

# Population size in segregating generations in cotton breeding programs

*Daryl T. Bowman*

*Crop Science Department, North Carolina State University, Raleigh NC USA*

*Correspondence author [daryl\\_bowman@ncsu.edu](mailto:daryl_bowman@ncsu.edu)*

## ABSTRACT

*Private and public cotton breeders in the U.S.A. make around 100 genetic combinations annually. In the first segregating generation ( $F_2$ ), an average of 1000 plants are examined. This number then doubles in the next generation. With cotton having 52 chromosomes, are 1000  $F_2$  plants sufficient to begin selection? Will there be a great probability that a transgressive segregate is identified? Is linkage so tight that small numbers of progeny is sufficient or does that dictate very large numbers of progeny? In other words, is the current approach of making large numbers of genetic combinations with few progeny per cross the most efficient breeding method for upland cotton? If we assume that we have information on the breeding value of the parents, e.g. yield performance, then we may make fewer genetic combinations and look at larger progeny numbers. If we assume we have no information on breeding value then we may want to make more crosses and look at fewer progeny per cross. The probability of producing a transgressive segregate is proportional to the number of loci involved, linkage, and crossing over rates. The data suggests that populations much larger than 1000  $F_2$  plants are required or that selection begin in the  $F_3$  or later generation when the population number can be expanded and homogeneity is approached.*

## Introduction

Plant breeders must decide what approach to take in their programs given the resources available. Should they make a large number of crosses (genetic combinations) and look at few progeny from each cross? Should they make fewer crosses and look at more progeny per cross? Does the size of the segregating generation depend on the information on each parent?

Typically, plant breeders make many genetic combinations each year. The International Centre for Maize and Wheat Improvement (CIMMYT) makes over 12,000 genetic combinations (crosses) each year (Maredia and Byrlee, 1999) while the International Rice Research Institute made over 2,000 genetic combinations (IRRI, 1985), and the Plant Breeding Institute, UK made over 1200 genetic combinations annually during the 1980s (Bingham and Lupton, 1987). In the U.S.A., the typical private cotton (*Gossypium hirsutum* L.) breeder makes 105 genetic combinations each year, while public breeders make nearly the same (102) (Bowman, 2000). Both private and public cotton breeders average around 1,000  $F_2$  plants from each cross. Private breeders look

at nearly 2,000  $F_3$  plants/cross while the median number for public breeders is 760  $F_3$  plants. In the  $F_4$  generation, the numbers are 1280 and 1036 for private and public cotton breeders, respectively. Crosses are being discarded at various stages; otherwise, the nursery may occupy 10 hectares or more.

Witcombe and Virk (2001) were not able to identify any programs, outside their own, that made few genetic combinations and looked at large population sizes per cross. They reported on low cross programs in rice (*Oryza sativa* L.), pearl millet (*Pennisetum typhoides* (B.) Stapf and Hubb) and maize (*Zea mays* L.). Baker (1984) stated that there were highly successful breeding programs, which make only 10-20 crosses annually although he gave no specifics. In cotton, Kadapa (1995) was successful in looking at few crosses but large numbers of plants (22,000) per population. Troyer (1996) provided empirical evidence of an effective program that did not make a relatively large number of genetic combinations in maize.

The approach one takes in deciding whether to make a large number of crosses and examine few progeny/cross; or a small number of crosses and examine many progeny/cross depends, in part, on several assumptions. If we assume we have information on the value of the parents, then we may make fewer crosses but look at more progeny per cross (Witcombe and Virk, 2001). Yonezawa and Yanagata (1978) calculated the relative efficiency of differing  $F_2$  population size and expected frequency of desirable genotypes. They assumed no information on the breeding value of the parents and concluded that the number of crosses should be increased at the expense of the size of the  $F_2$  population.

Baenziger and Peterson (1992) suggested that one could use performance data to predict parental value. The assumption is that the trait of interest, primarily yield, is controlled by additive effects. For self-pollinating crops, as a group, this would be true.

