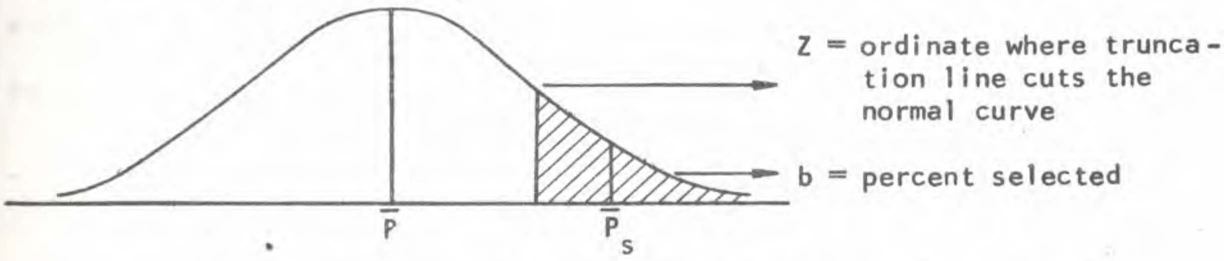
FOREST TREES AS A POPULATION OF GENOTYPES

Robert W. Touchberry

My views on forest tree genetics are those of an animal breeder, and not a forest tree breeder. They pertain mainly to variations between trees of the same species in the same plantation. Whether it be trees of different species or trees of the same species, the geneticist is concerned with estimating the breeding value of a tree or a group of trees under a rather loosely specified environment. He is interested in developing trees with breeding values equal to or exceeding those of trees now living.

Suppose that one goes into a stand of trees of a given age and measures (M) trees for some characteristic, such as diameter at a given height or fiber length at a given location. It is likely that there would be much variation between trees for each of these characteristics. Suppose that only the top 10 percent of these (M) trees were allowed to reproduce. How much would the average of the descendants of these trees $(\overline{P}_s)$ be expected to exceed the average of the descendants of all (M) trees $(\overline{P})$ if all (M) trees were allowed to reproduce? This expected genetic gain can be shown to be $\Delta G = b_{P_F I} \frac{Z}{b} \sigma_I$, where ΔG is the expected genetic gain and $b_{P_F I}$ is the regression of the performance of a future offspring (P_F) of a tree on the index (I) used to measure the performance of the mother tree. The fraction $\frac{Z}{b}$ can be obtained from tables giving the areas and ordinates of the standardized normal curve. It is the number of standard deviations by which the mean of the selected group exceeds the mean of the whole group. The diagram on the following page should explain the fraction $\frac{Z}{b}$. The standard deviation σ_I refers to the standard deviation of the index measures of the mother trees.



The regression $b_{P_F I}$ is a function of the genetic and environmental parts of the phenotypic variance (σ_P^2) of the (M) mother trees and the correlation between I and the genotype of the mother trees.

The Genetic Gain When Asexual Reproduction is Involved

The phenotypic variation (σ_P^2) between trees can be expressed as follows: $\sigma_P^2 = \sigma_H^2 + \sigma_E^2 + \sigma_{HE}^2$, where σ_H^2 is the hereditary variance, σ_E^2 the environmental variance, and σ_{HE}^2 the variance associated with the interaction between heredity and environment. The phenotypic variance (σ_P^2) cannot be divided into its component parts just by measuring the trees and deriving the variance between trees.

If one could hold the environment absolutely constant, then the variation between trees would be $\sigma_H^2 + \sigma_{HE}^2$. This state can be approached by making (K) cuttings of each of the (M) trees, rooting them, and setting them in a suitable field plot design. After removing the environmental variation associated with blocks, one can then obtain the following mean squares:

Source of variation	E.M.S.
Between clones	$\sigma_E^2 + K\,\sigma_C^2$
Between trees of the same clone	σ_E^2

The component of variance for clones (σ_C^2) then approaches $\sigma_H^2 + \sigma_{HE}^2$.

Now if one takes these same (M) clones and sets (K) cuttings of each clone at (L) different locations within a region, then one can obtain estimates of σ^2_H and σ^2_{HE}. For example, one could obtain the following mean squares:

Source	E.M.S.
Locations	$\sigma^2_E + K\,\sigma^2_{LC} + K\,M\,\sigma^2_L$
Clones	$\sigma^2_E + K\,\sigma^2_{LC} + K\,L\,\sigma^2_C$
Locations x clones	$\sigma^2_E + K\,\sigma^2_{LC}$
Within subclasses	σ^2_E

The component σ^2_C is an estimate of σ^2_H, and σ^2_{LC} is an estimate of σ^2_{HE}.

