MATING DESIGN CONSIDERATIONS - HOW MANY CROSSES DO WE REALLY NEED TO TEST?

Randy Johnson[1]

Abstract:--The impact of increasing the number of crosses per parent on the efficiency of backwards selection was examined using Monte Carlo simulation. Both the efficiency of reselection and its associated variance leveled off after two to three crosses per parent.

Because so few crosses appear to be needed to estimate parental GCA values, a quasi-complementary mating design option was investigated where only a subset of the full-sib families were planted into replicated field trials, while the remainder were established in less expensive pure-family full-sib blocks. Breeding values for families planted only in blocks were generated from the family data available from the replicated trials. Compared to using only the field tested families, significant additional gains could be made through the increased selection intensity resulting from the additional families planted only in blocks.

Keywords: Complementary mating design, GCA test, selection, testing

INTRODUCTION

Many breeding programs in the Pacific Northwest are beginning to establish second-generation progeny tests. These second-generation programs have tended to consolidate numerous first-generation programs in an effort to reduce costs. It was recognized that many of the 126 separate first-generation programs in western Oregon and Washington may have been too conservative with respect to ensuring adapted planting stock. Another way to reduce costs is to utilize more efficient methods of testing. Two questions which arise when discussing testing methods are: 1) How many times must a parent be crossed in order to obtain good general combining ability (GCA) information for seed orchard roguing (backwards selection)? and, 2) How can gains from among-family selection (forwards selection) be increased with limited resources?

Typically, Northwest Tree Improvement Cooperative (NWTIC) second-generation breeding population sizes range from 250 to 300 selections, which are crossed twice to produce 250 to 300 full-sib families. These selections have come from 1,500 to 3,000 first-generation half-sib families. This has resulted in very high family selection intensities (e.g. 300 out of 3,000). Next generation we will have substantially less family selection intensity, something along the lines of the best 120 out of 300. This will result in a significant drop in gain per generation because gains from family selection will decrease (family selection is a large source of gain due to its much higher heritability than within-family heritability).

There is some concern in the Pacific Northwest that the ability to rogue seed orchards will be jeopardized with only two crosses per parent. Typically, forest tree breeding programs cross a selection three or more times or use a separate GCA test to establish the breeding values of parents. However, depending on the efficiency of reselection desired, the literature suggests that either two or three crosses for a parent may be sufficient to estimate its breeding value (Burdon and van Buijtenen 1990, Johnson 1998a, Lindgren 1977).

With so few crosses per parent needed to estimate parental GCA values, one should be able to predict breeding values of full-sib families planted only in unreplicated block plantings using data from related full-sib families in replicated field tests (i.e., using other full-sib families in which the parents are used). A quasi-

---

complementary mating design option is being considered in the Pacific Northwest. Instead of using a separate GCA test, typical of most complementary designs, a subset of full-sib families will be planted in replicated field trials (the "GCA" test), while the remaining full-sib families will only be planted in less expensive pure-family blocks.

The objectives of this paper are to review how reselection efficiency (backwards selection) is affected by the number of times a parent is crossed and to examine the increased gains from a quasi-complementary design where family selection intensity is increased by using additional full-sib families planted in pure-family blocks ("untested" families).

METHODS

Backwards selection efficiency

Monte Carlo simulations were used to examine correlations between estimated breeding values and actual genetic values of clones in a seed orchard. Gains from roguing an orchard 50% were calculated by examining the genetic value of the orchard before and after roguing. A detailed description of the procedures and results are found in Johnson (1998a). The process used SAS software (SAS 1990) to generate independent normal distributions for genotypic values and environmental deviations. Phenotypes of individuals and family means were constructed by summing the genotypic values and environmental deviations (phenotype = genotype + environment). The process allows one to estimate breeding values using phenotypes (family means) and then correlate the estimated breeding values with the true genotypes. The basic genetic model assumed additive (GCA) and dominance (SCA) variation, but no interaction (epistatic) components of genetic variation. The base-line variance components represented narrow-sense heritabilities of 0.25 on a single site and 0.19 across sites. Dominance variance was set to 35% of the additive variance.

The base model assumed that 24 seed-orchard clones were selected from a first-generation program of 300 half-sib families. Twelve clones were randomly assigned to each of two sets. Crossing was limited to within sets and used a partial diallel mating design. GCA values of the parents were estimated with best linear prediction (BLP) solutions using the family means of the partial-diallel progeny tests. Afterwards, the worst 12 clones, based on their estimated GCA values, were rogued from the orchard without respect to breeding group. Estimated breeding values were correlated with the actual breeding values of the parents and the actual genetic gain from roguing was noted. This was repeated 200 times and used to generate means and standard deviations of the 200 correlations and gain estimates.

