Forwards vs. Backwards Selection for Seed Orchards and Cooperative Second-Generation Breeding in the US Pacific Northwest

Randy Johnson and Keith Jayawickrama

Executive Summary

Gain from various orchard strategies were modeled. The scenario tested 2,000 first-generation open-pollinated families, from which orchards of 20 selections were formed, using either parents, progeny or both. This was followed by a second-generation breeding population in which 200 full-sib families were tested followed by a second-generation orchard of 20 selections.

The results showed that a 1.5 generation seed orchard (recruit from many first-generation open-pollinated testing programs with lots of parents from which to choose) using parents would give more gain than all-progeny orchards and is essentially equal to the gain from selecting both progeny and parents. However, the situation was changed in the second cycle; in many cases progeny will have the highest expected gain. Gains from a second cycle of breeding and testing 200 full-sib families (and choosing the best 20 parents or progeny) approached gains from testing 2,000 open-pollinated families and selecting the top individuals from the top 200 families. This is reassuring given that the second cycle will cost only around 10% of the first cycle, if costs per planted tree remain constant.

There appeared to be good justification for selecting based on age-6 data, establishing an orchard, and roguing based on age-12 data rather than waiting for age-12 data to begin building the orchard.

Introduction

The question has arisen whether to use "backwards" selections (i.e. tested parents), "forwards" selections (in this case, mainly untested progeny from open-pollinated tests), or a combination of parents and progeny, both in seed orchards and in breeding. Those are the three main options currently available for cooperative breeding programs and seed orchards in the US Pacific Northwest. Which option will give the most gain, and which is the most reliable? The questions are important, since several 1.5 generation orchards are being established and we are in a period of intense activity for second-generation breeding and testing. They have been explored in some publications (e.g. Burdon and Kumar in press, Hodge 1985 and 1997, Hodge and White 1993, Ruotslainen and Lindgren 1998) ; however, a study focusing on the situation in the US PNW is also worthwhile. We therefore tried to shed some light on this issue for the benefit of NWTIC members.

When using parents, we have a good idea of what to expect because we have already tested their progeny. We also have good precision when selecting parents since across-site family mean heritabilities (h^2_f) tend to be between 0.6 and 0.8 for many cooperative first-generation test series. The expected gain from selecting families is a direct function of family mean heritability (expected genetic value = h^2_f x family mean). Assuming a family-mean heritability of 0.75, the correlation between the family mean and its genetic value is $r = \div h^2_f = h_f = \div 0.75 = 0.87$. If we deploy seed by family (i.e., seed orchard parent) and want to match families and sites, then it is always good to be sure we have the appropriate family.

We can also select progeny for seed orchards. In theory we increase gain every generation of breeding so the next generation of selections should be better than the last. A concern with forward

selections is that the breeding value of each selection has considerable variation associated with it (unlike parental selections). This is because selecting the best trees within families is relatively imprecise (within-family heritability = h_w^2 = 0.15-0.25). One solution is to select several trees from the best families (after testing) because at least one selection usually ranks highly. This is strictly a function of increasing the sample size (population size) to reduce the extra variation associated with forward selections. It is also possible that individual trees have been wrongly labeled and mapped - such errors usually have more serious consequences on forward selections than on selecting families. Mislabeling 1% of the trees thought to belong to a half-sib family will have little effect on the ranking of the parent, however, if we mistakenly use a mislabeled forward selection it could affect gain considerably. So, while identifying good families is fairly fool-proof, we aren't sure we have the best progeny selections. Another drawback to an orchard of forward selections is that you don't have data to match families to sites.

There is also a difference in selection efficiency between selecting individuals from open-pollinated families and full-sib families. In open-pollinated families we can pick the right female parent with good precision, but have no data on the male parent. Full-sib families allow us to choose the right female parent and male parent. In theory, the variation we select upon can be partitioned into three parts; additive genetic variation associated with the female parent (1/4), that associated with the male parent (1/4) and within-family variation (1/2). In open-pollinated trials we select with the efficiency of the family mean heritability (h_f^2) on only the female parent, and the within-family heritability (h_w^2) on the rest. With full-sib families we can select both the female parent and male parent with the efficiency of h_{fm}^2; therefore one would expect more gain from selecting in full-sib trials than open-pollinated trials.

