ted in quantitative inheritance), but several transgressions are evident. Most noteworthy are higher percentages of the citronellol group and car-3-ene, and lower ones of α-terpineol and the sesquiterpenes. These terpenes did not appear to have any significance in describing geographic variation of natural populations (von RUDLOFF, 1972, 1973, 1973, 1975) (von RUDLOFF and REHFELDT, 1980). Similar conclusions could be drawn from the leaf oil compositions of the narrow crosses (parent trees 28, 36, 102, and 109). When one of the parents has a small amount of interior intermediacy (e.g. trees 83, 102, 109, and 517 with 5—6% camphene group), this feature is lost in the crosses. We have reported earlier on the abrupt reduction of the camphene group percentages (dominant in the Rocky Mountain variety) in the zones of overlap of the two varieties when sampling the more westerly stands (von RUDLOFF and REHFELDT, 1980).

Conclusion

In coastal Douglas-fir there is a high degree of variability in the leaf oil terpene patterns, which indicates a high degree of heterozygosity. Only in the dwarf form of the S_2 generation is a trend towards homozygosity evident. Female dominance does not appear to be a major factor in leaf oil terpene inheritance. Hence, in chemosystematic studies utilizing leaf oil terpene compositions, variable patterns from type A through B to C (von RUDLOFF, 1973) can be expected in coastal populations. The small degree of intermediacy with the interior variety found in some coastal locations may be lost in the next generation.

References

HEAMAN, J. C.: Research Division, British Columbia Ministry of Forestry, Victoria (1970). — HUNT, R. S. and von RUDLOFF, E.: Chemosystematic studies in the genus *Abies*. I. Leaf and twig oil analysis of alpine and balsam firs. Can. J. Bot. 52, 477—487 (1974). — KEPNER, R. E., ELLISON, B. O. and MULLER, C. J.: Sesquiterpene hydrocarbons in needles of *Pseudotsuga menziesii*. Phytochem. 14: 808 (1975). — ORR-EWING, A. L.: The collection of Douglas-fir inbreds at the Cowichan Lake Experiment Station, Vancouver Island. Brit. Col. Min. of Forests, Res. Note. 8 (EP474) 1977). — von RUDLOFF, E. and REHFELDT, G. E.: Chemosystematic studies in the genus *Pseudotsuga*. IV. Inheritance and geographic variation in the leaf oil terpenes of Douglas-fir from the Pacific Northwest. Can. J. Bot. 58, 546-556 (1980). — von RUDLOFF, E.: Volatile leaf oil analysis in chemosystematic studies of North American conifers. Biochem. System. and Ecol. 2, 131—167 (1975). — von RUDLOFF, E.: Chemosystematic studies in the genus *Pseudotsuga*. III. Population differences in British Columbia as determined by volatile leaf oil analysis. Can. J. For. Res. 3, 443—452 (1973). — von RUDLOFF, E.: Chemosystematic studies in the genus *Pseudotsuga*. I. Leaf oil analysis of coastal and Rocky Mountain varieties of Douglas-fir. Can. J. Bot. 50, 1025—1040 (1972). — von RUDLOFF, E.: Chemosystematic studies in the genus *Pseudotsuga*. II. Geographical variation in the terpene composition of the leaf oil of Douglas-fir. Pure Appl. Chem. 34, 401—410 (1973).

Number of Offspring and Plot Sizes Required for Progeny Testing

By P. P. COTTERILL[1]) and J. W. JAMES[2])

Summary

The accuracy of progeny testing is examined in relation to numbers of offspring tested and plot size used in the field design. Two mathematical solutions are presented, together with results of subsampling studies employing data for *Pinus radiata* in New Zealand and Australia. The mathematical solutions determine either the size of the differences between families in a progeny test which should be detected as statistically significant, or the probability that the family having the best true mean will also have the best observed mean in a progeny test. The general recommendation from both the mathematical and subsampling studies is that 10 to 20 individuals per family, growing in single-tree plots or two-tree non-contiguous plots, are sufficient to reliably evaluate each family.

Key words: Progeny testing, family means, plot size, field design.

Zusammenfassung

Die Genauigkeit einer Nachkommenschaftsprüfung wird in bezug auf die Anzahl der getesteten Nachkommen und die Anzahl der Bäume je Parzelle, die im Freilandversuch notwendig sind, untersucht. Hierzu werden zwei algebraische Lösungen vorgeschlagen, zusammen mit den Ergebnissen aus den Stichprobenentnahmestudien unter Verwendung der Daten für *Pinus radiata* in Neuseeland und Australien. Es wird allgemein empfohlen, daß 10—20 Nachkommen pro Familie, die in Einzelbaumparzellen oder in Parzellen mit je zwei Bäumen, aufwachsen, für eine zuverlässige Auswertung jeder Familie genügen.

Résumé

La précision du test de la descendance est examinée par rapport aux nombres de descendants testés et au nombre d'arbres sur une parcelle usée pour les essais sur le terrain. L'auteur propose deux solutions algébraiques et énonce à la fois les résultats des études du sous-échantillonage, en employant les données pour les pins *(Pinus radiata)* en Nouvelle-Zélande et en Australie. Il préconise généralement que 10 à 20 descendants par famille, qui croissent sur les parcelles supportant une ou deux arbres, suffisent pour évaluer d'une maniére digne de confiance chaque famille.