Now let's consider the correlation between the genotype of the mother tree and the average performance of the (K) or (L) (K) cuttings made from that tree.

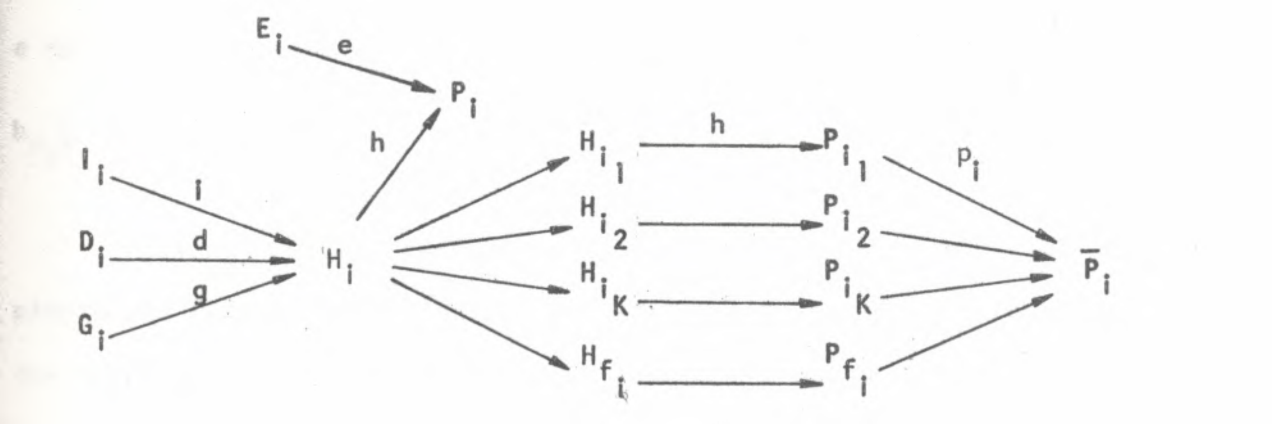


Figure 1. A path diagram of the relationship between the genotype of a mother tree (H_i) and the average of the phenotypes (Pi) of (K) cuttings from that tree. h = the square root of heritability and p_i = $\frac{1}{\sqrt{K\,[1 + (K-1)\;t]}}$, where $t = r_{P_{ij}P_{ik}}$; $\sigma_{H_i} = h\,\sigma_P$; $\sigma_{\bar{P}} = \sigma_P\sqrt{\frac{1+(K-1)\,t}{K}}$; $b_{H_i\bar{P}_i} = r_{H_i\bar{P}_i}\frac{\sigma_{H_i}}{\sigma_{\bar{P}_i}} = h^2\left[\frac{K}{1 + (K-1)\;t}\right]$.

Now if by the experimental design used in planting the cuttings (H_i), we have successfully removed any environmental correlation between the phenotypes P_{ij}, then $t = h^2$. If all of the cuttings are planted at one location, $h^2 = \frac{\sigma^2_H + \sigma^2_{HE}}{\sigma^2_H + \sigma^2_{HE} + \sigma^2_E}$, and the regression of the genotypes of the mother trees on the average performance of the cuttings is $b_{H_i\overline{P}_i} =$

$$\frac{\sigma^2_H + \sigma^2_{HE}}{\sigma^2_H + \sigma^2_{HE} + \frac{\sigma^2_E}{K}}.$$

The regression of the genotypes of the mother trees (H_i) on the $\overline{P}_i$ is shown by line A in Figure 2 for various numbers of cuttings (K) per tree and assuming that σ^2_H accounts for 40 percent of the phenotypic variance, σ^2_{HE} 10 percent, and σ^2_E 50 percent.