Methods

Computer simulation allows you to generate populations by first making genotypes and then adding environmental variation. You then select on the phenotypes (family means or individual values), and see what happens to the genotype (the actual genetic gains). Our simulations also considered using an early assessment (we chose age-6) and a later assessment (age-12). Age-age correlations were estimated with age-5 and age-11 height data from the Nehalem series, where the age-age genetic correlation (r_a) was estimated to be 0.716 and the age-age environmental correlation (r_e) was estimated to be 0.37. Comparing that with Johnson et al. (1997), age-age genetic correlations reported were: 5-10 = 0.69, 7-15 = 0.85, 10-20 = 0.90. For the simulations we assumed an age 6 and 12 assessment with the following correlations: r_a = 0.72, r_e = 0.37. The baseline breeding programs modeled the case starting with 2,000 open-pollinated families; though we also briefly examined first-generation simulations with differing numbers of starting OP families. This number of 2,000 families is representative of the number of families from which a typical second-generation metacooperative was formed.

Building on the first-generation of 2,000 OP families, we then selected the top 200 families based on family means (a 10% selection intensity as in the BZERC strategy). From each of these 200 families we chose the best tree based on its phenotype to go to the second generation. These 200 selections were then crossed in a disconnected 2 x 2 factorial mating design (this results in the same number of crosses per parent as the pair-matings in the BZERC strategy, and is easier to simulate).

Different trial designs were modeled for each generation to mimic the differences in cooperative first- and second-generation trials. First-generation trials had eight progeny test sites with 12 trees per family per site; second-generation trials had six progeny test sites with 20 trees per family

per site. Genetic variation was partitioned such that narrow sense heritability at a site was 0.25 and the type B genetic correlation among sites was r=0.70. The additive variance was set to 10 and dominance variance was 4.

For both the first- and second-generation programs, we looked at seed orchard gains from nine different selection options. For each generation, we used only the progeny test information from that generation, i.e., we did not use first-generation data when selecting second-generation parents or progeny. For a given set of data, seed orchard candidates could be progeny, parents or both. For each of these options we examined three selection age scenarios:

• Select the top 20, limited to 1 selection per family, using age-6 data.

• Select the top 40 (no more than 2 per family) on age-6 data, and rogue to the best 20 using age-12 data.

• Select the top 20 selections, limited to 1 selection per family, using age-12 data.

Gains were derived from age-12 genetic values in standard deviation units, scaled to 30% gain for the scenario modeling the first generation where the best 20 parents or progeny are selected on age 12 data. 30% gain is a reasonable target for age-12 or age-15 volume gain from a 20-clone orchard based on 2,000 tested families. For the option where we selected both parents and progeny with age-12 data, we noted the number of parents selected (as opposed to progeny).

Simulations were run 225 times, and the average gain and standard deviation of the gains were generated. The percentage of times that the 20 best progeny were better than the 20 best parents was also noted for all three of the selection age comparisons.

Results And Discussion

The average gains indicate that, for the first-generation program simulated, parental orchards are generally superior to orchards based entirely on open-pollinated progeny (Table 4, Figure 2). The opposite held in the next generation with full-sib crosses (Table 4, Figure 3). In either generation, using a combination of progeny (forwards) and parental (backwards) selections yielded the highest gains (on average).

In the first generation simulation, progeny orchards were superior to parent orchards in only 6% to 16% of the runs (Table 4). Not only were there higher gains in the parental orchards, but there was less variation associated with the gain estimates (Figure 2). The combined orchards using age-12 data averaged 19 parents and only one progeny. The advantage of using parental-selections increases with increasing numbers of starting parents (OP families) (Table 5). This is because as more parents are tested, a larger number will be found that have truly superior breeding values. With small number of starting parents, a progeny orchard could well be superior.