Introduction

A major problem in progeny testing is "How many individuals are required to adequately evaluate each family?" One solution is given by ROBERTSON (1957) with particular reference to animal breeding. This author provides estimates of genetic gain from progeny testing where a limit exists on the total number of offspring that can be tested, and the breeder has the opportunity to compromise between number of families tested (which determines the intensity of subsequent selection among families) and number

[1]) CSIRO Division of Forest Research, The Cunningham Laboratory, 306 Carmody Road, St. Lucia, Queensland 4067, Australia.

[2]) School of Wool and Pastoral Sciences, The University of New South Wales, P.O. Box 1, Kensington, New South Wales 2033, Australia.

of individuals per family (which determines the accuracy of the evaluation of each family). Genetic gain was found to be maximised by testing as few as five individuals per family and thereby devoting more resources to testing more families. ROBERTSON's solution represents a balance between choice (i. e. the more families tested the greater the choice) and accurarcy of the evaluation of each family.

Progeny testing is not always viewed as a balance between choice and accuracy. For instance, it is common in tree breeding to have a fairly definite number of families available for testing (i.e. a fixed choice) and the breeder is primarily concerned with accuracy. In other words, the tree breeder is interested in knowing how many offspring need to be tested in order to achieve reliable evaluations of each family (or parent) so the best may be chosen for future breeding and seed orchards. KUNG (1972) has attempted to relate the sensitivity of progeny tests in detecting significant differences among families of trees with the number of families tested and number of individuals per family. However, the measures of sensitivity used by KUNG do not give a particularly clear indication of the accuracy of progeny testing in identifying best families.

Another problem in progeny testing is "What combinations of plot size and number of plot replications are optimum for evaluating families?" (We assume here that the number of individuals per family is fixed and an increase in plot size must be at the expense of a decrease in plot replication.) WRIGHT and FREELAND (1960) have studied the influence of plot size in forestry experiments in general and concluded that single-tree plots are efficient for detecting significant differences between treatments. CORRELL (1978) concluded that for field experiments with *Pinus radiata* in South Australia, a row of four trees is the optimal plot size but little statistical efficiency is lost using plot sizes of between about one and 10 trees. CORRELL measured statistical efficiency in terms of either the variance between plot means or the standard error of the difference between treatment means. LEE (1972a, b) and BARNES and SCHWEPPENHAUSER (1979) subsampled progeny data for *P. nigra* or *P. patula* to examine the effects of plot size specifically for progeny testing but their results are difficult to appraise because the number of individuals per family was not or could not be held constant. JOHNSTONE and SAMUEL (1974) report experiments where progeny tests of *Larix* spp. and *Picea sitchensis* were repeated for different plot sizes and numbers of replications. Some treatments in the *Picea sitchensis* experiment maintained a constant 36 or 40 individuals per family and, under these circumstances, plot sizes of three to nine trees tended to minimise variance between plots. JOHNSTONE and SAMUEL do not report what influence these plot sizes and replications had on more direct measures of the accuracy of family evaluations (i.e. family rankings).

This article attempts to provide general recommendations on the number of offspring required to accurately evaluate each family, and the optimum plot size and replications for planting these individuals in the field. Two mathematical approaches are used to give direct measures of the accuracy of family evaluations when the number of individuals per family is varied. The theory is supported by a subsampling study employing data for *P. radiata* D. Don in New Zealand and Australia. Another subsampling study employing data for *P. radiata* in Australia is used to determine optimum combinations of plot size and replication.

Mathematical Solutions

The following algebra overcomes difficulties in combining several traits by assuming that the progeny test is for one trait only. This single trait may, however, be an index which combines information from several traits. The trait used as the selection criterion is assumed to be normally distributed.

Solution 1: Differences which should be Detected Between-Families

Statistical texts give the size of samples required to compare two treatments as —

$$n = (Z_\alpha + Z_\beta)^2 \, (2 \, \sigma_e^2/\delta^2) \tag{1}$$

(e.g. SNEDECOR and COCHRAN 1967, Equation 4.13.1). In the case of progeny testing, n represents the number of individuals in each family, δ the size of the difference between the true means of any two families in the test, σ_e^2 the variance within families, and $(Z_\alpha + Z_\beta)^2$ is a constant. Values of $(Z_\alpha + Z_\beta)^2$ are given by SNEDECOR and COCHRAN (1967) and reflect the probability that the progeny test will show a statistically significant difference between the observed means of two families when the difference between their true means is δ. The higher the probability and level of significance the breeder may require for the progeny test to show a significant difference between the family means, the higher the value of $(Z_\alpha + Z_\beta)^2$.

The within-family variance in Equation 1 can be written

$$\sigma_e^2 = (1-t) \, \sigma^2 \tag{2}$$

where t represents the intra-class correlation and σ^2 the phenotypic variance. Assuming an additive model; $t = \frac{1}{4}h^2$ for half sib families and $\frac{1}{2}h^2$ for full-sibs (h^2 is the individual heritability of the trait being evaluated).

Substituting Equation 2 into 1 it is possible to solve for, δ/σ, the minimum difference in phenotypic standard deviations which should be able to be detected between the means of two families with a probability and level of significance corresponding to $(Z_\alpha + Z_\beta)^2$ —

$$\delta/\sigma = (Z_\alpha + Z_\beta)^2 \, \sqrt{(2-2t)/n} \tag{3}$$

Solutions to Equation 3 are plotted against n in *Figure 1* for circumstances where the offspring are either half-sib or full-sib, h^2 is 0.05 or 0.8, and the breeder requires 80% or 95% probabilities of detecting a statistically significant difference between families at the 5% level.