Quasi-complementary mating design for forwards family selection

The genetic model used to evaluate the use of additional "untested" full-sib families in a quasi-complementary mating design assumed a single trait with additive and dominance variation and GxE. The test design examined was typical of second-generation NWTIC programs: six sites with 20 trees per family tested in a single-tree plot design. Additive variance was set at 24, dominance at 8, additive-by-site at 9, dominance-by-site at 3 and environmental at 76. This resulted in a narrow sense heritability of 0.20 and a heritability of family means of 0.77.

Estimates of gain from family selection were calculated with index selection formulae. This method assumes that a family's breeding value is estimated using a selection index. For example, for a series of 2x2 factorials, the following steps were taken.

Crossing design:

Parents	A	B	Parents	E	F
C	1	2	G	5	6
D	3	4	H	7	8

To calculate the appropriate index weightings for family 1 (AxC), the following matrices were constructed.

P=variance-covariance matrix of family means:

	Family 1	Family 2	Family 3	Family 4
Family 1	$\sigma^2_{family\ means}$	$\frac{1}{4}\sigma^2_a$	$\frac{1}{4}\sigma^2_a$	0
Family 2	$\frac{1}{4}\sigma^2_a$	$\sigma^2_{family\ means}$	0	$\frac{1}{4}\sigma^2_a$
Family 3	$\frac{1}{4}\sigma^2_a$	0	$\sigma^2_{family\ means}$	$\frac{1}{4}\sigma^2_a$
Family 4	0	$\frac{1}{4}\sigma^2_a$	$\frac{1}{4}\sigma^2_a$	$\sigma^2_{family\ means}$

G = genetic covariance of family 1 with the families in the P matrix:

	Family 1
Family 1	$\frac{1}{2}\sigma^2_a$
Family 2	$\frac{1}{4}\sigma^2_a$
Family 3	$\frac{1}{4}\sigma^2_a$
Family 4	0

From these two equations the index weights were calculated as: $b = P^{-1}{}'G$ and the gain from family selection was: i b'Pb / (b'Pb)$^{1/4}$ = i (bTb)$^{1/2}$, where i is the selection intensity and (b^{s}Pb)^v is the standard deviation of the index (Smith 1936, Hazel 1943, see Lin 1978 for discussion).

Predicting gain from a family which is **not** being field tested follows a similar procedure. For the case of parent A crossed with parent E, families from both 2x2 factorials were used in the P_2 matrix (an 8x8 matrix) and G_2 array (8x1) as follows:

This process is analogous to obtaining the BLP solution simultaneously for all the families in both test types. Gains from tested and "untested" families were estimated per unit of selection intensity (i) using the equation above.

$$P_2 = \left| \begin{array}{c|c} P & 0 \\ \hline 0 & P \end{array} \right| \quad G_2' = [\tfrac{1}{4}\sigma^2_a,\ 0,\ \tfrac{1}{4}\sigma^2_a,\ 0,\ \tfrac{1}{4}\sigma^2_a,\ 0,\ \tfrac{1}{4}\sigma^2_a,\ 0]$$

To examine the merits of utilizing "untested" families, I first assumed that 100 parents were crossed in one of five different mating designs (single-pair, 2x2 factorial, and 3-, 4- and 5-parent diallel), with all such families being field tested. The best 25 families from the crossing design were assumed to be selected in the following generation. Gain estimates from family selection were calculated using the formula above.

To approximate gain from using additional "untested" families, I determined at what selection intensities gains from the tested and "untested" family groups were equal and kept the total number of families selected at 25. For example: using a 2x2 factorial, the gain from choosing the best 25 out of 100 families (i=1.259) without any additional "untested" families is: i (b'Pb)$^{1/2}$ $_{tested}$ = 1.259 x 3.131 = 3.942 units. If we have an

additional 100 "untested" families, gains from the two groups are approximately equal when 16 out of 100 are selected from the tested group and 9 out of 100 are selected from the "untested" group. Gain from the best 16/100 tested families (i=1.504) is 1.504 x 3.131 = 4.709 units and the gain from the best 9/100 "untested" families *(1=1.777)* is: i (**b'Pb**) $_{untested =}$ 1.777 x 2.683 = 4.768 units. The increased gain over only using the tested families was approximated as (4.709 - 3.942) / 3.942 x 100 = 19.5%.