Table 4. Percentage of times that a progeny (forwards selections) orchard was superior to a parental (backwards selections) orchard for a first –generation open-pollinated program and a second-generation control-pollinated program (see text for details).

Orchard Scenario	Breeding program type	
	1st-generation open-pollinated	2nd-generation control-pollinated
Best 20 based on age-6 data	15 %	93 %
Best 40 based on age-6 data, roguing to best 20 on age-12 data	6 %	84 %
Best 20 based on age-12 data	16 %	95 %

Start with 2,000 open pollinated families, and select the following:

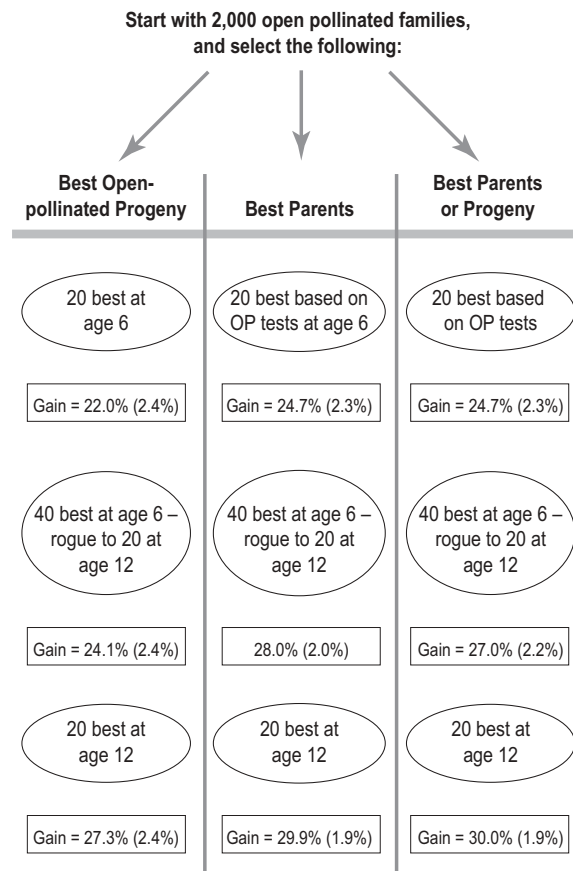| Best Open-pollinated Progeny	Best Parents	Best Parents or Progeny
20 best at age 6	20 best based on OP tests at age 6	20 best based on OP tests
Gain = 22.0% (2.4%)	Gain = 24.7% (2.3%)	Gain = 24.7% (2.3%)
40 best at age 6 – rogue to 20 at age 12	40 best at age 6 – rogue to 20 at age 12	40 best at age 6 – rogue to 20 at age 12
Gain = 24.1% (2.4%)	28.0% (2.0%)	Gain = 27.0% (2.2%)
20 best at age 12	20 best at age 12	20 best at age 12
Gain = 27.3% (2.4%)	Gain = 29.9% (1.9%)	Gain = 30.0% (1.9%)

Figure 2. Gains from a first-generation testing program: for a 20-clone, 1.5 generation orchard constructed by one of nine methods. Standard deviations of gain estimates given in parentheses.