The minimum difference δ/σ which should be detected between families decreases logarithmically as the number of individuals per family is increased *(Figure 1)*. The curves have flattened appreciably by about n = 10 individuals and after about n = 20 there is very little reduction in δ/σ with further increases in n. We conclude from *Figure 1* that around 10 individuals per family represents a lower limit for reasonably accurate evaluation of families, while 20 individuals per family represents an upper limit beyond which further testing of offspring will probably lead to little increase in accuracy. The 10 to 20 individuals per family is greater than the five individuals recommended by ROBERTSON (1957) for maximising genetic gain from selection among families, but ROBERTSON's calculations show that testing 10 to 20 individuals will usually lead to a response from selection which is within 10% of the maximum. When the cost of progeny testing is taken into account through benefit-cost analyses (such as those used by HILL 1974 for breed-comparison experiments) the re-
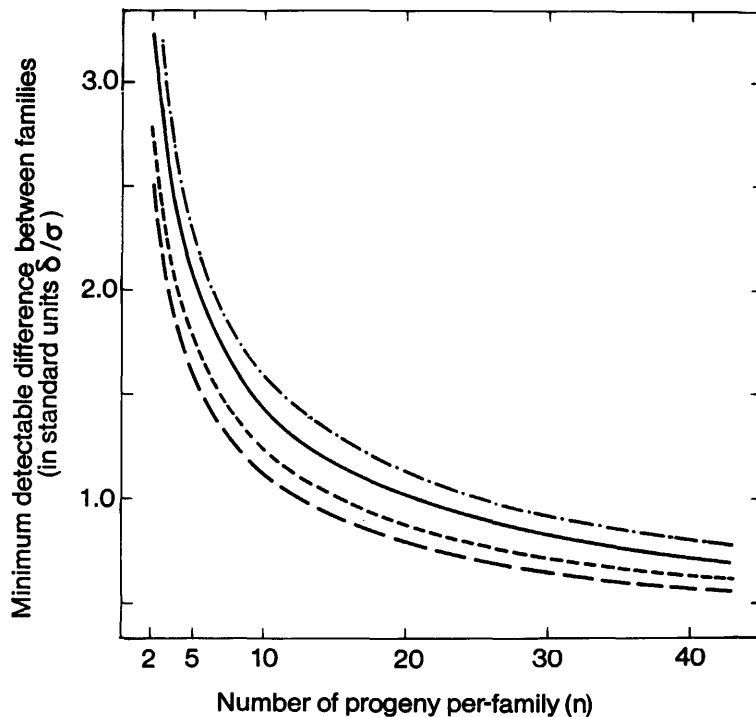
Figure 1. — Minimum differences δ/σ (in units of phenotypic standard deviation) between the means of families in a progeny test which should have either a 95% or 80% probability of being found statistically significant at the 5% level. Minimum differences between means of half-sib families are plotted against number of individuals per family n for the circumstances of 95% probability and either $h^2 = 0.05$ (——·——) or $h^2 = 0.8$ (————); or for the circumstances of 80% probability and either $h^2 = 0.05$ (— — — — —) or $h^2 = 0.8$ (— —). Minimum differences between means of full-sib families are plotted against n for the circumstances of 95% probability and either $h^2 = 0.05$ (—— · ——) or $h^2 = 0.8$ (— — — — —).

commendation may again favour less than 10 individuals per family, but this would depend on the cost structure of particular breeding operations.

At any level of n, the minimum difference δ/σ is reduced when either the probability required for detecting a significant difference is decreased, or the heritability of the trait being evaluated is increased (Figure 1). When heritability is high ($h^2 = 0.8$) the use of full-sib instead of half-sib offspring also reduced δ/σ (Figure 1) because the full-sibs had less within-family variation (σ^2_e; Equation 2). However, under conditions of low heritability ($h^2 = 0.05$) there is very little difference in variation within full-sibs compared with half-sibs, and consequently the use of full-sib offspring did not substantially reduce δ/σ.

The usefulness of the parameter δ/σ and Figure 1 for predicting the accuracy of family evaluations is best illustrated by example. Consider a progeny test for diameter at say 4½ years in P. radiata in South Australia. It is known that the trait has a heritability of around 0.15 and a phenotypic standard deviation of around $\sigma = 2.2$ cm (Cotterill and Zed 1980). We may use Figure 1 to calculate that n = 20 half-sib individuals per family should allow an 80% probability of detecting a minimum difference between-families of at least $\delta = 1.98$ cm (i.e. $\delta/\sigma = 0.9$ for $h^2 = 0.05$, Figure 1). For n = 10 half-sibs the difference should be $\delta = 2.64$ cm (i.e. $\delta/\sigma = 1.2$), and for five half-sibs $\delta = 3.96$ cm ($\delta/\sigma = 1.8$).

A major shortcoming of this mathematical Solution 1 to predict the accuracy of progeny testing is that "accuracy" is measured in terms of the probability that the classical null hypothesis of no difference between families is rejec-

ted. However, the objective of progeny testing is to evaluate families to determine which are best, not determine the statistical significance of the differences which exist among families (Hill 1974). The next mathematical solution we present does not attempt to determine probabilities of rejecting the null hypothesis of equal means, but gives the expected probability that correct decisions will be made in identifying best families.