There are volumes of performance data on most commercial cotton cultivars but very little on germplasm lines. However, in cotton, performance of the parent is not always predictive of progeny performance. 'DES 119' and 'Deltapine 90' performed well and their progeny also performed well as evidenced by the number of cultivars developed using them as parents (Calhoun *et al.*, 1997). However, 'Stoneville 474' had shown outstanding yield performance in the 1990s, and it has not appeared in the pedigrees of any commercial cultivar other than as a recurrent parent. One private breeder (personal communication) disclosed that all progeny from crosses with Stoneville 474 were useless, and he has since discarded all material with that cultivar. This is an exception; generally better-yielding cotton cultivars end up in the pedigrees of better-yielding cotton cultivars the next generation. This would support the idea that one could be very selective in choos-

ing parents and look at more progeny per cross.

Sometimes breeders choose parents based on genetic diversity. Thus, their pedigree determines their value. However, choice of parents based on diversity as measured by coefficient of parentage may or may not be predictive. Murphy *et al.* (1986) concluded from a study of soft, red winter wheat (*Triticum vulgare*) cultivars that parents should first be selected on the basis of performance, and then on genetic distance based upon multivariate analyses of measured traits, and finally on coefficient of parentage. In soybean, (*Glycine max* L.) where there is a positive correlation between genetic distance and genetic variance, the coefficient of parentage is useful (Marjarrez-Sandoval *et al.*, 1997). However, in cotton there wasn't a linear relationship between coefficient of parentage between parents of the final cross and acceptance by the grower as evidenced by the hectareage planted (Van Esbroeck and Bowman, 1998); the coefficient of parentage ranged from 0.0 to 0.875. Thus, this knowledge may not prove useful when choosing a breeding strategy except for a few select species.

### Probabilities

For quantitative traits, the probability of producing a transgressive segregate is proportional to the number of loci involved ( $2^n$ ). Allard (1999) looked at the maximum population size required to include the desired genotype with a polygenic trait at linkage equilibrium. Theory suggests large  $F_2$  populations are needed. To have a high probability of recovering a desired genotype it would require a population equal to  $\log(\text{probability level})/\log(1-1/2^n)$ . For the 0.05 probability level this would require a population approximately three times the minimum required. As the number of loci increases, the required population size increases quickly. Snee (1977) proposed that one look at 1684  $F_2$ -derived  $F_3$  lines with each line containing 228 plants to be 98% sure of selecting a desirable genotype from a cross of two different wheat cultivars varying for yield at 21 different loci. This would result in a total population size of 383,952 plants. The reasoning for going to the  $F_3$  generation is based on the unreliability of selection of superior phenotypes in the  $F_2$  generation.

Examples of large segregating generation sizes to recover a desired genotype include cotton and sudan grass (*Sorghum bicolor* (L.) Moench). Abdel-Nabi *et al.* (1965) found only one transgressive segregate out of 1731  $F_3$  plants from a strong-fibered Acala by a weak-fibered Upland cotton cross. Burton (1951) found only one of 30,000  $F_2$  plants with disease resistance and desired plant characteristics of sudan grass. An  $F_2$  population of 3000 tomato (*Lycopersicon esculentum* Miller) plants is thought to be required to ensure a recombinant outyielding the parent (Williams, 1959).

## Heritability

Heritability of yield on an individual plant basis is extremely low for most crops; cotton is no exception. In one study by Murray and Verhalen (1969), broad-sense heritability for yield on an individual plant basis was 75% in one year and 0% the next. Heritability estimates for lint yield on a progeny-row basis, however, has ranged from 0.29 (Baker and Verhalen, 1995) to 0.66 (Miller *et al.*, 1958). It has been suggested that the breeder go to the  $F_3$  generation to begin selection since heritability would be higher due to increased homozygosity.

