highest values gave only very marginal increases in SD, so details are not shown here or in Tables 5 and 6 and Figures.
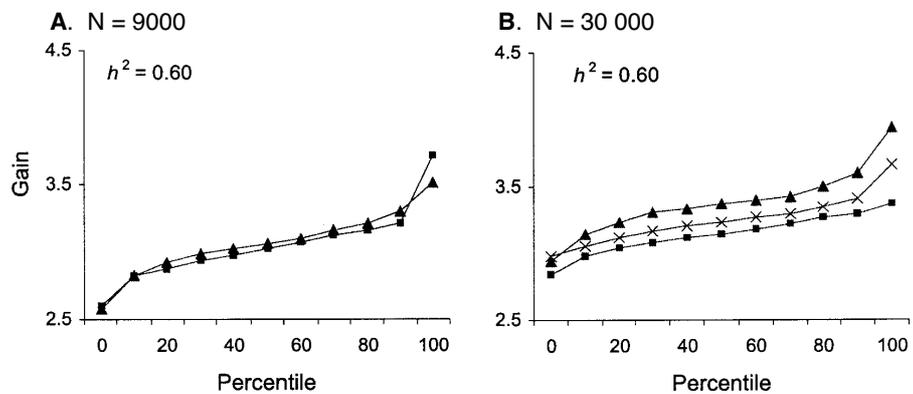

FIG. 1–Plots of percentile distributions (1%, 10%, 20% … 90%, 100%) for gain from (forwards selection in half-sib families) *vs* backwards selection, for *N* (total offspring) = 9000 and 30 000, for different *n* (selections made) in forwards selection, and h^2 (heritability) = 0.6. Backwards selection plotted in solid squares. Forwards: ▲, *n* = 15; ×, *n* = 30.

TABLE 5–Comparisons of coefficients of variation (%) of gain among individual simulation runs, for different h^2 (heritabilities), d (σ^2_D/σ^2_A) (where relevant), and n (numbers of selections), for the various selection scenarios, with N (total offspring) = 9000.

h^2	Category	Forwards option	d	n 15	n 30	n 60	n 120
0.1	Backwards	–	–	13.0	–	–	–
	Forwards	Base half-sibs	–	17.1	14.0	12.0	10.8
		Select polymix	–	16.0	12.8	10.3	9.5
		Select pair-cross	0	10.4	8.5	8.7	9.2
			0.3	13.8	12.5	11.9	11.8
			1.0	14.5	13.0	12.0	11.9
0.2	Backwards	–	–	7.6	–	–	–
	Forwards	Base half-sibs	–	12.2	9.7	7.3	5.7
		Select polymix	–	11.4	9.6	8.7	8.4
		Select pair-cross	0	8.1	7.3	6.5	6.2
			0.3	10.9	10.4	9.4	8.9
			1.0	10.2	9.7	9.2	9.2
0.4	Backwards	–	–	5.3	–	–	–
	Forwards	Base half-sibs	–	8.2	6.4	4.7	4.2
		Select polymix	–	7.4	6.6	6.1	5.6
		Select pair-cross	0	7.2	6.5	5.9	5.5
			0.3	6.6	6.1	5.6	5.2
0.6	Backwards	–	–	5.7	–	–	–
	Forwards	Base half-sibs	–	6.0	4.6	4.2	3.2
		Select polymix	–	5.7	4.4	4.0	3.7
		Select pair-cross	0	5.8	5.3	4.8	4.5

Similarly, in comparing percentile distributions of gain for different numbers of selections made within the three forwards-selection options, there were also very few crossovers. Almost all of them occurred towards the very lowest percentiles, such that the differences in mean gain were reflected mainly in divergences towards the top percentiles.

Varying total number of offspring (N) did not greatly change the picture, beyond what was already evident from comparing among coefficients of variation for gain (Tables 5 and 6).

DISCUSSION

We review separately the issues related mainly to the construction and validity of the model and those related mainly to practical applications, noting that there is some overlap between these categories.