Solution 2: Probability of Detecting the Best Family

It is possible to estimate the probability that the best family (i.e. the family having the best true mean) will also have the best observed mean in a progeny test with n individuals per family. This represents the probability that d, the difference between the observed mean ($\overline{X}_1$) of the best family and the observed mean ($\overline{X}_2$) of another family, is greater than zero (i.e. P (d) > 0; where $d = \overline{X}_1 - \overline{X}_2$). Assuming the population of d's is normally distributed with mean Ψ and variance $2\sigma^2_e/n$, the probability P (d) > 0 is given by the area under the standard normal curve to the right of a point Z at which d = 0 —

$$Z = (d-\psi)/\sqrt{2\sigma_e^2/n} \tag{4}$$

$$= -\psi/\sqrt{2\sigma_e^2/n}; \text{ where } d = 0 \tag{5}$$

Substituting Equation 2 —

$$Z = (-\psi/\sigma)\sqrt{n/(2-2t)} \tag{6}$$

The approach is similar to the method adopted by Jones (1980) for determining optimum sample size for prize exhibitions with animals. Probabilities corresponding to different values of Z in Equation 6 can be derived from cumulative standard tables. For instance, *Table A3* in Snedecor and Cochran (1967) gives the area under the standard normal curve from the origin to a point Z. The total area to the right of the negative values for Z derived from Equation 6 would be the area given in *Table A3* plus 0.5.

Table 1 gives probabilities determined from Equation 6 for a range of values of n and ψ/σ, under circumstances where the offspring are half-sib and $h^2 = 0.4$. The value ψ/σ is of course the difference in standard deviations between the true means of the two families being compared. Preliminary calculations revealed that the probabilities in *Table 1* are essentially the same for full-sib offspring and high or low heritabilities. It is apparent from *Table 1* that 10 to 20 individuals per family give the breeder a very good chance of making correct decisions from the results of progeny testing. For instance, 20 individuals per family can be expected to give a 95% chance of detecting the best family when that family has a true mean one-half of a phenotypic standard deviation greater than the true mean of another family (i.e. $\psi/\sigma = 0.5$; *Table 1*). The corresponding probability for n = 10 individuals is 88%, while for n = 5 the probability falls to 80%.

A very large number of individuals per family is required to give reasonable probabilities of identifying the best family when its true mean is only one-quarter of a standard deviation greater than other families (i.e. n = 50 individuals are required to give a 91% probability; *Table 1*). In practice, it is unreasonable for breeders to expect this level of accuracy from progeny testing. In fact James (1975) demonstrated that it is not even necessary to measure traits to within one-quarter of a standard deviation. This author calculates that for progeny tests or other large field experiments an acceptable unit of measurement is around one-third of a standard deviation. Thus, if diameter at $4^{1}/_{2}$ years has a phenotypic standard deviation of 2.2 cm, measurement of the trait in units of 0.7 cm (or rounded to units of say 0.5 cm) would be sufficiently accurate.

Subsampling Radiata Pine Progeny Data

Family Rankings for Different Numbers of Offspring

Height and diameter data for first- generation progeny tests of *P. radiata* in the Kaingaroa Forest (North Island of New Zealand) and the Myora Forest (south-eastern South Australia) have been employed to measure the extent to which rankings of families (i.e. accuracy of family evaluations) are altered by reducing the number n of individuals used to calculate the family mean. The object is to determine, for actual data, how few offspring may be tested before family evaluations become unacceptably erroneous.

The progeny test at Kaingaroa was measured six years after planting and the data we used is from 25 half-sib families. These half-sib families are polycross, produced by controlled-pollination using a mix of 10 pollens. The field design of the progeny test is 10 replications of non-contiguous plots of approximately five single-trees (planted at a spacing of 4 × 4 m), with a total of n = 33 to 50 individuals measured for each family. The mean height of trees in the test was 8.1 m ($\sigma = 1.0$ m) and the mean diameter 15.8 cm ($\sigma = 2.3$ cm). The Myora progeny test was measured $4^{1}/_{2}$ years after planting and we have used data from 44

Table 1. — Probability that a family having a true mean δ/σ standard units greater than another family will have the best observed mean in a progeny test with n half-sib progeny per-family. The heritability of the trait being evaluated is 0.4.

Standard difference between true means (ψ/σ)	Number of individuals per family (n)						
	2	5	10	15	20	30	50
0.25	60	66	72	77	80	85	91
0.50	70	80	88	93	95	98	100
0.75	79	89	96	99	99	100	100
1.00	85	95	99	100	100	100	100

families; 33 of which are full-sib and the remainder half-sib (open-pollinated in a clonal orchard). The field design is randomised complete blocks with 30 replications of single-tree plots (planted at a spacing of 2.1 × 2.1 m), with n = 29 to 30 individuals measured per family. The mean height of trees was 9.9 m ($\sigma = 1.0$ m) and the mean diameter 13.9 cm ($\sigma = 2.2$ cm).

