The Need for Rapid Assays of Protein Quality

M. W. Adams

Genetic bases of quality may range from simple to complex, necessitating in the more complex cases the screening of very large numbers of lines in order to have a reasonable chance of identifying desirable genetic combinations. Because of the large numbers involved, and the brief analytical time available between successive generations, breeders are generally more interested in relative precision and quick results rather than a high degree of accuracy at the expense of time. Quality tests conducted in early segregating gen-erations should require only a few grams of material since the amount available in early generations is limited and other tests may also have to be conducted with the same populations. In summary, plant breeders concerned with the improvement of protein quality would like rapid, simple, inexpensive tests of quality components that are reasonably accurate, that correlate well with biological value, and that consume only small amounts of material.

With the belated general recognition that genes play a powerful role in determining nutritive potential of storage proteins in food crops, plant breeders around the world have rushed to fill the protein void by selecting or building genotypes that in some way result in storage proteins of enhanced nutritional value. Following the first flush of success from the Op-2 system in maize, there has come the sobering reassessment that in the other cereal grains, the root crops, and in the grain legumes instant success via a major gene like Op-2 may be expecting too much of nature. Such genetic variation in protein quality factors as has been found—and it has been found when large numbers of varieties or strains have been examined—suggests continuous variability of the kind determined by numerous genes of individually small effects. Success in producing a high-protein quality variety would, on this assumption, depend on accumulating in a given genotype all the favorable polygenes affecting quality.

THE TASK

If this were the case genetically, it would be necessary to examine, with some quality assay system, a very large number of lines over several generations. This could be a slow tedious job.

The plant breeder, therefore, asks of the protein quality analyst that there be devised *rapid, simple, inexpensive* tests of quality components that *are reasonably accurate and that require only a few grams of material.*

JUSTIFICATION

The plant breeder, whose goals include genetic improvement of protein quality, is confronted particularly by four technical problems. The purpose of this paper is to examine these problems as they relate to the needs of an assay system.

(a) The Genetic Problem. This results from the low expected frequency of recovery in segregating generations of genotypes that combine genes favoring protein quality along with the numerous genes necessary for disease resistance, agronomic performance, and consumer acceptance.

Although it was suggested in the introductory statement that several genes might be involved in determining protein quality, in most cases we simply do not know that number, nor do we have in most cases more than a rough idea of whether the genes act additively or show recessivity, dominance, or interaction in their effects.

In the simplest possible case, one gene pair in the recessive state, the breeder can expect to recover the desired genotype in the segregating generation (F_2) in one-fourth of the plants of that generation. But to recover this type one time with a 99% probability of not missing it when it is in fact present requires 16 plants, not four. The precision of the test itself may require that more plants be screened than indicated here.

Let us examine the outcome in a slightly more realistic situation. Suppose as many as five independent gene pairs in the recessive state were required to produce the level of protein quality desired. The expected rate of recovery in an F_2 generation or an unselected F_3 is approximately one out of a thousand. To be assured of success in detecting the desired genotype, with a probability of 99% against failure, the analyst would have to assay some 5000 plants (calculated 4713). If in the same population the breeder is also selecting for other performance characteristics, he requires a very large population indeed. In such a case he would select for the easily recognized traits first, and seek to reduce the number of plants or progenies to a manageable level before attempting the quality analyses.

Quite possibly, however, the breeder would choose the backcross system of varietal improvement, wherein the recurrent parent is chosen to carry the genes required for disease resistance, agronomic performance, and consumer properties. The nonrecurrent parent is chosen as a source from which to introduce the protein-quality genes. The breeder depends upon the recurrent introduction of genes carried by the recurrent parent in the crosses so that they are expected to be recovered almost automatically after several generations of backcrossing. The genes for quality must be selected for in each cycle. Here it makes an important difference whether the genes are dominant, additive, or recessive. If they show additivity or dominance, screening may be done in each backcross generation, and if the genes are also nonlinked with respect to each other, the expected rate of recovery depends upon the value $(\frac{1}{2})^n$, where n = the number of gene pairs involved. For 5 gene pairs, some 150 plants (calculated 145) in each backcross generation would have to be assayed to be assured at the 99% level of probability that the desired genotype had not been missed. A population of this size is manageable.

If the relevant genes are recessive, then a generation of self-fertilization must be interposed after each backcross to expose the recessive genes so that their effects on quality can be detected. But the probability of recovery of each favorable gene pair now is $\frac{1}{8}$, and for five gene pairs it is $(\frac{1}{8})^5$. At the 99% level of assurance of having not missed the best combination, some 150,000 plants (calculated 150,909) would have to be assayed, a considerable task.

The number, X, of plants in a segregating population that must be assayed is calculable from the formula $X = \log (1 - pd)/\log (1 - pr)$, where pd = the probability of detection, set at 0.95 or 0.99, and pr = the expected probability of recovery of the desired combination in a particular segregating population. The value is based on genetic expectations.