The regression of the performance of a future cutting (P_{F_i}) from a mother tree on the average performance $(\overline{P}_i)$ of (K) previous cuttings is

$$b_{P_F\overline{P}_i} = r_{P_F\overline{P}_i} \frac{\sigma_{P_F}}{\sigma_{\overline{P}_i}} = \frac{\sigma^2_H + \sigma^2_{HE}}{\sigma^2_H + \sigma^2_{HE} + \frac{\sigma^2_E}{K}}.$$

If (K) of the cuttings from each of the (M) mother trees are planted at each of (L) locations within a region, then the regression of the genotypes of the mother trees (H_i) on the average performances of the (K) (L) cuttings $(\overline{P}_i)$ is $b_{H_i\overline{P}_i} = \frac{\sigma^2_C}{\sigma^2_C + \frac{\sigma^2_{CL}}{L} + \frac{\sigma^2_E}{KL}}$, which approaches

$$\frac{\sigma^2_H}{\sigma^2_H + \frac{\sigma^2_{HE}}{L} + \frac{\sigma^2_E}{KL}}$$ as (L) increases.

For those environmental factors that are alike at all locations, σ^2_C would include σ^2_{HE}. As (L) increases, the probability that environmental factors will be alike at all locations decreases; consequently, σ^2_C approaches σ^2_H and σ^2_{CL} approaches σ^2_{HE}.