The Kaingaroa data were reduced from 10 replications of around five-tree non-contiguous plots to 10 replications of exactly four, three, two and single-tree plots, and then to five and finally two replications of single-tree plots; giving n = 33 to 50 (complete data), 40, 30, 20, 10, 5 and 2 individuals per family. The Myora data were reduced by deleting replications of the single-tree plots to give n = 29 to 30 (complete data), 20, 10, 5 and 2 individuals. The subsampling of both progeny tests was at random. Two independent subsamples were actually taken for each level of n and both gave essentially the same results (i.e. essentially the same rankings of families). Rank correlations (described below for each pair of replicate subsamples were averaged and the means presented in *Figure 2*. Progeny tests that employ single-tree plots were used deliberately in this study to avoid confounding of subsample and plot sizes.

Spearman's rank correlations (Snedecor and Cochran 1967) have been used to quantify the degree of association be-
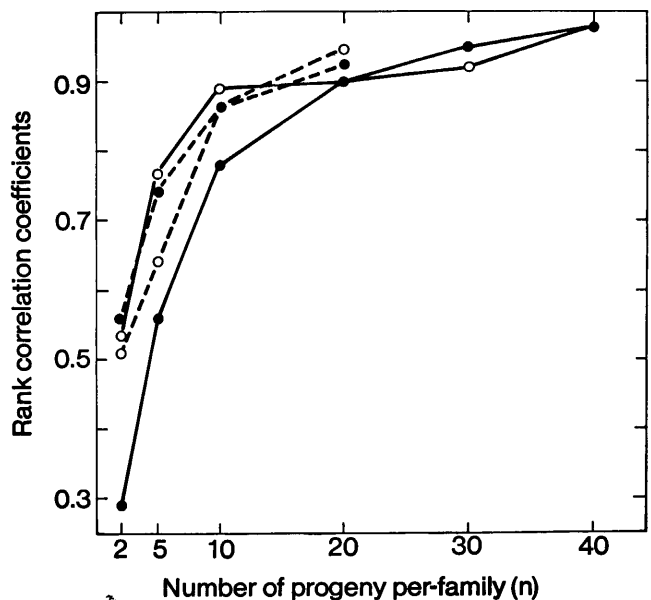

Figure 2. — Correlations between rankings of families calculated for complete data compared with rankings calculated for subsamples of 40, 30, 20, 10, 5 or 2 individuals per family. The data were height (●) and diameter (○) measurements for *Pinus radiata* progeny tests in the Kaingaroa (————) and Myora (— — — —) Forests. Complete data for the Kaingaroa test were 33 to 50 individuals and for the Myora test 29 to 30 individuals per family.

tween rankings of families calculated 'for complete data compared with their rankings calculated for each level of n, and results plotted in *Figure 2*. It was encouraging to find close agreement not only between the two traits and two sites studied, but also between the general results of this subsampling study and the previous mathematical solutions. Subsamples of n = 20 or more individuals per family were found to rank families in essentially the same order as the complete data (rank correlations were consistently above 0.9; *Figure 2*). Subsamples of n = 10 individuals provided correlations of between 0.78 and 0.89, which indicate reasonably close agreement in the rankings of families. However, sampling fewer than 10 individuals per family led to a marked decline in the reliability of family evaluations. Rank correlations for n = 5 individuals ranged from 0.77 to an unacceptably low 0.56.

Ordinary product-moment correlations were also calculated between family means estimated for complete data and means estimated for each level of n, and the absolute values of these correlations were found to be almost identical to corresponding rank correlations. The reason for preferring rank correlations in this study is because we consider the main purpose of progeny testing is to rank families so the best are chosen for future breeding.

Lee (1972a, b) found trends similar to those in *Figure 2* for subsamples of height, specific gravity and tracheid length data of 27 seed sources of *P. nigra*. In this case product-moment correlations of over 0.96 were observed between seed source means calculated for n = 40 individuals per family (complete data) and corresponding means calculated for either n = 30 or 20 individuals, while correlations for n = 10 remained high at between 0.86 and 0.93. Lee did not consider samples of fewer than 10 individuals per family.

Another way of quantifying the reliability of family rankings is to list say the top 20% of families according to each level of n and count how many of these families are also ranked in the top 20% for the complete data. The top 20% of families represent the top five families for the Kaingaroa progeny test and approximately the top 10 families for the Myora test. *Table 2* reveals that subsamples of 10 or more individuals per family were successful in detecting at least four of the top five families for the Kaingaroa test and at least eight of the top 10 families for the Myora test. The subsamples of five individuals per family were surprisingly successful in correctly identifying three and seven of the top families for the Kaingaroa and Myora tests, respectively. These findings suggest that the accuracy of the n = 10 (and perhaps the n = 5) samples would be

sufficient for most practical breeding purposes. The results in *Table 2* tend to support the conclusion of Robertson (1957) that genetic gain from progeny testing is maximised by measuring no more than five individuals per family in order to devote more resources to testing more families.

Family Rankings for Different Plot Size and Replication

Here we subsampled height and diameter data from another first-generation progeny test of *P. radiata* in the Myora Forest to determine the accuracy of family evaluations when either 12 or 24 individuals per family are tested under different combinations of plot size and numbers of plot replications. Subsamples of n = 12 and 24 individuals have been chosen because they are convenient denominators for the particular set of data, and these numbers of offspring approximate the range we recommend for progeny testing. Other studies of plot size in progeny testing are based on larger values for n (e.g. Johnstone and Samuel 1974; n = 36 or 40). The following subsampling of experimental data to determine optimum plot size will tend to bias against smaller plots because no account is taken of the effect of decreased block size and therefore more efficient blocking under smaller plots. Hence the method is useful mainly for placing an upper limit on plot size (Correll 1978).