In reality a breeder would simply select among all the available families using their index values. When data are standardized, such that test means equal zero, the index values represent gains. This can be shown as follows:

$$\text{gain} = i \, (\text{b`Pb})^{\frac{1}{2}}$$

$$i \text{ for a particular family} = \frac{(\text{family index value - mean index value})}{(\text{standard deviation of the index})}$$

$$\text{mean index value} = 0$$

therefore, gain expected from selecting a family =

$$\frac{(\text{family index value - 0})}{(\text{b`Pb})^{\frac{1}{2}}} * (\text{b`Pb})^{\frac{1}{2}} = \text{family index value}$$

This process assumes that gain from within-family selection would be the same for both the field tested families and the families planted in pure-family blocks. Estimated gain from forward selections comes from both the family and within-family components. If the within-family gains were less from the families established in pure-family blocks, then much fewer selections would come from them than what is assumed here.

A crude financial analysis was evaluated for operational NWTIC Douglas-fir breeding programs in the Pacific Northwest. These programs are designed to cross each parent twice in a design very similar to a 2x2 factorial. Approximately 300 advanced-generation parents represent the breeding population for these programs. The resulting 300 families will be tested on 6 to 9 sites throughout each testing zone. Family selection in the following generation is estimated to be on the order of 120 out of 300. Assuming six test sites and using our estimates of additive, dominance, and environmental variation; estimated gains from family selection and within-family selection were approximately equal. Therefore, gain levels in growth for the mainline breeding program for this simulation were set at 5% for both family and within-family selection. This would result in a 10% gain from the mainline program of 300 families. Any increase in gain from using "untested" families only affected the 5% gain from family selection. Estimated discounted costs for a typical program with 6 sites and using a discount rate of 3.5% was $800,000 (David Todd, pers. comm.). The additional costs of breeding, establishing and maintaining a family block was estimated to be $370.

RESULTS

Backwards Selection Efficiency

The added efficiency of making more crosses per parent dropped markedly after only two crosses (Table 1). In addition to examining average gain, the odds of achieving that level of gain should also be examined. Risk is a function of the variation of the estimate (standard deviation and coefficient of variation in Table 1). The coefficient of variation of the correlation coefficients stabilized after three crosses per parent and

after two crosses for the percent increase in orchard gain (Table 1). The trends for the percentiles were similar to the trends for the means (Johnson 1998a).

Table 1. Correlations between estimated breeding values and actual genetic values and estimates of gain from roguing. Means, standard deviations, coefficient of variation, and minimums are reported.

	Crosses per parent					
Variable	1	2	3	4	5	6
Correlation coefficient						
Mean	0.655	0.849	0.905	0.917	0.924	0.926
S.D.	0.129	0.067	0.051	0.048	0.046	0.046
C.V.	0.179	0.079	0.056	0.052	0.049	0.050
Min	0.276	0.573	0.716	0.735	0.741	0.727
% increase in orchard gain from orchard roguing						
Mean	28.2	35.7	37.9	38.3	38.4	38.6
S.D.	8.69	8.65	8.68	8.50	8.75	8.59
C.V.	0.31	0.24	0.23	0.22	0.23	0.22
Min	6.7	8.7	13.3	14.5	8.9	17.2

It would therefore appear that three crosses per parent would be sufficient to determine a parent's breeding value. The correlation between the estimated breeding value and the actual genetic value averaged 0.905 for three crosses per parent (Table 1); with six crosses the correlation only increased to 0.926. Two crosses per parent was more efficient than some may think; the average correlation among estimated breeding value and actual genetic value was 0.849 for the example here. Increasing the dominance variance did not have major effects on the results (Johnson 1998a).

Forwards Family Selection Using a Quasi-Complementary Design

The relative efficiency for "untested" vs. tested families ranged from 71% to 99% (Table 2). The gain per unit of selection efficiency for a single-pair mating design was surprisingly efficient (71%) at choosing families not field tested (Table 2). The efficiency of selecting "untested" families is approximately 98% of that for selecting tested families after a parent is field tested in 3 crosses (4-parent diallel). However, more gain will always occur with tested families than with "untested" families. For example, adding 50 "untested" families to a 4-parent diallel increased gain per unit i from 4.715 to 5.074 (Table 3), an increase of 7.6%. Increasing the number of tested families increased gain per unit i from 4.715 to 5.288 (4-parent vs. 5-parent diallel in Table 3), an increase of 12.2%.

A thorough economic analysis is needed before one can decide whether planting additional family blocks is warranted. This requires estimates of predicted gains, values of those gains, and appropriate discount rates. These vary by organization so I chose to compare marginal gains from planting family blocks with total gain from the mainline program for typical NWTIC breeding programs given the previously stated assumptions.

Table 2. Estimated gains from family selection per unit of selection intensity (i), i.e. $(b'Pb)^{1/2}$.