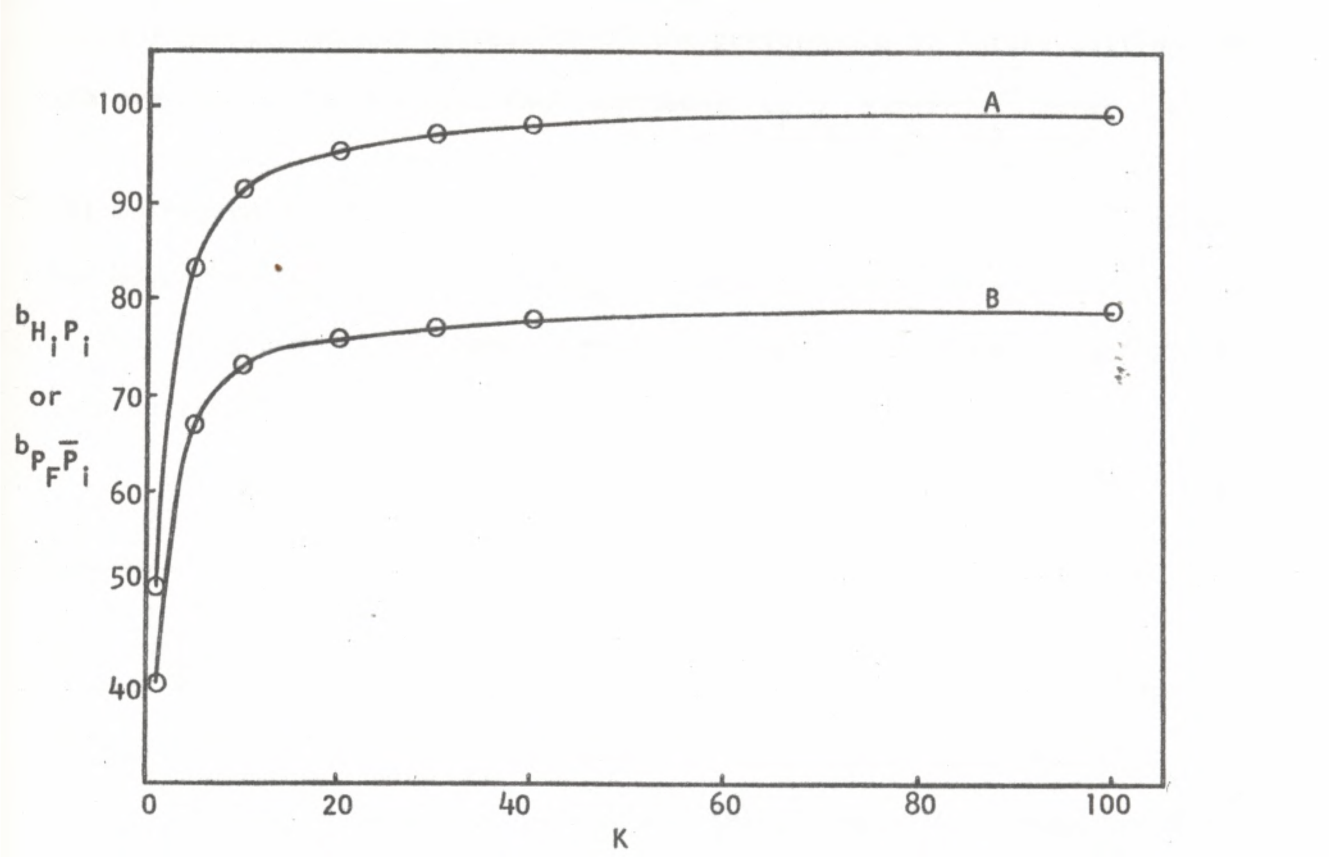


Fig. 2. The regressions of the genotype of the mother tree and the phenotype of future cuttings of the mother tree on the average of K cuttings from mother tree.

$\sigma_H^2 = .40$, $\sigma_{HE}^2 = .10$, $\sigma_E^2 = .50$.

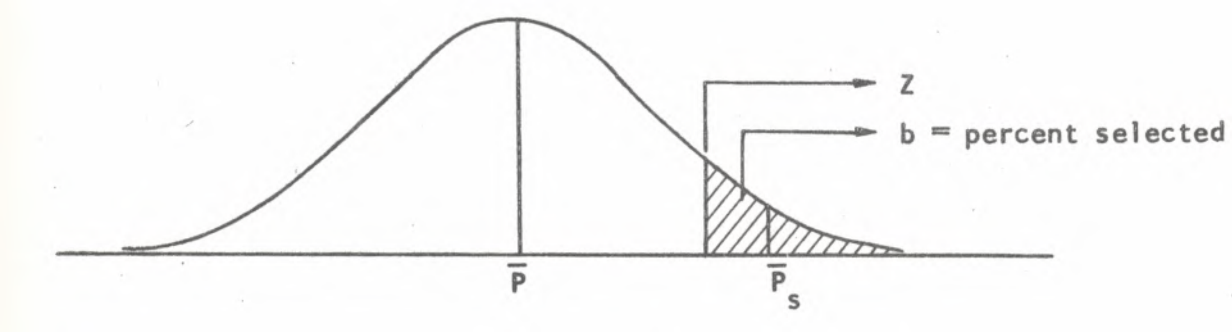
Plotted Values

K	$b_{H_i \bar{P}_i}$	$b_{P_F \bar{P}_i}$
1	.50	.40
5	.83	.67
10	.91	.73
20	.95	.76
30	.97	.77
40	.98	.78
100	.99	.79

If the test cuttings are all planted at one location, but future cuttings of the (M) mother trees are planted at various locations throughout a region, then the regression of the performance of future cuttings on the average of the test cuttings approaches $b_{P_F\bar{P}_I} = \frac{\sigma_H^2}{\sigma_H^2 + \sigma_{HE}^2 + \frac{\sigma_E^2}{K}}$. Line (B) in Figure 2 shows the regression of P_F on $\bar{P}_i$, assuming that σ_H^2 accounts for 40 percent of the variance, σ_{HE}^2 10 percent, and σ_E^2 50 percent. Again it should be emphasized that the tests are made at one location and the future cuttings are to be planted at other locations.

The real regression is probably between lines A and B of Figure 2, $\frac{\sigma_H^2}{\sigma_H^2 + \sigma_{HE}^2 + \frac{\sigma_E^2}{K}} < b_{P_F\bar{P}_i} < \frac{\sigma_H^2 + \sigma_{HE}^2}{\sigma_H^2 + \sigma_{HE}^2 + \frac{\sigma_E^2}{K}}$, as it is not likely that all of σ_{HE}^2 is ever removed from the numerator or that all of σ_{HF}^2 is ever included in the numerator.

Suppose that we go back to our (M)- tested mother trees and consider only the top 10 percent of those trees. How much would we expect the mean diameter or fiber length of the top 10 percent (Ps)to exceed the mean. of all (M) trees $(\bar{P})$. This can be obtained by the formula $\bar{P}_S - \bar{P} = \frac{Z}{b}\sigma_I$



where Z is the ordinate where the line truncating the top b percent cuts the normal curve and σ_1 is the standard deviation of the index used to measure the performance of the mother trees. For example, if we use the mean diameter of (K) cuttings of each of the (M) trees as a measure of diameter at a given age, then σ_I is $\sigma_P \sqrt{\frac{1 + (K-1)\ t}{K}}$, where σ_P is the standard deviation of the diameters of the test cuttings at a given age.