It is hard to find an OP progeny better than an outstanding parent because considerably more within-family gain is needed such a parent than an average patent. Consider the two open-pollinated (OP) families depicted in Figure 4. Both families A and B are better than the population average. The breeding value of the female parent for each of these families is twice the deviation of the family mean from the population average. This is because we assume that the pollen parents of an OP family represent the population average; and the family mean is the result of the average of female parent and the pollen average. When calculating the breeding value of a within-family selection (a progeny) we only get half the female parent's breeding value since only half its genes come from the female parent; therefore, the family gain component is represented by the family

Start with 2,000 open pollinated families	→	Select best 200 families
Gain = 0		

Select best progeny per family	First-stage Gain = 21.1%

Generate 2 x 2 factorial mating design, grow out full-sib families, and select the following:

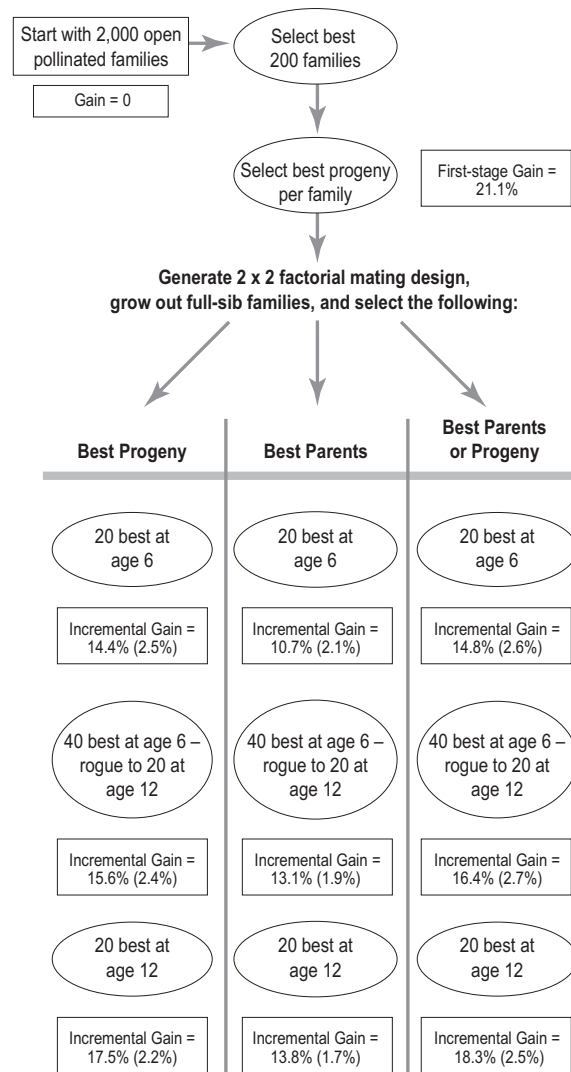| Best Progeny	Best Parents	Best Parents or Progeny
20 best at age 6	20 best at age 6	20 best at age 6
Incremental Gain = 14.4% (2.5%)	Incremental Gain = 10.7% (2.1%)	Incremental Gain = 14.8% (2.6%)
40 best at age 6 – rogue to 20 at age 12	40 best at age 6 – rogue to 20 at age 12	40 best at age 6 – rogue to 20 at age 12
Incremental Gain = 15.6% (2.4%)	Incremental Gain = 13.1% (1.9%)	Incremental Gain = 16.4% (2.7%)
20 best at age 12	20 best at age 12	20 best at age 12
Incremental Gain = 17.5% (2.2%)	Incremental Gain = 13.8% (1.7%)	Incremental Gain = 18.3% (2.5%)

Figure 3. Gains from second-generation testing using 2 crosses per parent, where a 20 clone orchard is constructed by one of nine methods. Standard deviations of gain estimates given in parentheses.

mean. Gain from within-family selection must exceed the gain from the family selection in order for the progeny to be superior to the parent.

When we examine the full-sib crosses in the second generation of simulations, orchards using the best progeny are more often superior to orchards using the best parents (Figure 3). In the combined second-generation orchard, selecting on age-12 data, there were an average of four parents and 16 progeny in the second-generation (ranged from 0 to 8). The reason that progeny were generally better than parents is that gain from family-selection when using full-sib

Table 5. Average gains from progeny and parental orchard, the percentage of times progeny orchard is best, and percent of progeny in a combined orchard; for a first-generation open-pollinated breeding program with different numbers of starting parents, i.e. open-pollinated families. All orchard selections were based on age-12 data and selecting 20 clones.