The progeny data we have used includes 28 half-sib families (open-pollinated in clonal orchards) which were measured $4\frac{1}{2}$ years after planting. The field design of the progeny test is randomised complete blocks with 12 replications of six-tree row plots; n = 66 to 72 individuals were measured per family. The mean height of trees was 9.8 m ($\sigma = 1.0$ m) and the mean diameter was 14.4 cm ($\sigma = 2.2$ cm). In the case of the n = 12 subsamples, the data were reduced to the following combinations of plot size and plot replication: single-tree plots × 12 replications (1-tree × 12), 2-tree (row) × 6, 3-tree (row) × 4, 4-tree (row) × 3, and 6-tree (row) × 2. The n = 24 subsamples were 2-tree (row) × 12, 3-tree (row) × 8, 4-tree (row) × 6, and 6-tree (row) × 4. The subsampling of block replications was at random and two subsamples were taken for each plot size × replication treatment. The means of correlations for each pair of replicate subsamples are presented in *Figure 3*.

Rank correlations were determined between rankings of families calculated for complete data (6-tree × 12, with n = 66 to 72) and rankings calculated for the subsamples of 12 or 24 individuals per family. We must assume that the original plot size of six trees provides accurate rankings of families when a large number of progeny (i.e. n = 66 to 72) are tested. This assumption is supported by the findings of Johnstone and Samuel (1974) and Correll (1978).

Figure 3 demonstrates that for n = 12 individuals the single and two-tree plot designs gave the most reliable rankings of families. The somewhat lower magnitude of the actual rank correlations for these 1-tree × 12 and 2-tree × 6 subsamples, compared with corresponding correlations for n = 10 in *Figure 2*, can probably be attributed to reduced variation between families which was evident in the present progeny test. For n = 24 individuals the plot sizes of three or fewer trees in a row were found to rank families in essentially the same order as the complete data *(Figure 3)*.

Our general recommendation is that plot sizes of two trees or less should be used where progeny testing is based on around 10 to 20 individuals per family. Larger plot sizes entail correspondingly fewer plot replications and can lead to an unacceptable decline in accuracy of

Table 2. — Number of the top 20% of families according to complete data which were also ranked in the top 20% according to subsamples of data. The top 20% of families for the Kaingaroa and Myora progeny tests are the top five and 10 families, respectively.

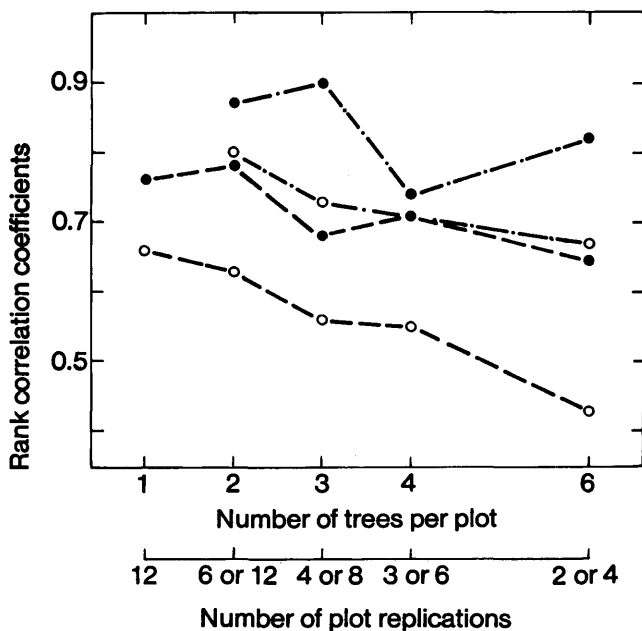| Trait	Number of individuals per family (n)					
	2	5	10	20	30	40
Kaingaroa test: Number of top 5 families identified correctly						
Height	2	3	4	4	4	4
Diameter	3	3	4	4	4	5
Myora test: Number of top 10 families identified correctly						
Height	6	7	8	10		
Diameter	5	7	8	9		

Figure 3. — Correlations between rankings of families calculated for complete height (●) and diameter (○) data (6-tree row plots × 12 replications) for a *Pinus radiata* progeny test in the Myora Forest compared with rankings calculated for subsamples of 12 (———) or 24 (——·——) individuals per family. The subsamples of 12 individuals included the following combinations of plot size and replications: 1-tree plot × 12 replications, 2-tree × 6, 3-tree × 4, 4-tree × 3 and 6-tree × 2. The subsamples of 24 individuals included 2-tree × 12, 3-tree × 8, 4-tree × 6 and 6-tree × 4.

family evaluations. Of course this recommendation is made from a study of just one species on one site, but the conclusions are largely commonsense and there seems little reason for doubt. Where progeny testing employs substantially more than 20 individuals per family, larger plot sizes appear to be acceptable (JOHNSTONE and SAMUEL, 1974).