If in the future we plant only cuttings from the top 10 percent of the mother trees, how much would we expect the mean of the future cuttings from the top 10 percent (Psf) to exceed the mean of future cuttings from all other trees ($\bar{P}$). This can be obtained by the formula $\bar{P}_{SF} - \bar{P} = h^2 \frac{Z}{b} \sigma_I$, where h^2 is the heritability of the characteristic involved.

When the characteristic is measured directly on the mother tree and selected on the basis of the mother tree, the gain is proportional to

$$\frac{\sigma^2_H}{\sigma^2_H + \sigma^2_{HE} + \sigma^2_E} (1.755) \sqrt{\sigma^2_H + \sigma^2_{HE} + \sigma^2_E}.$$

When the characteristic is measured by taking the mean diameter of (K) test cuttings of each mother tree, the gain is proportional to

$$h^2 \frac{Z}{b} \sigma_I = \frac{\sigma^2_H}{\sigma^2_H + \sigma^2_{HE} + \frac{\sigma^2_E}{K}} (1.755) \sqrt{\sigma^2_H + \sigma^2_{HE} + \frac{\sigma^2_E}{K}} = \frac{\sigma^2_H\ (1.755)}{\sqrt{\sigma^2_H + \sigma^2_{HE} + \frac{\sigma^2_E}{K}}}.$$

Comparing the relative progress of the two, we get:

$$\frac{\dfrac{\sigma^2_H}{\sqrt{\sigma^2_H + \sigma^2_{HE} + \sigma^2_E}}}{\dfrac{\sigma^2_H}{\sqrt{\sigma^2_H + \sigma^2_{HE} + \frac{\sigma^2_E}{K}}}} = \frac{\sqrt{\sigma^2_H + \sigma^2_{HE} + \frac{\sigma^2_E}{K}}}{\sqrt{\sigma^2_H + \sigma^2_{HE} + \sigma^2_E}}.$$

With the numerical values for σ^2_H, σ^2_{HE}, and σ^2_E that were assumed, it can be shown that selecting on a basis of the mother tree's own phenotype is from 50 to 100 percent as effective as selecting on the basis of the average of K cuttings from that tree.

The Genetic Gain When Sexual Reproduction Is Involved

In most cases it is likely that seed from the selected trees is going to be harvested and planted to reproduce the future trees. The correlation between mother trees and seedlings is less than that between mother trees and cuttings. When reproduction is by cuttings (asexual reproduction), we are concerned with heritability in the broad sense $\left(\frac{\sigma_H^2}{\sigma_P^2}\right)$. When reproduction is from seed (sexual reproduction), we are concerned with heritability in the narrow sense $\left(\frac{\sigma_G^2}{\sigma_P^2} = h^2 g^2\right)$. The hereditary variance (σ_H^2) can be divided into σ_G^2, the additive genetic variance; σ_D^2, the variance caused by dominance deviations; and σ_I^2, the variance caused by epistatic deviations. Heritability in the narrow sense is that fraction of the phenotypic variance accounted for by the additive genetic variance.

What is the correlation between the average of (K) offspring of a mother tree and the breeding value of the mother tree? it is assumed that a random sample of pollen from many trees collected on the stigmata of the mother trees or that mating was random. Figure 3 illustrates the correlation between the mean of the seedlings and the breeding value of the mother tree.

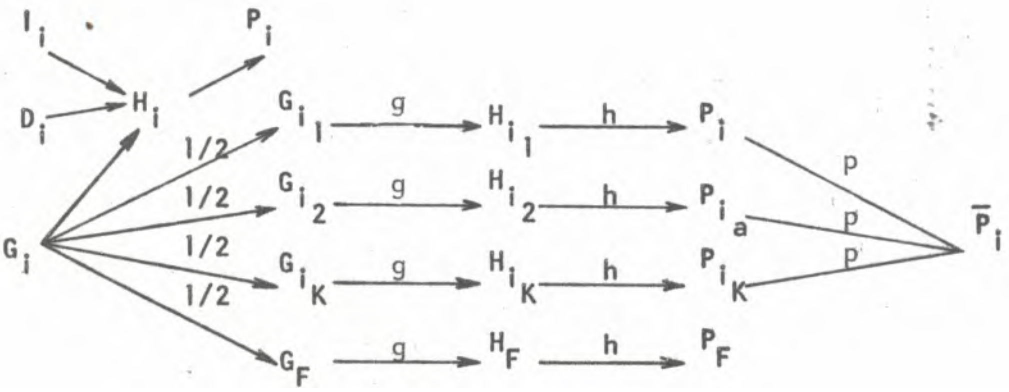


Figure 3. Path coefficient diagram of the relationship between the breeding value (G1) of a mother tree and the average Pi of (K) of its seedling offspring.

