LONG-TERM TREE BREEDING

Hyun Kang[1/]

Abstract.--Analysis of variance techniques are not useful for developing long-term tree breeding strategies. Therefore, tree breeders must use the information obtained from animal and crop experiments, and population genetics theories to develop long-term tree breeding strategies.

Additional keywords: Population size, gene model.

Many tree breeders are vitally interested in developing advanced generation breeding techniques because their breeding populations are entering the second generation. The development of such techniques is complex and requires a classification of the associated problems. Presented in this paper are examples of the information obtained from crop and animal genetic experiments and population genetics theories that can be used to develop long-term tree breeding strategies.

ANALYSIS OF VARIANCE

During the past 30 years, we have learned a great deal about analysis of variance. It has become evident that the existence of genetic variance is fundamental to successful tree breeding. We have also learned how to design mating schemes and field experiments and how to predict genetic gains for both direct and indirect selection systems. These techniques are designed to develop a picture of the variance structure of the breeding population(s) at a given point in time. The information obtained has then been used to help make breeding decisions for the immediately following generation. However, any linear projection of future population improvement based on the analysis of the picture at a given point in time is subject to severe errors if applied for more than three or four generations because the evolutionary process of populations is so complex and dynamic.[2/] Furthermore, the analysis of variance classifies variances but does not describe the causes of the variances; consequently, many breeding principles cannot be developed by the use of analysis of variance.

TREE BREEDING BY INTUITION

Assuming that the proposition about the shortcomings of the analysis of variance is valid, are there other approaches which are as suitable to develop long-term tree breeding strategies in a fashion similar to the way the analysis of variance was to short-term tree breeding plans? The answer

---

[1/] Population Geneticist, Forestry Sciences Laboratory, North Central Forest Experiment Station, USDA, Forest Service, Rhinelander, Wisconsin 54501.

[2/] Many discussions of this kind can be found in articles published by Alan Robertson between 1960-1970. For example, Robertson (1961).

depends on the definition of the term "suitable". If it represents the accuracy of the point estimates of target parameters of the future, then the answer is "no". In developing long-term tree breeding strategies, the target parameters are not as important as they are in short-term tree breeding. For example, it is not realistic to ask what the average biomass production of loblolly pine at a typical North Carolina piedmont would be 4,000 years from now. It is more important to determine how to condition the breeding population so that it will express its maximum potential at the time of equilibrium while maintaining the capability to absorb shocks such as sudden changes in environments, intensity of forest management, breeding objectives, etc. Therefore, in long-term tree breeding, we are more interested in finding out the sufficient minimum conditions necessary for breeding rather than in setting definite targets for the future.

A good alternative to analysis of variance in long-term tree breeding appears to be the collective intuition of breeders and population geneticists. Although we might think that it is old, simple, and often irrelevant, collective intuition was the foundation on which we developed our estimation theory.

SOURCES OF INTUITION

To develop a long-term tree breeding strategy we must free our minds momentarily from the breeding populations and specific objectives we now have. Even if we know the genetic properties of our present breeding populations, they represent only a single static point of all the possible combinations of the factors within the boundary conditions. Such an understanding is not of much help in learning and using the population dynamics necessary for long-term breeding.

We can improve our intuition for advanced generation breeding in two ways: (1) by evaluating long-term breeding results obtained from crop and animal experiments, and (2) by learning and applying theoretical population genetics.

The breeding conditions of a crop (or animal) breeding experiment are likely to be different from that for trees. Likewise, the conditions are likely to be different among different crop (or animal) experiments. Therefore, results from a single experiment normally are not sufficient for making generalizations. It is possible, however, to find a more deterministic common denominator when several experimental results are assessed. A good example of this is from recurrent selection studies where many published results indicate that on the average the improvement of two phenotypic standard deviations of generation zero is easily obtained before the population average plateaus, regardless of the initial conditions and breeding techniques used. The example on selection for high oil content in corn from the textbook by Allard (1960) is a representative case of recurrent selection results (Figure 1). The figure vividly shows what is possible through recurrent selection. The area of overlap between the performances of individuals at generation zero and at generation ten is very small. Clearly, a tree breeder at generation ten cannot afford to go back to a natural population(s) for addition of genes to keep the genetic base broad because it will be difficult to find an individual that is as good as the average of the individuals in the tenth generation breeding population. Some other facts suggested by the graph are: (1) Continuous introduction of new individuals into