	Starting number of open-pollinated families			
	250	500	1,000	2,000
Progeny orchard gain (%)	22.6	24.0	25.8	27.4
Parental orchard gain (%)	21.1	24.3	27.4	30.0
% of times progeny orchard gives more gain than a parental orchard	67%	45%	24%	16%
Average number of progeny selections in a combined orchard	53%	32%	14%	4%

families obtains gain from precise selection of both the female and male parents . When two superior parents are crossed, the difference between the family mean and the best parent's breeding value is much less than that with OP families. This is shown in Figure 4; less within-family gain is needed to bridge the distance from the AxB full-sib family mean to the breeding value of parent B.

As expected, gains based on age-12 data were higher than gains based on age-6 data. However, establishing an orchard based on age-6 data and roguing at age 12 approached (within 1 to 3%) the gain from waiting to age-12 to start establishing orchards.

Gains from the second cycle of breeding and testing (200 full-sib families + orchard selection) of approached gains from the first cycle (as defined by testing 200 open-pollinated families and selecting the best progeny from 200 families) at 18% compared to 21%. This is reassuring given that the second cycle will cost around 10% the cost of the first cycle, if costs per planted tree are considered constant from the first- to the second generations. The highest cumulative gain from two cycles was 39.4%, 9.4% above the highest gain from the first cycle. This gain of 9.4% does not include unknown (but potentially large) gains from expanding breeding zones and using faster-growing sources from 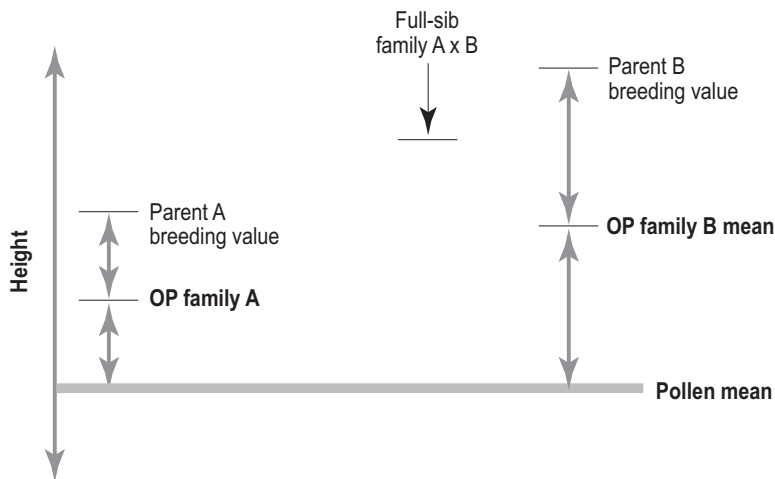outside the first generation breeding zone. The simulation mimics a situation where all breeding zones have the same mean; it is likely that the second cycle will pinpoint seed sources which are genetic "hotspots" and which can be deployed widely throughout the second-generation breeding zone.

It is worth pointing out that in each cycle, we can generate a certain amount of gain from breeding and testing a relatively large breeding population, and

Figure 4. Relationship between a half-sib family mean and the parents' breeding value

extract further gain by "creaming off" the best for orchards (the production population). For the first cycle in the US PNW, the sum of those two gains could be very large (if we are very selective when establishing orchards) since a very large number of families were tested. For the second cycle, we would get gains from two cycles of testing and a subsequent "creaming off" step; the sum of those three gains will be larger than (but not double) the gains from an elite 1.5 generation orchard.