The use of single-tree plots in progeny testing has been a source of argument since they were recommended by WRIGHT and FREELAND (1960). Controversy has centred on the amount of heterogeneity which may be created in the genetic portion of error variance when single-tree plots are employed, and the effect (if any) such heterogeneity would have on analyses of variance (SHIUE and PAULEY 1961 versus FRANKLIN 1971). This controversy is of little concern where progeny testing is used solely for the purpose of evaluating and ranking families as distinct from determining the statistical significance of the differences that exist between families. Missing values, which invariably occur with single-tree plots, can also complicate analyses of variance but provided their frequency is not too great they should have little effect on the reliability of family evaluations.

Discussion

Numbers of offspring and plot size for progeny testing have so far been studied from the standpoint of identifying best families for future breeding. However, the progeny tests themselves may have secondary functions, such as providing data to estimate genetic paramters or serving as a population for future selection. We will briefly discuss these secondary functions and their influence on numbers of offspring and plot size. We also consider progeny testing on more than on site, and the special case of all-or-nothing traits.

ROBERTSON (1959) has determined that the optimum number of half-sib individuals per family for estimating heri-

tability is approximately $n = 4/h^2$. Since the inheritance of most commercial traits in forestry is around $h^2 = 0.4$ to 0.2 the optimum n should be around 10 to 20 individuals. This conclusion is supported by calculations of HARRIS (1964). Accurate estimation of low heritabilities (below say $h^2 = 0.1$) will obviously require many more offspring but, in terms of say expected genetic gain from individual selection, these heritabilities are often written off as being effectively zero (even though the economic importance of the trait may be high). The accurate estimation of genetic correlations appears to require more offspring than estimation of inheritance (ROBERTSON 1959) but as ROBERTSON has indicated, we may not need to know the genetic correlation as accurately as the heritability. In general there seems little conflict, as far as numbers of offspring and plot size is concerned, between accurate evaluation of families and estimation of genetic parameters.

A fundamental conflict does occur where progeny tests are used as breeding populations for future selection because, in advanced-generation breeding programs, selection is within-families as well as among families (BURDON und SHELBOURNE 1971). Maximising genetic gain from individual selection within-families requires as many individuals as possible in each family (to ensure the highest possible intensity of selection) and large plots (to increase the efficiency of comparisons among individuals; BURDON et al., 1977). Considerations of cost may favour progeny tests doubling as a breeding population, but considerations of experimental design certainly favour specialised progeny tests to identify best families and specialised breeding populations to facilitate selection within these best families (e.g. COTTERILL 1984). The progeny tests may have 10 to 20 individuals per family planted in one or two-tree plots, while the breeding population should have as many individuals as possible per family which could be planted in very large unreplicated plots.

Where progeny testing is conducted across multiple sites, we recommend that each test have sufficient individuals per family (i.e. $n = 10$ to 20) to allow it to stand alone. This both provides insurance against failure of other tests and allows flexibility to either select families (parents) for performance at a particular site, or pool results and use mean performance across sites at the selection criterion. These options for selection are of course relevant where family × site interactions are important.

HILL (1974) has suggested that all-or-nothing traits may be a special case for determining how many offspring are required to evaluate genotypic performance. All-or-nothing traits are binomially distributed and therefore violate the assumption of normality required for the mathematical Solutions 1 and 2. However, in practice, the problem does not appear to be too serious. The single-tree plot Myora data which we subsampled for this study included an all-or-nothing score for forking (i.e. score 0 = no fork, 1 = forked main stem) and although results are not reported in *Figure 2* the trait was found to behave similarly to height and diameter, with 10 individuals per family providing reasonably accurate rankings of families. Other traits not reported for the Myora and Kaingaroa data were subjective scores (using 5 or 9-point scales) of stem straightness and crown form. These scores had approximately normal distributions and behaved similarly to the growth traits.

General Recommendation

Our general recommendation is that 10 to 20 individuals per family, growing in single-tree plots or two-tree non-

contiguous plots, should provide reasonably accurate progeny testing. Little additional accuracy is achieved from testing more offspring and indeed genetic gain and benefit-cost solutions may favour fewer than 10 individuals per family. Of course allowance has to be made for deaths or perhaps artificial thinnings when deciding how many offspring to actually plant when establishing a progeny test.