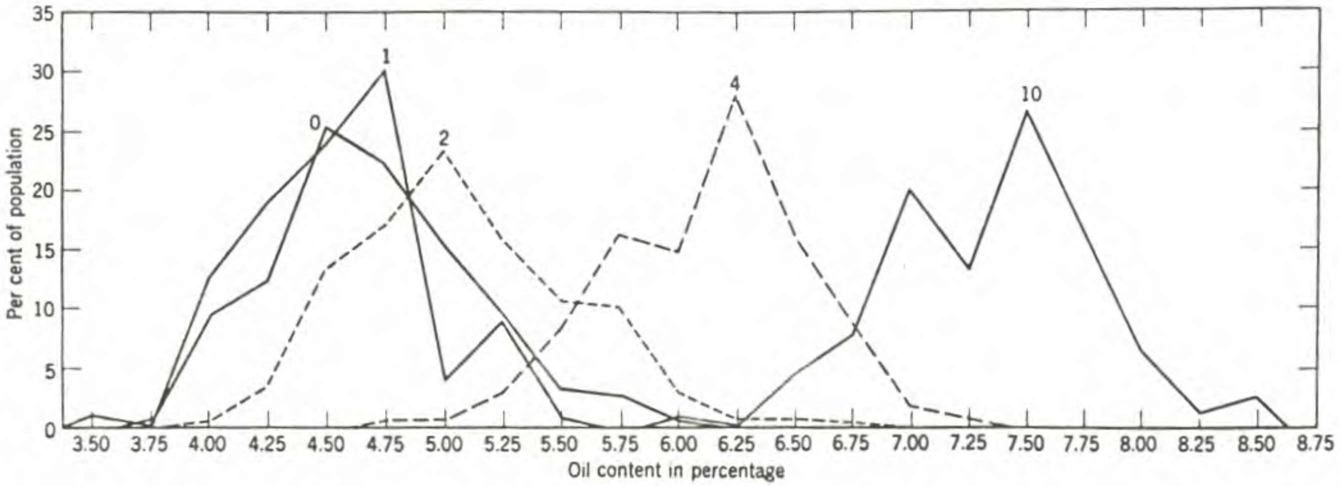

FIGURE - **1.** Selection for high oil content in corn. The frequency distribution of the original population is designated by 0 and the frequency distributions after various numbers of generations of selection by numbers. The mean oil content increased steadily with selection. Transgressive segregants appeared in the second-selected generation, and by the tenth-selected generation all individuals of the selected generations transgressed the original population. (After Smith, 1908.)

The above is an exact replica of Figure 16-1 of Allard (1960).
The reference for Smith (1908) from which it came is not found in the book.
(Reprinted by permission of John Wiley & Sons, Inc.).

the breeding population (open breeding population) is not a desirable option because it slows down the speed of the average population advance. (2) Repeated vegetative propagation preceded by selection is not an efficient long-term tree breeding method because the maximum gain after ten cycles of selection is limited by the distribution of the base population. (3) The maintenance of parallel breeding populations (whose ancestors may or may not overlap) under similar environments is a potential way of preparing for the future gene exchange need.

The recurrent selection research results, however, are not very useful for determining the necessary breeding population size. Most of the studies are not designed to answer this question and also it is not possible to obtain answers to some questions related to population size by experimental means.

Breeding population size can be determined through the use of population genetics theory. It is important, however, to understand that recommendations based on theoretical population genetics cannot be stated as positively as those of the previous example. The theories are developed from simple genetic models, while the structure of a real population is complex.[3/] As long as the limitation of the model is made clear from the beginning, population genetics theories are useful for improving intuitive breeding programs for forest trees.

Four simplified models will be used to illustrate the necessary population size for breeding. The breeding population size represents the number of selected parents used for mating.

(a) Major gene model

Under this model, the phenotypic expression of the trait of interest is determined by the genotype without environmental modification. If one locus influences the trait and the alleles act additively, the necessary population size will be one individual for monoecious species and two individuals for dioecious species. The population will reach its maximum potential in the trait of interest in one generation.

(b) Minor (quantitative) gene model

Under this model, a trait is influenced by genes at more than one loci and the expression of the genes is modified by the environment. If no linkage and no epistasis among genes at different loci is assumed (a simplifying but rather unrealistic assumption), it is possible to interpret the performance of the trait of interest in terms of the fixation probability of a favorable allele at a single locus. The probability is determined by initial gene frequency (q), degree of dominance (h), population size (N), and selection coefficient(s). Under gene action models other than over-dominance, the trait average is maximum when all the loci are fixed with the most desirable allele at each locus. The formula for computing the fixation probability

---

[3/] A model does not have to be as complex as the real object it imitates; the main advantage of using models lies in their simplicity to show general trends.

(u(q)) is:

$$u(q) = \frac{\int_0^q \exp[-2Ns(2h-1)(1-x)x-2Nsx]dx}{\int_0^1 \exp[-2Ns(2n-1)(1-x)x-2Nsx]dx} \qquad \text{Kimura (1957)}$$

For simplicity, the formula can be evaluated as an additive model only (Table 1). The necessary population size decreases as the selection proportion decreases and the initial gene frequency increases. The necessary population size increases more rapidly between initial gene frequency .01 and .05 than for any other interval. Therefore, if we define the economic break-even point of initial gene frequency as .05 and ignore the rare alleles with smaller frequencies, we can conclude that the necessary population size for a minor gene is 85. A good review of this subject was made by Rawlings (1970).