In those cases where the individual genes affecting protein quality cannot be clearly identified, breeders usually fall back upon another interpretation, that of polygenic

Crop and Soil Sciences Department, Michigan State University, East Lansing, Michigan 48823.

inheritance. In these cases, variation in quality factors is assumed to be a joint consequence of numerous genes, which individually have small and additive effects upon the quality of protein, and of environmental forces. Upon statistical treatment of data from genetic families varying in quality, it is possible to partition the variation into a portion due to additively acting genes, and a residual portion attributable to more complex gene effects and to effects of the environment. This ratio of the additive genetic portion to the total variability is termed *heritability*, symbolized h^2. Heritability values range from 0 to 100%. When h^2 is high, individual selection is more effective than when h^2 is low, but while gains from selection are predictable, the breeder generally cannot predict with any accuracy the expected frequency of recovery of the favorable genotype. The recourse open to the breeder is recurrent selection, in which, in whatever sized population the breeder can handle, selection of the better or best individuals is made, these serving as parents to be intercrossed to produce the next generation. This process is repeated over several cycles until the desired level of quality is attained.

The numbers of plants to be assayed, still very large in the long view, are relatively small on a cycle basis, and allow steady progress toward improved quality to be made without calling for a crash program involving immense resources and effort at any one time.

I think I have presented sufficient alternative examples to make clear that the number of plants that must be quality assayed can vary greatly depending upon the genetic situation, but that the number in the more realistic cases could be very large. Breeders have ways, in the recurrent selection system, of improving the odds of recovery over several generations, so breeding for higher quality protein is still seen as a feasible activity.

(b) Problems of Time and Accuracy. In breeding for quality factors we are generally more interested in relative precision and quick results from tests than we are in absolute accuracy at the expense of time. Let us consider a hypothetical but typical example in winter wheat as grown in the north central United States. There is a 6–7-week period between final harvest in late July and seeding time in mid-September. During that interval, harvest samples must be recleaned, evaluated for milling and baking characteristics, and, if protein quality is a goal, also assayed for one or more protein quality factors and/or biological value, selections made, and seed prepared for planting. If 3 days are taken up in running 40 samples through the amino acid analyzer and calculating results, 1 day for determining particle size index on 40 samples, another day for measuring gluten strength, and a day to do the Kjeldahl nitrogens, it is estimated that some 250 lines can be screened to this extent by one technician in the time span allowable. Unless the frequency of desirable genotypes is already fairly high, a breeding program for protein quality will not make worthwhile progress under these conditions.

Even with only 5 gene pairs segregating in a particular F_2 population, some 3000 lines (calculated 2922) would have to be assayed to recover the preferred combination one time, with a 95% level of probability of detection. With the schedule suggested above—250 samples screened by one technician—the equivalent of 12 technicians working 6 days a week would be required to complete the screening of 3000 entries in 7 weeks.

Furthermore, the breeder ordinarily will select not just the top line in a population but several lines at the upper end of the curve of variation. These will be grown another year in one or more locations and retested for quality components. Absolute accuracy, then, is not required in the initial screening. What is required in an assay method is enough relative precision to distinguish lines of differing genetic potential.

The requirement for rapidity in the assay is a relative one also. When the time interval between harvest and

planting is short, then speed of the assay is critical. For annual crops planted in the spring, there is a 6-8-month time interval available for assay, and speed of the individual assay is less important.

Rapidity of evaluation is also tied in with how much biological relevance the breeder demands of the test. An automated amino acid analysis requiring 4 days for 40 samples is obviously more rapid than a rat or swine test lasting 28 days, but biological relevance of the outcome data is clearly superior in the animal tests. Bacterial bioassays requiring 2 days and the meadow vole test requiring a 5-day feeding period appear to be valuable compromises between the contradictory demands for both speed and relevance.

For logistic reasons—time, number of lines to be assayed, amount of material needed, expense, and relevance—the breeder most likely will resort to the chemical or bacterial methods in the early stages of a breeding program, when the number of lines is high and grams of seed per line are limited, and will turn to the actual feeding trials at a later stage when only definitely promising lines need to be evaluated.

(c) Instability of a Quality Factor over Time or Location. There is mounting evidence that, just as yields of grain and total nitrogen show strain by environment interaction, the level of certain quality factors also varies among varieties grown at different locations or in different seasons. Quality testing, therefore, should be conducted on material produced in different seasons or locations, before final quality selections are made. This only compounds the requirements for rapidity in assaying. And the assay must be specific for particular quality components, not general, since in the latter case, selection applied to a population might be directed toward one constellation of genes in one season and toward a different set of genes the next.

(d) Characteristics and Requirements of the Quality Test Itself. The time required per sample and the relative precision of the assay have already been mentioned. The choices made on these points will probably be reflected in the amount of material needed per test and cost of the procedure. But these two factors are important in their own right. Breeders, as a general rule, never have enough material per genetic entity to carry out all the testing they think desirable. This applies particularly to the early generations of selection. Consequently, they favor test procedures that consume the least number of seeds.

Of course, breeders also want an inexpensive test. Some of the instrumentation in the analytical laboratory is expensive to purchase, but because of reliability and accuracy with proper maintenance, thousands of samples can be analyzed at a relatively low unit cost. But for the assays based on small animals, the per unit cost is usually much higher. For this reason, the requirement of much more test material per animal, and the extra time factors, breeders will employ the small animal tests only in the final stages of a selection program.

CONCLUSION

It appears, then, that the breeder, at successive stages in his quality improvement program, will want to employ several different assay systems, varying in complexity, precision, seed requirements, and cost.

The most immediate task, and possibly the most formidable, is to design a test system that is rapid, inexpensive, and reasonably accurate but not necessarily highly precise, that correlates well with biological value, and that can be applied to small quantities of seeds of large populations.