Model-related Issues

The numbers of 300 plus-trees and the values of N (total offspring) and n (genotypes selected) were chosen to simulate realistic breeding scenarios, N = 30 000 representing little concession to manageability for simulation. However, N = 30 000 would be appropriate

TABLE 6–Comparisons of coefficients of variation (%) of gain among individual simulation runs, for different h^2 (heritabilities), d (σ^2_D/σ^2_A) (where relevant), and n (numbers of selections), for the various selection scenarios, with N (total offspring) = 30 000.

h^2	Category	Forwards option	d	n			
				15	30	60	120
0.1	Backwards	–	–	8.6	–	–	–
	Forwards	Base half-sibs	–	13.4	10.8	8.4	7.8
		Select polymix	–	12.7	11.0	10.3	9.9
		Select pair-cross	0	9.5	8.9	8.8	8.0
			0.3	10.7	9.7	8.8	8.3
			1.0	11.5	9.5	9.6	9.5
0.2	Backwards	–	–	6.8	–	–	–
	Forwards	Base half-sibs	–	11.1	8.3	6.4	5.5
		Select polymix	–	8.6	7.6	7.0	6.8
		Select pair-cross	0	8.0	7.2	6.6	6.5
			0.3	9.4	8.8	7.8	7.6
			1.0	10.0	8.8	9.2	9.1
0.4	Backwards	–	–	4.5	–	–	–
	Forwards	Base half-sibs	–	7.2	5.4	4.5	4.0
		Select polymix	–	7.7	6.2	5.2	4.9
		Select pair-cross	0	6.2	5.6	4.5	4.6
			0.3	6.0	5.5	5.2	4.9
0.6	Backwards	–	–	3.9	–	–	–
	Forwards	Base half-sibs	–	5.2	4.0	3.1	2.7
		Select polymix	–	5.3	4.7	4.2	3.9
		Select pair-cross	0	4.9	4.1	3.9	4.0

for replicating trials across sites, which would be indicated where genotype × site interaction is significant. With marked interaction, backwards selection would compare more favourably than has been indicated here.

Comparisons made between gains from backwards selection and those from forwards selection within base half-sibs (Fig. 1) tended to exaggerate the overlap between their respective gains. This is because the gains were explored through independent runs, giving independent sampling errors. In fact, where the choice arises between these options, it will involve material that is subject to common sampling errors for plus-tree selection, which will tend to reduce actual overlap. The same would apply to comparing gain from backwards selection from select polycross (not explicitly addressed) with gain from forwards selection within the same.

Comparing select pair-cross with forwards selection in base half-sibs or in select polycross, the arbitrary difference in restriction on numbers selected per family, which was based on the fact that pollen parents were known in select pair-cross, probably operated slightly in favour of it for gain.