Table 1. The necessary population size to fix the desirable allele in the population with 95% assurance*

Heritability .2
Number of loci 100

Initial gene freq.(%)	Proportion selected (%) under a truncation system				
	1	5	10	25	50
.01	281	364	426	590	937
.05	56	73	85	118	187
.1	28	36	43	59	94
.25	11	15	17	24	38
.5	6	7	8	12	18
.75	3	4	5	6	10
.9	-	2	2	3	4

*Additional steps, which were not detailed in the text, were used to develop the table using Kimura's formula.

(c) Independent neutral gene model

The neutral gene model is important to gene conservation. Many genes appear to be selectively neutral to a trait of interest, but they could play an important role if breeding conditions, product desires, or environment change. Because these genes are selectively neutral, the average probability of loss is not influenced by selection or by the population size. The probability of loss is 1-q. The population size, however, influences the speed of gene loss -- i.e., the smaller the population, the quicker the allele will be lost from the population (average). We can take advantage of the difference in rate to determine the necessary breeding population size. For example, a breeder might be willing to ignore the consequences of having a small population after a certain number of generations, say 100 generations or 4,000 years. Under this assumption the necessary population size can be obtained by modifying a formula given by Kimura and Ohta (1969) as follows:

$$N = -\frac{t(1-q)}{4q(\log_e q)}$$

This formula is subject to the condition that N -¼- (Table 2), If we have neutral genes with q = .05, it will be necessary to have N 159 to maintain this gene in the population for 100 generations,

Table 2. Population size necessary to maintain neutral genes with initial frequency q for t generations*

t (gen)	Initial frequency of neutral genes (q)								
	.005	.01	.02	.03	.04	.05	.1	.2	.5
10	-	-	-	-	-	-	10	6	4
20	-	107	63	46	37	32	20	12	7
30	282	161	94	69	56	48	29	19	11
40	376	215	125	92	75	63	39	25	14
50	469	269	157	115	93	79	49	31	18
60	563	322	188	138	112	95	59	37	22
70	657	376	219	161	130	111	68	43	25
80	751	430	251	184	149	127	78	50	29
90	845	484	282	207	168	143	88	56	32
100	939	537	313	231	186	159	98	62	36

*t: Average extinction time in generations.

(d) Linked neutral gene model[3]

In this model a neutral gene is linked with a selective locus. Namkoong and Roberds (In preparation) suggested that when the selective locus has additive gene effect,a good general rule of thumb would be to double the size of the population required when an independent neutral gene model was used. In our example the necessary population size would be 318.

All the gene action types described by the four models are likely to exist simultaneously in a real population. Therefore, the breeder can choose from a large range of population sizes. The size selected will depend on which model is given the most emphasis. A conservative breeder may wish to have a large population size, say 318, to maintain neutral genes in the population for many generations. But this size would be unnecessarily large if model b was given the most emphasis. On the other hand, the breeder could pick 85 as the breeding population size, but this number would be too small for models c and d. Therefore, the choice of population size is subjective.

The application of results obtained from the models is limited, and the population sizes discussed above should not be taken as recommendations. There is no reason why the population size should be 318, 85, etc., and the numbers should be considered as bits of information breeders can use to improve their intuition to breed trees. For example, breeders agree that gene conservation can be achieved by maintaining a large breeding population, but

---

[3] This model was suggested by Dr. Gene Namkoong, and the author is grateful to him for making the information available before its publication.

the definition of "large" ranges from 20 to multi-thousands, The study of models c and d do not generate a fixed number that would define a large population. By studying the models, the breeders can appreciate the fact that the sufficient breeding population size would most likely be in the range of hundreds.

Today, tree breeders are at a crossroad. They have begun to realize that they have to evaluate their breeding systems in respect to time (long-term breeding) as well as space (single-generation breeding). The addition of this new dimension leaves them with a great deal more uncertainties than before. Yet, because experimental means are not available, they have to depend entirely on intuition to develop future breeding population(s). We can improve our intuition by learning the success and failure of crop and animal breeders, and by studying population dynamics.

ACKNOWLEDGMENTS

I thank Drs. Bruce Zobel and Gene Namkoong for their comments on the manuscript.

LITERATURE CITED

Allard, R. W. 1960. Principles of Plant Breeding. John Wiley and Sons, N.Y. 485 pp.

Kimura, M. 1957. Some problems of stochastic process in genetics. Ann. Math. Stat. 28: 882-901.

Kimura, M., and Ohta, T. 1969. The average number of generations until fixation of a mutant gene in a finite population. Genetics 61: 763-771.

Namkoong, G., and Roberds, J. H. Short-term loss of neutral alleles in small-population breeding. (In preparation),

Rawlings, J. O. 1970. Present status of research on long and short-term recurrent selection in finite populations. Proceedings of the Working Group on Quantitative Genetics, Section 22, IUFRO. August 18-19, 1969, Raleigh, North Carolina, pp. 1-15.

Robertson, A. 1961. Some comments on quantitative genetic theories. In W. D. Hanson and H. F. Robinson (eds.) Statistical genetics and plant breeding, p. 108-115. NAS-NRC, Washington, D.C.