It is worth comparing these results with the publications referred to previously. Hodge and White (1993) found the top ranking parental selections to have higher breeding values than the top ranking offspring selections (from full-sib crosses) with populations of 201 and 2001. One difference with their study was that the full-sibs were random crosses, with no selection of the parents (unlike the BZERC model in which only the top 10% were crossed). Ruotsalainen and Lindgren (1998) similarly showed that where one exerts a high selection intensity, the best parents will beat the best open-pollinated progeny. For example, with a heritability of 0.2 and a selection intensity of 1 in a 100, the backward selection had a higher breeding value than the highest-ranked open-pollinated family selection in 25% of the families. Burdon and Kumar (in press, modeling a situation with 300 tested first-generation parents) found results very similar to ours. In their case, the best 15 (of 300) first-generation parents beat the best 15 forward selections from the 300 open-pollinated families, but were in turn beaten by the best 15 progeny in full-sib trials.

Reference to the southeastern USA (where an extra cycle of breeding and testing has been completed) can be instructive. For example, the third cycle slash pine breeding population of 466 selections is to contain 71 backward selected first-generation parents, 95 backward selected second-generation parents and 300 untested third-cycle selections (White et al. 2003).

We have a chance to further boost the gain from the second cycle of testing by doing every-thing possible to increase the heritability in the tests being planted. We assumed an individual heritability of 0.25, but with good site selection, preparation and maintenance we may hope to increase that. Some within-site heritabilities of 0.5 were obtained in the best first-generation tests, although such within-site estimates are potentially biased upward by genotype x environment interaction.

Implications

Most cooperative 1.5 generation seed orchards (such as the Interim Dallas orchard grafted in 2002) recruit from many first-generation open-pollinated testing programs and have lots of parents from which to choose. In these cases, using parental selections would give more gain than all-progeny orchards and is essentially equal to the gain from selecting both progeny and parents. With small number of starting parents, a progeny orchard may be superior.

However, progeny will be more important in achieving gain after the second cycle of breeding and testing. Combined orchards will be the norm, and in many cases progeny will have the highest expected gain.

Gains from the second cycle of breeding and testing (200 full-sib families + orchard selection) of approached gains from the first cycle (as defined by testing 200 open-pollinated families and selecting the best progeny from 200 families) at 18% compared to 21%. This is reassuring given that the second cycle would cost only around 10% of the first cycle, if costs per planted tree remain constant.

There appeared to be good justification for selecting based on age-6 data, establishing an orchard, and roguing based on age-12 data rather than waiting for age-12 data to begin orchard establishment.

NWTIC cooperators have an opportunity to further boost the gain from the second cycle of testing by doing everything possible to increase the heritability in the tests being planted (with good site selection, preparation and maintenance).

Caveats and Context

As in many simulations, the main benefit is to compare the relative merits of different options, rather than to predict the absolute values. Thus we can infer the relative gain from backwards vs. forwards selections from this exercise, but it is not designed to give an accurate estimate of gain for a given second-generation breeding program or orchard.

As in many simulations, changing the assumptions and input parameters would impact the results. For example, increasing the individual heritability from 0.25 to 0.50 would increase the gain from forward selections faster than it would increase the gain from backward selection.

References

Burdon, R.D. and Kumar, S. Expected gains and their distributions from 1.5- and second-generation selection: stochastic simulation. Silvae Genetica (in press).

Hodge, G. R. 1985. Parent vs. offspring selection: A case study. Pp. 145–154 in Proc. 18th Southern Forest Tree Improvement Conference, May 21–23, Long Beach, Mississippi.

Hodge, G.R. 1997. Selection procedures with overlapping generations. Pp. 199–206 in Burdon, R.D.; Moore, J.M. (Ed.) "IUFRO '97 Genetics of Radiata Pine". Proceedings of NZ FRI-IUFRO Conference 1-4 December and Workshop 5 December, Rotorua, New Zealand. FRI Bulletin No. 203.

Hodge, G.R. and White, T.L. 1993. Advanced-generation wind-pollinated seed orchard design. New Forests 7: 213–236.

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

White, T., Huber, D. and Powell, G. 2003. Three cycles of slash pine improvement. 27th Biennial Southern Forest Tree Improvement Conference, June 24–27 2003, Oklahoma State University, Stillwater, OK, USA.