The comparisons explored here have not addressed epistatic gene effects. If epistasis is important, there could be some loss of the apparent genetic gain from backwards selection
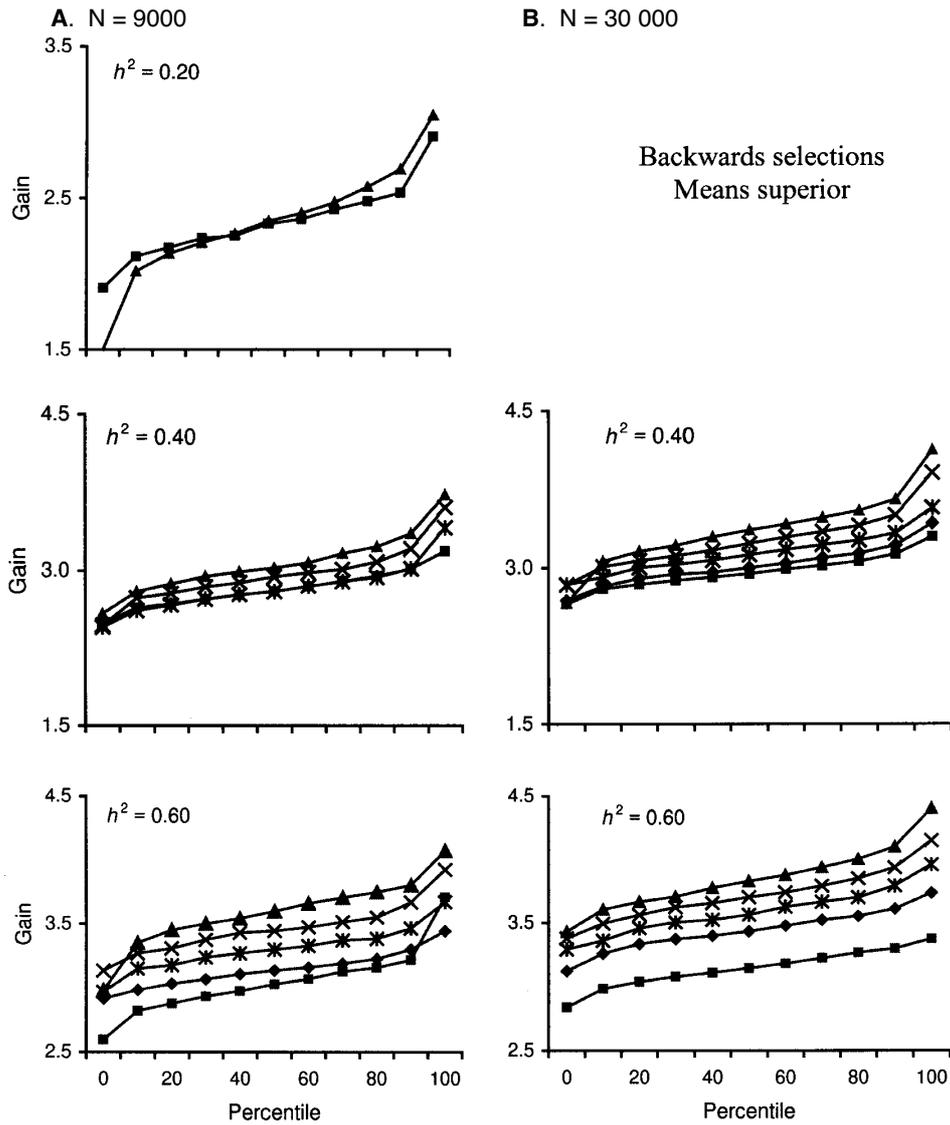
A. N = 9000

B. N = 30 000

Backwards selections
Means superior

FIG. 2–Plots of percentile distributions (1%, 10%, 20% ... 90%, 100%) for gain from forwards
selection in select-pollen polycrosses *vs* backwards selection, for *N* (total offspring) = 9000
and 30 000, for different *n* (selections made) in forwards selection, and h^2 (heritability) = 0.1,
0.2, 0.4, and 0.6. Backwards selection plotted in solid squares. Forwards: ▲, *n* = 15; ×, *n* = 30;
∗, *n* = 60; ♦, *n* = 120.