$r_{G_i\bar{P}_i} = K\ (1/2)\ (h)\ (g)\ p$; t = correlation between P_{ij} and p = $\frac{1}{\sqrt{K\,[1 + (K-1)\ t]}}$ so $r_{G_i\bar{P}_i} = \frac{hg}{2}\sqrt{\frac{K}{1 + (K-1)\ t}}$. If by planting in an appropriate design the environmental correlations between the P_{ij} are removed, then $t = \frac{h^2g^2}{4}$ or 1/4 of heritability in the narrow sense. Then

$$r_{G_i\bar{P}_i} = \frac{hg}{2}\sqrt{\frac{K}{1 + (K-1)\ \frac{h^2g^2}{4}}} \quad \text{and} \quad b_{G_i\bar{P}_i} = r_{G_i\bar{P}_i}\ \frac{\sigma_{G_i}}{\sigma_{\bar{P}_i}} =$$

$$\frac{h^2g^2}{2}\ \frac{K}{1 + (K-1)\ \frac{h^2g^2}{4}}.$$

How can heritability in the narrow sense be estimated? Suppose that (K) seedlings from each of (M) mother trees are planted in an appropriate field plot design at one location. From such an experiment, the mean squares below can be obtained:

Source	Expected mean square
Between half-sib groups	$\sigma_E^2 + K\ \sigma_S^2$
Within half-sib groups	σ_E^2

If the design has been successful in removing the random environmental variations, then σ_S^2 approaches $1/4\ \sigma_G^2 + 1/16\ \sigma_{AA}^2 + 1/64\ \sigma_{AAA}^2$, etc. The terms σ_{AA}^2 and σ_{AAA}^2 refer to the epistatic variance arising from the interaction of additive deviations. Since all but the first term of the above expression are likely to be small, then $\sigma_S^2 \longrightarrow 1/4\ \sigma_G^2$. Likewise, σ_W^2 approaches 3/4 of $\left[\sigma_G^2 + \sigma_D^2 + \sigma_I^2 + \sigma_E^2 + \sigma_{HE}^2\right]$ or $\left[\sigma_P^2 - 1/4\ \sigma_G^2\right]$. Consequently, 4t is an estimate of heritability in the narrow sense. 4t =

$$\frac{4\ \sigma_S^2}{\sigma_S^2 + \sigma_W^2} = \frac{\sigma_G^2}{\sigma_P^2}.$$

The regression of the breeding value of a mother tree (G_i) on the average performance of (K) of its seedling offspring is $b_{G_i\bar{P}_i} = \dfrac{\sigma_G^2}{2\ \sigma_P^2}$

$$\frac{K}{1 + (K-1)\dfrac{\sigma_G^2}{4\ \sigma_P^2}} = \frac{2\ K\ h^2 g^2}{4 + (K-1)\ h^2 g^2}.$$

The regression of the phenotype of a future seedling P_F of the same mother tree G_i on $\bar{P}_i$ is $b_{P_F\bar{P}_i}$ =

$$\frac{K\ \sigma_G^2}{4\ \sigma_P^2 + (K-1)\ \sigma_G^2}.$$

If σ_G^2 is .25 and $\sigma_P^2 = 1.00$, then the regression of G_i on $\bar{P}_i$ is $b_{G_i\bar{P}_i} = \dfrac{2\ K\ (.25)}{4 + .25\ (K-1)}$.

In Figure 4 the size of these regressions is shown for various values of (K). If heritability in the narrow sense increases or decreases, then the regressions increase or decrease, but $b_{P_F\bar{P}_i}$ remains exactly one-half of $b_{G_i\bar{P}_i}$. As (K) becomes infinitely large, $b_{G_i\bar{P}_i}$ approaches 2 and $b_{P_i\bar{P}_i}$ approaches 1.