Baker (1984) made the argument for population sizes of 500-1000 based on probability of fixing alleles (Comstock, 1978; Bailey and Comstock, 1976), heritability on a plot basis, and expectations of finding the maximum value in a sample of varying size. Increasing the population size from 1000 to 5000 did not appreciably improve the chances of selecting the desired genotypes. He assumed a maximum of 2000 lines (crosses x lines/cross) and the goal was to select five best lines in five best crosses. His model predicted 50-100 crosses were needed. Thus, he would need to test 20-40 lines/cross. Baker (1984) also noted that as cross-numbers go up then heritability of cross means go down while the opposite is true when you decrease cross numbers. Selection intensity within crosses increase as you reduce cross numbers.

### Recombination rates

The frequency of crossing over impacts the probability of finding transgressive segregates. In *Brassica napus*, Howell *et al.* (1996) reported only an average of 11 recombinations in a map distance of 1204.5 cm; this was two thirds of the mapped genome. In pearl millet, Witcombe and Virk (2001) found from 1 to 10 recombinants when looking at 56% of the mapped genome; they concluded that large blocks of parental genetic material are inherited in the  $F_2$  generation. With low probability of recombinants occurring, large populations are needed to improve the probability of finding the desired recombinants.

If the two parents have nearly equal number of favorable alleles at different loci, then the probability of finding a transgressive segregate improves. If, on the other hand, one parent contributes the majority of the favorable alleles then there is less likelihood of finding an improved trans-segregate. It may be worthy to note that genetic maps have revealed the cotton genome to be very high for recombination. Consider that the tetraploid cotton genome is similar in size, based on DNA amount to maize, yet the genetic map for cotton is greater than 5000 cM. The genetic map of maize is slightly over 3000 cM (Brubaker *et al.*, 1999; Lacape *et al.*, 2003). Even though this is true, there is very little anecdotal evidence that this high recombination rate

translates to measurable phenotypic variation in large numbers.

## Conclusion

Arguments have been made for maximizing cross-number while sacrificing progeny number as well as judiciously choosing fewer parents and examining more progeny per cross. Theoretical probabilities of fixing alleles and expectations of finding the maximum values support the current practice of looking at 1,000  $F_2$  plants. However, with low rates of phenotypic variation, low heritability, quantitative traits, and some knowledge of parental value, I would argue that cotton breeders should make fewer crosses and examine more progeny and that selection should begin in the  $F_3$  generation. I look at 10-12,000  $F_3$  plants from each cross and make very few crosses each year. I should note that I also look for three morphological traits, which necessitate large  $F_3$  progeny number. Two traits are dominant and the third is a double recessive. Therefore, the selection pressure is 3.5% just on the morphological traits. Although I have taken this approach, it is yet to be seen if it is any more successful than the approach used by most cotton breeders.

Very little has been mentioned about the ability to recognize superior phenotypes in segregating generations. Large population sizes only help to ensure the probability that the extremes are present. Since most traits of interest are quantitative in nature, the extremes may be difficult to identify.