that is indicated by general combining abilities, in producing seedling offspring. Less clear
would be the impact of epistasis on gains achieved from forwards selection. However, if
superior general combining abilities contain an element of minority advantage (relating to
allele-frequency effects combined with dominance), then the expected gains would
certainly be inflated, especially those from backwards selection. In any event, an additivity
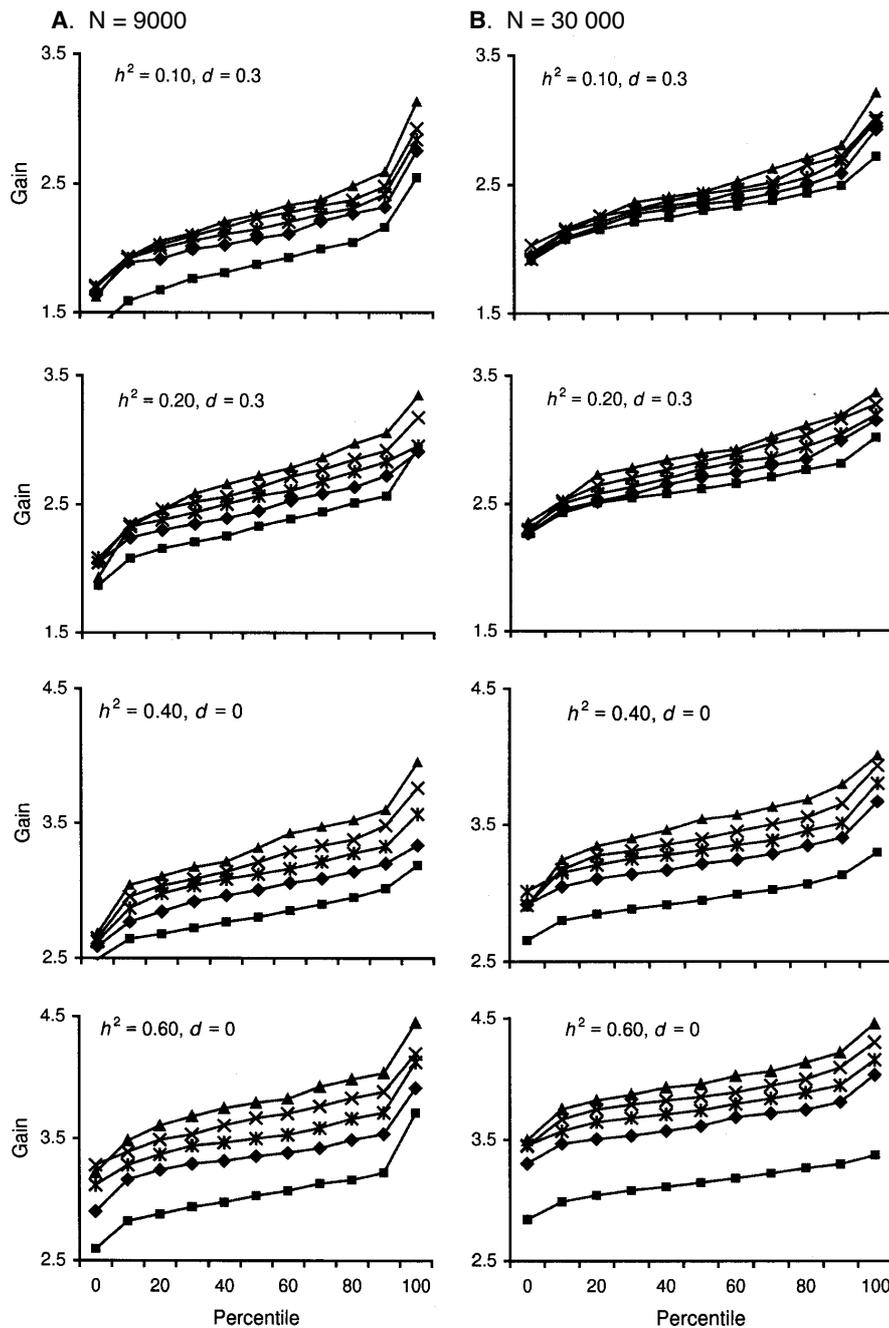
FIG. 3–Plots of percentile distributions (1%, 10%, 20% ... 90%, 100%) for gain from forwards selection in pair-crosses (d (σ^2_D/σ^2_A) = 0 or 0.3, as deemed appropriate) vs backwards selection, for N (total offspring) = 9000 and 30 000, for different n (selections made) in forwards selection, and h^2 (heritability) = 0.1, 0.2, 0.4, and 0.6. Backwards selection plotted in solid squares. Forwards: ▲, n = 15; ×, n = 30; *, n = 60; ♦, n = 120.

plus dominance genetic model would be an oversimplification, not that the mis-specification of the non-additive genetic component of the model necessarily has any serious affect on the model's power for predicting genetic gain. The sampling variation of gains for select pair-cross, as illustrated in Fig. 3, will be under-estimated because it was based on $d = 0$.