Crossing Design	Crosses per parent	Families generated	Gains from tested families	Gains from "untested" families	Relative efficiency ("untested" / tested)
Single-pair	1	50	3.032	2.144	0.71
2x2 factorial	2	100	3.131	2.683	0.86
3-parent diallel	2	100	3.070	2.905	0.95
4-parent diallel	3	150	3.169	3.102	0.98
5-parent diallel	4	200	3.232	3.195	0.99

Table 3. Estimated gain from family selection for five mating designs, with and without utilizing "untested" pure family blocks. All options reflect selecting the best 25 families the following generation from a population generated by 100 parents.

Cross Type	Number of additional "untested" families			
	0	50	100	150
Single-pair	2.386	3.214	3.581	3.854
2x2 Factorial	3.942	4.340	4.709	4.919
3-parent diallel	3.865	4.427	4.823	5.050
4-parent diallel	4.715	5.074	5.413	5.609
5-parent diallel	5.288	5.559	5.785	6.047

Examination of the marginal costs and gains (benefits) showed that the cost/benefit ratio of establishing an additional 200 full-sib families in pure-family blocks was less than that of the mainline breeding program (Table 4). The mainline program resulted in cost/benefit ratio of $80k per unit of gain ($800k110%), while the cost/benefit ratio of planting 200 pure-family blocks was $57.3k per unit gain ($74k / 1.29% gain). The cost/benefit ratio of increasing the number of pure-family blocks from 150 to 200 was smaller than the mainline program ($72.8k / 1% gain vs. $80k / 1% gain in Table 4). The cost/benefit ratio of increasing the number families in pure-family blocks from 200 to 250 was larger than that of the mainline population ($89.2k / 1% gain vs. $80k / 1% gain). The increase in gain from utilizing pure-family blocks came without increasing the benefit of within-family selection; a source of gain which was credited to the mainline program.

DISCUSSION

In the Pacific Northwest we are still using our operational progeny tests to help delineate proper breeding/deployment zones. We are testing families over a diverse array of environments; thus increasing our costs to test individual families. This has resulted in testing fewer families than in the previous generation. The need to use common families over the region has also limited our ability to use true complementary mating designs. Differing testers in each testing zone would prohibit common families from being tested over wide areas.

family blocks. The gain estimates presented here assumed that gain from within-family selection was the same for the tested and "untested" groups of families.

Estimated gain from family selection using a true complementary design with a separate GCA test on 6 sites was estimated to be 3.151 per unit i. This was similar to the gains from testing full-sib families (Table 2). The drawback to a complementary design with its separate GCA test comes from the inability to use the field tested half-sib families for selection purposes; however, it is much simpler to estimate breeding values of family block plantings with the simple GCA tests. Depending on costs of putting in trials and family blocks, the quasi-complementary design may be a better value for the dollar. Assuming equal costs to put out either a 300 family GCA test or 300 full-sib families ($800k); the cost to put in a GCA test and 500 family blocks is $985k; compared to $874k for the quasi-complementary design with 300 tested and 200 "untested" families. Predicted gains (assuming 5% for family and 5% for within-family gains) were 11.77% for the complementary design and 11.29% for the quasi-complementary design. Total gain was larger for the complementary design, but cost per unit gain was less for the quasi-complementary design ($77,414 vs. $83,687 per 1% gain). Programs may wish to consider this design in lieu of traditional complementary mating designs. As with any decision, an organization must examine the type of testing design and the number of family blocks to plant on the basis of their estimates of gain expectation, discount rates, and acreage over which the gains will apply.

ACKNOWLEDGMENTS

LITERATURE CITED

Burdon, R.D. and J.P. van Buijtenen. 1990. Expected efficiencies of mating designs for reselection of parents. Can. J. For. Res. 20: 1664-1671.

Cotterill, P.P. and N. Jackson. 1989. Gains expected from clonal orchards under alternate breeding strategies. For. Sci. 35:183-196.

Hazel, L.N. 1943. The genetic basis for construction of selection indexes. Genetics 28:476-490.

Johnson, G.R. 1998a. Parental GCA Testing - How many crosses per parent? Can. J. For. Res. 28: 540-545.

Johnson, R. 1998b. Breeding design considerations for coastal Douglas-fir. Gen. Tech. Rep. PNW-GTR-411. Portland, OR: USDA-FS, Pacific Northwest Research Station. 34 p.

Lin C.Y. 1978. Index selection for genetic improvement of quantitative characters. Theor. Appl. Genet. 52:49-56.

Lindgren, D. 1977. Genetic gain by progeny testing as a function of cost. *In* Third World Consultation on Forest Tree Breeding, Canberra, Australia. Food and Agriculture Organization and IUFRO Pap. 6.9, pp. 1223-1235.

SAS Institute Inc. 1990. SAS/STAT User's Guide, Version 6, 4th Edition. Cary, NC: SAS Institute Inc.

Smith, H.G. 1936. A discriminant function for plant selection. Ann. Eugen. 7:240-250.