## References

- Allard, R.W. (1999). Principles of Plant Breeding. Second Edition. John Wiley and Sons, New York.
- Abdel-Nabi, H., Jones, J.E. and Tipton, K.W. (1965). Studies on the inheritance of fiber strength and fiber elongation in the  $F_3$  generation of a cross between two varieties of Upland cotton. In Proceedings of the Seventeenth Annual Cotton Improvement Conference, Ed. H.H. Ramey, National Cotton Council of America Memphis, TN. pp. 80-89.
- Baenziger, P.S. and Peterson, C.J. (1992). Genetic variation: Its origin and use for breeding self-pollinated species. In H.T. Stalker and J.P. Murphy (Eds.), Plant Breeding in the 1990's, pp. 69-92. CAB International Wallingford.
- Bailey, T.B. Jr. and Comstock, R.E. (1976). Linkage and the synthesis of better genotypes in self-fertilizing species. *Crop Sci.*, **16**: 363-370.
- Baker, J.L. and Verhalen, L.M. (1975). Heterosis and combining ability for several agronomic and fiber properties among selected lines of upland cotton. *Cott. Grow. Res.*, **52**: 209-223.
- Baker, R.J. (1984). Quantitative genetic principles in plant breeding. In J.P. Gustafson, ed. Gene Manipulation in Plant Improvement. Plenum Press, New York.
- Bingham, J. and Lupton, F.G.H. (1987). Production of new varieties: an integrated research approach to plant breeding. In F.G.H. Lupton (ed.) Wheat Breeding. pp. 487-537. Chapman and Hall, London.
- Bowman, D.T. (2000). Attributes of public and private cotton breeding programs. *J. Cot. Sci.*, **4**: 1-7.
- Brubaker, C.L., Paterson, A.H., and Wendel, J.F. (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome*, **42**: 184-203.
- Burton, G.W. (1951). The adaptability and breeding of suitable grasses for the Southeastern states. *Adv. Agron.*, **3**: 197-241.
- Calhoun, D.S., Bowman, D.T. and May, O.L. (1997). Pedigree of upland and pima cotton cultivars released between 1970 and 1995. Bull. 1069. Miss. Agric. and For. Exp. Stn. Mississippi State, MS.
- Comstock, R.E. (1978). Quantitative genetics in maize breeding. pp. 191-206 In Maize breeding and genetics. D.B. Walden, ed. John Wiley and Sons, New York.
- Howell, P.M., Marshall, D.F. and Lydiate, D.J. (1996). Towards developing intervarietal substitution lines in Brassica napus using marker-assisted selection. *Genome*, **39**: 348-358.
- IRR1, (1985). Parentage of IRR1 crosses IR1-IR 50,000. The International Rice Research Institute (IRRI), Los Banos, Philippines.
- Kadapa, S.N. (1995). Composite-crossing methodology pays rich dividend in cotton breeding – a resume. *Indian J. Genet.*, **55**: 290-301.
- Lacape, J.M., Nguyen, T.B., Thibivilliers, S., Bojinov, B., Courtois, B., Cantrell, R.G., Burr, B., and Hau, B. (2003). A combined RFLP-SSS-SFLP map of tetraploid cotton based on a *Gossypium hirsutum* x *Gossypium barbadense* backcross population. *Genome*, **46**: 612-626.
- Manjarrez-Sandoval, P. Carter, T.E. Jr., Webb, D.M. and Burton, J.W. (1997). RFLP, genetic similarity estimates, and coefficient of parentage as genetic variance predictors for soybean yield. *Crop Sci.*, **37**: 698-703.
- Maredia, M.K. and Byerlee, D. (1999). The global wheat improvement system: prospects for enhancing efficiency in the presence of spillovers. CIMMYT Research Report No. 5. International Centre for Wheat and Maize Improvement, Mexico, D.F.
- Miller, P.A., Williams, J.C., Robinson, H.F. and Comstock, R.E. (1958). Estimates of genotypic and environmental variances and covariances in upland cotton and their implications in selection. *Agron. J.*, **50**: 126-131.
- Murphy, J.P., Cox, T.S. and Rodgers, D.M. (1986). Cluster analysis of red winter wheat cultivars based upon coefficient of parentage. *Crop Sci.*, **26**: 672-676.
- Murray, J.C. and Verhalen, L.M. (1969). Genetics of earliness, yield, and fiber properties in cotton (*Gossypium hirsutum* L.). *Crop Sci.*, **9**: 752-755.

- Sneep, J. (1977). Selection for yield in early generations for self-fertilizing crops. *Euphytica*, **26**: 27-30.
- Troyer, A.F. (1996). Breeding widely adapted, popular maize hybrids. *Euphytica*, **92**: 163-174.
- Van Esbroeck, G.A. and Bowman, D.T. (1998). Cotton germplasm diversity and its importance to cultivar development. *J. Cot. Sci.*, **2**: 121-129.
- Williams, W. (1959). Isolation of "pure lines" from F<sub>1</sub> hybrids of tomato and the problem of heterosis in inbreeding crop species. *J. Agric. Sci.*, **53**: 347-353.
- Witcombe, J.R. and Virk, D.S. (2001). Number of crosses and population size for participatory and classical plant breeding. *Euphytica*, **122**: 451-462.