The combinations of heritability (h^2) and ratios of dominance : additive genetic variance (d) chosen for Fig. 3, were among those deemed to be most likely, based on commonsense and the results of Carson (1986) and Cotterill *et al.* (1987).

Simple truncation selection, whereby offspring of all selections would be equally represented in the resulting plantation crops, was assumed for all selection, subject to the restrictions on selection from individual families in forwards selection. However, it would be possible to combine all the schemes considered with differential representation of selections in a seed orchard, with the representation of individuals decreasing as one moves down the order of family ranking (Lindgren & Matheson 1986). This can give a better combination of genetic gain and breadth of genetic base (Lindgren *et al.* 1989, 1993; Lindgren & Wei 1994). Thus both backwards and forwards selection, as simulated here, would have had a degree of sub-optimality, although the latter would have helped to ensure greater genetic diversity at higher *n*. A comprehensive optimisation, however, was not the goal of this study; instead, the goal was a fair comparison between forwards and backwards selection, which we consider to have been achieved.

The use of internally generated estimates of genetic parameters for making forwards selection was a conservative procedure, since it incurred an element of exposure to effects of estimation errors that could possibly be reduced by using prior information on the parameters. However, this exposure is likely to be limited — partly because the population sizes and configurations (300 plus-trees, and $N = 9000$ or 30 000) are conducive to good estimates of the parameters, especially at higher heritabilities; and partly because the sensitivity of gain to the consequent weighting of family information is likely to be low, given the absence of any negative intercorrelation among orthogonal effects. Further protection against errors of estimating genetic parameters is likely to be afforded by the restriction on number of selections made in any one family. There will be some downward bias in estimation of heritabilities (Appendix Steps FH4, FP6, and FR6), through not adjusting for truncation of parental variance through initial plus-tree selection. However, this bias is likely to be minor, given the low heritabilities assumed in the base population, and any consequent over-emphasis on among-family selection would tend to be countered by the restrictions on selections per family.

For forwards selection we have considered dependence on information on just two genetic parameters, heritability and to a lesser extent *d*. The errors of estimating such parameters appearing unlikely to be very important. However, if multi-trait selection is involved, with adverse between-trait genetic correlations, or with genetic correlations between traits of contrasting rankings for economic worth and heritability, errors of estimating genetic parameters could be more critical.

Increasing base-population heritability to equal that in the progeny trials would increase the response to initial plus-tree selection by a factor of $\sqrt{2}$, making forwards selection within base half-sibs compare even less favourably with the other options. However, gain in the second phase of selection would be reduced through greater truncation of additive genetic variance among parents.

Application Issues

Relatedness among selections tends to increase theoretical inbreeding, with loss of expected gain in seed orchards from resulting inbreeding depression (Olsson *et al.* 2001). This may be mitigated by choice of orchard layout (e.g., Cannon & Low 1994), or even averted by production systems based on controlled crossing. Moreover, one may adjust expected gains for likely inbreeding depression (Lindgren & Mullin 1997), while actual inbreeding may well be less than theoretical values. Furthermore, with the restrictions on related selections even the theoretical inbreeding coefficients will be low, and actual levels of inbreeding would be reduced by any effective preferential outcrossing. That said, the select polymix option would be the most susceptible to inbreeding, and therefore subject to slightly greater "slippage" of genetic gain.

Our model assumed just two discrete generations, in the base population and the offspring of sections from it, which need not be imposed in practice. Indeed, there are ways in which elements of classical forwards and backwards selection can be used in conjunction. One way is to use a mix of the best forwards and backwards selections. Pointers for optimising such an approach have been given by Ruotsalainen & Lindgren (1998), although their simulations usually showed no great genetic-gain advantage over the single-method options.

For more advanced generations the simulation will be more complicated, but the underlying issue of how imprecise information on individuals may trade-off against any theoretical advantages of forwards selection should remain the same.