Acknowledgements

Literature Cited

Barnes, R. D. and Schweppenhauser, M. A.: Genetic control of 1, 5-year-old traits in *Pinus patula* Schiede et Deppe and a comparison of progeny test methods. Silvae Genet. 28: 156—67 (1979). — Burdon, R. D. and Shelbourne, C. J. A.: Breeding populations for recurrent selection: conflicts and possible solutions. N.Z. J. For. Sci. 1: 174—93 (1971). — Burdon, R. D., Shelbourne, C. J. A. and Wilcox, M. D.: Advanced selection strategies. Proc., 3rd Wrld. Consul. For. Tree Breed., Canberra. pp 1137—47 (1977). — Correll, R. L.: The optimal plot size for *Pinus radiata* establishment trails. Mathematical Project, South Australian Instit. of Technology. 40 pp (1978). — Cotterill, P.P.: A plan for breeding radiata pine. Silvae Genet. 33: 84—90 (1984). — Cotterill, P. P. and Zed, P. G.: Estimates of genetic parameters for growth and form traits in four *Pinus radiata* D. Don progeny tests in South Australia. Aust. For. Res. 10: 155—67 (1980). — Franklin, E. C.: Statistical validity of single-tree plot in forest genetics research. Silvae Genet. 20: 73—5 (1971). — Harris, D. L.: Expected and predicted progress from index selection involving estimates of population parameters. Biometrics 20: 46—72 (1964). — Hill, W. G.: Size of experiments for breed or strain comparisons. Proc. Work. Symp. Breed Eval. and Crossing Exp., Zeist. pp 43—54 (1974). — James, J. W.: Genetical considerations in large field experiments. In: Developments in Field Experiment Design and Analysis (Ed. V. J. Bofinger and J. L. Wheeler). Commonwealth Agric. Bureaux, Bulletin 50. pp 155—67 (1975). — Johnstone, R. C. B. and Samuel, C. J. A.: Experimental design for forest tree progeny tests with particular reference to plot size and shape. Proc. Joint IUFRO Meet., S.02.04.1-3, Stockholm. pp 357-76 (1974). — Jones, L. P.: Probability of correctly ranking strains. In: The Design and Conduct of Production Competitions for Comparison of Genotypes. Standing Committee on Agric., Canberra, Tech. Rept. Series No. 6, Appendix 2 (1980). — Kung, F. H.: Experimental design and analysis for half-sib progeny test. Proc., Meet. of Work. Party on Prog. Testing. Macon, Georgia. pp 82—102 (1972) — Lee, Chen Hui: Statistical efficiency in wood quality study based on a black pine plantation. Silvae Genet. 21: 249—50 (1972a). — Lee, Chen Hui: Statistical efficiency in black pine progeny test. Proc., Meet. of Work. Party on Prog. Testing, Macon, Georgia. pp 73—81 (1972b). — Robertson, A.: Optimum group size in progeny testing and family selection. Biometrics 13: 442—50 (1957). — Robertson, A.: The sampling variance of the genetic correlation coefficient. Biometrics 15: 469—85 (1959). — Shiue, Cherng-Jiann and Pauley, S. S.: Some considerations on the statistical design for provenance and progeny tests in tree improvement programs. Forest Sci. 7: 116—22 (1961). — Snedecor, G. W. and Cochran, W.G.: Statistical Methods. 6th Edit. The Iowa State Univ. Press, Ames, Iowa (1967). — Wright, J. W. and Freeland, F. D.: Plot Size and Experimental Efficiency in Forest Genetic Research. Michigan State Univ., Agric. Exp. Stat., Tech. Bull. 280. 28 pp (1960).

The Use of Competition Indices in Advanced-Generation Selection

By G. A. Tuskan[1]) and C. R. McKinley[2])

Texas A & M University, College Station, Texas 77843, USA

Summary

Two first-generation open-pollinated progeny tests of loblolly pine *(Pinus taeda* L.) were used to test the usefulness of four competition indices in advanced-generation selection procedures. In the five-year-old test all competition indices demonstrated an overall lack of competition. However, among the basal area classes within the five-year-old test there were differences in the competitive environment. Below basal area of 4.6 m²/ha there was no competition among adjacent trees, between basal areas 4.6 and 9.2 m²/ha competition appeared to stimulate growth, and above basal area 9.2 m²/ha competition acted to diminish growth. In the 15-year-old test, within all basal area classes, all four indices functioned well as indicators of the past competitive environment. The best overall indicator was the Proportional Area Index; the Weighted Tally appeared to be the most practical. The changes in the competitive environment at age five, and the consistently negative relationship between each of the above four indices and the subject tree measurements indicates that individuals within families could be selected based upon 1) the subject tree's measurements prior to the onset of competition and 2) those individuals, after the onset of competition, whose measurements occur above the regression line of the competition index predicting the subject tree's diameter or height.

Key words: Interspecific-competition, advanced-generation selection, *Pinus taeda,* progeny testing.

1) Assistant Professor, Dept. of Horticulture and Forestry, North Dakota University, Fargo, North Dakota 58105, USA.

2) Assistant Professor, Dept. of Forest Science, Texas A & M University.

Zusammenfassung

Die Ergebnisse aus zwei Nachkommenschaftsprüfungen bei *Pinus taeda* L. der ersten Generation wurden ausgewertet, um zu prüfen, in welcher Weise vier verschiedene Konkurrenzindizes in Selektionsverfahren für die folgenden Generationen nützlich sein können. In der Nachkommenschaftsprüfung im Alter 5 zeigte sich,, daß alle vier Indizes noch keine Konkurrenz aufzeigen konnten. Es konnten jedoch Unterschiede im Konkurrenzverhalten beobachtet werden, wenn die Bäume nach dem Querschnitt am Stammfuß klassifiziert worden waren. Unterhalb eines Stammquerschnittes von 4,6 m²/ha gab es keine Konkurrenz zwischen benachbarten Bäumen, bei Bäumen mit einem Stammquerschnitt zwischen 4,6 und 9,2 m²/ha scheint Konkurrenz das Wachstum zu fördern und bei einem Stammquerschnitt über 9,2 m²/ha vermindert Konkurrenz das Wachstum. In dem Test im Alter 15 konnten die vier Konkurrenzindizes die Konkurrenzsituation früherer Jahre gut beschreiben und zwar innerhalb aller Stammquerschnittsklassen. Der beste Konkurrenzindex war der Proportional Area Index, der Weighted Tally Index erschien aber am praktischsten. Die Änderung in der Konkurrenzsituation bei fünfjährigen Bäumen und die negative Be-