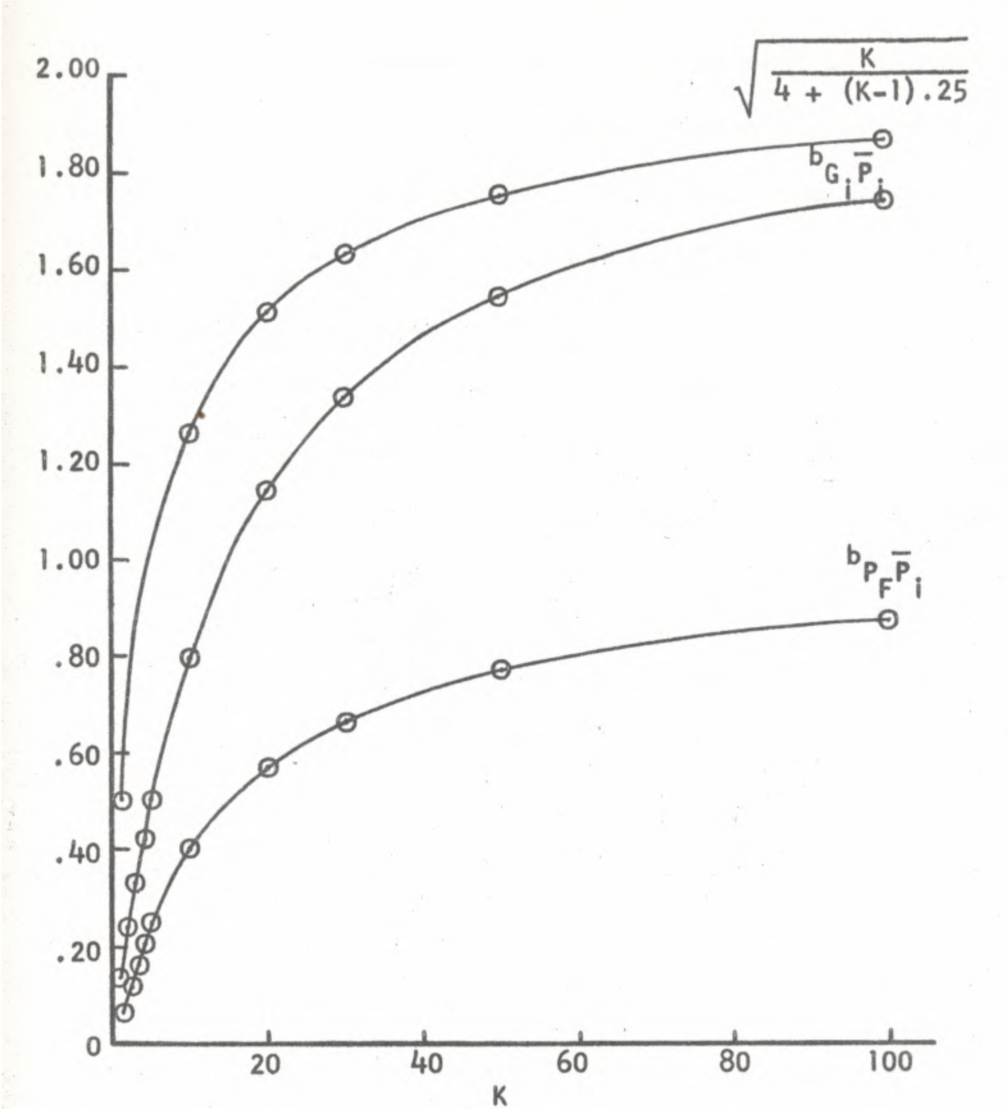


Fig. 4. The regression of the breeding value of the mother tree (Gi) on the average performance of K seedling daughters of the mother tree (Pi) and the regression of a future seedling daughter (P_F) on the average of K seedling daughters of the same mother tree.