In comparing options, the timing of propagation and selection, which can depend strongly on the reproductive biology of the species, is potentially an issue. If open-pollinated (OP) families from the plus-trees are easily obtainable and approximate closely to half-sibs, gains from backwards selection may be captured more rapidly; this would arise through progenies being available sooner and selection being feasible at closer to a mature age, and therefore more accurate, at a given date than in control-pollinated offspring. However, if OP families do not approximate at all well to half-sibs, they may be very inefficient for backwards selection but much less so for forwards selection.

If OP families are essentially half-sibs, it may be easy to use them for backwards selection and controlled pair-crossing for forwards selection. However, if OP families depart substantially from half-sibs, then involving each parent in multiple pair-crosses may be indicated (Burdon & van Buijtenen 1990; van Buijtenen & Burdon 1990). Indeed, if in addition to OP families departing substantially from half-sibs, controlled crossing is difficult and heritability is high, backwards selection may not be worthwhile at all.

The greater genetic variability that may be expected among forwards than among backwards selections may cause concern. However, it is very unlikely to be a decisive consideration. The phenotypic variance of a crop produced by a seed orchard will generally be dominated by the sum of the within-family segregational variance and environmental variance, and very much by the latter if heritability is low. If heritability is high, forwards selection will give greatly superior gain, which will be accompanied by modest variation in breeding values of seed-orchard parents, although the segregational genetic variance will still arise.

The crossovers of percentile plots for gains from forwards selection with differing n were very limited, and confined mainly to the lowest percentiles. This is reassuring, since $n = 15$ was chosen to represent a safe minimum number for a production population, primarily in connection with risks of post-deployment failure of individual clones. This fits well with what is becoming a broad consensus (Burdon 2001), namely that little additional gain in risk spread accrues from deploying more than 20 unrelated clones, with such extra protection against risk having to be weighed up against progressive erosion of expected genetic gain. This congruence of figures is all the more noteworthy, given that much of the risk analysis associated with clonal forestry has addressed serious outcomes rather than relatively mild suboptimality of performance.

Possible concerns over the practicalities of using large numbers of clones in seed orchards may disappear in the light of the lack of any clear advantage in increasing effective numbers of orchards clones above about 15. Logistical issues may arise in providing for revising lists of seed-orchard clones, from forwards or backwards selection, as progeny trials get older and provide more reliable information, but they are beyond the scope of this paper.

CONCLUSIONS

The comparisons among scenarios for mean gain are in line with findings from past studies (as reviewed by Ruotsalainen & Lindgren 1998), while comparisons of standard deviations among selections and of among-run coefficients of variation in gain are not surprising. Making these comparisons, however, provided the platform for studying the percentile distributions for gain among replicate runs, which yielded the salient findings:

(1) Provided n (forwards selections made) ≥15, any expected superiority of forwards selection will almost always be realised, despite sampling errors associated with relatively imprecise estimation of individual breeding values in forwards selection;

(2) Additional protection from effects of selection errors through increasing n beyond about 15 in forwards selection will seldom outweigh the associated sacrifice in selection intensity.

ACKNOWLEDGMENTS

REFERENCES

BURDON, R.D. 2001: Genetic diversity and disease resistance: some considerations for research, breeding, and deployment. *Canadian Journal of Forest Research 31*: 596–606.

BURDON, R.D.; VAN BUIJTENEN, J.P. 1990: Expected efficiencies of alternative mating designs for reselection of parents. *Canadian Journal of Forest Research 20*: 1664–1771.

CANNON, P.G.; LOW, C.B. 1994: A computer-aided test layout for open-pollinated populations of insect-pollinated tree species. *Silvae Genetica 43*: 265–267.

CARSON, M.J. 1986: Control-pollinated seed orchards of the best general combiners — a new strategy for radiata pine improvement. Pp. 144–149 *in* Williams, T.S.; Wratt, G.S. (Ed.) "Proceedings, Plant Breeding Symposium DSIR 1986". *Agronomy Society of New Zealand Special Publication No. 5.*

COTTERILL, P.P.; DEAN, C.A.; VAN WYK, G. 1987: Additive and dominance genetic effects in *Pinus pinaster*, *P. radiata* and *P. elliottii* and some implications for breeding. *Silvae Genetica 36*: 221–232.

LINDGREN, D.; MATHESON, A.C. 1986: An algorithm for increasing the genetic quality of seed from seed orchards by using the better clones in higher proportions. *Silvae Genetica 15*: 173–177.

LINDGREN, D.; MULLIN, T.J. 1997: Balancing gain and relatedness in selection. *Silvae Genetica 46*: 124–129.

LINDGREN, D.; WEI, R.-P. 1994: Gain versus effective number. Pp. 164–177 *in* Lee, S.J. (Ed.) "Progeny Testing and Breeding Strategies", Proceedings, Nordic Group for Tree Breeding, Edinburgh, 6–10 October 1993, Forestry Authority, Edinburgh.

LINDGREN, D.; LIBBY, W.J.; BONDESSON, F.L. 1989: Development to plantations and proportions of clones with special emphasis on maximising gain at a constant diversity. *Theoretical and Applied Genetics 77*: 825–831.

LINDGREN, D.; WEI, R.-P.; BONDESSON, F.L. 1993: Optimal selection from families. *Heredity 70*: 619–621.

OLSSON, T.; LINDGREN, D.; LI, B. 2001: Balancing genetic gain and relatedness in seed orchards. *Silvae Genetica 50*: 222–227.

RUOTSALAINEN, S.; LINDGREN, D. 1998: Predicting genetic gain of forward and backward selection in forest tree breeding. *Silvae Genetica 47*: 42–50.

VAN BUIJTENEN, J.P.; BURDON, R.D. 1990: Expected efficiencies of mating designs for advanced-generation selection. *Canadian Journal of Forest Research 20*: 1648–1663.

APPENDIX 1

DETAILS OF SIMULATION STEPS

Initial general steps

(1) On the basis of the specified genetic parameters, <u>generate the initial base population</u> by drawing random samples of the component effects (note application of h^2_0 with corresponding $\sigma^2 p_0$), and summing them in each case to produce 30 000 phenotypic values.

(2) <u>Select the top 300 phenotypes</u>, recording their breeding values (mean ΔG_0).

Backwards selection using "base half-sib" families

B3. <u>Generate progeny-test population</u>**.** For each of the initial 300 selections: draw $^1/_2 A$, and add to it the stochastic sample of within-family effects, the latter being the sum of effects, $W_{AH,}$ W_{DH}, and E (which sum to W_H), of variances $^3/_4 \sigma^2 A$, σ^2_D, and σ^2_E respectively. Generate $N/300$ such offspring, and derive their mean. Save A values for individuals ($^1/_2 A + W_{AH}$) and the "phenotypic" effect for each family (F_H).

B4. <u>Select the top 15 families</u>**,** on F_H, and average the additive genetic values of the parents concerned to give ΔG_B, the gain from backwards selection for the run, and save.

Forwards selection within base half-sib families

FH4. <u>Estimation of genetic parameters.</u> From Step B3:

   (i)   Estimate among- and within-family variances (σ^2_{HS} and σ^2_{WH} respectively),

   (ii) Thence estimate heritabilities of family means (h^2_{FH}) and within-family effects (h^2_{WH}) according to $h^2_{FH} = (1 + {}^3/_n)\sigma^2_{HS}/(\sigma^2_{HS} + \sigma^2_{WH}/p)$ and $h^2_{WH} = 3\sigma^2_{HS}/\sigma^2_{WH}$, p being the number of individuals per family (= $N/300$).

FH5. <u>Selection of individual offspring.</u>

   (i)   For each individual offspring calculate a selection index (I), of the form $I = F_H.\hat{h}^2_{FH} + W_H.\hat{h}^2_{WH}$,

   (ii) On the basis of I, select the prescribed number (n) of offspring, with the restriction that no more than $n/10$ individuals (rounding upwards to an integer as required) come from any one family.

   (iii) For each simulation run, average A for the individuals selected to give ΔG_{FH}, the gain, and save.

Forwards selection within select-pollen polycrosses ("select polymix")

FP3. <u>Choice of pollen parents.</u> Select top 30 phenotypes among the initial plus-trees, record additive genetic values (A_{SPPm}) of each.