Plotted Values

K	$b_{G_i\bar{P}_i}$	$b_{P_F\bar{P}_i}$	$\sqrt{\frac{K}{4+(K-1).25}}$
1	.125	.0625	.50
2	.24	.120	--
3	.33	.165	--
4	.42	.210	--
5	.50	.25	--
10	.80	.40	1.26
20	1.14	.57	1.51
30	1.33	.665	1.63
50	1.54	.77	1.75
100	1.74	.87	1.86

Now if we choose the best 10 percent of the mother trees on the basis of the phenotypes of the mother trees themselves and produce seedlings from seed of these trees, the expected genetic gain will be $\overline{G}_{P_S} - \overline{G}_{P_I} = \frac{h^2 g^2}{2} \frac{Z}{b} \sigma_P$. If we assume $\sigma_P^2 = 1$ and $\sigma_G^2 = .25$, then the expected gain = .125 (1.755) (1.0).

If we choose the best 10 percent of the mother trees on the basis of the average of 20 seedlings of each mother tree and then plant only the seed from the top 10 percent of these trees, the genetic gain will be $\frac{K \sigma_G^2}{4 \sigma_P^2 + (K-1) \sigma_G^2} \frac{Z}{b} \sigma_P \sqrt{\frac{1 + (K-1) t}{K}} = \frac{.25}{2} \sqrt{\frac{K}{4 + (K-1) .25}} (1.755)$.

From a comparison of these two, it can be seen that progress per generation from selection based on seedling progeny tests is $\sqrt{\frac{K}{4 + (K-1) .25}}$ times as great as selection based on the phenotypes of mother trees. A line for $\sqrt{\frac{K}{4 + (K-1) .25}}$ is plotted in Figure 4 for various values of (K). As heritability in the narrow sense increases, the relative value of a progeny test as an aid to selection decreases.

In general, selection on a basis of progeny tests increases the generation interval over that when selection is based on the phenotypes of the mother trees. If the generation interval with progeny testing is equal to or more than $\sqrt{\frac{K}{4 + (K-1) h^2 g^2}}$ times the generation interval when selection is based on the phenotypes of the mother trees, then progress per year is greater for selection based on the phenotypes of the mother trees.

Genetic Gain When Many Characteristics Are Involved

We have been talking so far about selecting for one characteristic at a time. But the forester would be interested in improving several characteristics. Suppose, for example, that he is selecting for five independent characteristics. What fraction of trees would be in the top 10 percent for all characteristics? $(1/10)^5 = 1/100,000$. If there were 10 characteristics, the fraction would be 1/10,000,000,000. Finding such elite trees would be a tremendous job. Just to look at 10,000,000,000 trees, allowing 1 minute per tree, would require 8,148 man years for 8 hours per day 5 days per week. One might even question whether such an elite tree actually exists.

This suggests that the forester should be using an index for choosing seed trees. This index should be constructed from the phenotypic and genetic variances and covariances of the various characteristics and their relative economic values. This would seem to be the most efficient way of simultaneously selecting for many characteristics.

Such an index would be $I = A_1\hat{G}_{x_1} + A_2\hat{G}_{x_2} + \dots + A_n\hat{G}_{x_n}$, where the A_i are the relative economic values and the G_{x_i} are the estimated genetic or hereditary values.

Each hereditary value can be estimated from a set of multiple regression equations:

$$\sigma^2_{x_1} b_{11} + \sigma_{x_1x_2} b_{12} + \dots \sigma_{x_1x_n} b_{1n} = \sigma^2_{g_1}$$

$$\sigma_{x_1x_2} b_{11} + \sigma^2_{x_2} b_{12} + \dots + \sigma_{x_2x_n} b_{1n} = \sigma_{g_1g_2}$$

$$\sigma_{x_1x_n} b_{11} + \sigma_{x_2x_n} b_{12} + \dots + \sigma^2_{x_n} b_{1n} = \sigma_{g_1g_n}$$

By solving this set of equations, we would obtain values for b_{11} b_{12} . . . + b_{1n} and we could write the equation $\hat{G}_1 = b_{11}X_1 + b_{12}X_2 + b_{13}X_3 + \dots + b_{1n}X_n$. Similarly, we could find prediction equations for the other $\hat{G}_i$ and then multiply each $\hat{G}_i$ by its corresponding A_i and have a logical index for choosing trees.

To me it seems that a forester must know something about the relative sizes of σ^2_G, σ^2_H, σ^2_{HE}, σ^2_E, and σ^2_P for the characteristics with which he is concerned and the genetic and phenotypic covariances between these characters. Until he knows these, I don't see how he can intelligently embark on a tree breeding program.