FP4. <u>Generate offspring.</u> For each seed parent, generate $N/300$ individual offspring as follows:

   (i)   Draw offspring at random from the pool of 30 parents, each with the following component effects:

   • $^1/_2 A_f$, i.e., the seed-parent contribution

- $\frac{1}{2}A_{SPPm}+D_{fm}$, i.e., the pollen-parent plus specific combining ability contribution (σ^2_{Dfm} which equals $\frac{1}{4}\sigma^2_D$)
  - within-[full-sib]family effects, W_{AP}, W_{DP}, and E effects drawn from variances of $\frac{1}{2}\sigma^2_A$, $\frac{3}{4}\sigma^2_D$, and σ^2_E, and sum the same plus $\frac{1}{2}A_{SPPm}$ and D_{fm} to give $W_{P'}$

(ii) Sum $\frac{1}{2}A_f$ from the seed parent, $\frac{1}{2}A_{SPPm}$ from the pollen parent, and W_{AP}, to give additive genetic values of individuals, and save

(iii) Sum $\frac{1}{2}A_f$, and W_P to give the phenotypic value, and save

(iv) For each seed parent, obtain mean of phenotypic values of offspring

(v) Use same to derive family- and within-family effects (F_P and W_P respectively), and save.

FP5. Estimation of genetic parameters. From the Step FP4 data: as for Step FH4, except that the respective heritabilities are denoted h^2_{FP} and h^2_{WP} respectively (there will be slight upward bias in estimatjng h^2_{WP} and, to a lesser extent, h^2_{FP}).

FP6. Selection of individuals.

(i) For each individual calculate a selection index (I) of the form $I = F_P.\hat{h}^2_{FP} + W_P.\hat{h}^2_{WP}$

(ii) Select n of top-ranked individuals for I, with no more than $n/10$ individuals to come from any one family

(iii) Average A for the selected individuals to give ΔG_{FP}, the gain from forwards selection for the run, and save.

Forwards selection within random pair-crosses ("select pair-cross")

FR3. Generate pair-crosses. Allocate the 300 first-generation to 150 random pairs.

FR4. Generate offspring from the pair-crosses. For each of the 150 pair-crosses:

(i) Average the additive genotypic values of the parents, to give A_F, and add to each an effect D_F of variance $\frac{1}{4}\sigma^2_D$, to give G_R, a family genotypic effect

(ii) Generate the family, by adding to G_R, $N/150$ individual values which are the sums of effects, A_{WR}, D_{WR}, and E (summing to W_R), of variances $\frac{1}{2}\sigma^2_A$, $\frac{3}{4}\sigma^2_D$, and σ^2_E respectively

(iii) Obtain the phenotypic effect for the family (F_R) as $G_R + \sum W_R/(2p)$, and the individual within-family phenotypic effects (W_R)

(iv) Derive the individual additive genetic values ($A = A_{FR} + A_{WR}$).

FR3. Estimation of genetic parameters. From the pair-cross data:

(i) Estimate "empirically", from the population generated in Step FR2, among- and within-family variances (σ^2_{GR} and σ^2_{WR} respectively),

(ii) Estimate heritabilities of family means (h^2_F) and within-family effects (h^2_W), according to the ratios $\hat{h}^2_F = (1 + 1/p)\sigma^2_{GR}/(\sigma^2_{GR} + \sigma^2_{WR}/(2p))$, and $\hat{h}^2_W = \sigma^2_{GR}/(\sigma^2_{WR})$, respectively.

FR4. <u>Selection of individuals.</u>

(i) For each individual construct a selection index (I) of the form $I = F_R.\hat{h}^2{}_{FR} + W_R.\hat{h}^2{}_W$

(ii) Select the prescribed number (n) of top-ranked individuals for I, with the restrictions that no more than $n/5$ individuals come from any one family

(iii) Average A for the selected individuals to give ΔG_{FR}, the gain from backwards selection for the run, and save.